MATHEMATICAL PROGRAMMING STUDIES

MATHEMATICAL PROGRAMMING STUDY 27

A PUBLICATION OF
THE MATHEMATICAL PROGRAMMING SOCIETY

Stochastic Programming 84
Part I

Edited by A. PRÉKOPA and R.J.-B. WETS

April 1986

NORTH-HOLLAND – AMSTERDAM

This STUDY is also available to nonsubscribers in a book edition.

Printed in The Netherlands

PREFACE

One may think of stochastic programming as simply a subfield of nonlinear programming. The fact that the objective function or some of the constraints are expressed in terms of multidimensional integrals does not change the essence of the problem, at least in theory. However, it is precisely because the problem at hand demands the evaluation of multidimensional integrals that its nature is altered in a fundamental way from a numerical viewpoint.

Let us consider the following type of problem

$$\text{find } x \in S \subset \mathbb{R} \text{ that minimizes } F(x)$$

where $F(x) = E\{f(x, \omega)\} = \int f(x, \omega)\, dP(\omega)$, and S is a closed set determined by some constraints that could be of probabilistic type. For simplicity, let us consider the case in which only the objective is given by a multidimensional integral with respect to the probability measure P. Because the operator E has a 'smoothing' effect, this optimization problem usually has many desirable properties. For example, if the integrand f is convex in x, then so is F. If f is differentiable with respect to x or the measure P is absolutely continuous, then it usually turns out that F is also differentiable. Thus, in principle the problem could be solved by relying on some of the existing subroutines for nonlinear programming problems; all that is needed is to appeal to a multidimensional integration subroutine to evaluate the function F, its gradients, or subgradients, as the case may be. However, general purpose integration subroutines are available only for the 1-dimensional case. In 2-dimensions some serious difficulties already must be dealt with, and in 3-dimensions subroutines are available only for very special cases. Typically, a stochastic programming problem involves anywhere from 5 to 100 random variables, making it totally impossible to rely on existing numerical integration subroutines. Naturally, some efforts have been made to design multidimensional integration subroutines—and some of the papers in this collection report the progress made in that direction—but essentially they rely on sampling techniques (involving the generation of pseudo- or quasi-random numbers). This presupposes that the integrand is sufficiently easy to evaluate, and for stochastic programming models that is the exception, not the rule. The integrand $f(x, \omega)$ is often defined implicitly, for example as the optimal value of an optimization problem. Thus efficient procedures must avoid numerous evaluations of the integrand.

Although the search for reliable multidimensional integration subroutines has not been abandoned, the design of solution procedures for stochastic optimization problems has been chiefly oriented toward methods that in one way or another avoid coming to grips with this potential stumbling block. Excluding some very

specific classes of stochastic programming problems, the suggested solution strategies can be divided into two major categories:

- 'descent' methods that rely on directions determined by statistical estimates of the subgradients of F, and
- approximation methods that replace either the original distribution P by one that would be more manageable or the integrand f by a 'simpler' one that would make it possible to carry out the multidimensional integration.

All these possibilities are illustrated in the articles of this collection.

In Volume 1, the first three articles deal with evaluating multidimensional integrals as they arise in stochastic programming (Szantai, Niederreiter) or obtaining bounds for them (Gassman/Ziemba). The next group of three articles deal with approximation schemes. We start with a review of the existing results as well as some suggestions for implementation (Birge/Wets). Approximating by problem redefinition is illustrated in the article by Beale, Dantzig and Watson, whereas Andreatta and Runggaldier work by approximating the probability measure. Intimately related to approximation is the question of the stability of the problem under various perturbations, in particular perturbations of the probability distribution function. This is the subject of the contributions of Dupačová and Wang.

The remaining articles deal with specific procedures for solving particular or general stochastic programming problems. In Volume 1, the articles by Klein Haneveld and Qi deal with stochastic network problems. The structure of the problems plays a very important role in the procedures they suggest. In Volume 2, the first four articles deal with stochastic programs with recourse models (Nazareth/Wets, Wallace, Louveaux, Rockafellar/Wets). Next, Komaromi suggests a new dual-based procedure for solving problems with probabilistic constraints. The last three articles introduce modifications of the stochastic-gradient method to make the calculations of the step size more directly adaptive (Ruszczynski), to include nonstochastic descent information (Marti/Fuchs), and to allow for its application in the case where the decision variables themselves are probability measures (Gaivoronski).

The decision to submit these contributions in the form of a Study was made at the first meeting of COSP (Committee on Stochastic Programming) on December 1, 1983, at IIASA (International Institute for Applied Systems Analysis), Laxenburg, Austria. These two volumes could very well serve as commemorative issues to mark that occasion.

Andras Prékopa
Roger J.-B. Wets

CONTENTS

PART I
(this volume)

PART II
(Mathematical Programming Study 28)

Mathematical Programming Study 27 (1986) 1-16
North-Holland

EVALUATION OF A SPECIAL MULTIVARIATE GAMMA DISTRIBUTION FUNCTION

T. SZÁNTAI
L. Eötvös University of Budapest, Budapest, Hungary

Received 1 April 1984
Revised manuscript received 25 February 1985

In this paper we describe two different methods for the calculation of the bivariate gamma probability distribution function. One of them is based on a direct numerical integration and the other on a series expansion in terms of Laguerre polynomials. In the multivariate case we propose a Monte Carlo method. Our method can be used for other types of multivariate probability distributions too. In the special case of the multivariate normal distribution the computer experiments show that our method has the same efficiency as other known methods. In the last paragraph we briefly describe the possible applications of the proposed algorithms in stochastic programming.

Key words: Multivariate Probability Distribution, Multivariate Integration, Simulation, Stochastic Programming.

1. Introduction

A. Prékopa and T. Szántai introduced a special multivariate gamma distribution [10]. This is the probability distribution of the random vector

$$x = Ay \tag{1.1}$$

where the random vector y has independent, standard gamma distributed components and A is a matrix of 0 and 1 entries. A continuous probability distribution is called standard gamma distribution if it has the following type of probability density:

$$\frac{1}{\Gamma(\vartheta)} z^{\vartheta-1} \mathrm{e}^{-z} \quad \text{if } z > 0$$

and zero if $z \leq 0$; ϑ is a positive constant. The matrix A consists of all possible nonzero column vectors having components 0 and 1.

The fitting of this distribution to empirical data means that we choose the parameter values ϑ belonging to the components of the random vector y in such a way that the sampling expectations and sampling covariances of the random vector x coincide with those corresponding to the random vector Ay. In this construction many components of y will have a degenerated standard gamma distribution in practice, i.e. many of the parameter values ϑ will be equal to zero. In this paper we do not deal with the problem of fitting the above-mentioned multivariate gamma distribution to empirical data but we give an algorithm for the calculation of the values of the probability distribution function.

First we describe two different methods for the calculation of the bivariate gamma probability distribution function. These are then used to produce good lower and upper bounds for the same function. Finally we develop a Monte Carlo simulation algorithm for the evaluation of the multivariate gamma probability distribution function. Results of computer experiments are described in Section 5.

Our algorithm can be used for other types of multivariate distributions as well. We only need an efficient procedure for calculating the values of the univariate and bivariate marginal distribution functions. In fact we also have computational experience for the case of the multivariate normal distribution. The results show that in this case our algorithm has the same efficiency as other known algorithms (see e.g. Deák [2]).

2. Evaluation of the bivariate gamma probability distribution function

Every bivariate marginal distribution of the multigamma distribution introduced by Prékopa and Szántai [10] has the following structure:

$$\begin{aligned} x_1 &= y_1 + y_2 \ , \\ x_2 &= y_1 \quad\ + y_3 \end{aligned} \tag{2.1}$$

where y_1, y_2 and y_3 have independent standard gamma distributions with parameters ϑ_1, ϑ_2 and ϑ_3, respectively.

We want to calculate the probability

$$P(x_1 < z_1, x_2 < z_2) \tag{2.2}$$

for all nonnegative real values z_1 and z_2.

The first method to calculate (2.2) is based on the simple fact that when conditioning on the value of y_1, the random variables x_1 and x_2 become independent. So we have

$$\begin{aligned} F(z_1, z_2) &= P(x_1 < z_1, x_2 < z_2) = P(y_1 + y_2 < z_1, y_1 + y_3 < z_2) \\ &= \int_0^\infty P(y_1 + y_2 < z_1, y_1 + y_3 < z_2 | y_1 = y) f_{\vartheta_1}(y)\, \mathrm{d}y \\ &= \int_0^\infty P(y_2 < z_1 - y, y_3 < z_2 - y | y_1 = y) f_{\vartheta_1}(y)\, \mathrm{d}y \\ &= \int_0^{\min(z_1, z_2)} F_{\vartheta_2}(z_1 - y) F_{\vartheta_3}(z_2 - y) f_{\vartheta_1}(y)\, \mathrm{d}y, \end{aligned} \tag{2.3}$$

where

$$f_\vartheta(x) = \frac{1}{\Gamma(\vartheta)} x^{\vartheta - 1} \mathrm{e}^{-x}, \quad \text{if } x > 0, \qquad F_\vartheta(x) = \frac{1}{\Gamma(\vartheta)} \int_0^x t^{\vartheta - 1} \mathrm{e}^{-t}\, \mathrm{d}t.$$

Thus for the calculation of (2.2) we only need to perform a one dimensional numerical integration and to calculate the values of the complete and incomplete gamma functions. Most computers have good subroutines for these calculations. However, there are difficulties if the parameter ϑ_1 has a value near zero because in this case the numerical integration becomes instable.

The second, more efficient method uses the Laguerre polynomial expansion of the joint probability density function of the random variables x_1, x_2 given by (2.1). This expansion can be derived by the application of the inverse Laplace transformation on the joint characteristic function of the random variables x_1, x_2. The joint characteristic function of x_1 and x_2 is

$$\begin{aligned}\varphi(t_1, t_2) &= E(e^{it_1x_1+it_2x_2}) = E(e^{it_1y_1+it_1y_2+it_2y_1+it_2y_3}) \\ &= E(e^{i(t_1+t_2)y_1})E(e^{it_1y_2})E(e^{it_2y_3}) \\ &= \frac{1}{(1-i(t_1+t_2))^{\vartheta_1}} \frac{1}{(1-it_1)^{\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_3}}.\end{aligned}$$

If we put this joint characteristic function into the formula of the inverse Laplace transformation

$$f(z_1, z_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-it_1z_1} e^{-it_2z_2} \varphi(t_1, t_2)\, dt_1\, dt_2$$

we get

$$\begin{aligned}f(z_1, z_2) &= \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-it_1z_1} e^{-it_2z_2} \frac{1}{(1-i(t_1+t_2))^{\vartheta_1}} \\ &\quad \times \frac{1}{(1-it_1)^{\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_3}} dt_1\, dt_2 \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-it_1z_1} e^{-it_2z_2} \frac{1}{(1-it_1)^{\vartheta_1+\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_1+\vartheta_3}} \\ &\quad \times \left[1 - \frac{it_1it_2}{(1-it_1)(1-it_2)}\right]^{\vartheta_1} dt_1\, dt_2.\end{aligned}$$

Applying here the binomial expansion we obtain

$$\begin{aligned}f(z_1, z_2) &= \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-it_1z_1} e^{-it_2z_2} (1-it_1)^{-(\vartheta_1+\vartheta_2)} (1-it_2)^{-(\vartheta_1+\vartheta_3)} \\ &\quad \times \left[1 + \vartheta_1 \frac{it_1it_2}{(1-it_1)(1-it_2)} + \vartheta_1(\vartheta_1+1)\frac{1}{2!}\right. \\ &\quad \left. \times \frac{(it_1)^2(it_2)^2}{(1-it_1)^2(1-it_2)^2} + \cdots\right] dt_1\, dt_2. \qquad (2.4)\end{aligned}$$

It is known (from the theory of Laplace transforms) that

$$f_\vartheta(z)=\frac{1}{\Gamma(\vartheta)}z^{\vartheta-1}\,\mathrm{e}^{-z}=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}tz}(1-\mathrm{i}t)^{-\vartheta}\,\mathrm{d}t,$$

$$f_{\vartheta+r}(z)=\frac{1}{\Gamma(\vartheta+r)}z^{\vartheta+r-1}\,\mathrm{e}^{-z}=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}tz}(1-\mathrm{i}t)^{-\vartheta-r}\,\mathrm{d}t,\quad R=0,1,\ldots,$$

$$\frac{\mathrm{d}^r f_{\vartheta+r}(z)}{\mathrm{d}z^r}=\frac{1}{\Gamma(\vartheta+r)}\frac{\mathrm{d}^r}{\mathrm{d}z^r}(z^{\vartheta+r-1}\,\mathrm{E}^{-z})$$

$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}(-1)^r(\mathrm{i}t)^r\,\mathrm{e}^{-\mathrm{i}tz}(1-\mathrm{i}t)^{-\vartheta-r}\,\mathrm{d}t,\quad r=0,1,\ldots.$$

Thus from (2.4) we get

$$f(z_1,z_2)=f_{\vartheta_1+\vartheta_2}(z_1)f_{\vartheta_1+\vartheta_3}(z_2)+\vartheta_1\frac{\mathrm{d}f_{\vartheta_1+\vartheta_2+1}(z_1)}{\mathrm{d}z_1}\frac{\mathrm{d}f_{\vartheta_1+\vartheta_3+1}(z_2)}{\mathrm{d}z_2}$$

$$+\frac{\vartheta_1(\vartheta_1+1)}{2!}\frac{\mathrm{d}^2f_{\vartheta_1+\vartheta_2+2}(z_1)}{\mathrm{d}z_1^2}\frac{\mathrm{d}^2f_{\vartheta_1+\vartheta_3+2}(z_2)}{\mathrm{d}z_2^2}+\cdots. \tag{2.5}$$

Now by the definition of the Laguerre polynomials (see e.g. [5, 10.12(5)])

$$\frac{\mathrm{d}^r f_{\vartheta+r}(z)}{\mathrm{d}z^r}=\frac{1}{\Gamma(\vartheta+r)}\frac{\mathrm{d}^r}{\mathrm{d}z^r}(z^{\vartheta+r-1}\,\mathrm{e}^{-z})$$

$$=\frac{1}{\Gamma(\vartheta+r)}r!z^{\vartheta-1}\,\mathrm{e}^{-z}L_r^{\vartheta-1}(z)$$

$$=\frac{r!}{\Gamma(\vartheta+r)}\Gamma(\vartheta)f_\vartheta(z)L_r^{\vartheta-1}(z),\quad r=0,1,\ldots,$$

and from (2.5) we can obtain the final form of the joint probability density function of the random variables x_1 and x_2:

$$f(z_1,z_2)=f_{\vartheta_1+\vartheta_2}(z_1)f_{\vartheta_1+\vartheta_3}(z_2)\Bigg\{1+\sum_{r=1}^{\infty}r!$$

$$\times\frac{\Gamma(\vartheta_1+r)}{\Gamma(\vartheta_1)}\frac{\Gamma(\vartheta_1+\vartheta_2)}{\Gamma(\vartheta_1+\vartheta_2+r)}\frac{\Gamma(\vartheta_1+\vartheta_3)}{\Gamma(\vartheta_1+\vartheta_3+r)}L_r^{\vartheta_1+\vartheta_2-1}(z_1)L_r^{\vartheta_1+\vartheta_3-1}(z_2)\Bigg\}. \tag{2.6}$$

This expansion is well known in the literature (see Eagleson [4], Lancaster [9], Sarmanov [11] and Diachenko [3]).

Now we can determine the joint probability distribution function of the random variables x_1, x_2. We know that the Laguerre polynomials satisfy the following differentiation formula (see e.g. [5, 10.12(28)] for $a=\vartheta$, $m=1$ and $n=r-1$)

$$\frac{\mathrm{d}}{\mathrm{d}x}[x^\vartheta\,\mathrm{e}^{-x}L_{r-1}^\vartheta(x)]=rx^{\vartheta-1}\,\mathrm{e}^{-x}L_r^{\vartheta-1}(x).$$

From this by integration we get

$$x^{\vartheta} e^{-x} L_{r-1}^{\vartheta}(x) = r \int_0^x t^{\vartheta-1} e^{-t} L_r^{\vartheta-1}(t)\, dt,$$

i.e.,

$$\int_0^x f_{\vartheta}(t) L_r^{\vartheta-1}(t)\, dt = \frac{1}{r} \frac{\Gamma(\vartheta+1)}{\Gamma(\vartheta)} f_{\vartheta+1}(x) L_{r-1}^{\vartheta}(x). \tag{2.7}$$

Using (2.6) and (2.7) we obtain the joint probability distribution function in the following manner:

$$F(z_1, z_2) = \int_0^{z_2} \int_0^{z_1} f(t_1, t_2)\, dt_1\, dt_2 = F_{\vartheta_1+\vartheta_2}(z_1) F_{\vartheta_1+\vartheta_3}(z_2))$$
$$+ \sum_{r=1}^{\infty} C(\vartheta_1, \vartheta_2, \vartheta_3, r) f_{\vartheta_1+\vartheta_2+1}(z_1) L_{r-1}^{\vartheta_1+\vartheta_2}(z_1) f_{\vartheta_1+\vartheta_2+1}(z_2) L_{r-1}^{\vartheta_1+\vartheta_3}(z_2), \tag{2.8}$$

where

$$c(\vartheta_1, \vartheta_2, \vartheta_3, r) = \frac{(r-1)!}{r} \frac{\Gamma(\vartheta_1+r)}{\Gamma(\vartheta_1)} \frac{\Gamma(\vartheta_1+\vartheta_2+1)}{\Gamma(\vartheta_1+\vartheta_2+r)} \frac{\Gamma(\vartheta_1+\vartheta_3+1)}{\Gamma(\vartheta_1+\vartheta_3+r)}$$
$$= \frac{(r-1)!}{r} \frac{(\vartheta_1+r-1)\cdots\vartheta_1}{(\vartheta_1+\vartheta_2+r-1)\cdots(\vartheta_1+\vartheta_2+1)(\vartheta_1+\vartheta_3+r-1)\cdots(\vartheta_1+\vartheta_3+1)}.$$

The Laguerre polynomials involved in (2.8) can be calculated by the recursive formula (see e.g. [5, 10.12(8)])

$$(r+1) L_{r+1}^{\vartheta}(x) - (2r+\vartheta+1-x) L_r^{\vartheta}(x) + (r+\vartheta) L_{r-1}^{\vartheta}(x) = 0, \quad r = 1, 2, \ldots$$

where

$$L_0^{\vartheta}(x) = 1, \qquad L_1^{\vartheta}(x) = \vartheta + 1 - x.$$

Regarding the convergence of the infinite series in (2.8) it is not difficult to check the following estimation:

$$C(\vartheta_1, \vartheta_2, \vartheta_3, r) = \frac{(r-1)!}{r}$$
$$\times \frac{(\vartheta_1+r-1)\cdots\vartheta_1}{(\vartheta_1+\vartheta_2+r-1)\cdots(\vartheta_1+\vartheta_2+1)(\vartheta_1+\vartheta_3+r-1)\cdots(\vartheta_1+\vartheta_3+1)}$$
$$= \frac{\vartheta_1+\vartheta_3}{r} \frac{(r-1)\cdots 1}{(\vartheta_1+\vartheta_2+r-1)\cdots(\vartheta_1+\vartheta_2+1)}$$
$$\times \frac{(\vartheta_1+r-1)\cdots\vartheta_1}{(\vartheta_1+\vartheta_3+r-1)\cdots(\vartheta_1+\vartheta_3)}$$

$$= \frac{\vartheta_1+\vartheta_3}{r}\left(1-\frac{\vartheta_1+\vartheta_2}{\vartheta_1+\vartheta_2+r-1}\right)\cdots\left(1-\frac{\vartheta_1+\vartheta_2}{\vartheta_1+\vartheta_2+1}\right)$$

$$\times\left(1-\frac{\vartheta_3}{\vartheta_1+\vartheta_3+r-1}\right)\cdots\left(1-\frac{\vartheta_3}{\vartheta_1+\vartheta_3}\right)$$

$$\leqslant \frac{\vartheta_1+\vartheta_3}{r}\prod_{j=1}^{r-1}\exp\left(-\frac{\vartheta_1+\vartheta_2}{\vartheta_1+\vartheta_2+j}\right)\prod_{j=0}^{r-1}\exp\left(-\frac{\vartheta_3}{\vartheta_1+\vartheta_3+j}\right)$$

$$= \frac{\vartheta_1+\vartheta_3}{r}\exp\left(-(\vartheta_1+\vartheta_2)\sum_{j=1}^{r-1}\frac{1}{\vartheta_1+\vartheta_2+j}\right)\exp\left(-\vartheta_3\sum_{j=0}^{r-1}\frac{1}{\vartheta_1+\vartheta_3+j}\right)$$

$$\leqslant \frac{\vartheta_1+\vartheta_3}{r}\exp(-(\vartheta_1+\vartheta_2)(\ln(\vartheta_1+\vartheta_2+r)-\ln(\vartheta_1+\vartheta_2+1))$$

$$\times\exp(-\vartheta_3(\ln(\vartheta_1+\vartheta_3+r)-\ln(\vartheta_1+\vartheta_3))$$

$$= \frac{\vartheta_1+\vartheta_3}{r}\left(\frac{\vartheta_1+\vartheta_2+1}{\vartheta_1+\vartheta_2+r}\right)^{\vartheta_1+\vartheta_2}\left(\frac{\vartheta_1+\vartheta_3}{\vartheta_1+\vartheta_3+r}\right)^{\vartheta_3}$$

$$= (\vartheta_1+\vartheta_3)^{\vartheta_3+1}(\vartheta_1+\vartheta_2+1)^{\vartheta_1+\vartheta_2}\frac{1}{r(\vartheta_1+\vartheta_2+r)^{\vartheta_1+\vartheta_2}}\frac{1}{(\vartheta_1+\vartheta_3+r)^{\vartheta_3}}$$

$$< (\vartheta_1+\vartheta_3)^{\vartheta_3+1}(\vartheta_1+\vartheta_2+1)^{\vartheta_1+\vartheta_2}\frac{1}{(r-1)^{1+\vartheta_1+\vartheta_2+\vartheta_3}}, \quad r=2,3,\ldots. \tag{2.9}$$

For the asymptotic behaviour of the Laguerre polynomials we have Fejér's formula (see e.g. [5, 10.15(1)])

$$L_r^{\vartheta}(x) = \Pi^{-1/2}\,e^{(1/2)x}x^{-(1/2)\vartheta-1/4}r^{(1/2)\vartheta-1/4}\cos(2(rx)^{1/2}-\tfrac{1}{2}\vartheta\pi-\tfrac{1}{4}\pi)$$
$$+O(r^{(1/2)\vartheta-3/4})$$

for real ϑ and fixed $x>0$ or uniformly in $0<\varepsilon\leqslant x\leqslant\omega<\infty$. Thus we have the estimation

$$L_{r-1}^{\vartheta_1+\vartheta_2}(z_1)L_{r-1}^{\vartheta_1+\vartheta_3}(z_2) \leqslant K(r-1)^{(1/2)(\vartheta_1+\vartheta_2)-1/4+(1/2)(\vartheta_1+\vartheta_3)-1/4}$$
$$= K(r-1)^{\vartheta_1+(1/2)(\vartheta_2+\vartheta_3-1)}, \quad r=2,3,\ldots, \tag{2.10}$$

where K is a constant that depends on the arguments z_1, z_2 and the parameters ϑ_1, ϑ_2, ϑ_3.

Now using (2.9) and (2.10) we finally obtain:

$$\sum_{r=1}^{\infty} C(\vartheta_1,\vartheta_2,\vartheta_3,r)f_{\vartheta_1-\vartheta_2+1}(z_1)L_{r-1}^{\vartheta_1+\vartheta_2}(z_1)f_{\vartheta_1+\vartheta_3+1}(z_2)L_{r-1}^{\vartheta_1+\vartheta_3}(z_2)$$

$$\leqslant \vartheta_1 f_{\vartheta_1+\vartheta_2+1}(z_1)f_{\vartheta_1\vartheta_3+1}(z_2)+K'\sum_{r=2}^{\infty}\frac{1}{(r-1)^{3/2+(1/2)(\vartheta_2+\vartheta_3)}} \tag{2.11}$$

where K' is another constant that depends on the arguments z_1, z_2 and the parameters ϑ_1, ϑ_2, ϑ_3.

This estimate proves the convergence of the infinite series in (2.8).

3. Lower and upper bounds on the value of multigamma probability distribution function

If we can calculate the bivariate and univariate marginal distribution function values then we can give good lower and upper bounds on the value of a multivariate distribution function. For in this case we can apply the so called Bonferroni inequalities to estimate the value of the distribution function. Concerning results on inequalities of this type see the papers by Galambos [6], Galambos and Mucci [7], Stathe, Pradhan and Shah [12].

We start the discussion by describing the main results of the last paper and then show how those results can be used for estimation of the value of a multivariate probability distribution function.

Let $A_1, A_2, \ldots, A_n$ be arbitrary events in a probability space. Let y_m be the probability of the occurrence of at least m of those events further

$$S_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \ldots A_{i_k}).$$

For every nonnegative integer k we define

$$U_k = S_1 - k \quad \text{and} \quad 2V_k = 2S_2 - k(k-1).$$

Stathe, Pradhan and Shah proved the following results.

Result 1

$$y_m \geq \frac{2V_{m-1} - (m-2)U_{m-1}}{n(n-m+1)}. \tag{3.1}$$

Result 2. *If* $2V_{m-1} < (n+m-2)U_{m-1}$, *then*

$$y_m \geq 2\frac{(k^*-1)U_{m-1} - V_{m-1}}{(k^*-m)(k^*-m+1)} \tag{3.2}$$

where k^*+m-3 *is the largest integer smaller than or equal to* $2V_{m-1}/U_{m-1}$.

Result 3

$$y_m \leq 1 + [(n+m-1)U_m - 2V_m]/mn. \tag{3.3}$$

Result 4. *If* $2V_m < (m-1)U_m$ *then*

$$y_m \leq 1 - 2\frac{(k^{**}-1)U_m - V_m}{(m-k^{**})(m-k^{**}+1)} \tag{3.4}$$

where $k^{**}+m-1$ *is the largest integer smaller than or equal to* $2V_m/U_m$.

Now let $x_1, x_2, \ldots, x_n$ be random variables and define the random events

$$A_i = \{x_i < z_i\}, \qquad \bar{A}_i = \{x_i \geq z_i\}.$$

As we have the equality

$$P(x_1 < z_1, \ldots, x_n < z_n) = P(A_1 \ldots A_n) = 1 - P(\bar{A}_1 + \cdots + \bar{A}_n),$$

we can apply the above results for the estimation of the probability $P(\bar{A}_1 + \cdots + \bar{A}_n)$. For this we only need those special cases of the above results when $m = 1$.

Below we give the lower and upper bounds in terms of

$$\bar{S}_1 = \sum_{i=1}^{n} P(\bar{A}_i) \quad \text{and} \quad \bar{S}_2 = \sum_{1 \leq i < j \leq n} P(\bar{A}_i \bar{A}_j).$$

Since we have the equalities

$$\bar{S}_1 = n - S_1 \quad \text{and} \quad \bar{S}_2 = \frac{n(n-1)}{2} - (n-1)S_1 + S_2$$

this means that the bounds are expressed only in terms of S_1, S_2, the one and two dimensional marginal probability distribution function values.

From Results 1 and 3 we derive

$$-\frac{n(\bar{S}_1 - 1) - 2\bar{S}_2}{n} \leq P(x_1 < z_1, \ldots, x_n < z_n) \leq 1 - \frac{2\bar{S}_2 + \bar{S}_1}{n^2}. \tag{3.5}$$

In addition, if $2\bar{S}_2 < (n-1)\bar{S}_1$, then according to Result 2 we can give the sharper upper bound

$$P(x_1 < z_1, \ldots, x_n < z_n) \leq 1 - 2\frac{(k^*-1)\bar{S}_1 - \bar{S}_2}{(k^*-1)k^*} \tag{3.6}$$

where k^* is the largest integer smaller than or equal to $2\bar{S}_2/\bar{S}_1 + 2$.

As the inequality $\bar{S}_2 < 0$ is never fulfilled unfortunately Result 4 doesn't yield a sharper lower bound.

We remark that by Theorem 2 of Galambos and Mucci [7] the inequality $2\bar{S}_2 \leq (n-1)\bar{S}_1$ is always true so we can apply the sharper upper bound (3.6) except for the very special case when $2\bar{S}_2 = (n-1)\bar{S}_1$. But in this case it can easily be seen that there is no difference between the two upper bounds.

4. An algorithm for the evaluation of the multivariate gamma probability distribution function

With the notations of the previous section we have

$$P(x_1 < z_1, \ldots, x_n < z_n) = 1 - P(\bar{A}_1 + \cdots + \bar{A}_n)$$
$$= 1 - \bar{S}_1 + \bar{S}_2 - \bar{S}_3 + \cdots + (-1)^n \bar{S}_n. \tag{4.1}$$

Now $\bar{S}_1$ and $\bar{S}_2$ can be expressed in terms of values of the one and two dimensional marginal probability distribution functions. In fact

$$\bar{S}_1 = \sum_{i=1}^{n} P(x_i \geqslant z_i) = n - \sum_{i=1}^{n} P(x_i < z_i) = n - \sum_{i=1}^{n} F_i(z_i) \tag{4.2}$$

and

$$\bar{S}_2 = \sum_{1 \leqslant i < j \leqslant n} P(x_i \geqslant z_i, x_j \geqslant z_j)$$
$$= \sum_{1 \leqslant i < j \leqslant n} (1 - P(x_i < z_i) - P(x_j < z_j) + P(x_i < z_i, x_j < z_j))$$
$$= \frac{n(n-1)}{2} - (n-1) \sum_{i=1}^{n} F_i(z_i) + \sum_{1 \leqslant i < j \leqslant n} F_{ij}(z_i, z_j). \tag{4.3}$$

The main idea of our algorithm is we can produce three different estimates of the distribution function in the same Monte Carlo simulation procedure. The first one is the direct relative frequency corresponding to the probability $P(x_1 < z_1, \ldots, x_n < z_n)$. In the second one we replace $\bar{S}_1$ in (4.1) by the expression (4.2) and then estimate the remaining part of the right hand side. In the third one we replace both $\bar{S}_1$ and $\bar{S}_2$ in (4.1) by the expressions (4.2) and (4.3), respectively and then estimate the remaining part of the right hand side. Let ν_0, ν_1 and ν_2 denote the above estimates, respectively. Then the three different estimates of the distribution function value are:

$$\hat{P}_0 = \nu_0, \qquad \hat{P}_1 = 1 - n + \sum_{i=1}^{n} F_i(z_i) + \nu_1,$$

$$\hat{P}_2 = \frac{(n-1)(n-2)}{2} - (n-2) \sum_{i=1}^{n} F_i(z_i) + \sum_{1 \leqslant i < j \leqslant n} F_{ij}(z_i, z_j) + \nu_2.$$

If we also estimate the covariance matrix of ν_0, ν_1 and ν_2 in the simulation procedure then, using this and the estimates $\hat{P}_0$, $\hat{P}_1$ and $\hat{P}_2$ we can give a very efficient estimation for the function value (4.1). Here we only give the final formulae as the method that we apply is well known in the literature (see Hammersley and Handscomb [8]). Let us introduce the notation

$$C = \begin{pmatrix} c_{00} & c_{01} & c_{02} \\ c_{01} & c_{11} & c_{12} \\ c_{02} & c_{12} & c_{22} \end{pmatrix}$$

for the empirical covariance matrix of ν_0, ν_1 and ν_2 and let

$$\lambda_0 = c_{01}(c_{22}-c_{12}) + c_{11}(c_{02}-c_{22}) + c_{12}(c_{12}-c_{02}),$$

$$\lambda_1 = c_{00}(c_{12}-c_{22}) + c_{01}(c_{22}-c_{02}) + c_{02}(c_{02}-c_{12}),$$

$$\lambda_2 = c_{00}(c_{12}-c_{11}) + c_{01}(c_{01}-c_{12}) + c_{02}(c_{11}-c_{01}).$$

Now we define the following weights

$$w_0 = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}, \qquad w_1 = \frac{\lambda_1}{\lambda_0+\lambda_1+\lambda_2}, \qquad w_2 = \frac{\lambda_2}{\lambda_0+\lambda_1+\lambda_2},$$

and give the final estimation of the distribution function value as follows

$$\hat{P} = w_0\hat{P}_0 + w_1\hat{P}_1 + w_2\hat{P}_2.$$

The empirical variance of the final estimation is

$$w_0^2 c_{00} + w_1^2 c_{11} + w_2^2 c_{22} + 2w_0w_1c_{01} + 2w_0w_2c_{02} + 2w_1w_2c_{12}.$$

The calculation of the estimates ν_0, ν_1 and ν_2 can be performed in an economic way. Suppose that the random numbers $x_1^{(s)}, \ldots, x_n^{(s)}$, $s = 1, \ldots, S$, have the same multivariate gamma distribution as the random variables $x_1, \ldots, x_n$.

According to Formula (1.1) such random numbers can easily be transformed from independent standard gamma distributed random numbers. For the generation of independent standard gamma distributed random numbers there are well tested efficient computer codes on most computers. These are based on the methods presented in the papers by Ahrens and Dieter [1] and Wallace [13]. Now if we count how many times all of the inequalities

$$x_1^{(s)} < z_1, \ldots, x_n^{(s)} < z_n \tag{4.4}$$

are satisfied and divide this number by the sample size S then we get the relative frequency value ν_0 for the given sample.

The situation is more complicated in the case of ν_1 and ν_2. In these cases many probability values are contained in the sums

$$\bar{S}_2 - \bar{S}_3 + \cdots + (-1)^n \bar{S}_n \tag{4.5}$$

and

$$-\bar{S}_3 + \cdots + (-1)^n \bar{S}_n. \tag{4.6}$$

However, we state that the calculation of ν_1 and ν_2 is no more difficult than that of ν_0 and one can make the calculations simultaneously that is one must check all of the inequalities (4.4) only once.

When calculating the value of ν_1 for every fixed s we must check the following inequalities

$$\begin{aligned}
&(x_i^{(s)} \geq z_i, x_j^{(s)} \geq z_j), \quad 1 \leq i < j \leq n, \\
&(x_i^{(s)} \geq z_i, x_j^{(s)} \geq z_j, x_k^{(s)} \geq z_k), \quad 1 \leq i < j < k \leq n, \\
&\quad \vdots \\
&(x_1^{(s)} \geq z_1, x_2^{(s)} \geq z_2, \ldots, x_n^{(s)} \geq z_n).
\end{aligned} \tag{4.7}$$

In view of (4.5) we must add plus one to the observed value every time a subsystem of inequalities with an even number of elements is satisfied and minus one when a subsystem of inequalities with an odd number of elements is satisfied. Instead of checking separately all of the inequality subsystems in (4.7) we can proceed as follows. Let us check again the inequalities (4.4) successively and let k denote the total number of violations. Now if $k \leq 1$ then the observed value is zero and for the other k values we simply can take the sum

$$\binom{k}{2} - \binom{k}{3} + \cdots + (-1)^k \binom{k}{k} = k - 1 \tag{4.8}$$

as the observed value.

The situation is similar in the case of ν_2. Then if $k \leq 2$ the observed value is zero and for the other k values we can take the sum

$$-\binom{k}{3} + \cdots + (-1)^k \binom{k}{k} = -\frac{k(k-1)}{2} + k - 1 = \tfrac{1}{2}(k-1)(2-k) \tag{4.9}$$

as the observed value.

If we add the observed values for all s and divide the sum by the simple size S we get the estimates ν_1 and ν_2, respectively.

Let i_0, i_1 and i_2 denote the observed values according to the estimates ν_0, ν_1 and ν_2. Then we can summarize the necessary calculations at every simulation step. First we must calculate the total number of violations of the inequalities (4.4). We remark one can stop checking the inequalities after the first violation when calculating only the estimate ν_0. Then we can take

$$\begin{aligned}
&i_0 = 1, \quad i_1 = 0, \quad i_2 = 0 && \text{if } k = 0, \\
&i_0 = 0, \quad i_1 = 0, \quad i_2 = 0 && \text{if } k = 1, \\
&i_0 = 0, \quad i_1 = 1, \quad i_2 = 0 && \text{if } k = 2, \\
&i_0 = 0, \quad i_1 = k - 1, \quad i_2 = \tfrac{1}{2}(k-1)(2-k) && \text{if } k \geq 3.
\end{aligned}$$

We must add to the cross products

$$\begin{aligned}
&i_0 i_0 \quad \text{according to } c_{00} \quad && \text{only if } k = 0, \\
&i_1 i_1 \quad \text{according to } c_{11} \quad && \text{only if } k \geq 2,
\end{aligned}$$

$i_2 i_2$ according to c_{22} only if $k \geq 3$,

$i_0 i_1$ according to c_{01} zero for all k,

$i_0 i_2$ according to c_{02} zero for all k,

$i_1 i_2$ according to c_{12} only if $k \geq 3$.

We remark the calculation of the cross product according to c_{00} is superfluous as $i_0 i_0 = 1$ if $k = 0$ that is one will get simply the value ν_0 as cross product.

Finally we give here an algorithmic description of our simulation procedure.

Step 1. Initialization.

Let

$$\nu_0 = 0, \quad \nu_1 = 0, \quad \nu_2 = 0;$$

$$c_{11} = 0, \quad c_{22} = 0, \quad c_{12} = 0;$$

$$s = 0.$$

Step 2. Generation of a new random vector.

Let $s = s + 1$; if $s > S$ then go to Step 6.

Generate the random numbers $x_1^{(s)}, \ldots, x_n^{(s)}$.

Step 3. Initialization of the cycle for checking the inequalities.

Let $k = 0$; $i = 0$.

Step 4. The cycle for checking the inequalities.

Let $i = i + 1$; if $i > n$ then go to Step 5.

If $x_i^{(s)} < z_i$ then repeat Step 4

else let $k = k + 1$ and also repeat Step 4.

Step 5. Update of the frequency values and the cross products.

If $k = 0$ then $\nu_0 = \nu_0 + 1$ and go to Step 2.

If $k = 1$ then go to Step 2.

If $k = 2$ then $\nu_1 = \nu_1 + 1$; $c_{11} = c_{11} + 1$ and go to Step 2.

If $k \geq 3$ then

$$i_1 = k - 1, \quad i_2 = \tfrac{1}{2}(k-1)(2-k);$$

$$\nu_1 = \nu_1 + i_1, \quad \nu_2 = \nu_2 + i_2;$$

$$c_{11} = c_{11} + i_1 i_1, \quad c_{22} = c_{22} + i_2 i_2, \quad c_{12} = c_{12} + i_1 i_2;$$

and go to Step 2.

Step 6. Calculation of the relative frequencies and their covariance matrix.

Let

$$\nu_0 = \nu_0 / S, \quad \nu_1 = \nu_1 / S, \quad \nu_2 = \nu_2 / S;$$

$$c_{00} = \nu(1 - \nu_0), \quad c_{11} = c_{11}/S - \nu_1^2, \quad c_{22} = c_{22}/S - \nu_2^2,$$

$$c_{01} = -\nu_0 \nu_1, \quad c_{02} = -\nu_0 \nu_2, \quad c_{12} = c_{12}/S - \nu_1 \nu_2.$$

Step 7. Calculation of the final estimation.
Let

$$P_0 = \nu_0,$$

$$P_1 = 1 - n + \sum_{i=1}^{n} F_i(z_i) + \nu_1,$$

$$P_2 = \frac{(n-1)(n-2)}{2} - (n-2) \sum_{i=1}^{n} F_i(z_i) + \sum_{1 \leq i < j \leq n} F_{ij}(z_i, z_j) + \nu_2,$$

$$\lambda_0 = c_{01}(c_{22} - c_{12}) + c_{11}(c_{02} - c_{22}) + c_{12}(c_{12} - c_{02}),$$

$$\lambda_1 = c_{00}(c_{12} - c_{22}) + c_{01}(c_{22} - c_{02}) + c_{02}(c_{02} - c_{12}),$$

$$\lambda_2 = c_{00}(c_{12} - c_{11}) + c_{01}(c_{01} - c_{12}) + c_{02}(c_{11} - c_{01}),$$

$$\lambda = \lambda_0 + \lambda_1 + \lambda_2,$$

$$w_0 = \lambda_0/\lambda, \qquad w_1 = \lambda_1/\lambda, \qquad w_2 = \lambda_2/\lambda,$$

$$P = w_0 P_0 + w_1 P_1 + w_2 P_2,$$

Let the empirical variance of the final estimation equal to

$$(w_0^2 c_{00} + w_1^2 c_{11} + w_2^2 c_{22} + 2 w_0 w_1 c_{01} + 2 w_0 w_2 c_{02} + 2 w_1 w_2 c_{12})/S.$$

5. Computer experiences

The computer runs were carried out on the VAX machine of the International Institute for Applied Systems Analysis. Some of the measured execution times are: floating add 1.0 μs, floating multiply 1.5 μs. The generation of random numbers uniformly distributed in (0, 1) was carried out by the IMSL subroutine GGUBS and one floating point number was produced in 115 μs.

The results of Table 1 represent the efficiency of the calculation of the bivariate gamma probability distribution function value by the series expansion (2.8). In this table we also give the results of the one-dimensional integration by (2.3), too. It can be seen that for ϑ_q values near zero the results of the one-dimensional integration are disappointing. The bivariate gamma probability distribution function values were calculated for the arguments $x = 6$, $y = 12$ and for the single parameter values $\vartheta_2 = 1$, $\vartheta_3 = 3$, whereas the values of the common parameter ϑ_1 are given in Table 1.

The speed and proper working of the subroutine for the calculation of the multigamma probability distribution function value was tested by evaluation of a large number of randomly generated examples. In Table 2 we give the results for dimensions 5-15. Four characteristics of the estimators are given as computed from a sample with 1000 elements:

Table 1
Results of the bivariate gamma probability distribution function value calculations

ϑ_1	Probability value by expansion	CPU time for series expansion	Probability value by one-dimensional integration	CPU time for one dimensional integration
5	0.546773	0.009667	0.546785	0.092834
4	0.707647	0.010833	0.707649	0.098333
3	0.842963	0.010167	0.842965	0.143333
2	0.934346	0.010833	0.934346	0.097000
1	0.980933	0.010500	0.980932	0.087833
0.5	0.991608	0.010667	0.990699	0.613333
0.1	0.996236	0.010667	0.733257[a]	0.553666
0.01	0.996930	0.010667	0.123711[a]	0.541500
0.001	0.996993	0.009333	0.013115[a]	0.529000
0.00001	0.997000	0.009167	0.001319[a]	0.542999
0	0.997000[b]	0.002000		

[a]The DCADRE IMSL subroutine for the one-dimensional integration failed.
[b]Exact value calculated by multiplication of the one-dimensional marginal distribution function values.

(i) the estimated value of the probability distribution function, (ii) the dispersion σ of the estimate, (iii) the elapsed processor time in seconds and (iv) the efficiency of the estimator P compared to the estimator P_0, that is the quantity $t_0\sigma_0^2/(t\sigma^2)$. The exact lower and upper bounds of the probability values are given as well.

6. Applications in stochastic programming

The theory of logarithmic concave measures was developed by A. Prékopa [14]. Due to this theory it became possible to handle joint probabilistic constraints in the stochastic programming problems. These constraints have the form

$$P(a_i x \geq \beta_i,\ i = 1, \ldots, m) \geq p, \tag{6.1}$$

where $\beta_1, \ldots, \beta_m$ have a logconcave distribution.

When solving numerically this type of stochastic programming problems we have to calculate the probability value (6.1). For this purpose one can apply multi-dimensional integration techniques. Unfortunately the multi-dimensional integration methods have extremely slow convergence for higher dimensions. In these cases we can apply Monte-Carlo methods only. This paper is a contribution to this field. We remark the lower and upper bounds derived in Section 3 are useful in the solution of the stochastic programming problems as well.

Some applications are described in the papers [15, 16 and 17].

Table 2

Results for the calculation of the multigamma probability distribution function values

Example	Estimator	$\hat{P}_0$	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}$	Lower bound	Upper bound
1. p	0.9830	0.9797	0.9843	0.9843	0.9791	0.9843	
	σ	0.0041	0.0020	0.0000	0.0000	—	—
$n=5$	t	24.1667	—	—	25.3333	—	—
	eff.	—	—	—	∞[a]	—	—
2.	p	0.9720	0.9734	0.9754	0.9747	0.9707	0.9774
	σ	0.0052	0.0032	0.0014	0.0004	—	—
$n=6$	t	31.1833	—	—	32.0167	—	—
	eff.	—	—	—	142.5058	—	—
3.	p	0.9750	0.9743	0.9688	0.9728	0.9678	0.9798
	σ	0.0049	0.0056	0.0100	0.0012	—	—
$n=7$	t	40.3500	—	—	40.6167	—	—
	eff.	—	—	—	13.8514	—	—
4.	p	0.9620	0.9643	0.9524	0.9615	0.9524	0.9729
	σ	0.0060	0.0083	0.0211	0.0020	—	—
$n=8$	t	49.8333	—	—	50.7500	—	—
	eff.	—	—	—	6.7068	—	—
5.	p	0.9420	0.9453	0.9392	0.9437	0.9392	0.9453
	σ	0.0074	0.0071	0.0210	0.0010	—	—
$n=9$	t	66.6833	—	—	68.0500	—	—
	eff.	—	—	—	50.2798	—	—
6.	p	0.9570	0.9637	0.9716	0.9669	0.9541	0.9758
	σ	0.0064	0.0056	0.0045	0.0011	—	—
$n=10$	t	71.3000	—	—	73.3500	—	—
	eff.	—	—	—	22.9346	—	—
7.	p	0.9893	0.9900	0.9900	0.9900	0.9893	0.9900
	σ	0.0033	0.0010	0.0000	0.0000	—	—
$n=11$	t	90.2000	—	—	93.4500	—	—
	eff.	—	—	—	∞[a]	—	—
8.	p	0.9973	0.9969	0.9973	0.9969	0.9969	0.9973
	σ	0.0014	0.0000	0.0000	0.0000	—	—
$n=12$	t	105.5833	—	—	105.9167	—	—
	eff.	—	—	—	∞[a]	—	—
9.	p	0.9884	0.9872	0.9884	0.9872	0.9872	0.9884
	σ	0.0033	0.0000	0.0000	0.0000	—	—
$n=13$	t	122.9500	—	—	124.2167	—	—
	eff.	—	—	—	∞[a]	—	—
10.	p	0.9860	0.9847	0.9863	0.9847	0.9847	0.9863
	σ	0.0037	0.0000	0.0000	0.0000	—	—
	t	141.8000	—	—	144.2833	—	—
	eff.	—	—	—	∞[a]	—	—
11.	p	0.9771	0.9776	0.9804	0.9776	0.9771	0.9804
	σ	0.0048	0.0010	0.0000	0.0000	—	—
	t	155.2833	—	—	158.9000	—	—
	eff.	—	—	—	∞[a]	—	—

[a]The value estimated by the Monte-Carlo simulation was zero so the efficiency is infinite.

References

[1] J.H. Ahrens and U. Dieter, "Computer methods for sampling from gamma, beta, Poisson and binomial distributions", *Computing* 12 (1974) 223-246.

[2] I. Deák, "Three digit accurate multiple normal probabilities", *Numerische Mathematik* 35 (1980) 369-380.

[3] Z.N. Diachenko, "Ob odnom vide dvumernoi funkcii-raspredelenia", *Trudy Lesotehnicheskoi Akademii im. S.M. Kirova* 94 (1962) 5-17.

[4] G.K. Eagleson, "Polynomial expansions of bivariate distributions", *Annals of Mathematical Statistics* 35 (1964) 1208-1215.

[5] A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, *Higher transcendental functions, Vol. II* (McGraw-Hill, New York, Toronto, London, 1953).

[6] J. Galambos, "Bonferroni inequalities", *The Annals of Probability* 5 (1977) 577-581.

[7] J. Galambos and R. Mucci, "Inequalities for linear combinations of binomial moments", *Publicationes Mathematicae Debrecen* 27 (1980) 263-268.

[8] J.M. Hammersley and D.C. Handscomb, *Monte Carlo methods* (Methuen, London, 1964).

[9] H.O. Lancaster, "Correlations and canonical forms of bivariate distributions", *Annals of Mathematical Statistics* 34 (1963) 532-538.

[10] A. Prékopa and T. Szántai, "A new multivariate gamma distribution and its fitting to empirical data", *Water Resources Research* 14 (1978) 19-24.

[11] I.O. Sarmanov, "Process gamma-korrelacii i ego svoistva", *Doklady Akademii Nauk SSSR* 191 (1970) 30-32.

[12] Y.S. Stathe, Meena Pradhan and S.P. Shah, "Inequalities for the probability of the occurrence of at least *m* out of *n* events", *Journal of Applied Probability* 17 (1980) 1127-1132.

[13] N.D. Wallace, "Computer generation of gamma random variates with non-integral shape parameters", *Communications of the ACM* 17 (1974) 691-695.

[14] A. Prékopa, "Logarithmic concave measures with application to stochastic programming", *Acta Scientiarum Mathematicarum* 32 (1971) 301-316.

[15] A. Prékopa and T. Szántai, "On multi-stage stochastic programming (with application to optimal control of water supply)", in: A. Prékopa, ed., *Colloquia Mathematica Societatis János Bolyai* 12. *Progress in operations research* (János Bolyai Mathematical Society, Budapest and North-Holland Publishing Company, Amsterdam, London, 1976) 733-755.

[16] A. Prékopa and T. Szántai, "On optimal regulation of a storage level with application to the water level regulation of a lake", in: A. Prékopa, ed., *Survey of mathematical programming*, Proceedings of the 9th International Mathematical Programming Symposium, Budapest, August 23-27, 1976 (Publishing House of the Hungarian Academy of Sciences, Budapest, 1978) 183-210.

[17] A. Prékopa and T. Szántai, "Flood control reservoir system design using stochastic programming", *Mathematical Programming Study* 9 (1978) 138-151.

Mathematical Programming Study 27 (1986) 17-38
North-Holland

MULTIDIMENSIONAL NUMERICAL INTEGRATION USING PSEUDORANDOM NUMBERS

Harald NIEDERREITER

Mathematical Institute, Austrian Academy of Sciences, Dr. Ignaz-Seipel-Platz 2, A-1010 Vienna, Austria

Received 18 January 1984
Revised manuscript received 17 July 1984

Practical implementations of Monte Carlo methods for multidimensional numerical integration use nodes derived from pseudorandom numbers. We present effective error bounds for Monte Carlo integration with nodes derived from the two most common types of pseudorandom numbers, namely linear congruential pseudorandom numbers and Tausworthe pseudorandom numbers. We compare the results with those obtained by the use of quasirandom nodes.

Key words: Monte Carlo Method, Numerical Integration, Linear Congruential Pseudorandom Numbers, Tausworthe Pseudorandom Numbers, Quasirandom Points.

1. Introduction

A standard method for the approximate calculation of high-dimensional integrals is the Monte Carlo method. It is based on approximating the integral by an average of values of the integrand at randomly selected nodes. If we normalize the integration domain to be the s-dimensional unit cube $I^s = [0, 1]^s$, then

$$\int_{I^s} f(\boldsymbol{t})\, d\boldsymbol{t} \approx \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{x}_n), \tag{1}$$

where the nodes $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1} \in I^s$ are obtained by taking N independent samples from the uniform distribution on I^s. The expected integration error is of the order of magnitude $N^{-1/2}$.

In practical implementations of Monte Carlo integration one actually does not use random nodes, but rather nodes derived from pseudorandom numbers (abbreviated PRN) that can easily be generated by the computer. The question arises whether for such deterministic nodes one can give effective error bounds for the approximation (1). We will show that recent work of the author on PRN can be used to establish satisfactory error bounds for nodes derived from the two most popular types of PRN, namely linear congruential (or Lehmer) PRN and digital (or Tausworthe) PRN. In fact, it will turn out that if the parameters in these methods are well chosen, then the error bounds are much smaller than the Monte Carlo bound $O(N^{-1/2})$.

The basic method of obtaining nodes in I^s from PRN is the following. Let x_0, $x_1, \ldots$ be uniform PRN in $I = [0, 1]$ and let s be a given dimension. Then we define

the points

$$\boldsymbol{x}_n = (x_n, x_{n+1}, \ldots, x_{n+s-1}) \in I^s \quad \text{for } n \geq 0. \tag{2}$$

It is one of the advantages of this method that the calculation of one new s-dimensional point just requires the calculation of one new coordinate, so that these points can be generated very quickly.

As usual in numerical integration, the error bounds depend on the regularity of the integrand. Already for integrands with a rather low degree of regularity, namely integrands of bounded variation, the bounds for (1) guarantee good convergence rates for the method. In fact, by a proper choice of parameters in the method of generating PRN one can obtain error bounds of the order $O(N^{-1}(\log N)^{c(s)})$ with $c(s)$ only depending on s. For integrands of bounded variation, the bounds for the integration error in (1) involve a quantity that measures the uniformity of distribution of the nodes, the so-called discrepancy. For arbitrary points $\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1} \in I^s$ we define the discrepancy

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) = \sup_{\boldsymbol{t} \in I^s} \left| \frac{1}{N} \# \{n: \boldsymbol{w}_n \leq \boldsymbol{t}\} - |\boldsymbol{t}| \right|,$$

where the vector inequality $\boldsymbol{w}_n \leq \boldsymbol{t}$ is interpreted coordinate-wise and $|\boldsymbol{t}|$ denotes the product of the coordinates of $\boldsymbol{t}$. Detailed information on the discrepancy is given in Kuipers and Niederreiter [9, Chapter 2]. If we are dealing with points $\boldsymbol{x}_n$ derived from PRN as in (2), we write

$$D_N^{(s)} = D_N(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}). \tag{3}$$

As there are various concepts of bounded variation for multivariable functions, we note that the concept to be used here is that of bounded variation in the sense of Hardy and Krause. We refer to Hua and Wang [6, Chapter 5], Kuipers and Niederreiter [9, Chapter 2], and Niederreiter [16, Section 2] for the definition and a discussion of this concept. The following result of Hlawka [5] is basic. A proof can also be found in Kuipers and Niederreiter [9, pp. 147–153]; for the version given here see Niederreiter [16, Section 2].

Theorem A. *If f is of bounded variation $V(f)$ on I^s in the sense of Hardy and Krause and $\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}$ are arbitrary points in I^s, then*

$$\left| \int_{I^s} f(\boldsymbol{t})\, d\boldsymbol{t} - \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{w}_n) \right| \leq V(f) D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}).$$

In some applications the exact value of $V(f)$ may be hard to determine. In such cases one will have to make do with upper bounds for $V(f)$ or with approximate values. Upper bounds for $V(f)$ can be obtained easily if f satisfies a Lipschitz condition on I^s (see [6, p. 101]). A numerical method for approximating $V(f)$ can be based on a standard discretization approach, namely to place a sufficiently fine

grid on I^s and calculate the variation $V(f; P)$ of f with respect to the partition P induced by the grid. This variation $V(f; P)$ is just a finite sum of values of difference operators (compare with [16, Section 2]). If f is continuous on I^s and the mesh size of the grid tends to zero, then $V(f; P)$ converges to the exact value of $V(f)$.

If we use the nodes $\boldsymbol{x}_n$ defined in (2), then Theorem A shows that we have to establish bounds for the discrepancy $D_N^{(s)}$ in (3). In Section 2 we will present such bounds for nodes $\boldsymbol{x}_n$ derived from linear congruential PRN and in Section 3 for nodes $\boldsymbol{x}_n$ derived from digital PRN.

It should be noted that there are analogs of Theorem A for other regularity classes, such as continuous functions or Riemann-integrable functions, and also for general integration domains. We refer to [16, Section 2] for details.

It is a known phenomenon in classical integration formulas such as the trapezoidal rule that periodic integrands tend to yield much smaller integration errors. This phenomenon occurs again in our context if nodes derived from linear congruential PRN are used. We will discuss this case in Section 4.

In the literature one can find constructions of various sets of nodes for the specific purpose of making the integration error in (1) small. These nodes make no pretense of passing statistical tests for randomness, and are therefore often called quasirandom points (compare with [16, Section 3]). They offer the advantage that they have very small discrepancy, but their calculation is more costly than that of the nodes in (2). In Section 5 we will compare the relative merits of quasirandom points and of nodes derived from PRN in more detail.

The following notation will be convenient throughout the paper. For a positive integer b put

$$C(b) = \{h \in \mathbb{Z}: -b/2 < h \leq b/2\}.$$

The set $C(b)$ forms a complete residue system modulo b in the sense of number theory. For a given dimension $s \geq 1$ we define $C_s(b) = \{(h_0, \ldots, h_{s-1}) \in \mathbb{R}^s: h_i \in C(b)$ for $0 \leq i \leq s-1$, not all $h_i = 0\}$. By a lattice point we mean a point in $\mathbb{R}^s$ with integer coordinates, i.e. an element of $\mathbb{Z}^s$. In particular, all the elements of $C_s(b)$ are lattice points.

2. Nodes derived from linear congruential PRN

Linear congruential PRN are generated in the following way. Let $M \geq 3$ and r be integers and let λ be an integer relatively prime to M with $2 \leq \lambda < M$. A sequence $y_0, y_1, \ldots$ of integers with $0 \leq y_n < M$ is generated by the recursion

$$y_{n+1} \equiv \lambda y_n + r \bmod M \quad \text{for } n = 0, 1, \ldots,$$

and the normalized numbers $x_n = y_n/M$ in the interval $I = [0, 1]$ are then linear congruential PRN with modulus M, multiplier λ, and increment r. In practice, M is taken to be a large prime or a large power of 2. The sequence $x_0, x_1, \ldots$ is always

purely periodic; let τ denote the length of the least period. Only those sequences of PRN with a large value of τ are of interest for simulation purposes. A detailed study of the properties of linear congruential PRN can be found in Knuth [7, Chapter 3].

From the discussion in Section 1, and in particular from Theorem A, it is clear that effective bounds for the integration error in (1) with nodes $\boldsymbol{x}_n$ derived from linear congruential PRN are available once we can establish bounds for the discrepancy $D_N^{(s)}$ in (3). Such bounds can, in fact, be found in several papers of the author [13; 15; 16, Section 11; 19]. We summarize the results for the most interesting cases. We can assume $s \geqslant 2$, since for $s = 1$ classical integration methods are preferable anyway.

The most convenient description of the bounds for $D_N^{(s)}$ is in terms of the so-called figure of merit $\rho^{(s)}(\lambda, b)$, which depends on the dimension s, the multiplier λ, and an integer b that either agrees with the modulus M or is a slightly reduced modulus. For more precise (but also more complicated) bounds we refer to the papers listed above. To define $\rho^{(s)}(\lambda, b)$, we first introduce the 'size function' $R(\boldsymbol{h})$ for lattice points $\boldsymbol{h} = (h_0, \ldots, h_{s-1})$ by

$$R(\boldsymbol{h}) = \prod_{i=0}^{s-1} \max(1, |h_i|). \tag{4}$$

Then we set

$$\rho^{(s)}(\lambda, b) = \min R(\boldsymbol{h}), \tag{5}$$

where the minimum is taken over all $\boldsymbol{h} \in C_s(b)$ with

$$\sum_{i=0}^{s-1} h_i \lambda^i \equiv 0 \bmod b.$$

Clearly, $\rho^{(s)}(\lambda, b)$ is a positive integer.

We consider now two typical cases leading to very large values of the least period τ. First, let M be prime, $r = 0$, $y_0 \neq 0$, and λ be such that modulo M it yields the largest possible multiplicative order $M - 1$ (i.e. λ is a so-called primitive root mod M); then $\tau = M - 1$. Second, let M be a power of 2, r odd, and $\lambda \equiv 5 \bmod 8$; then $\tau = M$. In both cases we have

$$D_\tau^{(s)} \leqslant \frac{c_s (\log M)^s}{\rho^{(s)}(\lambda, M)} \tag{6}$$

with an explicit constant c_s only depending on s. A possible value of c_s is $(2/\log 2)^s$. Another common case is the one where M is a power of 2, $r = 0$, y_0 odd, and $\lambda \equiv 5 \bmod 8$; here $\tau = M/4$ and

$$D_\tau^{(s)} \leqslant \frac{c_s (\log M)^s}{\rho^{(s)}(\lambda, \tau)}. \tag{7}$$

The bounds in (6) and (7) already represent (up to logarithmic factors) the true order of magnitude of $D_\tau^{(s)}$, since $D_\tau^{(s)}$ can also be bounded from below in terms of the reciprocal of the relevant figure of merit (see [15; 16, Section 11; 19]).

The bounds in (6) and (7) refer to the case where we exhaust the full period of the PRN. If only the first N points are taken, where $N < \tau$, then bounds for $D_N^{(s)}$ can still be obtained, but they are less satisfactory than those for $N = \tau$ (see [15; 16, Section 11; 19]). For instance, in the two cases pertaining to (6) we have

$$D_N^{(s)} \leq c_s\left(N^{-1}M^{1/2}(\log M)^{s+1} + \frac{(\log M)^s}{\rho^{(s)}(\lambda, M)}\right) \quad \text{for } 1 \leq N < \tau. \tag{8}$$

These bounds are, of course, only useful if one knows that the figure of merit $\rho^{(s)}(\lambda, b)$ can be made large for large b. In fact, it follows from [13, Theorem 4.4] that for any $s \geq 2$ and any prime b there exists a primitive root λ mod b with

$$\rho^{(s)}(\lambda, b) \geq \frac{c_s' b}{(\log b)^{s-1} \log\log b},$$

where the constant $c_s' > 0$ depends only on s. For the other case of interest, namely where b is a large power of 2, it was shown in [13, Theorem 6.5] that for $s = 2$ there always exists a $\lambda \equiv 5 \bmod 8$ with

$$\rho^{(2)}(\lambda, b) \geq \frac{c_2' b}{\log b}.$$

From the available tables for figures of merit (see e.g. Hua and Wang [6] and Maisonneuve [11]) one is led to expect that for any $s \geq 2$ and any b that is either a prime or a power of 2 there is always a λ yielding a value of $\tau \approx M$ and satisfying $\rho^{(s)}(\lambda, b) \geq c_s' b/(\log b)^{s-1}$. In the case (6), for instance, this will then give a bound for the integration error of the form $O(M^{-1}(\log M)^{2s-1})$ for integrands of bounded variation, where the number of nodes is $M-1$ or M. By a direct study of the discrepancy one can prove the existence of multipliers λ yielding even smaller bounds. For instance, if M is prime, $r = 0$, and $y_0 \neq 0$, then it follows from the proof of [13, Theorem 3.4 and Corollary 3.5] that for any $s \geq 2$ a primitive root λ mod M yields on the average

$$D_\tau^{(s)} = O(M^{-1}(\log M)^s \log\log M)$$

with $\tau = M - 1$. A similar result can be shown for $s = 2$ when M is a power of 2, $r = 0$, y_0 odd, and $\lambda \equiv 5 \bmod 8$, the average bound being $D_\tau^{(s)} = O(M^{-1}(\log M)^2)$ with $\tau = M/4$ (see [13, Theorem 7.3 and Corollary 7.4]). Again one can expect that for any $s \geq 2$ and any modulus M that is either a prime or a power of 2, there is a multiplier yielding $\tau \approx M$ and

$$D_\tau^{(s)} = O(M^{-1}(\log M)^s). \tag{9}$$

For integrands of bounded variation this will then give a bound for the integration error of the same order of magnitude.

Since multipliers yielding a value of $\tau \approx M$ produce on the average a small value of $D_\tau^{(s)}$, a search method used by Haber [3] in a related context can be employed here. Given a dimension $s \geq 2$ and a modulus M, pick a relatively small number of 'candidates' λ (with λ a primitive root mod M for M prime and $\lambda \equiv 5 \bmod 8$ for M a power of 2) and calculate $\rho^{(s)}(\lambda, M)$ for those λ. Then there is a good chance that the candidate λ yielding the largest value of $\rho^{(s)}(\lambda, M)$ is a good multiplier in the sense that the corresponding value of $D_\tau^{(s)}$ meets the bound (9). According to the numerical experience reported in Haber [3], a choice of 50 candidates should be quite sufficient. We note that a systematic search for optimal multipliers has so far only been carried out in dimension $s = 2$ and for moduli $M = 2^\alpha$ with $\alpha \leq 35$ (see Borosh and Niederreiter [1]). We refer to Knuth [7, Chapter 3] for tables of good, though not necessarily optimal, multipliers.

When generating linear congruential PRN for simulation purposes, it is customary to work with a very large modulus such as $M = 2^{35}$ and then use only a relatively small segment of the period of the sequence. For purposes of numerical integration this does not seem to be the best strategy since the bound (8) for parts of the period is rather weak. It is more advisable to exploit the good behavior of the discrepancy for the full period by choosing a smaller modulus and using all the points in the period of the sequence. The use of smaller moduli also facilitates the search for multipliers with large figures of merit. Needless to say, for very high dimensions s the logarithmic factors in the error bounds can cause problems unless sufficiently large moduli are employed.

3. Nodes derived from digital PRN

We consider now the discrepancy of points (2) derived from a sequence $x_0, x_1, \ldots$ of digital PRN. These PRN were proposed by Tausworthe [20] and are generated as follows. Let p be a prime (in practice often $p = 2$) and let $k \geq 2$ be an integer. A sequence $y_0, y_1, \ldots$ of integers with $0 \leq y_n < p$ is generated by the recursion

$$y_{n+k} \equiv a_{k-1} y_{n+k-1} + \cdots + a_0 y_n \bmod p \quad \text{for } n = 0, 1, \ldots,$$

where the a_i are constant integral coefficients. The initial values $y_0, \ldots, y_{k-1}$ should not all be 0. The characteristic polynomial

$$g(x) = x^k - a_{k-1} x^{k-1} - \cdots - a_0$$

of the recursion, considered as a polynomial over the finite field $F_p = \mathbb{Z}/p\mathbb{Z}$, is assumed to be primitive, that is, g is the minimal polynomial over F_p of a generator of the cyclic group F_q^*, the multiplicative group of the finite field F_q with $q = p^k$ elements. This guarantees that the sequence of y_n is a maximal period sequence, i.e. a sequence for which the length τ of the least period is as large as possible for fixed p and k, namely $\tau = p^k - 1$. We refer to Lidl and Niederreiter [10, Chapter 8] for these facts on sequences generated by linear recursions.

From the sequence $y_0, y_1, \ldots$ we obtain a sequence $x_0, x_1, \ldots$ of uniform PRN in the interval $I = [0, 1]$ in the following manner. We choose a positive integer $m \leq k$ and put

$$x_n = \sum_{j=0}^{m-1} y_{mn+j} p^{-j-1} \quad \text{for } n = 0, 1, \ldots,$$

that is, blocks of consecutive y_n of length m are interpreted as digits of x_n in the base p. The sequence of x_n is again periodic, with the length of the least period being $\tau / \gcd(m, \tau)$. In order to make the length of the least period as large as possible, we assume that $\gcd(m, \tau) = 1$, i.e. that m is relatively prime to τ. In this way, the length of the least period of the sequence of x_n is again $\tau = p^k - 1$.

With this sequence of x_n we set up the s-dimensional nodes $\boldsymbol{x}_n$ according to (2) and consider their discrepancy $D_N^{(s)}$ defined in (3). This discrepancy was studied in Niederreiter [18] for $s = 1$ and in Niederreiter [17] for $s > 1$ under more general hypotheses. For small dimensions s, i.e. for $s \leq k/m$, we get in our case

$$D_\tau^{(s)} < \frac{s}{p^m} + \frac{c_s}{\tau} m^s (\log p)^s$$

with an explicit constant $c_s > 0$ only depending on s. A possible value of c_s is $(2/\pi)^s$. This bound is independent of the characteristic polynomial g once g is chosen as a primitive polynomial. The situation changes for dimensions $s > k/m$, where the magnitude of $D_\tau^{(s)}$ depends strongly on the characteristic polynomial g. A bound for $D_\tau^{(s)}$ can be given in terms of a quantity $R^{(s)}(g, p, m)$ introduced in [17]. For m-dimensional lattice points $\boldsymbol{h} = (h_1, \ldots, h_m)$ define

$$P(\boldsymbol{h}) = \begin{cases} \dfrac{p^m - 1}{p^m - p^{m-1}} & \text{if } \boldsymbol{h} = \boldsymbol{0}, \\ \dfrac{1}{p^d |\sin(\pi/p) h_d|} + \dfrac{p^{m-d} - 1}{p^m - p^{m-1}} & \text{if } \boldsymbol{h} \in C_m(p), \end{cases}$$

where d is the largest subscript with $h_d \neq 0$. For an s-tuple $H = (\boldsymbol{h}_0, \ldots, \boldsymbol{h}_{s-1})$ of lattice points with $\boldsymbol{h}_i = \boldsymbol{0}$ or $\boldsymbol{h}_i \in C_m(p)$, define

$$P(H) = \prod_{i=0}^{s-1} P(\boldsymbol{h}_i).$$

With each $H = (\boldsymbol{h}_0, \ldots, \boldsymbol{h}_{s-1})$, where $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{im})$, we associate a polynomial

$$F_H(x) = \sum_{i=0}^{s-1} \sum_{j=1}^{m} h_{ij} x^{im+j-1}$$

over the finite field F_p. Then

$$R^{(s)}(g, p, m) = \sum_{H} P(H),$$

where the sum is extended over all H such that F_H is divisible by g and F_H is not the zero polynomial.

The quantity $R^{(s)}(g, p, m)$ is rather unwieldy for calculation, so we replace it by a somewhat simpler one. We note first that

$$P(\boldsymbol{h}) < 2 \text{ for } \boldsymbol{h} = \boldsymbol{0},$$

and that for $\boldsymbol{h} \in C_m(p)$ we can use $|\sin \pi t| \geq 2|t|$ for $|t| \leq \frac{1}{2}$ to obtain

$$P(\boldsymbol{h}) < \frac{1}{2p^{d-1}|h_d|} + \frac{2}{p^d} = \frac{p + 4|h_d|}{2p^d|h_d|} \leq \frac{3}{2p^{d-1}|h_d|}.$$

If we associate with $\boldsymbol{h} = (h_1, \ldots, h_m)$ the polynomial

$$G(x) = \sum_{j=1}^{m} h_j x^{j-1}$$

over F_p and define

$$r(G) = \begin{cases} \frac{1}{2} & \text{for } G = 0, \\ \frac{2}{3} p^{\deg(G)} \, |\text{leading coeff. of } G| & \text{for } G \neq 0, \end{cases}$$

where the leading coefficient of G is interpreted to be an element of $C_1(p)$, then

$$P(\boldsymbol{h}) < \frac{1}{r(G)} \quad \text{for all } \boldsymbol{h}. \tag{10}$$

We introduce now a quantity $R_1^{(s)}(g, p, m)$ as follows. We consider polynomials F over F_p with $\deg(F) < ms$. Such a polynomial can be represented uniquely in the form

$$F(x) = \sum_{i=0}^{s-1} G_i(x) x^{im} \quad \text{with } \deg(G_i) < m.$$

We set

$$R(F) = \prod_{i=0}^{s-1} r(G_i)$$

and define

$$R_1^{(s)}(g, p, m) = \sum_{F} R(F)^{-1},$$

where the sum is extended over all F with $0 \leq \deg(F) < ms$ that are divisible by g. From (10) we get then

$$R^{(s)}(g, p, m) < R_1^{(s)}(g, p, m).$$

It follows thus from [17, Théorème 2] that

$$D_\tau^{(s)} < \frac{s}{p^m} + \frac{c_s}{\tau} m^s (\log p)^s + R_1^{(s)}(g, p, m) \tag{11}$$

with an explicit constant $c_s > 0$ only depending on s. A possible value of c_s is $(2/\pi)^s$.

One can also establish bounds for $D_N^{(s)}$ with $N < \tau$. These bounds are analogous to (8) and thus comparatively weak (see [17]). Hence we have a situation similar

to that for linear congruential PRN (compare with the discussion at the end of Section 2). It is advisable to choose k so small that it becomes feasible to use all $\tau = p^k - 1$ nodes from the full period in the numerical integration.

Concerning the order of magnitude of $R_1^{(s)}(g, p, m)$ one can show that one can achieve basically the same order of magnitude as for the second term on the right-hand side of (11). Indeed, one proves in analogy with [17, Théorème 3] that for given s, m, p, and k there exists a primitive polynomial g over F_p of degree k with

$$R_1^{(s)}(g, p, m) < \frac{c_s}{\tau} m^{s+1} (\log p)^s \log\log \tau, \tag{12}$$

where the constant c_s depends only on s. This upper bound represents in fact the average order of magnitude of $R_1^{(s)}(g, p, m)$ when g runs through all primitive polynomials over F_p of degree k. The search for primitive polynomials g meeting the bound (12) can thus proceed in a similar way as the search for good multipliers λ in Section 2. Concretely, we select a relatively small number of primitive polynomials g over F_p of fixed degree k at random and calculate $R_1^{(s)}(g, p, m)$. The polynomial yielding the least value of this quantity can then be expected to be a good choice as a characteristic polynomial. This procedure is facilitated by the fact that extensive tables of primitive polynomials are already available; see Lidl and Niederreiter [10, Chapter 10]. With a primitive polynomial g satisfying (12) we obtain from (11),

$$D_\tau^{(s)} < c_s \left(\frac{1}{p^m} + \frac{1}{\tau} m^{s+1} (\log p)^s \log\log \tau \right)$$

with a suitable constant c_s. For instance, the choice $m = k$ yields

$$D_\tau^{(s)} = O\left(\frac{1}{\tau} (\log \tau)^{s+1} \log\log \tau \right),$$

and thus, by Theorem A, a bound for the integration error of the same order of magnitude for integrands of bounded variation. It is very likely that there exist choices for g for which this bound becomes $O(\tau^{-1}(\log \tau)^s)$, which would then be comparable with (9).

4. Periodic integrands

Let f be a complex-valued continuous function on $\mathbb{R}^s$ of period 1 in each variable. We will show that if (1) is used with nodes $\boldsymbol{x}_n$ derived from linear congruential PRN and if f satisfies additional regularity conditions, then much better error bounds than those in Section 2 can be obtained.

If f is as above, then it has a Fourier coefficient

$$c(\boldsymbol{h}) = \int_{I^s} f(\boldsymbol{t})\, e^{-2\pi i \langle \boldsymbol{h}, \boldsymbol{t} \rangle}\, d\boldsymbol{t}$$

associated with every s-dimensional lattice point $\boldsymbol{h}$, where $\langle \boldsymbol{h}, \boldsymbol{t} \rangle$ denotes the standard inner product in $\mathbb{R}^s$. By the Riemann-Lebesgue lemma, $c(\boldsymbol{h})$ tends to 0 as $\boldsymbol{h}$ moves away from the origin. We introduce function classes $E^k(A)$ depending on the rate of convergence of $c(\boldsymbol{h})$ to 0. For constants $k > 1$ and $A > 0$ we let $f \in E^k(A)$ if

$$|c(\boldsymbol{h})| \leq AR(\boldsymbol{h})^{-k} \quad \text{for all } \boldsymbol{h} \neq \boldsymbol{0}, \tag{13}$$

where $R(\boldsymbol{h})$ is as in (4). We note an important sufficient condition for f to belong to such a function class: if $k > 1$ is an integer and all the partial derivatives

$$\frac{\partial^{q_1 + \cdots + q_s} f}{\partial t_1^{q_1} \cdots \partial t_s^{q_s}} \quad \text{with } 0 \leq q_j \leq k-1 \text{ for } 1 \leq j \leq s$$

exist and are of bounded variation on I^s in the sense of Hardy and Krause, then $f \in E^k(A)$ with a value of A that can be given explicitly (see Zaremba [21]). The following result is basic.

Lemma 1. *The Fourier series of any $f \in E^k(A)$ is absolutely and uniformly convergent and represents f.*

Proof. The Fourier series of f is given by

$$F(\boldsymbol{t}) = \sum_{\boldsymbol{h}} c(\boldsymbol{h})\, e^{2\pi i \langle \boldsymbol{h}, \boldsymbol{t} \rangle},$$

where $\boldsymbol{h}$ runs through all s-dimensional lattice points. From (13) we get

$$\begin{aligned} \sum_{\boldsymbol{h}} |c(\boldsymbol{h})| &\leq |c(\boldsymbol{0})| + A \sum_{\boldsymbol{h} \neq \boldsymbol{0}} R(\boldsymbol{h})^{-k} < |c(\boldsymbol{0})| + A \sum_{\boldsymbol{h}} R(\boldsymbol{h})^{-k} \\ &= |c(\boldsymbol{0})| + A \sum_{h_0, \ldots, h_{s-1} \in \mathbb{Z}} \max(1, |h_0|)^{-k} \cdots \max(1, |h_{s-1}|)^{-k} \\ &= |c(\boldsymbol{0})| + A\left(\sum_{h \in \mathbb{Z}} \max(1, |h|)^{-k} \right)^s = |c(\boldsymbol{0})| + A\left(1 + 2 \sum_{h=1}^{\infty} h^{-k} \right)^s < \infty, \end{aligned}$$

thus the Fourier series is absolutely and uniformly convergent. In particular, F is continuous and has the same Fourier coefficients as f. From the completeness of the orthogonal system $\{e^{2\pi i \langle \boldsymbol{h}, \boldsymbol{t} \rangle} : \boldsymbol{h} \in \mathbb{Z}^s\}$ it follows that $F = f$.

We also need some basic principles about sequences $z_0, z_1, \ldots$ of integers satisfying a recursion

$$z_{n+1} = \lambda z_n + r \quad \text{for } n = 0, 1, \ldots, \tag{14}$$

where λ and r are given integers. Let M be a positive integer relatively prime to λ. If the sequence $z_0, z_1, \ldots$ is considered modulo M, then it is periodic, and we denote by $\tau(M)$ the length of the least period. If d is a positive divisor of M, then $\tau(M)$ is a multiple of $\tau(d)$. For typographical convenience we write now $e(t) = e^{2\pi i t}$ for real t.

Lemma 2. *Let $z_0, z_1, \ldots$ be a sequence of integers satisfying* (14), *and let M be relatively prime to λ. If there exists a divisor $d > 1$ of M such that $\tau(M) = d\tau(M/d)$, then*

$$\sum_{n=0}^{\tau(M)-1} e\left(\frac{z_n}{M}\right) = 0.$$

Proof. Put $P = \tau(M)$ and $Q = \tau(M/d)$. From (14) it follows that P is identical with the number of distinct least residues modulo M in the sequence of z_n, and similarly for Q. If $b_1, \ldots, b_Q$ are the distinct least residues modulo M/d of the sequence, then the least residues modulo M of the sequence are necessarily among the arithmetic progressions $\{b_j + hM/d : h = 0, 1, \ldots, d-1\}$, $1 \leq j \leq Q$. These yield $dQ = P$ distinct least residues modulo M, and so they produce all the P least residues modulo M of the sequence. Therefore

$$\sum_{n=0}^{\tau(M)-1} e\left(\frac{z_n}{M}\right) = \sum_{j=1}^{Q} \sum_{h=0}^{d-1} e\left(\frac{b_j}{M} + \frac{h}{d}\right) = \left(\sum_{j-1}^{Q} e\left(\frac{b_j}{M}\right)\right)\left(\sum_{h=0}^{d-1} e\left(\frac{h}{d}\right)\right) = 0,$$

since the second sum vanishes.

We consider now nodes $\boldsymbol{x}_n$ derived from linear congruential PRN. We use the same notation as in Section 2. We take a modulus $M = p^\alpha$ with a prime p and $\alpha \geq 2$, and in addition to the assumptions in Section 2 we suppose that $(\lambda - 1)y_0 + r \neq 0$. We write $p^c \| g$ if p^c is the largest power of p dividing the nonzero integer g. We define κ and ω by $p^\kappa \| (\lambda - 1)$ and $p^\omega \| ((\lambda - 1)y_0 + r)$. For an integer m relatively prime to λ we let ord (λ, m) be the least positive integer n such that $\lambda^n \equiv 1 \bmod m$. Then we define β as follows: if p is odd, then $p^\beta \| (\lambda^{\operatorname{ord}(\lambda, p)} - 1)$; if $p = 2$, then $2^\beta \| (\lambda^{\operatorname{ord}(\lambda, 4)} - 1)$. According to [13, p. 131] we have

$$\operatorname{ord}(\lambda, p^{\beta+j}) = p^j \operatorname{ord}(\lambda, p^\beta) \quad \text{for all } j \geq 0. \tag{15}$$

We set $\gamma = \beta + \omega - \kappa$, and we note that $\kappa \leq \beta$ implies $\gamma \geq 0$. For a dimension $s \geq 1$, a positive integer b and an arbitrary $k > 1$ we define

$$P_k^{(s)}(\lambda, b) = \sum R(\boldsymbol{h})^{-k},$$

where the sum is extended over all nonzero $\boldsymbol{h} = (h_0, \ldots, h_{s-1}) \in \mathbb{Z}^s$ satisfying

$$\sum_{t=0}^{s-1} h_t \lambda^t \equiv 0 \bmod b.$$

Theorem 1. *Let $M = p^\alpha$ be a modulus with a prime p and $\alpha \geq 2$. Suppose the conditions above are satisfied and also $\gamma < \alpha$. Let the nodes $\boldsymbol{x}_n$ be derived from the corresponding*

linear congruential PRN *according to* (2). *Then for any* $f \in E^k(A)$ *we have*

$$\left| \int_{I^s} f(\boldsymbol{t})\, \mathrm{d}\boldsymbol{t} - \frac{1}{\tau} \sum_{n=0}^{\tau-1} f(\boldsymbol{x}_n) \right| \leq A P_k^{(s)}(\lambda, p^{\alpha-\gamma}).$$

Proof. Let $u_0, u_1, \ldots$ be the sequence defined by $u_0 = y_0$ and $u_{n+1} = \lambda u_n + r$ for $n = 0, 1, \ldots$, and put $\boldsymbol{u}_n = (u_n, u_{n+1}, \ldots, u_{n+s-1})$. From the periodicity of f it follows that

$$f(\boldsymbol{x}_n) = f\left(\frac{1}{M} \boldsymbol{u}_n\right) \quad \text{for all } n.$$

Hence by Lemma 1,

$$\begin{aligned} \frac{1}{\tau} \sum_{n=0}^{\tau-1} f(\boldsymbol{x}_n) - \int_{I^s} f(\boldsymbol{t})\, \mathrm{d}\boldsymbol{t} &= \frac{1}{\tau} \sum_{n=0}^{\tau-1} f\left(\frac{1}{M} \boldsymbol{u}_n\right) - c(\boldsymbol{0}) \\ &= \frac{1}{\tau} \sum_{n=0}^{\tau-1} \sum_{\boldsymbol{h}} c(\boldsymbol{h}) e\left(\frac{1}{M} \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle\right) - c(\boldsymbol{0}) \\ &= \frac{1}{\tau} \sum_{\boldsymbol{h} \neq \boldsymbol{0}} c(\boldsymbol{h}) \sum_{n=0}^{\tau-1} e\left(\frac{1}{M} \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle\right). \end{aligned}$$

Using (13) we get

$$\left| \int_{I^s} f(\boldsymbol{t})\, \mathrm{d}\boldsymbol{t} - \frac{1}{\tau} \sum_{n=0}^{\tau-1} f(\boldsymbol{x}_n) \right| \leq A \sum_{\boldsymbol{h} \neq \boldsymbol{0}} R(\boldsymbol{h})^{-k} \left| \frac{1}{\tau} \sum_{n=0}^{\tau-1} e\left(\frac{1}{M} \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle\right) \right|. \tag{16}$$

Now fix $\boldsymbol{h} = (h_0, \ldots, h_{s-1}) \neq \boldsymbol{0}$ and put $z_n = \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle$. Then

$$z_{n+1} = \sum_{i=0}^{s-1} h_i u_{n+1+i} = \sum_{i=0}^{s-1} h_i (\lambda u_{n+i} + r) = \lambda z_n + r' \quad \text{for all } n, \tag{17}$$

where

$$r' = r \sum_{i=0}^{s-1} h_i.$$

Furthermore,

$$(\lambda - 1) z_0 + r' = z_1 - z_0 = \sum_{i=0}^{s-1} h_i (u_{i+1} - u_i),$$

and a straightforward induction shows

$$u_{n+1} - u_n = \lambda^n ((\lambda - 1) y_0 + r) \quad \text{for all } n,$$

hence

$$(\lambda - 1) z_0 + r' = \left(\sum_{i=0}^{s-1} h_i \lambda^i \right) ((\lambda - 1) y_0 + r). \tag{18}$$

Now let $\boldsymbol{h}$ be such that

$$\sum_{i=0}^{s-1} h_i \lambda^i \not\equiv 0 \bmod p^{\alpha-\gamma}. \tag{19}$$

If σ is defined by $p^{\sigma} \| \sum_{i=0}^{s-1} h_i \lambda^i$, then $\sigma < \alpha - \gamma$. Since $p^{\omega} \| ((\lambda-1)y_0 + r)$, we have $p^{\sigma+\omega} \| ((\lambda-1)z_0 + r')$ by (18). It follows from (17) and [12, Lemma 8] that the sequence $z_0, z_1, \ldots$ satisfies

$$\tau(p^{\alpha-1}) = \operatorname{ord}(\lambda, p^{\alpha-1-\sigma-\omega+\kappa}) = \operatorname{ord}(\lambda, p^{\beta+\alpha-\gamma-\sigma-1}),$$

and so

$$\tau(p^{\alpha-1}) = p^{\alpha-\gamma-\sigma-1} \operatorname{ord}(\lambda, p^{\beta})$$

by (15). Similarly, one shows that

$$\tau(M) = \tau(p^{\alpha}) = p^{\alpha-\gamma-\sigma} \operatorname{ord}(\lambda, p^{\beta}) = p\tau(p^{\alpha-1}).$$

Thus Lemma 2 implies

$$\sum_{n=0}^{\tau(M)-1} e\left(\frac{z_n}{M}\right) = 0.$$

From the definition of u_n it follows that $u_n \equiv y_n \bmod M$, hence $u_{n+\tau} \equiv u_n \bmod M$ and $z_{n+\tau} \equiv z_n \bmod M$ for all n. Consequently, τ is a multiple of $\tau(M)$, and so

$$\sum_{n=0}^{\tau-1} e\left(\frac{1}{M}\langle \boldsymbol{h}, \boldsymbol{u}_n \rangle\right) = \sum_{n=0}^{\tau-1} e\left(\frac{z_n}{M}\right) = 0$$

whenever $\boldsymbol{h}$ satisfies (19). Now we go back to (16) and use the identity above as well as the trivial bound for the exponential sums corresponding to those $\boldsymbol{h}$ not satisfying (19). This yields the result of the theorem.

We note the values of the crucial parameter γ in Theorem 1 for the special cases considered in Section 2. If M is a power of 2, r odd, and $\lambda \equiv 5 \bmod 8$, then $\gamma = 0$; if M is a power of 2, $r = 0$, y_0 odd, and $\lambda \equiv 5 \bmod 8$, then $\gamma = 2$. Thus in the upper bound in Theorem 1 we have either $P_k^{(s)}(\lambda, M)$ or $P_k^{(s)}(\lambda, M/4)$.

It is an important fact that $P_k^{(s)}(\lambda, b)$ can be bounded from above in terms of the figure of merit $\rho^{(s)}(\lambda, b)$ defined in (5). This can be done either by the method of Hua and Wang [6, pp. 143-145] or that of Zaremba [22, pp. 90-96], and in both cases we get an inequality of the form

$$P_k^{(s)}(\lambda, b) \leq C(k, s) \rho^{-k} (1 + \log \rho)^{s-1} \quad \text{for } s \geq 2, \tag{20}$$

where $\rho = \rho^{(s)}(\lambda, b)$ and the constant $C(k, s)$ depends only on k and s. From the information on figures of merit in Section 2 we see that we can expect to obtain values of $\rho^{(s)}(\lambda, b)$ that are of the order of magnitude b, up to logarithmic factors. For such choices of λ, (20) shows that $P_k^{(s)}(\lambda, b)$ is at most of the order of magnitude b^{-k}, up to logarithmic factors. According to Theorem 1, we get then a bound for the integration error of the same order of magnitude.

In the special cases considered above, namely M a power of 2, $\lambda \equiv 5 \bmod 8$, and r odd, or $r = 0$ and y_0 odd, it follows that if λ yields a large value of $\rho^{(s)}(\lambda, M)$ resp. $\rho^{(s)}(\lambda, M/4)$, then the bound for the integration error is of the order of magnitude M^{-k}, up to logarithmic factors. This error bound is the better the more regular f is, and is obviously a significant improvement on the error bounds in Section 2.

For prime moduli M the improvement for periodic integrands is less dramatic. We consider only the case where the multiplier λ yields the largest possible value of τ, namely when λ is a primitive root mod M. We suppose also that $(\lambda - 1)y_0 + r \not\equiv 0 \bmod M$. Then it follows from well-known results (see e.g. [12, Lemma 8]) that $\tau = M - 1$.

Theorem 2. *Let M be a prime modulus and suppose the conditions above are satisfied. Let the nodes $\boldsymbol{x}_n$ be derived from the corresponding linear congruential* PRN *according to* (2). *Then for any $f \in E^k(A)$ we have*

$$\left| \int_{I^s} f(\boldsymbol{t})\, d\boldsymbol{t} - \frac{1}{\tau} \sum_{n=0}^{\tau-1} f(\boldsymbol{x}_n) \right| \leq A\left(\frac{B(k, s)}{\tau} + P_k^{(s)}(\lambda, M) \right)$$

with a constant $B(k, s) > 0$ only depending on k and s.

Proof. The argument in the proof of Theorem 1 leading to (16) can be carried out without any changes. For fixed $\boldsymbol{h} = (h_0, \ldots, h_{s-1}) \neq \boldsymbol{0}$ put $z_n = \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle$. Then we have again (17) and (18). Now let $\boldsymbol{h}$ be such that

$$\sum_{i=0}^{s-1} h_i \lambda^i \not\equiv 0 \bmod M. \tag{21}$$

From the condition $(\lambda - 1)y_0 + r \not\equiv 0 \bmod M$ and (18) it follows then that the prime M does not divide $(\lambda - 1)z_0 + r'$. Hence [12, Lemma 8] implies that the sequence $z_0, z_1, \ldots$ satisfies $\tau(M) = \operatorname{ord}(\lambda, M) = M - 1 = \tau$. Consequently, the first $\tau(M)$ terms of the sequence of z_n, considered modulo M, run through all the least residues modulo M except one, thus

$$\left| \sum_{n=0}^{\tau-1} e\left(\frac{1}{M} \langle \boldsymbol{h}, \boldsymbol{u}_n \rangle \right) \right| = \left| \sum_{n=0}^{\tau(M)-1} e\left(\frac{z_n}{M} \right) \right| = 1$$

whenever $\boldsymbol{h}$ satisfies (21). Now we go back to (16) and use the identity above as well as the trivial bound for the exponential sums corresponding to those $\boldsymbol{h}$ not satisfying (21). This yields

$$\left| \int_{I^s} f(\boldsymbol{t})\, d\boldsymbol{t} - \frac{1}{\tau} \sum_{n=0}^{\tau-1} f(\boldsymbol{x}_n) \right| \leq A\left(\frac{1}{\tau} \sum_{\boldsymbol{h} \in \boldsymbol{Z}^s} R(\boldsymbol{h})^{-k} + P_k^{(s)}(\lambda, M) \right)$$

$$= A\left(\frac{1}{\tau} \left(1 + 2 \sum_{h=1}^{\infty} h^{-k} \right)^s + P_k^{(s)}(\lambda, M) \right),$$

where the last identity comes from the proof of Lemma 1. This completes the proof of the theorem.

By the remarks following Theorem 1, a proper choice of λ will yield a value of $P_k^{(s)}(\lambda, M)$ that is of the order of magnitude M^{-k}, up to logarithmic factors. Since $\tau = M - 1$ and $k > 1$, the error bound in Theorem 2 will then only be of the order of magnitude M^{-1}. This is still better, however, than the error bounds in Section 2.

The error bounds for periodic integrands in Theorems 1 and 2 are of such a good quality that, for the purpose of numerical integration, one need not choose the modulus M as large as is customary in pseudorandom number generation. Thus the remarks at the end of Section 2 apply, mutatis mutandis, also to the present situation. We note that for $M = 2^\alpha$, $2 \leq \alpha \leq 17$, tables of parameters λ yielding small values of $P_k^{(s)}(\lambda, M)$ for $k = 2, 4, 6$ and $2 \leq s \leq 8$ were recently compiled by Haber [3]; not all of the parameters in these tables satisfy $\lambda \equiv 5 \bmod 8$.

The fact that nodes derived from linear congruential PRN yield very precise integration methods for periodic integrands makes it worthwhile to consider the question of how to utilize such methods for nonperiodic integrands. Indeed, there is the possibility of employing the device of "periodization", i.e. of replacing a nonperiodic integrand by a periodic integrand without changing the value of the integral. Several periodization techniques are available in the literature; we refer to Hua and Wang [6, Chapter 6] and Zaremba [22, Section 3] for a detailed discussion of such methods. Periodization involves, of course, an additional computational effort, but this effort may well pay off through increased accuracy in the numerical integration method.

5. Comparison with quasirandom points

Sets of N nodes $\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}$ in I^s that are especially designed to give a small value of the discrepancy $D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1})$ are called quasirandom points (compare with Niederreiter [16, Section 3]). Such nodes will produce a small error bound in Theorem A. We discuss briefly the constructions that are available for dimension $s \geq 2$ and compare the results with those in Sections 2 and 3.

We consider only constructions that pertain to the case where the number N of nodes is prescribed. There are also constructions of infinite sequences of points in I^s such that any initial segment has small discrepancy (see [2; 16, Section 3]). Such sequences are useful in cases where one does not want to fix N in advance.

A classical construction of quasirandom points is that of Hammersley [4]. For integers $b \geq 2$ and $n \geq 0$ let

$$n = \sum_{i=0}^{k} a_i b^i$$

be the digit expansion of n in the base b and define

$$\phi_b(n) = \sum_{i=0}^{k} a_i b^{-i-1}.$$

Now choose bases $b_1, \ldots, b_{s-1}$ that are pairwise relatively prime and set

$$\boldsymbol{w}_n = \left(\phi_{b_1}(n), \ldots, \phi_{b_{s-1}}(n), \frac{n}{N} \right) \in I^s \quad \text{for } n = 0, 1, \ldots, N-1.$$

For these points we have

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) \leq C(b_1, \ldots, b_{s-1}) N^{-1} (\log N)^{s-1} \quad \text{for } N \geq 2,$$

with a constant $C(b_1, \ldots, b_{s-1})$ that is given roughly by

$$C(b_1, \ldots, b_{s-1}) \approx \prod_{i=1}^{s-1} \frac{b_i}{\log b_i}.$$

It is conjectured that the order of magnitude $N^{-1}(\log N)^{s-1}$ for the discrepancy of N points in I^s cannot be improved upon, and a proof of this conjecture is known for $s = 1$ and $s = 2$ (see the discussion in [16, Section 3]).

A recent construction of Faure [2] achieves the same order of magnitude for the discrepancy, but with a constant that is much smaller than $C(b_1, \ldots, b_{s-1})$. Let p be the least prime $\geq s$, and for $n \geq 0$ let

$$n = \sum_{i=0}^{k} a_i p^i$$

be the digit expansion of n in the base p. For $1 \leq h \leq s-1$ define

$$w_n^{(h)} = \sum_{i=0}^{k} c_i p^{-i-1},$$

where

$$c_i \equiv \sum_{j=i}^{k} (h-1)^{j-i} \binom{j}{i} a_j \mod p \quad \text{and} \quad 0 \leq c_i < p.$$

Then set

$$\boldsymbol{w}_n = \left(w_n^{(1)}, \ldots, w_n^{(s-1)}, \frac{n}{N} \right) \in I^s \quad \text{for } n = 0, 1, \ldots, N-1.$$

For these points we have

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) \leq B(s) N^{-1} (\log N)^{s-1} \quad \text{for } N \geq 2,$$

where the constant $B(s)$ is given roughly by

$$B(s) \approx \frac{1}{s!} \left(\frac{s}{\log s} \right)^s.$$

From Stirling's formula it follows immediately that $\lim_{s \to \infty} B(s) = 0$.

The constructions of Hammersley and Faure have the disadvantage that the calculation of the points is rather timeconsuming. There are, however, quasirandom points that can be calculated quickly and for which the discrepancy is only slightly larger than in the cases above. Let $g = (g_0, \ldots, g_{s-1}) \in \mathbb{Z}^s$ be a fixed lattice point and consider the points

$$\boldsymbol{w}_n = \left(\left\{\frac{n}{N} g_0\right\}, \ldots, \left\{\frac{n}{N} g_{s-1}\right\}\right) \in I^s \quad \text{for } n = 0, 1, \ldots, N-1, \tag{22}$$

where $\{t\} = t - \lfloor t \rfloor$ denotes the fractional part of t. The following gives an improvement on a result of the author [14].

Theorem 3. *For every integer $N \geq 2$ and every dimension $s \geq 2$ there exists a lattice point $g \in \mathbb{Z}^s$ with coordinates relatively prime to N such that the points in* (22) *satisfy*

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) < \frac{1}{N}(0.9 + \log N)^s + \frac{A(s)}{N},$$

where $A(s) = s - 1$ if N is a prime power and

$$A(s) = (1.24)(1.09)^s s + (1.16)(1.17)^s s + s - 1$$

if N is not a prime power.

Proof. With $C_s(N)$ denoting the set of lattice points defined in Section 1, we can apply [13, Lemma 2.2] to any points $\boldsymbol{w}_n$ of the form (22) and obtain

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) \leq \frac{s}{N} + \sum_{\boldsymbol{h} \in C_s(N)} \frac{1}{r(\boldsymbol{h}, N)} \left| \frac{1}{N} \sum_{n=0}^{N-1} e\left(\frac{1}{N} \langle \boldsymbol{h}, n\boldsymbol{g} \rangle\right) \right|,$$

where for $\boldsymbol{h} = (h_0, \ldots, h_{s-1}) \in C_s(N)$ we define

$$r(\boldsymbol{h}, N) = \prod_{i=0}^{s-1} r(h_i, N)$$

with

$$r(h, N) = \begin{cases} 1 & \text{if } h = 0, \\ N \sin \dfrac{\pi |h|}{N} & \text{if } h \neq 0. \end{cases}$$

Now

$$\sum_{n=0}^{N-1} e\left(\frac{1}{N} \langle \boldsymbol{h}, n\boldsymbol{g} \rangle\right) = \sum_{n=0}^{N-1} e\left(\frac{n}{N} \langle \boldsymbol{h}, \boldsymbol{g} \rangle\right)$$

has the value N if $\langle \boldsymbol{h}, \boldsymbol{g} \rangle \equiv 0 \bmod N$ and 0 otherwise. Therefore

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) \leq \frac{s}{N} + A(\boldsymbol{g}, N)$$

with

$$A(\boldsymbol{g}, N) = \sum \frac{1}{r(\boldsymbol{h}, N)},$$

where the sum is extended over all $\boldsymbol{h} \in C_s(N)$ with $\langle \boldsymbol{h}, \boldsymbol{g} \rangle \equiv 0 \bmod N$. From $\sin \pi t \geqslant 2t$ for $0 \leqslant t \leqslant \frac{1}{2}$ it follows that $r(h, N) \geqslant r(2h) := \max(1, 2|h|)$ for $|h| \leqslant N/2$, hence $r(\boldsymbol{h}, N) \geqslant R(2\boldsymbol{h})$ for all $\boldsymbol{h} \in C_s(N)$ with the notation in (4). Consequently,

$$D_N(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}) \leqslant \frac{s}{N} + B(\boldsymbol{g}, N) \tag{23}$$

with

$$B(\boldsymbol{g}, N) = \sum_{\substack{\boldsymbol{h} \in C_s(N) \\ \langle \boldsymbol{h}, \boldsymbol{g} \rangle \equiv 0 \bmod N}} R(2\boldsymbol{h})^{-1}.$$

Let G be the set of lattice points $\boldsymbol{g} \in C_s(N)$ for which each coordinate is relatively prime to N. The cardinality of G is $\phi(N)^s$, where ϕ is Euler's totient function. We consider

$$M(N) = \frac{1}{\phi(N)^s} \sum_{\boldsymbol{g} \in G} B(\boldsymbol{g}, N). \tag{24}$$

It follows then as in [14, pp. 214-215] that

$$M(N) = \frac{1}{N} \sum_{j=0}^{N-1} \left(\frac{1}{\phi(N)} \sum_{h \in C(N)} \sum_{\substack{g \in C(N) \\ \gcd(g, N) = 1}} e\left(\frac{j}{N} hg\right) r(2h)^{-1} \right)^s - 1.$$

The contribution from $j = 0$ to the sum above is $(1 + \frac{1}{2}L(N))^s$, where

$$L(N) = \sum_{h \in C_1(N)} |h|^{-1}.$$

Therefore,

$$M(N) = \frac{1}{N}(1 + \tfrac{1}{2}L(N))^s + \frac{1}{N} \sum_{j=1}^{N-1} \left(\frac{1}{\phi(N)} \sum_{h \in C(N)} \sum_{\substack{g \in C(N) \\ \gcd(g, N) = 1}} e\left(\frac{j}{N} hg\right) r(2h)^{-1} \right)^s - 1. \tag{25}$$

For fixed j, $1 \leqslant j \leqslant N - 1$, we consider

$$S_j = \sum_{h \in C(N)} \sum_{\substack{g \in C(N) \\ \gcd(g, N) = 1}} e\left(\frac{j}{N} hg\right) r(2h)^{-1}.$$

Separating the contribution from $h = 0$, we get

$$S_j = \phi(N) + \tfrac{1}{2} \sum_{h \in C_1(N)} \sum_{\substack{g \in C(N) \\ \gcd(g, N) = 1}} e\left(\frac{j}{N} hg\right) |h|^{-1}.$$

Decompose $b = \gcd(N, j)$ in the form $b = b_1 b_2$, where b_2 is the product of those highest prime powers in the canonical factorization of N that also divide b (we can

have $b_2 = 1$). Then it follows as in [14, pp. 211–213] that

$$S_j = \phi(N) + \tfrac{1}{2} b_1 \phi(b_2) \sum_d \mu(d) L(bd),$$

where μ is the Moebius function and the sum is extended over all positive divisors d of $p_1 \cdots p_u$, with $p_1, \ldots, p_u$ being the distinct prime factors of N that do not divide b_2. As in [14, p. 213] one shows that

$$|S_j| < \phi(N) \qquad \text{if } u = 1 \tag{26}$$

and

$$|S_j| < \phi(N) + \frac{1.35}{b} \quad \text{if } u \geq 2. \tag{27}$$

If N is a prime power, then always $u = 1$, and so from (25) and (26),

$$M(N) < \frac{1}{N}(1 + \tfrac{1}{2} L(N))^s + \frac{N-1}{N} - 1 = \frac{1}{N}(1 + \tfrac{1}{2} L(N))^s - \frac{1}{N}. \tag{28}$$

If N is not a prime power, we can always apply the worse estimate (27) for $|S_j|$. We note also that for each proper divisor b of N there are exactly $\phi(N/b)$ values of j with $1 \leq j \leq N-1$ and $\gcd(N, j) = b$. Hence from (25),

$$M(N) < \frac{1}{N}(1 + \tfrac{1}{2} L(N))^s + \frac{1}{N} \sum_{\substack{b \mid N \\ b < N}} \phi\left(\frac{N}{b}\right)\left(1 + \frac{1.35}{\phi(N) b}\right)^s - 1. \tag{29}$$

By [14, equation (31)] we have

$$L(N) < 2 \log N - 0.2 \quad \text{for } N \geq 6, \tag{30}$$

and so for a prime power $N \geq 6$ it follows from (28) that

$$M(N) < \frac{1}{N}(0.9 + \log N)^s - \frac{1}{N}.$$

The definition of $M(N)$ in (24) implies that there exists a $g \in G$ with

$$B(g, N) < \frac{1}{N}(0.9 + \log N)^s - \frac{1}{N}.$$

The result of the theorem for prime powers $N \geq 6$ follows now from (23). For $2 \leq N \leq 5$ the result of the theorem is trivial, since always $D_N(w_0, \ldots, w_{N-1}) \leq 1$, whereas

$$\frac{1}{N}(0.9 + \log N)^s + \frac{s-1}{N} > 1.$$

It remains to consider values of N that are not prime powers. From (29) we get

$$M(N) < \frac{1}{N}(1+\tfrac{1}{2}L(N))^s + \frac{1}{N}\sum_{\substack{b|N \\ b<N}} \phi\left(\frac{N}{b}\right) + \frac{1}{N}\sum_{\substack{b|N \\ b<N}} \phi\left(\frac{N}{b}\right) \times \left[\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1\right] - 1$$

$$= \frac{1}{N}(1+\tfrac{1}{2}L(N))^s + \frac{1}{N}\sum_{\substack{b|N \\ b<N}} \phi\left(\frac{N}{b}\right)\left[\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1\right] - \frac{1}{N}. \quad (31)$$

Now assume that $\phi(N) \geq 8$. Then for $b \geq 2$,

$$\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1 < \frac{1.35}{\phi(N)b}\left(1+\frac{1.35}{16}\right)^{s-1} s < \frac{(1.24)(1.09)^s s}{\phi(N)b}$$

and for $b = 1$,

$$\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1 < \frac{1.35}{\phi(N)}\left(1+\frac{1.35}{8}\right)^{s-1} s < \frac{(1.16)(1.17)^s s}{\phi(N)}.$$

Consequently,

$$\sum_{\substack{b|N \\ b<N}} \phi\left(\frac{N}{b}\right)\left[\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1\right]$$

$$< \frac{(1.24)(1.09)^s s}{\phi(N)} \sum_{\substack{b|N \\ 1<b<N}} \phi\left(\frac{N}{b}\right) b^{-1} + (1.16)(1.17)^s s$$

$$= \frac{(1.24)(1.09)^s s}{\phi(N)N} \sum_{\substack{b|N \\ 1<b<N}} \phi(b)b + (1.16)(1.17)^s s.$$

Now

$$\frac{1}{\phi(N)N} \sum_{\substack{b|N \\ 1<b<N}} \phi(b)b < \frac{1}{\phi(N)N}\sum_{b|N} \phi(b)b - 1 < 1$$

according to [14, equation (27)], hence

$$\sum_{\substack{b|N \\ b<N}} \phi\left(\frac{N}{b}\right)\left[\left(1+\frac{1.35}{\phi(N)b}\right)^s - 1\right] < (1.24)(1.09)^s s + (1.16)(1.17)^s s.$$

Going back to (31) and using (30), we obtain

$$M(N) < \frac{1}{N}(1+\tfrac{1}{2}L(N))^s + \frac{A(s)-s}{N} < \frac{1}{N}(0.9+\log N)^s + \frac{A(s)-s}{N}.$$

The definition of $M(N)$ in (24) implies that there exists a $g \in G$ with

$$B(g, N) < \frac{1}{N}(0.9 + \log N)^s + \frac{A(s) - s}{N}.$$

The result of the theorem for values of N that are not prime powers and satisfy $\phi(N) \geq 8$ follows now from (23). The values of N with $\phi(N) < 8$ that are not prime powers are exactly $N = 6$, 10, 12, 14, and 18. For these values of N the result of the theorem is trivial, since always $D_N(w_0, \ldots, w_{N-1}) \leq 1$, whereas

$$\frac{1}{N}(0.9 + \log N)^s + \frac{A(s)}{N} > 1.$$

The theorem is thus shown in all cases.

The proof of Theorem 3 is not constructive. There are, however, search algorithms for lattice points g for which the points in (22) have a small discrepancy; see Hua and Wang [6, pp. 95-97] and Korobov [8]. The book of Hua and Wang [6] also contains tables of such lattice points for dimensions $2 \leq s \leq 18$.

From the results in this section we see that the lowest order of magnitude that can be achieved for the discrepancy of N points in I^s is $N^{-1}(\log N)^{s-1}$, or $N^{-1}(\log N)^s$ if a simple construction is used. A comparison with the results in Sections 2 and 3 shows that, with a proper choice of parameters in pseudorandom number generation, nodes derived from PRN come very close to these lowest orders of magnitude. In fact, their discrepancy may only differ by a logarithmic factor from the lowest order of magnitude $N^{-1}(\log N)^{s-1}$. As noted in Section 1, nodes derived from PRN can, however, be calculated very quickly.

Note added in proof. Further information on the discrepancy of digital PRN can be found in H. Niederreiter, "The serial test for digital k-step pseudorandom numbers", to appear.

References

[1] I. Borosh and H. Niederreiter, "Optimal multipliers for pseudo-random number generation by the linear congruential method", *BIT* 23 (1983) 65-74.

[2] H. Faure, "Discrépance de suites associées à un système de numération (en dimension s)", *Acta Arithmetica* 41 (1982) 337-351.

[3] S. Haber, "Parameters for integrating periodic functions of several variables", *Mathematics of Computation* 41 (1983) 115-129.

[4] J.M. Hammersley, "Monte Carlo methods for solving multivariable problems", *Annals of the New York Academy of Sciences* 86 (1960) 844-874.

[5] E. Hlawka, "Funktionen von beschränkter Variation in der Theorie der Gleichverteilung", *Annali di Matematica Pura ed Applicata* (4) 54 (1961) 325-333.

[6] L.K. Hua and Y. Wang, *Applications of number theory to numerical analysis* (Springer-Verlag, Berlin, 1981).

[7] D.E. Knuth, *The art of computer programming. Vol. 2: Seminumerical algorithms, 2nd ed.* (Addison-Wesley, Reading, MA, 1981).

[8] N.M. Korobov, "On the computation of optimal coefficients" (in Russian), *Doklady Akademia Nauk SSSR* 267 (1982) 289-292.

[9] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences* (Wiley-Interscience, New York, 1974).

[10] R. Lidl and H. Niederreiter, *Finite fields*, Encyclopedia of mathematics and its applications, Vol. 20 (Addison-Wesley, Reading, MA, 1983).

[11] D. Maisonneuve, "Recherche et utilisation des 'bons treillis'. Programmation et résultats numériques", in: S.K. Zaremba, ed., *Applications of number theory to numerical analysis* (Academic Press, New York, 1972) pp. 121-201.

[12] H. Niederreiter, "On the distribution of pseudo-random numbers generated by the linear congruential method. III", *Mathematics of Computation* 30 (1976) 571-597.

[13] H. Niederreiter, "Pseudo-random numbers and optimal coefficients", *Advances in Mathematics* 26 (1977) 99-181.

[14] H. Niederreiter, "Existence of good lattice points in the sense of Hlawka", *Monatshefte für Mathematik* 86 (1978) 203-219.

[15] H. Niederreiter, "The serial test for linear congruential pseudo-random numbers", *Bulletin of the American Mathematical Society* 84 (1978) 273-274.

[16] H. Niederreiter, "Quasi-Monte Carlo methods and pseudo-random numbers", *Bulletin of the American Mathematical Society* 84 (1978) 957-1041.

[17] H. Niederreiter, "Applications des corps finis aux nombres pseudo-aléatoires", in: Séminaire de théorie des nombres 1982-1983 (Université de Bordeaux I, Talence, 1983) exposé 38.

[18] H. Niederreiter, "The performance of k-step pseudorandom number generators under the uniformity test", *SIAM Journal on Scientific and Statistical Computing* 5 (1984) 798-810.

[19] H. Niederreiter, "The serial test for pseudo-random numbers generated by the linear congruential method", *Numerische Mathematik* 46 (1985) 51-68.

[20] R.C. Tausworthe, "Random numbers generated by linear recurrence modulo two", *Mathematics of Computation* 19 (1965) 201-209.

[21] S.K. Zaremba, "Some applications of multidimensional integration by parts", *Annales Polonici Mathematicae* 21 (1968) 85-96.

[22] S.K. Zaremba, "La méthode des 'bons treillis' pour le calcul des intégrales multiples", in: S.K. Zaremba, ed., *Applications of number theory to numerical analysis* (Academic Press, New York, 1972) pp. 39-119.

Mathematical Programming Study 27 (1986) 39-53
North-Holland

A TIGHT UPPER BOUND FOR THE EXPECTATION OF A CONVEX FUNCTION OF A MULTIVARIATE RANDOM VARIABLE

H. GASSMANN and W.T. ZIEMBA
University of British Columbia, Vancouver, Canada

Received 22 September 1983
Revised manuscript received 20 September 1984

Edmundson and Madansky have developed an upper bound for the expectation of a convex function of a multivariate random variable. The bound exists under rather general conditions. However, it is computable for $n \geq 3$ only when the random variables are independent. This paper develops a procedure for computing this bound using a linear program. We also show how to extend and sharpen the bound by utilizing it on subsets of the sample space. Moreover, the bound is applicable to arbitrary convex domains which may be unbounded. The new bounds along with Jensen's Inequality may be applied on subsets using the procedures developed by Huang, Ziemba, and Ben-Tal to yield a procedure for obtaining upper and lower bounds on the expectation of a convex function of a multivariate random variable to an arbitrary degree of accuracy. The results are useful in a wide variety of optimization applications. Some numerical work is provided to illustrate the use of the new bounds.

Key words: Upper Bounds, Convex Functions, Approximations for Stochastic Programs.

1. Introduction

The problem of finding the expected value of a convex function of a multivariate random variable has received considerable attention in the literature [1, 2, 8-12]. In principle, it requires the evaluation of a multiple integral of the form

$$\bar{\phi} \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(X_1, \ldots, X_n)\, \mathrm{d}F(X_1, \ldots, X_n), \tag{1}$$

where ϕ is a convex function, $\mathbb{R}^n \to \mathbb{R}$ and F is a probability distribution on $\mathbb{R}^n$. The solution of (1) presents enormous computational difficulties if n is larger than about 5, particularly when ϕ is difficult to evaluate.

This paper extends an idea of Edmundson and Madansky [10] to find upper bounds for (1), which, together with the known lower bounds arising from the use of Jensen's inequality [9], can be used to find bounds on $\bar{\phi}$, sharp in the sense that the new upper bound is the best bound available with only first moment information on the probability distribution. The new bounds can be calculated for arbitrary

This research was partially supported by Natural Sciences and Engineering Research Council of Canada grant 67-7147. Without implicating them, the authors would like to thank R. Wets and the referees for the helpful comments on an earlier version of this paper.

distributions, where previously the Madansky bounds were only available when the X_i were independent. See [8, 12] for surveys of this work.

The plan of the paper is as follows. Section 2 reviews the basic ideas of Edmundson-Madansky bound construction in the univariate case and then develops the new bound in the multivariate case. Section 3 provides numerical examples to illustrate the bound. The extension of the bound to subsets to make it arbitrarily sharp is discussed in Section 4. The generalization to unbounded domains appears in Section 5. Examples illustrating the upper bound as well as Jensen's Inequality lower bounds on subsets are discussed in Section 6. A computer code for performing the bounds calculations for bivariate normal distributions is described in Section 7. Final conclusions are discussed in Section 8.

2. Upper bounds

Let $X \equiv (X_1, \ldots, X_n)'$ be a multivariate random variable on the probability space (Ω, Σ, P) with distribution function F and finite mean $\bar{x}$. Let U denote the support of X, i.e., the smallest closed subset contained in $\mathbb{R}^n$ such that $\int_U \mathrm{d}F = 1$.

Assume U is convex, and let $\phi: U \to \mathbb{R}$ be a convex function which is integrable with respect to $\mathrm{d}F$, so that the expected value $\bar{\phi} := E_x\phi(x)$ exists and is finite.

Suppose first that U is compact.

A discussion of upper bounds in the case $n = 1$ is contained in [2]. In this case U is the interval $[a, b]$. This implies the estimate, see Fig. 1,

$$\bar{\phi} \leq \frac{(b-\bar{x})\phi(a)+(\bar{x}-a)\phi(b)}{b-a}.$$

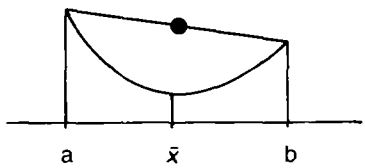


Fig. 1. Madansky's upper bound in one dimension.

It is easy to extend this result to the case of a random vector ($n > 1$) with *independent* components. For U is an n-dimensional rectangle of the form

$$U = \prod_{i=1}^{n} [a_i, b_i]$$

and independence then yields the following estimate [9]:

$$\bar{\phi} \leq \sum_{i_1=0}^{1} \cdots \sum_{i_n=0}^{1} \left(\prod_{k=1}^{n} \delta_{i_k}^k \right) \phi(d_{i_1}^1, \ldots, d_{i_n}^n) \tag{2}$$

where $d_0^i = a_i$, $d_1^i = b_i$, $\delta_0^k = (b_k - \bar{x}_k)/(b_k - a_k)$ and $\delta_1^k = (\bar{x}_k - a_k)/(b_k - a_k)$.

If the X_i are correlated, one has to proceed differently.

Let $\operatorname{Gr}\phi$ denote the graph of ϕ, i.e., the subset of $\mathbb{R}^{n+1}$ given by $\operatorname{Gr}\phi = \{(x, y) | x \in U, y = \phi(x)\}$, and let H be its convex hull. Then $(\bar{x}, \bar{\phi}) \in H$ whenever $\bar{x}$ and $\bar{\phi}$ both exist, since integration with respect to a probability measure can be viewed as simply taking convex combinations. Hence one immediately has the upper bound $\bar{\phi} \leq \sup\{y | (\bar{x}, y) \in H\}$. For the case of an n-dimensional rectangle this was pointed out by Madansky [10], whose notation $H^*(x) := \sup\{y | (x, y) \in H\}$ we adopt.

The following result gives a computable upper bound on $\bar{\phi}$ regardless of the shape of U.

Theorem 1. *Let V be a bounded convex polyhedron containing U and let $\{v_1, \ldots, v_m\}$ be the extreme points of V. Assume ϕ can be extended to a convex function mapping V into $\mathbb{R}$. (This extended map will be denoted by ϕ as well.) Then*

$$\bar{\phi} \leq z := \max\left\{ \sum_{i=1}^{m} \phi(v_i)\lambda_i \,\middle|\, \lambda_i \geq 0, \sum \lambda_i = 1, \sum \lambda_i v_i = \bar{x} \right\}.$$

Proof. Let $(x, y) \in H$. Then there exist k, $\alpha_i \geq 0$ and $u_i \in U$, $i = 1, \ldots, k$ such that $\sum_{i=1}^{k} \alpha_i = 1$, $\sum_{i=1}^{k} \alpha_i u_i = x$, $\sum_{i=1}^{k} \alpha_i \phi(u_i) = y$. Now for every i, there exist $\mu_{ij} > 0$, $j = 1, \ldots, m$ such that $\sum_j^m \mu_{ij} = 1$, $\sum_{j=1}^{m} \mu_{ij} v_j = u_i$, i.e., $x = \sum_i \sum_j \alpha_i \mu_{ij} v_j$. Since ϕ is convex

$$y = \sum_{i=1}^{k} \phi(u_i)\alpha_i = \sum_i \alpha_i \phi\left(\sum_j \mu_{ij} v_j\right) \leq \sum_i \sum_j \alpha_i \mu_{ij} \phi(v_j).$$

Setting $\hat{\lambda}_j = \sum_i \alpha_i \mu_{ij}$ yields $x = \sum_j \hat{\lambda}_j v_j$, $\sum_j \hat{\lambda}_j = 1$, $\hat{\lambda}_j \geq 0$, and

$$y \leq \sum_j \hat{\lambda}_j \phi(v_j) \leq \max\left\{ \sum_j \phi(v_j)\lambda_j \,\middle|\, \lambda_j \geq 0, \sum \lambda_j = 1, \sum \lambda_j v_j = x \right\}.$$

In particular, this holds for $(x, y) = (\bar{x}, H^*(\bar{x}))$ (which is an element of H). □

Remarks. 1. If U is a polyhedral set, then $H^*(\bar{x}) = z$. This holds in particular for n-dimensional rectangles, so Theorem 1 subsumes Madansky's result as a special case. Figure 2 illustrates the upper bound when $n = 2$.

2. If $n = 1$, then $U = [a, b]$, so $z = \max\{\lambda_1\phi(a) + \lambda_2\phi(b) | \lambda_1 + \lambda_2 = 1, \lambda_1 a + \lambda_2 b = \bar{x}, \lambda_1, \lambda_2 \geq 0\}$. This implies $\lambda_1 = (b - \bar{x})/(b - a)$, $\lambda_2 = 1 - \lambda_1 = (\bar{x} - a)/(b - a)$, whence $z = (b - \bar{x})/(b - a)\phi(a) + (\bar{x} - a)/(b - a)\phi(b)$, which is the usual Edmundson-Madansky bound.

3. When V is a simplex, there is a *unique* convex combination of the extreme points $\{v_i\}$ which yields $\bar{x}$. The maximization in Theorem 1 is therefore trivial.

4. Birge and Wets [3; Proposition 5.1] give a related result using a family $\{\nu_x(\cdot), x \in U\}$ of probability measures on ext U, the set of extreme points of U, such that

$$\int_{\text{ext } U} s\nu_x(ds) = x \quad \text{for all } x \in U \tag{3}$$

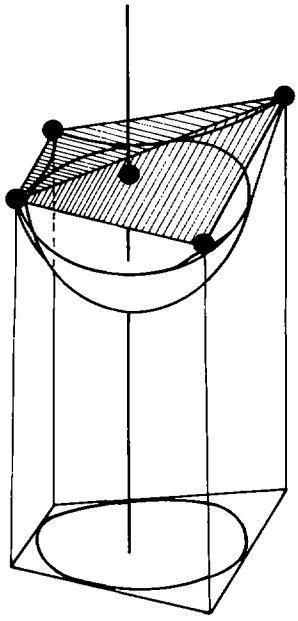


Fig. 2. The new bound when $n = 2$.

This condition makes their result more difficult to apply, although it is in general sharper than Theorem 1. The λ_i of Theorem 1 can be interpreted as a probability measure on ext $V = \{v_1, \ldots, v_m\}$ such that (3) holds at $x = \bar{x}$ only.

3. Numerical examples

$$(1) \qquad \phi(x, y) = e^x, \qquad f(x, y) = \begin{cases} 3(\sqrt{1 - x^2 - y^2})/2\pi & \text{if } x^2 + y^2 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then $U = \{(x, y) \mid x^2 + y^2 \leq 1\}$ and

$$\bar{\phi} = \frac{3}{2\pi} \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} e^x \sqrt{1 - x^2 - y^2}\, dy\, dx = \frac{3}{e} \approx 1.0364.$$

Jensen's inequality yields the lower bound $\bar{\phi} \geq \phi(\bar{x}) = \phi(0, 0) = 1$. There are many polyhedra containing U which give rise to the following upper bounds (see Fig. 3):

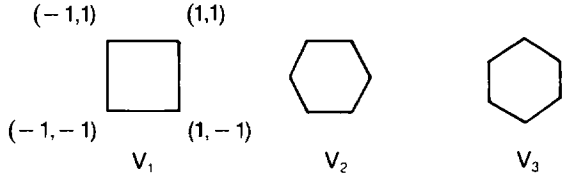


Fig. 3.

$$V_1 = \mathrm{Co}\{(-1,-1),(1,-1),(-1,1),(1,1)\}, \qquad z_1 = \frac{e^2+1}{2e} \approx 1.54308.$$

$$V_2 = \mathrm{Co}\{(-a,-1),(a,-1),(2a,0),(a,1),(-a,1),(-2a,0)\}$$

where $a = \frac{1}{3}\sqrt{3}$,

$$z_2 = \tfrac{1}{2}(e^{2/3\sqrt{3}} + e^{-2/3\sqrt{3}}) \approx 1.74411 > z_1.$$

$$V_3 = \mathrm{Co}\{(-1,-a),(-1,1),(0,2a),(1,a),(1,-a),(0,-2a)\}, \qquad z_3 = z_1.$$

V_3 is obtained from V_2 by a 90 degree rotation. Moreover, the area of V_2 is smaller than the area of V_1 ($2\sqrt{3}$ vs 4 square units) and V_2 seems a 'better' approximation to U, although it yields a less sharp upper bound. V_1 and V_3 provide the same bound.

(2) $$\phi(x,y) = (x+y)^2 + x + 2y, \qquad f(x,y) = \begin{cases} 1/\pi & \text{if } x^2+y^2 \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\bar{\phi} = \frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} ((x+y)^2 + x + 2y)\,dx\,dy = \frac{4\sqrt{2}}{3\pi} \approx 0.60021. \tag{5}$$

Lower bound: $\bar{\phi} \geq \phi(0,0) = 0$.

Using V_1 as in example 1, $z_1 = 4$.

$$V_4 = \mathrm{Co}\{(-1,-1),(\tfrac{1}{2},-1),(1,-\tfrac{1}{2}),(1,1),(-\tfrac{1}{2},1),(-1,\tfrac{1}{2})\}, \qquad z_4 = z_1 = 4.$$

$$V_5 = \mathrm{Co}\{(-\tfrac{1}{2},-1),(1,-1),(1,\tfrac{1}{2}),(\tfrac{1}{2},1),(-1,1),(-1,-\tfrac{1}{2})\}, \qquad z_5 = 2.25.$$

Although V_4 and V_5 have the same shape and area, they yield different upper bounds. See Fig. 4.

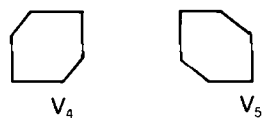


Fig. 4.

Remark. For the independent case the estimates obtained from Theorem 1 are not as sharp as the upper bounds in (2). For example, let

$$\phi(x,y) = (x+y)^2 + x + 2y, \qquad F(x,y) = \begin{cases} 1 & \text{if } 0 \leq x, y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then (2) yields the upper bound $\bar{\phi} \leq \frac{1}{4}[\phi(0,0) + \phi(0,1) + \phi(1,0) + \phi(1,1)] = 3$, whereas one computes, using Theorem 1:

$$\max\{3\lambda_2 + 2\lambda_3 + 7\lambda_4 \mid \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1,\ \lambda_3 + \lambda_4 = \tfrac{1}{2},\ \lambda_2 + \lambda_4 = \tfrac{1}{2}\}$$

$$= \max\{\tfrac{7}{2} - 2\lambda_2 \mid 0 \leq \lambda_2 \leq \tfrac{1}{2}\} = 3.5.$$

Jensen's inequality yields the lower bound $\bar{\phi} \geq \phi(\frac{1}{2},\frac{1}{2}) = 2.5$.

4. Improving the upper bound: Applying it on subsets

The bound can be sharpened by considering a partition of V into convex polyhedra $V^j, j = 1, \ldots, k$ such that $V = \bigcup_{j=1}^{k} V^j$, $V^i \cap V^j = \emptyset$ if $i \neq j$.

Let $\eta_j \equiv \int_{V^j} \mathrm{d}F(x)$, $\bar{x}_j \equiv 1/\eta_j \int_{V^j} x \, \mathrm{d}F(x)$, $j = 1, \ldots, k$, and z_j be the upper bound on $E\{\phi \mid V^j\}$ obtained from Theorem 1. Then we have

Theorem 2

$$\sum_{j=1}^{k} \eta_j \phi(\bar{x}_j) \leq \bar{\phi} \leq \sum_{j=1}^{k} \eta_j z_j. \tag{6}$$

Remarks. If $\eta_j = 0$ set $\bar{x}_j = \bar{x}$ to remove ambiguities. The lower bounds are derived from Jensen's Inequality on subsets [9].

Example (see Figure 5)

$$\phi(x, y) = (x+y)^2 + x + 2y, \qquad f(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$V = [-1, 1] \times [-1, 1], \qquad V^1 = \mathrm{Co}\{(1, 1), (1, -1), -1, 1)\},$$

$$V^2 = \mathrm{Co}\{(1, -1), (-1, 1), (-1, -1)\}.$$

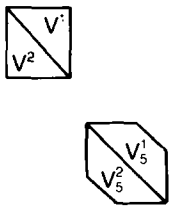


Fig. 5.

Then $\eta_1 = \eta_2 = \frac{1}{2}$, $\bar{x}_1 = 2\sqrt{2}/3\pi\ (1, 1)$, $\bar{x}_2 = -\bar{x}_1$.

$z_1 = 7\sqrt{2}/3\pi$, $z_2 = \sqrt{2}/3\pi$, and hence $\bar{\phi} \leq 8\sqrt{2}/3\pi \approx 1.20042$.

Using instead $\bar{V}^1 = \mathrm{Co}\{(-1, -1), (1, -1), (1, 1)\}$ and $\bar{V}^2 = \mathrm{Co}\{(-1, -1), (-1, 1), (1, 1)\}$ yields $\bar{\phi} \leq 4 - 8\sqrt{2}/3\pi \approx 2.79958$.

Subdividing V_5 into $V_5^1 = \mathrm{Co}\{(-1, 1), (\frac{1}{2}, 1), (1, \frac{1}{2}), (1, -1)\}$ and $V_5^2 = \mathrm{Co}\{(-1, 1), (-1, -\frac{1}{2}), (-\frac{1}{2}, -1), (1, -1)\}$, yields $\bar{\phi} \leq 6\sqrt{2}/3\pi \approx 0.90032$.

5. Generalization to unbounded domains

Throughout this section, assume that the domain U is convex but possibly unbounded and $\int_U \phi \, \mathrm{d}F$ exists. Then the theory of improper integrals implies that

for every $\varepsilon > 0$ there exists a compact convex $\tilde{U} \subset U$ such that $|\int_{U\setminus\tilde{U}} \phi \, dF| < \varepsilon$, $\int_{U\setminus\tilde{U}} dF < \varepsilon$. $A\setminus B$ denotes the complement of B relative to A. The techniques of Sections 2 and 4 can then be applied to $\tilde{U}$ in place of U. One then obtains the upper bound

$$\bar{\phi} = E_U\phi = \int_{\tilde{U}} dF \cdot E\{\phi | \tilde{U}\} + \int_{U\setminus\tilde{U}} dF \cdot E\{\phi | U\setminus\tilde{U}\} \leq E\{\phi | \tilde{U}\} + \varepsilon$$

as well as the lower bound

$$\bar{\phi} \geq (1-\varepsilon)E\{\phi | \tilde{U}\} + \int_{U\setminus\tilde{U}} dF \cdot E\{\phi | U\setminus\tilde{U}\} \geq (1-\varepsilon)E\{\phi | \tilde{U}\} - \varepsilon.$$

Since it is often difficult to find a suitable $\tilde{U}$ for given ε, a different method is desirable. Such a method is available under the following assumptions:

(i) U is contained in (the translate of) a *pointed cone* in $\mathbb{R}^n$. A cone C is pointed if $x \in C$, $-x \in C \Rightarrow x = 0$.

(ii) ϕ satisfies the linear growth condition: $\lim_{t\to\infty} \phi(x+td)/\|x+td\| \leq \delta < \infty$, $\forall x \in U$, $\forall$ directions d of U. Recall that d is a direction of U, $d \in \Delta$, if $x + td \in U$ $\forall x \in U$, $t \geq 0$ and $d \neq 0$.

Theorem 3. *Assume there exists a polyhedron V with extreme points $\{v_1, \ldots, v_k\}$ and extreme directions $\{d_1, \ldots, d_l\}$ such that $U \subset V$. If ϕ can be extended to a convex function on V, and assumptions* (i) *and* (ii) *hold, then*

$$\bar{\phi} \leq \max\left\{ \sum_{i=1}^{k} \lambda_i \phi(v_i) + \sum_{j=1}^{l} \delta\mu_j \|d_j\| \;\middle|\; \lambda \geq 0, \mu_j \geq 0, \right.$$

$$\left. \sum \lambda_i = 1, \sum_i \lambda_i v_i + \sum_j \mu_j d_j = \bar{x} \right\}. \tag{7}$$

Proof. We will show $\forall (x, y) \in H$,

$$y \leq \max\left\{ \sum_i \lambda_i \phi(v_i) + \sum_j \delta\mu_j \|d_j\| \;\middle|\; \lambda_i \geq 0, \mu_j \geq 0, \sum \lambda_i = 1, \sum_i \lambda_i v_i + \sum_j \mu_j d_j = x \right\}.$$

The result then follows from the fact that $(\bar{x}, \bar{\phi}) \in H$.

Let $(x, y) \in H$ and assume without loss of generality that $\|d_j\| = 1, j = 1, \ldots, l$. Then there exist $\gamma_i \geq 0$, $u_i \in U$, $i = 1, \ldots, I$ such that

$$\sum \gamma_i = 1, \qquad \sum \gamma_i u_i = x, \qquad \sum \gamma_i \phi(u_i) = y.$$

Now for all i, there exist $\alpha_{ij} \geq 0$, $\beta_{ij} \geq 0$ such that

$$\sum_j^k \alpha_{ij} = 1, \; u_i = \sum_{j=1}^{k} \alpha_{ij} v_i + \sum_{j=1}^{l} \beta_{ij} d_j.$$

Therefore

$$x = \underbrace{\sum_i \sum_j \gamma_i \alpha_{ij} v_j}_{=: w'} + \underbrace{\sum_i \sum_j \gamma_i \beta_{ij} d_j}_{=: d'}.$$

Note that d is a direction of U for every $i = 1, \ldots, I$. Therefore

$$y = \sum \gamma_i \phi(u_i) = \sum_i \gamma_i \phi\left(\sum_j \alpha_{ij} v_j + \sum_j \beta_{ij} d_j\right)$$

$$\leq \sum_i \left\{ \gamma_i \phi\left(\sum_j \alpha_{ij} v_j\right) + \delta \left\| \sum_j \beta_{ij} d_j \right\| \right\} \leq \sum_i \sum_j \gamma_i \alpha_{ij} \phi(v_j) + \sum_i \sum_j \gamma_i \beta_{ij} \delta$$

by convexity of ϕ and the triangle inequality. Then set

$$\hat{\lambda}_j = \sum_i \gamma_i \alpha_{ij}, \quad j = 1, \ldots, k, \qquad \hat{\mu}_j = \sum_i \gamma_i \beta_{ij}, \quad j = 1, \ldots, l,$$

and observe that

$$\hat{\lambda}_j \geq 0, \qquad \sum_j \hat{\lambda}_j = 1, \qquad \hat{\mu}_j \geq 0, \qquad \sum_{j=1}^{k} \hat{\lambda}_j v_j + \sum_{j=1}^{l} \hat{\mu}_j d_j = x. \tag{8}$$

In particular, then,

$$y \leq \max \sum_{j=1}^{k} \lambda_j \phi(v_j) + \sum_{j=1}^{l} \mu_j \delta \quad \text{over all } \lambda_j, \mu_j \text{ satisfying (8).} \quad \square$$

Remarks. 1. If both d and $-d$ are directions of V, then the linear program (7) is unbounded. This situation is ruled out by assumption (i).

2. For bounded V, the set of extreme direction is empty, i.e. $l = 0$. Theorem 3 then reduces to Theorem 1.

3. In one dimension, V has the form $[a, \infty)$ (or $(-\infty, b]$). Taking $d = 1$, one obtains

$$\max\{\lambda \phi(a) + \mu \delta \mid \lambda \geq 0, \mu \geq 0, \lambda = 1, \lambda a + \mu = \bar{x}\} = \phi(a) + \delta(\bar{x} - a).$$

This expression was derived in [2].

In practice it may be difficult to check whether assumption (ii) is satisfied for every x and d. Assumptions (i), (ii) can be replaced by the following set of assumptions:

(i') V is contained in the translate of an orthant in $\mathbb{R}^n$.

(ii') V satisfies the growth condition:

$$\max_{i=1,\ldots,k} \lim_{t \to \infty} \frac{\phi(v_i + t d_j)}{\| v_i + t d_j \|} = \delta_j < \infty, \quad j = 1, \ldots, l.$$

This leads to a slightly sharper bound:

Theorem 4. *If V and ϕ are as in Theorem 3 and (i') and (ii') hold, then*

$$\bar{\phi} = \max\left\{ \sum_{i=1}^{k} \lambda_i \phi(v_i) + \sum_{j=1}^{l} \mu_j \delta_j \| d_j \| \;\middle|\; \lambda_i \geq 0,\ \mu_j > 0,\ \sum \lambda_i = 1, \sum_i \lambda_i v_i + \sum_j \mu_j d_j = \bar{x} \right\}.$$

The proof depends on two lemmas which are geometrically obvious in two dimensions and easy to prove in general.

Define $\Phi(x, d) = \lim_{t\to\infty} \phi(x+td)/\|x+td\|$ if $x \in U$, $d \in \Delta$, and the limit exists.

Lemma 1. *$\Phi(x, d)$ is defined for all $x \in \mathrm{Co}\{v_i\}$, $d \in \Delta$.*

Lemma 2. *Φ is convex in x and d separately (but not jointly).*

The fact that the angle between any two directions of V is at most $\pi/2$ is crucial in the proof of convexity of Φ in the second variable.

Proof of Theorem 4. Observe that, by convexity of ϕ,

$$\Phi(x, d) = \lim_{t\to\infty} \frac{\phi(x+td)}{\|x+td\|} = \lim_{t\to\infty} \frac{\phi(x+td) - \phi(x)}{t\|d\|} = \sup_{t>0} \frac{\phi(x+td) - \phi(x)}{t\|d\|}$$

i.e. $\phi(x+td) - \phi(x) \leq \Phi(x, d)t\|d\|$ for all $t > 0$, $x \in \mathrm{co}\{v_i\}$, $d \in \Delta$. In particular, with notation as in Theorem 3,

$$\begin{aligned}
y &= \sum_i \gamma_i \phi\left(\sum_j \alpha_{ij} v_j + \sum_{j'} \beta_{ij'} d_{j'} \right) \\
&\leq \sum_i \gamma_i \left[\phi\left(\sum_j \alpha_{ij} v_j \right) + \Phi\left(\sum_j \alpha_{ij} v_j, \sum_{j'} \beta_{ij'} d_{j'} \right) \|d^i\| \right] \\
&\leq \sum_i \sum_j \gamma_i \alpha_{ij} \phi(v_j) + \sum_i \sum_j \sum_{j'} \gamma_i \alpha_{ij} \beta_{ij'} \Phi(v_j, d_{j'}) \|d_{j'}\| \quad \text{(by Lemma 2)} \\
&\leq \sum_j \hat{\lambda}_j \phi(v_j) + \sum_{j'} \mu_{j'} \delta_{j'} \\
&\leq \max\left\{ \sum_j \lambda_j \phi(v_j) + \sum_{j'} \mu_{j'} \delta_{j'} \;\middle|\; \lambda_j \geq 0, \mu_{j'} \geq 0, \sum \lambda_j = 1, \sum \lambda_j v_j + \sum \mu_{j'} d_{j'} = x \right\}.
\end{aligned}$$

□

Remarks. 1. As in Section 4, the upper bound can be sharpened through the use of a partition $\{V^1, \ldots, V^k\}$ of V. Formula (6) still holds.

2. If assumption (i') is violated, e.g. if $U = \mathbb{R}^n$, the use of subsets may still be possible. Let $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$ and consider the partition of U into the 2^n subsets U^i given by $U^1 = \{x \in U \mid x_1 \leq t_1, \ldots, x_n \leq t_n\}$, $U^2 = \{x \in U \mid x_1 \leq t_1, \ldots, x_{n-1} \leq t_{n-1}, x_n > t_n\}$, etc.

6. Examples over unbounded regions

1. $$f(x,y)=\begin{cases} e^{-x}\cdot 1/2\sqrt{x} & \text{if } |y|\le\sqrt{x},\ x>0,\\ 0 & \text{otherwise,}\end{cases}$$

$$\phi(x,y)=\frac{x}{2}+\frac{1}{x+2}+\sqrt{x}-\sqrt{2x-y^2} \quad \text{if } |y|\le\sqrt{x}.$$

In this example, U is the area inside the parabola $y^2=x$, and ϕ can be extended to a convex function on the right half-plane by linear continuation, i.e. setting

$$\phi(x,y)=\frac{x}{2}+\frac{1}{x+2}-\sqrt{x}+|y| \quad \text{if } |y|>\sqrt{x}.$$

The expected value of (x, y) is easily computed to be $(1, 0)$, and the lower bound $\bar{\phi}\ge\phi(1,0)=\frac{11}{6}-\sqrt{2}\approx 0.41912$ obtains. (The value of $\bar{\phi}$ is approximately 0.60618.) Different upper bounds for different polyhedra V are easily computed using Theorem 4, as indicated in Table 1.

Table 1
Bounds on $\bar{\phi}$ for some choices of V

Polyhedron	Extreme points	Extreme rays	δ_1	δ_2	Upper bound on $\bar{\phi}$
V_1	$(0,\frac{1}{4}),(0,-\frac{1}{4})$	$y=\pm(x+\frac{1}{4})$	$\frac{3}{4}\sqrt{2}$	$\frac{3}{4}\sqrt{2}$	2.25
V_2	$(0,1),(0,-1)$	$y=\pm(\frac{1}{4}x+1)$	$3/\sqrt{17}$	$3/\sqrt{17}$	2.25
V_3	$(0,\frac{1}{2}),(0,-\frac{1}{2})$	$y=\pm(\frac{1}{2}x+\frac{1}{2})$	$2/\sqrt{5}$	$2/\sqrt{5}$	2.00
V_4	$(0,\pm\frac{1}{4}),(0,\pm 1)$	$y=\pm(\frac{1}{4}x+1)$	$3/\sqrt{17}$	$3/\sqrt{17}$	1.50

By subdividing these sets the upper bounds can be improved as follows. We subdivide V_3 into two sets and subdivide V_4 three different ways. This yields four different estimates which are given in Table 2.

2. $$\phi(x,y)=2.28027(\sqrt{x^2+(y^2/4+1)(0.41896)}-x)$$

The surface described by $(x, y, \phi(x, y))$ is the upper half of a hyperboloid of two sheets, rotated around the y-axis through an angle of -0.3 radians to remove some of the symmetries. This form was chosen because it satisfies the growth condition in an obvious manner. Assume that x and y are normally distributed with zero mean, unit variance and correlation coefficient $r\ne\pm 1$. Here $U=\mathbb{R}^2$, so an upper bound is not directly available without forming a partition. Jensen's inequality yields the lower bound $\bar{\phi}\ge\phi(0,0)\approx 1.49388$. Tables 3 and 4 give various upper bounds based on different partitions of $\mathbb{R}^2$ when the correlation between x and y is 0 and +0.5, respectively.

The calculations and timing were performed on an Amdahl 470 V/8 at the University of British Columbia.

Table 2
Subset bounds on $\bar{\phi}$ for various partitions

Subset	Extreme points	Extreme rays	δ_1	δ_2	Measure	$(\bar{x}, \bar{y})$	Subset upper bound	Upper bound on $\bar{\phi}$
V_{31}	$(0, \frac{1}{2})$, $(0, 0)$	$y = x/2 + \frac{1}{2}$, $y = 0$	$2/\sqrt{5}$	$\frac{1}{2}$	0.5	$(1, \sqrt{\pi}/4)$	1.44311	
								1.44311
V_{32}	$(0, -\frac{1}{2})$, $(0, 0)$	$y = -(x/2 + \frac{1}{2})$, $y = 0$	$2/\sqrt{5}$	$\frac{1}{2}$	0.5	$(1, \sqrt{\pi}/4)$	1.44311	
V_{41}	$(0, 0)$, $(0, \frac{1}{4})$, $(1, \frac{5}{4})$	$y = 0$, $y = x/4 + 1$	0.5	$3/\sqrt{17}$	0.5	$(1, \sqrt{\pi}/4)$	1.10472	
								1.10472
V_{42}	$(0, 0)$, $(0, \frac{1}{4})$, $(0, -\frac{5}{4})$	$y = 0$, $y = -x/4 - 1$	0.5	$3/\sqrt{17}$	0.5	$(1, \sqrt{\pi}/4)$	1.10472	
V_{43}	$(0, \pm\frac{1}{4})$, $(1, \pm\frac{5}{4})$	—		—	$1/1/e$	$(1 - 1/e - 1, 0)$	0.89267	
								1.23871
V_{44}	$(1, \pm\frac{5}{4})$	$y = \pm(x/4 + 1)$	$3/\sqrt{17}$	$3/\sqrt{17}$	$1/e$	$(2, 0)$	1.8333	
V_{45}	$(0, \pm\frac{1}{4})$, $(1/2, \pm\frac{3}{4})$	—		—	$1 - 1/\sqrt{e}$	(0.22925, 0)	0.72382	
V_{46}	$(1/2, \pm\frac{3}{4})$, $(1, \pm\frac{5}{4})$	—		—	$1/\sqrt{e} - 1/e$	(0.72925, 0)	0.87190	
								1.08181
V_{47}	$(1, \pm\frac{5}{4})$, $(2, \pm\frac{3}{2})$	—		—	$1/e - 1/e^2$	(1.418, 0)	1.22366	
V_{48}	$(2, \pm\frac{3}{2})$	$y = \pm(x/4 + 1)$	$3/\sqrt{17}$	$3/\sqrt{17}$	$1/e^2$	$(3, 0)$	2.25008	

7. A computer code

The FORTRAN program EMG2 [6] computes upper and lower bounds on the expected value of a convex function of a bivariate normal distribution to within a tolerance prescribed by the user.

The basic idea is as follows: An upper and lower bound is computed on each of the four quadrants in $\mathbb{R}^2$. Then a weighted difference is formed on each quadrant

Table 3

Bounds on Problem 2 when $\rho = 0$

$\bar{\phi} = 2.657$

Subset	Measure	Mean X	Mean Y	Number of extreme points	Number of extreme directions	Upper bound	Lower bound
U_{11}	0.2500	−0.7979	−0.7979	1	2	5.818	4.326
U_{12}	0.2500	−0.7979	−0.7979	1	2	5.818	4.326
U_{21}	0.2500	0.7979	0.7979	1	2	2.090	0.598
U_{22}	0.2500	0.7979	0.7979	1	2	2.090	0.598

Summary: $3.954 \geq \bar{\phi} \geq 2.462$

Partition:

U_{12}	U_{22}
U_{11}	U_{21}

Subset	Measure	Mean X	Mean Y	Number of extreme points	Number of extreme directions	Upper bound	Lower bound
U_{11}	0.0252	−1.5251	−1.5251	1	2	8.053	7.590
U_{12}	0.0542	−1.5251	−0.4599	2	1	7.608	7.441
U_{13}	0.0542	−1.5251	0.4599	2	1	7.608	7.441
U_{14}	0.0252	−1.5251	1.5251	1	2	8.053	7.590
U_{21}	0.0542	−0.4599	−1.5251	2	1	3.689	3.238
U_{22}	0.1165	−0.4599	−0.4599	4	0	3.237	2.946
U_{23}	0.1165	−0.4599	0.4599	4	0	3.237	2.946
U_{24}	0.0542	−0.4599	1.5251	2	1	3.689	3.238
U_{31}	0.0542	0.4599	−1.5251	2	1	1.541	1.090
U_{32}	0.1165	0.4599	−0.4599	4	0	1.089	0.798
U_{33}	0.1165	0.4599	0.4599	4	0	1.089	0.798
U_{34}	0.0542	0.4599	1.5251	2	1	1.541	1.090
U_{41}	0.0252	1.5251	−1.5251	1	2	0.928	0.465
U_{42}	0.0542	1.5251	−0.4599	2	1	0.482	0.316
U_{43}	0.0542	1.5251	0.4599	2	1	0.482	0.316
U_{44}	0.0252	1.5251	1.5251	1	2	0.928	0.465

Summary: $2.903 \geq \bar{\phi} \geq 2.587$
0.021 seconds CPU time
9 function evaluations

Partition:

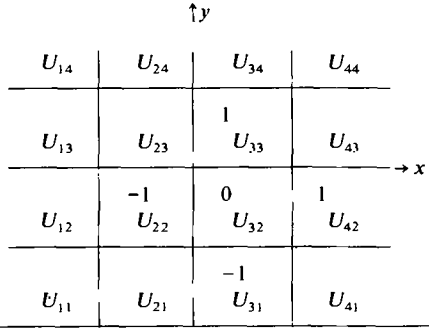


Table 4

Bounds on Problem 2 when $\rho = +0.5$

$\bar{\phi} = 2.648$

Subset	Measure	Mean X	Mean Y	Number of extreme points	Number of extreme directions	Upper bound	Lower bound
U_{11}	0.3333	−0.8976	−0.8976	1	2	6.359	4.757
U_{12}	0.1667	−0.5984	−0.5984	1	2	4.737	3.492
U_{21}	0.1667	0.5984	0.5984	1	2	1.941	0.696
U_{22}	0.3333	0.8976	0.8976	1	2	2.164	0.564

Summary: $3.954 \geq \bar{\phi} \geq 2.472$

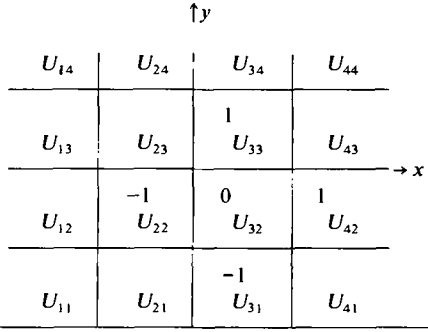
Partition:

U_{12}	U_{22}
U_{11}	U_{21}

Subset	Measure	Mean X	Mean Y	Number of extreme points	Number of extreme directions	Upper bound	Lower bound
U_{11}	0.0625	−1.6364	−1.6364	1	2	8.656	8.105
U_{12}	0.0649	−1.4831	−0.5255	2	1	7.418	7.258
U_{13}	0.0275	−1 3979	0.3776	2	1	7.005	6.868
U_{14}	0.0038	−1.3316	1.3316	1	2	7.005	6.702
U_{21}	0.0649	−0.5255	−1.4831	2	1	3.890	3.456
U_{22}	0.1141	−0.4720	−0.4720	4	0	3.284	2.993
U_{23}	0.1079	−0.4254	0.4254	4	0	3.107	2.816
U_{24}	0.0275	−0.3776	1.3979	2	1	3.303	2.907
U_{31}	0.0275	0.3776	−1.3979	2	1	1.539	1.143
U_{32}	0.1079	0.4254	−0.4254	4	0	1.119	0.828
U_{33}	0.1411	0.4720	0.4720	4	0	1.078	0.787
U_{34}	0.0649	0.5255	1.4831	2	1	1.435	1.001
U_{41}	0.0038	1.3316	−1.3316	1	2	0.783	0.481
U_{42}	0.0275	1.3979	−0.3776	2	1	0.474	0.337
U_{43}	0.0649	1.4831	0.5255	2	1	0.489	0.329
U_{44}	0.0625	1.6364	1.6364	1	2	1.011	0.460

Summary: $2.902 \geq \bar{\phi} \geq 2.580$
0.021 seconds CPU time
9 function evaluations

Partition:



to determine whether the overall upper and lower bounds ($\sum_i \eta_i u_i$ and $\sum_i \eta_i l_i$, respectively, where η_i is the measure of the ith subset) differ by less than the tolerance supplied by the user. If this is not the case, for some region, then this region is subdivided into four subregions and the process repeated. This continues up to a maximum of six subdivisions or until the tolerance criterion is met.

8. Final remarks

The lower bound obtained from Jensen's inequality seems to be consistently much better than the upper bounds resulting from Theorems 1 and 4. Hausch and Ziemba [8] also found this to be the case in their experiments with independent random variables. To obtain good estimates it therefore may be necessary to partition V into a relatively large number of subsets. This has the disadvantage that the measure and mean have to be calculated on every subset, requiring in principle a large number of numerical integrations. Therefore the method of bounding may only be competitive with a direct evaluation of the integrand (1) if the value of ϕ is difficult to compute. Even then it is essential to have efficient algorithms at hand which calculate the measures and means on each subset. For normal distributions this could be effected by using Monte-Carlo methods along the lines of Deák [4, 5]. An algorithm for the bivariate normal case appears in Gassmann [7].

References

[1] A. Ben-Tal and E. Hochman, "More bounds on the expectation of a convex function of a random variable", *Journal of Applied Probability* 9 (1972) 803-812.

[2] A. Ben-Tal, C.C. Huang and W. Ziemba, "Bounds on the expectation of a random variable: With applications to stochastic programming", Working Paper No. 280, Faculty of Commerce, University of British Columbia (Vancouver, B.C., 1975).

[3] J. Birge and R.J-B. Wets, "Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse", this volume, pp. 54-102 (North-Holland, Amsterdam, 1986).

[4] I. Deák, "Computation of multiple normal probabilities", in: P. Kall and A. Prékopa, eds., Recent results in stochastic programming (Springer Verlag, New York, 1980) pp. 107-120.

[5] I. Deák, "Three digit multiple normal probabilities", *Numerische Mathematik* 35 (1980) 369-380.

[6] H. Gassmann, "Computer code EMG2 for computing upper and lower bounds on the expectation of a convex function of a bivariate normal distribution", mimeo, Faculty of Commerce, University of British Columbia (Vancouver, B.C., 1983).

[7] H. Gassmann, "Means of bivariate normal distributions over a rectangle: An efficient algorithm", Working Paper No. 971, Faculty of Commerce, University of British Columbia (Vancouver, B.C., 1983).

[8] D.B. Hausch and W.T. Ziemba, "Bounds on the value of information in uncertain decision problems II", *Stochastics* 10 (1983) 181-217.

[9] C.C. Huang, W.T. Ziemba and A. Ben-Tal, "Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming", *Operations Research* 25 (1977) 315-325.

[10] A. Madansky, "Bounds on the expectation of a convex function of a multivariate random variable", *Annals of Mathematical Statistics* 30 (1959) 743-746.

[11] M.D. Perlman, "Jensen's inequality for a convex vector-valued function on an infinite dimensional space", *Journal of Multivariate Analysis* 4 (1974) 52-65.

[12] J. Pfanzagl, "Convexity and conditional expectations", *Annals of Probability* 2 (1974) 490-494.

[13] R. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., Mathematical programming 1982: The state of the art (Springer-Verlag, Berlin, 1983) pp. 566-603.

Mathematical Programming Study 27 (1986) 54-102
North-Holland

DESIGNING APPROXIMATION SCHEMES FOR STOCHASTIC OPTIMIZATION PROBLEMS, IN PARTICULAR FOR STOCHASTIC PROGRAMS WITH RECOURSE

John R. BIRGE
Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

Roger J.-B. WETS
Department of Mathematics, University of Kentucky, Lexington, KY, USA

Received 15 November 1983
Revised manuscript received 38 August 1984

Various approximation schemes for stochastic optimization problems, involving either approximates of the probability measures and/or approximates of the objective functional, are investigated. We discuss their potential implementation as part of general procedures for solving stochastic programs with recourse.

Key words: Stochastic Optimization, Approximation Methods, Stochastic Programs with Recourse

1. Introduction

We consider a simple abstract version of stochastic optimization problems and present general results for approximating both the probability measure and the objective functional. The error bounds are derived as the case may be from convexity, sublinearity, linearity, and monotonicity properties of the objective functional. The probability measure approximations are generally discrete measures that should allow for easy calculation of the objective at each value. The functional approximations are appropriate linearizations of the objective functional. The guidelines are provided by a class of problems known as stochastic programs with (fixed) recourse. It also conditions the implementation methodology of the general results. It is anticipated that these approximations can be used together in stochastic optimization solution procedures. We also report on some experimental computational results in the last section.

We take

$$\text{find } x \in \mathbb{R}^n \text{ that minimizes } E_f(x) = E\{f(x, \xi(x, \xi(\omega)))\} \tag{1.1}$$

as a prototype for the class of *stochastic optimization problems* under investigation.

Supported in part by grants of the National Science Foundation.

where

$$E\{f(x, \xi(\omega))\} = \int f(x, \xi(\omega))P(\mathrm{d}\omega), \tag{1.2}$$

ξ is a random vector which maps the probability space $(\Omega, \mathcal{A}, P)$ on to $(\mathbb{R}^N, \mathcal{B}^N, F)$ with F the distribution function and $\Xi \in \mathbb{R}^N$ the support of the probability measure induced by ξ (i.e. Ξ is the set of possible values assumed by ξ), and $f: \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is an extended real-valued function. Assume:

$$\text{for all } x,\ \omega \to f(x, \xi(\omega)) \text{ is measurable,} \tag{1.3}$$

and the following integrability condition:

$$\text{if } P[\omega \,|\, f(x, \xi(\omega)) < +\infty] = 1 \quad \text{then } E_f(x) < +\infty. \tag{1.4}$$

We refer to $E_f = E\{f(\cdot, \xi(\omega))\}$ as an *expectation functional.* Note that it can also be expressed as a Lebesgue–Stieltjes integral with respect to F:

$$E_f(x) = \int_{\mathbb{R}^N} f(x, \zeta)\, \mathrm{d}F(\zeta). \tag{1.5}$$

A wide variety of stochastic optimization problems fit into this (abstract) framework; in particular *stochastic programs with (fixed) recourse* [41]

$$\text{find } x \in \mathbb{R}^{n_1}_+ \text{ such that } Ax = b, \text{ and } z = cx + \mathcal{Q}(x) \text{ is minimized} \tag{1.6}$$

where A is an $m_1 \times n_1$-matrix, $b \in \mathbb{R}^{m_1}$,

$$\mathcal{Q}(x) = E\{Q(x, \xi(\omega))\} = \int Q(x, \xi(\omega))P(\mathrm{d}\omega), \tag{1.7}$$

and the *recourse function* is defined by

$$Q(x, \xi(\omega)) = \inf_{y \in \mathbb{R}^{n_2}_+} \{q(\omega)y \,|\, Wy = h(\omega) - T(\omega)x\}. \tag{1.8}$$

The $(m_2 \times n_2)$-matrix W is called the *recourse matrix.* For each ω: $T(\omega)$ is $m_2 \times n_1$, $q(\omega) \in \mathbb{R}^{n_2}$ and $h(\omega) \in \mathbb{R}^{m_2}$. Piecing together the stochastic components of the problem yields a vector $\xi \in \mathbb{R}^N$ with $N = n_2 + m_2 + (m_2 \times n_1)$, and

$$\xi = (q_1, \ldots, q_{n_2}, h_1, \ldots, h_{m_2}, t_{11}, \ldots, t_{1n_1}, t_{21}, \ldots, t_{m_2 n_1}).$$

We set

$$f(x, \xi) = \begin{cases} cx + Q(x, \xi) & \text{if } Ax = b,\ x \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.9}$$

Provided the recourse problem is a.s. bounded, i.e.

$$P[\omega \,|\, \exists \pi \text{ such that } \pi W \geq q(\omega)] = 1, \tag{1.10}$$

which we assume henceforth, the function Q and thus also f, does not take on the value $-\infty$. The measurability of $f(x, \cdot)$ follows directly from that of $\xi \mapsto Q(x, \xi)$ [1, Section 3]. If ξ has finite second moments, then $\mathcal{Q}(x)$ is finite whenever $\omega \mapsto Q(x, \xi(\omega))$ is finite [41, Theorem 4.1] and this guarantees condition (1.4).

Much is known about problems of this type [41]. The properties of f, as defined through (1.9), quite often motivate and justify the conditions under which we obtain various results. The relevant properties are

$$(h, T) \mapsto Q(x, \xi = (q, h, T)) \text{ is a piecewise linear convex function for all feasible } x \in K = K_1 \cap K_2, \tag{1.11}$$

where

$$K_1 = \{x \mid Ax = b, x \geq 0\},$$

$$K_2 = \{x \mid \forall \xi(\omega) \in \Xi, \exists y \geq 0 \text{ such that } Wy = h(\omega) - T(\omega)x\},$$

$$q \mapsto Q(x, \xi = (q, h, T)) \text{ is a concave piecewise linear function,} \tag{1.12}$$

and

$$x \mapsto Q(x, \xi) \text{ is a convex piecewise linear function} \tag{1.13}$$

which implies that

$$x \mapsto \mathcal{Q}(x) \text{ is a Lipschitzian convex function,} \tag{1.14}$$

finite on K_2, as follows from the integrability condition on $\xi(\cdot)$.

When T is nonstochastic, or equivalently does not depend on ω, it is sometimes useful to work with a variant formulation of (1.6). With $T = T(\omega)$ for all ω, we obtain

$$\text{find } x \in \mathbb{R}^{n_1}_+, \chi \in \mathbb{R}^{m_2} \text{ such that } Ax = b,\ Tx = \chi, \text{ and } z = cx + \Psi(\chi) \text{ is minimized} \tag{1.15}$$

where

$$\Psi(\chi) = E\{\psi(\chi, \xi(\omega))\} = \int \psi(\chi, \xi(\omega)) P(\mathrm{d}\omega) \tag{1.16}$$

and

$$\psi(\chi, \xi(\omega)) = \inf\{q(\omega)y \mid Wy = h(\omega) - \chi, y \in \mathbb{R}^{n_2}_+\}. \tag{1.17}$$

This formulation stresses the fact that choosing x corresponds to generating a *tender* $\chi = Tx$ to be 'bid' against the outcomes $h(\omega)$ of random events. The functions ψ and Ψ have basically the same properties as Q and $\mathcal{Q}$, replacing naturally the set K_2 by the set $L_2 = \{\chi = Tx \mid x \in K_2\} = \{\chi \mid \forall h(\omega) \in \Xi_h, \exists y \geq 0 \text{ such that } Wy = h(\omega) - Tx\}$. The function f is now given by

$$f((x, \chi), \xi) = \begin{cases} cx + \psi(\chi, \xi) & \text{if } Ax = b,\ Tx = \chi,\ x \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.18}$$

A significant number of applications have the function ψ separable, i.e. $\psi(\chi, \xi) = \sum_{i=1}^{m_2} \psi_i(\chi_i, \xi_i)$ such as in stochastic programs with *simple recourse* ([41, Section 6], for the nonlinear version cf. [49]). This will substantially simplify the implementation of various approximation schemes described below. When separability is not at hand, it will sometimes be useful to introduce it, by constructing appropriate approximates for ψ or Q, see Section 3.

Another common feature of stochastic optimization problems, that one should not lose track of when designing approximation schemes, is that the random behavior of the stochastic elements of the problem can often be traced back to a few *independent* random variables. Typically

$$\xi(\omega) = \zeta_1(\omega)\xi^1 + \zeta_2(\omega)\xi^2 + \cdots + \zeta_M(\omega)\xi^M \tag{1.19}$$

where the

$$\{\zeta_i : \Omega \to \mathbb{R};\ i = 1, \ldots, M\}$$

are independent real-valued random variables, and

$$\xi^i = (q^i_1, \ldots, q^i_{n_2}, h^i_1, \ldots, h^i_{m_2}, t^i_{11}, \ldots, t^i_{m_2 n_1})$$

are fixed vectors. In fact many applications—such as those involving scenario analysis—involve just one such random variable $\zeta(\cdot)$; naturally, this makes the components of the random vector $\xi(\cdot)$ highly dependent. Last, but not least, in many practical cases, we do not have adequate statistics to model with sufficient accuracy joint phenomena involving intricate relationships between the components of ξ. Hence, we shall devote most of our attention to the *independent case*, remaining at all times very much aware of the construction (1.19).

This will serve as background to our study of approximation schemes for calculating

$$E_f(x) = \int f(x, \xi(\omega))P(\mathrm{d}\omega).$$

After discussing general convergence results (Section 2), we begin our study with a description of possible approximates of f in the context of stochastic programs with recourse (Section 3). We then examine the possibility of obtaining lower or upper bounds on E_f by means of discretization (of the probability measure) using conditional expectations (Section 4), measures with extremal support (Section 5), extremal measures (Section 6) or majorizing probability measures (Section 7). In each case we also sketch out the implementation of the results in the framework of stochastic programs with recourse, relying in some cases on the approximates to f obtained in Section 3. In Section 8, we give some further error bounds for $\inf E_f$ that require the actual calculation of $E_f(x)$ at some points.

2. Convergence results

The purpose of this section is to free us at once from any further detailed argumentation involving convergence of solutions, infima, and so on. To do so we rely on the tools provided by epi-convergence. Let $\{g; g^\nu, \nu = 1, \dots\}$ be a collection of functions defined on $\mathbb{R}^n$ with values in $\bar{\mathbb{R}} = [-\infty, +\infty]$. The sequence $\{g^\nu, \nu = 1, \dots\}$ is said to *epi-converge* to g if for all $x \in \mathbb{R}^n$, we have

$$\liminf_{\nu\to\infty} g^\nu(x^\nu) \geq g(x) \quad \text{for all } \{x^\nu, \nu = 1, \dots\} \text{ converging to } x, \tag{2.1}$$

and

$$\text{there exists } \{x^\nu, \nu = 1, \dots\} \quad \text{converging to } x \text{ such that}$$
$$\limsup_{\nu\to\infty} g^\nu(x^\nu) \leq g(x). \tag{2.2}$$

Note that any one of these conditions implies that g, the *epi-limit* of the g^ν, is necessarily lower semicontinuous. The name epi-convergence comes from the fact that the functions $\{g^\nu, \nu = 1, \dots\}$ epi-converge to g if and only if the sets $\{\text{epi } g^\nu, \nu = 1, \dots\}$ converge to $\text{epi } g = \{(x, \alpha) \mid g(x) \leq \alpha\}$; for more details consult [44, 1]. Our interest in epi-convergence stems from the following properties [2].

2.3. Theorem. *Suppose a sequence of functions $\{g^\nu, \nu = 1, \dots\}$ epi-converges to g. Then*

$$\limsup_{\nu\to\infty} (\inf g^\nu) \leq \inf g, \tag{2.4}$$

and, if

$$x^k \in \operatorname{argmin} g^{\nu_k} = \{x \mid g^{\nu_k}(x) \leq \inf g^{\nu_k}\}$$

for some subsequence of functions $\{g^{\nu_k}, k = 1, \dots\}$ and $x = \lim_{k\to\infty} x^k$, it follows that

$$x \in \operatorname{argmin} g \quad \textit{and} \quad \lim_{k\to\infty} (\inf g^{\nu_k}) = \inf g.$$

Moreover, if $\operatorname{argmin} g \neq \emptyset$, *then* $\lim_{\nu\to\infty}(\inf g^\nu) = \inf g$ *if and only if* $x \in \operatorname{argmin} g$ *implies the existence of sequences $\{\varepsilon_\nu \geq 0, \nu = 1, \dots\}$ and $\{x^\nu, \nu = 1, \dots\}$ with*

$$\lim_{\nu\to\infty} \varepsilon_\nu = 0 \quad \textit{and} \quad \lim_{\nu\to\infty} x^\nu = x,$$

such that for all $\nu = 1, \dots,$

$$x_\nu \in \varepsilon_\nu\text{-}\operatorname{argmin} g^\nu = \{x \mid g^\nu(x) \leq \inf g^\nu + \varepsilon_\nu\}.$$

2.5. Corollary. *Suppose a sequence of functions $\{g^\nu, \nu = 1, \dots\}$ epi-converges to g, and there exists a bounded set D such that*

$$\operatorname{argmin} g^\nu \cap D \neq \emptyset$$

for all ν sufficiently large. Then

$$\lim_{\nu\to\infty}(\inf g_\nu) = \inf g$$

and the minimum of g is attained at some point in the closure of D.

Proof. Since D is bounded, there exists a bounded sequence $\{x^\nu, \nu = 1, \dots\}$ with

$$x^\nu \in \operatorname{argmin} g^\nu \cap D.$$

This means that a subsequence converges $\{x^{\nu_k}, k = 1, \dots\}$ to a point x both in the closure of D and in argmin g as follows from epi-convergence. Theorem 2.3 also yields

$$\lim_{k\to\infty} g^{\nu_k}(x^{\nu_k}) = g(x) = \inf g.$$

There remains only to argue that the entire sequence $\{(\inf g^\nu), \nu = 1, \dots\}$ converges to $\inf g$. But this simply follows from the observation that the preceding argument applied to any subsequence yields a further subsequence converging to $\inf g$. □

The following proposition provides very useful criteria for verifying epi-convergence.

2.6. Proposition [3, Proposition 3.12]. *Suppose $\{g^\nu : \mathbb{R}^n \to \bar{\mathbb{R}}, \nu = 1, \dots\}$ is a collection of functions pointwise converging to g, i.e. for all x, $g(x) = \lim_{\nu\to\infty} g^\nu(x)$. Then the g^ν epi-converge to g, if they are monotone increasing, or monotone decreasing with g lower semicontinuous.*

For expectation functionals, we obtain the next assertion as a direct consequence of the definition of epi-convergence and Fatou's lemma.

2.7. Theorem. *Suppose $\{f; f^\nu, \nu = 1, \dots\}$ is a collection of functions defined on $\mathbb{R}^n \times \Omega$ with values in $\mathbb{R} \cup \{+\infty\}$ satisfying conditions (1.3) and (1.4), such that for all $\xi \in \Xi$ the sequence $\{f^\nu(\cdot, \xi), \nu = 1, \dots\}$ epi-converges to $f(\cdot, \xi)$. Suppose moreover that the functions f^ν are absolutely bounded by uniformly integrable functions. Then the expectation functionals E_{f^ν} epi-converge to E_f.*

When instead of approximating the functional f, we approximate the probability measure P, we get the following general result that suits our needs in most applications, see [45, Theorem 3.9], [46, Theorem 3.3].

2.8. Theorem. *Suppose $\{P_\nu, \nu = 1, \dots\}$ is a sequence of probability measures converging in distribution to the probability measure P defined on Ω, a separable metric space with $\mathcal{A}$ the Borel sigma-field. Let*

$$(x, \omega) \mapsto f(x, \xi(\omega) : \mathbb{R}^n \times \Omega \to \mathbb{R} \cup \{+\infty\}$$

be continuous in ω for each fixed x in K, where

$$K = \{x \mid E_f(x) < +\infty\} = \{x \mid f(x, \xi(\omega)) < +\infty, \text{ a.s.}\} \neq \emptyset,$$

and locally Lipschitz in x on K with Lipschitz constant independent of ω. Suppose moreover that for any $x \in K$ and $\varepsilon > 0$ there exists a compact set S_ε and ν_ε such that for all $\nu \geq \nu_\varepsilon$

$$\int_{\Omega \setminus S_\varepsilon} |f(x, \xi(\omega))| P_\nu(\mathrm{d}\omega) < \varepsilon, \tag{2.9}$$

and with $V = \{\omega \mid f(x, \omega) = +\infty\}$, $P(V) > 0$ if and only if $P_\nu(V) > 0$. Then the sequence of expectation functionals $\{E_f^\nu, \nu = 1, \dots\}$ epi- and pointwise converges to E_f, where

$$E_f^\nu(x) = \int f(x, \xi(\omega)) P_\nu(\mathrm{d}\omega).$$

Proof. We begin by showing that the E_f^ν pointwise converge to E_f. First let $x \in K$ and set

$$g(\omega) = f(x, \omega).$$

From (2.9), it follows that for all $\varepsilon > 0$, there is a compact set S_ε and index ν_ε such that for all $\nu \geq \nu_\varepsilon$

$$\int_{\Omega \setminus S_\varepsilon} |g(\omega)| P_\nu(\mathrm{d}\omega) < \varepsilon.$$

Let $M_\varepsilon = \sup_{\omega \in S_\varepsilon} |g(\omega)|$. We know that M_ε is finite since S_ε is compact and g is continuous, recall that $x \in K$. Let g^ε be a truncation of g defined by

$$g^\varepsilon(\omega) = \begin{cases} g(\omega) & \text{if } |g(\omega)| \leq M_\varepsilon, \\ M_\varepsilon & \text{if } g(\omega) > M_\varepsilon, \\ -M_\varepsilon & \text{if } g(\omega) < -M_\varepsilon. \end{cases}$$

The function g^ε is bounded and continuous and for all $\omega \in \Omega$

$$|g^\varepsilon(\omega)| \leq |g(\omega)|.$$

Hence from the convergence in distribution of the P_ν

$$\lim_{\nu \to \infty} \left[\beta_\nu^\varepsilon = \int_\Omega g^\varepsilon(\omega) P_\nu(\mathrm{d}\omega) \right] = \int_\Omega g^\varepsilon(\omega) P(\mathrm{d}\omega) = \beta^\varepsilon \tag{2.10}$$

and also for all $\nu \geq \nu_\varepsilon$,

$$\int_{\Omega \setminus S_\varepsilon} g^\varepsilon(\omega) P_\nu(\mathrm{d}\omega) < \varepsilon.$$

Now let

$$\beta_\nu = E_f^\nu(x) = \int_{S_\varepsilon} g(\omega) P_\nu(\mathrm{d}\omega) + \int_{\Omega \setminus S_\varepsilon} g(\omega) P_\nu(\mathrm{d}\omega).$$

We have that, for all $\nu \geq \nu_\varepsilon$,

$$|\beta_\nu - \beta_\nu^\varepsilon| = \left| \int_{\Omega \setminus S_\varepsilon} [g(\omega) - g^\varepsilon(\omega)] P_\nu(\mathrm{d}\omega) \right| < 2\varepsilon$$

and also that

$$|E_f(x) - \beta^\varepsilon| < 2\varepsilon.$$

Combining the two last inequalities with (2.10) shows that for all $\varepsilon > 0$, there exists ν_ε such that for all $\nu \geq \nu_\varepsilon$

$$|E_f(x) - \beta_\nu| < 6\varepsilon,$$

and thus for all $x \in K$,

$$\lim_{\nu \to \infty} E_f^\nu(x) = E_f(x).$$

If $x \notin K$, this means that

$$P[V = \{\omega \,|\, f(x, \xi(\omega)) = +\infty\}] > 0$$

which also means that for all ν

$$P_\nu(V) > 0,$$

from which it follows that for all ν

$$\lim_{\nu \to +\infty} E_f^\nu(x) = +\infty = E_f(x).$$

And thus, for all $x \in \mathbb{R}^n$, $E_f(x) = \lim_{\nu \to \infty} E_f^\nu(x)$. This gives us not only pointwise convergence, but also condition (2.2) for epi-convergence.

To complete the proof, it thus suffices to show that condition (2.1) is satisfied for all $x \in K$. The function $x \mapsto f(x, \xi(\omega))$ is Lipschitzian on K, with Lipschitz constant L independent of ω. For any pair x, x^ν in K, we have that for all ω

$$|f(x, \xi(\omega)) - f(x^\nu, \xi(\omega))| \leq L \operatorname{dist}(x, x^\nu)$$

which implies that

$$f(x, \xi(\omega)) - L \operatorname{dist}(x, x^\nu) \leq f(x^\nu, \xi(\omega)).$$

Let us now take x^ν as part of a sequence $\{x^\nu, \nu = 1, \dots\}$ converging to x. Integrating

on both sides of the preceding inequality and taking $\liminf_{\nu\to\infty}$, we get

$$E_f(x) = \lim_{\nu\to\infty} E_f^\nu(x) - L \lim_{\nu\to\infty} \operatorname{dist}(x, x^\nu)$$

$$= \liminf_{\nu\to\infty} (E_f^\nu(x) - L \operatorname{dist}(x, x^\nu))$$

$$\leq \liminf_{\nu\to\infty} E_f^\nu(x^\nu),$$

which completes the proof. □

2.11. Application. *Suppose $\{P_\nu, \nu = 1, \dots\}$ is a sequence of probability measures that converge in distribution to P, all with compact support Ω. Suppose*

$$\mathcal{Q}^\nu(x) = \int Q(x, \xi(\omega)) P_\nu(\mathrm{d}\omega)$$

with the recourse function Q defined by (1.8) *and $\mathcal{Q}$ by* (1.7). *Then the $\mathcal{Q}^\nu$ both epi- and pointwise converge to $\mathcal{Q}$.*

It suffices to observe that the conditions of Theorem 2.8 are satisfied. The continuity of $Q(x, \xi)$ with respect to ξ (for $x \in K_2$) follows from (1.11) and (1.12). The Lipschitz property with respect to x is obtained from [41, Theorem 7.7]; the proof of that theorem also shows that the Lipschitz constant is independent of ξ, consult also [40].

2.12. Implementation. From the preceding results it follows that we have been given great latitude in the choice of the probability measures that approximate P. However, in what follows we concern ourselves almost exclusively with discrete probability measures. The basic reason for this is that the form of $f(x, \xi)$—or $Q(x, \xi)$ in the context of stochastic programs with recourse—renders the numerical evaluation of E_f (or E_f^ν) possible only if the integral is actually a (finite) sum. Only in highly structured problems, such as for stochastic programs with simple recourse [42], may it be possible and profitable to use other approximating measures.

3. Approximating the recourse function Q

When f is convex in ξ, it is possible to exploit this property to obtain simple but very useful lower bounding approximates for E_f.

3.1. Proposition. *Suppose $\xi \mapsto f(x, \xi)$ is convex, $\{\xi^l, l = 1, \dots, \nu\}$ is a finite collection of points in Ξ, and for $l = 1, \dots,$*

$$v^l \in \partial_\xi f(x, \xi^l),$$

i.e. v^l *is a subgradient of* $f(x, \cdot)$ *at* ξ^l. *Then*

$$E_f(x) \geq E\{\max_{1 \leq l \leq \nu} [v^l \xi(\omega) + (f(x, \xi^l) - v^l \xi^l)]\}. \tag{3.2}$$

Proof. To say that v^l is a subgradient of the convex function of $f(x, \cdot)$ at ξ^l, means that

$$f(x, \xi) - f(x, \xi^l) \geq v^l(\xi - \xi^l).$$

Since this holds for every l, we obtain

$$f(x, \xi) \geq \max_{1 \leq l \leq \nu} [v^l \xi + (f(x, \xi^l) - v^l \xi^l)].$$

Integrating on both sides yields (3.2). □

3.3. Application. *Consider the stochastic program with recourse* (1.6) *and suppose that only h and T are stochastic. Let* $\{\xi^l = (h^l, T^l), l = 1, \ldots, \nu\}$ *be a finite number of realizations of h and T,* $x \in K_2$ *and for* $l = 1, \ldots, \nu$,

$$\pi^l \in \operatorname{argmax}[\pi(h^l - T^l x) \mid \pi W \leq q].$$

Then

$$\mathcal{Q}(x) \geq E\{\max_{1 \leq l \leq \nu} \pi^l(h(\omega) - T(\omega)x)\}. \tag{3.4}$$

This is a direct corollary of Proposition 3.1. We give an alternative proof which could be of some help in the design of the implementation. Since $x \in K_2$, for every $\xi = (h, T)$ in Ξ, the linear program

$$\text{find } \pi \in \mathbb{R}^{m_2} \text{ such that } \pi W \leq q \text{ and } w = \pi(h - Tx) \text{ is maximized} \tag{3.5}$$

is bounded, given naturally that it is feasible as follows from assumption (1.10). Hence, for $l = 1, \ldots, \nu$,

$$Q(x, \xi^l) = \pi^l(h^l - T^l x),$$

and moreover since π^l is a feasible solution of the linear program (3.5), for all $\xi \in \Xi$,

$$Q(x, \xi) \geq \pi^l(h - Tx).$$

Since this holds for every l,

$$Q(x, \xi) \geq \max_{1 \leq l \leq \nu} \pi^l(h - Tx).$$

Integrating on both sides yields (3.4).

3.6. Implementation. In general finding the maximum for each ξ, in expression (3.4)—or equivalently for each $(h, T) \in \Xi$—could be much too involved. But we may assign to each π^l a subregion of Ξ, without resorting to (exact) maximization.

The lower bound may then not be as tight as (3.4), but we can refine it by taking successively finer and finer partitions. However, one should not forget that (3.4) involves a rather simple integral and the expression to the right could be evaluated numerically to an acceptable degree of accuracy, without major difficulties. The calculation of this lower bound imposes no limitations on the choice of the ξ^l. However, is is obvious that a well-chosen spread of the $\{\xi^l, l = 1, \ldots, \nu\}$ will yield a better approximation. For example, the ξ^l could be the conditional expectation of $\xi(\cdot)$ with respect to a partition $\mathcal{S} = \{S_l, l = 1, \ldots, \nu\}$ of Ξ which assigns approximately the same probability to each S_l. The use of a larger collection of points, i.e. increasing ν, will also yield a better lower bound.

3.7. Convergence. *Suppose that $\xi \mapsto f(x, \xi)$ is convex and $E_f(x)$ is lower semicontinuous. For each $\nu = 1, \ldots,$ let $\mathcal{S}^\nu = \{S_l^\nu, l = 1, \ldots, L_\nu\}$ denote a partition of Ξ with*

$$\xi^{\nu l} = E\{\xi(\omega) \mid S_l^\nu\},$$

the conditional expectation of $\xi(\cdot)$ given S_l^ν. Suppose moreover that $\mathcal{S}^\nu \subset \mathcal{S}^{\nu+1}$ and that

$$\lim_{\nu \to \infty} \Big(\max_{1 \leq l \leq L_\nu} P[\omega \mid \xi(\omega) \in S_l^\nu] \Big) = 0. \tag{3.8}$$

Then, with $v^{\nu l} \in \partial_\xi f(x, \xi^{\nu l})$ and

$$E_f^\nu(x) = E\{ \max_{1 \leq l \leq L_\nu} [v^{l\nu} \xi(\omega) + f(x, \xi^{\nu l}) - v^{\nu l} \xi^{\nu l}] \}, \tag{3.9}$$

we have that the sequence of functions $\{E_f^\nu, \nu = 1, \ldots\}$ is monotone increasing, and, for all x,

$$E_f(x) = \lim_{\nu \to \infty} E_f^\nu(x).$$

Hence the sequence $\{E_f^\nu, \nu = 1, \ldots\}$ is both pointwise- and epi-convergent.

Proof. From Proposition 3.1, it follows that $E_f^\nu \leq E_f$ for all ν. The inequality

$$E_f^\nu \leq E_f^{\nu+1} \leq E_f$$

then follows simply from the fact that $\mathcal{S}^{\nu+1} \supset \mathcal{S}^\nu$. Now observe that

$$\max_{1 \leq l \leq L_\nu} [v^{\nu l} \xi + f(x, \xi^{\nu l}) - v^{\nu l} \xi^{\nu l}] \geq g^\nu(x, \xi) \tag{3.10}$$

where g^ν is defined as follows:

$$g^\nu(x, \xi) = v^{\nu l} \xi + f(x, \xi^{\nu l}) - v^{\nu l} \xi^{\nu l} \quad \text{if } \xi \in S_l^\nu.$$

It follows that

$$E_f^\nu(x) \geq E\{g^\nu(x, \xi(\omega))\} = \sum_{l=1}^{L_\nu} P[\xi(\omega) \in S_l^\nu] f(x, \xi^{\nu l})$$

which gives us

$$E_f(x) \geq \lim_{\nu\to\infty} E_f^\nu(x) \geq \lim_{\nu\to\infty} E\{g^\nu(x, \xi(\omega))\} = E_f(x);$$

the last equality following from assumption (3.8).

We have thus shown that the sequence $\{E_f^\nu, \nu = 1, \dots\}$ is monotone increasing and pointwise converges, and this implies epi-convergence, see Proposition 2.6. □

If $f(x, \cdot)$ is *concave*, the equality in (3.2) is reversed and, instead of a lower bound on E_f, we obtain an upper bound. In fact, we can again use Proposition 3.1, but this time applied to $-f$.

3.11. Application. *Consider the stochastic program with recourse* (1.6) *and suppose that only the vector q is stochastic. Let* $\{\xi^l = q^l, l = 1, \dots, \nu\}$ *be a finite number of realizations of q,* $x \in K_2$ *and, for* $l = 1, \dots, \nu$,

$$y^l \in \operatorname{argmin}[q^l y \mid Wy = p - Tx, y \geq 0].$$

Then

$$\mathcal{Q}(x) \leq E\{\min_{1\leq l\leq \nu} q(\omega)y^l\}. \tag{3.12}$$

This is really a corollary of Proposition 3.1. A slightly different proof proceeds as follows: Note that for all $q = \xi \in \Xi$, for every l, y^l is a feasible, but not necessarily optimal, solution of the linear program

$$\text{find } y \in \mathbb{R}^{n_2}_+ \text{ such that } Wy = p - Tx \text{ and } w = qy \text{ is minimized.}$$

Hence

$$Q(x, \xi) \leq \min_{1\leq l\leq \nu} qy^l$$

from which (3.12) follows by integrating on both sides.

The remarks made about Implementation 3.6 and the arguments used in Convergence 3.7 still apply to the concave case since we are in the same setting as before provided we work with $-f$ or $-Q$.

Proposition 3.1 provides us with a lower bound for E_f when $\xi \mapsto f(x, \xi)$ is convex. The next result yields an upper bound.

3.13. Proposition. *Suppose* $\xi \mapsto f(x, \xi)$ *is convex,* $\{\xi^l, l = 1, \dots, \nu\}$ *is a finite collection of points in* Ξ. *Then*

$$E_f(x) \leq E_f^\nu(x) = \int f^\nu(x, \xi(\omega))P(\mathrm{d}\omega) \tag{3.14}$$

where

$$f^\nu(x, \xi) = \inf_{\lambda\in\mathbb{R}^\nu_+}\left[\sum_{l=1}^{\nu} \lambda_l f(x, \xi^l) \,\middle|\, \xi = \sum_{l=1}^{\nu} \lambda_l \xi^l, 1 = \sum_{l=1}^{\nu} \lambda_l\right]. \tag{3.15}$$

If the function $\xi \mapsto f(x, \xi)$ is sublinear, the f^ν can be defined as follows:

$$f^\nu(x, \xi) = \inf_{\lambda \in \mathbb{R}^\nu_+} \left[\sum_{l=1}^{\nu} \lambda_l f(x, \xi^l) \,\middle|\, \xi = \sum_{l=1}^{\nu} \lambda \xi^l \right]. \tag{3.16}$$

(*Note that $f^\nu(x, \xi)$ is $+\infty$ if the corresponding program is infeasible.*)

Proof. Convexity implies that for all $\lambda_1 \geq 0, \ldots, \lambda_\nu \geq 0$ with $\sum_{l=1}^{\nu} \lambda_l = 1$, and $\xi = \sum_{l=1}^{\nu} \lambda_l \xi^l$ we have

$$f(x, \xi) \leq \sum_{l=1}^{\nu} \lambda_l f(x, \xi^l) \tag{3.17}$$

from which (3.14) follows using (3.15). Sublinearity (convexity and positive homogeneity) also yields (3.17) but this time without $\sum_{l=1}^{\nu} \lambda_l = 1$, and this in turn yields (3.14) using (3.16) this time. □

3.18. Application. *Ray functions.* Consider the stochastic program with recourse in the form (1.15) and suppose that only h is stochastic, i.e., with fixed matrix T and recourse cost coefficients q. Now suppose that for given χ, we have the values of $\{\psi(\chi, \xi^l) \,|\, \xi^l = h^l, l = 1, \ldots, \nu\}$ for a finite collection of realizations of $h(\cdot)$. Let $\xi \in \Xi$ and define

$$\psi^\nu(\chi, \xi) = \inf_{\lambda \in \mathbb{R}^\nu_+} \left[\sum_{l=1}^{\nu} \lambda_l \psi(\chi, \xi^l) \,\middle|\, \xi = \chi + \sum_{l=1}^{\nu} \lambda_l (\xi^l - \chi) \right]. \tag{3.19}$$

Then

$$\Psi(\chi) \leq \Psi^\nu(\chi) = \int \psi^\nu(\chi, \xi(\omega)) P(\mathrm{d}\omega).$$

The above follows from the second part of Proposition 3.13 provided we observe that from the definition (1.17) of ψ, we have that

$$h \mapsto \psi(\chi, h + \chi)$$

is sublinear. From this it follows that for any $\lambda \in \mathbb{R}^\nu_+$

$$\psi(\chi, (\xi - \chi) + \chi) \leq \sum_{l=1}^{\nu} \lambda_l \psi(\chi, (\xi^l - \chi) + \chi)$$

whenever

$$\xi - \chi = \sum_{l=1}^{\nu} \lambda_l (\xi^l - \chi),$$

and this leads to the construction of ψ^ν in (3.19). □

3.20. Implementation. Finding for each ξ, the optimal value of a linear program as required by the definition of ψ^ν in (3.19), could involve much more work than is

appropriate to invest in the computation of an upper bound. One way to remedy this is to subdivide Ξ such that each ξ is automatically assigned to a particular region spanned by a subset of the $\{\xi^l, l = 1, \ldots, \nu\}$ or to the subset whose points are such that

$$\xi - \chi \notin \operatorname{pos}(\xi^1 - \chi, \ldots, \xi^\nu - \chi) = \left\{ t \,\middle|\, t = \sum_{l=1}^{\nu} \lambda_l(\xi^l - \chi),\ \lambda \in \mathbb{R}_+^\nu \right\}.$$

One case in which all of this falls nicely into place is when a stochastic program with recourse of type (1.15) can be approximated by stochastic program with *simple recourse* [41, Section 6] where the function $\psi(\chi, \xi)$ is separable,

$$\psi(\chi, \xi) = \sum_{i=1}^{m_2} \psi_i(\chi_i, \xi_i) \tag{3.21}$$

and

$$\psi_i(\chi_i, \xi_i) = \inf[q_i^+ y_i^+ + q_i^- y_i^- \,|\, y_i^+ - y_i^- = h_i - \chi_i,\ y_i^+ \geq 0,\ y_i^- \geq 0], \tag{3.22}$$

here $\xi_i = (q_i^+, q_i^-, h_i)$. The function Ψ is then also separable and can be expressed as

$$\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i),$$

where

$$\Psi_i(\chi_i) = E\{\psi_i(\chi_i, \xi_i(\omega))\}.$$

(This is the *linear* version of the simple recourse problem.)

3.23. Application. *Approximation by simple recourse.* Consider a stochastic program with recourse of the type (1.15), with only h stochastic and complete recourse [1, Section 6]. This means that the recourse matrix W is such that

$$\operatorname{pos} W = \{t \,|\, t = Wy,\ y \geq 0\} = \mathbb{R}^{m_2},$$

i.e. the recourse problem is feasible whatever be h or χ. For $i = 1, \ldots, m_2$, define

$$q_i^+ = \inf\{qy \,|\, Wy = e^i,\ y \geq 0\}, \tag{3.24}$$

and

$$q_i^- = \inf\{qy \,|\, Wy = -e^i,\ y \geq 0\}, \tag{3.25}$$

where e^i is the unit vector with a 1 in the ith position, i.e.

$$e^i = (0, \ldots, 0, 1, 0, \ldots, 0)^T.$$

The recourse function $\psi(\chi, \xi)$ is then approximated by the recourse function (3.21) of a stochastic program with simple recourse using for q_i^+ and q_i^- the values defined by (3.24) and (3.25). This is a special case of the ray functions built in Application 3.18; each $(\xi - \chi)$ falls in a given orthant and is thus automaticallly assigned a

particular positive linear combination of the chosen points $(\pm e^i - \chi,\ i = 1, \ldots, m_2)$. To improve the approximation we have to introduce additional vectors ξ^l, which brings us back to the more general situation described in Application 3.18.

3.26. Application. Consider a stochastic program with recourse of type (1.15), with only q stochastic. The function

$$q \mapsto \psi(\chi, q = \xi) : \mathbb{R}^{m_2} \to \bar{\mathbb{R}}$$

is not only concave and polyhedral (1.12), it is also positively homogeneous. For any finite collection $\{\xi^l = q^l,\ l = 1, \ldots, \nu\}$ we have that

$$\psi(\chi, q) \geq \sup_{\lambda \in \mathbb{R}^\nu_+} \left[\sum_{l=1}^{\nu} \lambda_l \psi(\chi, q^l) \,\middle|\, q = \sum_{l=1}^{\nu} \lambda_l q^l \right]. \tag{3.27}$$

This again follows directly from Proposition 3.13; note that $\psi(\chi, q^l) = q^l y^l$ where $y^l \in \operatorname{argmin}[q^l y \mid Wy = h - \chi,\ y \geq 0]$.

3.28. Implementation. Calculating for each q, the upper bound provided by (3.27) may be prohibitive. We could assign each $q \in \Xi$ to some subregion of Ξ spanned by the positive combinations of some of the $\{q^l,\ l = 1, \ldots, \nu\}$. Such a bound is easier to obtain but is not as sharp as that generated by (3.27).

Another approach to constructing upper and lower bounds for stochastic programs with recourse is to rely on the *pairs programs* as introduced in [6, Section 4]. One relies again on convexity properties and once again one needs to distinguish between (h, T) stochastic, and q stochastic. To begin with, let us consider h, and T stochastic. For every $(h, T) = \xi \in \Xi$, and $(\hat{h}, \hat{T}) = \hat{\xi} \in \operatorname{co} \Xi$ (the convex hull of Ξ), let

$$\begin{aligned} \rho(\hat{\xi}, \xi) = \inf[cx + \hat{p} q \hat{y} + (1 - \hat{p}) q y_\xi] \\ \text{such that} \quad Ax \qquad\quad &= b, \\ \hat{T}x + W\hat{y} \quad &= \hat{h}, \\ Tx + W y_\xi &= h, \\ x \geq 0,\ \hat{y} \geq 0,\ y_\xi &\geq 0, \end{aligned} \tag{3.29}$$

with $\hat{p} \in [0, 1]$. If (1.6) is solvable, so is (3.29) as follows from [41, Theorem 4.6]. Suppose x^0 solves (1.6) and for all $\xi = (h, T)$, let

$$y^0(\xi) \in \operatorname*{argmin}_{y \in \mathbb{R}^{n_2}_+} [qy \mid Wy = h - Tx].$$

It is well-known that $y^0(\xi)$ can be chosen so that $\xi \mapsto y^0(\xi)$ is measurable [41, Section 3]. Now suppose

$$\bar{\xi} = (\bar{h}, \bar{T}) = E\{\xi\} \quad \text{and} \quad \bar{y} = E\{y^0(\xi)\}.$$

The triple $(x^0, \bar{y}, y^0(\xi))$ is a feasible, but not necessarily optimal, solution of the linear program (3.29) when $(\hat{h}, \hat{T}) = (\bar{h}, \bar{T})$. Hence

$$\rho(\bar{\xi}, \xi) \leq cx^0 + \hat{p}q\bar{y} + (1-\hat{p})qy^0(\xi)$$

and integrating on both sides, we obtain

$$E\{\rho(\bar{\xi}, \xi)\} \leq cx^0 + \mathcal{Q}(x^0). \tag{3.30}$$

This bound can be refined in many ways: first, instead of just using one point $\bar{\xi}$, one could use a collection of points obtained as conditional expectations of a partition of Ξ, and create a pairs program for each subregion of Ξ. Second, instead of just one additional point $\hat{\xi}$, we could use a whole collection $\{\hat{\xi}^1, \ldots, \hat{\xi}^\nu\}$ to build a program of the type (3.29). This is described in detail in [6] for the case when only h is stochastic but can easily be generalized to the case h and T stochastic.

When only q is stochastic, we consider a dual version of (3.29), viz. for every $q = \xi \in \Xi$ and $\hat{q} = \hat{\xi} \in \operatorname{co} \Xi$, let

$$\begin{aligned} \rho^d(\hat{\xi}, \xi) = \sup[\sigma b + \hat{p}\hat{\pi}h + (1-\hat{p})\pi_\xi h] \\ \text{such that} \quad \sigma A + \hat{\pi}T &\leq c, \\ \hat{\pi}W &\leq q, \\ \pi_\xi W &\leq q \end{aligned} \tag{3.31}$$

with $\hat{p} \in [0, 1]$. The same arguments as above with $\hat{\xi} = \bar{\xi}$, but relying this time on the dual [37] of problem (1.6), lead to

$$E\{\rho^d(\bar{\xi}, \xi)\} \geq cx^0 + \mathcal{Q}(x^0) := \inf(c\cdot + \mathcal{Q}). \tag{3.32}$$

4. Discretization of the probability measure P throughout conditional expectations

Jensen's inequality for convex functions is the basic tool to obtain lower bounds for E_f when $f(x, \cdot)$ is convex or upper bounds when E_f is concave. Here, it leads to the use of (molecular) probability measures concentrated at conditional expectation points. In the context of stochastic programming this was first done by Madansky [31] and further refined by Huang, Ziemba and Ben-Tal [20] and Kall [22].

4.1. Proposition. *Let $\mathcal{S} = \{S^l, l = 1, \ldots, \nu\}$ be a partition of Ξ, with*

$$\xi^l = E\{\xi(\omega) \mid S^l\} \quad \text{and} \quad p_l = P[\xi(\omega) \in S^l].$$

Suppose first that $\xi \mapsto f(x, \xi)$ is convex. Then

$$E_f(x) \geq \sum_{l=1}^{\nu} p_l f(x, \xi^l). \tag{4.2}$$

If $\xi \mapsto f(x, \xi)$ is concave, then

$$E_f(x) \leq \sum_{l=1}^{\nu} p_l f(x, \xi^l). \tag{4.3}$$

Proof. Follows from the iterated application of Jensen's inequality: $f(x, E\{\xi(\omega)\}) \leq E\{f(x, \xi(\omega))\}$ when $f(x, \cdot)$ is convex; consult [35]. □

4.4. Application. Consider the stochastic program with recourse with only h and T stochastic. With $\mathcal{S} = \{S^l, l = 1, \ldots, \nu\}$ a partition of Ξ and for $l = 1, \ldots, \nu$, let

$$\xi^l = (h^l, T^l) = E\{(h(\omega), T(\omega)) \mid S^l\}$$

and $p_l = P[\xi(\omega) \in S^l]$. As follows from (1.11) and (4.2), we obtain

$$\sum_{l=1}^{\nu} p_l Q(x, \xi^l) \leq \mathcal{Q}(x), \tag{4.5}$$

and thus if

$$z^\nu = \inf_{x \in \mathbb{R}^{n_1}} \left\{ cx + \sum_{l=1}^{\nu} p_l Q(x, \xi^l) \,\middle|\, Ax = b, x \geq 0 \right\}$$

where

$$Q(x, \xi^l) = \inf_{y \in \mathbb{R}^{n_2}} \{qy \mid Wy = h^l - T^l x, y \geq 0\},$$

we have that

$$z^\nu \leq z^* = \inf\{cx + \mathcal{Q}(x) \mid Ax = b, x \geq 0\}.$$

Each z^ν is thus a lower bound for the optimal value of the stochastic program. (An alternative derivation of (4.5) relying on the dual of the recourse problem that defines $Q(x, \xi)$ appears in [7].)

4.6. Convergence. Suppose $\mathcal{S}^\nu = \{S^l, l = 1, \ldots, \nu\}$ for $\nu = 1, \ldots$, are partitions of Ξ with $\mathcal{S}^\nu \subset \mathcal{S}^{\nu+1}$ and chosen so that the P_ν, $\nu = 1, \ldots$ converge in distribution to P. The P_ν are the (molecular) probability distributions that assign probability $p_l = P[\xi(\omega) \in S^l]$ to the event $[\xi(\omega) = \xi^l]$ where ξ^l is the conditional expectation (with respect to P) of $\xi(\cdot)$ given that $\xi(\omega) \in S^l$. The epi-convergence of the $\{\mathcal{Q}^\nu, \nu = 1, \ldots\}$ to $\mathcal{Q}$, with the accompanying convergence of the solutions, follows from Theorem 2.8, where

$$\mathcal{Q}^\nu(x) = \sum_{l=1}^{\nu} p_l Q(x, \xi^l) = \int Q(x, \xi(\omega)) P_\nu(\mathrm{d}\omega).$$

To make use of these results we need to develop a sequential partitioning scheme for Ξ, i.e. given a partition $\mathcal{S}^\nu$ of Ξ how should it be refined so as to improve the approximation to $\mathcal{Q}$ as much as possible. P. Kall has also worked out various refinement schemes [24] that overlap and complement these given here.

4.7. Implementation. *Stochastic programs with simple recourse*, with h stochastic, q and T are fixed. Recall that for a stochastic program with simple recourse Ψ takes on the form:

$$\Psi(\chi) = E\left\{\sum_{i=1}^{m_2} \psi_i(\chi_i, \xi_i)\right\},$$

where $\xi_i = h_i$ and, as follows from (3.21),

$$\psi_i(\chi_i, \xi_i) = \begin{cases} q_i^+(h_i - \chi_i) & \text{if } h_i \geq \chi_i, \\ q_i^-(\chi_i - h_i) & \text{if } h_i \leq \chi_i. \end{cases}$$

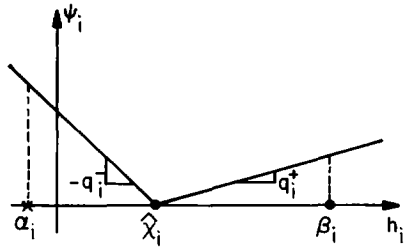


Fig. 4.8. The function $\psi_i(\chi_i, \cdot)$.

Let $[\alpha_i, \beta_i]$ be the support of the realizations of $h_i(\cdot)$, possibly an unbounded interval. If we are only interested in a lower bound for Ψ that approximates it as closely as possible at the point $\hat{\chi}$, then the optimal partitioning of $[\alpha_i, \beta_i]$ is given by

$$S_i^1 = [\alpha_i, \hat{\chi}_i) \quad \text{and} \quad S_i^2 = [\hat{\chi}_i, \beta_i].$$

In this way the approximating function Ψ_i^a takes on the form:

$$\Psi_i^a(\chi_i) = \begin{cases} q_i^+ \bar{h}_i - q_i^+ \chi_i & \text{if } \chi_i \leq h_i^1, \\ (q_i^+ h_i^2 p_{i2} - q_i^- h_i^1 p_{i1}) + (q_i^- p_{i1} - q_i^+ p_{i2})\chi_i & \text{if } h_i^1 \leq \chi_i \leq h_i^2, \\ -q_i^- \bar{h}_i + q_i \chi_i & \text{if } \chi_i \geq h_i^2 \end{cases}$$

where, for $l = 1, 2$,

$$h_i^l = E\{h_i(\omega) \mid S^l\} \quad \text{and} \quad p_{il} = P[h_i(\omega) \in S^l],$$

and $\bar{h}_i = E\{h_i(\omega)\}$. Note that

$$\Psi_i(\hat{\chi}_i) = q_i^- \int_{h_i(\omega) < \hat{\chi}_i} (\hat{\chi}_i - h_i(\omega)) P(\mathrm{d}\omega) + q_i^+ \int_{h_i(\omega) \geq \hat{\chi}_i} (h_i(\omega) - \hat{\chi}_i) P(\mathrm{d}\omega)$$

$$= \Psi_i^a(\hat{\chi}_i).$$

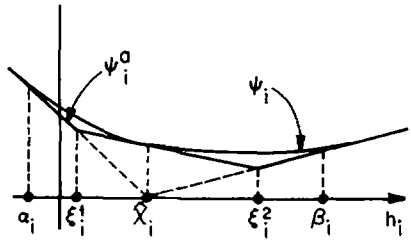


Fig. 4.9. The function Ψ^a.

Thus $\Psi_i^a \leq \Psi_i$ with equality holding for $\chi_i \leq \alpha_i$, $\chi_i \geq \beta_i$ and at $\chi_i = \hat{\chi}_i$. If the interval $[\alpha_i, \beta_i]$ has already been partitioned into ν intervals $\{[\alpha_i^0 = \alpha_i, \alpha_i^1), \ldots, [\alpha_i^{\nu-1}, \alpha_i^\nu = \beta_i]\}$ and $\hat{\chi}_i \in [\alpha_i^l, \alpha_i^{l+1})$. Then again the optimal subdivision of the interval $[\alpha_i^l, \alpha_i^{l+1})$ into $[\alpha_i^l, \hat{\chi}_i)$ and $[\hat{\chi}_i, \alpha_i^{l+1})$ yields an exact bound for Ψ_i at $\hat{\chi}_i$. An alternative is to split the interval under consideration around $\hat{\chi}_i$ such that $\hat{\chi}_i$ turns out to be the conditional expectation of the new region. This would provide a quite good bound for Ψ_i in the neighborhood of $\hat{\chi}_i$ and this would be very useful if the value of χ_i is not expected to change much in the next iterations.

4.10. Implementation. *General recourse matrix* W, with h stochastic; q and T are fixed. The function

$$h \mapsto \psi(\chi, h)$$

is not separable, it is convex and polyhedral (1.11). Note also that

$$h \mapsto \psi(\chi, h - \chi)$$

is a sublinear function. Because of this we shall say that $\psi(\chi, \cdot)$ is *sublinear with root at* χ. We assume that $\Xi \subset \mathbb{R}^{m_2}$ is a rectangle and that we are given a partition $\{S^l, l = 1, \ldots, \nu\}$ illustrated in Fig. 4.11. We shall take it for granted that the next

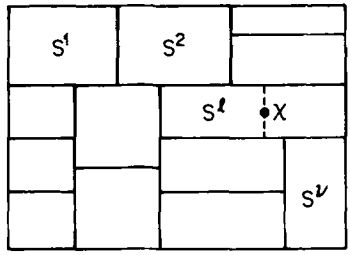


Fig. 4.11. Partition $\mathcal{S} = \{S^1, \ldots, S^\nu\}$ of Ξ.

partition of Ξ will be obtained by splitting one of the cells S^l. Other partitioning strategies may be used but this single cell approach has the advantage of increasing only marginally the linear program that needs to be solved in order to obtain the lower bound.

(i) Let us first consider the case when $\chi \in S^l \subset \Xi$. We plan to split S^l with a hyperplane containing χ and parallel to a face of S^l, or equivalently parallel to a hyperplane bounding the orthants. To do this, we study the behavior of $h \mapsto \psi(\chi, h)$ on each edge E_k of the cell S^l. Let

$$h \mapsto \theta_k(h) = \psi(\chi, h) : E_k \to \mathbb{R}.$$

This is a piecewise linear convex function. The possible shape of this function is illustrated in Fig. 4.12; by χ^p we denote the orthogonal projection of χ on E_k. If

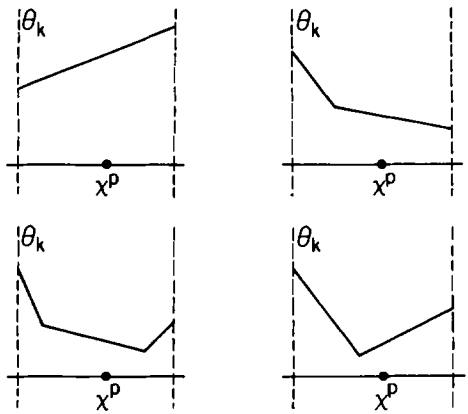


Fig. 4.12. The function θ_k on E_k.

θ_k is linear on E_k, it means that we cannot improve the approximation to θ_k by splitting S^l so as to subdivide E^k. On the other hand if the slopes of θ_k at the end points are different, then splitting S^l so as to subdivide E_k would improve the approximation to Ψ. On the subdivided cells, the resulting functions θ_k would be close to, if not actually, linear. Among all edges E_k, we would then choose to partition the cell S^l so as to subdivide the edge E_k that exhibits for θ_k the largest difference of slopes at the end points. What we need to know are the subgradients of the function

$$h \mapsto \psi(\chi, h)$$

at each vertex $\{h^s, s = 1, \ldots, r\}$ of the cell S^l. This is obtained by solving the linear programs

$$\text{find } \pi \in \mathbb{R}^{m_2} \quad \text{such that } \pi W \leq q \text{ and } w_s = \pi(h^s - \chi) \text{ is maximized} \qquad (4.13)$$

for $s = 1, \ldots, r$. The optimal π^s is a subgradient of $\psi(\chi, \cdot)$ at h^s [41, Proposition 7.12]. From this we obtain the directional subderivative of $\psi(\chi, \cdot)$ in each coordinate direction (which are the slopes of the functions θ_k); they are simply the components of the vector $\{\pi_i^s, i = 1, \ldots, m_2\}$. We now construct a subdivision of S^l with a hyperplane passing through χ and orthogonal to the edge of S^l that exhibits

maximum slope difference. If the underlying probability structure is such that the random vector $h(\cdot)$ is the sum of a few random variables, such as described by (1.19), the calculation of the directional subderivatives of $\xi \mapsto \psi(\chi, \xi)$ again begins with the calculation of the optimal solution of the programs (4.13) each h^s being obtained as the map of a vertex of S^l through the map (1.19). To obtain the subderivatives, we again need to use this transformation.

(ii) We now consider the case when $\chi \notin \Xi$. This time we cannot always choose a hyperplane passing through χ that generates a further subdivision of some cell S^l. Even when this is possible, it might not necessarily improve the approximation, the function $\xi \mapsto \psi(\chi, \xi)$ being linear on that cell for example. Ideally, one should then search all cells S^l and each edge in any given cell to find where the maximum gain could be realized. Generally, this is impractical. What appears reasonable is to split the cell with maximum probability, on which $\psi(\chi, \cdot)$ is not linear.

Concerning the implementation of this partitioning technique, we are seeking the best possible approximation to Ψ in the neighborhood of χ. We are thus working with the implicit assumption that we are in a neighborhood of the optimal solution and that χ will not change significantly from one iteration to the next. If this is the case, and the problem is well-posed, then we should not really have to deal with case (ii), since it would mean that the optimal tender χ^0 would be such that we would consistently underestimate or overestimate the demand!

4.14. Application. Consider the stochastic program with recourse with only q stochastic. With $\mathcal{S} = \{S^l, l = 1, \ldots, \nu\}$ a partition of Ξ, and for $l = 1, \ldots, \nu$, let

$$\xi^l = q^l = E\{q(\omega) \mid S^l\}$$

and $p_l = P[\xi(\omega) \in S^l]$. As follows from (1.12) and (4.3) we have

$$\sum_{l=1}^{\nu} p_l Q(x, \xi^l) \geq \mathcal{Q}(x). \tag{4.15}$$

Thus, with

$$z^\nu = \inf_{x \in \mathbb{R}^{n_1}} \left\{ cx + \sum_{l=1}^{\nu} p_l Q(x, \xi^l) \;\middle|\; Ax = b, x \geq 0 \right\}$$

where

$$Q(x, \xi^l) = \inf_{y \in \mathbb{R}^{n_2}} [q^l y \mid Wy = h - Tx, y \geq 0],$$

we have that

$$z^\nu \geq z^* = \inf[cx + \mathcal{Q}(x) \mid Ax = b, x \geq 0].$$

Each z^ν is thus an upper bound for the optimal value of the stochastic program.

4.15. Implementation. The function

$$q \mapsto \psi(\chi, q) = \max\{\pi(\xi - \chi) \mid \pi W \leq q\}$$

is polyhedral and sublinear. What changes from one χ to the next are the slopes of this function, so we cannot use the present χ as a guide for the design of the approximation. One possibility in this case is to simply subdivide a cell of the partition with maximum probability.

5. Discrete probability measures with support on extreme points

The maximum of a convex function on a compact convex set is attained at an extreme point; moreover, the function value at any point (of its domain) obtained as a convex combination of extreme points is dominated by the same convex combination of the function values at those extreme points. These elementary facts are used in the construction of measures that yield upper bounds for the expectation functional E_f.

5.1. Proposition. *Suppose $\xi \mapsto f(x, \xi)$ is convex, Ξ the support of the random variable $\xi(\cdot)$ is compact, and let* ext Ξ *denote the extreme points of* co Ξ, *the convex hull of Ξ. Suppose moreover that for all ξ, $\nu(\xi, \cdot)$ is a probability measure defined on* (ext Ξ, $\mathscr{E}$) *with $\mathscr{E}$ the Borel field, such that*

$$\int_{\text{ext}\,\Xi} e\nu(\xi, \mathrm{d}e) = \xi,$$

and the multifunction

$$\omega \mapsto \nu(\xi(\omega), A)$$

is measurable for all $A \in \mathscr{E}$. Then

$$E_f(x) \leq \int_{\text{ext}\,\Xi} f(x, e)\lambda(\mathrm{d}e) \tag{5.2}$$

where λ is the probability measure on $\mathscr{E}$ defined by

$$\lambda(A) = \int_{\Omega} \nu(\xi(\omega), A)P(\mathrm{d}\omega). \tag{5.3}$$

Proof. The convexity of $f(x, \cdot)$ implies that for the measure ν

$$f(x, \xi) \leq \int_{\text{ext}\,\Xi} f(x, e)\nu(\xi, \mathrm{d}e).$$

Substituting $\xi(\cdot)$ for ξ and integrating both sides with respect to P yields the desired inequality (5.2). □

5.4. Corollary. *Suppose $\xi \mapsto f(x, \xi)$ is convex, Ξ the support of the random variable $\xi(\cdot)$ is compact, and let* ext Ξ *denote the extreme points of* co Ξ, *the convex hull of*

Ξ. *Then*

$$E_f(x) \leq \sup_{e \in \operatorname{ext} \Xi} f(x, e) = f(x, e_x). \tag{5.5}$$

Proof. Simply follows from $f(x, e_x) \geq f(x, \xi)$ for all $\xi \in \Xi$, or we could use Proposition 5.1 with ν concentrated on e_x. □

5.6. Application. Consider the stochastic program with recourse (1.6) with only h and T stochastic. Assume that Ξ the support of the random variables $h(\cdot)$ and $T(\cdot)$ is compact, with

$$\operatorname{ext} \Xi = \{\xi^l = (h^l, T^l), l = 1, \ldots, L\}$$

the extreme points of co Ξ. We explicitly assume that L is finite. As usual

$$Q(x, \xi^l) = \min\{qy \mid Wy = h^l - T^l x, y \geq 0\}.$$

Then with

$$\xi^s \in \operatorname{argmax}\{Q(x, \xi^l), l = 1, \ldots, L\},$$

as follows from (5.5), we have that $\mathcal{Q}(x) \leq Q(x, \xi^s)$. Hence

$$\begin{aligned} z^* &= \inf\{cx + \mathcal{Q}(x) \mid Ax = b, x \geq 0\} \\ &\leq \inf\{cx + qy \mid Ax = b, T^s x + Wy = h^s, x \geq 0, y \geq 0\}. \end{aligned} \tag{5.7}$$

This is a very crude bound that can easily be improved by partitioning Ξ. Say $\mathcal{S} = \{S^k, k = 1, \ldots, \nu\}$ is a partition and for each k we compute $\xi^k \in \operatorname{argmax}_{\xi \in S^k} Q(x, \xi)$. Then

$$\mathcal{Q}(x) = \sum_{k=1}^{\nu} \int_{S^k} Q(x, \xi(\omega)) P(\mathrm{d}\omega) \leq \sum_{k=1}^{\nu} \int_{S^k} Q(x, \xi^k) P(\mathrm{d}\omega).$$

With $p_k = P[\xi(\omega) \in S^k]$, we obtain

$$z^* \leq \inf\left\{ cx + \sum_{k=1}^{\nu} p_k q y^k \;\middle|\; Ax = b, T^k x + W y^k = h^k, x \geq 0, y^k \geq 0 \right\}. \tag{5.8}$$

The potential use of this inequality as an approximation tool for solving stochastic programs with recourse was pointed out by Kall and Stoyan [23].

5.9. Application. We take the same situation as in Application 5.6. Let us define a probability measure $\nu(\xi, \cdot)$ on ext $\Xi = \{\xi^1, \ldots, \xi^L\}$, i.e. scalars $\{p_1(\xi), \ldots, p_L(\xi)\}$ such that $\sum_{l=1}^{L} p_l(\xi) = 1$ and

$$\xi = (h, T) = \sum_{l=1}^{L} p_l(\xi) \cdot (h^l, T^l). \tag{5.10}$$

Then

$$Q(x, \xi) \leq \sum_{l=1}^{L} p_l(\xi) Q(x, \xi^l)$$

and

$$\mathcal{Q}(x) \leq \sum_{l=1}^{L} \bar{p}_l Q(x, \xi^l) \tag{5.11}$$

where, for each $l = 1, \ldots, L$,

$$\bar{p}_l = \int p_l(\xi(\omega)) P(\mathrm{d}\omega).$$

The $\{\bar{p}_l, l = 1, \ldots, L\}$ determine a probability measure on ext Ξ. The right-hand term of (5.11) may however be quite difficult to compute since the dependence of the p_l on ξ may not be easy to express: they must be chosen so as to satisfy (5.10). There are some important cases when all of this can be worked out relatively easily. We review them next.

(i) Ξ (or co Ξ) *is a simplex.* Each ξ in Ξ has a *unique* representation in terms of the extreme points ext Ξ, viz. in terms of its barycentric coordinates. For example, if co Ξ is the fundamental simplex in $\mathbb{R}^N$ whose extreme points are

$$\{0, (e^j, j = 1, \ldots, N\} \quad \text{with } e^j = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathsf{T}},$$

then each point

$$\xi = (\xi_1, \ldots, \xi_N) \text{ in } \Xi$$

has the barycentric representation

$$p_0(\xi) = 1 - |\xi| \quad (p_l = \xi_l, l = 1, \ldots, N)$$

where $|\xi| = \sum_{l=1}^{N} \xi_l$. All other cases can be reduced to this example by an invertible linear transformation.

(ii) Ξ *is an interval.* This is a special case of the preceding one. Let $\Xi = [\alpha, \beta]$, then $\xi = (1-p)\alpha + p\beta$ with $0 \leq p \leq 1$; and thus $p = (\xi - \alpha)/(\beta - \alpha)$ from which it follows that

$$(1 - \bar{p}) = (\beta - \bar{\xi})/(\beta - \alpha) \quad \text{and} \quad \bar{p} = (\bar{\xi} - \alpha)/(\beta - \alpha)$$

with $\bar{\xi} = E\{\xi\}$. Thus

$$\mathcal{Q}(x) \leq (1 - \bar{p}) Q(x, \alpha) + \bar{p} Q(x, \beta). \tag{5.12}$$

This inequality is due to Edmundson. Madansky [31] used it in the context of stochastic programs with simple recourse random right-hand sides h. A much refined version of this upper bound can be obtained by partitioning the interval $[\alpha, \beta]$ and computing for each subinterval the corresponding version of (5.12). The expression for the $\bar{p}$ will now involve conditional expectations. For stochastic programs with simple recourse this was carried out by Ben-Tal and Hochmann [4], Huang, Ziemba, and Ben-Tal [20] and by Kall and Stoyan [23]; see also [17].

(iii) *Ξ is a rectangle and $\xi \mapsto Q(x,\xi) = \sum_{i=1}^{N} Q_i(x,\xi_i)$ is separable.* This is the case of stochastic programs with simple recourse with stochastic h, for example. Then

$$\mathcal{Q}(x) = \sum_{i=1}^{N} \int Q_i(x, \xi_i(\omega)) P(\mathrm{d}\omega).$$

We can now find bounds for $\mathcal{Q}$ by seeking bounds for each $Q_i(x,\cdot)$ separately. We are in the situation considered in (ii). The inequality (5.12) becomes

$$\mathcal{Q}(x) \leq \sum_{i=1}^{N} [(1-\bar{p}_i) Q_i(x,\alpha_i) + \bar{p}_i Q_i(x,\beta_i)] \tag{5.13}$$

where $[\alpha_i, \beta_i]$ is the support of the random variable $\xi_i(\cdot)$ and $\bar{p}_i$ defined as above.

(iv) *Ξ is a rectangle and the random variables are independent.* Let $F_i : \mathbb{R} \to [0,1]$ be the distribution function of the random variable ξ_i. We have that

$$\mathcal{Q}(x) = \int_{\alpha_N}^{\beta_N} \mathrm{d}F_N(\xi_N) \cdots \int_{\alpha_1}^{\beta_1} \mathrm{d}F_1(\xi_1) Q(x,(\xi_1,\ldots,\xi_N))$$

where $\Xi = \times_{i=1}^{N} [\alpha_i, \beta_i]$. With $\xi_2,\ldots,\xi_N$ fixed, for each $\xi_1 \in [\alpha_1,\beta_1]$, it follows from convexity that

$$Q(x,(\xi_1,\xi_2,\ldots,\xi_N)) \leq \frac{\beta_1-\xi_1}{\beta_1-\alpha_1} Q(x,(\alpha_1,\xi_2,\ldots,\xi_N)) + \frac{\xi_1-\alpha_1}{\beta_1-\alpha_1} Q(x(\beta_1,\xi_2,\ldots,\xi_N)).$$

Integrating on both sides with respect to $\mathrm{d}F_1$, and with $\bar{\xi}_1 = E\{\xi_1\}$ we have

$$\int_{\alpha_1}^{\beta_1} \mathrm{d}F_1(\xi_1) Q(x,(\xi_1,\xi_2,\ldots,\xi_N)) \leq \frac{\beta_1-\bar{\xi}_1}{\beta_1-\alpha_1} Q(x,(\alpha_1,\xi_2,\ldots,\xi_N)) + \frac{\bar{\xi}_1-\alpha_1}{\beta_1-\alpha_1} Q(x,(\beta_1,\xi_2,\ldots,\xi_N)).$$

We can now repeat this process for ξ_2, considering the two functions

$$\xi_2 \mapsto Q(x,(\alpha_1,\xi_2,\ldots,\xi_N), \qquad \xi_2 \mapsto Q(x,(\beta_1,\xi_2,\ldots,\xi_N).$$

One obtains

$$\begin{aligned}
&\int_{\alpha_2}^{\beta_2} \mathrm{d}F_2(\xi_2) \int_{\alpha_1}^{\beta_1} \mathrm{d}F_1(\xi_1) Q(x,(\xi_1,\xi_2,\ldots,\xi_N)) \\
&\quad \leq [(\beta_1-\alpha_1)(\beta_2-\alpha_2)]^{-1} [(\beta_1-\bar{\xi}_1)(\beta_2-\bar{\xi}_2) Q(x,(\alpha_1,\alpha_2,\ldots,\xi_N)) \\
&\qquad + (\beta_1-\bar{\xi}_1)(\bar{\xi}_2-\alpha_2) Q(x,(\alpha_1,\beta_2,\ldots,\xi_N)) \\
&\qquad + (\bar{\xi}_1-\alpha_1)(\beta_2-\bar{\xi}_2) Q(x,(\beta_1,\alpha_2,\ldots,\xi_N)) \\
&\qquad + (\bar{\xi}_1-\alpha_1)(\bar{\xi}_2-\alpha_2) Q(x,(\beta_1,\beta_2,\ldots,\xi_N))].
\end{aligned}$$

Doing this, in turn, for every $\xi_3, \ldots, \xi_N$ yields an upper bound for $\mathcal{Q}$ of the following type:

$$\mathcal{Q}(x) \leq \prod_{i=1}^{N} (\beta_i - \alpha_i)^{-1} \sum_{\gamma \in G} \left(\prod_{i=1}^{N} |\bar{\xi}_i - \gamma_i| Q(x, (\gamma_1, \ldots, \gamma_N)) \right) \tag{5.14}$$

where $|\cdot|$ is absolute value, and G is the collection of 2^N vectors defined by

$$G = \{\gamma = (\gamma_1, \ldots, \gamma_N) \,|\, \gamma_i = \alpha_i \text{ or } \beta_i,\ i = 1, \ldots, N\}.$$

One can also interpret (5.14) as follows: Let ext $\Xi = \{\xi^l, l = 1, \ldots, L = 2^N\}$ and now define on ext Ξ a probability measure ν which assigns probability p_l to ξ^l, where

$$p_l = \prod_{i=1}^{N} (1 - (|\bar{\xi}_i - \xi_i^l| / (\beta_i - \alpha_i))).$$

Note that this probability measure, suggested first in [23], yields an upper bound for $\mathcal{Q}$ that does not require passing through a transformation assigning to each ξ a particular combination of the extreme points.

(v) *Ξ is a polytope*, possibly a rectangle. Let $\mathcal{S} = \{S^k, k = 1, \ldots, \nu\}$ be a simplicial decomposition of Ξ, i.e. the partition is generated by a complex whose cells are simplices. Then in each cell we are in the situation described in (i). On each one we have an upper bound of type (5.11) for

$$\int_{\{\xi(\omega) \in S^k\}} Q(x, \xi(\omega)) P(\mathrm{d}\omega)$$

which we can then add up to obtain a bound for $\mathcal{Q}$. The bounds can be improved by refining the partition, for example. Another way is to consider for each ξ not just one possible representation, but look for the smallest upper bound given by a number of possible simplicial decompositions. Again, let $\{\xi^l, l = 1, \ldots, L\} = \text{ext } \Xi \subset \mathbb{R}^N$, and $\mathcal{P}$ the sets of all $(N+1)$-subsets of ext Ξ. Let $\mathcal{P}(\xi)$ be the elements of $\mathcal{P}$ such that ξ belongs to their convex hull. Then

$$Q(x, \xi) \leq \min_{\{\xi^{l_0}, \ldots, \xi^{l_N}\} \in \mathcal{P}(\xi)} \left\{ \sum_{j=0}^{N} p_j(\xi) Q(x, \xi^{l_j}) \,\middle|\, \sum_{j=0}^{N} p_j(\xi) \xi^{l_j} = \xi \right\}. \tag{5.15}$$

Integrating on both sides, after replacing ξ by $\xi(\omega)$, gives the desired upper bound on $\mathcal{Q}(x)$, and thus also on z^* as defined by (5.7). A last suggestion, in this general case, is for Ξ a rectangle but the $\{\xi_j(\cdot), j = 1, \ldots, N\}$ not independent. We still have that for all j, $\xi_j \mapsto Q(x, \xi)$ is convex. Set $j = 1$. Using (5.5), and with F the distribution function of $\xi(\cdot)$ on Ξ, we have that

$$\begin{aligned} \mathcal{Q}(x) &= \int Q(x, \xi)\, \mathrm{d}F(\xi) \\ &\leq \max_{\{\hat{\xi} = (\hat{\xi}_2, \ldots, \hat{\xi}_N) \,|\, \hat{\xi} \in \hat{\Xi}\}} \int_{\alpha_1}^{\beta_1} Q(x, (\xi_1, \hat{\xi}_2, \ldots, \hat{\xi}_N))\, \mathrm{d}F(\xi_1, \hat{\xi}) \end{aligned}$$

$$= \max_{\{\hat{\xi}=(\hat{\xi}_2,\ldots,\hat{\xi}_N)\,|\,\hat{\xi}\in\hat{\Xi}\}} \left[\left(\frac{\beta_1 - \bar{\xi}_1(\hat{\xi})}{\beta_1 - \alpha_1} \right) Q(x, (\alpha_1, \hat{\xi})) \right.$$

$$\left. + \left(\frac{\bar{\xi}_1(\hat{\xi}) - \alpha_1}{\beta_1 - \alpha_1} \right) Q(x, (\beta_1, \hat{\xi})) \right] \tag{5.16}$$

where $\bar{\xi}_1(\hat{\xi})$ is the conditional expectation of $\xi_1(\cdot)$ given $\hat{\xi}$ and where $\hat{\Xi}$ is the projection of Ξ into $\mathbb{R}^{N-1}$ through its first component. A bound of this type can be computed for each j and then we should choose the smallest one to bound $\mathcal{Q}$.

5.17. Application. Consider the stochastic program with recourse (1.15) with only q stochastic. If we now assume that Ξ, the support of the random variable $\xi(\cdot) = q(\cdot)$, is compact, all the bounds obtained for $\mathcal{Q}$ when h and T are stochastic have their counterparts in this case, except that this time we get lower bounds instead of upper bounds.

5.18. Implementation. We are in the same situation as in Section 4. Given a partition (simplicial decomposition, interval subdivision, or a rectangular cell splitting case such as illustrated by Fig. 4.11) the question which arises is to find a refinement of the partition that adds only a few cells and improves the approximation as much as possible. In practice, this boils down, as in Implementation 4.10, to subdividing just one cell. The piecewise linear character of $\xi \mapsto Q(x, \xi)$ plays the predominant role; as a matter of fact, all the arguments used to justify subdivision by a hyperplane passing through χ still apply. We would thus follow the same strategies as those suggested in Implementations 4.7, 4.10 and 4.15. The situation is illustrated by considering the simple recourse case (with h stochastic). Then $h_i \mapsto \psi_i(\chi_i, h_i)$ is a one-dimensional piecewise linear function. If $[\alpha_i, \beta_i]$ is the support of $h_i(\cdot)$, we have as a first bound

$$\Psi_i(\chi_i) \leq \frac{\beta_i - \bar{h}_i}{\beta_i - \alpha_i} \psi_i(\chi_i, \alpha_i) + \frac{\bar{h}_i - \alpha_i}{\beta_i - \alpha_i} \psi_i(\chi_i, \beta_i).$$

Subdividing $[\alpha_i, \beta_i]$ at $\hat{\chi}_i$ we get

$$\Psi_i(\chi_i) \leq \gamma_\alpha \psi_i(\chi_i, \alpha_i) + \hat{\gamma} \psi_i(\chi_i, \hat{\chi}_i) + \gamma_\beta \psi_i(\chi_i, \beta_i)$$

where

$$\gamma_\alpha = (\hat{\chi}_i - E\{h_i(\cdot) \,|\, h_i(\omega) \in [\alpha_i, \hat{\chi}_i)\})/(\hat{\chi}_i - \alpha_i),$$

$$\gamma_\beta = (E\{h_i(\cdot) \,|\, h_i(\omega) \in [\hat{\chi}_i, \beta_i]\} - \hat{\chi}_i)/(\beta_i - \hat{\chi}_i),$$

$$\hat{\gamma} = (E\{h_i \,|\, [\alpha_i, \hat{\chi}_i)\} - \alpha_i)(\beta_i - E\{h_i \,|\, [\hat{\chi}_i, \beta_i])/(\hat{\chi}_i - \alpha_i)(\beta_i - \hat{\chi}_i).$$

This is a much tighter bound for general χ_i and equality holds if $\chi_i = \hat{\chi}_i$. To illustrate what is going on, compare the graph of the approximating function a_1 to $\psi_i(\chi_i, \cdot)$ before subdividing at $\hat{\chi}_i$ and the graph of a_2 after subdivision (see Fig. 5.19).

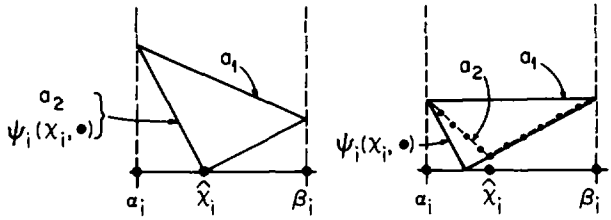


Fig. 5.19. Splitting $[\alpha_i, \beta_i]$ at $\hat{\chi}_i$.

5.20. Convergence. The same argument as that used to obtain Convergenc 4.6 in Section 4, again relying on Theorem 2.8, applies to this case. Except here, we start with Ξ compact and for all l, ξ^l is the extreme point of the cell S^l at which $\xi \mapsto Q(x, \xi)$ attains its maximum or minimum, depending on $Q(x, \cdot)$ being convex or concave.

6. Extremal probability measures

The use of extremal measures to construct upper and lower bounds is intimately related to a number of questions usually raised in the context of stochastic optimization with partial or incomplete information. In order to find a bound for

$$E_f(x) = \int f(x, \xi(\omega)) P(\mathrm{d}\omega)$$

we intend to replace P by another probability measure, say P_ν, that automaticallly guarantees

$$E_f(x) \geq \int f(x, \xi(\omega)) P_\nu(\mathrm{d}\omega) = E_f^\nu(x), \tag{6.1}$$

or its converse. One way to do this is to find a measure P_ν in a certain class $\mathscr{P}$ of probability measures on $(\Omega, \mathscr{A})$, which contains P, and that maximizes (or minimizes) the linear functional $P' \mapsto \int f(x, \xi(\omega)) P'(\mathrm{d}\omega)$ on the set $\mathscr{P}$. Since by assumption $P \in \mathscr{P}$, we have

$$\inf_{P' \in \mathscr{P}} \int f(x, \xi(\omega)) P'(\mathrm{d}\omega) \leq E_f(\chi) \leq \sup_{P' \in \mathscr{P}} \int f(x, \xi(\omega)) P'(\mathrm{d}\omega). \tag{6.2}$$

Note that the measures that minimize or maximize the preceding expressions in general depend on x, but not always. And if they do, quite often the same measure remains extremal for a relatively large neighborhood of x.

To exploit (6.2) in the search of upper and lower bounds for stochastic programming problems, the choice of $\mathscr{P}$ is of utmost importance. On one hand we want $\mathscr{P}$ to be 'narrow' enough to give us a measure in the immediate neighborhood of P; on the other hand, the chosen measure P_ν should be such that finding $E_f^\nu(x)$ is easy. In the context of the applications we have in mind, this means that P_ν should

be a discrete measure. One possibility is to define $\mathcal{P}$ as a neighborhood of P such as

$$\mathcal{P} = \{\text{prob. meas. } Q \,|\, \sup_{A \in \mathcal{A}} |P(A) - Q(A)| \leq \varepsilon\}, \tag{6.3}$$

or even

$$\mathcal{P} = \left\{\text{prob. meas } Q \,|\, \sup_{\theta \in \Theta} \left| \int \theta(\omega) P(\mathrm{d}\omega) - \int \theta(\omega) Q(\mathrm{d}\omega) \right| \leq \varepsilon \right\}, \tag{6.4}$$

where Θ is a class of test functions. Or with F the distribution function of $\xi(\cdot)$ defined on $\mathbb{R}^N$

$$\mathcal{P} = \{\text{dist. funct. } G \,|\, \sup_{z} |F(z) - G(z)| \leq \varepsilon\}. \tag{6.5}$$

The class $\mathcal{P}$ can be further restricted by limiting the acceptable class of measures to those having finite molecular support, etc.

The construction of bounds through extremal measures will however follow a quite different course; $\mathcal{P}$ will be defined by a finite number of equalities and inequalities which lead automatically to extremal measures with *finite* support. For a number of reasons that will become apparent later on, it is easier to work here with $\Xi \subset \mathbb{R}^N$ as the support of the measure P (technically, $\xi(\cdot)$ is then the identity map and $\Omega = \Xi$). So let $\mathcal{P}$ be the set of probability measures Q on $(\Xi, \mathcal{B}^N)$ that satisfy

$$\int_{\Xi} v_i(\xi) Q(\mathrm{d}\xi) \leq \alpha_i, \quad i = 1, \ldots, s, \tag{6.6}$$

$$\int_{\Xi} v_i(\xi) Q(\mathrm{d}\xi) = \alpha_i, \quad i = s+1, \ldots, M, \tag{6.7}$$

where M is finite and *the v_i are bounded continuous functions.* We shall always assume that $P \in \mathcal{P}$, i.e. satisfies the relations (6.6) and (6.7). The problem of finding a measure that satisfies these conditions and maximizes or minimizes

$$\int_{\Xi} v_0(\xi) Q(\mathrm{d}\xi) \tag{6.8}$$

where $v_0(\xi) = f(x, \xi)$, can be viewed as a *generalized moment problem* [28]. For problems of this type, we have the following general result:

6.9. Theorem. *Suppose Ξ is compact. Then the set $\mathcal{P}$ is convex and compact (with respect to the weak* topology), and $\mathcal{P} = \text{cl co}(\text{ext } \mathcal{P})$. Assuming that v_0 is continuous relative to Ξ, then $Q \mapsto \int v_0(\xi) Q(\mathrm{d}\xi)$ attains its optimum (maximum or minimum) at an extreme point of $\mathcal{P}$. Moreover the extremal measures of $\mathcal{P}$ are precisely those having*

finite (molecular) support $\{\xi^1, \ldots, \xi^L\}$ *with* $L \leq M+1$ *such that the vectors*

$$\begin{bmatrix} v_1(\xi^1) \\ \vdots \\ v_M(\xi^1) \\ 1 \end{bmatrix}, \ldots, \begin{bmatrix} v_1(\xi^L) \\ \vdots \\ v_M(\xi^L) \\ 1 \end{bmatrix}$$

are linearly independent.

Except for the presence of inequalities in the definition of $\mathcal{P}$, this result can be found in [26, Theorem 2.1]. Kemperman [27] shows that the supremum can be obtained if v_0 is upper semicontinuous, the v_i, $1 \leq i \leq s$, are lower semicontinuous and the v_i, $s+1 \leq i \leq m$, are continuous (with all functions appropriately dominated). We choose more restrictive assumptions since we work with both maxima and minima. Dupačová's minimax approach to stochastic programming [48] led to the investigation of extremal measures. She obtained results of this type [12, 13] for related moment problems. The proof we give here, based on generalized linear programming [10, Chapter 24], is due to Ermoliev, Gaivoronsky and Nedeva [15]. It is reproduced here because it is constructive and used in the sequel.

Proof. Suppose the constraints (6.6), (6.7), and

$$\int Q(\mathrm{d}\xi) = 1$$

are consistent, otherwise there is nothing to prove. The convexity of $\mathcal{P}$ is easy to verify, the compactness for the space of measures on a compact metric space follows from Prohorov's Theorem, and $\mathcal{P} = \mathrm{cl\ co}(\mathrm{ext}\ \mathcal{P})$ from the Krein-Milman Theorem about the representation of the elements of a convex set as convex combinations of its extremal points.

Now suppose that $\{\xi^1, \ldots, \xi^\nu\}$ is a finite collection of points of Ξ that we view as part of the potential support of the extremal measure that maximizes (6.8); the case of minimization of (6.8) involves the same arguments and does not need to be dealt with separately. The question now is to assign to these points $\{\xi^1, \ldots, \xi^\nu\}$ a probability distribution that maximizes (6.8). This can be expressed as a linear programming problem, with variables $\{p_1, \ldots, p_\nu\}$, formulated as follows:

$$\text{find } p_1 \geq 0, \ldots, p_\nu \geq 0$$

such that

$$\sum_{t=1}^{\nu} p_t = 1,$$

$$\sum_{t=1}^{\nu} v_i(\xi^t) p_t \leq \alpha_i \quad \text{for } i = 1, \ldots, s,$$

$$\sum_{l=1}^{\nu} v_i(\xi^l)p_l = \alpha_i \quad \text{for } i = s+1, \ldots, M, \text{ and}$$

$$z = \sum_{l=1}^{\nu} v_0(\xi^l)p_l \text{ is maximized.} \tag{6.10}$$

Assuming the points $\{\xi^1, \ldots, \xi^\nu\}$ have been picked so as to make this problem feasible, it is then also solvable. Let $\{p_l^\nu, l = 1, \ldots, \nu\}$ denote the optimal solution and let

$$(\theta^\nu, \pi_1^\nu, \ldots, \pi_s^\nu, \pi_{s+1}^\nu, \ldots, \pi_M^\nu) = (\theta^\nu, \pi^\nu)$$

be the simplex multipliers associated at the optimum to the constraints of (6.10). The measure determined by

$$Q_\nu[\xi^l] = p_l^\nu, \quad l = 1, \ldots, \nu,$$

is the desired extremal measure, unless some ξ in Ξ can be found such that

$$v_0(\xi) - \sum_{i=1}^{M} \pi_i^\nu v_i(\xi) - \theta^\nu > 0. \tag{6.11}$$

This follows directly from the optimality criteria for linear programs, when we note that each ξ in Ξ potentially generates a column that could be added to (6.10). If (6.11) holds for some ξ, let

$$\xi^{\nu+1} \in \operatorname{argmax}\left[v_0(\xi) - \sum_{i=1}^{M} \pi_i^\nu v_i(\xi) \,\middle|\, \xi \in \Xi \right]. \tag{6.12}$$

The existence of $\xi^{\nu+1}$ is not in question since the v_i are continuous and Ξ is compact. Adding the column

$$\begin{bmatrix} 1 \\ v_1(\xi^{\nu+1}) \\ \vdots \\ v_M(\xi^{\nu+1}) \\ v_0(\xi^{\nu+1}) \end{bmatrix}$$

to linear program (6.10) is guaranteed to yield a new solution $\{p_l^{\nu+1}, l = 1, \ldots, \nu+1\}$ and a measure

$$Q_{\nu+1}[\xi^l] = p_l^{\nu+1}, \quad l = 1, \ldots, \nu+1,$$

such that

$$\int v_0(\xi) Q_\nu(\mathrm{d}\xi) < \int v_0(\xi) Q_{\nu+1}(\mathrm{d}\xi).$$

Repeating this until the point ξ generated by (6.12) fails to satisfy (6.11) yields the extremal measure that maximizes (6.8). Since this is generalized linear programming,

the convergence proof of Dantzig [10, Chapter 24] applies; a variant appears in [43, Chapter 11.B] which can be exploited to obtain convergence of a modified procedure that only requires verification of (6.11) up to ε [15, Theorem 5], a most desirable feature in practice.

To complete the proof of the theorem, it suffices to observe that the optimal basis, associated with the solution of (6.10) will involve at most $M+1$ columns of the type $[1, v_1(\xi^t), \ldots, v_M(\xi^t)]^T$ that are linearly independent. And this holds for every possible objective $\int v_0(\xi)Q(d\xi)$, which by varying v_0 yields all extreme points of $\mathcal{P}$. □

Theorem 6.9 can now be applied to a variety of cases. The simplest one is $\Xi=[\alpha,\beta]\subset\mathbb{R}^1$, and the only condition is that the expectation with respect to P_ν should match the expectation $\bar{\xi}$ with respect to P. The problem reads:

find Q a measure on $(\Xi, \mathcal{B}^1)$

such that $Q\geq 0$, $\int_\alpha^\beta Q(d\xi)=1$, $\int_\alpha^\beta \xi Q(d\xi)=\bar{\xi}$

and $\int_\alpha^\beta v_0(\xi)Q(d\xi)$ is maximized.

Using the mechanism of the algorithm for generating P_ν, in particular (6.12), it is not difficult to see that

$$\text{with } v_0 \text{ concave}, \quad P_\nu\{\bar{\xi}\}=1, \tag{6.13}$$

$$\text{with } v_0 \text{ convex}, \quad P_\nu\{\alpha\}=\frac{\beta-\bar{\xi}}{\beta-\alpha},\ P_\nu\{\beta\}=\frac{\bar{\xi}-\alpha}{\beta-\alpha}. \tag{6.14}$$

This result and extensions thereof involving conditional expectation conditions, variance and unimodality conditions have been obtained and then applied to stochastic programming problems by Dupačová [12, 13, 14] and Cipra [9]. Observe that the extremal measure defined by (6.14) is precisely the discrete measure with support on extreme points obtained in Application 5.6 when Ξ is an interval (Case (ii)) and $\xi\mapsto v_0(\xi)=Q(x,\xi)$ is convex. In fact, many of the results obtained in Sections 4 and 5 can be recovered by a judicious application of Theorem 6.9, most often by relying on the further characterization of the support of the extremal measure given by the next theorem.

6.15. Theorem. *Suppose $\Xi\subset\mathbb{R}^N$ is compact,*

$$\mathcal{P}=\left\{Q\,\middle|\,\int_\Xi \xi Q(d\xi)=\bar{\xi}=\int_\Xi \xi P(d\xi)\right\},$$

and $v_0: \Xi \to \mathbb{R}$ *is convex. Then there exists*

$$P_\nu \in \operatorname*{argmax}_{Q \in \mathcal{P}} \int_\Xi v_0(\xi) Q(d\xi)$$

with finite (molecular) support $\{\xi^1, \ldots, \xi^L\}$ *with* $L \leq N+1$ *and*

$$\{\xi^1, \ldots, \xi^L\} \subset \operatorname{ext}(\operatorname{co} \Xi) \subset \Xi.$$

Proof. From Theorem 6.9 we already know that P_ν can be chosen with finite support in Ξ. Suppose

$$\xi^l \in \Xi \backslash \operatorname{ext}(\operatorname{co} \Xi) \quad \text{and} \quad P_\nu\{\xi^l\} = p_l > 0.$$

Then there exist $\{\xi^{lj} \in \operatorname{ext}(\operatorname{co} \Xi), j = 1, \ldots, J\}$ and nonnegative scalars $\{\lambda_j, j = 1, \ldots, J\}$ such that

$$\xi^l = \sum_{j=1}^{J} \lambda_j \xi^{lj}, \qquad \sum_{j=1}^{J} \lambda_j = 1.$$

By convexity of v_0,

$$p_l v_0(\xi^l) \leq \sum_{j=1}^{J} p_l \lambda_j v_0(\xi^{lj}).$$

Thus replacing P_ν which assigns probability p_l to ξ^l with P'_ν which assigns probability 0 to ξ^l and for $j = 1, \ldots, J$, probability $p_l \lambda_j$ to ξ^{lj} we have

$$\int v_0(\xi) P_\nu(d\xi) \leq \int v_0(\xi) P'_\nu(d\xi)$$

but still

$$\int \xi P'_\nu(d\xi) = \int \xi P_\nu(d\xi) = \bar{\xi}.$$

The argument shows that the search for an extremal measure can, in this case, be restricted to those having their support included in $\operatorname{ext}(\operatorname{co} \Xi)$. But this is a compact subset of $\mathbb{R}^N$. We complete the proof by applying Theorem 6.9 with

$$\mathcal{P} = \left\{ Q \text{ prob. meas. on } \operatorname{ext}(\operatorname{co} \Xi) \,\middle|\, \int \xi Q(d\xi) = \bar{\xi} \right\}. \qquad \square$$

6.16. Corollary [31; 17, Theorem 1]. *Suppose* K *is polytope that includes* Ξ, $\mathcal{P} = \{Q \mid \int_\Xi \xi Q(d\xi) = \bar{\xi}\}$ *and* $v_0: K \to \mathbb{R}$ *is convex. Then*

$$\int v_0(\xi) P(d\xi) \leq \max_{\lambda \geq 0} \left[\sum_{j=1}^{L} \lambda_j v_0(e^j) \,\middle|\, \sum_{j=1}^{L} \lambda_j e^j = \bar{\xi}, \sum_{j=1}^{L} \lambda_j = 1 \right] \tag{6.17}$$

where $\{e^1, \ldots, e^L\} = \operatorname{ext} K$.

Proof. Of course

$$\sup_{Q\in\mathcal{P}} \int v_0(\xi)Q(\mathrm{d}\xi) \leq \sup_{Q\in\mathcal{P}_K} \int v_0(\xi)Q(\xi)$$

where $\mathcal{P}_K = \{Q \text{ prob. meas on } K \mid \int_K \xi Q(\mathrm{d}\xi) = \bar{\xi}\} \supset \mathcal{P}$. We then apply Theorem 6.15 with Ξ replaced by K and $\mathcal{P}$ by $\mathcal{P}_K$. □

Reformulating this in terms of f and E_f this becomes:

6.18. Proposition. *Suppose $\xi \mapsto f(x, \xi)$ is convex. Then*

$$E_f(x) \leq \max_{\lambda \geq 0} \left[\sum_{j=1}^{L} \lambda_j h(e^j) \;\middle|\; \sum_{j=1}^{L} \lambda_j e^j = \bar{\xi},\ \sum_{j=1}^{L} \lambda_j = 1 \right] \tag{6.19}$$

where $\{e^1, \ldots, e^L\}$ is a (finite) collection of points in $\mathbb{R}^N$ such that $\mathrm{co}(e^1, \ldots, e^L) \supset \Xi$ and $h: \mathrm{co}(e^1, \ldots, e^L) \to \bar{\mathbb{R}}$ is a convex function such that

$$h \geq f(x, \cdot) \quad \text{on } \Xi.$$

In general, however, i.e. when other constraints than first moment conditions are part of the definition of $\mathcal{P}$ or the function v_0 is not convex or concave, what limits the use of Theorem 6.9 in practice is solving (6.12)! In general, the function

$$\xi \mapsto v_0(\xi) - \sum_{i=1}^{M} \pi_i^{\nu} v_i(\xi) \tag{6.20}$$

is neither convex nor concave, if some v_i is nonlinear, since the π_i^{ν} are not restricted in sign. The remainder of this section is concerned with how to handle this global optimization problem in the context of stochastic programs with recourse. We begin with the simplest case.

6.21. Application. Consider the stochastic program with *simple recourse* with random right-hand sides h, i.e. of the form (1.15) with $\psi(\chi, \xi)$ defined through (3.21) and (3.22). The problem is then separable and the function v_0 can be expressed as

$$v_0(\xi) = \sum_{j=1}^{m_2} v_{0j}(\xi_j) = \sum_{j=1}^{m_2} \psi_j(\chi_j, \xi_j)$$

and consequently is also separable. Since only marginal density information would be required in evaluating (6.8), the only sensible (generalized moment) conditions of type (6.6) or (6.7) would involve no more than one component of ξ at a time. Thus finding the maximum (or the minimum) of $v_0 - \sum_{i=1}^{M} \pi_i^{\nu} v_i$ is reduced to N $(= m_2)$ one-dimensional maximization problems that can be handled in practice in a number of ways, see also [8].

6.22. Implementation. We have to solve

$$\text{find } \xi \in [\alpha, \beta] \text{ such that } z = v_{0j}(\xi) - \sum_{i=1}^{M} \pi_i^{\nu} v_{ij}(\xi) \text{ is maximized.} \tag{6.23}$$

We consider the case when $M = 2$, $v_{ij}(\xi) = \xi$, $v_{2j}(\xi) = \xi^2$; we want to match the first two moments. The function $\phi(\xi) = v_{0j}(\xi)$ is convex (1.11). Reformulating (6.23) we have:

$$\text{find } \xi \in [\alpha, \beta] \quad \text{such that } \phi(\xi) - \pi_1^{\nu}\xi - \pi_2^{\nu}\xi^2 \text{ is maximized.} \tag{6.24}$$

If $\pi_2^{\nu} \leqslant 0$, the objective function is convex, in which case we only need to examine its values at the boundary points of the interval. If $\pi_2^{\nu} > 0$, the interval can be divided up into regions of convexity and concavity and on each one the maximum can be found by conventional methods. Another possibility when higher moments are involved, is to use the bounds on the expected value of a convex function, obtained by Don [11], for a class of sample-based probability measures. The optimal points of density $\xi^1, \ldots, \xi^l$ and the associated probabilities $p_1, \ldots, p_l$ are then straightforward to calculate provided the measure P has certain symmetries. When this last condition is not satisfied, we could still use the so-generated discrete measure to initialize an algorithmic procedure for solving (6.23).

6.25. Application. Consider the stochastic program with recourse (1.15) with random right-hand sides h. Suppose

$$\mathcal{S} = \{\Xi_i, i = 1, \ldots, M\}$$

is a partition of Ξ, for every $i = 1, \ldots, M$, v_i is the indicator function of Ξ_i and $\alpha_i = E\{h(\cdot) \mid \Xi_i\}$ is the conditional expectation of $h(\cdot)$ given Ξ_i. For $i = 1, \ldots, M$, let

$$p_i = P(\Xi_i) = P[h(\cdot) \in \Xi_i]$$

and again let $v_0(\xi) = \Psi(\chi, \xi)$. The problem of maximizing v_0 subject to (6.7) is then decomposable, in that each subregion Ξ_i can be dealt with separately. Indeed,

$$\int v_0(\xi) P(\mathrm{d}\xi) = \sum_{i=1}^{M} \int_{\Xi_i} v_0(\xi) P(\mathrm{d}\xi)$$

and thus the original problem decomposes into M subproblems of the type

$$\text{find a probability measure } Q_i \text{ on } \Xi_i$$

$$\text{such that } \int_{\Xi_i} \xi Q_i(\mathrm{d}\xi) = \alpha_i$$

$$\text{and } \int_{\Xi_i} v_0(\xi) Q_i(\mathrm{d}\xi) \text{ is maximized (or minimized).} \tag{6.26}$$

With P_i^ν the optimal solution to (6.26), the desired measure is given by

$$P^\nu = \sum_{i=1}^{M} p_i P_i^\nu.$$

Solving (6.26) is in principle not any easier than solving the general problem, except that we are only dealing with linear functions v_i (which means that the convexity of v_0 yields the convexity of the objective function of the subproblem (6.20)) and if the partition $\mathcal{S}$ of Ξ is left to us, we can choose it so that it corresponds to linear pieces of $\xi \mapsto \Psi(\chi, \xi)$.

6.27. Application. Consider the stochastic program with recourse (1.15) with random right-hand sides h, with the $h_i(\cdot)$ independent random variables for $i = 1, \ldots, m_2$; we also have that $v_0(\xi) = \Psi(\chi, \xi)$. With the independence of the random variables comes the separability of the constraints (6.6) and (6.7). We would thus have a relatively easy problem to solve if it were not for the intricate relationship between the $\xi_i = h_i$ that appears in the objective v_0.

6.28. Implementation. If we are interested in the probability measure that minimizes $\int v_0(\xi) Q(\mathrm{d}\xi)$ we can rely on the approximation to Ψ provided by Application 3.3. We have that

$$v_0(\xi) \geq \max_{1 \leq l \leq L} \eta^l(\xi - \chi)$$

where $\chi = Tx$ and, as in Application 3.3,

$$\eta^l \in \operatorname{argmax}[\eta(\xi^l - \chi) \mid \eta W \leq q]$$

for $\{\xi^l = h^l, l = 1, \ldots, L\}$ a finite number of realizations of $h(\cdot)$. Minimizing the function (6.20) that appears in the subproblem can then be expressed as

$$\text{find } \theta \in \mathbb{R}_1 \text{ and } (\xi_i \in \Xi_i, i = 1, \ldots, m_2)$$

$$\text{such that } \theta \geq \eta_l(\xi - \chi), l = 1, \ldots, L,$$

$$\text{and } \theta - \sum_{i=1}^{M} \pi_i^\nu v_i(\xi) \text{ is minimized.}$$

If, for example, the functions v_i correspond to first- and second-order moments, then this is a quadratic program, not necessarily convex. To solve it, we can rely on existing subroutines [16].

6.29. Implementation. If in Application 6.27 we are interested in the probability measure that maximizes $\int \Psi(\chi, \xi) Q(\mathrm{d}\xi)$, we rely instead on the approximation to $\Psi(\cdot, \chi)$ which comes from Application 3.23, which gives a separable function v_0, actually of the same type as for stochastic programs with simple recourse. This brings us to the case already studied in Application 6.21 and Implementation 6.22.

We note that the use of the approximating functions for the recourse function makes the calculation of extremal measures a reasonable undertaking. Otherwise, the global optimization problem of finding the $\xi \in \Xi$ that maximizes (or minimizes) the function $v_0(\cdot) - \sum_{i=1}^{M} \pi_i^{\nu} v_i(\cdot)$ would be too involved to solve exactly as a subproblem in obtaining error bounds. Finally, we observe that all the results derived here could be extended to h, T and q stochastic; each case, however, requires a separate analysis to take full advantage of the properties of the problem under consideration. As more information is gathered about these types of approximation and resulting bounds, we expect to see a more detailed analysis of each case. The use of these techniques in an overall scheme for solving stochastic programs with recourse also needs further study, here we have limited ourselves to finding extremal measures that yield the best possible lower and upper bounds for a given x or χ. Changing x only affects the function v_0 and this may affect the extreme points (see [12] and [13]). Often, however, all that may be needed when passing from x to another is a recalculation of the weight factors $p_1, \ldots, p_\nu$ in (6.10) the points $\{\xi^1, \ldots, \xi^\nu\}$ remaining unchanged. Moreover at each new x, it may not be necessary to solve the generalized moment problem to optimality.

6.30. Convergence. To obtain convergence, we need to consider sequences of generalized moment problems with an increasing number of restrictions on the moments of $\xi(\cdot)$. This must be done such that a sequence of extremal measures $\{P_\nu, \nu = 1, \ldots\}$ is obtained that converges in distribution to P. We may, for example, fit additional conditional mean information as in Section 4. We can then apply Theorem 2.8.

7. Majorizing probability measures

The role that convexity played in obtaining many of the bounds in the previous sections is taken over here by order preserving properties. The approximations are based on *stochastic ordering* [32, Chapter 17]. They are especially useful because of their simple calculability. The use of majorizing measures to approximate stochastic programs was first advocated in [46].

We denote by $\preceq_C$ the *partial ordering induced by* the closed convex pointed cone C on $\mathbb{R}^N$. We write

$$t^1 \preceq_C t^2 \quad \text{if} \quad t^2 - t^1 \in C \subset \mathbb{R}^N$$

and say that t^1 precedes t^2 (with respect to $\preceq_C$). A random vector $\xi^1 : \Omega \to \mathbb{R}^N$ *stochastically precedes* the random vector $\xi^2 : \Omega \to \mathbb{R}^N$ (with respect to $\preceq_C$) if

$$P\{\omega \,|\, \xi^1(\omega) \preceq_C \xi^2(\omega)\} = 1;$$

we write $\xi^1(\cdot) \preceq_C \xi^2(\cdot)$. A function ϕ from $\mathbb{R}^N$ into $\mathbb{R} \cup \{+\infty\}$ is *order preserving* with respect to $\preceq_C$ if

$$t^1 \preceq_C t^2 \quad \text{implies} \quad \phi(t^1) \leq \phi(t^2).$$

For ϕ order preserving and $\xi^1(\cdot) \preceq_C \xi^2(\cdot)$, obviously

$$E\{\phi(\xi^1(\omega))\} \leq E\{\phi(\xi^2(\omega))\}. \tag{7.1}$$

From this follows directly

7.2. Proposition. *Suppose $\xi \mapsto f(x, \xi)$ is order preserving with respect to $\preceq_C$ and for $i = 1, 2$, $\xi^i(\cdot) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^N, \mathcal{B}^N, F_i)$ are two random vectors such that $\xi^1(\cdot)$ stochastically precedes $\xi^2(\cdot)$. Then*

$$E_f^1(x) = \int f(x, \xi^1(\omega)) P(\mathrm{d}\omega) \leq \int f(x, \xi^2(\omega)) P(\mathrm{d}\omega) = E_f^2(\chi). \tag{7.3}$$

7.4. Application. Consider the stochastic program with recourse (1.15) with only $h(\cdot)$, the right-hand sides, stochastic. Let

$$\operatorname{pos} W = \left\{ t \,\middle|\, t = \sum_{j=1}^{n_2} W^j y_j,\ y_j \geq 0 \right\},$$

the convex cone generated by the columns of W, see (1.17). Let $\{t^l \in \mathbb{R}^{m_2},\ l = 1, \ldots, L\}$ be a frame for this polyhedral cone, i.e. the vectors t^l are positively linearly independent and $\operatorname{pos}(t^l, l = 1, \ldots, L) = \operatorname{pos} W$. Suppose that for all $l = 1, \ldots, L$, and $\xi \in \Xi$, the function

$$\lambda \mapsto \psi(\chi, \xi + \lambda t^l) : \mathbb{R}_+ \to \bar{\mathbb{R}}$$

is monotone increasing and that pos W is pointed. Then, if $\xi^L(\cdot) \preceq_{\operatorname{pos} W} \xi(\cdot)$

$$\Psi^L(\chi) := \int \psi(\chi, \xi^L(\omega)) P(\mathrm{d}\omega) \leq \Psi(\chi) \tag{7.5}$$

and, if $\xi(\cdot) \preceq_{\operatorname{pos} W} \xi^U(\cdot)$

$$\Psi(\chi) \leq \Psi^U(\chi) := \int \psi(\chi, \xi^U(\omega)) P(\mathrm{d}\omega). \tag{7.6}$$

This all follows directly from Proposition 7.2. It suffices to verify that the conditions imply that $\xi \to \psi(\chi, \xi)$ is order preserving with respect to $\preceq_{\operatorname{pos} W}$, details are worked out in [46, Proposition 3.2].

Below, in Application 7.8, we give an example where the monotonicity of ψ in each direction t^l can be verified directly. In other cases, one may have to rely on various properties of the problem at hand. The construction of the random variables $\xi^L(\cdot)$ and $\xi^U(\cdot)$ relies on subdividing the range of $\xi(\cdot)$ into subsets generated by the partial ordering $\preceq_{\operatorname{pos} W}$. This is done in [46, Section 3]. Convergence can be obtained by relying on finer and finer subdivisions of Ξ and by relying on a special form of Theorem 2.8. We shall concentrate instead on questions of implementability and special cases.

7.7. Application. Consider the stochastic program with recourse (1.15) with only $q(\cdot)$ stochastic. Let

$$D(W) = \{y \mid y \geq \pi W \text{ for some } \pi \in \mathbb{R}^{m_2}\}$$

and let $\{u^l, l = 1, \ldots, L\}$ be a frame for the convex polyhedral cone $D(W)$. Suppose that for all $l = 1, \ldots, L$ and $\xi \in \Xi$, the function

$$\lambda \mapsto \psi(\chi, \xi + \lambda u^l) : \mathbb{R}_+^N \to \bar{\mathbb{R}}$$

is monotone increasing, and

$$\xi^L(\cdot) \preceq_{D(W)} \xi(\cdot) \preceq_{D(W)} \xi^U(\cdot).$$

Then

$$\Psi^L(\chi) := \int \psi(\chi, \xi^L(\omega)) P(\mathrm{d}\omega) \leq \Psi(\chi) \tag{7.8}$$

and

$$\Psi(\chi) \leq \Psi^U(\chi) := \int \psi(\chi, \xi^U(\omega)) P(\mathrm{d}\omega). \tag{7.9}$$

To apply Proposition 7.2, we need to show that the monotonicity of $\lambda \mapsto \psi(\chi, \xi + \lambda u^l)$ for $l = 1, \ldots, L$ implies that $\xi \mapsto \psi(\chi, \xi)$ is order preserving. Suppose $\xi^1 \preceq_{D(W)} \xi^2$, then $\xi^2 - \xi^1 \in D(W)$ which means that

$$\xi^2 = \xi^1 + \sum_{l=1}^{L} \alpha_l u^l$$

for some scalars $\alpha_l \geq 0$. Relying on the monotonicity of ψ in each coordinate, we obtain:

$$\psi(\chi, \xi^1) \leq \psi(\chi, \xi^1 + \alpha_1 u^1) \leq \psi(\chi, \xi^1 + \alpha_1 u^1 + \alpha_2 u^2) \leq \cdots \leq \psi(\chi, \xi^2).$$

Note that h and q stochastic can be handled simultaneously provided naturally that the conditions laid out in Applications 7.4 and 7.7 be satisfied; this suggests some of the advantages of this approach. The real utility of this approach is, however, in the separable case.

7.8. Application. Consider a stochastic program with simple recourse with random right-hand sides h, i.e. of the form (1.15) with $\psi(\chi, \xi)$ separable as defined by (3.21) and (3.22). Suppose that for $i = 1, \ldots, m_2$, $q_i^+ \geq 0$ and $q_i^- \geq 0$, and define $\xi_i^L(\cdot)$ and $\xi_i^U(\cdot)$ as follows:

$$\xi_i^L(\cdot) \geq \xi_i(\cdot) \quad \text{on } \{\omega \mid \xi_i(\omega) \leq \chi_i\},$$

$$\xi_i^L(\cdot) \leq \xi_i(\cdot) \quad \text{otherwise}$$

and

$$\xi_i^U(\cdot) \leq \xi_i(\cdot) \quad \text{on } \{\omega \mid \xi_i(\omega) \leq \chi_i\},$$

$$\xi_i^U(\cdot) \geq \xi_i(\cdot) \quad \text{otherwise.}$$

Then

$$\Psi^{L}(\chi) := \sum_{i=1}^{m_2} \int \psi_i(\chi_i, \xi_i^{L}(\omega)) P(\mathrm{d}\omega) \leq \Psi(\chi)$$

and

$$\Psi(\chi) \leq \Psi^{U}(\chi) := \sum_{i=1}^{m_2} \int \psi_i(\chi_i, \xi_i^{U}(\omega)) P(\mathrm{d}\omega).$$

To see this, observe that the functions

$$\xi_i \mapsto \psi_i(\chi_i, \xi_i)$$

are monotone decreasing on $(-\infty, \chi_i]$, and monotone increasing on $[\chi_i, +\infty)$ since

$$\psi_i(\chi_i, \xi_i) = \begin{cases} q_i^-(\chi_i - \xi_i) & \text{if } \xi_i \leq \chi_i, \\ q_i^+(\xi_i - \chi_i) & \text{if } \xi_i \geq \chi_i. \end{cases}$$

Therefore, $-\psi_i(\chi_i, \cdot)$ is order preserving with respect to $\preceq_{\mathbf{R}_+}$ when $\xi_i \leq \chi_i$ and $\psi_i(\chi_i, \cdot)$ is order preserving with respect to $\preceq_{\mathbf{R}_+}$ when $\xi_i \geq \chi_i$. We apply Proposition 7.2 to obtain

$$-\int_{-\infty}^{\chi_i} \psi_i(\chi_i, \xi_i(\omega)) P(\mathrm{d}\omega) \leq -\int_{-\infty}^{\chi_i} \psi_i(\chi_i, \xi_i^{L}(\omega)) P(\mathrm{d}\omega)$$

and

$$\int_{\chi_i}^{\infty} \psi_i(\chi_i, \xi_i^{L}(\omega)) P(\mathrm{d}\omega) \leq \int_{\chi_i}^{\infty} \psi_i(\chi_i, \xi_i(\omega)) P(\mathrm{d}\omega).$$

Adding up these two inequalities and then summing with respect to i yields the assertion involving $\xi^{L}(\cdot)$. The symmetric inequality with $\xi^{U}(\cdot)$ is obtained similarly.

7.9. Implementation. The search for random variables $\xi_i^{L}(\cdot)$ and $\xi_i^{U}(\cdot)$ that yield the desired inequalities, can be carried out in terms of the distribution functions F_i^{L} and F_i^{U} induced by these random variables. Let F_i be the distribution function of $\xi_i(\cdot)$ with support $[\alpha_i, \beta_i]$. The conditions become

$$F_i^{L} \leq F_i \leq F_i^{U} \quad \text{on } (-\infty, \chi_i],$$

$$F_i^{L} \geq F_i \geq F_i^{U} \quad \text{on } [\chi_i, \infty).$$

Figure 7.10 gives an example of a discrete distribution F_i^{U} that could be used to approximate F_i. As usual, we are only interested in discrete approximations. Our goal is thus to find best discrete approximates that are below or above F. Since $F_i^{L} = F_i = F_i^{U}$ at χ_i we can find the best approximating distribution function that is below (or above) F_i on each segment $(-\infty, \chi_i]$ and $[\chi_i, +\infty)$ separately. And since below or above is just a question of reversing signs, we may as well consider the

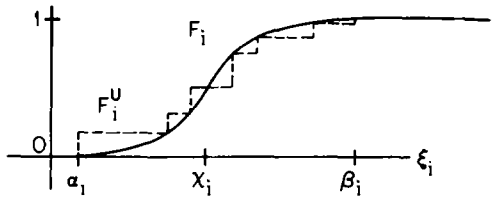


Fig. 7.10. Majorizing distribution function F_i^U.

problem at hand in the following framework:

find a distribution function $\hat{F}: \mathbb{R} \to [0, 1]$

such that $\hat{F} \leq F$, $\hat{F}$ is piecewise constant with at most L jumps

and $\int_{-\infty}^{+\infty} |F(s) - \hat{F}(s)| \, ds$ is minimized. (7.11)

We have defined best approximation in terms of the l_1-norm. Recalling that $\hat{F} \leq F$, we have

$$\int |F(s) - \hat{F}(s)| \, ds = \int F(s) \, ds - \int \hat{F}(s) \, ds$$

and thus we may as well simply maximize $\int \hat{F}(s) \, ds$ subject to $\hat{F} \leq F$. If $z_1, \ldots, z_L$ are the points of discontinuity of $\hat{F}$, it is easy to verify that $\hat{F}$ cannot be optimal unless at those points $\hat{F}(z_l) = F(z_l)$, $l = 1, \ldots, L$. Taking these observations into account, Problem (7.11) becomes

find $\alpha = z_0 \leq z_1 \leq z_2 \leq \cdots \leq z_l \leq z_{L+1} = \beta$

such that $\rho(z) = \sum_{l=1}^{L} (z_{l+1} - z_l)(F(z_l) - F(z_{l-1}))$ is maximized, (7.12)

where $[\alpha, \beta]$ is the support of the distribution function F. Note that ρ is not convex. Even with $L = 1$, when (7.12) reads

find $z \in [\alpha, \beta]$

such that $\rho(z) = (\beta - z)(F(z) - F(\alpha))$ is maximized, (7.13)

the solution is not necessarily unique, in fact the solution set may be a disconnected set of points. Assuming that F is twice differentiable with F' denoting the corresponding density, we have that z^* is optimal if

$$\frac{F(z^*) - F(\alpha)}{F'(z^*)} = \beta - z^* \geq 2 \frac{F'(z^*)}{F''(z^*)} \tag{7.14}$$

which in general has a multiplicity of solutions. To solve (7.12) we propose a heuristic that sequentially adjusts the jump points $z_1, \ldots, z_L$.

Step 0. Pick L points (for example with equal quantiles) in (α, β). Set $l = 0$.

Step 1. Set $l = l + 1$. Readjust z_l using the formula:

$$z_l^n \in \operatorname*{argmax}_{z \in [z_{l-1}^n, z_{l+1}]} (z_{l+1} - z)(F(z) - F(z_{l-1}^n)).$$

Solve using (7.14) exploring the local optima. Restart Step 1 if $l < L$; otherwise, go to Step 2.

Step 2. Stop if for all $l = 1, \ldots, L$, $|z_l^n - z_l| < \varepsilon$. Otherwise, return to Step 1 with $z_l := z_l^n$ for $l = 1, \ldots, L$ and $l = 0$.

This algorithm converges (a monotone increasing sequence bounded above by $\int F(s)\,ds$) but not necessarily to the optimal solution, this depends on the initial choice of $z_1, \ldots, z_L$.

An alternative approach to finding the best approximating discrete distribution function is to enter the points $z_1, \ldots, z_L$ with associated weights. These may correspond to the values of the recourse function, for example. With $v(\cdot)$ as the weighting function, Problem (7.12) becomes

$$\text{find } \alpha = z_0 \leq z_1 \leq \cdots \leq z_L \leq z_{L+1} = \beta$$

$$\text{such that } \rho(z) = \sum_{l=1}^{L} [v(z_{l+1}) - v(z_l)][F(z_l) - F(z_{l+1})] \text{ is maximized.} \tag{7.15}$$

In the case $L = 1$, we have a formula for the optimal z^* that corresponds to (7.14), and for the general case the same algorithm, with the obvious modifications, can be used as a heuristic. We could also use generalized programming, as in Section 6, to solve Problem (7.12) or (7.15). The problem corresponding to (7.25) is then

$$\text{find } p_j \geq 0, j = 1, \ldots, \nu$$

$$\text{such that } \sum_{i=1}^{j} p_i \leq F(z_j), j = 1, \ldots, \nu, \text{ and } \sum_{j=1}^{\nu} v(z_j) p_j \text{ is maximized} \tag{7.16}$$

where $p_j = \hat{F}(z_j) - \hat{F}(z_{j-1})$. For $v(z_j) \geq 0$, which is usually the case, the optimal solution is $p_j = F(z_j) - F(z_{j-1})$. The optimal dual variables associated to (7.16) are defined by

$$\sigma_\nu = v(z_\nu), \qquad \sigma_j = v(z_j) - \sum_{l=j+1}^{\nu} \sigma_l.$$

To add a new point $z_{\nu+1}$ that generates a new column of (7.16), we need to solve:

$$\max_{1 \leq j \leq \nu} \left[\max_{z \in [z_{j-1}, z_j]} \left\{ v(z) - \sum_{l=j+1}^{\nu} \sigma_l \right\} \right]. \tag{7.17}$$

This approach however does not lend itself easily to a fixed upper bound on the number of discontinuities of $\hat{F}$. It could be used to initialize the procedure suggested earlier.

7.18. Implementation. When Ψ is not separable, we can still proceed as in Implementation 7.9, if we first replace the recourse function by its simple recourse approximate, cf. Application 3.18, at least when seeking an upper bound for $\Psi(\chi)$.

8. Further bounds involving E_f or $\mathcal{Q}$

In this concluding section, we just want to record a number of bounds that require the evaluation of the objective functional E_f at some points. The use of these results is thus limited by our capability of evaluating E_f (or its gradient) with sufficient accuracy.

To begin, let us simply observe that for all $x \in \mathbb{R}^N$

$$\inf E_f \leq E_f(x), \tag{8.1}$$

which gives us a readily available upper bound. Using the subgradient inequality for convex functions we have:

8.2. Proposition. *Suppose $x \mapsto f(x, \xi): \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function. Then for any pair x, $\bar{x}$ in $\mathbb{R}^n$,*

$$E_f(x) - E_f(\bar{x}) \geq \bar{v} \cdot (x - \bar{x}) \tag{8.3}$$

with $\bar{v} \in \partial E_f(\bar{x})$, provided the set $\partial E_f(\bar{x})$ of subgradients of E_f at $\bar{x}$ is nonempty.

Proof. Simply observe that $f(\cdot, \xi)$ convex implies that E_f is convex which then implies (8.3). □

8.4. Application. Consider the stochastic program with recourse (1.6) with only $h(\cdot)$ stochastic. Then from [41, Corollary 7.16], we know that with $h = \xi$:

$$-E\{\pi(x, \xi)\}T \in \partial \mathcal{Q}(x),$$

where $\pi(x, \cdot): \Omega \to \mathbb{R}^{m_2}$ is a measurable function such that

$$\pi(x, \xi) \in \operatorname{argmax}\{\pi(\xi - Tx) \mid \pi W \leq q\}.$$

Thus, with f as defined by (1.9), we obtain

$$\mathcal{Q}(\hat{x}) \geq \mathcal{Q}(x) + E\{\pi(x, \xi(\omega))\}T(x - \hat{x}). \tag{8.5}$$

8.5. Implementation. Except for some special cases such as stochastic programs with simple recourse, evaluating $\mathcal{Q}(x)$ or $E\{\pi(x, \xi)\}$ is not feasible, but suppose that $\mathcal{Q}^L \leq \mathcal{Q}$ where $\mathcal{Q}^L$ has been obtained by relying on an approximating measure P_L. Then for any $\hat{x}$ we have that

$$\mathcal{Q}(\hat{x}) \geq \mathcal{Q}^L(\hat{x}) \geq \mathcal{Q}^L(x) + \int \pi(x, \xi(\omega))T(x - \hat{x})P_L(\mathrm{d}\omega),$$

with $\mathcal{Q}^L(x) = \int Q(x, \xi) P_L(d\xi)$. The term on the right can now be calculated and gives us a lower bound.

8.6. Application. Consider the stochastic program with recourse (1.15) with $q(\cdot)$ and $h(\cdot)$ stochastic. As usual

$$\Psi(\chi) = E\{\psi(\chi, \xi(\omega))\},$$

but let us now also define ρ as follows:

$$\rho(\chi) = \inf[cx \mid Ax = b, T = \chi, x \geq 0].$$

The stochastic program can then be formulated:

$$\text{find } \chi \in \mathbb{R}^{m_2} \quad \text{such that } \rho(\chi) + \Psi(\chi) \text{ is minimized.} \tag{8.7}$$

Suppose $\hat{\chi}$ is a point at which both ρ and Ψ are finite, and suppose $\hat{v} \in \partial\rho(\hat{\chi})$; the convexity of ρ follows from standard results in parametric linear programming. Let $\check{\chi}$ be such that

$$-\hat{v} \in \partial\Psi(\check{\chi}).$$

Assume such a point exists. For any $\chi \in \mathbb{R}^{m_2}$, it follows from the subgradient inequality for convex functions, in particular (8.3), that

$$\rho(\chi) - \rho(\hat{\chi}) \geq \hat{v}(\chi - \hat{\chi}) \quad \text{and} \quad \Psi(\chi) - \Psi(\check{\chi}) \geq -\hat{v}(\chi - \check{\chi}).$$

Adding up these two inequalities, we obtain that for all χ,

$$\rho(\chi) + \Psi(\chi) \geq \rho(\hat{\chi}) + \psi(\check{\chi}) - \hat{v}(\hat{\chi} - \check{\chi})$$

and hence

$$\inf(\rho + \Psi) \geq \rho(\hat{\chi}) + \Psi(\hat{\chi}) - \hat{v}(\hat{\chi} - \check{\chi}). \tag{8.8}$$

We have thus a lower bound for the infimum of the stochastic program.

We note that inequality (8.8) also follows from a duality argument. Assuming that all operations are well-defined:

$$\begin{aligned}\inf(\rho + \Psi) &= -(\rho + \Psi)^*(0) = -(\rho^* \square \Psi^*)(0) \\ &= -\inf_v(\rho^*(v) + \Psi^*(-v)) \\ &\geq -\rho^*(v) - \Psi^*(-v) \quad \text{for all } v,\end{aligned}$$

where $*$ denotes conjugacy and $\square$ inf-convolution. Inequality (8.8) now follows from the preceding one with $v = \hat{v}$ and observing that:

$$\rho^*(\hat{v}) = \hat{v}\hat{\chi} - \rho(\hat{\chi}), \qquad \Psi^*(-\hat{v}) = -\hat{v}\check{\chi} - \Psi(\check{\chi}).$$

This also shows that inequality (8.8) is sharp since

$$\inf(\rho + \Psi) = \sup_v[-\rho^*(v) - \Psi^*(-v)].$$

8.9. Implementation. Let us illustrate the use of this inequality in the case of a stochastic program with simple recourse with stochastic right-hand sides $h(\cdot)$. Suppose χ^e is, for possibly heuristic reasons, believed to be a good guess at the optimal tender (certainty equivalent). Let us now solve the linear program

$$\begin{aligned} &\text{find } x \in \mathbb{R}^n_+,\ u^+ \in \mathbb{R}^{m_2}_+,\ u^- \in \mathbb{R}^{m_2}_+ \\ &\text{such that } Ax \qquad\qquad\quad = b, \\ &\qquad\quad Tx + \ u^+ - \ u^- = \chi^e \quad \text{and} \\ &\qquad\quad cx + q^+u^+ + q^-u^- = z \quad \text{is minimized} \end{aligned} \tag{8.10}$$

where q^+, q^- are as usual the recourse costs. Let $(\hat{x}, \hat{u}^+, \hat{u}^-)$ be the optimal solution, and $(\hat{\sigma}, \hat{\pi})$ the associated simplex multipliers. Then

$$\hat{\pi} \in \partial\rho(\hat{\chi})$$

with $\hat{\chi} = T\hat{x}$ and ρ as defined in Application 8.6. Moreover, $\hat{\pi} \in [-q^-, q^+]$ as follows from the optimality conditions, and thus there exists $\check{\chi}$ such that

$$-\hat{\pi} \in \partial\Psi(\check{\chi})$$

as follows from the formula for subgradients of the recourse function in the simple recourse case [21, Chapter III, Section 4]. If for $i = 1, \ldots, m_2$, F_i denotes the distribution function of the random variable $h_i(\cdot)$,

$$\hat{\pi}_i = q_i^+ - q_i F_i(\check{\chi}_i)$$

where $q_i = q_i^+ + q_i^-$. With z^0 the optimal value of the stochastic program (1.15) we have

$$z^0 \geq c\hat{x} + \Psi(\check{\chi}) - \hat{\pi}(\hat{\chi} - \check{\chi}).$$

Let $\hat{z} = c\hat{x} + \Psi(\hat{\chi})$ which with the above yields

$$0 \leq \hat{z} - z^0 \leq \Psi(\hat{\chi}) - \Psi(\check{\chi}) + \hat{\pi}(\hat{\chi} - \check{\chi}). \tag{8.11}$$

In the case at hand, this becomes [33, Chapter III, Section 4]

$$0 \leq \hat{z} - z^0 \leq \sum_{i=1}^{m_2} q_i \int_{\check{\chi}_i}^{\hat{\chi}_i} (\zeta - \hat{\chi}_i)\, \mathrm{d}F_i(\zeta),$$

which is known as *Williams' inequality*. Let us point out that the path followed to obtain this last inequality, using (8.8) is quite different from the original proof of Williams [47] and should clarify the underpinnings of this result.

9. Preliminary computational report

The objective functional and probability measure approximations presented above are intended to be used together in solution procedures for stochastic programs. The characteristics of each objective functional approximation make it especially

amenable for use with certain probability measure approximations. We consider here the case of stochastic programs with recourse and discuss the merits of various objective functional-probability measure approximation pairs. The pairs which appear to be the most promising are marked by lines in Fig. 9.1.

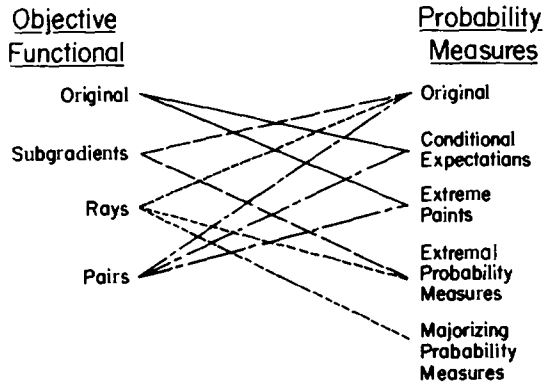


Fig. 9.1. Approximation pairs.

The original objective functional refers to the evaluation of $Q(\chi, \xi(\omega))$ as in (1.7). Since the integral in (1.7) may be extremely difficult to evaluate, the original probability measure must often be approximated. The simplest and most straightforward approximations that only require convex objective functionals are probably the conditional expectations and extreme point approximations. These approximations are linked to the original objective functional in Fig. 9.1 because of their simplicity and generality of application.

The subgradient objective approximation allows the original probability measure to be used when each region in the support Ξ is assigned a specific subgradient. In this case, the difficult integration in (1.7) is reduced to evaluating the probability of and the conditional expectation over each subgradient's region, where these regions can be chosen to be appropriately easy to evaluate. Subgradients are also well-suited to the extremal probability measure approximations because they can make the solution of (6.12) easier. If Ξ is partitioned into regions of concavity or convexity of $\sum_{i=1}^{M} \pi_i^r v_i(\xi)$, then the maximum can be found by checking optimality conditions for each of the subgradients without explicitly considering $v_0(\xi)$ in the solution.

The use of the ray function approximation is useful in similar circumstances to the subgradient approximation. In using the original probability measure, the simple recourse approximation 3.23 can be used in conjunction with an algorithm such as in Wets [42] to solve (1.6) using the ray approximation of $Q(x, \xi(\omega))$. The ray function approximation is also well-suited to the extremal probability measure approximation because the simple recourse formulation can be used to make the problem separable so that we are in the setting of Application 6.16. A further

advantage of this simple recourse formulation is that it can be used to achieve the conditions for the use of majorizing probability measures as in Application 7.8.

The pairs function approximation is basically a method for reducing the computational burden of a large deterministic equivalent program to (1.6). If the original distribution is finite, but perhaps too large for the deterministic equivalent to be easily solvable, then the pairs approximation may be used to obtain a bound by solving a number of small, closely-related programs. For continuous distributions, the pairs problem may also be used as a bounding technique. It may consider a large number of regions with conditional expectation or extreme point bounds and again obtain results through the solution of a series of small programs.

Computational studies in each of these areas are planned but only limited experience is currently available. Hausch and Ziemba [18], for example, have reported on the use of the original objective functional with conditional expectations and extremal support approximations. They found that these approximations provided good bounds for some simple production examples. They also noted that the improvement in the accuracy of the bounds is not necessarily strictly monotonic in the number of regions used in the approximation. Kall [24] has reported similar results in using these approximations as part of a solution procedure for (1.6). His initial experiments have shown that optimal solutions to (1.6) can often be obtained with very few regions used in the evaluation of the approximations. In some cases, however, the procedure of refining the partition in the region of greatest probability did not produce any improvement in the bounds and many iterations were required.

We have also observed this behavior in our initial investigations of the conditional expectation and extremal support applications. No improvements occur when refinements are made in regions of linearity of the recourse function. This led us to proposing the refinement procedure in 4.10 which produces monotonic improvement in the approximations. Our initial experiments have also shown that few refinements are necessary in many examples. This occurs because the recourse function is often 'flat', having little variation across the different regions in the partition of Ξ. In some examples, the recourse function, however, has more variation and more iterations are required.

We have begun limited experiments with the use of extremal probability measures and second moment conditions on the probability measures. Our initial results indicate that the master problem (6.10) and the subproblem (6.19) for simple recourse problems can be solved quickly with convergence established within five or six subproblem solutions. The resulting bounds provide a more precise interval around the simple recourse function than do a single expectation and extremal support bound. The utility of this bound as part of a solution procedure is still to be evaluated.

The emphasis in this paper has been on two-stage stochastic programs with recourse. The general functional form in (1.2), however, can also be applied to multistage stochastic linear programs. These problems have a wide variety of applications, including financial planning (Kallberg, White and Ziemba [25], Kusy and Ziemba [29]) and energy modeling (Louveaux and Smeers [30]). Their theoretical

properties are closely linked to those of two-stage programs and have been investigated, for example, by Olsen [33, 34] and Rockafellar and Wets [36, 38, 39]. Some initial results on approximation in multi-stage stochastic programs appear in Birge [5]. The application of the methods presented here in the multistage context should provide a foundation for the analysis and solution of these widely applicable models.

References

[1] H. Attouch, *Variational convergence for functions and operators*, Research Notes in Mathematics (Pitman, London, to appear).

[2] H. Attouch and R. Wets, "Approximation and convergence in nonlinear optimization", in: O. Mangasarian, R. Meyer and S. Robinson, eds., *Nonlinear Programming* 4 (Academic Press, New York, 1981), pp. 367-394.

[3] H. Attouch and R. Wets, "A convergence theory for saddle functions", *Transactions of the American Mathematical Society* 280 (1983) 1-41.

[4] A. Ben-Tal and E. Hochman, "More bounds on the expectation of a random variable", *Journal of Applied Probability* 9 (1972) 803-812.

[5] J. Birge, "Stochastic dynamic linear programs", Ph.D. Dissertation, Department of Operations Research, Stanford University (Stanford, CA, 1980).

[6] J. Birge, "The value of the stochastic solution in stochastic linear programs with fixed recourse", *Mathematical Programming* 24 (1982) 314-325.

[7] J. Birge and R. Wets, "Approximation and error bounds in stochastic programming", in: Y. Tong, ed., Inequalities in statistics and probability (Institute of Mathematical Statistics, Hayward, CA, 1984) pp. 178-186.

[8] A. Cabot, "Variations on a cutting plane method for solving concave minimization problems with linear constraints", *Naval Research Logistics Quarterly* 21 (1974) 265-279.

[9] T. Cipra, "Class of unimodal distributions and its transformations", *Časopis pro Pěstování Matematiky* 103 (1978) 17-26.

[10] G. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, 1963).

[11] H. Don, "Bounds on the expected value of a convex function in a sample-based class of probability measures", IIASA notes (Laxenburg, Austria, 1983).

[12] J. Dupačová, "Minimax stochastic programs with nonconvex separable penalty functions", in: A. Prékopa, ed., *Progress in operations research* (North-Holland, Amsterdam, 1976) pp. 303-316.

[13] J. Dupačová, "Minimax approach to stochastic linear programming and the moment problem", *Zeitschrift für Angewandte Mathematik und Mechanik* 58 (1978) 466-467.

[14] J. Dupačová, "The minimax approach to stochastic programming and an illustrative application", *Stochastics* 11 (1984).

[15] Y. Ermoliev, A. Gaivoronsky and C. Nedeva, "Stochastic optimization problems with partially known distribution functions", *SIAM Journal on Control and Optimization*, to appear.

[16] V. Garcia-Palomares and A. Restuccia, "Global space quadratic algorithm for solving a system of mixed equalities and inequalities", *Mathematical Programming* 21 (1981) 290-300.

[17] H.I. Gassman and W.T. Ziemba, "A tight upper bound for the expectation of a convex function of a multivariate random variable", *Mathematical Programming Study* (1985), forthcoming.

[18] D. Hausch and W. Ziemba, "Bounds on the value of information in uncertain decision problems II", *Stochastics* 10 (1983) 181-217.

[19] C. Huang, I. Vertinsky and W. Ziemba, "Sharp bounds on the value of perfect information", *Operations Research* 25 (1977) 315-325.

[20] C. Huang, W. Ziemba and A. Ben-Tal, "Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming", *Operations Research* 25 (1977) 315-325.

[21] P. Kall, *Stochastic linear programming* (Springer-Verlag, Berlin, 1976).

[22] P. Kall, "Computational methods for solving two-stage stochastic linear programming problems", *Zeitschrift fuer Angewandte Mathematik und Physik* 30 (1979) 261-271.

[23] P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds", *Mathematische Operationsforschung und Statistik Series Optimization* 13 (1982) 431-447.
[24] P. Kall, Lecture at IIASA, Laxenburg, Austria, May 1983.
[25] J. Kallberg, R. White and W. Ziemba, "Short term financial planning under uncertainty", *Management Science* 28 (1982) 670-682.
[26] A. Karr, "Extreme points of certain sets of probability measures, with applications", *Mathematics of Operations Research* 8 (1983) 74-85.
[27] J.H.B. Kemperman, "On a class of moment problems", in: L.M. Le Cam, J. Neyman, E.L. Scott, eds., 6th Berkeley symposium on mathematical statistics and probability, volume II: Probability theory (University of California Press, Berkeley-Los Angeles, 1972) 101-126.
[28] M. Krein and A. Nudelman, *The Markov moment problem and extremal problems*, Translations of Mathematical Monographs, 50 (American Mathematical Society, Providence, 1977).
[29] M. Kusy and W. Ziemba, "A bank asset and liability management model", Collaborative Paper CP-83-59, IIASA (Laxenburg, Austria, 1983).
[30] F. Louveaux and Y. Smeers, "A stochastic model for electricity generation", in: W. Haefele and L.K. Kirchmayer, eds., *Modelling of large-scale energy systems*, IIASA Proceedings Series 12 (Pergamon, Oxford, 1981) pp. 313-320.
[31] A. Madansky, "Inequalities for stochastic linear programming problems", *Management Science* 6 (1960) 197-204.
[32] A. Marshall and I. Olkin, *Inequalities: Theory of majorization and its applications* (Academic Press, New York, 1979).
[33] P. Olsen, "Multistage stochastic programs with recourse: the equivalent deterministic problem", *SIAM Journal on Control and Optimization* 14 (1976) 495-517.
[34] P. Olsen, "When is a multistage stochastic programming problem well defined?", *SIAM Journal on Control and Optimization* 14 (1976) 518-527.
[35] M. Perlman, "Jensen's inequality for a convex vector-valued function in an infinite dimensional space", *Journal of Multivariate Analysis* 4 (1974) 52-65.
[36] R.T. Rockafellar and R. Wets, "Nonanticipativity and L^1-martingales in stochastic optimization problems", *Mathematical Programming Study* 6 (1976) 170-187.
[37] R.T. Rockafellar and R. Wets, "Stochastic convex programming: basic duality", *Pacific Journal of Mathematics* 62 (1976) 173-195.
[38] R.T. Rockafellar and R. Wets, "Stochastic convex programming: relatively complete recourse and induced feasibility", *SIAM Journal on Control and Optimization* 14 (1976) 574-589.
[39] R.T. Rockafellar and R. Wets, "The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints", *SIAM Journal on Control and Optimization* 16 (1978) 16-30.
[40] J. Wang, "Distribution sensitivity analysis for stochastic programs with recourse", *Mathematical Programming*, to appear.
[41] R. Wets, "Stochastic programs with fixed recourse: The equivalent deterministic program", *SIAM Review* 16 (1974) 309-339.
[42] R. Wets, "Solving stochastic programs with simple recourse II", in: *Proceedings of the 1975 conference on information sciences and systems* (The Johns Hopkins University, Baltimore, 1975).
[43] R. Wets, *Grundlagen konvexer Optimierung*, Lecture Notes in Economics and Mathematical Systems, 137 (Springer-Verlag, Berlin, 1976).
[44] R. Wets, "Convergence of convex functions, variational inequalities and complementary problems", in: R. Cottle, F. Giannessi and J.-L. Lions, eds., *Variational inequalities and complementary problems* (J. Wiley and Sons, New York, 1980) pp. 375-403.
[45] R. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming: The state-of-the-art* 1982 (Springer-Verlag, Berlin, 1983) pp. 566-603.
[46] R. Wets, "Modeling and solution strategies for unconstrained stochastic optimization problems", *Annals of Operations Research* 1 (1984).
[47] A. Williams, "Approximation formulas for stochastic linear programming", *SIAM Journal on Applied Mathematics* 14 (1966) 668-677.
[48] J. Žáčková, "On minimax solutions of stochastic linear programming problems", *Časopis pro Pěstování Mathematiky* 91 (1966) 423-430.
[49] W. Ziemba, "Stochastic programs with simple recourse", in: P. Hammer and G. Zoutendijk, eds., *Mathematical programming in theory and practice* (North-Holland, Amsterdam, 1974) pp. 213-273.

Mathematical Programming Study 27 (1986) 103-117
North-Holland

A FIRST ORDER APPROACH TO A CLASS OF MULTI-TIME-PERIOD STOCHASTIC PROGRAMMING PROBLEMS

E.M.L. BEALE
Scicon Limited, Brick Close, Kiln Farm, Milton Keynes MK11 3EJ, England

G.B. DANTZIG
Stanford University, Stanford, CA 94305 USA

R.D. WATSON
Scicon Limited, Brick Close, Kiln Farm, Milton Keynes MK11 3EJ, England

Received 7 May 1984
Revised manuscript received 4 January 1985

There are many types of multi-time-period stochastic programming problems. In particular, there are problems where activities in one time period provide inventories or new capacities of uncertain magnitude for use in the next time period. One approach is then to ignore the uncertainties and solve a deterministic model using mean values. A slightly more sophisticated approach is to make first-order corrections to allow for the uncertainty. This paper suggests a strategy for computing such corrections. The problem of implementing this strategy is then studied by considering some very simple examples. These examples suggest that it may be seriously misleading to assume that all the relevant random variables are normally distributed unless the variance is small compared with the mean. This is because in reality the random variables are nonnegative. Fortunately the approach also works if the variables are assumed to have Gamma distributions.

Key words: Multi-Time Periods, Stochastic Programming, Exploration Activities, Approximate Solutions, Gamma Distributions.

1. Introduction

The task of finding an effective general method for solving multi-time-period stochastic linear programming problems has remained an elusive one since the problem was first posed, see Dantzig (1955). If the random variables have a discrete joint distribution, then the problem of choosing values for the first-time-period decision variables to optimize the expected value of the objective function reduces in principle to one of solving a large linear-programming problem. But the size of this problem becomes unmanageable except in very special cases. Beale et al. (1980) proposed a method for finding an approximate solution to problems with random right-hand sides representing uncertain sales demands, and this work has been extended by Ashford (1982). But other problem structures require other approaches. In some problems there are activities that provide either inventories or capacities for use in the next time period, where the extent of the new resource is a nonnegative random variable. The mathematical structure of the problem is then as follows:

If x_1 denotes the set of values of the first stage decision variables $x_{j1}(j = 1, \ldots, J_1)$ and x_R denotes the set of values of the remaining decision variables $x_{jt}(j = 1, \ldots, J_t)$ for $t > 1$, the problem is to choose x_1 to minimize

$$\sum_{j=1}^{J_1} a_{0j1}x_{j1} + E_{\alpha}\left(\min_{x_R} \sum_{t=2}^{T}\sum_{j=1}^{J_t} a_{0jt}x_{jt}\right)$$

subject to

$$\begin{aligned} &\sum_j a_{ij1}x_{j1} = b_{i1} \quad (i = 1, \ldots, I_1), \\ &-\sum_j \alpha_{ijt}x_{jt-1} + \sum_j a_{ijt}x_{jt} = b_{it} \quad (t = 2, \ldots, T,\ i = 1, \ldots, I_t), \end{aligned} \tag{1.1}$$

where the a_{ijt} and b_{it} are known constants, but the α_{ijt} are nonnegative random variables with a known joint probability distribution. We make the further assumption that the vectors of coefficients of the variables x_{jt-1} are statistically independent. In other words, while $\alpha_{i_1 j_1}$ and $\alpha_{i_2 j_2}$ may be correlated, $\alpha_{i_1 j_1 t_1}$ is independent of $\alpha_{i_2 j_2 t_2}$ unless $j_1 = j_2$ and $t_1 = t_2$.

We assume that the decision variables x_{j1} must be chosen without any further information, but that, for $t > 1$, the α_{ijt} will be known before the x_{jt} have to be chosen. We also assume that the problem has complete recourse, or in other words that there are no feasible sets of values of α_{iju+1} and x_{ju} for $u < t$ that make the problem infeasible with respect to the decision variables x_{jt}.

We are particularly interested in the special case where $0 \leq x_{jt} \leq 1$ for all variables having random coefficients. These variables then represent whether or not some activity with uncertain outcome is selected. The variable should then be restricted to integer values, but the LP relaxation may already provide adequate guidance.

We propose an approximate solution strategy that can be considered as a combination of linear and dynamic programming concepts (following Dantzig (1955)). The approach is conceptually similar to that used by Beale et al. (1980) and extended by Ashford (1982), and its effectiveness could usefully be tested by the type of simulation study made by Ashford. But the detailed assumptions and approximations are different. The problem of implementing this strategy is then studied by considering some very simple examples, taken from Watson (1983).

2. A recursive formulation

The state of the system when the x_{jt} have to be chosen is summarized by the vector $\boldsymbol{\beta}_t$, whose components β_{it} are defined by

$$\beta_{it} = b_{it} + \sum_j \alpha_{ijt}x_{jt-1} \quad (i = 1, \ldots, I_t). \tag{2.1}$$

If $V_t(\boldsymbol{\beta}_t)$ denotes the minimum value of $E(\sum_{u=t}^{T}\sum_j a_{0ju}x_{ju})$, when the x_{ju} are chosen to satisfy (1.1) for all $u \geq t$ and β_{it} is defined by (2.1), then we have the recurrence

relation, for $t = T+1, T, \ldots, 2$,

$$V_{t-1}(\boldsymbol{\beta}_{t-1}) = \min\left(\sum_j a_{0jt-1} x_{jt-1} + E(V_t(\boldsymbol{\beta}_t)) \right), \tag{2.2}$$

where the term $E(V_t(\boldsymbol{\beta}_t))$ is dropped for $t = T+1$, minimization being over all nonnegative x_{jt-1} satisfying

$$\sum_j a_{ijt-1} x_{jt-1} = \beta_{it-1} \quad (i = 1, \ldots, I_{t-1}) \tag{2.3}$$

with $\boldsymbol{\beta}_t$ being defined by (2.1), and the expectation being conditional on the given value of $\boldsymbol{\beta}_{t-1}$.

If we define β_{i1} as b_{i1}, then $V_1(\boldsymbol{\beta}_1)$ defines the optimum value of the objective function, and the corresponding values of x_{j1} are optimum.

3. The approximation

The first stage of the approximation procedure uses an idea introduced by Beale et al. (1980) and extended by Ashford (1982). To avoid working with separate subproblems defined by (2.2) and (2.3) for different values of $\boldsymbol{\beta}_{t-1}$, these subproblems are combined into a single subproblem. The quantities x_{jt-1} then represent the mean values of the decision variables averaged over all possible values of $\boldsymbol{\beta}_{t-1}$. Since the problem is linear in these variables, these mean values must satisfy (2.3) with the β_{it-1} replaced by their mean values, and we neglect the further constraints on the variables for individual realizations of $\boldsymbol{\beta}_{t-1}$. On the other hand, the variability in the vector $\boldsymbol{\beta}_t$ must allow for the variability in the input conditions for time period $t-1$ as well as the variability in the coefficients α_{ijt}. We therefore compute the first derivatives of the optimum values of the decision variables x_{jt-1} with respect to the components of $\boldsymbol{\beta}_{t-1}$. This provides linear approximations to the values of these variables as functions of these components, and hence we can use (2.1) to estimate the mean and covariance matrix for the components of $\boldsymbol{\beta}_t$ in terms of the mean and covariance matrix for the components of $\boldsymbol{\beta}_{t-1}$ and the means, variances and covariances of the α_{ijt}.

The mean values of the x_{jt} are denoted throughout by $\bar{x}_{jt}$. When $t = 1$, x_{jt} and $\bar{x}_{jt}$ are synonymous.

The next stage is to find an approximation to $E(V_t(\boldsymbol{\beta}_t))$ that is both adequate and computationally tractable. Even when $t = T$ this quantity represents the expected value of a wait-and-see stochastic programming problem, whose accurate solution in general requires a massive computing effort. And for smaller values of t the objective function has an additional component, being the expected value of the next wait-and-see stochastic programming problem.

So some further approximation is needed. A possible approach is suggested by considering the special case where only the first component β_{1t} of $\boldsymbol{\beta}_t$ is unknown.

Then we know from the theory of parametric linear programming that $V_t(\boldsymbol{\beta}_t)$ is a convex piecewise-linear function of β_{1t}. Further progress may be possible if we can approximate this function by a simple analytic form. One might first consider a convex quadratic approximation. If

$$V_t(\beta_t) = \theta_{0t} + \theta_{1t}\beta_{1t} + \theta_{2t}\beta_{1t}^2 \tag{3.1}$$

for some values of θ_{0t}, θ_{1t} and θ_{2t} with $\theta_{2t} > 0$, then

$$E(V_t(\boldsymbol{\beta}_t)) = \theta_{0t} + \theta_{1t}\mu_t + \theta_{2t}\mu_t^2 + \theta_{2t}\sigma_t^2, \tag{3.2}$$

where μ_t and σ_t^2 denote the mean and variance of β_{1t}, which can be computed as functions of the $\bar{x}_{jt-1}$ from (2.1) given the means and variances of the α_{1jt}.

This approximation implies that, when compared with the deterministic model where each α_{1jt} is replaced by its mean, $E(V_t)$ is increased by $\theta_{2t}\sigma_t^2$. It is therefore desirable to reduce this variance, but there is no particular incentive to choose a different value of μ_t except to the extent that a smaller value of μ_t may produce a smaller value of σ_t^2. This implication follows essentially from the symmetry of the function defined by (3.1). Such an approximation is therefore inappropriate for problems where it is important to achieve at least some specified minimum value of β_{1t}. In a stochastic model μ_t must then exceed this minimum value.

Now β_{1t} cannot be less than b_{1t}, because we assume that all $\alpha_{1jt} \geq 0$, but β_{1t} has no obvious upper bound. Since the problem has full recourse, we may assume an alternative approximation in which $V_t(\beta_{1t})$ is asymptotically linear as β_{1t} tends to infinity. Thus, as $\beta_{1t} \to \infty$,

$$V_t(\boldsymbol{\beta}_t) \to \theta_{0t} + \theta_{1t}\beta_{1t}$$

for some values of the parameters θ_{0t} and θ_{1t}. It may often be the case that $\theta_{1t} = 0$, but we do not need to assume this.

A convenient analytical formula for a convex function with this asymptotic behaviour is

$$V_t(\boldsymbol{\beta}_t) = \theta_{0t} + \theta_{1t}\beta_{1t} + \theta_{2t}\exp(-\theta_{3t}\beta_{1t}), \tag{3.3}$$

where θ_{2t} and θ_{3t} are positive. Then $E(V_t(\boldsymbol{\beta}_t))$ is related to the mean μ_t and variance σ_t^2 of β_{1t} by the formula

$$E(V_t(\boldsymbol{\beta}_t)) = \theta_{0t} + \theta_{1t}\mu_t + \theta_{2t}M_{\beta_{1t}}(-\theta_{3t}),$$

where $M_{\beta_{1t}}(s)$ denotes $E(\exp(s\beta_{1t}))$. The function $M_{\beta_{1t}}(s)$ is therefore the Moment Generating Function for β_{1t}. It is always convex, with $M_{\beta_{1t}}(0) = 1$.

If β_{1t} is normally distributed with mean μ_t and variance σ_t^2, then

$$M_{\beta_{1t}}(s) = \exp(\mu_t s + \tfrac{1}{2}\sigma_t^2 s^2). \tag{3.4}$$

But, since β_{1t} is assumed to be greater than or equal to b_{1t}, the assumption that it is normally distributed may be unrealistic. If we assume, following Watson (1983),

that $\zeta = \beta_{1t} - b_{1t}$ has a Gamma distribution, with probability density

$$\frac{(\zeta/k)^{n-1}e^{-\zeta/k}}{k\Gamma(n)} \quad \text{for } \zeta \geq 0,$$

the mean of the distribution is nk, the variance nk^2, and the moment generating function for ζ is

$$M_\zeta(s) = (1-ks)^{-n},$$

and hence

$$M_{\beta_{1t}}(s) = (1-ks)^{-n}\exp(b_{1t}s). \tag{3.5}$$

To compare this with the normal case, it is useful to express (3.5) in the form

$$M_{\beta_{1t}}(s) = \exp(\mu_t s + \tau_t\sigma_t^2 s^2), \tag{3.6}$$

where

$$\mu_t = b_{1t} + nk \quad \text{and} \quad \sigma_t^2 = nk^2.$$

Comparing (3.5) and (3.6) we see that

$$\tau_t\sigma_t^2 s^2 = -n\ln(1-ks) + b_{1t}s - \mu_t s.$$

so

$$\tau_t = -\frac{n}{\sigma_t^2 s^2}\ln(1-ks) - \frac{\mu_t - b_{1t}}{\sigma_t^2 s} = -\frac{1}{(ks)^2}\ln(1-ks) - \frac{1}{ks}. \tag{3.7}$$

This shows that τ_t is a slowly varying function of ks, where k is the ratio of the variance of $\beta_{1t} - b_{1t}$ to its mean. If ks is small, then $\ln(1-ks)$ can be expanded as a power series and we find that

$$\tau_t = \tfrac{1}{2} + \tfrac{1}{3}(ks) + \tfrac{1}{4}(ks)^2 + \cdots,$$

and hence (3.6) reduces to (3.4) when $k \to 0$.

In practice, although μ_t and σ_t^2 are functions of the decision variables and therefore k is also a function of the decision variables, we may be able to make an adequate first approximation to k and hence to τ_t a priori. Since $s = -\theta_{3t}$, it is always negative and τ_t is always less than its limiting value of $\frac{1}{2}$.

Given (3.6) we can write

$$E(V_t(\boldsymbol{\beta}_t)) = \theta_{0t} + \theta_{1t}\mu_t + \theta_{2t}\exp(-z_{1t}), \tag{3.8}$$

where

$$z_{1t} = \theta_{3t}\mu_t - \tau_t\theta_{3t}^2\sigma_t^2. \tag{3.9}$$

If α_{1jt} has mean $\bar{\alpha}_{1jt}$ and variance v_{1jt}, we see from (2.1) that

$$\mu_t = b_{1t} + \sum_j \bar{\alpha}_{1jt}\bar{x}_{jt-1}$$

and

$$\sigma_t^2 = \sum_j v_{1jt}\bar{x}_{jt-1}^2, \quad \text{if we neglect the variability in } x_{jt-1}.$$

But if the x_{jt-1} are now regarded as random variables, being approximately linear functions of the β_{it-1}, then the formula for μ_t is unaffected, but the formula for σ_t^2 becomes

$$\sigma_t^2 = \sum_j v_{1jt}\bar{x}_{jt-1}^2 + \eta_t,$$

where

$$\eta_t = \operatorname{var}(\sum_j \bar{\alpha}_{ijt}x_{jt-1}) + \sum_j v_{1jt}\operatorname{var}(x_{jt-1}).$$

Given an estimate of the covariance matrix for the components of $\boldsymbol{\beta}_{t-1}$ and estimates of the derivatives of x_{jt-1} with respect to the components of $\boldsymbol{\beta}_{t-1}$, we can compute η_t and treat this as a constant when computing appropriate values for the $\bar{x}_{jt-1}$.

So (3.8) and (3.9) reduce to

$$E(V_t(\boldsymbol{\beta}_t)) = \theta_{0t} + \theta_{1t}b_{1t} + \sum_j \theta_{1t}\bar{\alpha}_{1jt}\bar{x}_{jt-1} + \theta_{2t}\exp(-z_{1t}),$$

where

$$z_{1t} = \theta_{3t}b_{1t} + \sum_j (\theta_{3t}\bar{\alpha}_{1jt}\bar{x}_{jt-1} - \tau_t\theta_{3t}v_{1jt}\bar{x}_{jt-1}^2) - \tau_t\theta_{3t}^2\eta_t.$$

Thus $E(V_t(\boldsymbol{\beta}_t))$ is a linear function of the $\bar{x}_{jt-1}$ plus a convex function of the scalar argument z_{1t}, where z_{1t} is itself a sum of concave functions of the $\bar{x}_{jt-1}$. Thus $\exp(-z_{1t})$ is a convex function of $\bar{x}_{jt-1}$. So under these assumptions the problem is convex and separable. It becomes even easier if we can restrict attention to the case where $0 \leq x_{jt-1} \leq 1$ and interior values are of little interest. We can then approximate $\bar{x}_{jt-1}^2$ by $\bar{x}_{jt-1}$ and write

$$z_{1t} = \theta_{3t}b_{1t} + \sum_j (\theta_{3t}\bar{\alpha}_{1jt} - \tau_t\theta_{3t}^2v_{1jt})\bar{x}_{jt-1} - \tau_t\theta_{3t}^2\eta_t. \tag{3.10}$$

Note that if there are n identical investment opportunities defined by the same values of a_{ijt-1}, $\bar{\alpha}_{1jt}$ and v_{1jt}, we may let $\bar{x}_{jt-1}$ be a decision variable between 0 and n defining the number of these opportunities selected. Provided that the random variables are statistically independent, (3.10) will still hold.

But what shall we do if several components of $\boldsymbol{\beta}_t$ are unknown? We may assume that the first J_t are unknown. We might then be willing to assume that the main effects of the uncertainties in each component are approximately additive. More precisely, we might generalize (3.3) to read

$$V_t(\boldsymbol{\beta}_t) = \theta_{0t} + \sum_i \theta_{1it}\beta_{it} + \sum_i \theta_{2it}\exp(-\theta_{3it}\beta_{it}).$$

A further generalization can be made without making the analysis more complicated if we write

$$V_t(\boldsymbol{\beta}_t) = \theta_{0t} + \sum_i \theta_{1it}\beta_{it} + \sum_k \theta_{2kt} \exp(-\gamma_{kt}), \tag{3.11}$$

where

$$\gamma_{kt} = \sum_i \theta_{3kit}\beta_{it}, \tag{3.12}$$

provided that all $\theta_{2kt} \geq 0$ and all $\theta_{3kit} \geq 0$. This last assumption assures the correct asymptotic behaviour of $V_t(\boldsymbol{\beta}_t)$.

It is perhaps worth noting that this model requires assumptions about the univariate distributions of the random variables γ_{kt}, but does not require explicit assumptions about any multivariate distributions. This is convenient because there are not many suitable families of multivariate distributions other than the normal and lognormal, although the multivariate gamma distribution of Prékopa and Szantai (1978) should be noted.

If the mean of α_{ijt} is $\bar{\alpha}_{ijt}$, and the covariance of α_{i_1jt} and α_{i_2jt} is $c_{i_1i_2jt}$, then the mean value of γ_{kt} is

$$\sum_i \theta_{3kit} b_{it} + \sum_i \sum_j \theta_{3kit}\bar{\alpha}_{ijt}\bar{x}_{jt-1},$$

and its variance is $\sum_j v_{kjt}\bar{x}_{jt-1}^2 + \eta_{kt}$, where

$$v_{kjt} = \sum_{i_1} \sum_{i_2} \theta_{3ki_1t}\theta_{3ki_2t}c_{i_1i_2jt}, \tag{3.13}$$

and

$$\eta_{kt} = \operatorname{var}\left(\sum_i \sum_j \theta_{3kit}\bar{\alpha}_{ijt}x_{jt-1}\right) + \sum_j \sum_{i_1} \sum_{i_2} \theta_{3ki_1t}\theta_{3ki_2t}c_{i_2i_2j_t} \operatorname{var}(x_{jt-1}).$$

So if we again assume that the random variables have Gamma distributions such that the parameters τ_{kt} can be approximated a priori, and that the η_{kt} are estimated a priori from preliminary estimates of the covariance matrix for the components of $\boldsymbol{\beta}_{t-1}$ and the partial derivatives of the x_{jt-1}, then

$$E(V_t(\boldsymbol{\beta}_t)) = \theta_{0t} + \sum_i \theta_{1it}b_{it} + \sum_i \sum_j \theta_{1it}\bar{\alpha}_{ijt}\bar{x}_{jt-1} + \sum_k \theta_{2kt} \exp(-z_{kt}), \tag{3.14}$$

where

$$z_{kt} = \sum_i \theta_{3kit}b_{it} + \sum_i \sum_j \theta_{3kit}\bar{\alpha}_{ijt}\bar{x}_{jt-1} - \tau_{kt} \sum_j v_{kjt}\bar{x}_{jt-1}^2 - \tau_{kt}\eta_{kt}, \tag{3.15}$$

and again we may replace $\bar{x}_{jt-1}^2$ by $\bar{x}_{jt-1}$ if $0 \leq x_{jt-1} \leq 1$.

Given the values of θ_{0t}, θ_{1it}, θ_{2kt} and θ_{3kit}, the computation of optimum values of the $\bar{x}_{jt-1}$ therefore reduces to a convex separable-programming problem.

4. Estimating the parameters

Computational experiments will be needed to assess the value of this approach for any particular class of problem, and in particular how to choose the parameters θ_{0t}, θ_{1it}, θ_{2kt} and θ_{3kit}. A possible way to do this is now outlined.

Returning to the recursive formulation of Section 2, we see that, given the values of β_{iT}, the optimum decision variables x_{jT} for the last time period can be found by solving a linear programming problem. But, even given the values of β_{iT-1}, the optimum decision variables x_{jT-1} can only be found if we have a formula for $E(V_T(\boldsymbol{\beta}_T))$. This formula is given by (3.13), (3.14) and (3.15), but only in terms of parameters defining the approximate formula for $V_t(\boldsymbol{\beta}_T)$.

For any given vector $\boldsymbol{\beta}_T$, we can compute $V_T((\boldsymbol{\beta}_T)$ as the optimum objective function value of the last time period LP. And we can also compute $-\partial V_T/\partial\beta_{iT}$ as the shadow price π_{iT} on the ith constraint in (2.3). If (3.11) and (3.12) were exact, we would therefore have the relationships

$$\pi_{it} = \frac{\partial V_t(\boldsymbol{\beta}_t)}{\partial \beta_{it}} = -\theta_{1it} + \sum_k \theta_{2kt}\theta_{3kit} \exp(-\gamma_{kt}) \tag{4.1}$$

for $t = T$.

Given values of the quantities on the left hand sides, these relationships can be used to estimate the parameters.

So we can estimate the parameters by solving the last time period LP for a suitable set of values of the vector $\boldsymbol{\beta}_T$ and choosing the parameters so that (4.1) is approximately true.

The whole process can now be applied recursively to each earlier time-period in turn.

In the first instance we may set all the η_{kt} to their lower bounds of zero. Having computed first approximations to the values of the $\bar{x}_{jt}$ and the derivatives of x_{jt} with respect to the components of $\boldsymbol{\beta}_t$, we can compute estimated covariance matrices for the components of $\boldsymbol{\beta}_t$ and for the quantities x_{jt} for $t = 2, 3, \ldots, T$, since

$$\operatorname{cov}(\beta_{i_1 t}\beta_{i_2 t}) = \sum_{j_1}\sum_{j_2} \bar{\alpha}_{i_1 j_1 t}\bar{\alpha}_{i_2 j_2 t} \operatorname{cov}(x_{j_1 t-1}x_{j_2 t-1}) + \sum_j c_{i_1 i_2 jt}(\bar{x}^2_{jt-1} + \operatorname{var}(x_{jt-1})),$$

and $\operatorname{cov}(x_{j_1 t}x_{j_2 t})$ can be estimated from the covariance matrix for the components of $\boldsymbol{\beta}_t$ and the derivatives of x_{jt} with respect to these components.

These formulae can be applied immediately for $t = 2$, since $\operatorname{cov}(x_{j_1 1}x_{j_2 1}) = 0$ for all j_1 and j_2, and hence they can be applied recursively for each successive larger value of t. Hence more realistic values for the η_{kt} can be derived.

How might this be implemented?

It seems natural to start by solving a T-time-period LP problem in which all α_{ijt} are replaced by their mean values. This will produce a trial set of values for all x_{jt}, which can be used to compute approximate mean values, say $\tilde{\boldsymbol{\beta}}_t$ for $\boldsymbol{\beta}_t$ from (2.1).

We should ensure that the parameters satisfy (4.1) exactly when $\boldsymbol{\beta}_t = \tilde{\boldsymbol{\beta}}_t$. The model will then be correct when all the variances are very small, even when it is solved recursively as indicated in Section 2. We can choose θ_{0t-1} and the θ_{1it-1} by minimizing (2.2) subject to (2.3) and (2.1) when the β_{it-1} (for $i \leq J_{t-1}$) are all increased indefinitely. There should be no great difficulty in finding other plausible values of the β_{it-1} for which a good fit is desirable.

Even given the data to be fitted, the choice of parameter values may not be easy. But if we choose to minimize the sum of squares of the discrepancies between the left and right hand sides of (4.1), then, for given values of the θ_{3kit} the problem reduces to a quadratic programming problem. One might therefore choose a large number of values of k with fixed coefficients θ_{3kit}, and rely on the fitting program not choosing positive weights θ_{2kt} for more than a few of them - exploiting the fact that no θ_{2kt} may be negative.

5. Examples

To provide some practical insight into the problems of finding convenient analytical approximations to the solution to wait-and-see stochastic programming problems, two families of related very simple examples were studied by Watson (1983). The first represents an idealization of an oil exploration and recovery program. There are two time periods. In the first time period, exploration is carried out in one or both of two types of prospect. Each exploration venture may find either a large field, or a small field or nothing. In the second time period, oil may be produced, up to a maximum total capacity. For convenience this total capacity is defined as one unit.

The financial data are as follows:

Let B_i denote the net benefit from recovering a unit of oil from a field of Size i ($i = 1$ or 2), where $B_1 > B_2$.

Let C_j denote the cost of one unit of exploration activity of Type j ($j = 1$ or 2).

The probabilistic data are based on the assumption that there is a probability π_{ij} that one unit of exploration of Type j will find a field of Size i. The probabilities are assumed to be independent for different ventures. On this basis we can compute means and variances of any linear functions of the random variables occurring in the problem when the amounts of exploration are integer. We assume that the same formulae apply for fractional amounts of exploration.

A convenient mathematical formulation for this problem is as follows:

Let x_{j1} denote the amount of exploration activity of Type j.

Let x_{i2} denote the amount of oil produced from fields of Size i.

The problem is to choose x_{11} and x_{21} to maximize

$$-C_1x_{11}-C_2x_{21}+E(V(\boldsymbol{\beta}))$$

where

$$V(\boldsymbol{\beta}) \text{ is } \max(B_1x_{12}+B_2x_{22}))$$

subject to

$$x_{12} \leqslant \beta_{12},$$

$$x_{22} \leqslant \beta_{22},$$

$$x_{12}+x_{22} \leqslant 1.$$

and β_{12} has mean

$$\pi_{11}x_{11}+\pi_{12}x_{21}$$

and variance (5.1)

$$\pi_{11}(1-\pi_{11})x_{11}+\pi_{12}(1-\pi_{12})x_{21},$$

while $\beta_{12}+\beta_{22}$ has mean

$$(\pi_{11}+\pi_{21})x_{11}+(\pi_{12}+\pi_{22})x_{21} \tag{5.2}$$

and variance

$$(\pi_{11}+\pi_{21})(1-\pi_{11}-\pi_{21})x_{11}+(\pi_{12}+\pi_{22})(1-\pi_{12}-\pi_{22})x_{21}.$$

It turns out that β_{12} and $\beta_{12}+\beta_{22}$ are the relevant random variables. It has already been noted that the general approach of this paper requires data (or assumptions) about the distribution of various linear functions of the original random variables, which can be computed without assuming that the variables are statistically independent, but we do not need to known the joint distribution of the variables explicitly. In our example it would be unrealistic to assume that β_{12} and β_{22} were independent unless the π_{ij} were small.

Now if we consider a general wait-and-see linear programming problem with random right-hand sides and constant coefficients elsewhere, the tableau can be written as

$$x_0 = c_{00}+\sum_k c_{0k}\beta_k-\sum_j \bar{a}_{0j}x_j,$$

$$X_i = c_{i0}+\sum_k c_{ik}\beta_k-\sum_j \bar{a}_{ij}x_j \quad (i=1,\ldots,m),$$

where x_0 denotes the objective function value to be maximized, X_i denotes the ith basic variable, the β_k denote the random right hand sides and x_j denotes the jth nonbasic variable.

If $\hat{\beta}_k$ denotes a realistic value of β_k, for which the problem is feasible, then we can find a tableau such that

$$\hat{X}_i = c_{i0}+\sum_k c_{ik}\hat{\beta}_k \geqslant 0 \quad \text{for all } i.$$

If we now assume that there is a negligible probability that $c_{i0}+\sum_k c_{ik}\beta_k$ will be negative for more than one value of i, then the value of the solution to our LP problem will be

$$(c_{00}+\sum_k c_{0k}\beta_k)+\sum_i f_i(c_{i0}+\sum_i c_{ik}\beta_k) \tag{5.3}$$

where the $f_i(z_i)$ are the functions that are zero when $z_i \geq 0$ and can be computed by parametric programming for negative values of z_i. In fact if none of the variables has an upper bound,

$$f_i(z_i) = \min_{j,\bar{a}_{ij}<0} \left(\frac{\bar{a}_{0j}}{|\bar{a}_{ij}|}\right) z_i,$$

whenever $z_i < 0$.

Now our problem has the particularly convenient property that this approximation holds globally. We find that

$$V(\boldsymbol{\beta}) = B_1 + (B_1 - B_2) f(\beta_{12} - 1) + B_2 f(\beta_{12} + \beta_{22} - 1), \tag{5.4}$$

where

$$f(z) = \min(0, z). \tag{5.5}$$

Our general theory suggests that the function $f(z)$ should be approximated by an expression of the form

$$\hat{f}(z) = \phi_1 \exp(-\phi_2 z). \tag{5.6}$$

Since $z \geq -1$, it is natural to compute ϕ_1 and ϕ_2 to minimize the mean square difference between $f(z)$ and $\hat{f}(z)$ for $z \geq -1$. This leads to the values

$$\phi_1 = -0.1373, \qquad \phi_2 = 2.1491.$$

If we now assume that both $\beta_{12}-1$ and $\beta_{12}+\beta_{22}-1$ are normally distributed, it is reasonably easy to compute the expected net benefit and the corresponding values of x_{11} and x_{21} in the three following ways:

We can use LP, taking the mean values of β_{11} and β_{22} given by (5.1) and (5.2) and ignoring their variances.

We can use the true variances and the exponential approximation (5.6) to $f(z)$.

We can in this particular instance compute $E(V(\boldsymbol{\beta}))$ exactly and hence find optimum values of x_{11} and x_{21} as an unconstrained optimization exercise.

This was done using the following numerical data:

$$B_1 = 10.0, \quad C_1 = 1.5, \quad \pi_{11} = 0.2, \quad \pi_{12} = 0.1,$$
$$B_2 = 2.0, \quad C_2 = 1.0, \quad \pi_{21} = 0.2, \quad \pi_{22} = 0.5.$$

The results were as shown in Table 1. This shows that all is not well. At first sight, the value -1.4776 is absurd, because we know in fact that if $x_{11} = x_{21} = 0$ then the objective function will be zero. The error arises from an underestimate of $f(-1)$.

Table 1
Normally distributed random variables

Solution method	Objective function	x_{11}	x_{21}
LP	2.5000	5.0000	0.0000
Exponential	−1.4776	0.0000	0.8403
Exact	1.2158	0.6686	1.5034

One way of improving the approximation would be to choose the parameters ϕ_1 and ϕ_2 to give a better estimate for $f(z)$ for values of z near -1. But the fundamental cause of the error is the fact that the normal approximation to the distribution of β_{12} gives a finite probability that $\beta_{12}<0$, which is quite large in this instance; and this causes trouble because the exponential approximation to $f(z)$ is very inaccurate when $z<-1$.

If we replace the normal distributions of β_{12} and $\beta_{12}+\beta_{22}$ by Gamma distributions, as suggested in Section 3, we find that k, defined as the variance divided by the mean of the Gamma variate, is between 0.800 and 0.900 for β_{12} and between 0.400 and 0.600 for $\beta_{12}+\beta_{22}$. The values of τ are then given by (3.7) with $s=-\phi_2=-2.1491$.

This leads to a value of τ between 0.229 and 0.243 for β_{12} and between 0.277 and 0.324 for $\beta_{12}+\beta_{22}$. These ranges are both small, but both some way from the value 0.5 implied by the normal approximations.

When the model was recalculated assuming that both β_{12} and $\beta_{12}+\beta_{22}$ have Gamma distributions, the results were as shown in Table 2. The true objective function value associated with each solution method was calculated from the exact formula using the suggested values of x_{11} and x_{21}. The exponential method used values of τ derived from the LP values of x_{11} and x_{21}, which are 0.243 for β_{12} and 0.277 for $\beta_{12}+\beta_{22}$.

Table 2
Gamma distributed random variables

Solution method	objective function	x_{11}	x_{21}	True objective function value
LP	2.5000	5.0000	0.0000	−0.2756
Exponential	0.6383	2.3109	0.4223	0.8429
Exact	1.0050	0.0000	2.0424	1.0050

A comparison of Tables 1 and 2 shows that if the exact model for the function $f(z)$ is used, then the results do not depend much on whether the more convenient normal approximations to the distributions of β_{11} and $\beta_{12}+\beta_{22}$ or the more realistic Gamma distributions are used. It also shows that if the approximate exponential

model is used, the results are appreciably more realistic with the Gamma distributions. In larger stochastic programming models, the corresponding functions $f(z)$ may well look more like exponential functions than the function that occurs here, and in these circumstances the use of the Gamma distributions seems definitely worth while.

The fact that the optimum values of x_{11} and x_{21} are not accurately determined by our approximate method is not too worrying: the calculations show that the true objective function is very flat near the optimum.

Similar results were obtained using other values of the parameters B_i, C_j and π_{ij}. If $C_1 = 2.0$ but the parameters are otherwise unaltered the results were as shown in Table 3. This is one example where the exponential approximation gives slightly worse results than the LP approximation. This is disappointing, but the important result is that the true objective function value associated with the solution derived using the exponential approximation is always close to the calculated objective function, while that associated with the solution derived using the deterministic LP approximation is sometimes considerably smaller, as in our first problem.

Table 3

Solution method	Objective function	x_{11}	x_{21}	True objective function value
LP	1.6667	0.0000	1.6667	0.9807
Exponential	0.2866	0.0000	2.6762	0.9370
Exact	1.0050	0.0000	2.0424	1.0050

Finally we considered another model in which there is no exact solution of the form (5.3). This has the same distributions for β_{11} and β_{22} but requires the maximization of $E(V(\boldsymbol{\beta}))$ where

$$V(\boldsymbol{\beta}) = \max Bx_{12}$$

subject to

$$x_{12} \leqslant \beta_{12},$$

$$x_{12} \leqslant \beta_{22},$$

$$x_{12} \leqslant 1.$$

Here

$$V(\boldsymbol{\beta}) = B \min(\beta_{12}, \beta_{22}, 1). \tag{5.7}$$

We have the formula

$$V(\boldsymbol{\beta}) = B + Bf(\beta_{12} - 1) + Bf(\beta_{22} - 1), \tag{5.8}$$

which holds as long as either $\beta_{22} \geqslant 1$ or $\beta_{22} \geqslant 1$, but underestimates $V(\boldsymbol{\beta})$ when $\beta_{12} < 1$ and $\beta_{22} < 1$.

We calculated the solutions as if (5.8) were true globally, and also estimated the true objective function value associated with each solution by simulation, using the formula (5.7). The simulations were run 1000 times and the estimated standard errors are quoted. The parameter values were:

$$B = 14, \qquad C_1 = 1.5, \qquad C_2 = 1.0.$$

In all cases β_{12} and β_{22} were assumed to have Gamma distributions. The results were as shown in Table 4. The results suggest that, although the estimated objective function value is pessimistic, since (5.8) often underestimates $V(\boldsymbol{\beta})$, the values of x_{11} and x_{21} recommended by the exponential method appear to be quite good. Indeed the exponential approximation apparently gives slightly better results than the 'Exact' solution using (5.8), but our simulation results are not conclusive in this respect, and we did not consider that this particular issue was important enough to deserve further investigation.

Table 4

Solution method	Objective function	x_{11}	x_{21}	True objective function value from (5.7)
LP	6.5000	5.0000	0.0000	-0.2442 ± 0.1463
Exponential	0.2962	2.7132	3.6361	0.9225 ± 0.1509
'Exact', using (5.8)	0.6341	2.3739	3.8141	0.8834 ± 0.1539

6. Conclusion

The toy problems studied in Section 5 illustrate some of the potential, as well as some of the difficulties, of implementing the general strategy suggested in Section 3. We believe that the approach deserves to be studied further. The use of toy problems where exact results can be obtained helps to illuminate the method, and has led to what we believe is an important practical conclusion, namely that the use of normal approximations to nonnegative distributions is dangerous unless σ^2/μ is small, and that Gamma approximations should be preferred.

References

R.W. Ashford, "Multi-time-period stochastic programming for medium term production planning", Ph.D. Thesis, The School of Industrial and Business Studies, University of Warwick (Coventry, England, 1982).

E.M.L. Beale, J.J.H. Forrest and C.J. Taylor, "Multi-time-period stochastic programming", in: M.A.H. Dempster, ed., *Stochastic Programming* (Academic Press, London, 1980) pp. 387-402.

G.B. Dantzig, "Linear programming under uncertainty", *Management Science* 1 (1955) 197-206.

A. Prékopa and T. Szantai, "A new multivariate gamma distribution and its fitting to empirical streamflow data", *Journal of Water Resources* 14 (1978) 19-25.

R.D. Watson, "An approximate method of solution of the random resources multi-time-period stochastic programming problem", MSc Project Report, Mathematics Department, Imperial College, University of London (London, 1983).

Mathematical Programming Study 27 (1986) 118-132
North-Holland

AN APPROXIMATION SCHEME FOR STOCHASTIC DYNAMIC OPTIMIZATION PROBLEMS

Giovanni ANDREATTA
Dipartimento di Statistica, Università di Padova, I-35100-Padova, Italy

Wolfgang J. RUNGGALDIER*
Seminario Matematico, Università di Padova, I-35100 Padova, Italy

Received 10 January 1984
Revised manuscript received 6 August 1984

An approximation approach with computable error bounds is derived for a class of stochastic dynamic optimization problems that are too complex to be exactly solvable by straightforward dynamic programming. In particular, a problem arising from oil exploration is considered: for this problem, using the proposed approach, computational results are derived and compared to those obtained by means of other recent approximation schemes.

Key words: Optimization Under Uncertainty, Bayesian Approach, Dynamic Stochastic Optimization, Stochastic Control, Bayesian Dynamic Programming, Approximations, Error Bounds, Exploratory Oil Drilling.

1. Introduction

In this paper we study from a computational point of view a class of finite horizon dynamic stochastic optimization problems that includes a particular application to oil exploration. Such problems can most naturally be formulated as Bayesian stochastic control problems [5, 12, 16]. While most related work in the literature deals with Markovian or stationary models, here the problem is formulated in the framework of non-Markovian models and this is the subject of Section 2. For all such models a natural solution procedure is given by the method of stochastic dynamic programming (see e.g. [5, 12]). From a computational point of view however, dynamic programming may not be feasible due to the familiar problem of the 'curse of dimensionality', namely the problem of having the natural state space too large and this problem is particularly severe in non-Markovian models such as the one to be described here.

The purpose of this paper is to present for the model of Section 2, a procedure to construct an approximate solution along with an explicit error bound and this will be done in Section 3.

* Partially supported by the Italian Ministry of Education within the project 40% "Calcolo stocastico e sistemi dinamici stocastici".

The basic idea underlying our approach is that of shortening the horizon of action and changing the terminal value function. Such an idea is already discussed by Hinderer in [6] and [7] in the context of Markovian decision models with discounting. To apply his approach to the class of problems described in Section 2, one has to take as state space at every stage the entire set of admissible histories, so that the amount of computing is roughly of the same order of magnitude as is needed to exactly solve the original problem. A numerical comparison is given in Section 5.

Our approach can also be transformed to fit into a general framework proposed by Whitt in [19] for infinite horizon Markovian decision models with discounting. In fact, the basic idea underlying our approach is equivalent to restricting the dynamic programming to a subset of the set of all admissible histories and replacing the optimal value function outside this subset by upper and lower bounds, which is the basic idea underlying Whitt's approach. The main difference is that Whitt does not provide any procedure for defining the subsets and finding the bounds, whereas we explicitly present a constructive procedure for the non-Markovian finite horizon model described in Section 2.

For a very general theoretical framework to obtain bounds for dynamic programs, including those of Hinderer and Whitt, we refer the reader to [17].

In Section 4 further results are shown that are designed to improve the algorithmic aspects of our approach in various applications. In the last Section 5, a short description of the particular application to exploratory oil drilling is given along with numerical results and comparisons. For this application, an approximation approach along the lines of [18] and [7] has been studied in [13] exploiting the special structure of monotone Markov decision models (see e.g. [14]). In contrast to that approach, the present one is simpler to apply and requires considerably less computation.

2. The model

Consider a system that may be in one of a finite number M of possible *states*.

$$S := \{s_1, s_2, \ldots, s_M\} \tag{2.1}$$

with a prior probability p_i^0 of being in state s_i $(1 \leq i \leq M)$.

Assume the possibility is given to act on the system in a finite number of ways with the result of an action being both an observation described by a random variable, whose distribution depends on the actual state of the system, as well as a reward. Given a finite number N of periods, the purpose is to choose for every period, on the basis of the past history of actions and observations, an action so as to maximize the total expected reward.

To make the above more precise and building mainly on [5], we introduce various notations:

A: Finite set of a-priori possible *actions*, whose generic element is denoted by a.

$X_n \in \Gamma \subset \mathbb{R}^{k+1} (n = 1, \ldots, N)$: Random variable describing the *observation* in the generic period n; we assume Γ a finite set.

$\tilde{H}_n := A \times \Gamma \times A \times \Gamma \times \cdots \times \Gamma$ ($2n$ factors, $n = 1, \ldots, N$): Set of all a priori possible *histories* of actions and observations up to and including period $n \leq N$, whose elements are $h_n = (a_1, x_1, a_2, x_2, \ldots, a_n, x_n)$. By convention, set $\tilde{H}_0 := \{h_0\}$, i.e. 'no prior information'.

In general, the actions that are admissible in period n ($1 \leq n \leq N$) will depend on the past history, e.g. they may be subject to satisfy a series of constraints; so let

$$A_n(h_{n-1}) \subset A \tag{2.2}$$

denote the set of *admissible actions* in period n, given h_{n-1}.

The set H_n of *admissible histories* up to the period n is then defined recursively by

$$\begin{aligned} &H_0 := \tilde{H}_0, \\ &H_n := \{(h_{n-1}, a_n, x_n) \,|\, h_{n-1} \in H_{n-1},\ a_n \in A_n(h_{n-1}),\ x_n \in \Gamma\}. \end{aligned} \tag{2.3}$$

By *strategy* in period n we mean a mapping

$$f_n : H_{n-1} \to A \quad \text{with } f_n(h_{n-1}) \in A_n(h_{n-1}) \tag{2.4}$$

i.e. a rule, giving an admissible action for every history. An admissible *policy* is then a sequence $f = \{f_n\}_{n=1,\ldots,N}$ of strategies. We denote

F: set of all amissible policies, and, for any $h_n = (a_1, x_1, a_2, x_2, \ldots, a_n, x_n) \in H_n$,

$$F(h_n) := \{f \in F \,|\, f_m(h_{m-1}) = a_m;\ m = 1, 2, \ldots, n\}.$$

For $s \in S$, $h_n \in H_n$, $a \in A_{n+1}(h_n)$ assume given

$$P_a\{X_{n+1} = x_{n+1} \,|\, h_n, s\}, \quad x_{n+1} \in \Gamma, \tag{2.5}$$

namely the probability of observing x_{n+1} in period $n+1$ as a consequence of action a, given a current history h_n and given that the actual state of the system is s. We understand that this probability is zero if h_n cannot occur when the state is s.

For $f \in F(h_n)$ we now let

$$p_f(h_{n+1} \,|\, h_n, s) := \begin{cases} P_a\{X_{n+1} = x_{n+1} \,|\, h_n, s\} & \text{if } h_{n+1} = (h_n, a, x_{n+1}) \text{ with } a = f_{n+1}(h_n), \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Notice that $p_f(h_{n+1} \,|\, h_n, s)$ actually depends on f only through $a = f_{n+1}(h_n)$, so that we may equivalently write $p_a(h_{n+1} \,|\, h_n, s)$. From (2.6) we can compute for all $n > m \geq 0$ and $f \in F(h_m)$ the conditional probabilities $p_f(h_n \,|\, h_m, s)$. Furthermore let

$$p_f(h_n \,|\, s) := p_f(h_n \,|\, h_0, s). \tag{2.7}$$

Then, using the a priori distribution p_i^0, we have

$$p_f(h_n) = \sum_{i=1}^{M} p_f(h_n \,|\, s_i) p_i^0 \tag{2.8}$$

as well as

$$p_f(s_i \,|\, h_n) = \frac{p_f(h_n \,|\, s_i) p_i^0}{p_f(h_n)}. \tag{2.9}$$

Notice that, given h_n, the probability $p_f(s_i|h_n)$ does not depend explicitly on f, so that we may also write $p(s_i|h_n)$. Using (2.9), we finally have for $f \in F(h_m)$ and $n > m \geq 0$:

$$p_f(h_n|h_m) = \sum_{i=1}^{M} p_f(h_n|h_m, s_i)p(s_i|h_m). \tag{2.10}$$

For any given $h_n = (a_1, x_1, \ldots, a_n, x_n) \in H_n$ let now h_m^n $(1 < m \leq n)$ be defined as

$$h_m^n := (a_1, x_1, \ldots, a_m, x_m). \tag{2.11}$$

Then define, for $0 \leq m \leq n \leq N$ and $f \in F(h_m)$,

$$H_n(h_m, f) := \{h_n \in H_n | h_m^n = h_m,\ h_i^n = (h_{i-1}^n, f_i(h_{i-1}^n), x_i);\ m < i \leq n,\ x_i \in \Gamma\}, \tag{2.12}$$

$$H_n(h_m, f, s) := \{h_n \in H_n(h_m, f) | p_f(h_n|h_m, s) > 0\} \tag{2.13}$$

and notice that in many applications $H_n(h_m, f, s)$ can be considerably smaller than $H_n(h_m, f)$.

In the sequel we shall use expectations, whose meaning is as follows (G is any measurable function):

$$E_f\{G(h_n)|h_m, s\} = \sum_{h_n \in H_n(h_m, f, s)} G(h_n)p_f(h_n|h_m, s),\ m < n, f \in F(h_m), \tag{2.14}$$

$$E_f\{G(h_n)|h_m\} = \sum_{i=1}^{M} E_f\{G(h_n)|h_m, s_i)\}p(s_i|h_m),\quad f \in F(h_m), \tag{2.15}$$

$$E_f\{G(h_n)\} = E_f\{G(h_n)|h_0\}. \tag{2.16}$$

When $n = m+1$, letting $a = f_{m+1}(h_m)$, we shall prefer to write

$$E_a\{G(h_{m+1})|h_m, s\} \text{ instead of } E_f\{G(h_{m+1})|h_m, s\}$$

as well as

$$E_a\{G(h_{m+1})|h_m\} \text{ instead of } E_f\{G(h_{m+1})|h_m\}.$$

Finally, let $r_n(h_n)$ be the *reward* in period $n \leq N$, given the history h_n, and define

$$R_n(h_N) := \sum_{i=n}^{N} r_i(h_i^N). \tag{2.17}$$

By convention let

$$R_{N+1}(h_N) = 0. \tag{2.18}$$

Our original problem can then be described as that of finding $f^* \in F$ such that, for any $f \in F$,

$$E_{f^*}\{R_1(h_N)\} \geq E_f\{R_1(h_N)\}. \tag{2.19}$$

3. The approximation approach

We start with some preliminaries. For $f \in F(h_{n-1})$ let

$$V_n(h_{n-1}, f) := E_f\{R_n(h_N) | h_{n-1}\} \tag{3.1}$$

and

$$V_n(h_{n-1}) := \max_{f \in F(h_{n-1})} V_n(h_{n-1}, f). \tag{3.2}$$

Our problem can then be reformulated as that of finding $f^* \in F$ such that

$$V_1(h_0, f^*) = V_1(h_0) = \max_{f \in F} V_1(h_0, f) \tag{3.3}$$

and the dynamic programming algorithm (see e.g. [12, Section 5]) becomes

$$\begin{aligned} V_n(h_{n-1}) &= \max_{f \in F(h_{n-1})} E_f\{r_n(h_n) + V_{n+1}(h_n) | h_{n-1}\} \\ &= \max_{a \in A_n(h_{n-1})} E_a\{r_n(h_n) + V_{n+1}(h_n) | h_{n-1}\} \end{aligned} \tag{3.4}$$

for $n = 1, \ldots, N$, with (see (2.18)) $V_{N+1}(h_N) = 0$.

This algorithm allows us to compute the sequence of optimal values $\{V_n(h_{n-1})\}_{n=1,\ldots,N}$ for all admissible histories and to determine the optimal policy f^* which is given by

$$f_n^*(h_{n-1}) = \arg\max_{a \in A_n(h_{n-1})} E_a\{r_n(h_n) + V_{n+1}(h_n) | h_{n-1}\}. \tag{3.5}$$

However, we have to perform the computations in (3.4) starting from $n = N$ down to $n = 1$ taking into account all admissible histories, whose number increases exponentially with n. Assume now that for a certain, possibly small n $(1 < n < N)$, we know $V_n(h_{n-1})$ for all $h_{n-1} \in H_{n-1}$. Then we may restrict the dynamic programming only to the periods from 1 to $n < N$ and still obtain the optimal total expected reward $V_1(h_0)$. Now, if $V_n(h_{n-1})$ is not known and we do not want to perform all the computations required by dynamic programming, we may be satisfied with obtaining good approximations to the optimal value $V_1(h_0)$. For this purpose we may proceed in analogy to what is done by different techniques for infinite horizon Bayesian Markov control models in [16, Section 6], or for two stage recourse problems in [8] (see also the overview in [4]) where approximations to the optimal value are obtained by approximating the recourse function. We shall in fact show that for a given n $(1 \leq n \leq N)$ it is relatively easy to compute upper and lower bounds for $V_n(h_{n-1})$ that we shall denote by $U_n(h_{n-1})$ and $L_n(h_{n-1})$ respectively. Then, using dynamic programming over the periods from 1 to n starting from $U_n(h_{n-1})$ and $L_n(h_{n-1})$ respectively, by the monotonicity of the dynamic programming operator, we obtain in the first period an upper bound $U_1^n(h_0)$ and a lower bound $L_1^n(h_0)$, for $V_1(h_0)$. More precisely, in this Section we shall prove the following

Theorem 3.1. *For any given* n $(1 \leq n \leq N+1)$

$$L_1^n(h_0) \leq V_1(h_0) \leq U_1^n(h_0) \tag{3.6}$$

and, defining

$$\varepsilon^n := U_1^n(h_0) - L_1^n(h_0), \tag{3.7}$$

the sequence ε^n *is decreasing to* $\varepsilon^{N+1} = 0$.

The theorem will follow immediately from the Propositions 3.1 and 3.2 below. Choosing the approximating policy f^n as in (3.18) below will imply that

$$V_1(h_0, f^n) = L_1^n(h_0) \tag{3.8}$$

and, by virtue of Theorem 3.1, we then have

Corollary 3.1. *Given* n *with* $1 \leq n \leq N+1$,

$$V_1(h_0) - V_1(h_0, f^n) \leq U_1^n(h_0) - L_1^n(h_0) = \varepsilon^n.$$

The policy f^n will be constructed after the proof of Proposition 3.2. We first derive the upper bounds $U_n(h_{n-1})$. For this purpose let, for all $n = 1, \ldots, N$,

$$u_n(h_{n-1}, s) := \max_{f \in F(h_{n-1})} E_f\{R_n(h_N) \mid h_{n-1}, s\} \tag{3.9}$$

and define

$$U_n(h_{n-1}) := \sum_{i=1}^{M} u_n(h_{n-1}, s_i) p(s_i \mid h_{n-1}) \tag{3.10}$$

letting $U_{N+1}(h_N) = 0$.

From their definitions we immediately have

$$V_n(h_{n-1}) \leq U_n(h_{n-1}). \tag{3.11}$$

In fact,

$$\begin{aligned} V_n(h_{n-1}) &= \max_{f \in F(h_{n-1})} \sum_{i=1}^{M} E_f\{R_n(h_N) \mid h_{n-1}, s_i\} \cdot p\{s_i \mid h_{n-1}\} \\ &\leq \sum_{i=1}^{M} p(s_i \mid h_{n-1}) \max_{f \in F(h_{n-1})} E_f\{R_n(h_N) \mid h_{n-1}, s_i\} \\ &= U_n(h_{n-1}). \end{aligned} \tag{3.12}$$

Notice that the idea of conditioning on the underlying state s to obtain upper bounds has close analogies to the method of 'perceptive dynamic programming' introduced by Platzman [11] in the context of partially observable Markov decision processes. This idea which cannot be deduced from the framework in [19] is also implicit in the approximations proposed in [16, Section 6] in the context of infinite horizon Bayesian Markov control models. Notice also that the difference $U_n(h_{n-1}) -$

$V_n(h_{n-1})$ may be interpreted as the expected value of perfect information relative to the state s of the system.

For n given $(1 \leq n \leq N+1)$, define for $m \leq n$ the functions $U_m^n(h_{m-1})$ recursively as

$$\begin{aligned} &U_n^n(h_{n-1}) := U_n(h_{n-1}), \\ &U_m^n(h_{m-1}) := \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + U_{m+1}^n(h_m) \,|\, h_{m-1}\} \end{aligned} \tag{3.13}$$

setting $U_{N+1}(h_N) = 0$. Notice that $U_n^{N+1}(h_{n-1}) = V_n(h_{n-1})$. Furthermore we have

Lemma 3.1. *Given* $a \in A_n(h_{n-1})$,

$$E_a\{r_n(h_n) + u_{n+1}(h_n, s) \,|\, h_{n-1}, s\} \leq u_n(h_{n-1}, s).$$

Proof. Let $\bar{f} \in F(h_n) \subset F(h_{n-1})$ be such that

$$E_{\bar{f}}\{R_{n+1}(h_N) \,|\, h_n, s\} = u_{n+1}(h_n, s).$$

Notice that, if $h_n = (h_{n-1}, \bar{a}, x_n)$, then $\bar{f}_n(h_{n-1}) = \bar{a}$; therefore

$$\begin{aligned} &E_{\bar{a}}\{r_n(h_n) + u_{n+1}(h_n, s) \,|\, h_{n-1}, s\} \\ &\quad = E_{\bar{a}}\{r_n(h_n) + E_{\bar{f}}\{R_{n+1}(h_N) \,|\, h_n, s\} \,|\, h_{n-1}, s\} \\ &\quad = E_{\bar{a}}\{E_{\bar{f}}\{r_n(h_n) + R_{n+1}(h_N) \,|\, h_n, s\} \,|\, h_{n-1}, s\} \\ &\quad = E_{\bar{f}}\{E_{\bar{f}}\{R_n(h_N) \,|\, h_n, s\} \,|\, h_{n-1}, s\} = E_{\bar{f}}\{R_n(h_N) \,|\, h_{n-1}, s\} \\ &\quad \leq \max_{f \in F(h_{n-1})} E_f\{R_n(h_N) \,|\, h_{n-1}, s\} = u_n(h_{n-1}, s). \end{aligned}$$

Since $\bar{a}$ is any element in $A_n(h_{n-1})$, the result then follows. □

Proposition 3.1. *For all* m *and* n *with* $1 \leq m \leq n \leq N$ *and all* $h_{m-1} \in H_{m-1}$

$$\begin{aligned} V_m(h_{m-1}) = U_m^{N+1}(h_{m-1}) &\leq \cdots \leq U_m^{n+1}(h_{m-1}) \\ &\leq U_m^n(h_{m-1}) \leq \cdots \leq U_m(h_{m-1}). \end{aligned}$$

Proof. By induction on $p = n - m$. By Lemma 3.1, $\forall a \in A_m(h_{m-1})$, $E_a\{r_m(h_m) + U_{m+1}(h_m) \,|\, h_{m-1}\} \leq U_m(h_{m-1})$ therefore

$$U_m^{m+1}(h_{m-1}) = \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + U_{m+1}(h_m) \,|\, h_{m-1}\} \leq U_m(h_{m-1}).$$

Thus the assertion is true for $p = 1$. Assume it is true for $p \geq 1$; then

$$\begin{aligned} U_m^{n+1}(h_{m-1}) &= \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + U_{m+1}^{n+1}(h_m) \,|\, h_{m-1}\} \\ &\leq \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + U_{m+1}^n(h_m) \,|\, h_{m-1}\} = U_m^n(h_{m-1}) \end{aligned}$$

from which the result follows. □

Remark 3.1. With analogous proofs it can be shown that the results of Lemma 3.1 and Proposition 3.1 also hold if instead of (3.9) we let

$$u_n(h_{n-1}, s) := \max_{f \in F(h_{n-1})} \max_{h_N \in H_N(h_{n-1}, f, s)} R_n(h_N). \tag{3.14}$$

In this way the bounds will be less sharp, but computation is further reduced. □

Next we turn to deriving the lower bounds $L_n(h_{n-1})$. To this effect choose, for n $(1 \leq n \leq N)$ and $h_{n-1} \in H_{n-1}$ given, any policy $f \in F(h_{n-1})$ and let

$$L_n(h_{n-1}) := E_f\{R_n(h_N) \mid h_{n-1}\}. \tag{3.15}$$

The choice of such f could be based on simple heuristic arguments, or restricted to those f, if any, for which (3.15) is easy to compute. It immediately follows that

$$V_n(h_{n-1}) \geq L_n(h_{n-1}). \tag{3.16}$$

For n given $(1 \leq n \leq N+1)$, define for $m \leq n$ the functions $L_m^n(h_{m-1})$ recursively as

$$\begin{cases} L_n^n(h_{n-1}) := L_n(h_{n-1}), \\ L_m^n(h_{m-1}) := \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + L_{m+1}^n(h_m) \mid h_{m-1}\}, \end{cases} \tag{3.17}$$

setting $L_{N+1}(h_N) = 0$. Notice that $L_n^{N+1}(h_{n-1}) = V_n(h_{n-1})$. Furthermore we have

Proposition 3.2. *For all m and n with $1 \leq m \leq n \leq N$ and all $h_{m-1} \in H_{m-1}$,*

$$V_m(h_{m-1}) = L_m^{N+1}(h_{m-1}) \geq \cdots \geq L_m^{n+1}(h_{m-1}) \geq L_m^n(h_{m-1}) \geq \cdots \geq L_m(h_{m-1}).$$

Proof. We proceed by induction on $p = n - m$. For $p = 0$, letting $a_m = f_m(h_{m-1})$,

$$\begin{aligned} L_m(h_{m-1}) &= E_{a_m}\{r_m(h_m) + L_{m+1}^{m+1}(h_m) \mid h_{m-1}\} \\ &\leq \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + L_{m+1}^{m+1}(h_m) \mid h_{m-1}\} = L_m^{m+1}(h_{m-1}). \end{aligned}$$

Assume now the assertion is true for $p \geq 0$; then

$$\begin{aligned} L_m^n(h_{m-1}) &= \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + L_{m+1}^n(h_m) \mid h_{m-1}\} \\ &\leq \max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + L_{m+1}^{n+1}(h_m) \mid h_{m-1}\} = L_m^{n+1}(h_{m-1}) \end{aligned}$$

from which the result follows. □

Given n $(1 \leq n \leq N)$, we finally construct the approximating policy f^n for which $V_1(h_0, f^n) = L_1^n(h_0)$. It is easily seen that such policy is given by

$$f_m^n(h_{m-1}) = \begin{cases} f_m(h_{m-1}) & \text{for } m \geq n, \\ \arg\max_{a \in A_m(h_{m-1})} E_a\{r_m(h_m) + L_{m+1}^n(h_m) \mid h_{m-1}\} & \text{for } m < n. \end{cases} \tag{3.18}$$

4. Additional results

A convenient way to apply in practice the approximation approach described in the previous Section is the following: Fix an $\varepsilon > 0$, the accepted level for the approximation error, and determine n^* $(1 \leq n^* \leq N+1)$ such that $U_1^{n^*} - L_1^{n^*} < \varepsilon$ (by Theorem 3.1 such n^* always exists). So far, with only the results of the previous Section, we would either have to guess n^* or to compute all U_1^n, L_1^n starting from $n = 1$ until the condition is met.

The purpose of this Section is to provide, under the explicit assumption of a discount factor β, additional results, that will allow us to more conveniently determine such n^*. The main result of this Section is the following Theorem 4.1, whose proof is an immediate consequence of Proposition 4.1 below.

Theorem 4.1. *For* $1 \leq m < n \leq N+1$,

$$(U_1^m - L_1^m) - \sum_{i=m}^{n-1} \beta^{i-1} E_i \leq U_1^n - L_1^n \leq (U_1^m - L_1^m) - \sum_{i=m}^{n-1} \beta^{i-1} e_i$$

where $(1 \leq i \leq N)$

$$E_i := \sup_{H_{i-1}} (U_i^i - U_i^{i+1}) + \sup_{H_{i-1}} (L_i^{i+1} - L_i^i), \tag{4.1}$$

$$e_i := \inf_{H_{i-1}} (U_i^i - U_i^{i+1}) + \inf_{H_{i-1}} (L_i^{i+1} - L_i^i). \tag{4.2}$$

The practical application of Theorem 4.1 for the purpose of this Section is given by the following Algorithm 4.1, that by Theorem 3.1 certainly terminates.

Algorithm 4.1

Step 1: Fix $m \geq 1$.

Step 2: Compute $U_1^m - L_1^m$.
Terminate with $n^* = m$ if

$$U_1^m - L_1^m < \varepsilon;$$

otherwise.

Step 3: Let $n = m+1$.

Step 4: Compute E_{n-1}, e_{n-1}.
Terminate with $n^* = n$ if

$$U_1^m - L_1^m - \sum_{i=m}^{n-1} \beta^{i-1} e_i < \varepsilon;$$

otherwise.

Step 5: Let $m = n$ and return to Step 2 if

$$U_1^m - L_1^m - \sum_{i=m}^{n-1} \beta^{i-1} E_i < \varepsilon;$$

otherwise increase n by 1 and return to Step 4.

The rest of this Section is devoted to the proof of Theorem 4.1 and to an additional remark. Given the assumption on the discount factor, we may rewrite $R_n(h_N)$ in (2.17) as

$$R_n(h_N) = \sum_{i=n}^{N} \beta^{i-n} r_i(h_i^N) \tag{4.3}$$

and the dynamic programming algorithm in (3.4) becomes

$$V_n(h_{n-1}) = \max_{a \in A_n(h_{n-1})} E_a\{r_n(h_n) + \beta V_{n+1}(h_n) \mid h_{n-1}\} \tag{4.4}$$

with the recursions for U_m^n in (3.13) and for L_m^n in (3.17) modified accordingly.

Letting M_n denote the set of all real functions defined on H_n, the dynamic programming algorithm leads to an operator D_n $(n = 1, \ldots, N)$

$$D_n : M_n \to M_{n-1}$$

defined as $(v_{n+1} \in M_n)$

$$(D_n v_{n+1})(h_{n-1}) := \max_{a \in A_n(h_{n-1})} E_a\{r_n(h_n) + \beta v_{n+1}(h_n) \mid h_{n-1}\}. \tag{4.5}$$

Also let

$$D_n^i : M_n \to M_{n-i} \quad (1 \leq i \leq n)$$

be given by

$$(D_n^i v_{n-i})(h_{n-i}) := (D_{n-1}^{i-1} D_n v_{n+1})(h_{n-i}) \tag{4.6}$$

with $D_n^1 = D_n$.

It then follows immediately (see e.g. [6]) that for all $n = 1, \ldots, N$, all i $(1 \leq i \leq n)$ and all v_n, $w_n \in M_n$:

$$v_n \leq w_n \Rightarrow D_n^i v_n \leq D_n^i w_n, \tag{4.7}$$

$$D_n^i(v_n + c) = D_n^i v_n + \beta^i c, \tag{4.8}$$

$$\beta^i \inf_{H_n}(v_n - w_n) \leq D_n^i v_n - D_n^i w_n \leq \beta^i \sup_{H_n}(v_n - w_n) \tag{4.9}$$

where $c \in \mathbb{R}$ is a given constant.

Lemma 4.1. *For all* i $(1 \leq i \leq N)$,

(i) $\beta^{i-1} \inf_{H_{i-1}}(U_i^i - U_i^{i+1}) \leq U_1^i - U_1^{i+1} \leq \beta^{i-1} \sup_{H_{i-1}}(U_i^i - U_i^{i+1})$,

(ii) $$\beta^{i-1}\inf_{H_{i-1}}(L_i^{i+1}-L_i^i)\leq L_1^{i+1}-L_1^i\leq\beta^{i-1}\sup_{H_{i-1}}(L_i^{i+1}-L_i^i).$$

Proof. (i) $U_1^i=D_{i-1}^{i-1}U_i^i=D_{i-1}^{i-1}U_i^{i+1}+(D_{i-1}^{i-1}U_i^i-D_{i-1}^{i-1}U_i^{i+1})$ from which, by (4.9),

$$U_1^i\begin{cases}\leq U_1^{i+1}+\beta^{i-1}\sup_{H_{i-1}}(U_i^i-U_i^{i+1}),\\ \geq U_1^{i+1}+\beta^{i-1}\inf_{H_{i-1}}(U_i^i-U_i^{i+1}),\end{cases}$$

(ii) $$L_1^{i+1}=D_{i-1}^{i-1}L_i^{i+1}=D_{i-1}^{i-1}L_i^i+(D_{i-1}^{i-1}L_i^{i+1}-D_{i-1}^{i-1}L_i^i)$$

from which, by (4.9),

$$L_1^{i+1}\begin{cases}\leq L_1^i+\beta^{i-1}\sup_{H_{i-1}}(L_i^{i+1}-L_i^i),\\ \geq L_1^i+\beta^{i-1}\inf_{H_{i-1}}(L_i^{i+1}-L_i^i).\end{cases}\qquad\square$$

From Lemma 4.1 we immediately have

Proposition 4.1. *For* $1\leq m<n\leq N+1$,

(i) $$U_1^m-\sum_{i=m}^{n-1}\beta^{i-1}\sup_{H_{i-1}}(U_i^i-U_i^{i+1})\leq U_1^n\leq U_1^m-\sum_{i=m}^{n-1}\beta^{i-1}\inf_{H_{i-1}}(U_i^i-U_i^{i+1}),$$

(ii) $$L_1^m+\sum_{i=m}^{n-1}\beta^{i-1}\inf_{H_{i-1}}(L_i^{i+1}-L_i^i)\leq L_1^n\leq L_1^m+\sum_{i=m}^{n-1}\beta^{i-1}\sup_{H_{i-1}}(L_i^{i+1}-L_i^i).\qquad\square$$

Remark 4.1. Consider the operator T_i $(1\leq i\leq N)$,

$$T_i: M_{i-1}\to M_i$$

defined by $(v_i\in M_i)$

$$(T_iv_{i-1})(h_i)=v_{i-1}(h_{i-1}^i).\tag{4.10}$$

Then it can be easily seen that Theorem 4.1 also holds with E_i and e_i defined as

$$E_i:=\sup_{H_{i-1}}(U_i^i-D_iT_iU_i^i)+\sup_{H_{i-1}}(D_iT_iL_i^i-L_i^i)$$
$$+\beta\left[\sup_{H_{i-1}}(T_iU_i^i-U_{i+1}^{i+1})\sup_{H_{i-1}}(L_{i+1}^{i+1}-T_iL_i^i)\right],\tag{4.11}$$

$$e_i:=\inf_{H_{i-1}}(U_i^i-D_iT_iU_i^i)+\inf_{H_{i-1}}(D_iT_iL_i^i-L_i^i)$$
$$+\beta\left[\inf_{H_{i-1}}(T_iU_i^i-U_{i+1}^{i+1})+\inf_{H_{i-1}}(L_{i+1}^{i+1}-T_iL_i^i)\right].\tag{4.12}$$

Depending on the particular application one may run Algorithm 4.1 with E_i and e_i defined either by (4.1) and (4.2) or by (4.11) and (4.12). $\square$

5. An application to oil exploration

5.1. Oil discovery process model

A petroleum *prospect* is a geologic anomaly that is conceived of as containing hydrocarbons and forms a target for drilling. A petroleum *play* is a collection of prospects within a geographic region, all of which share certain common geologic attributes. A prospect is said to be a field if it contains hydrocarbons, a *dry hole* otherwise.

It is never a priori certain whether or not a prospect is a field or a dry hole and there is uncertainty also concerning the amount of petroleum (*magnitude*) contained in a field. Exploratory drilling is used to discover the fields among the prospects as well as their magnitudes.

The *discovery process model* is assumed to be essentially that presented in [10] except for minor modifications. More precisely, we make the following assumptions:

Assumption I. Multinomial magnitude distribution. The magnitudes of the fields, as deposed by Nature, are realizations of mutually independent, identically distributed multinomial random variables (taking on k possible values).

Assumption II. Sampling without replacement and proportional to random magnitude. That is, if there are N_i fields of magnitude B_i ($i = 1, 2, \ldots, k$), the probability that the magnitude of the first discovered field is B_j is given by: $N_jB_j/(\sum_{i=1}^{k} N_iB_i)$.

Assumption III. There are N prospects in the play among which N_0 are known to be dry holes. Drilling successes and failures occur via hypergeometric sampling of these N prospects.

For a justification of assumption I see [15]. Assumption II has been suggested and extensively tested by Kaufman et al. [3, 9].

5.2. The oil discovery process as a stochastic dynamic optimization problem

The oil discovery process may be formulated as a dynamic stochastic optimization problem. In fact, since drilling is costly and the outcomes, and therefore also the rewards, from any drilling effort are uncertain, it makes sense to distribute the drilling effort over time and to determine sequentially the optimal number of wells to be drilled in each of a finite number of periods so as to maximize the total expected reward.

We now formulate such problem as a problem within the class defined in Section 2 and give a numerical example for which in the next subsection computational results are reported and compared with those from other approximation approaches.

States. A state is a $(k+1)$-tuple $(N_0, N_1, \ldots, N_k)$ with the following meaning: N_0 is the number of dry holes and $N_1, \ldots, N_k$ are the numbers of fields of magnitude $B_1, \ldots, B_k$ respectively as deposed by Nature. In the numerical example for subsec-

tion 5.3 we assume: $k = 3$, $N_0 = 10$, $N_0 + N_1 + N_2 + N_3 = 30$, therefore the total number of possible states is $M = 231$.

State probabilities p_i^0. The a priori distribution p_i^0 is assumed to be multinomial (see Assumption I in subsection 5.1).

Actions. The action a_n to be taken in period n simply means the number of wells to be drilled in that period. In the numerical example, if n_0, n_1, n_2, n_3 are the numbers of dry holes and of fields of magnitude B_1, B_2, B_3 respectively already discovered prior to period n, then a_n must be less than or equal to $30-(n_0+n_1+n_2+n_3)$. We further assume $A = \{a \mid 0 \leq a \leq 3\}$.

Observations. An observation $x_n = (x_n^0, x_n^1, x_n^2, x_n^3)$ is a 4-tuple, where x_n^0 is the number of dry holes and x_n^1, x_n^2, x_n^3 are the numbers of fields of magnitude B_1, B_2, B_3 respectively discovered in period n as a consequence of action a_n. Obviously, given a_n, we only need to consider those x_n for which $x_n^0 + x_n^1 + x_n^2 + x_n^3 = a_n$. For example, if $a_n \leq 3$, only 35 such histories need to be considered.

Histories. A history h_n is the collection of all actions and observations up to period n. Notice that the cardinality of H_n is exponentially growing with n. In the numerical example $|H_n| \simeq (35)^n$.

For the oil discovery problem considered here, it was possible to find a 'sufficient statistic' so that the number of different histories that had to be taken into account was reduced to: $|H_1| = 35$, $|H_2| = 210$, $|H_3| = 715$, $|H_4| = 1820$, $|H_5| = 3876$ etc. For details see [1].

Horizon. The most natural choice for N is to take it equal to $N_0 + N_1 + N_2 + N_3$, so that in the numerical example $N = 30$.

Reward. Letting $h_n = (h_{n-1}, a, x)$, the reward function can be chosen as $r_n(h_n) = r(a, x)$, and, for the numerical example, we explicitly assume a discount factor $\beta = .8$.

Probabilities. Given Assumption II, the computation of probabilities such as $p_f(h_n \mid h_m, s)$ and $p_f(h_n \mid h_m)$ is by no means easy both in terms of memory and computing time requirements. For computational details we refer to [1]; the basic idea is presented in [2].

Upper and lower bounds. In the numerical example, $u_n(h_{n-1}, s)$ is computed according to (3.14) and, for the computation of the lower bound $L_n(h_{n-1})$, the policy f is chosen so that $f_i(h_{i-1}) = 0$ for all $i = n, \ldots, N$.

5.3. Numerical results and comparisons

The bounds derived in Section 3, have been computed for an instance of the problem just described and are shown in Table 5.1 for different truncation periods of the horizon of action.

These bounds have been compared with bounds derived from Whitt [19] and Hinderer [7]. Whitt's lower and upper bounds depend on the particular choice of functions e and g outside the designated subset. Since he does not give any constructive procedure for finding them, we chose $e = 0$ which appears to be most appropriate for the given problem and as function g we used an improvement to

Table 5.1
Bounds to the optimal value $V_1 = 28.69$

n (truncation period)	L_1^n lower bound	U_1^n upper bound	Hinderer upper bound	Whitt upper bound
2	8.27	49.72	84.66	78.98
3	14.54	43.65	71.53	67.83
4	19.25	39.17	60.45	58.83
5	22.69	35.91	50.83	51.63
6	25.15	33.59	42.20	45.93
7	26.81	31.98	36.17	41.53
8	27.84	30.88	33.02	38.16
9	28.39	30.14	31.16	35.55
10	28.63	29.64	29.78	33.59
11	28.68	29.30	28.97	32.10
12	28.69	29.05	28.73	30.95
13	28.69	28.86	28.70	30.11
14	28.69	28.73	28.69	29.47
15	28.69	28.70	28.69	29.00
16	28.69	28.69	28.69	28.75
17	28.69	28.69	28.69	28.70
18	28.69	28.69	28.69	28.69

the particular g given in the Remark of p. 182 in [19]. Hinderer's bounds (see Theorem 3.1 p. 296 in [7]) also depend on the choice of a function $\hat{V}_0$; in our problem, a natural choice is $\hat{V}_0 = 0$.

With this choice of e and $\hat{V}_0$, the lower bounds of Whitt and Hinderer both coincide with ours. Therefore in Table 5.1 only Whitt's and Hinderer's upper bounds are reported. In the example our upper bounds are uniformly better than Whitt's, whereas they are numerically worse than Hinderer's for $n = 11, \ldots, 15$; however, to compute Hinderer's bounds, when truncating at period n, one has to take into account all possible histories belonging to H_N, not only those belonging to H_n, which makes their computation extremely heavy, if not impossible.

References

[1] G. Andreatta, "Computational problems arising in PPMWOR sampling", LADSEB-CNR internal report 10/83 (Padova, Italy, 1983).

[2] G. Andreatta and G.M. Kaufman, "Estimation of finite population properties when sampling is without replacement and proportional to magnitude", Working paper EL 82-027WP, Energy Laboratory, MIT (Cambridge, MA, 1983).

[3] E. Barouch and G.M. Kaufman, "Estimation of undiscovered oil and gas", in: P. Dax, ed., Mathematical aspects of production and distribution of energy, Proceedings of the symposia in applied mathematics, volume XXI, (American Mathematical Society, Providence, R.I., 1977) pp. 77-91.

[4] J. Birge and R.J.-B. Wets, "Designing approximation schemes for stochastic programs with recourse", IIASA working paper WP-83-111 (Laxenburg, Austria, 1983).

[5] K. Hinderer, "Foundation of non-stationary dynamic programming with discrete-time parameter", *Lecture Notes in Operations Research and Mathematical Systems* 33 (Springer, Berlin, 1970).
[6] K. Hinderer, "Estimates for finite-stage dynamic programs", *Journal of Mathematical Analysis and Applications* 55 (1976) 207-238.
[7] K. Hinderer, "On approximate solutions of finite-stage dynamic programs", in: M. Puterman, ed., *Dynamic programming and its applications* (Academic Press, London, 1979) pp. 289-317.
[8] P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds", *Matematische Operationsforschung und Statistik Series Optimization* 13 (1982) 431-447.
[9] G.M. Kaufman and J.-W. Wang, "Model mis-specification and the Princeton study of volume and area of oil fields and their impact on the order of discovery", Working paper EL-80-003WP, Energy Laboratory, MIT (Cambridge, MA, 1980).
[10] G.M. Kaufman, W. Runggaldier and Z. Livne, "Predicting the time-rate of supply from a petroleum play", in: J. B. Ramsey, ed., *The economics of exploration for energy resources* (JAI Press, Greenwich, CT, 1981) pp. 69-102.
[11] L.K. Platzman, "Finite-memory estimation and control of finite probabilistic systems", Ph.D. thesis, MIT, also MIT Report ESL-R-723 (Cambridge, MA, 1977).
[12] U. Rieder, "Bayesian dynamic programming", *Advances in Applied Probability* 7 (1975) 330-348.
[13] W.J. Runggaldier and F. Spizzichino, "Approximations and bounds for a generalized optimal stopping problem", *Zeitschrift für Operations Research* 26 (1982) 143-155.
[14] R.F. Serfozo, "Monotone optimal policies for Markov decision processes", *Mathematical Programming Study* 6 (1976) 202-215.
[15] J.L. Smith and G.L. Ward, "Maximum likelihood estimates of the size distributions of North Sea oil deposits", Working Paper EL-80-027WP, Energy Laboratory, MIT (Cambridge, MA, 1980).
[16] K.M. van Hee, *Bayesian control of Markov chains*, Mathematical Centre Tracts 95 (Amsterdam, 1978).
[17] K.H. Waldman, "On bounds for dynamic programs", *Mathematics of Operations Research* (to appear).
[18] W. Whitt, "Approximations of dynamic programs-I", *Mathematics of Operations Research* 3 (1978) 231-243.
[19] W. Whitt, "Approximations of dynamic programs-II", *Mathematics of Operations Resarch* 4 (1979) 179-185.

Mathematical Programming Study 27 (1986) 133-144
North-Holland

STABILITY IN STOCHASTIC PROGRAMMING WITH RECOURSE. CONTAMINATED DISTRIBUTIONS

J. DUPAČOVÁ
Charles University, Prague, Czechoslovakia

Received 1 December 1983
Revised manuscript received 27 July 1984

In the paper, stability of the optimal solution of a stochastic program with recourse with respect to small changes of the underlying distribution of random coefficients is considered. As a tool, contamination of the given distribution by another one is suggested and the original stability problem is thus reduced to that with linearly perturbed objective function. The theory of perturbed Kuhn-Tucker points and strongly regular equations is used to get explicit formulas for Gâteaux differentials of optimal solutions under different assumptions. Possible exploitation of the results for further robustness studies is indicated.

Key words: Stochastic Programming, Incomplete Information, Stability, Robustness, Contamination, Simple Recourse Problem.

Consider the following stochastic programming problem:

> Maximize $f(x; F) = E_F\{c(x) - \varphi(x; z)\}$ on the set $\mathscr{X}$ where $\mathscr{X} \subset \mathbb{R}^n$ is a nonempty closed convex set of admissible solutions, $c: \mathscr{X} \to \mathbb{R}^1$ is a given function, F is a given joint probability distribution of a random vector z on $(\mathscr{Z}, \mathscr{B}_Z)$, $\mathscr{Z} \subset \mathbb{R}^l$, $\varphi: \mathscr{X} \times \mathscr{Z} \to \mathbb{R}^1$ is a given nonnegative function such that $\varphi(x; z)$ are measurable for all $x \in \mathscr{X}$. (1)

An example of (1) is when a nonlinear program

$$\begin{aligned} &\text{maximize} && c(x) \\ &\text{subject to} && g_k(x; z) \geq 0,\ 1 \leq k \leq m,\ x \in \mathscr{X}, \end{aligned}$$

contains random parameters in $g_k(x; z)$, $1 \leq k \leq m$, and the decision $x \in \mathscr{X}$ has to be chosen before the values of these random parameters are observed. The function $\varphi(x; z)$ evaluates the loss corresponding to the case that the chosen $x \in \mathscr{X}$ does not fulfil the constraints $g_k(x; z) \geq 0$, $1 \leq k \leq m$ for the observed values of the random parameters.

The essential results concerning the objective function in (1) are summarized in the following lemma (see e.g. [10]):

Lemma. *Let $\varphi: \mathscr{X} \times \mathscr{Z} \to \mathbb{R}^1$ be Lipschitz continuous on the set $\mathscr{X}$ for an arbitrary $z \in \mathscr{Z}$ and let the Lipschitzian constant $k_1(z)$ be integrable with respect to F. Let the gradient*

$\nabla_x \varphi(x; z)$ exist for $x \in \mathcal{X}$ almost surely with respect to F and let $E_F\varphi(x; z)$ be finite at least for one point $x \in \mathcal{X}$. Then $E_F\varphi(x; z)$ is Lipschitz continuous on $\mathcal{X}$, the gradient $\nabla_x E_F\varphi(x; z)$ exists for $x \in \mathcal{X}$ and

$$\nabla_x E_F\varphi(x; z) = E_F \nabla_x \varphi(x; z). \tag{2}$$

Remark 1. Under assumptions of the Lemma, the existence of the expectation

$$f(x; F) = E_F\{c(x) - \varphi(x; z)\}$$

is evidently guaranteed for all $x \in \mathcal{X}$. Under the additional assumption that $\varphi(\cdot; z)$ is convex on $\mathcal{X}$ for all $z \in \mathcal{Z}$ and c is concave, then the function $f(x; F)$ is concave, differentiable on $\mathcal{X}$.

As in stochastic linear programming, the optimal solution $x(F)$ of (1) (provided that it exists) depends on the assumed distribution F. In many real-life situations, however, the assumption of a completely known distribution F is hardly acceptable and the solution of (1) should be thus at least supplemented by a proper stability study with respect to F. In the robust case, a small change in the distribution F should cause only a small change of the optimal solution. In the preceding papers [4, 5], the first attempts were made to study stability of the optimal solution $x(F)$ of (1) with respect to the distribution F and its parameters through completing the approaches developed for nonlinear programming stability studies by suitable statistical methods. In this paper, local behaviour of $x(F)$ will be studied via t-contamination of F by a distribution G belonging to a specified set of distributions (see [3, 4] for special cases), i.e., instead of F, distributions of the form

$$F_t = (1-t)F + tG, \quad 0 \leq t \leq 1, \tag{3}$$

will be considered. In (3), F_t is called *distribution F t-contaminated by distribution G* and for our purpose, the Gâteaux differential

$$\mathrm{d}x(F; G-F) = \lim_{t \to 0^+} \frac{x(F + t(G-F)) - x(F)}{t}$$

of the optimal solution $x(F)$ at F in the direction of $G - F$ is of importance.

Disregarding the constraints (i.e., taking $\mathcal{X} = \mathbb{R}^n$), the optimal solution $x(F_t)$ of the program

$$\text{maximize } f(x; F_t) = E_{F_t}\{c(x) - \varphi(x; z)\} \tag{4}$$

should fulfil the system of n equations

$$\Psi(x; F_t) = 0$$

where (for F, G fixed) $\Psi : \mathbb{R}^n \times \langle 0, 1) \to \mathbb{R}^n$ and its components

$$\psi_j(x; F_t) = \frac{\partial}{\partial x_j} f(x; F_t), \quad 1 \leq j \leq n,$$

are assumed to exist for all j. Obviously

$$\Psi(x; F_t) = \Psi(x; F) + t[\Psi(x; G) - \Psi(x; F)], \quad 0 \leq t \leq 1.$$

Using the implicit function theorem, the Gâteaux differential $\mathrm{d}x(F; G-F)$ can be computed under suitable differentiability and regularity assumptions; taking into account that $\psi(x(F); F) = 0$, we get

$$\mathrm{d}x(F; G-F) = -D^{-1}\Psi(x(F); G)$$

where

$$D = \left(\frac{\partial \psi_j(x(F); F)}{\partial x_k}\right) = \left(\frac{\partial^2 f(x(F); F)}{\partial x_j\, \partial x_k}\right), \quad 1 \leq j, k \leq n.$$

To obtain the Gâteaux differential $\mathrm{d}x(F; G-F)$ of the optimal solution of (1), we shall use the theory of perturbed Kuhn-Tucker points and strongly regular equations developed in [12], [13]. In principle, it is possible to get Gâteaux differentials of optimal solutions for probabilistic constrained programs using similar tools.

The knowledge of the Gâteaux differential of $x(F)$ at F in the direction of $G-F$ is useful not only for the first order approximation of the optimal solutions corresponding to distributions belonging to a neighbourhood of F but also for deeper statistical conclusions on robustness, namely, in connection with statistical properties of the estimate $x(F_\nu)$ of $x(F)$, which is based on the empirical distribution F_ν. For the special choices $G = \delta_u$ (degenerated distributions concentrated at one point u), the Gâteaux differential $\mathrm{d}x(F; \delta_u - F)$ corresponds to the influence curve $\Omega_F(u)$ widely used in asymptotic statistics. Different characteristics of $\Omega_F(y)$ suggested in [9] measure the effect of contamination of the data by gross errors, the local effect of rounding or grouping of the observations, etc.

We shall concentrate upon obtaining formulas for the Gâteaux differentials under different assumptions leaving the detailed investigation of the statistical aspects to a forthcoming paper. We shall start with the general constrained case with

$$\mathscr{X} = \{x \in \mathbb{R}^n: g_i(x) \geq 0,\ 1 \leq i \leq m,\ h_r(x) = 0,\ 1 \leq r \leq p\};$$

the Lagrange function and the Kuhn-Tucker points will be denoted by

$$L(w; F) = f(x; F) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{r=1}^{p} v_r h_r(x),$$

$$w(F) = [x(F), u(F), v(F)] \text{ and } I(F) = \{i: g_i(x(F)) = 0\}.$$

Theorem 1. *For the program*

$$\text{maximize} \quad f(x; F) := E_F\{c(x) - \varphi(x; z)\} \quad \text{on the set } \mathscr{X} \tag{5}$$

assume

(i) $$\mathscr{X} = \{x \in \mathbb{R}^n: g_i(x) \geq 0,\ 1 \leq i \leq m,\ h_r(x) = 0,\ 1 \leq r \leq p\} \neq \emptyset, \tag{6}$$

g_i, $1 \leq i \leq m$, h_r, $1 \leq r \leq p$, *are twice continuously differentiable,*

(ii) *$c: \mathcal{X} \to \mathbb{R}^1$ is twice continuously differentiable.*

(iii) *The distribution F on $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$ and the function $\varphi: \mathcal{X} \times \mathcal{Z} \to \mathbb{R}^1$ fulfil the assumptions of the Lemma and the mean value $E_F\varphi(x; z)$ is twice continuously differentiable.*

(iv) *For the program (5) with $\mathcal{X}$ given by (6), Kuhn-Tucker conditions of the first and second order, the linear independence condition and the strict complementarity conditions are fulfilled for $w(F) = [x(F), u(F), v(F)] \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p$, and the matrix*

$$C = \nabla^2_{xx} L(w(F); F)$$

is nonsingular.

(v) *There is a neighbourhood $\mathcal{O}(x(F)) \subset \mathbb{R}^n$ on which φ and the distribution G on $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$ fulfil the assumptions of the Lemma and $E_G\varphi(x; z)$ is twice continuously differentiable on $\mathcal{O}(x(F))$.*

Then: (a) *There is a neighbourhood $\mathcal{O}(w(F)) \subset \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p$, a real number $t_0 > 0$ and a continuous function $w: \langle 0, t_0 \rangle \to \mathcal{O}(w(F))$, $w(0) = w(F)$ such that for any $t \in \langle 0, t_0)$, $w(t) = [x(t), u(t), v(t)]$ is the Kuhn-Tucker point of*

$$\max_{x \in \mathcal{X}} f(x; F_t) := E_{F_t}\{c(x) - \varphi(x; z)\} \tag{7}$$

for which the second order sufficient condition, the linear independence condition and the strict complementarity conditions are fulfilled.

(b) *The Gâteaux differential $\mathrm{d}x(F; G - F)$ of the isolated local maximizer $x(F)$ of (5), (6) in the direction of $G - F$ is given by*

$$\mathrm{d}x(F; G - F) = -D^{-1}\nabla_x L(w(F); G), \tag{8}$$

where

$$\begin{aligned} D^{-1} &= [I - C^{-1}P(P^TC^{-1}P)^{-1}P^T]C^{-1}, \\ P &= [\nabla_x g_i(x(F)), i \in I(F), \nabla_x h_r(x(F)), 1 \leq r \leq p] \end{aligned} \tag{9}$$

and I is the n-dimensional unit matrix.

Proof. The first assertion of Theorem 1 can be proved by means of the implicit function theorem as in [12, Theorem 2.1]. To prove the second assertion, we shall use the implicit function theorem once more. (See also [7] for a similar approach.)

For the sake of simplicity assume that $I(F) = \{1, \ldots, s\}$, denote by $\bar{u} \in \mathbb{R}^s$ the projection of u into R^s and define

$$\Psi: \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^p \times \langle 0, 1 \rangle \to \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^p,$$

a vector valued function whose components are

$$\psi_j(x, \bar{u}, v; t) = \frac{\partial f(x; F_t)}{\partial x_j} + \sum_{i=1}^{s} u_i \frac{\partial g_i(x)}{\partial x_j} + \sum_{r=1}^{p} v_r \frac{\partial h_r(x)}{\partial x_j}, \quad 1 \leq j \leq n,$$

$$\psi_{n+i}(x, \bar{u}, v; t) = g_i(x), \quad 1 \leq i \leq s,$$

$$\psi_{n+s+1}(x, \bar{u}, v; t) = h_r(x), \quad 1 \leq r \leq p.$$

Under our specification of $I(F)$, the system $\Psi(x, \bar{u}, v; t) = 0$ together with $u_i = 0$, $s+1 \leq i \leq m$, forms the local Kuhn-Tucker conditions of the first order for problem (7) with $t < t_0$ and according to (i), (ii), (iii), (v), there exists a neighbourhood $\mathcal{O}(x(F))$ on which $\Psi(x, \bar{u}, v; t)$ is continuously differentiable with respect to all variables. The matrix

$$D(x, \bar{u}, v; t) = \left(\frac{\partial \Psi(x, \bar{u}, v; t)}{\partial x_j},\ 1 \leq j \leq n, \quad \frac{\partial \Psi(x, \bar{u}, v; t)}{\partial u_i},\ 1 \leq i \leq s, \quad \frac{\partial \Psi(x, \bar{u}, v, t)}{\partial v_r},\ 1 \leq r \leq p \right)$$

$$= \begin{pmatrix} \nabla^2_{xx} L(w; F_t) & P \\ P^T & 0 \end{pmatrix}$$

with $w = (x, \bar{u}, 0, v)$ and $0 \in \mathbb{R}^{m-s}$ is nonsingular for $w = w(F)$ according to (iv). This implies the existence of continuous right-hand derivatives of $\bar{w}(t) = [x(t), \bar{u}(t), v(t)]$ at 0:

$$\frac{\mathrm{d}\bar{w}(0^+)}{\mathrm{d}t} = -D(\bar{w}(0); 0)^{-1} \frac{\partial}{\partial t} \Psi(\bar{w}(0); 0) \tag{10}$$

where

$$\frac{\partial}{\partial t} \Psi(\bar{w}(0); 0)$$

$$= \begin{pmatrix} \nabla_x f(x(F); G) - \nabla_x f(x(F); F) \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \nabla_x f(x(F); G) + \sum_{i=1}^{s} u_i(F) \nabla_x g_i(x(F)) + \sum_{r=1}^{p} v_r(F) \nabla_x h_r(x(F)) \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \nabla_x L(w(F); G) \\ 0 \end{pmatrix}. \tag{11}$$

Formula (8), (9) follows from (10), (11) by inversion of the block matrix $D(\bar{w}(0); 0) = \left(\begin{smallmatrix} C & P \\ P^T & 0 \end{smallmatrix}\right)$; the first equality in (11) follows from $\nabla_x L(w(F); F) = 0$.

Remark. Due to the fact that (7) is a special type of a linearly perturbed nonlinear program, the Gâteaux differentials of $x(F)$ at F both in the direction of $G-F$ and in the direction of G are equal: $\mathrm{d}x(F; G-F) = \mathrm{d}x(F; G)$.

For $\mathcal{X}$ polyhedral we get, as a special case of Theorem 1,

Theorem 2. *Let in the problem*

$$\max_{x \in \mathcal{X}} f(x; F) = E_F\{c(x) - \varphi(x; z)\} \tag{12}$$

the following assumptions be fulfilled:

(i) $\mathscr{X} = \{x \in \mathbb{R}^n: Px = p, x \geq 0\} \neq \emptyset$, $P(r, n)$, $p \in \mathbb{R}^r$ *are a given matrix of rank* r *and a given vector; let the vertices of* $\mathscr{X}$ *be nondegenerate.*

(ii) $c: \mathscr{X} \to \mathbb{R}^1$ *is twice continuously differentiable.*

(iii) *The distribution* F *on* $(\mathscr{Z}, \mathscr{B}_{\mathscr{Z}})$ *and the function* $\varphi: \mathscr{X} \times \mathscr{Z} \to \mathbb{R}^1$ *fulfil the assumptions of the Lemma and the mean value* $E_F \varphi(x; z)$ *is twice continuously differentiable.*

(iv) *There exists a Kuhn–Tucker point* $[x(F); \pi(F)]$ *for* (12) *such that the second order sufficient condition and the strict complementarity conditions are fulfilled. For* $J = \{j: x_j(F) > 0\}$, *the matrix*

$$C_J = \left(\frac{\partial^2 f(x(F); F)}{\partial x_j \, \partial x_k} \right), \quad j, k \in J, \tag{13}$$

is nonsingular.

(v) *There is a neighbourhood* $\mathscr{O}(x(F)) \subset \mathbb{R}^n$ *on which the function* φ *and the distribution* G *on* $(\mathscr{Z}, \mathscr{B}_{\mathscr{Z}})$ *fulfil assumptions of the Lemma and* $f(x; G)$ *is twice continuously differentiable on* $\mathscr{O}(x(F))$.

Then (a) *There are neighbourhoods* $\mathscr{O}_1(x(F)) \subset \mathscr{O}(x(F))$, $\mathscr{V}(\pi(F)) \subset \mathbb{R}^r$, *a real number* $t_0 > 0$ *and continuous functions*

$$x: \langle 0, t_0) \to \mathscr{O}_1(x(F)), \quad x(0) = x(F),$$

$$\pi: \langle 0, t_0) \to \mathscr{V}(\pi(F)), \quad \pi(0) = \pi(F),$$

such that for any $t \in \langle 0, t_0)$, $[x(t); \pi(t)]$ *is a Kuhn–Tucker point for the problem*

$$\max_{x \in \mathscr{X}} f(x; F_t) = E_{F_t}\{c(x) - \varphi(x; z)\} \tag{14}$$

with $F_t = (1 - t)F + tG$, $0 \leq t \leq 1$. *The second order sufficient condition and the strict complementarity conditions are fulfilled for* $[x(t); \pi(t)]$ *and*

$$x_j(F_t) = x_j(t) = 0, \quad j \notin J,$$

$$x_j(F_t) = x_j(t) > 0, \quad j \in J.$$

(b) *The vector* $\mathrm{d}x_J(F; G - F)$ *of the components of the Gâteaux differential of the isolated local maximizer* $x(F)$ *of* (12) *in the direction of* $G - F$ *for* $j \in J$ *is given by*

$$\mathrm{d}x_J(F; G - F) = -D_J^{-1} \Psi_J(x(G); \pi(F); G) \tag{15}$$

where

$$\Psi_J(x(F); \pi(F); G) = \left(\frac{\partial}{\partial x_j} f(x(F); G) + \sum_{k=1}^{r} p_{kj} \pi_k(F) \right)_{j \in J},$$

$$D_J^{-1} = [I_J - C_J^{-1} P_J^T (P_J C_J^{-1} P_J^T)^{-1} P_J] C_J^{-1},$$

$$P_J = (p_{kj})_{\substack{1 \leq k \leq r \\ j \in J}},$$

I_J *is a unit matrix of dimension* $s = \mathrm{card}\, J$ *and* C_J *is given by* (13).

The remaining components of the Gâteaux differential $\mathrm{d}x(F; G-F)$ *are equal to zero.*

In the special case of a simple recourse problem with random right-hand sides and with $\mathscr{X} = \mathbb{R}^n_+$, i.e., for

$$\max_{x \geq 0} E_F\left\{ c^{\mathrm{T}}x - \sum_{i=1}^{m} q_i\left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right)^+ \right\}, \tag{16}$$

we have the following theorem:

Theorem 3 [4]. *Assume*:

(i) *F is an m-dimensional continuous distribution of b for which* $E_F b$ *exists.*

(ii) *The optimal solution* $x(F)$ *of* (16) *exists and the strict complementarity conditions hold true. Let* $J = \{j : x_j(F) > 0\}$.

(iii) $q_i > 0$, $1 \leq i \leq m$, $A_J = (a_{ij})$, $1 \leq i \leq m$, $j \in J$ *has full column rank.*

(iv) *The marginal densities* f_i, $1 \leq i \leq m$, *are continuous and positive at the points* $X_i(F) = \sum_{j \in J} a_{ij}x_j(F)$, $1 \leq i \leq m$, *respectively.*

(v) *G is an m-dimensional distribution whose marginal distribution functions* G_i *have continuous derivatives in a neighbourhood of the points* $X_i(F) = \sum_{j \in J} a_{ij}x_j(F)$, $1 \leq i \leq m$, *respectively.*

Then (a) *There is a neighbourhood* $\mathcal{O}(x(F))$ *and a real number* $t_0 > 0$ *such that the program*

$$\max_{x \geq 0} E_{F_t}\left\{ c^{T}x - \sum_{i=1}^{m} q_i\left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right)^+ \right\} \tag{17}$$

with $F_t = (1-t)F + tG$ *has a unique optimal solution* $x(F_t) \in \mathcal{O}(x(F))$ *for any* $0 \leq t < t_0$, $x_j(F_t)$, $j \in J$ *are nonzero components of* $x(F_t)$ *and* $x_j(F_t) = 0$ *for* $j \notin J$.

(b) *Components of the Gâteaux differential of the optimal solution* $x(F)$ *at F in the direction of* $G-F$ *corresponding to the nonzero components of* $x(F)$ *are given by*

$$\mathrm{d}x_J(F; G-F) = (A_J^T K A_J)^{-1}(c_J - A_J^T k) \tag{18}$$

where $c_J = (c_j)_{j \in J}$, $k = (k_i)$, $1 \leq i \leq m$ *with*

$$k_i = q_i G_i\left(\sum_{h \in J} a_{ih}x_h(F) \right), \quad 1 \leq i \leq m,$$

and

$$K = \mathrm{diag}\left\{ q_i f_i\left(\sum_{h \in J} a_{ih}x_h(F) \right), 1 \leq i \leq m \right\}.$$

Theorem 3 illustrates, inter alia that the assumption of twice continuous differentiability of $f(x; F)$ can be fulfilled in practice. For detailed discussion of this question see [15].

Example. Let us compute the influence curve corresponding to the case considered in Theorem 3. Having solved the program (16) for the chosen distribution F, the set J, reduced matrices A_J, c_J and the diagonal matrix K are known. Let $u \neq Ax(F)$ be a chosen point and $G = \delta_u$. For the vector k we have $k = q(u; F)$, where

$$q_i(u; F) = q_i \quad \text{if } u_i < X_i(F),$$

$$= 0 \quad \text{otherwise.}$$

The influence curve $\Omega_F(u)$ is given by

$$\Omega_F(u) = (A_J^T K A_J)^{-1}(c_J - A_J^T q(u; F))$$

and to get its characteristics, e.g., the gross-error-sensitivity

$$\gamma^* = \sup_u \|\Omega_F(u)\|$$

means to solve a discrete optimization problem

$$\text{maximize} \sum_k \left[\sum_j \alpha_{kj} \left(c_j - \sum_i \delta_i q_i a_{ij} \right) \right]^2 \tag{19}$$

with zero-one variables δ_i, $1 \leq i \leq m$. (In (19), α_{kj}'s denote the corresponding elements of $(A_J^T K A_J)^{-1}$.)

The assumptions of strict complementarity play an essential role in the proofs of Theorems 1, 2 and 3. They guarantee that the interval $\langle 0, t_0)$ on which $w(t)$ (resp. $[x(t), \pi(t)]$) is the Kuhn-Tucker point of (7) (resp. of (14)) is nonempty. Alternatively, the strict complementarity conditions can be replaced by the strong second order sufficient condition [13] which gives the existence of continuous Kuhn-Tucker points on a nonempty interval $\langle 0, t_0)$. This approach was applied in imbedding methods [8] and it will be used to get parallel results in our case.

Without assuming the strict complementarity conditions in (5), (6) denote

$$I^+(F) = \{i: g_i(x(F)) = 0 \text{ and } u_i(F) > 0\},$$

$$I^0(F) = \{i: g_i(x(F)) = 0 \text{ and } u_i(F) = 0\}$$

and formulate *the strong second order sufficient condition* [13]:

For each $y \neq 0$ with

$$y^T \nabla_x g_i(x(F)) = 0, \quad i \in I^+(F), \qquad y^T \nabla_x h_r(x(F)) = 0, \quad 1 \leq r \leq p,$$

the inequality $y^T \nabla_{xx}^2 L(x(F), u(F), v(F)) y < 0$ holds true.

Theorem 4. *Let assumptions* (i)-(iii), (v) *of Theorem* 1 *be fulfilled and the assumption* (iv) *be replaced by*

(iv') *For the program* (5) *with* $\mathcal{X}$ *given by* (6), *the linear independence condition and the strong second order sufficient condition are fulfilled for the Kuhn-Tucker point*

$$w(F) = [x(F), u(F), v(F)].$$

Then: (a) *There is a neighbourhood* $\mathcal{O}(w(F)) \subset \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p$, *a real number* $t_0 > 0$ *and a continuous function*

$$w: \langle 0, t_0) \to \mathcal{O}(w(F)), \quad w(0) = w(F),$$

such that for any $t \in \langle 0, t_0)$, $w(t) = [x(t), u(t), v(t)]$ *is the Kuhn-Tucker point of*

$$\max_{x \in \mathcal{X}} f(x; F_t) := E_{F_t}\{c(x) - \varphi(x; z)\}.$$

(b) *The Gâteaux differential* $\mathrm{d}x(F; G-F)$ *is the unique solution of the quadratic program*

$$\begin{aligned} &\textit{maximize} && \tfrac{1}{2} x^T \nabla^2_{xx} L(w(F); F)x + x^T \nabla_x L(w(F); G) \\ &\textit{subject to} && x^T \nabla_x g_i(x(F)) = 0, \quad i \in I^+(F), \\ &&& x^T \nabla_x g_i(x(F)) \geq 0, \quad i \in I^0(F), \\ &&& x^T \nabla_x h_r(x(F)) = 0, \quad 1 \leq r \leq p, \end{aligned} \tag{20}$$

and $\mathrm{d}u(F; G-F)$, $\mathrm{d}v(F; G-F)$ *are the unique Lagrange multipliers for* (20) *related with* $\mathrm{d}x(F; G-F)$ *with zero components* $\mathrm{d}u_i(F; G-F)$ *for* $i \notin I(F)$.

Proof. The first part follows from [13, Theorem 2.1] and the second one is a variant of [8, Theorem 5].

Assuming strict complementarity condition valid for the optimal solution $\mathrm{d}x(F; G-F)$ of (20), one can get the new active set $I(F_t)$ for (7) with t small enough:

$$I(F_t) = I^+(F) \cup \{i \in I^0(F): \mathrm{d}x(F; G-F)^T \nabla_x g_i(x(F)) = 0\}$$

[8, Corollary 1]. For the special case of Theorem 2 we have

$$I^+(F) = \left\{j: x_j(F) = 0 \text{ and } \frac{\partial}{\partial x_j} f(x(F); F) + \sum_k p_{kj} \pi_k(F) < 0\right\},$$

$$I^0(F) = \left\{j: x_j(F) = 0 \text{ and } \frac{\partial}{\partial x_j} f(x(F); F) + \sum_k p_{kj} \pi_k(F) = 0\right\},$$

so that $J = \{1, \ldots, n\} - [I^+(F) \cup I^0(F)]$. The program (20) has the form

$$\begin{aligned} &\text{maximize} && \tfrac{1}{2} x^T \nabla^2_{xx} f(x(F); F)x + x^T \nabla_x f(x(F); G) \\ &\text{subject to} && x_j = 0, \quad j \in I^+(F), \qquad x_j \geq 0, \quad j \in I^0(F), \end{aligned}$$

and for $t > 0$ small enough, the new set $J(F_t) = \{j: x_j(F_t) > 0\}$ fulfils

$$J \subset J(F_t) \subset J \cup I^0(F).$$

In the simple case where $I^0(F) = \{j_0\}$, we have explicitly

Theorem 5. *Let assumptions* (i)-(iii), (v) *of Theorem* 2 *be fulfilled. Assume further the existence of a Kuhn-Tucker point* $[x(F), \pi(F)]$ *for* (12) *such that the strong second order sufficient condition is fulfilled and* $I^0(F) = \{j_0\}$.

Then: (a) *There exist a neighbourhood* $\mathcal{O}_1(x(F))$ *and a real number* $t_0 > 0$ *such that, for* $0 \leq t < t_0$,

$$x_j(t) = 0, \quad j \notin J^0 = J \cup I^0(F)$$

and $x(t)$ *is the isolated local maximizer of one of the following problems*:

$$\max\{f(x; F_t): x \in \mathcal{X}_J\} \quad \text{or} \quad \max\{f(x; F_t): x \in \mathcal{X}_{J^0}\}$$

where for $H \subset \{1, \ldots, n\}$,

$$\mathcal{X}_H = \{x \in \mathbb{R}^n: Px = p, x_j = 0, j \notin H\}.$$

(b) *Correspondingly, the components of the Gâteaux differential are*

$$\begin{pmatrix} \mathrm{d}x_H(F; G-F) \\ \mathrm{d}\pi(F; G-F) \end{pmatrix} = -\begin{pmatrix} C_H & P_H^T \\ P_H & 0 \end{pmatrix}^{-1} \begin{pmatrix} \Psi_H(x(F), \pi(F); G) \\ 0 \end{pmatrix}$$

where for $H = J$ *or* J^0

$$C_H = \left(\frac{\partial^2 f(x(F); F)}{\partial x_i \, \partial x_j}\right)_{i,j \in H}, \qquad P_H = (p_{kj})_{\substack{1 \leq k \leq r, \\ j \in H}}$$

$$\Psi_H(x(F), \pi(F); G) = \left(\frac{\partial f(x(F); G)}{\partial x_j} + \sum_{k=1}^{r} p_{kj}\pi_k(F)\right)_{j \in H}$$

The remaining components of $\mathrm{d}x(F; G-F)$ *for* $j \notin H$ *equal zero.*

By specifying the set $\mathcal{G}$ of distributions G under consideration, the effect of t-contamination of F by distributions belonging to $\mathcal{G}$ on the optimal solution $x(F)$ can be studied. As a rule, $F \in \mathcal{G}$. Typical examples are

1. F uniform distribution of the random vector z on a closed interval $I \subset \mathbb{R}^l$ and $\mathcal{G}$ the set of distributions such that

$$E_G z = E_F z \text{ and } P_G(z \in I) = 1 \qquad \forall G \in \mathcal{G}. \tag{21}$$

2. The marginal distributions F_i are normal $N(\mu_i, \sigma_i^2)$ and $\mathcal{G}$ is the set of distributions of the random vector z or $\mathbb{R}^l$ such that

$$E_G z_i = \mu_i, \quad \operatorname{var}_G z_i = \sigma_i^2, \quad 1 \leq i \leq l, \qquad \forall G \in \mathcal{G}. \tag{22}$$

In this context, the extremal distributions belonging to $\mathcal{G}$ are of main interest. For the derivative of the objective function in (7) or (14)

$$\frac{\partial}{\partial t} f(x; F_t) = f(x; G) - f(x; F)$$

we have, for all $G \in \mathcal{G}$,

$$\inf_{G \in \mathcal{G}} f(x; G) - f(x; F) \leq \frac{\partial}{\partial t} f(x; F_t) \leq \sup_{G \in \mathcal{G}} f(x; G) - f(x; F).$$

Let G^*, G^{**} be such that

$$\inf_{G\in\mathcal{G}} f(x; G) = f(x; G^*), \qquad \sup_{G\in\mathcal{G}} f(x; G) = f(x; G^{**}).$$

The local changes of $x(F)$ in the direction of G^*-F or $G^{**}-F$ give the extremal local decrease or increase of the optimal value of the objective function $f(x; F)$. The corresponding problem, for $G = G^*$,

$$\max_{x\in\mathcal{X}} f(x; (1-t)F + tG^*) = \max_{x\in\mathcal{X}} [(1-t)f(x; F) + t \inf_{G\in\mathcal{G}} f(x; G)]$$

can be evidently related to the Hodges-Lehman decision rule [14] or to the Nadeau-Theodorescu restricted Bayes strategies [11]. The existence of the extremal distributions G^*, G^{**} has been proved for wide classes of recourse problems and for various sets $\mathcal{G}$ of distributions, e.g., for the sets $\mathcal{G}$ given by (21) and (22). For details see [1, 2, 6].

Acknowledgement

I would like to thank an anonymous referee for suggestions which helped to improve upon an earlier version of the paper.

References

[1] J. Dupačová, "Minimaxová úloha stochastického lineárního programování a momentový problém", Ekonomicko-matematický obzor 13 (1977) 279-307. (Extended abstract: "Minimax approach to stochastic linear programming and the moment problem. Selected results", *Zeitschrift für Angewandte Mathematik und Mechanik* 58 (1978) T466-T467.)

[2] J. Dupačová, "On minimax decision rule in stochastic linear programming", in: A. Prékopa, ed., *Mathematical methods of operations research I* (Akademiai Kiadó, Budapest, 1980) pp. 38-48.

[3] J. Dupačová, "Stability studies in stochastic programs with recourse. The special case", *Zeitschrift für Angewandte Mathematik und Mechanik* 62 (1982) T369-T370.

[4] J. Dupačová, "Stability in stochastic programming with recourse", *Acta Universitatis Carolinae-Mathematica et Physica* 24 (1983) 23-34.

[5] J. Dupačová, "Stability in stochastic programming with recourse-Estimated parameters", *Mathematical Programming* 28 (1984) 72-83.

[6] J. Dupačová, "The minimax approach to stochastic programming and an illustrative application", to appear in *Stochastics*.

[7] A.V. Fiacco, "Sensitivity analysis for nonlinear programming using penalty methods", *Mathematical Programming* 10 (1976) 287-311.

[8] H. Gfrerer, J. Guddat and Hj. Wacker, "A globally convergent algorithm based on imbedding and parametric optimization", *Computing* 30 (1983) 225-252.

[9] F.R. Hampel, "The influence curve and its role in robust estimation", *Journal of the American Statistical Association* 69 (1974) 383-397.

[10] P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds", *Mathematische Operationsforschung und Statistik, Ser. Optimization* 13 (1982) 431-447.

[11] R. Nadeau and R. Theodorescu, "Restricted Bayes strategies for programs with simple recourse", *Operations Research* 28 (1980) 777-784.

[12] S.M. Robinson, "Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms", *Mathematical Programming* 7 (1984) 1-16.
[13] S.M. Robinson, "Strongly regular generalized equations", *Mathematics of Operations Research* 5 (1980) 43-62.
[14] H. Schneeweiss, *Entscheidungskriterien bei Risiko* (Springer-Verlag, Berlin, 1967).
[15] J. Wang, "Distribution sensitivity analysis for stochastic programs with recourse", to appear in *Mathematical Programming*.

Mathematical Programming Study 27 (1986) 145-152
North-Holland

LIPSCHITZ CONTINUITY OF OBJECTIVE FUNCTIONS IN STOCHASTIC PROGRAMS WITH FIXED RECOURSE AND ITS APPLICATIONS

Jinde WANG
Department of Mathematics, University of Nanjing, People's Republic of China

Received 10 Feburary 1984
Revised manuscript received 20 January 1985

This paper deals with Lipschitz continuity of the objective functions of two-stage stochastic program with fixed recourse w.r.t. the first stage variable x and the random vector ξ jointly. This is then used to study stability of the considered problem. Some results, especially the Lipschitz continuity of the infimal functional in ξ, are stronger than early ones.

Key words: Stochastic Programming, Stability, Infimal Function, Marginal Value, Lipschitz Continuity.

1. Introduction

This paper deals with two-stage stochastic programs of the following type:

$$\text{Min} \quad c'x + \int_{\Omega} Q(x, q(\omega), T(\omega), b(\omega))P(d\omega) = c'x + Q(x, q, T, b)$$

$$\text{s.t.} \quad x \in S \tag{1}$$

where $Q(x, q(\omega), T(\omega), b(\omega))$ is the optimal value of the second stage problem

$$\text{Min} \quad q(\omega)y$$

$$\text{s.t.} \quad Wy = b(\omega) - T(\omega)x, \ y \geq 0, \tag{2}$$

$x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$; b is an m-dimensional vector; c, q, W, T are vectors or matrices with consistent dimensions, W is non-stochastic while q, T, b are random and defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$; $q(\omega)$, $T(\omega)$, $b(\omega)$ are their sample values respectively.

Clearly the objective function of (1) is related to the involved random vector $\xi = (q, T, b)$. To indicate this relation we write the objective function as $Z(x, \xi)$. Now suppose we have a family of problems, corresponding to a family of random vectors in L_2^r, where L_2^r is the space of square integrable random vectors, taking values in R^r and $r = n_2 + mn_1 + m$. Then $Z(x, \xi)$ is a mapping from $\mathbb{R}^{n_1} \times L_2^r$ to $\mathbb{R}^1$. The product space $R^{n_1} \times L_2^r$ can be endowed with a distance d, defined by

$$d^2(a_1, a_2) = |x_1 - x_2|^2 + |\xi_1 - \xi_2|^2_{L_2^r},$$

where $a_1 = (x_1, \xi_1)$, $a_2 = (x_2, \xi_2)$ are two points of $\mathbb{R}^{n_1} \times L_2^r$ and $|\cdot|_{L_2^r}$ denotes the norm in L_2^r. In the next section we will show Lipschitz continuity of $Z(x, \xi)$ over the space $\mathbb{R}^{n_1} \times L_2^r$. This property turns out to be useful for studying stability problem for programs (1). In sections 3 and 4 we use it to investigate epi-convergence of $Z(x, \xi^{(n)})$ to $Z(x, \xi^{(0)})$, as $\xi^{(n)} L_2$-converge to $\xi^{(0)}$. Local Lipschitz continuity of the infima functional $\inf_{x \in S} Z(x, \xi)$ is established there.

As mentioned earlier we assume that q, T, b are square integrable thorughout this paper. This is not a special restriction for our discussion. In fact this assumption is also imposed for $Z(x, \xi)$ to be finite. Thus ξ and q, T, b are elements of L_2^r, $L_2^{n_2}$, $L_2^{mn_1}$, L_2^m respectively. In the context the superscripts of these L's will be omitted. The reader can identify to which space they belong.

2. Lipschitz continuity of $Z(x, \xi)$

First we introduce the following definitions:

Definition 1. A functional f, defined on L_2, is said to be locally Lipschitz continuous at $\xi^{(0)}$ if there exist positive numbers δ and h such that

$$|f(\xi) - f(\xi^{(0)})| \leq h \cdot |\xi - \xi^{(0)}|_{L_2}$$

holds for all ξ, satisfying $|\xi - \xi^{(0)}|_{L_2} < \delta$.

Definition 2. A mapping $Z(x, \xi)$, defined on $\mathbb{R}^{n_1} \times L_2$, is said to be locally Lipschitz continuous with respect to x and ξ jointly at point $(x^{(0)}, \xi^{(0)})$ if there exist positive numbers ξ and L such that

$$|Z(x, \xi) - Z(x^{(0)}, \xi^{(0)})| \leq L \cdot d((x, \xi), (x^{(0)}, \xi^{(0)}))$$

for all (x, ξ) in the ε-neighborhood of $(x^{(0)}, \xi^{(0)})$.

We assume that W is a complete recourse matrix. Then $Q(x, q(\omega), T(\omega), b(\omega))$ is either finite or $-\infty$. Since the latter case is meaningless in practice, we assume $Q(x, q(\omega), T(\omega), b(\omega))$ is finite on $\mathbb{R}^{n_1} \times \Omega$. Thus by square integrability of q, T, b, $Z(x, \xi)$ is finite on $\mathbb{R}^{n_1} \times L_2$ and the explicit formula of $Z(x, \xi)$ is

$$Z(x, \xi) = Z(x, q, T, b) = c'x + \sum_i \int_{U_i} q_i(\omega) W_i^{-1}(b(\omega) - T(\omega)x) P(\mathrm{d}\omega) \qquad (3)$$

where W_i is one of the optimal bases $\{W_i, i \in I\}$ of W, $q_i(\omega)$ is a sub-vector of $q(\omega)$, whose components correspond to the columns of W_i, and U_i, $i \in I$, are the decision regions of problem (2), i.e.,

$$U_i = V_i \Big\backslash \bigcup_1^{i-1} V_j, \quad V_j = \{\omega: W_j^{-1}(b(\omega) - T(\omega)x) \geq 0,$$

$$q(\omega) - q_j(\omega) W_j^{-1} W \geq 0\}.$$

As $c'x$ is a linear function of x, the Lipschitz continuity of $Z(x, \xi)$ is equivalent to that of the integral in (3). Let (x, ξ) and $(x^{(0)}, \xi^{(0)})$ be two points in $\mathbb{R}^{n_1} \times L_2$. First consider the following difference

$$\begin{aligned}
&Q(x, q(\omega), T(\omega), b(\omega)) - Q(x^{(0)}, q^{(0)}(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) \\
&\quad = Q(x, q(\omega), T(\omega), b(\omega)) - Q(x, q(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) \\
&\qquad + Q(x, q(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) - Q(x^{(0)}, q(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) \\
&\quad = + Q(x^{(0)}, q(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) - Q(x^{(0)}, q^{(0)}(\omega), T^{(0)}(\omega), b^{(0)}(\omega)) \\
&\quad = D_1 + D_2 + D_3.
\end{aligned}$$

In [6] it was porved that $Q(x, q(\omega), T(\omega), b(\omega))$ is convex in x, $T(\omega)$, $b(\omega)$ respectively and concave in $q(\omega)$. Using inequalities for convex (and concave) functions we can get the following estimates:

$$\begin{aligned}
|D_1| &= |q_i(\omega) W_i^{-1}(b(\omega) - T(\omega)x) - q_i(\omega) W_i^{-1}(b^{(0)}(\omega) - T^{(0)}(\omega)x)| \\
&\leq \max_{i \in I} |q_i(\omega) W_i^{-1}| \cdot \{|b(\omega) - b^{(0)}(\omega)| + |T(\omega) - T^{(0)}(\omega)| \cdot |x|\}, \qquad (4)
\end{aligned}$$

$$|D_2| \leq \max_{i \in I} |q_i(\omega) W_i^{-1}| \cdot |T^{(0)}(\omega)| \cdot |x - x^{(0)}|, \qquad (5)$$

$$|D_3| \leq \max_{i \in I} |W_i^{-1}(b^{(0)}(\omega) - T^{(0)}(\omega)x^{(0)})| \cdot |q(\omega) - q^{(0)}(\omega)|. \qquad (6)$$

Combining inequalities (4), (5), (6) and integrating w.r.t. ω, we obtain

$$\begin{aligned}
&|Q(x, q, T, b) - Q(x^{(0)}, q^{(0)}, T^{(0)}, b^{(0)})| \\
&\quad \leq h_1' |b - b^{(0)}|_{L_2} + h_2' |T - T^{(0)}|_{L_2} + h_3' |x - x^{(0)}| + h_4' |q - q^{(0)}|_{L_2} \qquad (7)
\end{aligned}$$

where

$$h_1' = \left\{ \int_\Omega \max_{i \in I} |W_i^{-1} q_i(\omega)|^2 P(\mathrm{d}\omega) \right\}^{1/2}, \quad h_2' = h_1' \cdot |x|, \quad h_3' = h_1' |T^{(0)}|_{L_2},$$

$$h_4' = \left\{ \int_\Omega \max_{i \in I} |W_i^{-1}(b^{(0)}(\omega) - T^{(0)}(\omega)x^{(0)}|^2 P(\mathrm{d}\omega) \right\}^{1/2}.$$

From these formulas it can be seen that h_4' is a constant related to $b^{(0)}$, $T^{(0)}$, $x^{(0)}$ and that h_1', h_2', h_3' depend on x, q, T and b.

Let $\bar{h}_1$ be defined by

$$\bar{h}_1 = \max_{i \in I} |W_i^{-1}| \cdot |q|_{L_2}.$$

Obviously $h_1' \leq \bar{h}_1 \leq \max_{i \in I} |W_i^{-1}| \cdot |q^{(0)}|_{L_2} + \varepsilon_1$, provided $|q - q^{(0)}|_{L_2}$ is small enough, where ε_1 is a small positive number.

Define

$$h_1 = \max_{i \in I} |W_i^{-1}| \cdot |q^{(0)}|_{L_2} + \varepsilon_1, \qquad h_2 = h_1 \cdot |x^{(0)}|, \qquad h_3 = h_1 \cdot |T^{(0)}|_{L_2}.$$

Then $h_1' \leq h_1$, $h_2' \leq h_2$, $h_3' \leq h_3$, provided $|q - q^{(0)}|_{L_2}$, $|T - T^{(0)}|_{L_2}$, $|x - x^{(0)}|$ are small enough. With $h_4 = h_4'$ we can reduce (7) to

$$|Q(x, q, T, b) - Q(x^{(0)}, q^{(0)}, T^{(0)}, b^{(0)})|$$
$$\leq h_1 \cdot |b - b^{(0)}|_{L_2} + h_2|T - T^{(0)}|_{L_2} + h_3|x - x^{(0)}| + h_4|q - q^{(0)}|_{L_2}.$$

This inequality can yield Lipschitz continuity of $Q(x, q, T, b)$ w.r.t. the all arguments jointly at $(x^{(0)}, q^{(0)}, T^{(0)}, b^{(0)})$. This implies the local Lipschitz continuity of $Z(x, \xi)$ w.r.t. x and ξ jointly at $(x^{(0)}, \xi^{(0)})$. We summarize the discussion as

Proposition 1. *In two-stage stochastic programs with complete recourse of type* (1) *if all random vectors are square integrable, then the objective $Z(x, \xi)$ is locally Lipschitz continuous in x and ξ jointly.*

In the sequel we need the following

Corollary 2. *For any fixed x, $Z(x, \xi)$ is locally Lipschitz continuous in ξ and the Lipschitz constant can be continuous in x.*

Proof. With the same procedure as above we get

$$|Z(x, \xi) - Z(x, \xi')| \leq h_1|b - b'|_{L_2} + h_2|T - T'|_{L_2} + h_4|q - q'|_{L_2} \tag{8}$$

where h_1, h_2, h_4 are defined in the same way as above. Therefore h_1 is independent of x, h_2 and h_4 are continuous in x.

3. Applications

The Lipschitz continuity given in section 2 can be of use in studying stability and constructing approximation procedure. First we show that it will lead to epi-convergence of $Z(x, \xi^{(n)})$ to $Z(x, \xi^{(0)})$ as $\xi^{(n)}$ converge to $\xi^{(0)}$ in L_2-sense, which further yields stability of the optimal solutions of two-stage stochastic programs of type (1). The definition and main results of epi-convergence are cited below.

Definition 3 [1]. A sequence of lower semi-continuous functions $f^{(n)}(x)$ epi-converges to function f iff for all x the following conditions are satisfied:

(A) $\liminf f^{(n)}(x^{(n)}) \geq f(x)$ for all sequences $x^{(n)}$ converging to x;

(B) there exists a sequence $x^{(n)}$ converging to x such that

$$\limsup f^{(n)}(x^{(n)}) \leq f(x).$$

Theorem 3 [1]. *If a sequence of lower semi-continuous functions $f^{(n)}$ epi-converges to f, then*

$$\limsup \operatorname{Argmin} f^{(n)} \subset \operatorname{Argmin} f$$

and $\operatorname{liminf}(\inf f^{(n)}) = \inf f$, *where* $\operatorname{Argmin} f = \{x: \min f = f(x)\}$. *Moreover if* $\operatorname{Argmin} f \neq \emptyset$ *and* $\inf f$ *is finite, then*

$$\lim(\inf f^{(n)}) = \inf f,$$

if and only if $\bar{x} \in \operatorname{Argmin} f$ *implies that there exist sequences* $\varepsilon^{(n)} > 0$ *with* $\varepsilon^{(n)} \to 0$ *and* $x^{(n)}$ *converging to* $\bar{x}$ *such that for all* n

$$x^{(n)} \in \varepsilon^{(n)} - \operatorname{Argmin} f^{(n)} = \{x: f^{(n)}(x) \leq \inf f^{(n)} + \varepsilon^{(n)}\}.$$

For details of epi-convergence theory, see [1] and [7].

A direct consequence of Proposition 1 and Theorem 3 is the following

Proposition 4. *Suppose there is a sequence of two-stage problems of type* (1) *with random vectors* $\xi^{(0)}$ *and* $\xi^{(n)}$, $n = 1, 2, \ldots$, *such that they are square integrable and* $\xi^{(n)}$ L_2*-converge to* $\xi^{(0)}$. *Then* $Z(x, \xi^{(n)})$ *converges pointwise and in epi-sense to* $Z(x, \xi^{(0)})$.

Proof. By the virtue of joint Lipschitz continuity of $Z(x, \xi)$ in x and ξ it is very easy to verify conditions A and B in the definition of epi-convergence and conditions for pointwise convergence.

To get stability of the optimal solution of problem (1), we need

Lemma 5. *If the sequence of lower semi-continuous functions* $f^{(n)}$ *epi-converges to* f *and* S *is a closed set in a Euclidean space, then* $F^{(n)}$ *epi-converge to* F, *where* $F = f + \psi_S$, $F^{(n)} = f^{(n)} + \psi_S$ *and*

$$\psi_S(x) = \begin{cases} 0 & \text{if } x \in S, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Immediate.

Combining the results in Theorem 3, Proposition 4 and Lemma 5 we obtain

Proposition 6. *Suppose the assumptions in Proposition* 4 *hold true. Then*

$$\operatorname{limsup} \operatorname*{Argmin}_{x \in S} Z(x, \xi^{(n)}) \subset \operatorname*{Argmin}_{x \in S} Z(x, \xi^{(0)})$$

and

$$\operatorname{liminf}\left(\inf_{x \in S} Z(x, \xi^{(n)})\right) = \inf_{x \in S} Z(x, \xi^{(0)}).$$

If there is a sequence $\bar{x}^{(n)} \in \operatorname{Argmin}_{x \in S} Z(x, \xi^{(n)})$, *converging to* $\bar{x} \in$

$\operatorname{Argmin}_{x\in S} Z(x, \xi^{(0)})$, *then*

$$\left|\inf_{x\in S} Z(x, \xi^{(n)}) - \inf_{x\in S} Z(x, \xi^{(0)})\right| \leq M|\xi^{(n)} - \xi^{(0)}|_{L_2},$$

i.e. the infima functional is locally Lipschitz continuous in ξ.

Proof. The first two assertions are direct consequences of the last three propositions. It remains to show the Lipschitz property of the infima functional.

Observe that

$$|Z(\bar{x}^{(n)}, \xi^{(n)}) - Z(\bar{x}, \xi^{(0)})| = \begin{cases} Z(\bar{x}^{(n)}, \xi^{(n)}) - Z(\bar{x}, \xi^{(0)}) \\ Z(\bar{x}, \xi^{(0)}) - Z(\bar{x}^{(n)}, \xi^{(n)}) \end{cases}$$

$$\leq \begin{cases} Z(\bar{x}, \xi^{(n)}) - Z(\bar{x}, \xi^{(0)}) & \text{if } Z(\bar{x}^{(n)}, \xi^{(n)}) \geq Z(\bar{x}, \xi^{(0)}), \\ Z(\bar{x}^{(n)}, \xi^{(0)}) - Z(\bar{x}^{(n)}, \xi^{(n)}) & \text{if } Z(\bar{x}, \xi^{(0)}) \geq Z(\bar{x}^{(n)}, \xi^{(n)}). \end{cases}$$

From the joint Lipschitz continuity of $Z(x, \xi)$ it follows that

$$|Z(\bar{x}, \xi^{(n)}) - Z(\bar{x}, \xi^{(0)})| \leq L(\bar{x}) \cdot |\xi^{(n)} - \xi^{(0)}|_{L_2}$$

and

$$|Z(\bar{x}^{(n)}, \xi^{(0)}) - Z(\bar{x}^{(n)}, \xi^{(n)})| \leq L(\bar{x}^{(n)}) \cdot |\xi^{(n)} - \xi^{(0)}|_{L_2}.$$

By Corollary 2, $L(x)$ is continuous in x. Hence when n is large enough there exists a positive number M such that

$$L(x) \leq M \quad \text{and} \quad L(\bar{x}^{(n)}) \leq M$$

or, equivalently,

$$|Z(\bar{x}^{(n)}, \xi^{(n)}) - Z(\bar{x}, \xi^{(0)})| \leq M \cdot |\xi^{(n)} - \xi^{(0)}|_{L_2}.$$

This inequality gives the desired Lipschitz continuity.

The results presented here are stronger than some early ones in the following senses. In [2] and [4] P. Kall gave error bounds for the difference $Z(x, \xi^{(n)}) - Z(x, \xi)$. Here we establish the Lipschitz continuity of $Z(x, \xi)$ in x and ξ jointly. In [8] R. Wets proved that for two-stage problems of type (1) if $\xi^{(n)}$ and ξ are distributed on a compact set, then $Z(x, \xi^{(n)})$ epi-converges to $Z(x, \xi^{(0)})$ as $\xi^{(n)}$ converges to $\xi^{(0)}$ in distribution. Here in proposition 6 the compactness assumption is unnecessary (but we need L_2-convergence of $\xi^{(n)}$ to $\xi^{(0)}$). The more important one is that we give Lipschitz continuity of the infima functional.

To obtain Lipschitz continuity of the infima functional under some conditions one also may resort to more sophisticated tools, e.g. the results in [5] by R. Rockafellar. Nevertheless the approach used here is much cheaper in the sense that the conditions are easier to be verified and the proof procedure is much simpler.

4. Constructing L_2-convergent sequences

In literature and practice of solving two-stage stochastic programs discretization approximation is widely used. Here we show that the sequences in this kind of approximation scheme do converge in L_2-sense.

Proposition 7. *Any random vector ξ in L_2 can be approximated by a sequence of simple random vectors $\xi^{(n)}$ in L_2-sense.*

Proof. It is a ready result of probability theory. If ξ is a one dimensional random variable, then $\xi^{(n)}$ can be constructed as follows

$$\xi^{(n)}(\omega)=\begin{cases}\dfrac{i-1}{2^n} & \text{if } \omega\in U_i^{(n)}=\left\{\omega\colon \dfrac{i-1}{2^n}\leqslant \xi(\omega)<\dfrac{i}{2^n}\right\},\\ & i=-n2^n,\ldots,0,1,\ldots,n2^n,\\ n & \text{if } \omega\in \underline{U}^{(n)}=\{\omega\colon n\leqslant \xi(\omega)\},\\ -n & \text{if } \omega\in \bar{U}^{(n)}=\{\omega\colon -n>\xi(\omega)\}.\end{cases}$$

For a multi-dimensional random vector it is sufficient to construct such a sequence for each component.

Another often used discretization approximation is through using conditional expectation, see [3]. We show that one may construct a sequence of conditional expectations, which L_2-converges to a given random vector. Again it is sufficient to show that for a random variable.

For a given partition $\tilde{T}$ of the sample space Ω into several parts $\Omega=\bigcup_{i=1}^{N}U_i$, where U_i are measurable sets, we denote by $\tilde{G}$ the σ-algebra generated by U_i, $i=1,2,\ldots,N$. Then the conditional expectation $E(\xi/\tilde{G})$ of ξ given $\tilde{G}$ is defined. Suppose there is a family of partitions $\tilde{T}_n=\{U_i^{(n)}, i=1,2,\ldots,N_n\}$. Let $E(\xi/G_n)$ be the corresponding sequence of conditional expectations, where G_n is the σ-algebra generated by $\{U_i, i=1,2,\ldots,N_n\}$.

Proposition 8. *Suppose ξ is square integrable and T_n, $n=1,2,\ldots$, are partitions, consisting of sets $\{\underline{U}^{(n)}, \bar{U}^{(n)}, U_i^{(n)}, i=-n2^n,\ldots,n2^n\}$, which are defined in Proposition 7. Then $E(\xi/G_n)$ L_2-converges to ξ, where G_n is the σ-algebra generated by T_n.*

Proof. First we show that $E(\xi/G_n)$ almost surely converges to ξ. By martingale theory $E(\xi/G_n)$ almost surely converges to $E(\xi/G_\infty)$, where G_∞ is the smallest σ-algebra containing all G_n, $n=1,2,\ldots$. Therefore for the present purpose it is sufficient to prove $\xi=E(\xi/G_\infty)$, or equivalently to prove $G=G_\infty$, where G is the σ-algebra induced by ξ.

From the construction of G_n we know that any set U of the form $U=\xi^{-1}(-\infty,a)$ is either an element of some G_n or a union of (finite or countably infinite number of) elements of certain G_n. As G_∞ is a σ-algebra, this set U must belong to G_∞.

From this fact follows that the σ-algebra G, generated by the family of sets U corresponding to all intervals $(-\infty, a)$, is contained in G_∞. The inverse inclusion relation is trivial. Thus we have $G = G_\infty$ or, equivalently, $E(\xi/G_\infty) = E(\xi/G) = \xi$.

Next we show that $E(\xi/G_n)$ L_2-converges to ξ. Note that $\xi \in L_2$ implies $\xi \in L_1$, i.e., $\int_\Omega |\xi(\omega)| P(\mathrm{d}\omega) < \infty$. Hence

$$\int_\Omega |E(\xi/G_n)| P(\mathrm{d}\omega) = \sum_i \int_{U_i^{(n)}} |E(\xi/G_n)| P(\mathrm{d}\omega)$$

$$= \sum_i \left| \int_{U_i^{(n)}} \xi(\omega) P(\mathrm{d}\omega) \right| \leq \int_\Omega |\xi(\omega)| P(\mathrm{d}\omega).$$

Then by a.e. convergence of $E(\xi/G_n)$ to ξ and the dominated convergence theorem we arrive at the conclusion that $E(\xi/G_n)$ L_2-converges to ξ.

The two L_2-convergent sequences discussed above are two examples. From what we have done in this section one may expect many other sequences, L_2-converging to the given random vector.

References

[1] H. Attouch and R. Wets, "Approximation and convergence in nonlinear optimization", in: O. L. Mangasarian, R. Meyer and S. Robinson, eds., *Nonlinear Programming* 4 (Academic Press, New York, 1981) pp. 367-394.

[2] P. Kall, "Approximations to stochastic programs with complete fixed recourse", *Numerische Mathematik* 22 (1974) 333-339.

[3] P. Kall, "Computational method for solving two stage stochastic linear programming problems", *Zeitschrift für Angewandte Mathematik und Physik* 30 (1979) 261-271.

[4] P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds", *Mathematische Operationsforschung und Statistik, ser. Optimization* 13 (1982) 431-447.

[5] R.T. Rockafellar, "Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming", *Mathematical Programming Study* 17 (1982) 28-66.

[6] D. Walkup and R. Wets, "Stochastic program with recourse", *SIAM Applied Mathematics* 15 (1967) 1299-1314.

[7] R. Wets, "Convergence of convex functions, variational inequalities, and convex optimization problems", in: R. Cottle, F. Giannessi and J. Lions, eds. (Wiley, New York, 1979) pp. 375-403.

[8] R. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel, B. Korte, eds., *Mathematical programming, the state of the art* (Springer-Verlag, Berlin, 1983) pp. 566-603.

Mathematical Programming Study 27 (1986) 153-182
North-Holland

ROBUSTNESS AGAINST DEPENDENCE IN PERT: AN APPLICATION OF DUALITY AND DISTRIBUTIONS WITH KNOWN MARGINALS

Willem K. KLEIN HANEVELD

University of Groningen, Institute of Econometrics, P.O. Box 800, 9700 AV Groningen, The Netherlands

Received 5 February 1984
Revised manuscript received 24 October 1984

A PERT-type project planning problem is considered, under the assumption (to be relaxed in Section 4) that the marginal distributions of the durations of the activities are known. Instead of the assumption of independence a minimax approach is proposed. A complete characterization of worst-case joint distributions, which by definition maximize the mean delay of the project completion time over a fixed target time T, is given. In the same framework also an optimal value for T is determined: it balances the costs of delay with the costs for large values of T in a two-stage stochastic program.

The main tool of analysis is duality. Worst-case distributions can be described as the solutions of a generalized transportation problem. The complementary slackness conditions of this linear program and its dual characterize the worst-case distributions by means of a condition on their supports. Due to the special structure, the dual problem can be reduced to a finite-dimensional convex program. By dualizing the reduced dual again, a flow problem on the PERT-network is derived. Optimal flows appear to be the criticality numbers of the worst-case distributions. In Section 2 special attention is paid to the characterization of the so-called NW Rule Solution for a generalized transportation problem.

Key words: Stochastic Programming, Duality, Transportation Problem, Minimax Approach, Distributions with Known Marginals, Dependent Random Variables, Worst-Case Distributions, Robustness, Stochastic Networks, Project Planning, PERT.

1. Introduction and summary

Suppose that a multivariate distribution is unknown, except for the fact that all one-dimensional marginals have been given. In general, there are many joint distributions compatible with the given marginals, one of them being defined by independence. In a worst-case analysis the following *distribution problem* is appropriate:

Find the joint distribution which maximizes the expected value of a given criterion function.

This optimization problem has a nice structure: it is a *linear programming* problem of a special type. In fact, it is a generalization of the well-known *transportation* problem. For details we refer to Section 2. The dual of the generalized transportation problem involves an upper approximation of the criterion function by a separable function. By means of duality theory it is possible to characterize the optimal joint distributions by means of a condition on their supports. Special attention is paid

in Section 2 to conditions under which the so-called NW Rule Solution generates an optimal solution. The name refers to the well-known device to get a starting basic feasible solution of a transportation problem.

The distribution problem and its dual have been studied in different areas of probability theory. In this paper it is applied to a stochastic programming problem arising from a project planning model. At stake is the decision on making a promise on the completion time of the project, whereas the durations of the activities of the project are random. In practical situations only very partial information on the joint distribution is available. PERT is a heuristic for this problem which is often used, but which has a poor probabilistic basis. We show in Section 3, that a minimax approach is possible; this leads to a distribution problem, where the criterion function represents the delay of the completion time of the project. Due to the special form of this criterion it is possible to use a reduced version of the dual problem which appears to be a finite dimensional convex minimization problem. By means of the dual solution the most unfavorable distributions can be characterized completely. This can be done in an appealing way by dualizing the reduced dual again, with a *flow* problem defined on the PERT network as a result. Optimal flows appear to be criticality numbers of the worst-case distributions.

In Section 4 we show that the results of Section 3 can be applied also if only partial knowledge on the marginals is available, as is not unusual in project planning. In Section 5 we give some numerical results, showing that the minimax approach to project planning is computationally feasible.

2. The distribution problem and its dual

Let n be a natural number, $n \geq 2$, let, for $i = 1, \ldots, n$, X_i be a Borel measurable subset of $\mathbb{R}$ and μ_i a probability measure on the Borel sets of $\mathbb{R}$ such that $\mu_i(X_i) = 1$. One might restrict X_i to the support of μ_i, or its convex hull, but this is not necessary. Defining $X := \prod_{i=1}^n X_i$, let $\mathcal{M}$ be the linear space of all finite, signed measures on the Borel sets of X, and denote the projection of $\mu \in \mathcal{M}$ on the coordinate space X_i by $\operatorname{proj}_i \mu$. For a Borel set $B \subset X_i$ one has $(\operatorname{proj}_i \mu)(B) := \mu(\prod_{j=1}^n E_j)$ where $E_i := B$ and $E_j := X_j$ for $j \neq i$. Then the *distribution problem* on X with criterion function $f: X \to \mathbb{R}$ and prescribed marginals μ_i, $i = 1, \ldots, n$, can be formulated as:

$$P: \qquad \underset{\mu \in \mathcal{M}}{\text{maximize}} \left\{ \int_X f \, d\mu : \operatorname*{proj}_i \mu = \mu_i, \, i = 1, \ldots, n, \, \mu \geq 0 \right\}.$$

Sometimes it is more convenient to formulate P in terms of n random variables ξ_i and the distribution functions F_i of μ_i; $F_i(x) := P(\xi_i < x)$. Then P can be rewritten as

$$P': \qquad \underset{H \in \mathcal{H}(F_1, \ldots, F_n)}{\text{maximize}} \; E_H f(\xi_1, \ldots, \xi_n)$$

where $\mathcal{H}(F_1, \ldots, F_n)$ denotes the set of all n-dimensional distribution functions with marginals $F_1, \ldots, F_n$.

Moreover, by transforming the criterion function it is possible to reformulate P' in such a way that the marginals are uniform distributions. Let F_i^{-1} be an arbitrary version of the inverse function of F_i. That is, F_i^{-1} is any real function on $(0,1)$ satisfying $F_i^{-1}(t) = x \Rightarrow F_i(x) \leq t \leq F_i(x+0)$. Then it is easy to show, that the distribution function of $F_i^{-1}(\omega_i)$ is F_i if ω_i is a random variable with uniform distribution on $(0,1)$, $U(0,1)$. Consequently, the distribution of $(F_1^{-1}(\omega_1), \ldots, F_n^{-1}(\omega_n))$ is feasible for P'. Moreover, *every* $H \in \mathcal{H}(F_1, \ldots, F_n)$ can be represented as the distribution of $(F_1^{-1}(\omega_1), \ldots, F_1^{-1}(\omega_n))$ for a random vector $(\omega_1, \ldots, \omega_n)$ with $U(0,1)$ marginals [39, 33], so that P' is equivalent to

$$P'': \quad \underset{G \in \mathcal{H}(U,\ldots,U)}{\text{maximize}} \; E_G f(F_1^{-1}(\omega_1), \ldots, F_n^{-1}(\omega_n)).$$

In fact, it is even possible to reformulate P'' in terms of only one $U(0,1)$ random variable ω_0:

$$P''': \quad \underset{\varphi}{\text{maximize}} \; Ef(\varphi_1(\omega_0), \varphi_2(\omega_0), \ldots, \varphi_n(\omega_0)),$$

where φ_i is an arbitrary so-called *rearrangement* of F_i^{-1} (see [33]). For our purposes the formulations P and P' are the most appropriate; we shall switch the notations freely.

It is clear that problem P is trivial if the criterion function f is *separable* on X, that is, if $f(x) = \sum_{i=1}^n f_i(x_i)$ $\forall x = (x_1, \ldots, x_n) \in X$. Then $\int f \, d\mu$ depends only on μ via the marginals μ_i which are fixed in P. Therefore a separable function f does not discriminate between the feasible solutions and does not deserve the name criterion function. If f is not separable on X, then still separable functions are important in the analysis of problem P. For example, a sufficient condition for the finiteness of the optimal value $\sup P$ of P is that $|f|$ has a separable upper bound which is integrable:

$$|f(x)| \leq \sum_{i=1}^n e_i(x_i) \;\; \forall x \in X, \quad \text{with} \int_{X_i} e_i \, d\mu_i < \infty \;\; \forall i. \tag{1}$$

Moreover, it will be shown that problem P is related by duality to an upper approximation of the criterion function by a separable function.

Before introducing the dual problem, let us first notice that the objective function as well as the constraints in P are linear in μ, so that P is a linear programming (LP) problem. It has a special structure; in fact, it is a generalized n-dimensional transportation problem. Indeed, for $n := 2$, $X_1 := \{s_1, \ldots, s_a\}$, $X_2 := \{t_1, \ldots, t_b\}$, $p_i := \mu_1(\{s_i\})$, $q_j := \mu_2(\{t_j\})$, $x_{ij} := \mu(\{s_i, t_j)\})$, $c_{ij} := f(s_i, t_j)$, P reduces to the well-known *transportation problem*

$$P_0: \quad \underset{x \in \mathbb{R}^{ab}}{\text{maximize}} \left\{ \sum_{i=1}^a \sum_{j=1}^b c_{ij} x_{ij} : \sum_{j=1}^b x_{ij} = p_i, \, i = 1, \ldots, a, \right.$$

$$\left. \sum_{i=1}^a x_{ij} = q_j, \, j = 1, \ldots, b, \, x_{ij} \geq 0 \;\forall i, j \right\}.$$

The dual problem of P_0 can be written as

$$D_0: \quad \underset{y_1 \in \mathbb{R}^a, y_2 \in \mathbb{R}^b}{\text{minimize}} \left\{ \sum_{i=1}^{a} p_i y_{1i} + \sum_{j=1}^{n} q_j y_{2j} : y_{1i} + y_{2j} \geq c_{ij} \ \forall i, j \right\}.$$

The dual problem of the distribution problem P can be formulated analogously. Let Y_i denote the linear space of real Borel functions on X_i which are integrable with respect to μ_i, and let Y be the linear space of all *separable* real functions on X determined by Y_i, $i = 1, \ldots, n$:

$$Y := \left\{ y \in \mathbb{R}^X : y(x_1, \ldots, x_n) = \sum_{i=1}^{n} y_i(x_i) \ \forall x \in X, y_i \in Y_i \ \forall i \right\}.$$

Generalizing D_0, we get as the dual for P the problem of finding a separable upper bound for the criterion function f with minimal expected value with respect to an arbitrary distribution $H \in \mathcal{H}(F_0, \ldots, F_n)$:

$$D: \quad \underset{y \in Y}{\text{minimize}} \left\{ \sum_{i=1}^{n} \int_{X_i} y_i \, d\mu_i : \sum_{i=1}^{n} y_i(x_i) \geq f(x_1, \ldots, x_n) \text{ for all } (x_1, \ldots, x_n) \in X \right\}.$$

If X is a finite set (so that each μ_i is finitely discrete) then obviously (P) and (D) are finite dimensional LP problems with max $P = \min D$. In the general case it is clear that for any $\mu \in \mathcal{M}$, $\mu \geq 0$, $\text{proj}_i \mu = \mu_i$, $i = 1, \ldots, n$, and for any $y \in Y$, $\sum_{i=1}^{n} y_i \geq f$, one has that

$$\int f \, d\mu \leq \int \left(\sum_{i=1}^{n} y_i \right) d\mu = \sum_{i=1}^{n} \int y_i \, d\mu_i,$$

so that $\sup P \leq \inf D$, which is *weak duality*, with equality if *complementary slackness* is true:

$$\mu\left(\left\{ x \in X : \sum_{i=1}^{n} y_i(x_i) = f(x_1, \ldots, x_n) \right\} \right) = 1.$$

Even strong duality can be proved:

Theorem 1. *Duality for P and D.*

Suppose that the criterion function f is upper semicontinuous, and that $|f|$ is bounded by an integrable separable function. Then the following statements are true.

(a) *Both P and D have optimal solutions, and they have the same finite optimal value:*

$$-\infty < \max P = \min D < +\infty. \tag{2}$$

(b) *The probability measure $\mu^* \in \mathcal{M}$ is optimal for P and the separable function $y^* \in Y$ is optimal for D if and only if*

(i) *μ^* has the required marginals μ_i, $i = 1, \ldots, n$;*

(ii) *y^* is an upper bound for f on X;*

(iii) *the support of μ^* is contained in the closure of the set X^* on which y^* coincides with f:*

$$X^* := \left\{ x \in X : \sum_{i=1}^{n} y_i^*(x_i) = f(x_1, \ldots, x_n) \right\}. \tag{3}$$

Proof. (a) As is seen from (1), $|\sup P| < \infty$. The other statements follow from a theorem of Kellerer ([18], proved in [19]). This theorem is much more general than ours (see the next remark) except for the fact that we allow for an integrable separable lower bound on f instead of a constant. However, the result remains true (see [19], private communication).

(b) This is a direct reformulation of primal and dual feasibility, and complementary slackness. □

Remark. The assumption $X_i \subset \mathbb{R}$ is not necessary for Theorem 1. For example, if X_i is compact metrizable and f is continuous then it is possible to use standard duality theory to prove max P = min D [9, 31]. Kellerer generalizes this result to Hausdorff spaces X_i and upper semicontinuous functions f, which are bounded from below, whereas sup P = min D is established for a large class of function f (analytic with respect to all upper semicontinuous functions on X). In the proof an application of Choquet's capacity theorem is a crucial tool.

In different degrees of generality, the pair of dual problems P and D has been studied extensively in probability theory. For example, with f being the indicator function of $\{x \in \mathbb{R}^2 : |x_1 - x_2| \leq \varepsilon\}$ problem P is a major tool in proving a famous theorem of Strassen: the minimum Ky Fan distance between two random variables with given distributions equals the Prokhorov distance of their distributions ([36, 34, 11, 14]; related results in [18, 37, 38]). Closely related is the problem of existence of probability measures under conditions including prescribed one dimensional marginals [15, 16, 17, 36, 10]. For the theory of variance reduction the problem P gives an appropriate framework with e.g. $f(\xi_1, \ldots, \xi_n) = -(\sum_{i=1}^{n} (\xi_i - E(\xi_i)))^2$, and several inequalities can be derived using P and D [31, 32, 30, 1, 39, 21]. In the literature mentioned often a statement related to problem P is derived by solving D explicitly. In this paper we follow the same line in deriving results for a minimax approach to a stochastic programming model for a project planning problem (see Section 3).

One feasible solution for P is the joint distribution specified by *independence* $\mu^{\text{IND}}(\prod_{i=1}^{n} B_i) := \prod_{i=1}^{n} \mu_i(B_i)$. It is interesting to realize, that the independent solution can only be optimal for P in trivial cases. Indeed, theorem 1 shows, that μ^{IND} is optimal for P iff Borel sets $S_i \subset X_i$, $i = 1, \ldots, n$, exist such that $\mu_i(S_i) = 1$ $\forall i$ and f is separable on $\prod_{i=1}^{n} S_i$. In such a case every feasible solution for P is in fact optimal. If the criterion function f is not separable one may expect that the optimal solutions of P show heavy dependence between the marginals.

One other special feasible solution for P deserves attention. Let μ^0 be the probability measure on X defined by the distribution function

$$H^0(z_1, \ldots, z_n) := \min_{i=1,\ldots,n} F_i(z_i), \quad z \in \mathbb{R}^n. \tag{4}$$

It is easy to verify that H^0 indeed is a distribution function with marginals F_i. We shall call μ^0 the *NW Rule Solution* or the *NW Rule Distribution* of P, since in the case of the transportation problem P_0 it reduces to the well-known basic feasible solution determined by the North West Corner Rule. In terms of P'' the NW Rule Solution corresponds to the uniform distribution on the main diagonal in the unit cube, that is $\omega_1 = \omega_2 = \cdots = \omega_n$. In other words, H^0 is the distribution function of the random vector

$$(F_1^{-1}(\omega), F_2^{-1}(\omega), \ldots, F_n^{-1}(\omega)), \quad \omega \text{ is } U(0, 1). \tag{5}$$

In P''' the NW Rule Solution is described by $\varphi_i = F_i^{-1}$ $\forall i$: no real rearrangement is made.

Since μ^0 depends only on the marginals but not on the criterion function f, it is not to be expected to be optimal for P in general. However, it is possible to give conditions on f which guarantee that μ^0 is optimal for all specifications of μ_i, $i = 1, \ldots, n$. The function f is called *L-superadditive* ([23]) if for all $x \in \mathbb{R}^n$, all unit vectors e_i and e_j, $i \neq j$, and all nonnegative scalars α and β

$$f(x + \alpha e_i + \beta e_j) + f(x) \geq f(x + \alpha e_i) + f(x + \beta e_j).$$

Theorem 2. *If $\mu_1, \ldots, \mu_n$ have first moments and f is L-superadditive then the NW Rule Solution μ^0 is optimal for P.*

Proof. See [33]. □

The well-known result that the NW Rule Solution maximizes the correlation coefficient of two random variables can be formulated as a consequence of this theorem: take $f(x_1, x_2) = x_1 \cdot x_2$. For the transportation problem P_0 the NW Rule Solution is, as can easily be verified,

$$x_{kl}^0 = \left[\min\left(\sum_{i=1}^{k} p_i, \sum_{j=1}^{l} q_j \right) - \max\left(\sum_{i=1}^{k-1} p_i, \sum_{j=1}^{l-1} \right) \right]^+ \tag{6}$$

where $[z]^+ := \max(0, z)$.

Corollary 3. *The NW Rule Solution* (6) *is optimal for P_0 for all values of $p_i \geq 0$, $q_j \geq 0$ with $\sum_{i=1}^{a} p_i = \sum_{j=1}^{b} q_j$ $(= 1)$ if and only if, for each $k = 1, \ldots, a-1$ and $l = 1, \ldots, b-1$,*

$$c_{kl} + c_{k+1,l+1} \geq c_{k+1,l} + c_{k,l+1}. \tag{7}$$

Proof. Sufficiency follows from Theorem 2 (a direct proof based on Theorem 1 is

also not difficult). Necessity can be proved by contradiction: if (7) does not hold for $(\bar{k}, \bar{l})$ then (6) is not optimal if $p_{\bar{k}} = p_{\bar{k}+1} = \frac{1}{2}$ and $q_{\bar{l}} = q_{\bar{l}+1} = \frac{1}{2}$. □

As a typical application of Corollary 3 let us interpret P_0 as a *dynamic production* model. Then p_i is the production capacity in period i, q_j is the demand in period j, and x_{ij} is the production in period i destined for demand in period j, with corresponding unit cost c_{ij}. Both indices i and j are ordered in a natural way, namely increasing with time. The NW Rule can then be interpreted as a FIFO type of strategy: the oldest demands have highest priority. According to corollary 3 it gives the minimal cost solution for all values of the capacity parameters p_i and demand parameters q_i (provided the problem is balanced: $\sum_i p_i = \sum_j q_j$) if for all (k, l) the cost vector c satisfies

$$c_{kl} + c_{k+1,l+1} \leq c_{k+1,l} + c_{k,l+1}.$$

This is true for example if $c_{ij} = d + h(i-j)$, h convex. On the other hand, under the same conditions for c the total costs are maximized under the LIFO strategy. The corresponding solution can be called the NE Rule Solution since high values of j are coupled with low values of i and vice versa. By the way, generalizations of the NE Rule Solution to general n are not as obvious as in the NW case, see e.g. [31].

From (6) useful formulas for μ^0 and H^0 can be derived. For example, let $-\infty \leq a_i < b_i \leq \infty$, $-\infty \leq z_i \leq \infty$, $I, I_i \subset \{1, \ldots, n\}$ then the probability of each rectangle under H^0 can be expressed in terms of the marginal distribution functions:

$$\begin{aligned}
&P_{H^0}(a_i \leq \xi_i < b_i,\ i \in I) = \left[\min_I F_i(b_i) - \max_I F_i(a_i) \right]^+, \\
&P_{H^0}(a_i < \xi_i \leq b_i,\ i \in I) = \left[\min_I F_i(b_i + 0) - \max_I F_i(a_i + 0) \right]^+, \\
&P_{H^0}(\xi_i < z_i,\ i \in I_1;\ \xi_i \leq z_i,\ i \in I_2;\ \xi_i \geq z_i,\ i \in I_3;\ \xi_i > z_i,\ i \in I_4) \\
&\quad = \left[\min\left(\min_{I_1} F_i(z_i), \min_{I_2} F_i(z_i + 0) \right) \right. \\
&\qquad \left. - \max\left(\max_{I_3} F_i(z_i), \max_{I_4} F_i(z_i + 0) \right) \right]^+.
\end{aligned} \tag{8}$$

Hence, for any $z \in \mathbb{R}^n$ and any $h, k \in \{1, \ldots, n\}$ we have

$$\begin{aligned}
&P_{H^0}(\xi_h < z_h,\ \xi_k \geq z_k) \cdot P_{H^0}(\xi_h \geq z_h,\ \xi_k < z_k) = 0, \\
&P_{H^0}(\xi_h < z_h,\ \xi_k > z_k) \cdot P_{H^0}(\xi_h \geq z_h,\ \xi_k \leq z_k) = 0, \\
&P_{H^0}(\xi_h \leq z_h,\ \xi_k > z_k) \cdot P_{H^0}(\xi_h > z_h,\ \xi_k \leq z_k) = 0.
\end{aligned} \tag{9}$$

This is a characteristic property of the NW Distribution. It holds also for disjunct index sets I_1 and I_2 instead of $\{h\}$ and $\{k\}$. Hence, H^0 has positive probability mass in at most one of each pair of opposite translated orthants in $\mathbb{R}^n$, except if the

positive and negative orthants are concerned. For that case we have a characteristic maximality property:

Theorem 4. *For arbitrary* $z \in \mathbb{R}^n$ *and* $I_1, I_2 \subset \{1, \ldots, n\}$, *define* A_1 *and* A_2 *by*

$$A_1 := \{x \in \mathbb{R}^n : x_i < z_i \text{ for } i \in I_1 \text{ and } x_i \leq z_i \text{ for } i \in I_2\},$$

$$A_2 := \{x \in \mathbb{R}^n : x_i > z_i \text{ for } i \in I_1 \text{ and } x_i \geq z_i \text{ for } i \in I_2\}.$$

Then $\max_{\mu \in \mathcal{M}} \{\mu(A_k) : \text{proj}_i \mu = \mu_i \ \forall i, \mu \geq 0\} = \mu^0(A_k)$, $k = 1, 2$. *Consequently, the NW Rule Distribution maximizes the probability of all translated positive or negative orthants.*

Proof. By directly using (8), or from Theorem 2 by showing that the indicator function of A_k is L-superadditive. □

If we call a distribution H with marginals $F_1, \ldots, F_n$ *positively quadrant dependent* if $H(z_1, \ldots, z_n) \geq \prod_{i=1}^n F_i(z_i)$ for all $z \in \mathbb{R}^n$ (see [22]), then Theorem 4 characterizes the NW Rule Distribution as the *maximally* positively quadrant dependent distribution compatible with the fixed marginals.

Finally, we remark that from Theorem 4 it follows that the NW Rule Distribution is optimal for P if $f(x)$ or $f(-x)$ is the distribution function of a finite measure on the Borel sets of $\mathbb{R}^n$, as proved in an earlier version of this paper (or as a consequence of [30]).

3. Worst-case analysis in project planning

Suppose that one has to make a decision ('promise') T on the unknown completion time τ of a project. Reduction of T is profitable at a rate $c \geq 0$, but if the promise is not kept the delay $[\tau - T]^+$ is penalized with $q \geq c$ per time unit. Assuming that τ can be considered as a random variable with known distribution, the simple model (which is in fact a stochastic program with simple recourse)

$$\underset{T}{\text{minimize}} \{cT + q \cdot E[\tau - T]^+\} \tag{10}$$

is a reasonable starting point for the analysis of the decision on T. T^* is optimal for this 'production model' ($T :=$ production, $\tau :=$ demand, $c :=$ unit production cost, $q :=$ unit shortage cost) iff $T^* \in \Phi^{-1}(1 - \alpha)$, where $\alpha := c/q$ and Φ is the distribution function of τ, that is iff

$$P(\tau > T^*) \leq \alpha \leq P(\tau \geq T^*). \tag{11}$$

If there is only partial knowledge on the distribution of τ, one may introduce the

minimax variant of (10),

$$\underset{T}{\text{minimize}} \left\{ \alpha T + \sup_{\Phi \in \mathscr{A}} E_{\Phi}[\tau - T]^+ \right\}, \quad 0 \leq \alpha \leq 1, \tag{12}$$

where $\mathscr{A}$ is the family of distributions of τ compatible with the partial information.

Remark. Although (10) is practically equivalent to the chance constrained (CC) model $\min\{T: P(\tau > T) \leq \alpha\}$, it is not true that (12) is equivalent to the minimax CC model

$$\text{minimize} \left\{ T: \sup_{\Phi \in \mathscr{A}} P_{\Phi}(\tau > T) \leq \alpha \right\}, \quad 0 \leq \alpha \leq 1. \tag{13}$$

In (12) the amount of the delay is important, not only its sign as in (13). If T^* and Φ^* solve (12) then $E_{\Phi}[\tau - T^*]^+ \leq E_{\Phi^*}[\tau - T^*]^+$ for all $\Phi \in \mathscr{A}$: in this sense the distribution Φ^* one is looking for in (12) can be called a *worst-case distribution*. It depends on T^*, hence on α. However, Φ^* is not necessarily a worst-case distribution in the sense of (13). Of course, $P_{\Phi^*}(\tau > T^*) \leq \alpha$, see (11), but the existence of a $\Phi \in \mathscr{A}$ with $P_{\Phi}(\tau > T^*) > \alpha$ is not excluded.

Usually, a project is represented by a directed graph with one source and one sink. The arcs $i = 1, \ldots, n$ represent activities which have to be completed in order for the project to be completed, and the precedence relations between the activities are given by the network. Let B_j, $j = 1, \ldots, p$, be the index set of all activities at path j from source to sink, then $B_1, \ldots, B_p$ are different, nonempty sets whose union is $\{1, \ldots, n\}$. Denote the duration of activity i by ξ_i. Then the project completion time is given by $\tau = R(\xi)$ where

$$R(x) := \max_{j=1,\ldots,p} \sum_{i \in B_j} x_i, \quad x \in \mathbb{R}^n. \tag{14}$$

If the maximum is attained by j, path j is called *critical*. We shall denote the set of critical paths by $\mathrm{CP}(\xi)$,

$$\mathrm{CP}(x) := \arg \max_{j=1,\ldots,p} \sum_{i \in B_j} x_i, \quad x \in \mathbb{R}^n. \tag{15}$$

The set of *critical activities* is then $\mathrm{CA}(\xi)$,

$$\mathrm{CA}(x) := \bigcup \{i: i \in B_j, j \in \mathrm{CP}(x)\}, \quad x \in \mathbb{R}^n. \tag{16}$$

Obviously, each activity on a critical path is critical. The reverse is also true, in the sense that

$$j \in \mathrm{CP}(x) \Leftrightarrow i \in \mathrm{CA}(x) \text{ for all } i \in B_j.$$

This is due to the fact that if two critical paths j and k contain the same activity i one gets another critical path by replacing in j the activities before (or after) i by those of k.

In our analysis we shall assume that the network structure of the project is known, whereas the time durations ξ_i are random variables. One would like to be able to calculate the probability distribution of $R(\xi)$; in particular, $ER(\xi)$, $P(R(\xi) > T)$ and $E[R(\xi) - T]^+$ for certain values of T. Also *criticality indices* as $P(j \in \mathrm{CP}(\xi))$, $P(j \in \mathrm{CP}(\xi), R(\xi) > T)$ and $P(i \in \mathrm{CA}(\xi))$, $P(i \in \mathrm{CA}(\xi), R(\xi) > T)$ are important in the analysis of the random case. However, there are two practical difficulties. First of all, the joint distribution of $\{\xi_1, \ldots, \xi_n)$ has to be specified completely. Moreover, even if this is possible, the amount of calculations is prohibitive for networks of a reasonable size (see e.g. [24, 6]). Under the assumption of independence of $\xi_1, \ldots, \xi_n$ it is possible to derive bounds for the expectation of $R(\xi)$ [8, 5, 12] and for its distribution function [20, 35, 28, 13] from the marginal distributions. We shall, however, *not* assume independence, and analyze the 'worst-case' problem (12), where $\mathscr{A}$ is the family of all distributions of $\tau = R(\xi)$ compatible with the knowledge about the marginal distributions. In this section we shall assume that for each ξ_i the complete distribution function F_i, $F_i(t) = P(\xi_i < t)$ for $t \in \mathbb{R}$, is known, with finite mean value $\bar{\xi}_i$ and support $[a_i, b_i]$, $a_i = \lim_{\alpha \downarrow 0} F^{-1}(\alpha) \geq 0$, $b_i = \lim_{\alpha \uparrow 1} F_i^{-1}(\alpha) \leq +\infty$. (If $v = +\infty$, the notation $[u, v]$ has to be interpreted as $[u, \infty)$.) In Section 4 the assumption of the complete knowledge of the marginal distributions will be relaxed. It will appear that the worst-case analysis is tractable from a computational point of view. A similar problem has also been studied in [25] and [27] without explicit use of duality, but by Duality Theorem 1 the results become more transparent and more general.

Under the specification mentioned above problem (12) becomes

$$PP(\alpha): \quad \text{find} \quad \hat{h}(\alpha) := \inf_{t \in \mathbb{R}} \{\alpha T + h(T)\}, \quad 0 \leq \alpha \leq 1,$$

with

$$P(T): \quad h(T) := \sup_{H \in \mathscr{H}(F_1, \ldots, F_n)} E_H f(\xi_1, \ldots, \xi_n), \quad T \in \mathbb{R},$$

where

$$f(x) = f_T(x) := [R(x) - T]^+, \quad x \in \mathbb{R}^n. \tag{17}$$

Notice that the inner problem $P(T)$, which is of independent interest and which will be discussed first, is a distribution problem of the type P' (or P) of Section 2, with $X_i := \mathbb{R}$. Moreover, f satisfies the conditions of the Duality Theorem 1. Hence, the supremum is attained and the optimal value function $h(T)$ and the corresponding worst-case distributions for τ can be analyzed by means of the dual of $P(T)$

$$D(T): \quad \underset{y \in Y}{\text{minimize}} \{Ey : y \geq f\},$$

where Y is the set of separable real Borel functions on X, of which the components y_i are integrable with respect to F_i. It will appear that $D(T)$ can be reduced to a convex program in $\mathbb{R}^n$. In order to show this, we define for $i = 1, \ldots, n$

$$\tilde{Y}_i := \{y_i \in \mathbb{R}^{X_i} : \exists w_i \in \mathbb{R}, z_i \in \mathbb{R} \text{ s.t. } y_i(x_i) = w_i + [x_i - z_i]^+ \ \forall x_i \in X_i\}$$

and

$$\tilde{Y} := \{y \in \mathbb{R}^X : y = \sum_{i=1}^{n} y_i, y_i \in \tilde{Y}_i, i = 1, \ldots, n\}.$$

Since $\int |y_i(x_i)| \, dF_i(x_i) < \infty$ for all $y_i \in \tilde{Y}_i$, $\tilde{Y}$ is a subset of Y. Since for arbitrary but fixed $(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ the criterion function f, considered as a function of x_i only, belongs to $\tilde{Y}_i$, it is not surprising that in $D(T)$ only upper bounds $y \in \tilde{Y}$ for f are relevant.

Proposition 5. *For any $y \in Y$, $y \geq f$ there exists a $\tilde{y} \in \tilde{Y}$ such that $y \geq \tilde{y} \geq f$.*

Proof. Fix $y \in Y$ with $y \geq f$. Then

$$\infty > y_1(x_1) \geq f(x) - \sum_{i=2}^{n} y_i(x_i) > -\infty \quad \forall x \in \mathbb{R}^n. \tag{18}$$

Denote the vector $(x_2, \ldots, x_n) \in \mathbb{R}^{n-1}$ by v_1. Then $f(x) - \sum_{i=2}^{n} y_i(x_i)$ can be written as $w_1(v_1) + [x_1 - z_1(v_1)]^+$ for certain $w_1(v_1) \in \mathbb{R}$, $z_1(v_1) \in \mathbb{R}$ not depending on x_1. With $\tilde{y}_1$ defined by

$$\tilde{y}_1(x_1) := \sup_{v_1 \in \mathbb{R}^{n-1}} \{w_1(v_1) + [x_1 - z_1(v_1)]^+\}, \quad x_1 \in \mathbb{R}, \tag{19}$$

it follows from (18) that

$$\infty > y_1(x_1) \geq \tilde{y}_1(x_1) \geq f(x) - \sum_{i=2}^{n} y_i(x_i) > -\infty \quad \forall x \in \mathbb{R}^n. \tag{20}$$

We shall show that $\tilde{y}_1 \in \tilde{Y}_1$. Define

$$\tilde{w}_1 := \sup_{v_1 \in \mathbb{R}^{n-1}} w_1(v_1), \qquad \tilde{z}_1 := -\sup_{v_1 \in \mathbb{R}^{n-1}} (w_1(v_1) - z_1(v_1)) + \tilde{w}_1.$$

From (18) it follows that both suprema are finite. Moreover,

$$\begin{aligned}
\tilde{y}_1(x_1) &= \sup_{v_1} \max(w_1(v_1), w_1(v_1) - z_1(v_1) + x_1) \\
&= \max\left(\sup_{v_1} (w_1(v_1), x_1 + \sup_{v_1} (w_1(v_1) - z_1(v_1)) \right) \\
&= \max(\tilde{w}_1, \tilde{w}_1 + x_1 - \tilde{z}_1) = \tilde{w}_1 + [x_1 - \tilde{z}_1]^+
\end{aligned}$$

so that $\tilde{y}_1 \in \tilde{Y}_1$. From (20) we conclude that the replacement of y_1 by $\tilde{y}_1$ does not disturb the feasibility condition $y \geq f$. In the same way, $\tilde{y}_k \in \tilde{Y}_k$ can be constructed, successively for $k = 2, \ldots, n$, such that

$$y_k(x_k) \geq \tilde{y}_k(x_k) \geq f(x) - \sum_{i<k} \tilde{y}_i(x_i) - \sum_{i>k} y_i(x_i), \quad x \in \mathbb{R}^n.$$

For $k = n$ we have the desired result. □

Proposition 5 shows that in $D(T)$ one may restrict attention to functions $y \in \tilde{Y}$ rather than $y \in Y$. Doing this, a great simplification is possible: if $y_i(x_i) = w_i + [x_i - z_i]^+$ then

$$\sum_{i=1}^{n} y_i(x_i) \geq f(x)$$

is true for all $x \in \mathbb{R}^n$ iff it holds $x = z$; that is iff $\sum_{i=1}^{n} w_i \geq f(z)$. Hence $D(T)$ can be reduced to

$$\tilde{D}(T): \quad \underset{z \in \mathbb{R}^n}{\text{minimize}} \{f(z) + G(z)\},$$

where

$$G(z) := \sum_{i=1}^{n} G_i(z_i), \quad z \in \mathbb{R}^n, \quad \text{and}$$
$$G_i(z_i) := E[\xi_i - z_i]^+ = \int [x_i - z_i]^+ \, dF_i(x_i) = \int_{z_i}^{\infty} (1 - F_i(x_i)) \, dx_i, \quad z_i \in \mathbb{R}. \tag{21}$$

Clearly, G_i is a finite, nonnegative, nonincreasing, convex function on $\mathbb{R}$. Its asymptotes for $|z_i| \to \infty$ are given by $[\bar{\xi}_i - z_i]^+$; in fact, $G_i(z_i) = \bar{\xi}_i - z_i$ iff $z_i \leq a_i$ and $G_i(z_i) = 0$ iff $z_i \geq b_i$ (see Figure 1).

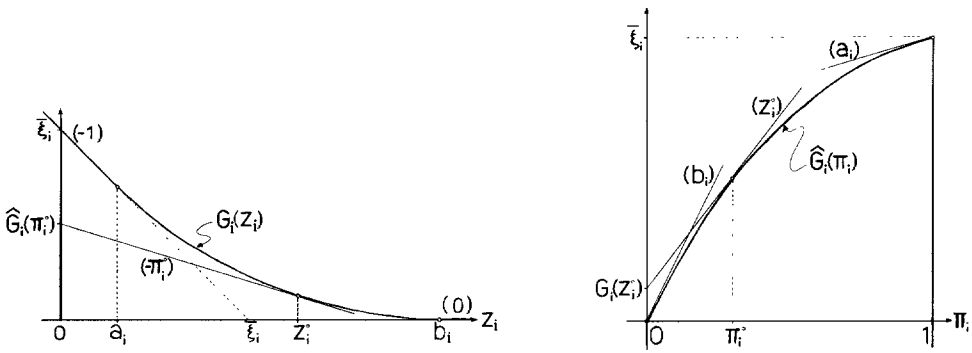


Fig. 1. The function G_i and its conjugate function $\hat{G}_i$. Slopes are indicated by $(\cdot)$; z_i^0 and π_i^0 are related by subgradient duality.

Theorem 6. *$D(T)$ and $\tilde{D}(T)$ are equivalent.*

(a) *For each $T \in \mathbb{R}$ the minimum in $D(T)$ is attained, and*

$$\min \tilde{D}(T) = \min D(T) = \max P(T) = h(T), \quad T \in \mathbb{R}.$$

(b) *If $z^* \in \mathbb{R}^n$ is optimal for $\tilde{D}(T)$ then $y^* \in \tilde{Y}$, defined by $y_i^*(x_i) := w_i + [x_i - z_i^*]^+$, $i = 1, \ldots, n$ with $\sum_{i=1}^{n} w_i = f(z^*)$, is optimal for $D(T)$. If $y^* \in Y$ is optimal for $D(T)$, then a $\tilde{y} \in \tilde{Y}$ exists which is also optimal for $D(T)$ with $P_{H^*}(\tilde{y} = y^*) = 1$ for any optimal distribution H^* for $P(T)$.*

Proof. Follows directly from Theorem 1 and (the discussion after) Proposition 5. □

The reduced dual $\tilde{D}(T)$ can be interpreted as the problem of finding reference values ('promises') z for the durations of the activities, in such a way that the delay (over T) of the project completion based on z, is balanced with the sum of the expected delays (over z) of the random activity durations. It is easy to prove, that an optimal solution z^* exists with $a_i \leq z_i^* \leq b_i$: the objective is a nonincreasing function of z_i if $z_i \leq a_i$, and nondecreasing if $z_i \geq b_i$. $\tilde{D}(T)$ is a convex program; in fact, it can be seen as a *simple recourse* model of stochastic programming, and all the preceding observations are well-known in that setting.

From Theorem 6a we know, that the optimal value function h of $P(T)$ and $D(T)$ can be written as

$$h(T) = \min_{z \in \mathbb{R}^n} \{[R(z) - T]^+ + G(z)\}. \tag{22}$$

This representation allows for a complete characterization of h (see Figure 2):

Proposition 7. *The function h is finite, nonnegative, nonincreasing and convex. Its asymptotes are given by $[T_c - T]^+$, where T_c is the finite number*

$$T_c := \inf_{z \in \mathbb{R}^n} (R(z) + G(z)) \geq R(\bar{\xi}).$$

In fact,

$$h(T) = T_c - T \quad \text{iff } T \leq T_1 := \sup\{R(z): z \text{ minimizes } R(z) + G(z)\},$$

$$h(T) = 0 \qquad \text{iff } T \geq T_0 := R(b).$$

Since all $a_i \geq 0$ it holds that $T_1 \geq 0$. Moreover, $T_1 \leq T_c \leq T_0 \leq \infty$, and T_0 is finite iff all b_i are finite.

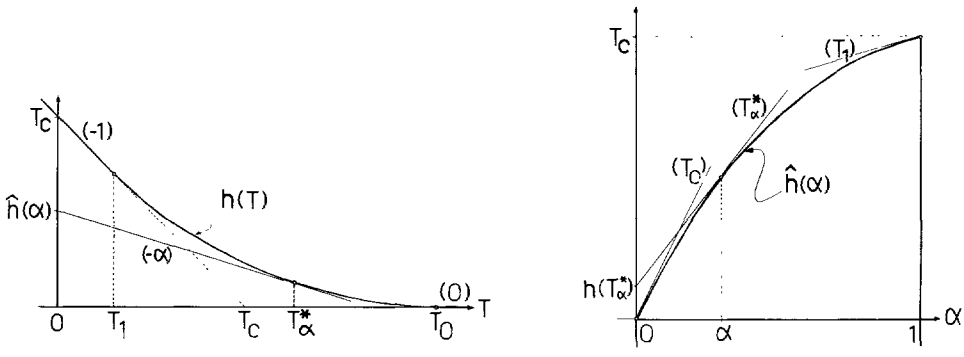


Fig. 2. The function h and its conjugate function $\hat{h}$. Slopes are indicated by $(\cdot)$; T_α^* and α are related by subgradient duality.

Proof. The first statement is trivial. Since for all $z \in \mathbb{R}^n$

$$\begin{aligned} R(z)+G(z) &\geq \max_j \sum_{i\in B_j} z_i + \sum_i [\bar{\xi}_i - z_i]^+ \\ &= \max_j \left\{ \sum_{i\in B_j} \bar{\xi}_i + \sum_{i\in B_j} [z_i - \bar{\xi}_i]^+ + \sum_{i\notin B_j} [\bar{\xi}_i - z_i]^+ \right\} \\ &\geq \max_j \sum_{i\in B_j} \bar{\xi}_i = R(\bar{\xi}) \end{aligned}$$

it follows that $T_c \geq R(\bar{\xi})$. The infimum in the definition of T_c is attained, since one might restrict the attention to $z \geq a$ (≥ 0), and $R(z)+G(z) \to \infty$ if $\|z\| \to \infty$, $z \geq 0$. Consequently, $T_1 \geq R(a) \geq 0$. Also,

$$h(T)+T = \min_{z\in\mathbb{R}^n} \{\max(R(z), T)+G(z)\} \geq T_c$$

and for T small enough the difference is arbitrarily small:

$$h(T)+T-T_c = \min_{z\in\mathbb{R}^n} \{\max(R(z)+G(z)-T_c,\ T+G(z)-T_c)\} \leq \varepsilon$$

if $T \leq R(z_\varepsilon) - \varepsilon$ for a $z_\varepsilon \in \mathbb{R}^n$ with $R(z_\varepsilon)+G(z_\varepsilon) \leq T_c + \varepsilon$. Therefore $\lim_{T\to-\infty}(h(T)+T) = T_c$; the limit is attained for $T \leq T_1$, but $h(T) > T_c - T$ if $T > T_1$, both by definition of T_1. Clearly $h(T) \geq 0$. Moreover, since $\lim_{z_i\to\infty} G_i(z_i) = 0$, there exists a $z'_\varepsilon \in \mathbb{R}^n$ with $G(z'_\varepsilon) \leq \varepsilon$ and

$$[R(z'_\varepsilon) - T]^+ + G(z'_\varepsilon) \leq \varepsilon \quad \text{if } T \geq R(z'_\varepsilon),$$

so that $\lim_{T\to\infty} h(T) = 0$. It is easy to show that $h(T) > 0$ for all $t \in \mathbb{R}$ unless all b_i are finite and $T \geq R(b)$. □

Corollary 8. *For $T \geq T_0$, $z^* \in \mathbb{R}^n$ is optimal for $\tilde{D}(T)$ iff $z^* \geq b$ and $R(z^*) \leq T$. For $T \leq T_1$, $z^* \in \mathbb{R}^n$ is optimal for $\tilde{D}(T)$ iff $z^* \in \arg\min_z (R(z)+G(z))$ and $R(z^*) \geq T$. In particular, $R(z^*) = T$ if z^* is optimal for $\tilde{D}(T)$ and $T = T_1$ or $T = T_0$.*

Proof. By substitution of $h(T) = [T_c - T]^+$ in (22). □

Note. $T_1 < T < T_0$ implies also $R(z^*) = T$, see Corollary 11.

The values $T \in \mathbb{R}$ with $T \geq T_0$ are not restrictive in problem $P(T)$ since in that case $P_H(R(\xi) \leq T) = 1$ for all feasible distributions H. On the other hand, the specification of T in $(-\infty, T_1]$ is very demanding in the sense that $P_{H^*}(R(\xi) \geq T) = 1$ for the worst-case distribution H^*. Both statements are easily derived by the substitution $h(T) = [T_c - T]^+$. In problem $PP(\alpha)$ the optimal T^*_α are characterized by $-\alpha \in \partial h(T^*_\alpha)$, so that $T^*_\alpha \in (T_1, T_0)$ if $\alpha \in (0, 1)$ with $\lim_{\alpha\downarrow 0} T^*_\alpha = T_0$, $\lim_{\alpha\uparrow 1} T^*_\alpha = T_1$. Values for T^*_α outside $[T_1, T_0]$ can only occur if $\alpha = 1$ (but then also T_1 itself is good) or if $\alpha = 0$ (but then also T_0 is good).

The number T_c has an interesting interpretation as the mean value of a certain distribution Φ_c. Indeed, the subgradient of h determines the function

$$\Phi_c(t) := 1 + \min\{u: u \in \partial h(t)\}, \quad t \in \mathbb{R}, \tag{23}$$

and from the characterization of h in Proposition 7 it follows that Φ_c is a probability distribution function of a random variable η with support in $[T_1, T_0]$. Representing h as the integral of its subgradient (see [29, Theorem 24.2]) we get

$$h(T) = \int_T^\infty (1 - \Phi_c(t))\, dt = \int [t - T]^+ \, d\Phi_c(t), \quad T \in \mathbb{R}, \tag{24}$$

so that

$$\int t \, d\Phi_c(t) = \lim_{T \to -\infty} \int \max(t, T)\, d\Phi_c(t) = \lim_{T \to -\infty} (h(T) + T) = T_c.$$

The distribution function Φ_c is interesting by itself since for all $T \in \mathbb{R}$

$$h(T) = \sup_{H \in \mathcal{H}(F_1, \ldots, F_n)} E_H[R(\xi) - T]^+ = E_{\Phi_c}[\tau - T]^+.$$

In general, there is *no* distribution of ξ in $\mathcal{H}(F_1, \ldots, F_n)$ for which $\tau = R(\xi)$ is distributed according to Φ_c. The existence of such a distribution is equivalent to having the *same* worst-case distribution for $P(T)$ *for all* $T \in \mathbb{R}$. Φ_c can be interpreted as the smallest upperbound for all feasible distributions of τ in the sense of *convex ordering* of probability distributions (see [25, 27]).

It is well-known, that for deterministic time durations z_i, $i = 1, \ldots, n$, the project completion time $R(z)$ can be described in terms of *flows* in the PERT network. Indeed, augment the network with an arc 0 from sink to source, and denote the flow in arc i by π_i, $i = 0, 1, \ldots, n$. A flow is called a *feasible circulation* of the augmented network if in each arc the flow is in the direction of the arc ($\pi_i \geq 0$) and if at each node the flow is conserved. Clearly, π_0 is then the total flow in the original network from source to sink, and $\pi_i \leq \pi_0 \; \forall i$. It is easy to see, that

$$R(z) = \max\left\{ \sum_{i=1}^n z_i \pi_i : \pi \in \Pi, \pi_0 = 1 \right\} \tag{25}$$

where Π is the polyhedron in $\mathbb{R}^{n+1}$ corresponding to all feasible circulations. A similar representation by flows is possible for the stochastic case. First we define for $i = 1, \ldots, n$ the 'concave conjugate' function $\hat{G}_i$ of G_i,

$$\hat{G}_i(\pi_i) := \inf_{z_i \in \mathbb{R}} \{\pi_i z_i + G_i(z_i)\}, \quad \pi_i \in \mathbb{R}, \qquad \hat{G}(\pi) = \sum_{i=1}^n G_i(\pi_i), \; \pi \in \mathbb{R}^{n+1}. \tag{26}$$

The graph of the function $\hat{G}_i$ is given in Figure 1.

Proposition 9. $\hat{G}_i(\pi_i) = -\infty$ *if* $\pi_i < 0$ *or* $\pi_i > 1$, *and*

$$\hat{G}_i(\pi_i) = \int_0^{\pi_i} F_i^{-1}(1-t_i)\,dt_i, \quad 0 \leq \pi_i \leq 1, \tag{27}$$

where F_i^{-1} *is any real function on* $(0,1)$ *satisfying* $F_i^{-1}(t) = x \Rightarrow F_i(x) \leq t \leq F_i(x+0)$. *On* $[0,1]$ $\hat{G}_i$ *is a finite nonnegative increasing concave function, with* $\hat{G}_i(0) = 0$, $\hat{G}_i(1) = \bar{\xi}_i$, *and with slope* b_i *in* $\pi_i = 0+$ *and slope* a_i *in* $\pi_i = 1-$. *Moreover,*

$$G_i(z_i) = \sup_{0 \leq \pi_i \leq 1} \{-\pi_i z_i + \hat{G}_i(\pi_i)\}, \quad z_i \in \mathbb{R}. \tag{28}$$

The minimizing z_i *in* (26) *and the maximizing* π_i *in* (28) *are related via subgradient duality*:

$$z_i \in \partial \hat{G}_i(\pi_i) \Leftrightarrow -\pi_i \in \partial G_i(z_i) \Leftrightarrow \hat{G}_i(\pi_i) = \pi_i z_i + G_i(z_i), \tag{29}$$

and these conditions are equivalent to

$$F_i(z_i) \leq 1 - \pi_i \leq F_i(z_i + 0). \tag{30}$$

Note. Whereas the subgradient of the convex function G_i is

$$\partial G_i(z_i) := \{\pi_i \in \mathbb{R}: G_i(x) \geq G_i(z_i) + \pi_i(x - z_i)\ \forall x \in \mathbb{R}\}, \tag{31}$$

the definition of the subgradient (or supergradient) of the concave function $\hat{G}_i$ is

$$\partial \hat{G}_i(\pi_i) := \{z_i \in \mathbb{R}: \hat{G}_i(y) \leq \hat{G}_i(\pi_i) + z_i(y - \pi_i)\ \forall y \in \mathbb{R}\}. \tag{32}$$

The equivalence in (29) is illustrated in Figure 1.

Proof of Proposition 9. (28) and (29) are a direct consequence of the fact that $-\hat{G}_i(-\pi_i)$ is the convex conjugate function of G (see [29] Theorem 23.5). Formula (27) is true, since it gives the correct supergradients (see (21) and (29)) and the correct value at $\pi_i = 0$ (see (26) and (29)). Since $G_i(z_i) \geq \bar{\xi}_i - z_i$ with equality for $z_i \leq a_i$ it follows that $\hat{G}_i(1) = \bar{\xi}_i$. Finally, the statement on the directional derivatives in 0 and 1 follows from (29), and (30) follows from (21) and (29). □

By performing a dualization we shall give a representation of the optimal value function h of $P(T)$ in terms of circulations in the augmented network. Similar results hold for the optimal value function $\hat{h}$ of $PP(\alpha)$, which can be interpreted as the concave conjugate of h (see also Figure 2).

Theorem 10

$$\tilde{P}(T): \quad h(T) = \max_{\pi \in \Pi} \{-\pi_0 T + \hat{G}(\pi): \pi_0 \leq 1\}, \quad T \in \mathbb{R},$$

$$DD(\alpha): \quad \hat{h}(\alpha) = \max_{\pi \in \Pi} \{\hat{G}(\pi): \pi_0 = \alpha\}, \quad 0 \leq \alpha \leq 1.$$

Moreover, h and $\hat{h}$ determine each other by conjugacy:

$$\hat{h}(\alpha) = \inf_{T \in \mathbb{R}} \{\alpha T + h(T)\}, \quad \alpha \in [0, 1], \tag{33}$$

$$h(T) = \sup_{\alpha \in [0,1]} \{-\alpha T + \hat{h}(\alpha)\}, \quad T \in \mathbb{R}. \tag{34}$$

In particular, the minimizing T in (33), *(i.e. the optimal solutions of $PP(\alpha)$) and the maximizing α in* (34) *are related via duality of subgradients (as defined in* (31) *and* (32); *see also Figure* 2):

$$T \in \partial \hat{h}(\alpha) \Leftrightarrow -\alpha \in \partial h(T) \Leftrightarrow \hat{h}(\alpha) = \alpha T + h(T). \tag{35}$$

These conditions are equivalent to

$$\Phi_c(T) \leq 1 - \alpha \leq \Phi_c(T + 0) \tag{36}$$

where Φ_c is the convex upperbound of the distribution of $R(\xi)$ as defined in (23). *Also, the slope of $\hat{h}$ in $\alpha = 0+$ is T_0 and in $\alpha = 1-$ it is T_1; and $\hat{h}(0) = 0$, $\hat{h}(1) = T_c$.*

Proof. From (22) it follows that

$$h(T) = \min_{y,z} \left\{ y + G(z) : \begin{bmatrix} -y \leq 0 \\ -y + \sum_{i \in B_j} z_i \leq T, j = 1, \ldots, p \\ y \in \mathbb{R}, z \in \mathbb{R}^n \end{bmatrix} \right\}. \tag{37}$$

This convex program is stable (the optimal value stays finite under perturbation of the right hand sides), therefore $h(T)$ is also the optimal value of the dual of (37),

$$h(T) = \max_{\lambda \in \mathbb{R}^{p+1}} \left\{ \inf_{(y,z) \in \mathbb{R}^{n+1}} L(y, z; \lambda) : \lambda \geq 0 \right\}$$

where the Lagrangian function is defined by

$$L(y, z; \lambda) = y + G(z) - T \sum_{j=1}^{p} \lambda_j - y \sum_{j=0}^{p} \lambda_j + \sum_{j=1}^{p} \lambda_j \sum_{i \in B_j} z_i. \tag{38}$$

The dual objective function is $-\infty$ if $\sum_{j=0}^{p} \lambda_j \neq 1$. If $\sum_{j=0}^{p} \lambda_j = 1$ then it is equal to $-\pi_0 T + \hat{G}(\pi)$ (see (26)) for $\pi_0 := \sum_{j=1}^{p} \lambda_j$ and $\pi_i = \sum_j \{\lambda_j : i \in B_j\}$, $i = 1, \ldots, n$. If we interpret λ_j as a flow through path j then π_i is the flow in arc i, $i = 1, \ldots, n$, and π_0 is the total flow from source to sink. Consequently, each $\lambda \in \mathbb{R}^{p+1}$ with $\lambda \geq 0$ and $\sum_{j=0}^{p} \lambda_j = 1$ determines a feasible circulation in the augmented network, with $\pi_0 \leq 1$. On the other hand, each circulation of this type can be decomposed into flows through paths, so that we might eliminate λ for π, leading to the formula in $\tilde{P}(T)$.

By definition we have (33); using (22) we get

$$\hat{h}(\alpha) = \min_{T,z} \{\alpha T + [R(z) - T]^+ + G(z)\}, \quad 0 \leq \alpha \leq 1, \tag{39}$$

so that without loss $T = R(z)$ and

$$PD(\alpha): \quad \hat{h}(\alpha) = \min_{z} \{\alpha R(z) + G(z)\}.$$

Completely similarly to the derivation of the formula in $\tilde{P}(T)$ from (37) the formula in $DD(\alpha)$ follows from $PD(\alpha)$. Since h behaves like G_i with F_i replaced by Φ_c, the remaining statements follow directly from Propositions 7 and 9. □

Corollary 11. *If $T_1 \leq T \leq T_0$ then each optimal solution z^* for $\tilde{D}(T)$ satisfies $R(z^*) = T$.*

Proof. If z^* solves $\tilde{D}(T)$, then $y := [R(z^*) - T]^+$ and $z := z^*$ solve the stable convex program

$$\min_{y,z} \{y + G(z): R(z) - y \leq T, -y \leq 0\}.$$

Let α and β be multipliers for the constraints, then the dual problem is

$$\max_{\alpha,\beta \geq 0} \inf_{y,z} \{-\alpha T + y(1 - \alpha - \beta) + G(z) + \alpha R(z)\}$$

so that $\beta = 1 - \alpha$ necessarily, and with $PD(\alpha)$ we get

$$\max_{0 \leq \alpha \leq 1} \left\{ -\alpha T + \min_z \{\alpha R(z) + G(z)\} \right\} = \max_{0 \leq \alpha \leq 1} \{-\alpha T + \hat{h}(\alpha)\}.$$

Optimal values for α and β are therefore determined by $T \in \partial \hat{h}(\alpha^*)$, that is $-\alpha^* \in \partial h(T)$, see (35), and $\beta^* = 1 - \alpha^*$. Consequently $\alpha^* > 0$ and $\beta^* > 0$ if $T_1 < T < T_0$, and the complementarity conditions imply $R(z^*) = T$. If $T = T_1$ or $T = T_0$ this result was already proved in Corollary 8. □

For $h(T)$ we have a 'project planning representation' in $\tilde{D}(T)$ and a 'flow representation' in $\tilde{P}(T)$ which are dual to each other. By reformulation of the Kuhn–Tucker conditions we get

Theorem 12. *Duality of $\tilde{P}(T)$ and $\tilde{D}(T)$, $T \in \mathbb{R}$.*

The circulation $\pi^* \in \Pi$ with $\pi_0^* \leq 1$ is optimal for $\tilde{P}(T)$ and $z^* \in \mathbb{R}^n$ is optimal for $\tilde{D}(T)$ iff the following conditions are true.

(a) (30) *holds for* $i = 1, \ldots, n$.

(b) *If* $\pi_i^* > 0$ *then* $i \in \mathrm{CA}(z^*)$, $i = 1, \ldots, n$.

(c) *If* $\pi_0^* > 0$ *then* $R(z^*) \geq T$, *if* $\pi_0^* < 1$ *then* $R(z^*) \leq T$.

For any π_0^*, $-\pi_0^* \in \partial h(T)$.

Proof. The Lagrangian L defined in (38) has a saddle point $(y^*, z^*; \lambda^*)$ iff the following conditions hold.

(i) Primal feasibility (and optimality): $y^* = [R(z^*) - T]^+$.

(ii) Dual feasibility: $\lambda^* \geq 0$, $\sum_{j=0}^{p} \lambda_j^* = 1$.

(iii) $0 \in \partial_{y,z} L(y^*, z^*; \lambda^*)$: (29) hence (30) holds with $\pi_i^* := \sum_j \{\lambda_j^*: i \in B_j\}$, $i = 1, \ldots, n$.

(iv) Complementary slackness: $\lambda_0^* y^* = 0$, $\lambda_j^* (\sum_{i \in B_j} z_i^* - y^* - T) = 0$, $j = 1, \ldots, p$, or, equivalently:

if $\lambda_0^* > 0$ then $R(z^*) \leq T$, and
if $\lambda_j^* > 0$ for a $j \geq 1$ then $R(z^*) \geq T$ and $j \in \mathrm{CP}(z^*)$.

Corresponding to every λ^* satisfying (ii) a $\pi^* \in \Pi$ with $\pi_0^* \leq 1$ can be defined by $\pi_i^* := \sum_j \{\lambda_j^* : i \in B_j\}$, $i = 1, \ldots, n$, and $\pi_0^* := \sum_{j=1}^{p} \lambda_j^* = 1 - \lambda_0^*$. Reversely, each $\pi^* \in \Pi$ with $\pi_0^* \leq 1$ can be decomposed (not uniquely, generally) into flows $\lambda_j^* \geq 0$ in paths $j = 1, \ldots, p$, and the same relations hold. It is an elementary exercise to show that these $\lambda \mapsto \pi$ transformations make (i)-(iv) equivalent to $a - c$. Use has to be made of: $j \in \mathrm{CP}(z^*) \Leftrightarrow i \in \mathrm{CA}(z^*)$ for all $i \in B_j$. Finally, if $\pi^* \in \Pi$ solves $\tilde{P}(T)$, then $\alpha := \pi_0^*$ solves

$$\max_{0 \leq \alpha \leq 1} \left\{ -\alpha T + \max_{\pi \in \Pi} \{\hat{G}(\pi) : \pi_0 = \alpha\} \right\} = \max_{0 \leq \alpha \leq 1} \{-\alpha T + \hat{h}(\alpha)\}$$

so that $-\pi_0^* \in \partial h(T)$, see (35). □

Similar relations exist between the solutions of $PD(\alpha)$ and $DD(\alpha)$.

Theorem 13. *Duality of $PD(\alpha)$ and $DD(\alpha)$, $0 \leq \alpha \leq 1$. The circulation $\pi^* \in \Pi$ with $\pi_0^* = \alpha$ is optimal for $DD(\alpha)$ and z^* is optimal for $DD(\alpha)$ iff the following conditions are true.*

(a) (30) *holds for* $i = 1, \ldots, n$.
(b) *If* $\pi_i^* > 0$ *then* $i \in \mathrm{CA}(z^*)$, $i = 1, \ldots, n$.

For any such z^*, $T := R(z^*)$ solves $PP(\alpha)$. All optimal solutions T_α^* of PP_α satisfy $T_\alpha^* \in \partial \hat{h}(\alpha)$, so that $T_1 < T_\alpha^* < T_0$ if $0 < \alpha < 1$.

Proof. The first statement is proved similarly as in Theorem 12. The last claims have been shown before (after Corollary 8, and (39)). □

We now shall characterize the worst-case distribution(s) H^*, the optimal solution(s) of $P(T)$. From Theorem 1 and 6 it is clear that the solutions z^* of $\tilde{D}(T)$ provide information on the support of H^*. Moreover, it is not surprising, that the optimal flows π^* in $\tilde{P}(T)$, which problem is the dual of the reduced dual of $P(T)$, can be interpreted in terms of H^*: it will appear that the flows are criticality numbers based on the distribution H^*. Define the following orthants in $\mathbb{R}^n$:

$$K_0 := \{x \in \mathbb{R}^n : x \leq 0\}, \qquad K_j := \{x \in \mathbb{R}^n : x_i \geq 0 \text{ if } i \in B_j \text{ and } x_i \leq 0 \text{ if } i \notin B_j\}$$

for $j = 1, \ldots, p$

and also the index sets

$$\mathrm{CP}_T(z) := \{0\} \qquad \text{if } R(z) < T,$$

$$\mathrm{CP}_T(z) := \{0\} \cup \mathrm{CP}(z) \qquad \text{if } R(z) = T,$$

$$\mathrm{CP}_T(z) := \mathrm{CP}(z) \qquad \text{if } R(z) > T.$$

Lemma 14. *For all* $x, z \in \mathbb{R}^n$

$$R(x) \leq R(z) + \sum_{i=1}^{n} [x_i - z_i]^+. \tag{40}$$

Equality holds iff $x - z \in \bigcup_{j \in CP(z)} K_j$. *If* $x - z \in K_j$ *and* $j \in CP(z)$ *then* $j \in CP(x)$ *too.*

Proof. The inequality (40) follows from

$$\sum_{i \in B_j} x_i \geqq \sum_{i \in B_j} z_i + \sum_{i=1}^{n} [x_i - z_i]^+ \quad \forall x \; \forall z \; \forall j. \tag{41}$$

In (41) equality holds iff $x - z \in K_j$. Therefore, $j \in CP(z)$ and $x - z \in K_j$ imply equality in (40), together with $j \in CP(x)$. On the other hand, if equality holds in (40) then for any $j \in CP(x)$

$$\sum_{i \in B_j} x_i = R(x) = R(z) + \sum_{i=1}^{n} [x_i - z_i]^+ \geqslant \sum_{i \in B_j} z_i + \sum_{i=1}^{n} [x_i - z_i]^+ \tag{42}$$

so that equality must be true in (41) and (42). Hence $x - z \in K_j$ and $j \in CP(z)$. □

Theorem 15. *In order that* $H^* \in \mathcal{H}(F_1, \ldots, F_n)$ *is optimal for* $P(T)$ *it is necessary and sufficient that its support is contained in* $X_T^*(z^*)$, *with*

$$X_T^*(z) := z + \bigcup_j \{K_j : j \in CP_T(z)\}, \tag{43}$$

where z^* *is any optimal solution of* $\tilde{D}(T)$.

Proof. From Theorems 1 and 6 it is clear that it is sufficient to prove that $X_T^*(z)$ is the same set as

$$W_T(z) := \left\{ x \in \mathbb{R}^n : [R(x) - T]^+ = [R(z) - T]^+ + \sum_{i=1}^{n} [x_i - z_i]^+ \right\}. \tag{44}$$

First we shall show that $X_T^*(z) \subset W_T(z)$. If $R(z) \leq T$ and $x \in z + K_0$ this is trivial since $x \leq z$ implies $R(x) \leq R(z) \leq T$ so that each term in (44) vanishes. If $R(z) \geqslant T$, $j \in CP(z)$ and $x \in z + K_j$ then

$$R(x) = R(z) + \sum_{i=1}^{n} [x_i - z_i]^+$$

(Lemma 14), so that

$$R(x) \geqslant T \text{ and } [R(x) - T]^+ = [R(z) - T]^+ + \sum_{i=1}^{n} [x_i - z_i]^+.$$

In order to prove the reverse inclusion, let $x \in W_T(z)$. Suppose first that $R(z) < T$. Then $[R(x) - T]^+ = \sum_{i=1}^{n} [x_i - z_i]^+$. From (40) we know however, that

$$R(x) \leq R(z) + \sum_{i=1}^{n} [x_i - z_i]^+ < T + \sum_{i=1}^{n} [x_i - z_i]^+.$$

Consequently, $R(x)-T<0$ and $\sum_{i=1}^{n}[x_i-z_i]^+=0$ so that $x\in z+K_0$. Suppose now that $R(z)=T$. If $x-z\in K_0$ we have nothing to prove, so assume $\sum_{i=1}^{n}[x_i-z_i]^+>0$. Since $x\in W_T(z)$ it follows that $R(x)\geq T$, and that equality must hold in (40), so that $x\in z+\bigcup_j\{K_j: j\in \mathrm{CP}(z)\}$. Finally suppose $R(z)>T$. Then $x\in W_T(z)$ implies also equality in (40), since now $[R(x)-T]^+=R(x)-T$ because of $[R(x)-T]^+\geq [R(z)-T]^+>0$. □

The structure of the support of a worst-case distribution H^* shows the dependence of the random variables $\xi_1,\ldots,\xi_n$ under H^*. For example, in the most interesting case ($T_1\leq T\leq T_0$) we have $T=R(z^*)$ and (43) indicates that with probability 1 there are only two possibilities for the realisations of ξ in relation to the promises z^*:

either all activities are ready in time: $\xi_i\leq z_i^*\ \forall i$ and $R(\xi)\leq T$,

or the project has a delay due to lateness of the activities of *one* critical subpath: $\exists j\in \mathrm{CP}(z^*)$, not necessarily unique, such that $\xi_i\leq z_i^*\ \forall i\notin B_j$, whereas $\xi_i\geq z_i^*\ \forall i\in B_j$ (and strict inequality holds at least for one $i_0\in B_j$) so that $j\in \mathrm{CP}(\xi)$ and $R(\xi)>T$.

Theorem 15 indicates that the worst-case distributions H^* might be constructed by conditioning upon $z^*+K_j, j\in \mathrm{CP}_T(z^*)$. Any probability distribution with support in $z^*+\bigcup_{j=0}^{p}K_j$ may be decomposed into

$$H(x)=\sum_{j=0}^{p}\lambda_j\cdot H_j(x),\quad x\in\mathbb{R}^n, \tag{45}$$

with $\lambda_j\geq 0$, $\sum_{j=0}^{p}\lambda_j=1$, where H_j is a probability distribution with support in z^*+K_j. The question arises: how to define λ_j and H_j in order that H solves $P(T)$? We will show that the conditions on λ_j can by characterized by means of the optimal solution(s) π^* for $\tilde{P}(T)$. The conditions on the H_j appear to be only conditions on their marginals. For the formulation of these conditions, we need two conditional distributions derived from F_i with use of π_i^* and z_i^*, $i=1,\ldots,n$, where π^* solves $\tilde{P}(T)$ and z^* solves $\tilde{D}(T)$. If $z_i^*\leq a_i$, then

$$\begin{aligned} G_i^-(x_i) &:= 0 \quad \text{if } x_i\leq z_i^*,\\ &:= 1 \quad \text{if } x_i>z_i^*. \end{aligned} \tag{46}$$

If $z_i^*>a_i$ (hence $0<F_i(z_i^*)\leq 1-\pi_i^*\leq F_i(z_i^*+0)$) we define for $x_i\in\mathbb{R}$

$$G_i^-(x_i):=\beta_i\cdot P_{F_i}(\xi_i<x_i|\xi_i<z_i^*)+(1-\beta_i)\cdot P_{F_i}(\xi_i<x_i|\xi_i\leq z_i^*) \tag{47}$$

where $\beta_i\in[0,1]$ satisfies

$$\frac{1}{1-\pi_i^*}=\frac{\beta_i}{F_i(z_i^*)}+\frac{1-\beta_i}{F_i(z_i^*+0)}. \tag{48}$$

Clearly, if $F_i(\{z_i^*\}) = 0$ then the value of β_i does not matter, but if $F_i(\{z_i^*\}) > 0$ we must have

$$\beta_i = \frac{F_i(z_i^*)}{1-\pi_i^*} \cdot \frac{F_i(z_i^*+0)-(1-\pi_i^*)}{F_i(\{z_i^*\})}.$$

It is easy to show that in all cases G_i^- is a probability distribution function with support in $(-\infty, z_i^*]$. If $z_i^* \geq b_i$ we define

$$\begin{aligned} G_i^+(x_i) &:= 0 \quad \text{if } x_i \leq z_i^*, \\ &:= 1 \quad \text{if } x_i > z_i^*. \end{aligned} \tag{49}$$

If $z_i^* < b_i$ (hence $F_i(z_i^*) \leq 1-\pi_i^* \leq F_i(z_i^*+0) < 1$) we define for $x_i \in \mathbb{R}$

$$G_i^+(x_i) := \gamma_i \cdot P_{F_i}(\xi_i < x_i | \xi_i \geq z_i^*) + (1-\gamma_i) P_{F_i}(\xi_i < x_i | \xi_i > z_i^*) \tag{50}$$

where $\gamma_i \in [0, 1]$ satisfies

$$\frac{1}{\pi_i^*} = \frac{\gamma_i}{1-F_i(z_i^*)} + \frac{1-\gamma_i}{1-F_i(z_i^*+0)}. \tag{51}$$

Again, if $F_i(\{z_i^*\}) = 0$ the value of γ_i does not matter, but if $F_i(\{z_i^*\}) > 0$ we must have

$$\gamma_i = \frac{1-F_i(z_i^*)}{\pi_i^*} \cdot \frac{F_i(z_i^*+0)-(1-\pi_i^*)}{F_i(\{z_i^*\})}.$$

In all cases, G_i^+ is a probability distribution function with support in $[z_i^*, \infty)$. Finally, define for $i = 1, \ldots, n$, $B_i^{-1} := \{j: 0 \leq j \leq p,\ i \in B_j\}$ with $B_0 := \emptyset$. Hence $0 \notin B_i^{-1}$ for all i; $B_i^{-1} \neq \emptyset$ since $\bigcup_{j=1}^p B_j = \{1, \ldots, n\}$.

Theorem 16. *Let z^* be optimal for $\tilde{D}(T)$.*

(a) *Let H be optimal for $P(T)$, and let $\lambda_j, j = 0, 1, \ldots, p$, be given in its representation (45). Then $\pi \in \mathbb{R}^{n+1}$, defined by*

$$\pi_0 := 1 - \lambda_0 = \sum_{j=1}^{p} \lambda_j, \qquad \pi_i := \sum_{j \in B_i^{-1}} \lambda_j, \quad i = 1, \ldots, n, \tag{52}$$

is optimal for $\tilde{P}(T)$.

(b) *Let π^* be optimal for $\tilde{P}(T)$. Then H, defined by (45) and (52) with $\pi = \pi^*$, is optimal for $P(T)$, if $H_j \in \mathcal{H}(G_1^j, \ldots, G_n^j)$, $j = 0, 1, \ldots, p$, where*

$$G_i^j := G_i^- \text{ if } i \notin B_j \quad \text{and} \quad G_i^j := G_i^+ \text{ if } i \in B_j.$$

Proof. (a) Let H solve $P(T)$. From Theorem 15 it follows that without loss we may take $\lambda_j = 0$ if $j \notin \mathrm{CP}_T(z^*)$. We shall show that π defined in (52) satisfies the conditions of Theorem 12. It is obvious that π is a feasible circulation in the augmented network, with $\pi_0 \leq 1$. If $\pi_0 < 1$ then $\lambda_0 > 0$ hence $0 \in \mathrm{CP}_T(z^*)$ so that $R(z^*) \geq T$. If $\pi_0 > 0$ then $\exists j \geq 1$ with $\lambda_j > 0$, so that $j \in \mathrm{CP}_T(z^*)$ and $R(z^*) \leq T$ follows. In the same way, if $\pi_i > 0$ for an $i \geq 1$ then $\exists j \geq 1$ with $i \in B_j$, $\lambda_j > 0$ so that

$j \in \mathrm{CP}(z^*)$ and $i \in \mathrm{CA}(z^*)$. (30) follows from the fact that F_i must be the i-th marginal of H. Indeed,

$$F_i(z_i^*) = P_H(\xi_i < z_i^*) = \sum_j \lambda_j P_{H_j}(\xi_i < z_i^*) \leq \sum_{j \notin B_i^{-1}} \lambda_j = 1 - \pi_i \tag{53}$$

where the inequality is a consequence of $P_{H_j}(z^* + K_j) = 1$ so that

$$\begin{aligned} P_{H_j}(\xi_i < z_i^*) &= 0 \quad \text{if } i \in B_j, \\ &\leq 1 \quad \text{if } i \notin B_j. \end{aligned} \tag{54}$$

Similarly

$$\begin{aligned} P_{H_j}(\xi_i \leq z_i^*) &\geq 0 \quad \text{if } i \in B_j, \\ &= 1 \quad \text{if } i \notin B_j, \end{aligned} \tag{55}$$

so that

$$F_i(z_i^* + 0) = P_H(\xi_i \leq z_i^*) = \sum_j \lambda_j P_{H_j}(\xi_i \leq z_i^*) \geq \sum_{j \notin B_i^{-1}} \lambda_j = 1 - \pi_i. \tag{56}$$

From (53) and (56) follows (30), which completes the proof.

(b) Because of Theorem 15 we have to show that H has the correct support and the correct marginals. Its support is contained in $X_T^*(z^*)$, see Theorem 12, since $\lambda_j > 0$ for a $j \geq 1$ implies $R(z^*) \leq T$ and $\pi_i^* > 0 \, \forall i \in B_j$. Hence $i \in \mathrm{CA}(z^*)$ for all $i \in B_j$, so that $j \in \mathrm{CP}(z^*)$, and $\lambda_0 > 0$ implies $\pi_0^* < 1$ so that $R(z^*) \geq T$. In order to show that F_i is the i-th marginal of H, it is sufficient to prove

$$P_H(\xi_i < x_i) = F_i(x_i) \qquad \text{for all } x_i < z_i^*, \tag{57}$$

$$P_H(\xi_i > x_i) = 1 - F_i(x_i + 0) \quad \text{for all } x_i > z_i^*. \tag{58}$$

Since $P_{H_j}(\xi_i < x_i) = 0$ if $x_i < z_i^*$ and $i \in B_j$, we have for $x_i < z_i^*$

$$P_H(\xi_i < x_i) = \sum_{j \notin B_i^{-1}} \lambda_j \cdot P_{H_j}(\xi_i < x_i) = (1 - \pi_i^*) P_{G_i^-}(\xi_i < x_i).$$

If $z_i^* \leq a_i$ the right-hand side vanishes, so that (57) holds in that case. Suppose now that $z_i^* > a_i$. Then from (47) and (48)

$$P_{G_i^-}(\xi_i < x_i) = \left(\frac{\beta_i}{F_i(z_i^*)} + \frac{1 - \beta_i}{F_i(z_i^* + 0)} \right) F_i(x_i) = \frac{1}{1 - \pi_i^*} F_i(x_i)$$

so that (57) is also true if $z_i^* > a_i$. Similarly, (58) follows from

$$P_H(\xi_i > x_i) = \sum_{j \notin B_i^{-1}} \lambda_j \cdot P_{H_j}(\xi_i > x_i) = \pi_i^* P_{G_i^+}(\xi_i > x_i), \; x_i > z_i^*,$$

using (49), (50) and (51). □

In general many worst-case distributions exist for $P(T)$, T fixed. Even if $\tilde{D}(T)$ and $\tilde{P}(T)$ have unique solutions (as occurs e.g. for $T_1 < T < T_0$ if $\hat{G}$ is strictly concave and differentiable; in particular if each marginal distribution has a positive

density on (a_i, b_i)) the corresponding λ_j are not unique, generally. Depending on the structure of the network one has no/little/ample freedom in decomposing a feasible circulation into flows through paths. Moreover, the freedom of choice of H_j in (45) is enormous. Although no attempt has been made to characterize all H_j, it is clear from Theorem 16 that it is sufficient to prescribe only their marginals. In fact, Theorem 16b specifies for each relevant j a distribution problem on $z^* + K_j$, comparable with P' or P in Section 2, with the difference that each feasible solution is good: one may take the independent solution, the NW Rule Solution, etc. The distinction between different worst-case distributions does not seem to be very important, since the relevant information is contained in π and λ which have an interpretation as criticality numbers. This is most easily seen for continuous distributions F_i as shown in the next theorem.

Theorem 17. *Suppose that F_i has a positive density function on $[a_i, b_i]$, for each $i = 1, \ldots, n$. Then $\tilde{P}(T)$ has a unique solution $\pi^* = \pi^*_T$ for $T_1 \leq T \leq T_0$, and each optimal solution $H^* = H^*_T$ for $P(T)$ satisfies then:*

$$P_{H^*}(R(\xi) > T) = \pi_0^* = 1 - \lambda_0^*, \tag{59}$$

$$P_{H^*}(R(\xi) > T, i \in \mathrm{CA}(\xi)) = \pi_i^*, \quad i = 1, \ldots, n, \tag{60}$$

$$P_{H^*}(R(\xi) > T, j \in \mathrm{CP}(\xi)) = \lambda_j^*, \quad j = 1, \ldots, p, \tag{61}$$

where $\lambda^* \in \Lambda := \{\lambda \in \mathbb{R}^{p+1}: \lambda \geq 0, \sum_{j=0}^{p} \lambda_j = 1\}$ is related to π^* by (52); λ^* may depend on H^*. In fact, for each $\lambda \in \Lambda$ satisfying (52) there exist corresponding worst-case distributions H^*.

Proof. Consider $\tilde{D}(T)$ for $T_1 \leq T \leq T_0$. Since necessarily $R(z^*) \leq T$ (Corollary 11), the set of optimal solutions of $\tilde{D}(T)$ can be written as

$$Z^* := \arg\min_{z \in \mathbb{R}^n} \{G(z): R(z) \leq T\}.$$

Obviously, $Z_i^* := \{u \in \mathbb{R}: \exists z^* \in Z^* \text{ with } z_i^* = u\}$ is a closed convex set, $i = 1, \ldots, n$. Since F_i is strictly increasing on $[a_i, b_i]$, Z_i^* can only be more than a singleton if it is contained in either $(-\infty, a_i]$ or $[b_i, \infty)$. In all cases, π_i^* is unique, since F_i is continuous (see (30)). Of coursé, uniqueness of π_i^*, $i = 1, \ldots, n$, implies uniqueness of π_0^*. In order to show (59)–(61), represent any worst-case distribution H^* by (45), with $\lambda := \lambda^* \in \Lambda$ satisfying (52). Since $R(z^*) = T$,

$$z^* + \operatorname{int} K_0 \subset \{x \in X_T(z^*): R(x) < T\} \subset z^* + K_0,$$

$$z^* + \operatorname{int} K_j \subset \{x \in X_T(z^*): R(x) > T, \{j\} = \mathrm{CP}(x)\} \subset z^* + K_j, \quad j \geq 1,$$

the events $\{R(x) = T\}$ and $\{|\mathrm{CP}(x)| \geq 2\}$ are contained in the (partly common) boundaries of the translated orthants, which have zero probability under H^* since

$F_i(\{z_i^*\}) = 0 \; \forall i$. Therefore,

$$P_{H^*}(R(\xi) < T) = P_{H^*}(R(\xi) \leq T) = P_{H^*}(z^* + K_0) = \lambda_0^*,$$

$$P_{H^*}(R(\xi) > T, j \in \mathrm{CP}(\xi)) = P_{H^*}(R(\xi) > T, \{j\} = \mathrm{CP}(\xi))$$
$$= P_{H^*}(z^* + K_j) = \lambda_j^*, \quad j \geq 1,$$

$$P_{H^*}(R(\xi) > T, i \in \mathrm{CA}(\xi)) = \sum_{j \in B_i^{-1}} P_{H^*}(R(\xi) > T, j \in \mathrm{CP}(\xi))$$
$$= \sum_{j \in B_i^{-1}} \lambda_j = \pi_i^*. \qquad \square$$

From the proof it is clear that (59)-(61) also hold for discrete marginals F_i, if z^* can be chosen such that $F_i(\{z_i^*\}) = 0 \; \forall i$; but this is not possible in general. If $F_i(\{z_i^*\}) > 0$ for some i the analysis is more difficult since the common boundaries of the translated orthants get positive probability. If in (45) the representation is changed into

$$H := \sum_j \{\hat{\lambda}_j \cdot \hat{H}_j : j \in \mathrm{CP}_T(z^*)\}$$

with $P_{\hat{H}_j}(z^* + \hat{K}_j) = 1$, where $\hat{K}_j \subset K_j$, $\bigcup_j \hat{K}_j = \bigcup_j K_j$, and $\hat{K}_j \cap \hat{K}_h = \emptyset$ for $j \neq h$, then it is not difficult to show, that (59)-(61) generalize to

$$P_{H^*}(R(\xi) > T) \leq \hat{\pi}_0 = 1 - \hat{\lambda}_0 \leq P_{H^*}(R(\xi) \geq T),$$

$$P_{H^*}(R(\xi) > T, \{j\} = \mathrm{CP}(\xi)) \leq \hat{\lambda}_j \leq P_{H^*}(R(\xi) \geq T, j \in \mathrm{CP}(\xi)), \quad j \geq 1,$$

$$P_{H^*}(R(\xi) > T, i \in \mathrm{CA}(\xi), |\mathrm{CP}(\xi)| = 1) \leq \hat{\pi}_i \leq P_{H^*}(R(\xi) \geq T, i \in \mathrm{CA}(\xi)),$$
$$i \geq 1.$$

As a final remark, we notice that the characterization of the optimal solution of $P(T)$ is at the same time applicable to problem $PP(\alpha)$; just substitute $T := T_\alpha^*$.

4. Worst-case marginals

In the previous section we assumed that the marginal distributions F_i are known completely. From a practical point of view this assumption is not realistic. In this section we shall show that the results on worst-case joint distributions can be combined in a natural way with a minimax approach to the marginals, too. Then the problem $P(T)$ is replaced by

$$h_{\mathcal{F}}(T) = \sup\left\{ \sup_{H \in \mathcal{H}(F_1, \ldots, F_n)} Ef(\xi) : F_i \in \mathcal{F}_i, \; i = 1, \ldots, n \right\},$$

where $\mathcal{F}_i$ is the class of all distribution functions F_i compatible with the partial

information. From the previous section we know, that

$$h_{\mathscr{F}}(T) = \sup_{\mathscr{F}_i} \min_{z \in \mathbb{R}^n} \left(f(z) + \sum_{i=1}^{n} E_{F_i}[\xi_i - z_i]^+ \right).$$

If $\mathscr{F}_i$ is such, that an $F_i^* \in \mathscr{F}_i$ exists, for which

$$\max_{F_i \in \mathscr{F}_i} E_{F_i}[\xi_i - z_i]^+ = E_{F_i^*}[\xi_i - z_i]^+ \quad \textit{for all } z_i \in \mathbb{R}^n,\ i = 1, \ldots, n,$$

then $\sup_{\mathscr{F}_i}$ and $\min_z$ may be interchanged. In that case the analysis of Section 3 is applicable directly, by taking for F_i the *worst-case marginal distribution* F_i^*.

Examples. If $\mathscr{F}_i$ is the class of all distributions with given bounded range $[a_i, b_i]$, then $F_i^*(\{b_i\}) = 1$, since $r(x) := [x - z]^+$ is nondecreasing. Less trivial is the situation, if also the mean value $\bar{\xi}_i$ is known; then the convexity of r implies that F_i^* is the two-point distribution (see e.g. [40, 4])

$$F_i^*(\{a_i\}) := \frac{b_i - \bar{\xi}_i}{b_i - a_i}, \qquad F_i^*(\{b_i\}) := \frac{\bar{\xi}_i - a_i}{b_i - a_i}. \tag{62}$$

The same author shows, that if $\mathscr{F}_i$ is the class of all symmetrical unimodular distributions on $[a_i, b_i]$, then F_i^* is the uniform distribution on $[a_i, b_i]$. If $\mathscr{F}_i$ is the class of all unimodular distributions on $[a_i, b_i]$ with a given mode m_i then F_i^* is the uniform distribution on $[m_i, b_i]$ (see [2]). This last specification of $\mathscr{F}_i$ is the most interesting, since it contains precisely the information which usually is supposed to be known in PERT networks.

5. Numerical calculations

The convex, polyhedral function $f(z) = [R(z) - T]^+$ can be represented as the solution of a LP program (as in (37), or in a more common dynamic formulation in terms of earliest times for all nodes). Therefore, $\tilde{D}(T)$ and $PD(\alpha)$ are simple recourse models for which good algorithms exist. For example, if all F_i are discrete then also each G_i can be put into a LP form, so that $\tilde{D}(T)$ and $PD(\alpha)$ are LP programs. Of course, also the original programs $P(T)$ and $PP(\alpha)$ are then finite LP problems but their size is tremendously larger. If the F_i are uniform distributions, $\tilde{D}(T)$ and $PD(\alpha)$ are quadratic with linear constraints; in the general case the objective $G(z)$ is separable convex. Similar statements hold for the flow formulations $\tilde{P}(T)$ and $DD(\alpha)$; e.g.

$$\hat{G}_i(\pi_i) = \max_{y} \left\{ \sum_{j=1}^{m} t_j y_j : \sum_{j=1}^{m} y_j = \pi_i,\ 0 \leq y_j \leq p_j\ \forall j \right\}$$

if F_i is the discrete distribution $F_i(\{t_j\}) = p_j$, $j = 1, \ldots, m$, with $t_1 > t_2 > \cdots > t_m$ and $\sum_{j=1}^{m} p_j = 1$, $p_j > 0$. Moreover, the flow structure may be exploited to devise special algorithms (see [3]). In fact, in [3] a simple recourse model, which is slightly more

general than $\tilde{D}(T)$, is taken as starting-point, and a well-known network flow algorithm is adjusted to solve the problem with discrete distributions.

In order to get an idea about the influence of alternative specifications the problem $DD(\alpha)$ has been solved for the project network of Figure 3. This network is a slight modification of the Electronic Module Development Project as described in [26, p. 216]. Apart from the network structure, the only information on the duration of activity i is given by a_i, m_i and b_i (see Figure 3), where $[a_i, b_i]$ is the support and m_i the mode of F_i. This is the 'standard' information needed for PERT. According to this heuristic, the distribution of the project completion time can be approximated by the normal distribution $N(\mu, \sigma^2)$, where $\mu := \sum_{i \in B_j} \bar{\xi}_i$, $\sigma^2 := \sum_{i \in B_j} \sigma_i^2$ for a $j \in \mathrm{CP}(\bar{\xi})$, with

$$\bar{\xi}_i := \frac{a_i + 4m_i + b_i}{6} \text{ and } \sigma_i := \frac{b_i - a_i}{6} \quad \text{for all } i.$$

In the example $\mu = 47.7$ and $\sigma = 3.3$. In spite of its charming simplicity it is not surprising that the PERT approach has been criticized for its dubious probabilistic justification. It seems to us that the minimax approach as worked out in Section 3 is an alternative. One might solve $P(T)$ (via $\tilde{D}(T)$ or $\tilde{P}(T)$) for one or more relevant values of T, or solve $PP(\alpha)$ (via $DD(\alpha)$ or $PD(\alpha)$) for one or more relevant values of α, or solve $P(T)$ or $PP(\alpha)$ for 'all' values in order to get the convex upperbound Φ_c. We solved $PP(\alpha)$ for a range of α-values, under four assumptions on the marginals:

A. F_i is triangular on $[a_i, b_i]$, with mode m_i.
B. F_i is uniform on $[m_i, b_i]$.
C. F_i is restricted to $\{a_i, b_i\}$, with mean value m_i, see (62).
D. F_i is restricted to $\{a_i, b_i, m_i\}$, with mean value m_i and $F_i(\{m_i\}) = \frac{1}{2}$.

Distribution A seems to be a reasonable distribution compatible with the underlying interpretation of the data $[a_i, m_i, b_i]$, whereas B is the worst-case unimodal distribution. In C and D the value for m_i is interpreted as the mean value $\bar{\xi}_i$ of ξ_i rather

Table 1

The optimal solutions T_α^* and the corresponding optimal values $\hat{h}(\alpha)$ for $PP(\alpha)$ for different distributions and α-values; $T_1 = T_1^*$, $T_c = \hat{h}(1)$, $T_0 = T_0^*$

	$\alpha = 1.0$	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
T_α^*											
A	46.8	49.3	50.9	52.4	54.0	55.6	57.4	59.4	61.9	65.3	76.0
B	58.0	59.4	60.7	62.2	63.6	65.1	66.5	68.2	70.2	72.4	76.0
C	41.4	45.4	46.0	46.0	51.0	56.0	56.0	71.0	71.0	76.0	76.0
D	42.1	42.1	44.1	44.1	48.6	49.6	52.0	55.0	57.0	71.0	76.0
$\hat{h}_\alpha$											
A	56.5	51.7	46.7	41.5	36.2	30.7	25.1	19.3	13.2	6.9	0.0
B	66.0	59.7	53.7	47.5	41.3	34.8	28.2	21.5	14.6	7.4	0.0
C	57.0	53.3	48.7	44.1	39.3	34.0	28.4	22.2	15.1	7.6	0.0
D	52.0	48.6	44.3	39.9	35.2	30.3	25.0	19.8	14.2	7.6	0.0

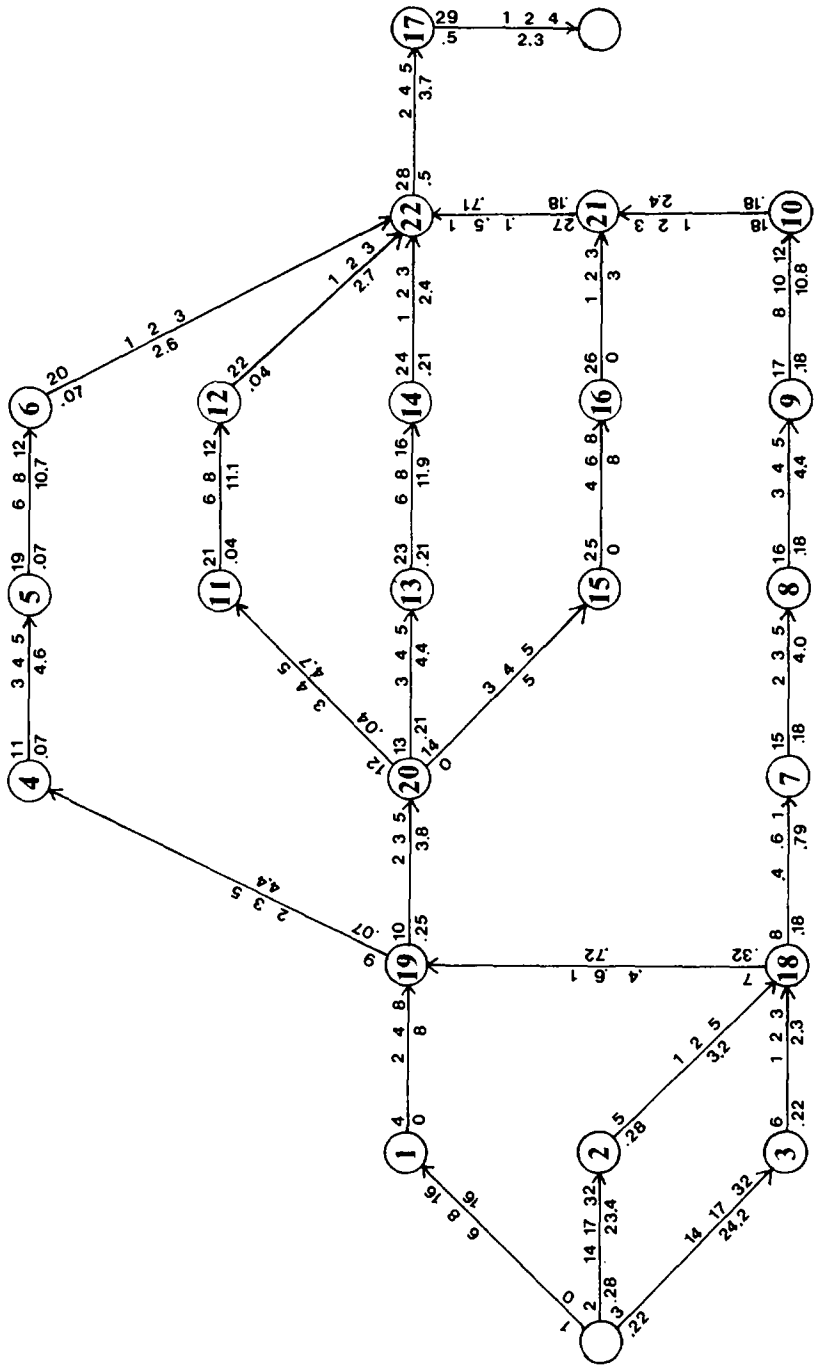


Fig. 3. The Electronic Module Development Project. Above each activity i the data i, a_i, m_i and b_i are given; underneath the solutions π_i^* and z_i^* for the triangular distribution A and $\alpha = 0.5$ are shown.

than its mode; C is the worst-case marginal for that situation, whereas D is a rather arbitrary 'unimodal' discrete distribution with the same mean value. The results are summarized in Table 1. The necessary calculations have been carried out by Evert Jan Bakker and Jan Blaakmeer, to whom I am much obliged.

Notice that Table 1 gives also quantile points of the convex upperbound distributions Φ_c: $1-\alpha$ as a function of T_α^*. Table 1 shows, that the uniform distribution B indeed is much less favorable than the triangular distribution A; the two-point distribution C gives much higher values for T_α^* than the three-point distribution D for not too large values for α. Of course, in any case the solutions coincide (approximately) for $\alpha =$ (approximately) 0. For distribution A and $\alpha = 0.5$ more details about the solutions are indicated in Figure 3: for each activity i the values for π_i^* and z_i^* are indicated. If, loosely speaking, we interpret α as the maximal acceptable 'risk' which we are ready to take under unfavorable conditions, π_i^* indicates the (uniquely) determined contribution of activity i to that risk. If $\pi_i^* = 0$ then activity i is not important in that respect, and the promise z_i^* can be taken maximal $= b_i$, but the more π_i^*/α is close to one, the more risk we have to take in activity i. By the way, this example illustrates also that the criticality numbers of the paths are far from unique: every flow passes node 18, and the flow in each sub-path from source to 18 and in each sub-path from 18 to sink is uniquely determined by π^*, but there are many possibilities to combine the flows in sub-paths to flow in paths.

References

[1] S. Cambanis, G. Simons and W. Stout, "Inequalities for $Ek(X, Y)$ when the marginals are fixed", *Zeitschrift für Wahrscheinlichkeitsrechnung und verwandte Gebiete* 36 (1976) 285-294.

[2] T. Cipra, "Class of unimodal distributions and its transformations", *Časopis pro pěstování matematiky, roč.* 103 (1978) 17-26.

[3] H.J. Cleef and W. Gaul, "Project scheduling via stochastic programming", *Mathematische Operationsforschung und Statistik Series Optimization* 13 (1982) 449-469.

[4] J. Dupačová, "Stochastic and dynamic programming: Experience in stochastic programming models", in: A. Prékopa, ed., *Survey of Mathematical Programming*, Proceedings of the 9th International Mathematical Programming Symposium (Publishing House of the Hungarian Academy of Sciences, Budapest, 1976), part 5, pp. 99-105.

[5] S.E. Elmaghraby, "On the expected duration of PERT type networks", *Management Science* 5 (1967) 299-306.

[6] S.E. Elmaghraby, *Activity networks* (John Wiley & Sons, New York, London, Sydney, Toronto, 1977).

[7] D.R. Fulkerson, "A network flow computation for project cost curves", *Management Science* 7 (1961) 167-178.

[8] D.R. Fulkerson, "Expected critical path length in PERT type networks", *Operations Research* 10 (1962) 808-817.

[9] N. Gaffke and L. Rüschendorf, "On a class of extremal problems in statistics", *Mathematische Operationsforschung und Statistik Series Optimization* 12 (1981) 123-135.

[10] N. Gaffke and L. Rüschendorf, "On the existence of probability measures with given marginals", *Statistics and Decisions*, to appear.

[11] U. Garcia-Palomares and E. Giné, "On the linear programming approach to the optimality property of Prokhorov's distance", *Journal of Mathematical Analysis and Applications*, 60 (1977) 596-600.

[12] W. Gaul, "Bounds for the expected duration of a stochastic project planning model", *Journal of Information and Optimization Sciences* 2 (1981) 45-63.
[13] W. Gaul, "Bounding distributions for random project-completion-times", in: M.J. Beckman, W. Eichhorn and W. Krelle, eds., *Mathematische Systeme in der Ökonomie* (Athenäum Verlag, Königstein/Ts, 1983) pp. 169-179.
[14] W.-R. Heilmann, "A mathematical programming approach to Strassen's theorem on distributions with given marginals", *Mathematische Operationsforschung und Statistik. Series Optimization* 12 (1981) 593-596.
[15] H.G. Kellerer, "Funktionen auf Produkträumen mit vorgegebenen Marginalfunktionen", *Mathematische Annalen* 144 (1961) 323-344.
[16] H.G. Kellerer, "Verteilungsfunktionen mit gegebenen Marginalverteilungen", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 3 (1964) 247-270.
[17] H.G. Kellerer, "Masstheoretische Marginalprobleme", *Mathematische Annalen* 153 (1964) 168-198.
[18] H.G. Kellerer, "Duality theorems for marginal problems and stochastic applications", in: *Proceedings of the 7th Braşov Conference on Probability Theory* (to appear, 1984).
[19] H.G. Kellerer, "Duality theorems for marginal problems", to be published (1984).
[20] G.B. Kleindorfer, "Bounding distributions for a stochastic acyclic network", *Operations Research* 19 (1971) 1586-1601.
[21] T.L. Lai and H. Robbins, "A class of dependent random variables and their maxima", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 42 (1978) 89-111.
[22] E.L. Lehman, "Some concepts of dependence", *The Annals of Mathematical Statistics* 37 (1966) 1137-1152.
[23] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its applications* (Academic Press, New York, London, Toronto, Sydney, San Francisco, 1979).
[24] J.J. Martin, "Distribution of the time through a directed, acyclic network", *Operations Research* 13 (1965) 46-66.
[25] I. Meilijson and A. Nádas, "Convex majorization with an application to the length of critical paths", *Journal of Applied Probability* 16 (1979) 671-677.
[26] J.J. Moder and C.R. Phillips, *Project management with CPM and PERT* (Rheinhold, New York, NY, 1964).
[27] A. Nádas, "Probabilistic PERT", *IBM Journal of Research and Development* 23 (1979) 339-347.
[28] P. Robillard and M. Traham, "The completion time of PERT networks", *Operations Research* 25 (1977) 15-29.
[29] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).
[30] L. Rüschendorf, "Inequalities for the expectation of Δ-monotone functions", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 54 (1980) 341-349.
[31] L. Rüschendorf, "Sharpness of Fréchet-bounds", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57 (1981) 293-302.
[32] L. Rüschendorf, "Random variables with maximum sums", *Advances in Applied Probability* 14 (1982) 623-632.
[33] L. Rüschendorf, "Solution of a statistical optimization problem by rearrangement methods", *Metrika* 30 (1983) 55-61.
[34] G. Schay, "Nearest random variables with given distributions", *The Annals of Probability* 2 (1974) 163-166.
[35] A.W. Shogan, "Bounding distributions for a stochastic PERT network", *Networks* 7 (1977) 359-381.
[36] V. Strassen, "The existence of probability measures with given marginals", *The Annals of Mathematical Statistics* 36 (1965) 423-439.
[37] A. Szulga, "On the Wasserstein metric", *Transactions of the eighth Prague Conference on information theory, statistical decision functions, random process* (Academia, Publishing house of the Czechoslovak Academy of Sciences, Prague, 1978), Volume B, pp. 267-273.
[38] A. Szulga, "On minimal metrics in the space of random variables", *Probability Theory and its Applications* 2 (1982) 401-405. Academy of Sciences of the USSR, Moscow.
[39] W. Whitt, "Bivariate distributions with given marginals", *The Annals of Statistics* 4 (1976) 1280-1289.
[40] J. Žačková, "On minimax solutions of stochastic linear programming problems", *Časopis pro pěstování matematiky, roč.* 91 (1966) 423-429.

Mathematical Programming Study 27 (1986) 183–190
North-Holland

AN ALTERNATING METHOD FOR STOCHASTIC LINEAR PROGRAMMING WITH SIMPLE RECOURSE

Liqun QI

*International Institute for Applied Systems Analysis, Schloss Laxenburg, A-2361 Laxenburg, Austria, and Computer Sciences Department, University of Wisconsin-Madison, Madison, WI 53706, USA**

Received 3 January 1984
Revised manuscript received 15 August 1984

Stochastic linear programming with simple recourse arises naturally in economic problems and other applications. One way to solve it is to discretize the distribution functions of the random demands. This will considerably increase the number of variables and will involve discretization errors. Instead of doing this, we describe a method which alternates between solving some n-dimensional linear subprograms and some m-dimensional convex subprograms, where n is the dimension of the decision vector and m is the dimension of the random demand vector. In many cases, m is relatively small. This method converges in finitely many steps.

Key words: Stochastic Linear Programming, Convex Program, Linear Program.

1. Introduction

The standard form of stochastic linear programs with simple recourse is as follows

$$\begin{aligned} \min \quad & cx + E(Q(x, \xi)) \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{SLP}$$

with

$$\begin{aligned} Q(x, \xi) = \min \quad & q^+ y^+ + q^- y^- \\ \text{s.t.} \quad & y^+ - y^- = \xi - Tx, \\ & y^+ \geq 0, \ y^- \geq 0, \end{aligned} \tag{RP}$$

where x, $c \in \mathbb{R}^n$, y^+, y^-, q^+, $q^- \in \mathbb{R}^m$, $b \in \mathbb{R}^k$, $A \in L(\mathbb{R}^n, \mathbb{R}^k)$, $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and ξ is an m-dimensional random vector with known distribution F, $\bar{\xi} < \infty$, $q = q^+ + q^- \geq 0$.

One of the economic interpretations of this model is as follows: x is the decision vector, which is nonnegative. $Ax = b$ is the resource constraint on x. T is the technology matrix which yields the linear transformation of the activities in finished products $w = Tx$. ξ represents demand. The goal is to minimize the total expected cost which consists of two parts: the production cost cx and the penalty cost. For

Sponsored by the National Science Foundation under Grant No. MCS-8200632.

* Current address: Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA.

the ith product, $i = 1, \ldots, m$, the output w_i is to be compared with the observed value ξ_i of the stochastic demand ξ_i. Any discrepancies between ξ_i and w_i, $i = 1, \ldots, m$, are penalized as follows:

if $\xi_i < w_i$ then the penalty is $q_i^-(w_i - \xi_i)$,

if $\xi_i \geq w_i$ then the penalty is $q_i^+(\xi_i - w_i)$.

The first eventuality corresponds to excess product, the second to shortage.

The equivalent deterministic problem can be written as

$$\begin{aligned} \min_{x,w} \quad & cx + \phi(w) \\ \text{s.t.} \quad & Ax = b, \\ & Tx = w, \\ & x \geq 0, \end{aligned} \tag{EDP}$$

where $w \in \mathbb{R}^m$ represents the output as mentioned above, $\phi(w) = \sum_{i=1}^m \phi_i(w_i)$ is separable convex and continuous,

$$\phi_i(w_i) = q_i^- \int_{\xi_i < w_i} (w_i - \xi_i)\, dF_i(\xi_i) + q_i^+ \int_{\xi_i > w_i} (\xi_i - w_i)\, dF_i(\xi_i),$$

F_i is the marginal distribution of ξ_i, $i = 1, \ldots, m$ [6, 12, 13].

Suppose that $m + k$ is small comparing with n. This assumption is practical. For example, in the stochastic transportation problem [1, 8, 15, 16, 17, 18], $n = m \times k$.

There are simplex-type methods to solve (EDP) by discretizing ϕ [3, 11] (for more references, see [10]). Usually, this will considerably increase the number of variables and will involve discretization errors. To avoid these two drawbacks, we describe here a method which alternates between solving some n-dimensional linear subprograms by fixing w and some m-dimensional convex subprograms by restricting n-m-k activities to zero levels. In Section 2, we define such subprograms and discuss their solvability. In Section 3, we give the algorithm. In Section 4, we prove the convergence of our method.

2. Subprograms

Definition 2.1. For any *fixed* $w \in \mathbb{R}^m$, the following linear program is called a linear subprogram of (EDP) and denoted by $L(w)$:

$$\begin{aligned} \min_{x} \quad & cx + \phi(w) \\ \text{s.t.} \quad & Ax = b, \\ & Tx = w, \\ & x \geq 0. \end{aligned} \tag{$L(w)$}$$

Without confusion, we also denote its optimal value by $L(w)$.

The constant term $\phi(w)$ in the objective function does not affect the solution. However, we keep it there for comparison with (EDP).

Theorem 2.2. *Suppose* (EDP) *is solvable. Then* $L(w)$ *is solvable if and only if it is feasible.*

Proof. 'Only if' is obvious. For the 'if' part, it suffices to know that $L(w)$ is bounded, then we know that $L(w)$ is solvable according to the duality theorem of linear programming. However, $L(w)$ cannot be unbounded, since otherwise (EDP) would be unbounded. This proves the theorem. □

If (EDP) is solvable, then for certain w^*, $L(w^*)$ will yield optimal solutions of (EDP). Such w^* is called a *certainty equivalent* of (EDP) [12, 14]. If we can find a certainty equivalent, then we have almost solved (EDP). Our method will provide a way to seek such a certainty equivalent.

Denote the support vector of x by $\operatorname{supp}(x)$.

$$\langle l = \operatorname{supp}(x)\rangle := \begin{cases} l_i = 1 & \text{if } x_i \neq 0, \\ l_i = 0 & \text{if } x_i = 0. \end{cases}$$

Definition 2.3. Let M be the set of all n-dimensional vectors with $m+k$ components being '1' and others being '0'. Suppose that $l \in M$. The following convex program is called a convex subprogram of (EDP) and is denoted by $C(l)$:

$$\begin{aligned} \min_{x,w} \quad & cx + \phi(w) \\ \text{s.t.} \quad & Ax = b, \\ & Tx = w, \\ & \operatorname{supp}(x) \leqslant l, \\ & x \geqslant 0. \end{aligned} \qquad C(l)$$

Without confusion, we also denote its optimal value by $C(l)$.

The following theorem is proved in [12]. Since it is important to our discussion, we still give a proof here.

Theorem 2.4. *Suppose* (EDP) *is solvable. Then there exists an* $l^* \in M$ *such that* $C(l^*)$ *yields an optimal solution of* (EDP).

Proof. Suppose (x^*, w^*) is an optimal solution of (EDP). Consider $L(w^*)$. $L(w^*)$ is solvable by Theorem 2.2. Suppose x^0 is a basic optimal solution of $L(w^*)$. Then there exists $l^* \in M$ such that $\operatorname{supp}(x^0) \leqslant l^*$. Then $C(l^*)$ yields an optimal solution (x^0, w^*) of (EDP). □

We call such l^* an *optimal support* of (EDP). If we find an optimal support of (EDP), we have also almost solved (EDP). The solvability of $C(l)$, however, is not so simple even if it is feasible and (EDP) is solvable. In fact, the solvability of (EDP) only implies the boundedness of $C(l)$.

Suppose $l \in M$. Then there are $m+k$ positive integers i_s, $s = 1, \ldots, m+k$ such that $1 \leq i_1 < i_2 < \cdots < i_{m+k} \leq n$ and

$$l_h = \begin{cases} 1 & \text{if } h = i_s \text{ for some } s, \\ 0 & \text{otherwise.} \end{cases}$$

Let x_e and c_e be the $m+k$ dimensional vectors consisting of the i_s-th components, $s = 1, \ldots, m+k$, of x and c. Let A_e and T_e be the submatrices of A and T, consisting of the i_s-th columns, $s = 1, \ldots, m+k$, of A and T. Then $C(l)$ is equivalent to

$$\begin{aligned} \min_{x_e, w} \quad & c_e x_e + \phi(w) \\ \text{s.t.} \quad & A_e x_e = b, \\ & T_e x_e = w, \\ & x_e \geq 0. \end{aligned} \tag{2.1}$$

Write $B = \binom{A}{T_e}$. This is an $(m+k) \times (m+k)$ square matrix. Suppose $G = B^{-1}$ exists. Then (2.1) is equivalent to

$$\begin{aligned} \min_{w} \quad & c_e G \begin{pmatrix} b \\ w \end{pmatrix} + \phi(w) \\ \text{s.t.} \quad & G \begin{pmatrix} b \\ w \end{pmatrix} \geq 0. \end{aligned} \tag{2.2}$$

Let $c_e G = (g, h)$, where $g \in \mathbb{R}^k$, $h \in \mathbb{R}^m$. Then (2.2) is equivalent to

$$\begin{aligned} \min_{w} \quad & \sum_{j=1}^{m} [\phi_j(w_j) + h_j w_j] + gb \\ \text{s.t.} \quad & G \begin{pmatrix} b \\ w \end{pmatrix} \geq 0. \end{aligned} \tag{2.3}$$

Another equivalent form of (2.1) is simply

$$\begin{aligned} \min_{x_e} \quad & c_e x_e + \phi(T_e x_e) \\ \text{s.t.} \quad & A_e x_e = b, \\ & x_e \geq 0. \end{aligned} \tag{2.4}$$

Remark 2.5. (2.3) is an m-variable convex program with linear constraints and a separable convex objective function. There are many methods to solve it [2, 4, 5, 7, 9]. We shall not discuss this here.

The solvability of $C(l)$ can be assured by some conditions on the random vector ξ.

Theorem 2.6. *Suppose $q \geq 0$ and Ω, the support of the random vector ξ, is compact. Then $C(l)$ is solvable for any $l \in M$ if it is feasible and* (EDP) is solvable.

Proof. If Ω is compact, then Ω_i, the support of ξ_i, $i = 1, \ldots, m$, are also compact. This also implies that $\bar{\xi}$ exists and thus ϕ is finite. Since (EDP) is solvable, (2.1), i.e., $C(l)$ is bounded. Then there exists feasible point sequence $\{(x_e^r, w^r) | r = 0, 1, 2, \ldots\}$ such that

$$\lim_{r \to \infty} [c_e x_e^r + \phi(w^r)] = \mathrm{Inf}\{C(l)\}.$$

If $\{(x_e^r, w^r) | r = 0, 1, 2, \ldots\}$ has a limiting point, then this limiting point will be an optimal solution of $C(l)$ since the feasible set of $C(l)$ is closed and the objective function of $C(l)$ is continuous. Suppose it has no limiting point. Then it has a limiting direction (x_e^c, w^c). Since the feasible set of $C(l)$ is closed convex and the objective function of $C(l)$ is continuous,

$$\lim_{\lambda \to \infty} [c_e(x_e^0 + \lambda x_e^c) + \phi(w^0 + \lambda w^c)] = \mathrm{Inf}\{C(l)\}.$$

Now the only case for (2.1) failing to be solvable is that

$$c_e(x_e^0 + \lambda x_e^c) + \phi(w^0 + \lambda w^c) > \mathrm{Inf}\{C(l)\} > -\infty \quad \forall \lambda \geq 0.$$

But this is impossible since $\phi(w^0 + \lambda w^c)$ is linear for λ sufficiently large (see 10.2 and 12.4 of [12]). □

3. The alternating algorithm

Algorithm 3.1. Starting from any $l^0 \in M$ or starting from any $w^0 \in \mathbb{R}^m$, do the following two procedures alternatively until the method stops in step 2. This (x^j, w^j) is an optimal solution.

1. From l^j, solve $C(l^j)$ to get an optimal solution w^j.
2. From w^j, solve $L(w^j)$ to get a basic optimal solution x^{j+1}. Pick $l^{j+1} \in$, such that

$$\mathrm{supp}(x^{j+1}) \leq l^{j+1}. \tag{3.1}$$

If there is more than one basic optimal solution of $L(w^j)$, we should choose x^{j+1} such that there exists $l^{j+1} \in M$ satisfying (3.1) and

$$C(l^{j+1}) < C(l^j). \tag{3.2}$$

If no such l^{j+1} can be found, or $l^{j+1} = l^j$, stop.

Remark 3.2. To start this algorithm, we can pick any $x^0 \in \{x | Ax - b, x \geq 0\}$ and let

$w^0 = Tx^0$. However, a good starting point should be a good estimate of the certainty equivalent w^*. According to our model, we can take $w^0 = \bar{\xi}$.

Remark 3.3. If $L(w^j)$ is nondegenerate, we can simply take

$$l^{j+1} = \text{supp}(x^{j+1}).$$

Remark 3.4. This algorithm yields a sequence $l^0, l^1, \ldots, l^j =$ an optimal support. We will prove in Section 5 that $l^r \neq l^s$ for $r \neq s$. In this sense, it looks like a pivoting method in M. However, l^{r+1} is not necessarily a neighbor vector of l^r. They may be different in more than two components.

4. The convergence theorem

Theorem 4.1 (Convergence Theorem). *Suppose that* $q \geq 0$, *that* $K = \{x | Ax = b, x \geq 0\} \neq \emptyset$, *that* (EDP) *is solvable and that the support of the random variable* ξ *is compact. Then Algorithm* 3.1 *is well-defined and stops in finitely many steps if it starts from a feasible point* w^0. *Furthermore,* (3.2) *holds for every* j *in this case.*

Proof. In Algorithm 3.1, $C(l^j)$ and $L(w^j)$ are always feasible if the algorithm begins from a feasible point. Therefore, by Theorems 2.2, 2.6 and the hypotheses of this theorem, we know that $C(l^j)$ and $L(w^j)$ are always solvable. In fact, it now suffices to prove that (3.2) holds for each l^j which is not an optimal support. Then we get the conclusion since M is finite. Suppose $C(l^j)$ has an optimal solution $(\bar{x}_e, w^j)$. Let $\bar{x}$ be the n-dimensional vector whose compenents consist of $\bar{x}_e$ and 0 correspondingly. If $\bar{x}$ is not an optimal solution of $L(w^j)$, then there exists an $l^{j+1} \in M$ with

$$C(l^{j+1}) \leq L(w^j) < C(l^j).$$

Thus, suppose $\bar{x}$ is an optimal solution of $L(w^j)$. We shall prove that there is an optimal solution of $L(w^j)$ such that the associated l^{j+1} satisfies (3.2). Suppose that w^* is a certainty equivalent. Let x^* be an optimal solution of $L(w^*)$. Since l^j is not an optimal support,

$$cx^* + \phi(w^*) < c\bar{x} + \phi(w^j).$$

Let

$$(x^\lambda, w^\lambda) = \lambda(\bar{x}, w^j) + (1-\lambda)(x^*, w^*), \quad 0 < \lambda < 1.$$

According to the convexity of ϕ, we have

$$cx^\lambda + \phi(w^\lambda) < c\bar{x} + \phi(w^j) \quad \forall 0 < \lambda < 1.$$

Suppose $\bar{x}^\lambda$ is a basic optimal solution of $L(w^\lambda)$. Then

$$c\bar{x}^\lambda + \phi(w^\lambda) \leq cx^\lambda + \phi(w^\lambda) < c\bar{x} + \phi(w^j).$$

Now pick $l^\lambda \in M$ corresponding to the optimal basis of $\bar{x}^\lambda$. We have

$$C(l^\lambda) \leq c\bar{x}^\lambda + \phi(w^\lambda) < c\bar{x} + \phi(w^j) = C(l^j).$$

Therefore

$$l^\lambda \neq l^j \quad \forall 0 < \lambda < 1.$$

Therefore, since M is finite, there exists a sequence $\{\lambda^r \mid r = 1, 2, \ldots\}$ such that $\lambda^r \to 1^-$, $l^{\lambda^r} \equiv l^{j+1}$ for some $l^{j+1} \in M$. Let B be the basic matrix of $\binom{A}{T}$, corresponding to l^{j+1}. Then $\hat{x} = B^{-1}\binom{b}{w^j} = \lim_{r\to\infty} B^{-1}\binom{b}{w^{\lambda^r}} = \lim_{r\to\infty} x^{\lambda^r} \geq 0$ exists. Therefore, l^{j+1} also corresponds to an optimal basic solution of $L(w^j)$, and

$$C(l^{j+1}) = C(l^{\lambda^r}) < C(l^j).$$

This proves the theorem. □

Corollary 4.2. *Let $\bar{x}$ be defined in the above proof. Then a necessary condition for w^j to be a certainty equivalent is that $\bar{x}$ be an optimal solution of $L(w^j)$. A sufficient condition for w^j to be a certainty equivalent is that $\bar{x}$ be the unique optimal solution of $L(w^j)$.* □

Acknowledgment

I am thankful to Professor S. M. Robinson for his suggestions.

References

[1] V. Balachandran, "Generalized transportation networks with stochastic demands: An operator theoretic approach", *Networks* 9 (1979) 169-184.

[2] M. Bazaraa and C.M. Shetty, *Nonlinear programming theory and algorithms* (John Wiley & Sons, New York, 1979).

[3] G.B. Danizig, *Linear programming and extensions* (Princeton University Press, Princeton, NJ, 1963).

[4] P.E. Gill and W. Murray, "Linearly constrained problems including linear and quadratic programming", in D. Jacobs, ed., *The state of the art in numerical analysis* (Academic Press, New York, 1977) pp. 313-363.

[5] P.E. Gill, W. Murray and M.H. Wright, *Practical optimization* (Academic Press, New York, 1980).

[6] P. Kall, *Stochastic linear programming* (Springer-Verlag, Berlin, 1976).

[7] L. Qi, "Base set strategy for solving linearly constrained convex programs", Technical Report 505, Computer Sciences Department, University of Wisconsin-Madison (Madison, WI, 1983).

[8] L. Qi, "Forest iteration method for stochastic transportation problem", Technical Report 522, Computer Sciences Department, University of Wisconsin-Madison (Madison, WI, 1983). Forthcoming in *Mathematical Programming Study*.

[9] S.M. Robinson, "Perturbed Kuhn-Tucker points and rates of convergence for a class a nonlinear programming algorithms", *Mathematical Programming* 7 (1974) 1-16.

[10] I.M. Stancu-Minasian and M.J. Wets, "A research bibliography in stochastic programming 1955-1975", *Operations Research* 24 (1976) 1078-1179.

[11] B. Strazicky, "Some results concerning an algorithm for the discrete recourse problem", in M.A.H. Dempster, ed., *Stochastic programming* (Academic Press, New York, 1980) pp. 263-274.

[12] R.J.-B. Wets, "Solving stochastic programs with simple recourse, I", Department of Mathematics, University of Kentucky (Lexington, KY, 1974).
[13] R.J.-B. Wets, "Stochastic programs with fixed recourse: the equivalent deterministic problem", *SIAM Review* 16 (1974) 309-339.
[14] R.J.-B. Wets, "Solving stochastic programs with simple recourse", *Stochastics* 10 (1983) 219-242.
[15] A.C. Williams, "A stochastic transportation problem", *Operations Research* 11 (1963) 759-770.
[16] D. Wilson, "On an a priori bounded model for transportation problems with stochastic demand and integer solutions", *AIEE Transaction* 4 (1972) 186-193.
[17] D. Wilson, "Tighter bounds for stochastic transportation problems", *AIEE Transaction* 5 (1973) 180-185.
[18] D. Wilson, "A mean cost approximation for transportation problems with stochastic demands", *Naval Research Logistics Quarterly* 22 (1975) 181-187.

MATHEMATICAL PROGRAMMING STUDIES

PREFACE

One may think of stochastic programming as simply a subfield of nonlinear programming. The fact that the objective function or some of the constraints are expressed in terms of multidimensional integrals does not change the essence of the problem, at least in theory. However, it is precisely because the problem at hand demands the evaluation of multidimensional integrals that its nature is altered in a fundamental way from a numerical viewpoint.

Let us consider the following type of problem

$$\text{find } x \in S \subset \mathbb{R} \text{ that minimizes } F(x)$$

where $F(x) - E\{f(x, \omega)\} = \int f(x, \omega)\, dP(\omega)$, and S is a closed set determined by some constraints that could be of probabilistic type. For simplicity, let us consider the case in which only the objective is given by a multidimensional integral with respect to the probability measure P. Because the operator E has a 'smoothing' effect, this optimization problem usually has many desirable properties. For example, if the integrand f is convex in x, then so is F. If f is differentiable with respect to x or the measure P is absolutely continuous, then it usually turns out that F is also differentiable. Thus, in principle the problem could be solved by relying on some of the existing subroutines for nonlinear programming problems; all that is needed is to appeal to a multidimensional integration subroutine to evaluate the function F, its gradients, or subgradients, as the case may be. However, general purpose integration subroutines are available only for the 1-dimensional case. In 2-dimensions some serious difficulties already must be dealt with, and in 3-dimensions subroutines are available only for very special cases. Typically, a stochastic programming problem involves anywhere from 5 to 100 random variables, making it totally impossible to rely on existing numerical integration subroutines. Naturally, some efforts have been made to design multidimensional integration subroutines—and some of the papers in this collection report the progress made in that direction—but essentially they rely on sampling techniques (involving the generation of pseudo- or quasi-random numbers). This presupposes that the integrand is sufficiently easy to evaluate, and for stochastic programming models that is the exception, not the rule. The integrand $f(x, \omega)$ is often defined implicitly, for example as the optimal value of an optimization problem. Thus efficient procedures must avoid numerous evaluations of the integrand.

Although the search for reliable multidimensional integration subroutines has not been abandoned, the design of solution procedures for stochastic optimization problems has been chiefly oriented toward methods that in one way or another avoid coming to grips with this potential stumbling block. Excluding some very

specific classes of stochastic programming problems, the suggested solution strategies can be divided into two major categories:

- 'descent' methods that rely on directions determined by statistical estimates of the subgradients of F, and
- approximation methods that replace either the original distribution P by one that would be more manageable or the integrand f by a 'simpler' one that would make it possible to carry out the multidimensional integration.

All these possibilities are illustrated in the articles of this collection.

In Volume 1, the first three articles deal with evaluating multidimensional integrals as they arise in stochastic programming (Szantai, Niederreiter) or obtaining bounds for them (Gassman/Ziemba). The next group of three articles deal with approximation schemes. We start with a review of the existing results as well as some suggestions for implementation (Birge/Wets). Approximating by problem redefinition is illustrated in the article by Beale, Dantzig and Watson, whereas Andreatta and Runggaldier work by approximating the probability measure. Intimately related to approximation is the question of the stability of the problem under various perturbations, in particular perturbations of the probability distribution function. This is the subject of the contributions of Dupačová and Wang.

The remaining articles deal with specific procedures for solving particular or general stochastic programming problems. In Volume 1, the articles by Klein Haneveld and Qi deal with stochastic network problems. The structure of the problems plays a very important role in the procedures they suggest. In Volume 2, the first four articles deal with stochastic programs with recourse models (Nazareth/Wets, Wallace, Louveaux, Rockafellar/Wets). Next, Komaromi suggests a new dual-based procedure for solving problems with probabilistic constraints. The last three articles introduce modifications of the stochastic-gradient method to make the calculations of the step size more directly adaptive (Ruszczynski), to include nonstochastic descent information (Marti/Fuchs), and to allow for its application in the case where the decision variables themselves are probability measures (Gaivoronski).

The decision to submit these contributions in the form of a Study was made at the first meeting of COSP (Committee on Stochastic Programming) on December 1, 1983, at IIASA (International Institute for Applied Systems Analysis), Laxenburg, Austria. These two volumes could very well serve as commemorative issues to mark that occasion.

Andras Prékopa
Roger J.-B. Wets

MATHEMATICAL PROGRAMMING STUDY 28

A PUBLICATION OF
THE MATHEMATICAL PROGRAMMING SOCIETY

Stochastic Programming 84
Part II

Edited by A. PRÉKOPA and R.J.-B. WETS

April 1986

NORTH-HOLLAND – AMSTERDAM

This STUDY is also available to nonsubscribers in a book edition.

Printed in The Netherlands

CONTENTS

PART I
(Mathematical Programming Study 27)

PART II
(this volume)

Mathematical Programming Study 28 (1986) 1-28
North-Holland

ALGORITHMS FOR STOCHASTIC PROGRAMS: THE CASE OF NONSTOCHASTIC TENDERS

J.L. NAZARETH
*International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria**

R.J.-B. WETS
Department of Mathematics, University of California at Davis, CA 95616, USA

Received 15 January 1984
Revised manuscript received 27 February 1985

We consider solution strategies for stochastic programs whose deterministic equivalent programs take on the form: Find $x \in \mathbb{R}^n$, $\chi \in \mathbb{R}^m$ such that $x \geq 0$, $Ax = b$, $Tx = \chi$ and $z = cx + \Psi(\chi)$ is minimized. We suggest algorithms based upon (i) extensions of the revised simplex method, (ii) inner approximations (generalized programming techniques), (iii) outer approximations (min-max) strategies.

Key words: Stochastic Programs with Recourse, Generalized Programming, Nonstochastic Tenders, Inner Linearization.

1. Introduction

The class of stochastic programs (with recourse) that we consider in this paper, and to which we refer as having nonstochastic tenders, arise as models for the following decision process. An (optimal) decision vector x must be selected when some of the parameters of the problem are only known in probability, i.e. only in a statistical sense, the actual cost depending in part on how well a transformation of x, $\chi = Tx$ matches a random demand or recourse vector p.

We think of χ as a *tender, nonstochastic* if the transformation T does not depend on the (unknown) values of the random parameters. For example, stochastic programs with simple recourse and fixed technology matrix are of this type. As we shall see in Section 2, for stochastic (linear) programs, the equivalent deterministic program can then be expressed as:

$$\begin{aligned} &\text{find} \quad x \in \mathbb{R}^n, \ \chi \in \mathbb{R}^m \quad \text{such that} \\ &Ax = b, \quad Tx = \chi, \quad x \geq 0, \\ &\text{and } z = cx + \Psi(\chi) \text{ is minimized.} \end{aligned} \tag{1.1}$$

The algorithms that we analyze could be viewed as procedures for convex programs of the type (1.1) that seek to take advantage of the special structure, and to some extent that view is certainly correct. In fact we expect that the suggested techniques

* Present address: CDSS, P.O. Box 4908, Berkeley, CA 94704, USA

will also be efficient whenever nonlinear optimization problems can be cast in the form (1.1). However, because stochastic programming problems present computational challenges of their own, it is their specific properties that are always in the background of our solution strategies. For example, our title is intended to suggest that *the major task of the solution procedure is the finding of optimal or nearly optimal tenders.*

In Section 2, we review briefly the properties of stochastic programs that will be used in the design of algorithmic procedures. In Section 3, we examine the issue of what information can be made available and its cost, and we also exhibit some important special cases when the objective and the underlying distribution functions are such, that the equivalent deterministic programs can be conveniently and inexpensively specified. We then turn to the three main solution strategies that we consider here. They are based upon

(i) extensions of the revised simplex method,
(ii) inner approximations (generalized programming techniques),
(iii) outer approximations (min-max strategies).

In order to give the essence of each solution strategy, we consider first, in Section 4, a very simple case, viz., equivalent linear programming formulations for finding the minimum of a convex piecewise linear function of one variable. In Sections 5, 6 and 7 we then go into each one as it applies to our class of stochastic programming problems. The detailed design and implementation of two specific algorithms can be found in a companion paper, Nazareth and Wets (1985).

2. Stochastic programs with recourse: Nonstochastic tenders

We consider stochastic (linear) programs of the type

$$\begin{aligned} &\text{find } x \in R^{n_1} \text{ such that} \\ &Ax = b, \quad x \geq 0, \\ &\text{and } z = E\{c(w)x + Q(x, w)\} \text{ is minimized,} \end{aligned} \tag{2.1}$$

where Q is calculated by finding for given decision x and event w, an optimal recourse $y \in \mathbb{R}^{n_2}$, viz.

$$Q(x, w) = \inf_{y \in C} [q(y, w) \mid Wy = p(w) - Tx]. \tag{2.2}$$

Here $A(m_1 \times n_1)$, $T(m_2 \times n_1)$, $W(m_2 \times n_2)$ and $b(m_1)$ are given (fixed) matrices, $c(\cdot)(n_1)$ and $p(\cdot)(m_2)$ are random vectors, $y \mapsto q(y, \cdot): \mathbb{R}^{n_2} \mapsto \mathbb{R}$ is a random finite-valued convex function and C is a convex polyhedral subset of $\mathbb{R}^{n_2}$, usually $C = \mathbb{R}^{n_2}_+$. Because W is nonstochastic one refers to (2.1) as having *fixed recourse.* Tenders are nonstochastic because T is fixed. With

$$c = E\{c(w)\} \quad \text{and} \quad \bar{Q}(x) = E\{Q(x, w)\}$$

we obtain the equivalent deterministic form of (2.1):

$$\begin{aligned} &\text{find} \quad x \in \mathbb{R}^{n_1} \quad \text{such that} \\ &Ax = b, \quad x \geq 0, \\ &\text{and } z = cx + \bar{Q}(x) \text{ is minimized.} \end{aligned} \tag{2.3}$$

We assume that the random elements of the problem are such that all quantities introduced are well-defined, with $\bar{Q}(x)$ finite, unless

$$P\{w \mid (p(w) - Tx) \notin W(C)\} > 0$$

where $W(C) = \{t = Wy \mid y \in C\}$, i.e. there is no feasible recourse with positive probability, in which case $\bar{Q}(x) = +\infty$. Detailed conditions have been made explicit in Wets, 1974; extensions to the multistage and nonconvex cases have been provided by P. Olsen, 1976 and J.B. Hiriart-Urruty, 1978 respectively.

As background to the algorithmic development, we review the basic properties of (2.3), proofs and further details can be found in Wets, 1974; see also Kall, 1982 for a compact treatment of stochastic prqgrams with *complete recourse*, i.e. when $W(C) = \mathbb{R}^{m_2}$ and consequently $\bar{Q}$ is everywhere finite.

2.4. Properties. *The function $\bar{Q}$ is lower semicontinuous and convex. It is Lipschitz if for (almost) all w, $y \mapsto q(y, w)$ is Lipschitz. Also the set*

$$K_2 = \{x \mid \bar{Q}(x) < +\infty\}$$

is a convex polyhedron that can be expressed as

$$K_2 = \{x \mid Dx \geq d\}$$

for some matrix D and vector d. Moreover if the distribution of the random elements of the problem is absolutely continuous then $\bar{Q}$ is differentiable relative to K_2.

Because $q(\cdot, w)$ is Lipschitz rather than linear, the assertion about $\bar{Q}$ being Lipschitz does not follow directly from Theorem 7.7 of Wets, 1974 but can be gathered from its proof, or see Wets, 1972, for example.

In the case of nonstochastic tenders it is useful to consider another representation of the deterministic equivalent program. Let us define χ and other associated quantities by

$$\chi = Tx, \quad \psi(\chi, w) = \inf_{y \in C} [q(y, w) \mid Wy = p(w) - \chi]$$

and

$$\Psi(\chi) = E\{\psi(\chi, w)\}.$$

Problem (2.3) is then cast in the form (1.1):

$$\begin{aligned} &\text{find} \quad x \in \mathbb{R}^{n_1},\ \chi \in \mathbb{R}^{m_2} \quad \text{such that} \\ &Ax = b, \quad Tx = \chi, \quad x \geq 0, \\ &\text{and } z = cx + \Psi(\chi) \text{ is minimized.} \end{aligned} \tag{2.5}$$

This program, more exactly the function Ψ, exhibits the same properties as those listed for $\bar{Q}$ under Properties 2.4. In particular it is finite for all χ such that $\chi = Tx$ and $x \in K_2$. Including these constraints explicitly in the formulation of the problem, we get

$$\begin{aligned} &\text{find} \quad x \in \mathbb{R}^{n_1},\ \chi \in \mathbb{R}^{m_2} \quad \text{such that} \\ &z = cx + \Psi(\chi) \text{ is minimized, and} \\ &Ax \quad = b, \\ &Dx \quad \geq d, \\ &Tx - \chi = 0, \\ &\quad x \geq 0, \end{aligned} \tag{2.6}$$

i.e. a convex program with Ψ finite on the feasible region. In what follows we shall assume that the constraints $Dx \geq d$ have been incorporated in the constraints $Ax = b$, $x \geq 0$, so that they will no longer appear explicitly, and that Ψ is finite on

$$\{\chi = Tx \mid Ax = b,\ x \geq 0\}.$$

Stochastic programs of this type are said to have relatively complete recourse (Wets, 1974, Section 6), a situation which is always obtained if the (induced) constraints, determining K_2, are incorporated in the original constraints.

When $W = I$ and $C = \mathbb{R}^{n_2}$ there is really no need to solve an optimization problem to know the optimal recourse and its associated cost. It is uniquely determined by the relation

$$y = p(w) - \chi, \quad Tx = \chi \quad \text{and} \quad \Psi(\chi) = E\{q(p(w) - \chi, w)\}.$$

The stochastic program is then said to be with *simple recourse*, which clearly implies complete recourse: $K_2 = \mathbb{R}^{n_1}$. Determining the value of Ψ at χ depends then on our capability of performing the multidimensional integration. Usually, the cost-function will be separable. However, if there is dependence between some of the components of the $p(\cdot)$-vector and the cost depends on the joint realizations, then one must necessarily resort to this more general form. Assuming that the integral is well-defined, we have that the subdifferential Ψ is given by

$$\partial\Psi(\chi) = -E\{\partial_y q(p(w) - \chi, w)\}$$

where $\partial_y q(\cdot, w)$ denotes the subdifferential with respect to the first variable. If the

convex function $y \mapsto q(y, w)$ is differentiable, then so is Ψ. The function Ψ is also differentiable if the measure is absolutely continuous. If the random variables are independent, then the multidimensional integration to obtain the value of Ψ or its gradient is reduced to a number of simple integrals on $\mathbb{R}^1$. This also occurs when there is separability.

If in addition to simple recourse, the recourse costs are *separable*, i.e., for all w,

$$q(y, w) = \sum_{i=1}^{m_2} q_i(y_i, w),$$

then

$$\begin{aligned}\Psi(\chi) &= E \sum_{i=1}^{m_2} q_i(p_i(w) - \chi_i, w) \\ &= \sum_{i=1}^{m_2} E q_i(p_i(w) - \chi_i, w) \\ &= \sum_{i=1}^{m_2} \Psi_i(\chi_i).\end{aligned}$$

Thus (2.5) becomes a convex *separable* program:

$$\begin{aligned}&\text{find} \quad x \in \mathbb{R}^{n_1},\ \chi \in \mathbb{R}^{m_2} \quad \text{such that} \\ &Ax = b, \quad Tx = \chi, \quad x \geq 0, \qquad (2.7) \\ &\text{and } cx + \sum_{i=1}^{m_2} \Psi_i(\chi_i) \text{ is minimized.}\end{aligned}$$

This latter optimization problem possesses many properties. Those that are directly relevant to our further development are summarized here below.

2.8. Properties. *For $i = 1, \ldots, m_2$, the functions Ψ_i are convex, finite-valued and thus continuous. If the random elements have a discrete distribution, the Ψ_i are piecewise linear when the $q_i(\cdot, w)$ are piecewise linear. On the other hand, if the marginals of $\{(q_i(\cdot, w), p_i(w)), i = 1, \ldots, m_2\}$ are absolutely continuous then the Ψ_i are differentiable. Moreover, if problem* (2.7) *is solvable it admits an optimal solution with no more than $m_1 + m_2$ positive entries in the x-vector.*

These properties are derived in Walkup and Wets (1970) (see also Ziemba, 1974), except the last assertion which was obtained by Murty (1968) in a somewhat modified context. A very simple proof appears in Wets (1979). In essence this latter proof says that when the *optimal tender* is known, (2.7) reduces to an LP and then appeals to a standard theorem of linear programming. Note that the result also holds when $\Psi(\chi)$ is not separable.

A version of (2.7) which has received a lot of attention, because of its direct amenability to efficient computational schemes and the many applications that can

be cast in this form, is when q_i is itself independent of w and piecewise linear with respect to y. More precisely $q_i(\cdot)$ is given by

$$q_i(y_i, w) = \begin{cases} -q_i^- y_i & \text{if } y_i \leq 0, \\ q_i^+ y_i & \text{if } y_i \geq 0, \end{cases} \tag{2.9}$$

with $q_i = q_i^+ + q_i^- \geq 0$, yielding the convexity of $y_i \mapsto q_i(y_i, w)$. In this case the function Ψ_i takes on a form particularly easy to describe. This is done in the next section.

3. Availability of information about the objective

The exact evaluation of $\bar{Q}$ or its gradient for general probability distribution μ, function q and recourse matrix W, might be prohibitively expensive, if at all possible. The difficulties come from two directions:

(i) for each w, evaluating $Q(x, w)$ involves solving a minimization problem, and
(ii) performing the multidimensional integration

$$Q(x) = \int Q(x, w)\mu(\mathrm{d}w).$$

For simple recourse, the evaluation of $Q(x, w)$, or equivalently $\psi(\chi, w)$, because T is fixed, is easy since the recourse is uniquely determined. When the recourse costs are also separable, the multidimensional integration is reduced to m_2 separate one-dimensional integrals. With T fixed, it takes the form:

$$\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i) = \sum_{i=1}^{m_2} \int q_i(p_i(w) - \chi_i, w) F_i(\mathrm{d}w)$$

where F_i is the marginal distribution function of the random elements appearing in this expression, and the integral $\int$ is a Lebesgue-Stieltjes integral. The subgradients of the convex function Ψ are then the (Cartesian) product of the subgradients of the Ψ_i which are themselves

$$\partial\Psi_i(\chi_i) = -\mathrm{cl} \int \partial_y q_i(p_i(w) - \chi_i, w) F_i(\mathrm{d}w), \tag{3.1}$$

at least when the problem satisfies the regularity conditions suggested at the beginning of Section 2. Hence cl denotes closure. In general $\partial_y q(p_i(w) - \chi_i, w)$ is multivalued, in fact closed convex valued, and the integral is then also a closed convex set. In particular if q is piecewise linear as in (2.9), we get the following expression:

$$\partial\Psi_i(\chi_i) = [q_i F_i^-(\chi_i) - q_i^+, q_i F_i(\chi_i) - q_i^+], \tag{3.2}$$

where $q_i = q_i^+ + q_i^-$,

$$F_i^-(z) = \mathrm{Prob.}[p_i(w) < z] \quad \text{and} \quad F_i(z) = \mathrm{Prob.}[p_i(w) \leq z].$$

We just note in passing that given (3.2), it implies that Ψ_i is differentiable whenever the distribution function F_i is continuous. In general, we have the following two representations for Ψ_i:

$$\Psi_i(\chi) - q_i^+ \bar{p}_i = (q_i F_i^-(\chi_i) - q_i^+)\chi_i - q_i \int_{\zeta < \chi_i} \zeta \, dF_i(\zeta)$$

$$= (q_i F_i(\chi_i) - q_i^+)\chi_i - q_i \int_{\zeta \leq \chi_i} \zeta \, dF_i(\zeta) \tag{3.3}$$

where $\bar{p}_i = E\{p_i(\cdot)\}$. If the distribution of $p_i(\cdot)$ is discrete, say with possible values

$$p_{i1}, p_{i2}, \ldots, p_{i,k_i},$$

with $p_{it} < p_{i,t+1}$, and with associated probabilities

$$f_{i1}, f_{i2}, \ldots, f_{i,k_i},$$

the function Ψ_i is piecewise linear. With $\sum_{t=1}^{0} = 0$, we obtain

$$F_i^-(\chi_i) = \sum_{t=1}^{\tau'-1} f_{it} \quad \text{where } \tau' = \min[k_i, \inf(t \mid p_{it} \geq \chi_i)],$$

and

$$F_i(\chi_i) = \sum_{t=1}^{\tau-1} f_{it} \quad \text{where } \tau = \min[k_i, \inf(t \mid p_{it} > \chi_i)].$$

Also

$$\int_{\zeta < \chi_i} \zeta \, dF_i(\zeta) = \sum_{t=1}^{\tau'} p_{it} f_{it} \quad \text{and} \quad \int_{\zeta \leq \chi_i} \zeta \, dF_i(\zeta) = \sum_{t=1}^{\tau} p_{it} f_{it}.$$

Note that $\tau' = \tau$ unless $\chi_i = p_{it}$ for some $t = 1, \ldots, k_i$ and then $\tau' = \tau - 1$. For $l = 0, \ldots, k_i$, we set

$$s_{il} = \left(\sum_{t=1}^{l} f_{it} \right) q_i - q_i^+$$

and

$$e_{il} = q_i^+ \bar{p}_i - q_i \left(\sum_{t=1}^{l} p_{it} f_{it} \right).$$

We obtain

$$\Psi_i(\chi_i) = \sup_{l=1,\ldots,k_i} (s_{il}\chi_i + e_{il}). \tag{3.4}$$

Observe that for any value of χ_i the supremum is attained by at most 2 linear forms. As we shall see in the subsequent sections, both (3.3) and (3.4) yield useful representations for Ψ_i leading to algorithmic procedures for problems involving

functions of this type. Still another representation of Ψ_i can be exploited in an algorithmic context. The value of Ψ_i is obtained as the solution of an optimization parametrized by χ_i. Let

$$d_{il} = p_{i,l+1} - p_{il} \text{ for } l = 1, \ldots, k_{i-1} \quad \text{and} \quad d_{i0} = p_{i1}.$$

Then

$$\begin{aligned} \Psi_i(\chi_i) - q_i^+ \bar{p}_i = \inf & \sum_{l=1}^{k_i} s_{il} y_{il} \\ \text{subject to} & \sum_{l=0}^{k_i} y_{il} = \chi_i, \\ & y_{i0} \leq d_{i0}, \\ & 0 \leq y_{il} \leq d_{il}, \quad l = 1, \ldots, k_{i-1}, \\ & 0 \leq y_{i,k_i}. \end{aligned} \tag{3.5}$$

To verify (3.5) it suffices to use the fact that the coefficients s_{il}, $l = 0, \ldots, k_i$ are strictly increasing and that consequently $y_{il} > 0$ only if $y_{i0} = d_{i0}$ and for $0 < t < l$, all y_{it} are at their upper bounds. Details are worked out in [Wets, 1983a, Proposition 1].

These expressions derived for Ψ_i taken in conjunction with the methods of Section 4 contain the germ of different algorithmic procedures embedded in them.

Before we turn to this, and in order not to lose sight of the fact that we are also interested in a more general class of problems, not simply stochastic programs with simple recourse with piecewise linear separable cost structure, we also describe a more general case. Suppose

$$\psi(\chi, w) = \inf_{y \geq 0} [qy \mid Wy = p(w) - \chi]$$

and χ is such that $\psi(\chi, w)$ is finite for all possible $p(w)$, i.e. the linear program defining $\psi(\chi, w)$ is feasible and bounded with probability 1. Then parametric analysis, in particular the Basis Decomposition Theorem (Wets, 1974), shows that there is a (simplical) decomposition of the sample space of $p(\cdot)$ (= the activity space),

$$S = \{S_h \subset \mathbb{R}^{m_2}, h = 1, \ldots, l\}$$

such that if $p(w) \in S_h$ then

$$\psi(\chi, w) = q_{(h)} W_{(h)}^{-1}(p(w) - \chi), \tag{3.6}$$

and with co denoting the convex hull,

$$\partial\psi(\chi, w) = -\text{co}\{q_{(h)} W_{(h)}^{-1} \mid p(w) \in S_h\}, \tag{3.7}$$

where $W_{(h)}$ is an invertible submatrix of W and $q_{(h)}$ the subvector of q corresponding

to the columns of $W_{(h)}$. Let S' be any partition generated by S. Then

$$\Psi(\chi) = \sum_h \int_{S'_h} q_{(h)} W_{(h)}^{-1}(p(w) - \chi) P(\mathrm{d}w)$$

and

$$\partial\Psi(\chi) = \sum_h \int_{S'_h} \partial\psi(\chi, w) P(\mathrm{d}w),$$

adding to the second term the normal cone to K_2 at χ, when χ is on the boundary of K_2 if we do not have relatively simple recourse or have not cast the original problem in that form, see (2.6). The potential use of the preceding formulas depends very much on how accurate one needs to be. Multidimensional integration over convex polyhedral cones can only be approached through sampling methods, cf. Deak (1980), Birge and Smith (1982), and the references given therein.

If $p(\cdot)$ is discretely distributed by which we mean here that it takes on a finite number of possible values, with

$$P[p(w) = p_k] = f_k, \quad k = 1, \ldots, N,$$

then the above formulas become simply sums, viz.

$$\Psi(\chi) = \sum_h \sum_{\{k \mid p_k \in S'_h\}} q_{(h)} W_{(h)}^{-1}(p_k - \chi) f_k \tag{3.8}$$

and a similar expression for $\partial\Psi$. To actually compute the above we can proceed via a sort of parametric analysis that we now describe. We refer to it as a *bunching* procedure. Let

$$\zeta_k = p_k - \chi, \quad k = 1, \ldots, N,$$

and suppose we have solved the linear program

$$\text{find} \quad y \geq 0 \quad \text{such that} \quad Wy = \zeta_1 \text{ and } qy \text{ is minimized,} \tag{3.9}$$

with optimal basis $W_{(1)}$ and associated subvector $q_{(1)}$ of q. S'_1, bunch 1, is defined by

$$S'_1 = \{q_k \mid W_{(1)}^{-1}\zeta_k \geq 0\}.$$

While constructing this set, identify those $\zeta_k \notin S'_1$ such that the vector $W_{(1)}^{-1}\zeta_k$ has the fewest number (and smallest) negative elements. Let ζ_{k_2} be such a vector. We find the optimal solution and a corresponding basis $W_{(2)}$ of the linear program (3.9) with ζ_{k_2} replacing ζ_1, by dual simplex pivoting, starting with the old basis $W_{(1)}$. The second bunch S'_2 is given by

$$S'_2 = \{p_k \notin S'_1 \mid W_{(2)}^{-1}\zeta_k \geq 0\}.$$

We continue in this fashion until all p_k have been bunched. Alternative procedures can be devised taking further advantage of the combinatorial structure of decompositions. How to do this so as to minimize the work involved needs further investigation.

In any case with the above we obtain the value of Ψ at χ, as well as a subgradient of Ψ at χ, viz.

$$\sum_h \sum_{\{k \mid p_k \in S'_h\}} (q_{(h)} W_{(h)}^{-1}) f_k \in \partial \Psi(\chi).$$

Observe that $q_{(h)} W_{(h)}^{-1}$, the vector of simplex multipliers, remains constant on S'_k. With

$$P_h = P[p_k \in S'_h],$$

the above becomes

$$\sum_h q_{(h)} W_{(h)}^{-1} P_h \in \partial \Psi(\chi).$$

This formula for a subgradient of Ψ is, in fact, independent of the form of the distribution of $p(\cdot)$, the problem being always the evaluations of P_h for a partitioning scheme constructed in the manner described above.

4. An illustration of each algorithmic approach

We consider the very simple problems of finding the unconstrained minimum of a 1-dimensional finite piecewise linear convex function ϕ defined on $[x_0, x_H]$ by reformulating the problem as an equivalent linear program. (This function ϕ could of course be minimized by some 1-dimensional search procedure or simply by a sort of slopes to find where they change sign, but this is not our real concern here.) There are at least three ways of formulating this equivalent linear program. Each contains the germ of a more general solution strategy considered in later sections.

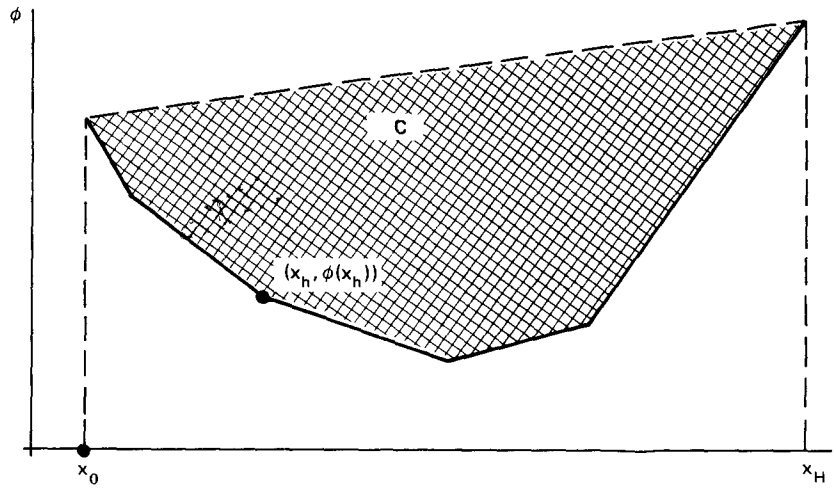


Fig. 4.1. The function ϕ.

The coordinates of the breakpoints of ϕ are denoted by

$$(x_h, \phi(x_h)), \quad h = 0, \ldots, H$$

and with slopes

$$s_h \text{ for } x \in [x_{h-1}, x_h], \quad h = 1, \ldots, H.$$

The convexity of ϕ implies that

$$s_1 \leq s_2 \leq \cdots \leq s_H. \tag{4.2}$$

With

$$\alpha_h = x_h - x_{h-1} \quad \text{and} \quad e_h = \phi(x_h) - s_h x_h,$$

the line segment on $[x_{h-1}, x_h]$ takes the form

$$\phi_h(x) := s_h x + e_h. \tag{4.3}$$

The bounded variable method

If we introduce a new variable y_h for each interval $[x_{h-1}, x_h]$ for any given value of x it is easily verified that on $[x_0, x_H]$,

$$\phi(x) = \phi(x_0) + \min\left[\sum_{h=1}^{H} s_h y_h \,\middle|\, x = x_0 + \sum_h y_h, \; 0 \leq y_h \leq \alpha_h, \; h = 1, \ldots, H\right]. \tag{4.4}$$

The assumption of convexity and hence (4.2) is of course crucial, since this means that y_1 is preferred to y_2, and y_2 to y_3 and so on in the minimization of (4.4). Hence at the optimum point in (4.3), $y_h > 0$ implies

$$y_1 = \alpha_1, y_2 = \alpha_2, \ldots, y_{h-1} = \alpha_{h-1}.$$

Minimizing ϕ on $[x_0, x_H]$ is equivalent to solving the following linear program:

$$\begin{aligned} &\text{find } y_h \in [0, \alpha_h] \text{ for } h = 1, \ldots, H \\ &\text{such that } z = \sum_{h=1}^{H} s_h y_h \text{ is minimized.} \end{aligned} \tag{4.5}$$

The optimal x^* is determined by

$$x^* = x_0 + \sum_h y_h^*$$

where $(y_h^*, h = 1, \ldots, H)$ is the optimal solution of (4.5).

Inner approximation

Referring to Fig. 4.1, any point (x, α) in the shaded region C, i.e. with $\alpha \geq \phi(x)$, can be written as a convex combination of the extreme points

$$(x_h, \phi(x_h)), \quad h = 0, \ldots, H.$$

For any given x it follows that

$$\phi(x) = \min\left[\sum_{h=0}^{H} \lambda_h \phi(x_h) \,\middle|\, \Sigma\lambda_h x_h = x,\ \Sigma\lambda_h = 1,\ \lambda_h \geq 0\right] \tag{4.6}$$

and thus minimizing ϕ on $[x_0, x_H]$ is equivalent to solving the following linear program

$$\begin{aligned} &\text{find} \quad \lambda_h \geq 0,\ h = 1, \ldots, H,\ \sum_{h=1}^{H} \lambda_h = 1 \\ &\text{such that } z = \sum_{h=0}^{H} \lambda_h \phi(x_h) \text{ is minimized.} \end{aligned} \tag{4.7}$$

The optimal x^* is determined by

$$x^* = \sum_{h=0}^{H} \lambda_h^* x_h$$

where $(\lambda_h^*, h = 0, \ldots, H)$ is the solution of (4.7). For an arbitrary convex function ρ with $\rho(x_h) = \phi(x_h)$, the function ϕ can be viewed as an inner linearization of ρ.

Outer approximation

Since ϕ is piecewise linear, we have that

$$\phi(x) = \max_{h=1,\ldots,H} \phi_h(x) \tag{4.8}$$

where the functions ϕ_h are defined by (4.3) when ϕ is expressed in this form, finding its minimum consists in solving the minimax problem:

$$\min_{x \in [x_0, x_H]} \max_{h=1,\ldots,H} \phi_h(x).$$

This is equivalent to solving the following linear program:

$$\begin{aligned} &\nu \in R \text{ and } x \in [x_0, x_H] \text{ such that} \\ &\nu \geq s_h x + e_h \text{ and} \\ &z = \nu \text{ is minimized.} \end{aligned} \tag{4.9}$$

The methods of inner and outer approximation rely on *dual representations of the epigraph* of ϕ, but one should note that the linear programs (4.7) and (4.9) are not dual linear programs.

Each of the above three problem manipulations delineates an approach to solving problems of the type (1.1). We consider each one in this more general setting in the next three sections.

5. Extensions of the revised simplex method

An algorithm for solving stochastic programs with simple recourse with separable piecewise linear recourse costs and discrete random variables has been proposed in Wets (1983a)[1], a write-up and computer code have been provided by Kallberg and Kusy (1976) and computational experience is reported in Kallberg, White and Ziemba (1982), Kusy and Ziemba (1981); cf. also Cleef (1981). From (3.4) we know that

$$\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i)$$

with each Ψ_i *polyhedral* with the slopes s_{ih} and $s_{i0} < s_{i1} < \cdots < s_{i,k_i}$. We can clearly apply the bounded variable method of Section 4 to the m_2-functions Ψ_i. Using (3.5) we obtain the following linear program, the analog of (4.5):

$$\begin{aligned} &\text{find} \quad 0 \leq x_j, \ j = 1, \ldots, n_1, \\ &\text{and for } i = 1, \ldots, m_2, \\ &0 \leq y_{il} \leq d_{il}, \quad l = 0, \ldots, k_{i-1} \quad \text{and} \quad y_{i,k_i} \geq 0 \end{aligned} \tag{5.1}$$

such that

$$A_i x = b_i, \quad i = 1, \ldots, m_1,$$

$$T_i x - \sum_{l=0}^{k_i} y_{il} = p_{i0}, \quad i = 1, \ldots, m_2, \quad \text{and}$$

$$z = cx + \sum_{i=1}^{m_2} \sum_{l=0}^{k_i} s_{il} y_{il} \text{ is minimized,}$$

where A_i and T_i are the ith rows of A and T, and, for $i = 1, \ldots, m_2$,

$$d_{i0} = p_{i1} - p_{i0},$$

with $p_{i0} \in]-\infty, p_{i1}[$ chosen so that for the optimal solution x^*, $T_i x^* > p_{i0}$ is guaranteed.

In seeking the solution of (5.1) using the revised simplex method for linear programs with simple upper bounds, let us assume that we have in hand a nondegenerate basic feasible solution say $(\tilde{x}, \tilde{y}_{il}, i = 1, \ldots, m_2, l = 0, \ldots, k_i)$ which yields the values of the functions Ψ_i, $i = 1, \ldots, m_2$ at a point $\tilde{\chi} = T\tilde{x}$ as follows: for each $i = 1, \ldots, m_2$,

$$\Psi_i(\tilde{\chi}_i) = \Psi_i(p_{i0}) + \sum_{h=0}^{k_i} s_{ih} \tilde{y}_{ih},$$

i.e. the $\tilde{y}_{il}$ actually solve the program (3.5) for $\chi_i = \tilde{\chi}_i$. It follows that if

$$0 < \tilde{y}_{il} < d_{il}$$

[1] This algorithm was proposed several years ago, but published only recently.

then

$$\hat{y}_{ih} = d_{ih} \quad \text{for } h < l,$$

$$\tilde{y}_{ih} = 0 \quad \text{for } h > l.$$

In Wets (1983a) such a solution is called a *perfect* basic solution. The fact that the optimal solution is of that type and that one can pass from a perfect basic solution to another can be argued as follows: Let $(\sigma, \pi) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ be the simplex multipliers associated with the solution at hand. To find the variable to be entered into the basis at the next iteration, we compute $\bar{s}_{ih}$, the reduced cost (the component of the reduced gradient corresponding to the current basis) associated with the variable y_{ih}, viz.

$$\bar{s}_{ih} = s_{ih} + \pi_i.$$

The $\{\bar{s}_{ih}, h = 0, \ldots, k_i\}$ are increasing, thus for each i among all variables at their lower bound ($h > l$), the variable $y_{i,l+1}$ is the one that yields potentially the greatest (marginal) improvement. Similarly, among all variables at their upper bound ($h < l$) the best candidate for decrease is $y_{i,l-1}$. It is also readily established that when either $y_{i,l+1}$ or $y_{i,l-1}$ is introduced into the basis then y_{il} will move to its appropriate bound and leave the basis. Thus the new basis will also be perfect. This property is not affected by exchanges between x-variables and y_{ih}-variables, unless degenerate cases are mishandled. A potential difficulty is that the algorithm could go through a great number of steps and associated basis changes if Ψ_i has many pieces. This can partially be overcome by an acceleration procedure that in one sweep makes a number of basis changes involving variables $\{y_{ih}, h = 0, \ldots, k_i\}$ for a given i, see Wets (1983a). This algorithm can thus be regarded as an application of the bounded variable revised simplex method with an acceleration step; in Wets (1983a) it is also shown how to exploit the structure of the problem to obtain a good starting basis.

A practical implementation of the above bounded variable method based upon the MINOS code, see Murtagh and Saunders (1978), is given in Nazareth and Wets (1985).

An extension and natural continuation of the above approach that seeks to avoid the difficulties associated with introducing bounded variables, and has the advantage of greater generality is to handle the tender χ explicitly rather than implicitly in terms of the basic variables y_{il}. This permits, for example, the distribution function F_i to be arbitrary and $q(\cdot, w)$ to be nonlinear and nonseparable. Let us return to the problem expressed in the form (2.5) or (2.6). Induced constraints are assumed to be incorporated into $Ax = b$, $x \geq 0$, and bounds on χ can be introduced if desired without substantially altering the following discussion.

As we have described under Properties 2.4 and 2.8, when the distribution function of the random elements is absolutely continuous, $\Psi(\chi)$ is differentiable. Various methods of smooth optimization can then be used or adapted directly to solve (2.5), see, for example, Ziemba (1970). The number of *nonlinear* variables χ in (2.5) is

determined by the number of technology rows, and this will normally be small compared with the number of linear variables, x. Under these circumstances, the *reduced gradient method* in the form given by Murtagh and Saunders, 1978, is particularly effective. To simplify the discussion let us write (2.5) as

$$\begin{aligned} &\text{minimize} && f(z) \\ &\text{s.t.} && \mathbb{A}z = \mathfrak{b}, \\ & && z_j \geq 0, \ i \leq j \leq m_1, \end{aligned} \tag{5.2}$$

where $z = (x, \chi)$, $f(z): cx + \Psi(\chi)$. $\mathbb{A} = \begin{bmatrix} A & 0 \\ T & -I \end{bmatrix}$ is an $m \times n$ matrix, with $m = m_1 + m_2$, and $\mathfrak{b} = \begin{pmatrix} b \\ 0 \end{pmatrix}$. Let us for this moment also assume that $f(z)$ is a *smooth* function.

Since $f(z)$ is nonlinear, we cannot claim that (5.2) has an optimal solution with at most m variables basic, i.e. off their bounds. In the MINOS implementation of the reduced gradient method (see Murtagh and Saunders, 1978) the variables at each cycle are partitioned into three groups $z = (z_B, z_S, z_N)$ representing m *basic* variables, s *superbasic* variables with $0 \leq s \leq n - m$, and $nb = n - m - s$ *nonbasic* variables respectively. Nonbasics are at their bounds. $\mathbb{A}$ is similarly partitioned as $\mathbb{A} = [B|S|N]$ where B is an $m \times m$ nonsingular matrix, S is an $m \times s$ matrix and N is an $m \times (n - m - s)$ matrix. Let $g = \nabla f(z)$ be similarly partitioned as $g = (g_B, g_S, g_N)$. A basic property of the reduced gradient method in this form is that an optimal solution exists for which the number of superbasics does not exceed the number of nonlinear variables, in this case m_2. Its proof is trivial, and indeed this is precisely the counterpart of Murty's, 1968 result discussed under (2.8).

Each cycle of the method can be viewed as roughly equivalent to (a) and (b) below:

(a) one or more iterations of a quasi-Newton method on an unconstrained optimization problem of dimension s. Here, a reduced gradient, say μ, is computed as

$$\mu = [-(B^{-1}S)^{\mathrm{T}} | I_{s \times s} | 0] \begin{pmatrix} g_B \\ g_S \\ g_N \end{pmatrix} \triangleq Z_S^{\mathrm{T}} g. \tag{5.3}$$

The columns of Z_S span the space in which the quasi-Newton step lies and normally rank $(Z_S) = s$. We can view this also in terms of active sets as follows: $\mathbb{A}z = \mathfrak{b}$ and $Z_N = 0$ are $m + (n - m - s)$ active constraints. If the columns of $\mathbb{N} = [\mathbb{A}^{\mathrm{T}} | I^J]$ are the normals to these constraints, with I^J denoting suitable columns of the $n \times n$ identity matrix $I_{n \times n}$, then $Z_S^{\mathrm{T}} \mathbb{N} = 0$.

To compute μ we can write (5.3) as

$$\mu = g_S - S^{\mathrm{T}} \Pi \tag{5.4}$$

where $\Pi^{\mathrm{T}} = g_B^{\mathrm{T}} B^{-1}$ is the vector of *prices*. The search direction, say p, is then defined by $p = -Z_S H \mu$, where H is a positive definite symmetric matrix defining a (variable) metric; a line search is usually carried out along p.

(b) an iteration of the revised simplex method on a linear program of dimension $m \times nb$. Here a central operation is the computation of Lagrange multipliers λ as components of a reduced gradient corresponding to constraints $Z_N \geq 0$. This is like temporarily reversing the roles of superbasics and nonbasics, and analogously to (5.3) gives

$$\lambda = [-(B^{-1}N)^{\mathrm{T}} | 0 | I_{nb \times nb}] \begin{pmatrix} g_B \\ g_S \\ g_N \end{pmatrix} \triangleq Z_N^{\mathrm{T}} g. \tag{5.5}$$

The components of λ computed from (5.5) are used differently from μ. They are used in a pricing out operation to revise the active set. Again λ can be conveniently computed using the above price vector Π by

$$\lambda = g_N - N^{\mathrm{T}}\Pi. \tag{5.6}$$

A special case of the reduced gradient method is the *convex simplex* method, see, for example, Zangwill (1969). Here $s = 0$, and

$$Z_N^{\mathrm{T}} = [-(B^{-1}N)^{\mathrm{T}} | I_{(n-m)\times(n-m)}]. \tag{5.7}$$

In place of (a) above we simply do a line search along a coordinate of Z_N determined from λ.

Let us now turn to the case of more immediate interest to us when $f(z)$ in (5.2) is nonsmooth. A single price vector Π can no longer be used to compute μ and λ. Instead the counterparts of (5.3) and (5.5) are to compute suitable vectors of steepest descent from the subdifferential[2] of $f(z)$ *in the appropriate reduced space.* This gives

$$\mu = \operatorname{argmin}\{\|\tilde{g}\|^2 : \tilde{g} = Z_S^{\mathrm{T}} g,\ g \in \partial f(z)\}$$

this can be computed as

$$\hat{g} = \operatorname{argmin}\{g^{\mathrm{T}}(Z_S Z_S^{\mathrm{T}})g : g \in \partial f(z)\}, \quad \mu = Z_S^{\mathrm{T}}\hat{g} \tag{5.8}$$

and

$$\lambda = \operatorname{argmin}\{\|\tilde{g}\|^2 : \tilde{g} = Z_N^{\mathrm{T}} g,\ g \in \partial f(z)\}$$

which can be computed as

$$\hat{g} = \operatorname{argmin}\{g^{\mathrm{T}}(Z_N Z_N^{\mathrm{T}})g : g \in \partial f(z)\}, \quad \lambda = Z_N^{\mathrm{T}}\hat{g}. \tag{5.9}$$

There is an alternative approach to (5.8) and (5.9) which is computationally more convenient but heuristic. Here a single 'gradient-like' vector is *first* computed from

$$\hat{g} = \operatorname{argmin}\{\|g\|^2 : g \in \partial f(z)\} \tag{5.10}$$

and μ and λ are then defined by

$$\mu = Z_S^{\mathrm{T}}\hat{g} = \hat{g}_S - S^{\mathrm{T}}\hat{\Pi}, \tag{5.11}$$

$$\lambda = Z_N^{\mathrm{T}}\hat{g} = \hat{g}_N - N^{\mathrm{T}}\hat{\Pi} \tag{5.12}$$

[2] We will not discuss here the need to replace $\partial f(x)$ by $\partial f_\varepsilon(x)$ the ε-subdifferential, which is needed in general for proofs of convergence, see Demyanov and Vasiliev (1981).

where $\hat{g} = (\hat{g}_B, \hat{g}_S, \hat{g}_N)$ and $\hat{\Pi}^T = \hat{g}_B^T B^{-1}$. Note that $\mu = 0$ implies $0 \in Z_S^T \partial f(z)$ and $\lambda = 0$ implies $0 \in Z_N^T \partial f(z)$ which are necessary optimality conditions. However, $\mu \neq 0$ is not necessarily a direction of descent in the reduced space. That is why we use the term heuristic for a straightforward extension of the convex-simplex or the reduced gradient method which uses (5.11) and (5.12). In order to ensure convergence one must modify this particular extension by drawing upon techniques of nonmonotonic optimization, see, for example, Shor (1983) or of bundle methods, see Lemarechal (1977).

Let us now come back to the application of these ideas to stochastic programs with complete recourse and nonstochastic tenders, and let us thus revert to our original notation (2.5). Wets (1976) first suggested a method for solving problems with simple recourse whose equivalent deterministic form is

$$\begin{aligned} \text{minimize} \quad & cx + \sum_{i=1}^{m_2} \Psi_i(\chi_i) \\ \text{s.t.} \quad & Ax = b, \\ & Tx - \chi = 0, \\ & x \geq 0. \end{aligned} \tag{5.13}$$

The subdifferential of $\Psi_i(\chi_i)$ is given by (3.1) and can be expressed as $\partial \Psi_i(\chi_i) = [c_i^-, c_i^+]$. It can be easily verified that this method is conceptually very closely related to the extension of the convex simplex method based upon (5.7), (5.10) and (5.12). An alternative is to use (5.7) and (5.9), which requires the solution of a suitable bound constrained quadratic program. In a similar vein, we could extend the reduced gradient method to solve (5.13) using either {(5.8), (5.9)} or {(5.10), (5.11), (5.12)}. In each case the special form of the objective function and constraints in (5.2) results in considerable simplification. For example, for the first possibility, namely, extension of the convex simplex method based upon (5.7), (5.10), (5.12), it is easily seen that (5.10) requires only the solution of

$$\text{for all } i, \text{ find} \quad c_i^- \leq \hat{g}_i \leq c_i^+ \quad \text{such that } |\hat{g}_i| \text{ is minimized,} \tag{5.14}$$

and this can be found explicitly.

This approach, based upon suitable adaptation and extension of MINOS carries across in a natural way to more general stochastic programming problems expressed in the form (2.5); the question of what information about the subdifferential can be provided becomes much more pressing, in particular, when we are outside the case of simple recourse. Very often it is necessary to resort to an approximation scheme that would provide upper and lower bounds for the solution, see Wets (1983b, Section 3) or accept the fact that the gradient can only be estimated such as in the methods of stochastic quasi-gradients, see Ermoliev (1983). We sketch out some possibilities in order to highlight the new obstacles that need to be overcome, and to stress the fact that there is a natural continuation of this approach that

provides solution procedures for more sophisticated stochastic programming problems.

Recall that $\Psi(\chi) = E\{\psi(\chi, w)\}$ and

$$\psi(\chi, w) = \inf_{y \in C} [q(y, w) \mid Wy = p(w) - \chi]$$

in this more general case, cf. Section 2. Let us consider only the linear case, i.e.,

$$\psi(\chi, w) = \inf_{y \geq 0} [q(w)y \mid Wy = p(w) - \chi]$$

$$= \sup_{\pi} [\pi(p(w) - \chi) \mid \pi W \leq q(w)].$$

For stochastic programs with complete recourse (that are bounded), $\psi(\cdot, w)$ is finite for all (possible) w and

$$\partial\Psi(\chi) = E\{\partial\psi(\chi, w)\}$$

where

$$\partial\psi(\chi, w) = \{-\pi \in \mathbb{R}^{m_2} \mid \pi \in \arg \max_{\pi W \leq q(w)} [\pi(p(w) - \chi)]\}.$$

Suppose the random variables have a discrete distribution with $p(\cdot)$ and $q(\cdot)$ taking on the values

$$\{(p^l, q^l), l = 1, \ldots, L\}$$

with probabilities f_l, $l = 1, \ldots, L$, then extensions of MINOS analogous to those described above for problem (5.13) must again solve structured quadratic programming problems. For example, the heuristic extension based upon (5.7), (5.10), (5.12) requires solution of

find $\nu \in \mathbb{R}^{m_2}$ such that $\nu^T \nu$ is minimized,

where

$$\nu^T = \sum_{l=1}^{L} f_l \pi^l \quad \text{and} \quad \pi^l W \leq q_l, \quad \pi^l(p^l - \chi) \geq \psi(\chi, (q^l, p^l)). \tag{5.15}$$

To solve this program efficiently we need to take advantage of its special structure, use the fact that for most l there is only a unique π^l that satisfies the inequalities and that for many l, π will be determined by the same basis of (W, I), and so on.

In general, when the random variables are not discretely distributed or when there are too many values for the discretely distributed random variables, it may not be possible to obtain complete information about $\Psi(\chi)$ or $\partial\Psi(\chi)$. We are then reduced to accepting approximates. There is at present no theory that allows us to deal directly with this case. What is needed is to extend the subgradient techniques, such as Demyanov (1968), Lemarechal (1978), Wolfe (1975) and, in particular, Bihain (1982) with appropriate reduced gradient calculations to handle this case. The convergence proofs could be derived by relying on the framework provided by nonlinear programming methods in the presence of noise, for example, Poljak (1978).

The question of how approximate the calculations of $\Psi(\chi)$ and $\partial\Psi(\chi)$ should be is still very much open to deeper investigation. We can even speculate at this point on the possibility of extension of the reduced gradient approach into the domain of stochastic quasi-gradient methods, see Ermoliev, 1983 which advocates the use of one or more sample points, say (p^s, q^s) to obtain $\partial\psi(\chi^s, (q^s, p^s))$ as a stochastic estimate of $\partial\Psi(\chi^s)$.

A more detailed discussion of some of the approaches outlined above can be found in Nazareth and Wets (1985).

6. Inner approximation

The algorithms we consider next use inner approximation of the type discussed in Section 4, see (4.7). After a general discussion of the algorithm, we consider first how it applies to problems with simple recourse and, as in Section 5, see how to extend the approach to more general classes of stochastic programs (with non-stochastic tenders).

The resulting algorithm is in effect the generalized programming technique, attributed by Dantzig (1963, Chapter 24) to P. Wolfe. Here we apply it to problems of type (2.5) taking advantage of the special structure and of the form of $\Psi(\chi)$. As a means of obtaining error bounds, Williams (1966) already suggested an approach of this nature, Parikh (1968) has described some of the details for simple recourse, and Ziemba (1972) gives particular application to a portfolio problem again with simple recourse, but apparently it has not been exploited as a general solution technique.

The algorithm as it applies to (1.1) or equivalently (2.5) can be summarized as follows:

Step 0. Find a feasible solution of $Ax^0 = b$, $x^0 \geq 0$

Set $\chi^0 = Tx^0$.

Choose $\chi^1, \ldots, \chi^\nu$ (a selection of tenders, $\nu \geq 0$).

Step 1. Solve the linear program:

$$\begin{aligned} \text{minimize} \quad & cx + \sum_{l=0}^{\nu} \lambda_l \Psi(\chi^l) = z \\ \text{subject to} \quad & Ax = b, \\ & Tx - \sum_{l=0}^{\nu} \lambda_l \chi^l = 0, \\ & \sum_{l=0}^{\nu} \lambda_l = 1, \\ & x \geq 0, \quad \lambda_l \geq 0 \text{ for } l = 0, \ldots, \nu. \end{aligned} \tag{6.1}$$

Let $(\sigma^\nu, \pi^\nu, \upsilon^\nu)$ be the (optimal) multipliers associated with the solution of (6.1).

Step 2. Find $\chi^{\nu+1} \in \operatorname{argmin}[\Psi(\chi) + \pi^\nu \chi]$.

If $\Psi(\chi^{\nu+1}) + \pi^\nu \chi^{\nu+1} \geq \upsilon^\nu$, stop: optimal.

Otherwise return to Step 1 with $\nu = \nu + 1$.

We have assumed here that for all π^ν generated in Step 1, the function $\chi \mapsto (\Psi(\chi) + \pi^\nu \chi)$ attains its minimum. There are naturally regularity conditions for stochastic programs that will guarantee this (Wets, 1974; Williams, 1966) but mostly we have done so to simplify the presentation and interpretation of the algorithm. Note that both upper and lower bounds for the infimum are available. Let z^ν denote the optimal value of z, and $(\lambda_l^\nu, l = 0, \ldots, \nu)$ the optimal values of the λ variables in (6.1). Then

$$z^\nu + [\Psi(\chi^{\nu+1}) + \pi^\nu \chi^{\nu+1}] - \sum_{l=0}^{\nu} \lambda_l^\nu [\Psi(\chi^l) + \pi^\nu \chi^l] \leq z^* \leq z^\nu \tag{6.2}$$

where z^* is the optimal value of the original program. The second inequality follows from the fact that (6.1) is an inner approximation, whereas the first one follows from Step 2 which implies that

$$-\pi^\nu \chi + [\Psi(\chi^{\nu+1}) + \pi^\nu \chi^{\nu+1}] \leq \Psi(\chi).$$

Adding cx and taking inf on both sides with respect to (x, χ) on the set $\{x \geq 0 \mid Ax = b, Tx = \chi\}$ yields the desired inequality, it suffices to observe that the first one of these two minimization problems admits for optimal solution the pair $(x^\nu, \sum_{l=0}^{\nu} \lambda_l^\nu \chi^l)$ with $(x^\nu, \lambda_l^\nu, l = 0, \ldots, \nu)$ the optimal solution of (6.1). Thus

$$0 \leq z^\nu - z^* \leq \max_{k=0,\ldots,\nu} \left[(\Psi(\chi^{\nu+1}) + \pi^\nu \chi^{\nu+1}) - \sum_{l=0}^{\nu} \lambda_l^\nu (\Psi(\chi^l) + \pi^\nu \chi^l) \right].$$

We interpret the algorithm as the search for a particular (optimal) tender χ^*. It is easy to see that if χ^* is part of the collection $\chi^0, \ldots, \chi^\nu$, then solving (6.1) will yield the optimal x^*. One reason for believing that this approach holds promise, is that in practice one should be able to initialize the algorithm with a good choice of tenders $\chi^0, \ldots, \chi^\nu$. The subsequent iterations can then be viewed as refinements of the original guesses. A line of further research is to find effective strategies for choosing initial tenders, see Birge and Wets (1983).

The convergence of the algorithm, with the following assumptions

(i) all tenders are retained, as part of (6.1), (6.3)

(ii) complete information is available about the function values of Ψ so that Step 2 can be carried out exactly, (6.4)

has been proved by Dantzig (1963, Chapter 24). Further, the algorithm applied to the convex program (2.5) is equivalent to a cutting plane algorithm applied to its dual. We can thus translate the results about retention of cuts (Topkis, 1970; Eaves

and Zangwill, 1971) into retention of tenders. In particular in our case, they imply that under suitable assumptions all tenders not associated with a basic variable λ_l can be dropped at the next iteration, without affecting the convergence proof of the algorithm.

A large number of tenders could be generated, although this is very unlikely in practice, especially if a good set of initial tenders is used. From a theoretical standpoint, however, and for reasons of sound implementation, it is worth examining the question of which tenders should be retained to enhance convergence. At iteration ν with the multipliers $(\sigma^\nu, \pi^\nu, \upsilon^\nu)$ we have for all tenders χ^l, $l=0,\ldots,\nu$,

$$\Psi(\chi^l)+\pi^\nu\chi^l \geq \upsilon^\nu. \tag{6.5}$$

At the next iteration, a tender $\chi^{\nu+1}$ is developed (several tenders could equally well be formed) and we need to resolve (6.1) with respect to $\chi^0,\ldots,\chi^{\nu+1}$. Suppose prior to the commencement of the next iteration $\nu+1$, a subset of tenders

$$\{\chi^l, l\in L\}\subset\{\chi^0,\ldots,\chi^{\nu+1}\}$$

must be found such that the optimal solution of (6.1) is unaffected. Since the (optimal) multipliers

$$(\sigma^{\nu+1}, \pi^{\nu+1}, \upsilon^{\nu+1})$$

are unknown at this stage, we formulate this problem as

$$\begin{aligned}&\text{given any } (\sigma,\pi,\upsilon) \text{ and a fixed index } k\\ &\text{find}\quad L\subset\{0,\ldots,\nu+1\}\quad\text{such that}\\ &\Psi(\chi^l)+\pi\chi^l\geq\upsilon\quad\text{for all } l\in L\\ &\text{implies}\quad \Psi(\chi^k)+\pi\chi^k\geq\upsilon.\end{aligned}\tag{6.6}$$

Let us write

$$D^l=\begin{bmatrix}\Psi(\chi^l)\\ \chi^l\\ -1\end{bmatrix},\quad b=\begin{bmatrix}\Psi(\chi^k)\\ \chi^k\\ -1\end{bmatrix}\quad\text{and}\quad I^1=\begin{bmatrix}1\\0\\0\end{bmatrix}$$

with $D=[D^0,\ldots,D^{\nu+1}]$.

6.7. Proposition. *A sufficient condition that*

$$\Psi(\chi^l)+\pi\chi^l\geq\upsilon\quad \textit{for all } l\in L \tag{6.8}$$

implies

$$\Psi(\chi^k)+\pi\chi^k\geq\upsilon \tag{6.9}$$

is that

$$b\in\operatorname{pos}[I^1; D^l, l\in L],$$

i.e. that b belongs to the positive hull of (or equivalently the convex cone, generated by) the columns of D corresponding to L and I^1.

Proof. To say that $b \in \text{pos}[I^1; D^l, l \in L]$ is to say that the linear system

$$b = \sum_{l \in L} D^l y_l + I^1 \alpha, \quad \alpha \geq 0, \quad y_l \geq 0, \quad l \in L,$$

is solvable. Thus the system

$$(\lambda, \pi, \sigma) D^l \geq 0, \quad l \in L, \quad \lambda \geq 0 \text{ and } (\lambda, \pi, \upsilon) \cdot b < 0$$

is not solvable, as follows from Farkas Lemma. Using now the definitions of D^l and b, we see that this implies that for a choice of variables $(\lambda = 1, \pi, \upsilon)$ satisfying (6.8), we necessarily must satisfy (6.9). □

The question raised in (6.6) can thus be translated into finding a minimum number of generators, i.e., a *frame*, for the convex polyhedral cone

$$\text{pos}[I^1; D^l, l = 0, \ldots, \nu + 1].$$

An algorithm for doing this is described by Wets and Witzgall (1967). Note also that it may be worthwhile to also eliminate tenders that have not been utilized in the solution of (6.1) on several prior iterations.

The use of this algorithm in the context of stochastic programming makes assumption (6.4) nontrivial. Even in the case of simple recourse, situations can arise when $\Psi(\chi)$ cannot be calculated exactly or the cost of calculating it could be excessive. For example, if $q(y, w)$ is nonlinear in y and the dependence on w is not simple (e.g. linear), the cost of evaluating

$$E\{q(p(w) - \chi, w)\}$$

could be very large. A similar situation could arise even after approximating the distribution functions by piecewise constant or piecewise linear distributions. In this case the generalized linear programming approach must be revised to include noisy functions and the question of convergence, both theoretical and practical, still needs further investigation.

In the case of simple recourse with separable cost the evaluation of the function Ψ presents no serious challenge since

$$\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i)$$

and each Ψ_i defined on $\mathbb{R}$ is given by a one-dimensional integral, viz.

$$\Psi_i(\chi_i) = \int q_i(p_i(w) - \chi_i, w) F_i(\mathrm{d}w)$$

with the subgradient given by (3.1). Special forms of q_i and F_i lead to even simpler representations for Ψ such as (3.3). Even more explicit is the expression obtained

in (3.4) and (3.5) in the case of piecewise linear recourse costs and piecewise constant distributions; for piecewise linear recourse costs and piecewise linear distributions see Wets (1975, Section 3); for even more detailed expressions for specific distributions consult Hansotia (1977). Note also that in this case Step 2 of the algorithm consists in finding for $i = 1, \ldots, m_2$, $\chi_i^{\nu+1}$ such that

$$-\pi_i^{\nu} \in \partial \Psi_i(\chi_i^{\nu+1}),$$

where the subgradient is given by (3.1). Again, in many cases it is possible to use the special forms of q_i and F_i to find efficient solutions procedures for the preceding relation. For example, in the situation covered by (3.3), the above becomes: find $\chi_i^{\nu+1}$ such that

$$\frac{q_i^+ - \pi_i^{\nu}}{q_i} \in [F_i^-(\chi_i^{\nu+1}), F_i(\chi_i^{\nu+1})].$$

It thus suffices to have a bracketing routine for finding the point at which the monotone function F_i passes through the value $(q_i^+ - \pi_i^{\nu})/q_i$. For some of these more simple cases it is easy to show that the generalized linear programming algorithm has finite termination.

In the more general case, when it is not feasible to compute the value of Ψ at χ exactly, see Section 3, there are basically two strategies available. The first one is to accept inaccurate evaluations of Ψ, view them as noisy observations of Ψ and rely on a convergence in probability argument (Poljak, 1978) with Step 2 being performed using, for example, the method of stochastic quasi-gradients, Ermoliev (1983). How to design an efficient and reliable algorithm that proceeds in this fashion has not been investigated yet.

The second approach is to proceed by approximations. By this we mean replace the original problem (2.1) by an approximate one, solve the approximating problem, obtain if possible bounds using this approximating solution and repeat the process with a refinement of the approximation if the bounds are not sufficiently tight. The subject of approximations, specially via discretization of the random variables, is reviewed in Wets (1983b) and will not be taken up here. We only want to raise some of the questions that need to be resolved before such a scheme could be made operational:

(i) How should the initial approximation be designed so as to obtain with minimal computational effort a 'good' approximate of the solution?

(ii) How to improve (refine) the approximation so as to 'maximize' the resulting improvement?

(iii) How to blend in, these successive approximations with the steps of the algorithm?

With regard to implementation and computational experience for the case of simple recourse see Nazareth (1983) and Nazareth and Wets (1985). See also Nazareth, 1983 for some further results and discussion of the complete recourse case.

7. Outer linearization

The third class of algorithms that we consider is based upon the outer approximation approach described in Section 4, see (4.9). We deal with this technique somewhat more briefly because for problems with simple recourse it appears at this time to be more limited in scope, whereas for more general classes of stochastic programs this approach is very close to the L-shaped algorithm (Van Slyke and Wets, 1969) which has already been studied extensively in the stochastic programming setting (Birge, 1982; Wets, 1983b, Section 2).

Consider first simple recourse with separable piecewise linear recourse cost and discrete random variables. From (3.4), the objective of the equivalent deterministic program is

$$cx + \sum_{i=1}^{m_2} \Psi_i(\chi_i), \tag{7.1}$$

where each Ψ_i is polyhedral with slopes s_{ih} and $s_{i0} < s_{i1} < \cdots < s_{i,k_i}$. Then using the outer approximation approach of Section 4, in particular (4.9) applied to each Ψ_i defined by (3.4) we obtain the counterpart of (5.1), namely

$$\begin{aligned}
&\text{find} \quad x \in \mathbb{R}^{n_1}_+,\ \chi \in \mathbb{R}^{m_2} \text{ and } \nu \in \mathbb{R}^{m_2} \quad \text{such that} \\
&\qquad Ax = b, \\
&\qquad Tx - \chi = 0, \\
&\qquad \nu_i \geq s_{il}\chi_i + e_{il}, \quad l = 0, \ldots, k_i,\ i = 1, \ldots, m_2, \text{ and} \\
&\qquad cx + \sum_{i=1}^{m_2} \nu_i \text{ is minimized.}
\end{aligned}$$

Since most of the constraints involving ν_i will be slack at optimality, a linear programming algorithm utilizing an active set strategy immediately suggests itself.

For the more general case when the above assumptions on linearity of the recourse objective and on the distribution are dropped but the recourse is still simple and the recourse objective separable, we again have the objective of the equivalent deterministic program of the form (7.1); now $\Psi_i(\chi_i)$ is no longer polyhedral. Let us assume that complete information is available about values and derivatives of $\Psi_i(\chi_i)$. Consistent with (4.8), we assume that we have the following representation for each Ψ_i.

$$\Psi_i(\chi_i) = \max_{j \in J_i} \Psi_{ij}(\chi_i)$$

where each Ψ_{ij} is a convex differentiable function and J_i a finite set of indices. For each χ_i, the value of $\max_{j \in J_i} \Psi_{ij}(\chi_i)$ is attained for a finite set $A(\chi_i) \subset J_i$ known as the *active set*, i.e., we are dealing with stochastic programs whose equivalent deterministic forms have (possibly) nondifferentiable objectives with explicitly known subdifferentials.

Problem (2.7) can thus be stated as

$$\begin{aligned}
&\text{find} \quad x \in \mathbb{R}^{n_1}_+,\ \chi \in \mathbb{R}^{m_2} \text{ and } \nu \in \mathbb{R}^{m_2} \quad \text{such that} \\
&Ax = b, \\
&Tx - \chi = 0, \\
&-\Psi_{ij}(\chi_i) + \nu_i \geq 0 \quad \text{for } j \in J_i,\ i = 1, \ldots, m_2, \text{ and} \\
&cx + \sum_{i=1}^{m_2} \nu_i \text{ is minimized.}
\end{aligned} \tag{7.2}$$

When $A(\chi_i)$ and $\partial\Psi_i(\chi_i)$ are known explicitly, the functions Ψ_{ij} can be obtained systematically and as needed.

One can utilize the method of successive linear approximation; see Frank and Wolfe (1956); Madsen and Schjaer-Jacobsen (1978), for example, with differentiable Ψ_i. This requires the solution of a sequence of linear programming problems of the form

$$\begin{aligned}
&\text{find} \quad x \in \mathbb{R}^{n_1}_+,\ \chi \in \mathbb{R}^{m_2} \text{ and } \nu \in \mathbb{R}^{m_2} \quad \text{such that} \\
&Ax = b, \\
&Tx - \chi = 0, \\
&-\nabla\Psi_i(\chi_i) \cdot \chi_i + \nu_i \geq (\Psi_i(\chi_i) - \nabla\Psi_i(\chi_i) \cdot \chi_i) \quad \text{for } i = 1, \ldots, m_2 \quad \text{and} \\
&cx + \sum_{i=1}^{m_2} \nu_i \text{ is minimized.}
\end{aligned} \tag{7.3}$$

The next approximation is obtained by linearization of the Ψ_i at $\chi_i^{\nu+1}$ where $\chi^{\nu+1}$ is the optimal value for χ in (7.3). (It is necessary to use an additional constraint to restrict the step size when implementing this.)

When second order differentiables (or good approximates thereof) can be computed, such as when the recourse costs and marginal distribution functions are piecewise linear (Wets, 1975) or other cases dealt with in the beginning of Section 3, one could proceed via quadratic approximations, as proposed for nonlinear programming by Wilson (1963) and Han (1976). Good reviews of both approaches can be found in Fletcher (1981), Powell (1978), Wierzbicki (1982); see also Han (1981). The code of Schittkowski (1980) to solve nonlinear programming problems by successive approximations using quadratic programs should also be studied for implementation in this setting. Because of the second order information necessary to carry out the steps, it may however only be possible to use it for a special class of stochastic programs with simple recourse.

In the more general situation when Ψ is not necessarily separable or smooth, and only subgradients can be calculated (or approximates thereof), one could consider *a cutting plane algorithm.* This would, for example, involve a sequence of

problems of the form

$$\begin{aligned} &\text{find} \quad x \in \mathbb{R}^{n_1}_+, \ \chi \in \mathbb{R}^{m_2} \text{ and } \nu \in \mathbb{R} \quad \text{such that} \\ &Ax = b, \\ &Tx - \chi = 0, \\ &-\pi^l \chi + \nu \geq (\Psi(\chi^l) - \pi^l \chi^l), \quad l = 1, \ldots, \nu, \quad \text{and} \\ &cx + \nu \ \text{is minimized.} \end{aligned} \tag{7.4}$$

Here $\pi^l \in \partial \Psi(\chi^l)$ for $l = 1, \ldots, \nu$. The solution of (7.4) yields a new tender $\chi^{\nu+1}$; we then need to compute $\pi^{\nu+1}$ and $\Psi(\chi^{\nu+1})$. This defines a new constraint to be added to (7.4). More sophisticated strategies based upon utilizing higher order information are given by Womersley (1981).

When the equation $Tx = \chi$ is used to delete χ from the formulation of (7.4) we are precisely in the L-shaped format Wets (1983b, Section 2) for which we already have experimental codes (Birge, 1982). It is, however, important to realize that introduction of the tender χ in (7.4) is useful from a computational viewpoint, since it will result in few nonzero elements in the representation of the problem. Note however that the associated cuts $-\pi^l \chi + \nu \geq \Psi(\chi^l) - \pi^l \chi^l$ are still dense.

In conclusion, we should point out that the algorithms that have been investigated in the preceding sections are those that seemed to us to be the most promising and most readily implementable. Clearly, other directions of algorithmic development are possible and should be pursued, see for example, Kallberg and Ziemba (1981a) for an approach based upon the Frank-Wolfe algorithm and Hogan (1973) and Kallberg and Ziemba (1981b) for an approach that uses directional derivatives in place of gradients.

Acknowledgements

We would like, in particular, to thank Drs. Hien Nguyen, Andre Bihain and Michael Saunders for their useful comments on an earlier version of this paper.

References

A. Bihain, "G.R.S.G. A general purpose N.D.O. Code", Technical Report, University of Namur (Namur, 1978).

J. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs", Technical Report 82-6, University of Michigan (Ann Arbor, 1982).

J. Birge and R. Smith, "New Monte-Carlo procedures for numerical integration", Technical Report, University of Michigan (Ann Arbor, 1982).

J. Birge and R. Wets, "Approximations and error bounds in stochastic programming", in: Y. Tong, ed., *Proceedings of the conference on inequalities in statistics and probability* (Institute of Mathematical Statistics, Hayward, CA, 1984).

H. Cleef, "A solution procedure for the two-stage stochastic program with simple recourse", *Zeitschrift für Operations Research* 25 (1981) 1-13.

G.B. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, NJ, 1963).

I. Deak, "Three digit accurate multiple normal probabilities", *Numerische Mathematik* 24 (1968) 342-380.

V.F. Demyanov, "Algorithms for some minimax problems", *Journal of Computer and System Sciences* 2 (1968) 342-380.

V.F. Demyanov and L.V. Vasiliev, *Nondifferentiable optimization* (Nauka, Moscow, 1981) [In Russian].

B.C. Eaves and W. Zangwill, "Generalized cutting plane algorithms", *SIAM Journal on Control* 9 (1971) 529-542.

Yu. Ermoliev, "Stochastic quasi-gradient methods and their application in systems optimization", *Stochastics* 8 (1983).

R. Fletcher, *Practical methods of optimization, Vol. 2: Constrained Optimization* (Wiley, Chichester, 1981).

M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3 (1956) 95-110.

S.P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems", *Mathematical Programming* 11 (1976) 263-282.

S.P. Han, "Variable metric methods for minimizing a class of nondifferentiable functions", *Mathematical Programming* 20 (1981) 1-12.

M. Hanscom, J.J. Strodiot and V.H. Nguyen, "Une approche de type gradient-reduit en optimization non-différentiable pour l'ordonnancement a moyen terme de la production d'énergie électrique", manuscript, 1981.

B. Hansotia, "Some special cases of stochastic programs with recourse", *Operations Research* 25 (1977) 361-363.

J.-B. Hiriart-Urruty, "Conditions necessaires d'optimalité pour un programme stochastique avec recourse", *SIAM Journal on Control and Optimization* 16 (1978) 125-130.

W. Hogan, "Directional derivatives for extremal-value functions with applications to the completely convex case", *Operations Research* 21 (1973) 188-209.

P. Kall, "Stochastic programming", *European Journal of Operations Research* 10 (1982) 125-130.

J. Kallberg and M. Kusy, "Code instruction for S.L.P.R., a stochastic linear program with simple recourse", Technical Report, University of British Columbia (Vancouver, 1976).

J. Kallberg and W. Ziemba, "An extended Frank-Wolfe algorithm with application to portfolio selection", in: P. Kall and A. Prekopa, eds., *Recent progress in stochastic programming* (Springer-Verlag, Berlin, 1981a).

J. Kallberg and W. Ziemba, "An algorithm for portfolio revision: theory, computation algorithm and empirical results", *Applications of Management Science* 1 (1981b) 267-291.

J. Kallberg, R. White and W. Ziemba, "Short term financial planning under uncertainty", *Management Science* 28 (1982) 670-682.

M. Kusy and W. Ziemba, "A bank asset and liability management model", Technical Report, University of British Columbia (Vancouver, 1981).

C. Lemarechal, "Bundle methods in nonsmooth optimization", in: C. Lemarechal and R. Mifflin, eds., *Nonsmooth optimization*, IIASA Proceedings Series, Vol. 3 (International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977) pp. 79-102.

C. Lemarechal, "Nonsmooth optimization and descent methods", Research Report 78-4, International Institute for Applied Systems Analysis (Laxenburg, Austria, 1978).

K. Madsen and H. Schjaer-Jacobsen, "Linearly constrained minimax optimization", *Mathematical Programming* 14 (1978) 208-223.

B. Murtagh and M. Saunders, "Large-scale linearly constrained optimization", *Mathematical Programming* 14 (1978) 41-72.

K. Murty, "Linear programming under uncertainty: a basic property of the optimal solution", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 10 (1968) 284-288.

L. Nazareth, "Algorithms based upon generalized linear programming for stochastic programs with recourse", to appear in Proceedings of IFIP Workshop on stochastic programming: algorithms and applications (Gargnano, Italy, September 16-21, 1983).

L. Nazareth and R.J.-B. Wets, "Stochastic programs with recourse: algorithms and implementation", IIASA Working Paper 1985 (forthcoming).

P. Olsen, "Multistage stochastic programming with recourse: the equivalent deterministic program", *SIAM Journal on Control and Optimization* 14 (1976) 495-517.

S.C. Parikh, "Lecture notes on stochastic programming", University of California (Berkeley, CA, 1968).

P. Poljak, "Nonlinear programming methods in the presence of noise", *Mathematical Programming* 14 (1978) 87-97.

M.J.D. Powell, "Algorithms for nonlinear constraints that use Lagrangian functions", *Mathematical Programming* 14 (1978) 224-248.

K. Schittkowski, *Nonlinear programming codes*, Lecture Notes in Economics and Mathematical Systems (Springer-Verlag, Berlin, 1980).

N.Z. Shor, "Generalized gradient methods of nondifferentiable optimization employing space dilatation operators", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical Programming: The state of the art* (Springer-Verlag, Berlin, 1983) pp. 501-529.

D. Topkis, "Cutting plane methods without nested constraint sets", *Operations Research* 18 (1970) 403-413.

R. Van Slyke and R. Wets, "*L*-shaped linear program with applications to optimal control and stochastic linear programs", *SIAM Journal on Applied Mathematics* 17 (1969) 638-663.

D. Walkup and R. Wets, "Stochastic programs with recourse: Special forms", in: H. Kuhn, ed., *Proceedings of the Princeton symposium on mathematical programming* (Princeton University Press, Princeton, 1970) pp. 139-162.

R. Wets and C. Witzgall, "Algorithms for frames and linearity spaces of cones", *Journal of Research of National Bureau of Standards, Section B* 71B (1967) 1-7.

R. Wets, "Stochastic programs with recourse: a basic theorem for multistage problems", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandete Gebiete* 21 (1972) 201-206.

R. Wets, "Stochastic programs with fixed recourse: the equivalent deterministic program", *SIAM Review* 16 (1974) 309-339.

R. Wets, "Solving stochastic programs with simple recourse II", *Proceedings of Johns Hopkins Symposium on Systems and Information*, 1975, pp. 1-6.

R. Wets, "A statistical approach to the solution of stochastic programs with (convex) simple recourse", manuscript, Department of Mathematics, University of Kentucky (Lexington, 1976).

R. Wets, "Solving stochastic programs with simple recourse", *Stochastics* 10 (1983a) 219-242.

R. Wets, "Stochastic programming: solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming: The state of the art* (Springer-Verlag, Berlin, 1983b) pp. 566-603.

A. Wierzbicki, "Lagrangian functions and nondifferentiable optimization", in: E. Nurminski, ed., *Progress in nondifferentiable optimization*, IIASA Collaborative Proceedings Series, CP-82-58 (1982) pp. 173-213.

A.C. Williams, "Approximation formulas for stochastic linear programming", *SIAM Journal on Applied Mathematics* 14 (1966) 668-677.

P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions", *Mathematical Programming Study* 3 (1975) 145-173.

R. Womersly, "Numerical methods for structural problems in nonsmooth optimization", Doctoral Thesis, University of Dundee, 1981.

W.I. Zangwill, *Nonlinear programming: A unified approach* (Prentice-Hall, Englewood Cliffs, 1969).

W.T. Ziemba, "Computational algorithms for convex stochastic programming with simple recourse", *Operations Research* 18 (1970) 415-431.

W.T. Ziemba, "Solving nonlinear programming problems with stochastic objective functions", *Journal of Financial and Quantitative Analysis* VII (1972) 1809-1827.

W.T. Ziemba, "Stochastic programs with simple recourse", in: P. Hammer and G. Zoutendijk, *Mathematical programming in theory and practice* (North-Holland, Amsterdam, 1974) pp. 213-274.

Mathematical Programming Study 28 (1986) 29-47
North-Holland

DECOMPOSING THE REQUIREMENT SPACE OF A TRANSPORTATION PROBLEM INTO POLYHEDRAL CONES

Stein W. WALLACE
Chr. Michelsen Institute, Fantoftvegen 38, N-5036 Fantoft, Norway

Received 5 January 1984
Revised manuscript received 30 August 1984

Solving a large number of linear programs with the same coefficient matrix, but different cost coefficients or right hand sides is often an important part of the solution procedures for dynamic and stochastic programs. This usually involves a decomposition of the parameter space, a decomposition which is dictated by the optimality criteria. The subregions correspond to (sub)bases of the coefficient matrix. We suggest an efficient procedure to generate such a decomposition in the following case: the linear programs are transportation problems with different right hand sides. We also report on our computational results.

Key words: Stochastic Programming, Transportation, Decomposition, Dual Enumeration, Polyhedral Cones.

1. Introduction

A linear programming problem is feasible if and only if the right hand side can be written as a nonnegative combination of the columns of the coefficient matrix. Or to put it another way, if and only if the right hand side is an element of the polyhedral cone generated by the columns of the coefficient matrix.

The purpose of this report is to give an algorithm for decomposing the polyhedral cone generated by the columns of the coefficient matrix of a transportation problem into a set of M-dimensional cones (where M is the number of rows in the matrix). These cones will have the property that all vectors (i.e. possible right hand sides) that belong to the same cone, will have the same dual solution (with respect to a given objective function).

This decomposition is useful if the same system is going to be solved for a large number of right hand sides. A clear-cut application of this is within stochastic programming with recourse when the recourse problem is a transportation problem such as in [4].

Because of the duality theorem and the above-mentioned property of the dual solutions within each cone, a method using decomposition might be an easy way to find the expected value of the objective function and to produce a subgradient. This is of course only true if it is easier to find the correct cone for a given right hand side than it is to find the solution directly. We will return to this in Section 5 of this paper.

When the problem at hand is large, we cannot generate all the cones in the decomposition. We therefore investigate the problem of characterizing the differences between 'large' and 'small' cones, such that we are able to create those that cover the largest collection of possible right hand sides.

The paper is organized as follows: In Section 2 we give the necessary formal definitions. The decomposition algorithm is presented in Section 3, with Section 4 outlining the problems connected to dual degeneracy. How to check if a given right hand side belongs to a given cone is explained in Section 5. We suggest a way of organizing the data in Section 6. Thereafter follow two sections on the number of cones and an attempt to define a 'large' cone. Finally in Section 9 we give an example.

2. Formal definition

Given a transportation problem with M_1 supply points and M_2 demand points, and making sure that total demand equals total supply, we get the following formulation when we have left out the redundant equation for supply point 1.

$$\text{Min} \quad \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} c_{ij} y_{ij}$$

subject to

$$\sum_{j=1}^{M_2} y_{ij} = S_i, \quad i = 2, \ldots, M_1,$$

$$-\sum_{i=1}^{M_1} y_{ij} = -D_j, \quad j = 1, \ldots, M_2,$$

$$y_{ij} \geqslant 0.$$

Let the coefficient matrix of the system be called N. It will have $(M_1 + M_2 - 1) = M$ rows and $M_1 * M_2 = N$ columns.

The following 2 definitions and theorem are taken from [6], the proof can be found in [3].

Definition 1. pos $N = \{t: Ny = t, y \geqslant 0\}$.

Definition 2. A *finite closed polyhedral complex* H will be any finite collection of closed convex polyhedra, called *cells* of H, such that

(i) if C is a cell of H, then every closed face of C is a member of H;

(ii) if C_1 and C_2 are distinct cells of H, then either they are disjoint, or one is a face of the other, or their intersection is a common face.

Theorem 3 (Basis Decomposition Theorem). *Let $P(t)$ denote the linear program*

$$\textit{Minimize} \quad cy$$

$$\textit{subject to} \quad Ny = t,$$

$$y \geqslant 0,$$

where c is fixed and N is a fixed $M \times N$ matrix of rank M. Then:

(i) *$P(t)$ is feasible if and only if t lies in* pos *N.*

(ii) *Either $P(t)$ is bounded for all t in* pos *N or $P(t)$ is unbounded for all t in* pos *N.*

(iii) *If $P(t)$ is bounded, there exists a decomposition of* pos *N into a finite closed polyhedral complex H whose cells are simplicial cones with vertex at the origin, and a one-to-one correspondence between the one-dimensional cells of H and selected columns of N which generate them, such that*

(a) *the closed M-dimensional cells of H cover* pos *N, and*

(b) *the M columns of N associated with the edges of a closed M-dimensional cell C of H constitute an optimal basis for all t in C.*

Our task is then to find such a finite closed polyhedral complex for the uncapacitated transportation problem. With 'cell' we shall hereafter always refer to *M*-dimensional cells of *H*.

We will now assume that the system is nondegenerate (it will be made clear what that means). In Section 4 we shall return to the problem without a nondegeneracy assumption and study a perturbation technique.

3. The decomposition algorithm

The method is based on the dual simplex method. Each cell is pos *N* represents a dual feasible basis. Going from one cell to another (from one basis to an adjacent basis) requires that one first picks an outgoing column and then finds the appropriate entering column. By doing this for all basic columns, we will find all the neighbour cells. Provided that they all exist, nondegeneracy tells us that there will be *M* of them.

The dual simplex method is oriented towards pivot selection in selected rows of the simplex tableau. The first question to answer is therefore: What is the significance of a row in the simplex tableau? In order to answer this we will refer to a result from [2].

Let a basis *B* be given in a network. Let arc(i, j) (arc (i, j) is the arc that goes from node i to node j) be a nonbasic arc, and let γ_{ij} be the cycle created when we add arc (i, j) to the basic tree. See Fig. 1. Furthermore, let arc(i, j) define the direction of the cycle, i.e. arc(i, j) is a forward arc. Let the $(n-1)$-vector $\mu(\gamma_{ij})$ be defined as

$$\mu_k(\gamma_{ij}) = \begin{cases} 1 & \text{if basic arc } k \text{ is a forward arc in } \gamma_{ij}, \\ -1 & \text{if basic arc } k \text{ is a reverse arc in } \gamma_{ij}, \\ 0 & \text{if basic arc } k \text{ is not in } \gamma_{ij}. \end{cases}$$

In our example in Fig. 1 assume the arcs have the following numbering: $(1, 2) \to 1$, $(3, 5) \to 2$ and $(1, 3) \to 3$. Then

$$\mu(\gamma_{25}) = (1, -1, -1, \mathbf{0})$$

where **0** represents all other basic arcs, since arc$(1, 2)$ is a forward arc and $(3, 5)$ and $(1, 3)$ are reverse arcs.

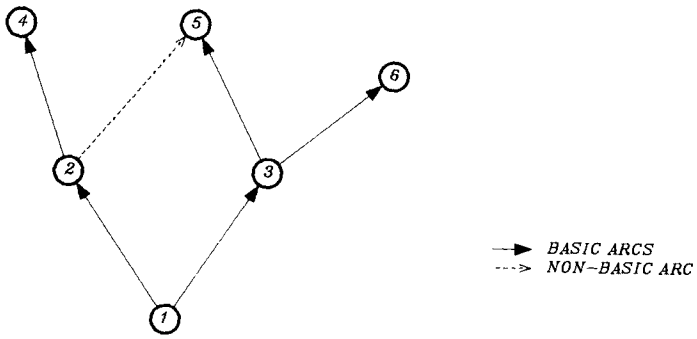


Fig. 1. The nonbasic arc(2, 5) creates a cycle $\gamma_{25} = (2, 5, 3, 1, 2)$ in the basic tree.

If we denote the column corresponding to arc(i, j) by α_{ij}, we then have

$$\mu(\gamma_{ij}) = -\boldsymbol{B}^{-1}\alpha_{ij} \tag{1}$$

i.e. the simplex tableau shows which arcs are in the cycle created by arc (i, j) and which orientation they have.

Therefore, the significance of row k is that if $a_{k,ij} = +1$ ($a_{k,ij}$ is the kth element of the column in the simplex tableau representing the arc from node i to node j), basic arc no. k is a reverse arc (note the minus sign in (1)) on the cycle created by arc(i, j), if $a_{k,ij} = -1$ it is a forward arc and if $a_{k,ij} = 0$ basic arc no. k is not part of the cycle. In other words, only arcs that use the leaving arc as forward arc in their cycles are candidates for entering the basis. (In the dual simplex method only negative entries in a row are candidates to become pivot elements).

So far we have not used any properties from the transportation structure of our problem. First we note that our graph is bipartite and that all supply points are connected with all demand points. See Fig. 2.

Proposition 4. *All cycles created by nonbasic arcs have an even number of arcs.*

Proof. *Follows from the fact that no arcs connect two supply or two demand points.*

Proposition 5. *Every second arc in all cycles created by nonbasic arcs is a forward arc.*

Therefore we can give an algorithm to find all arcs satisfying $a_{k,ij} < 0$ ($a_{k,ij} = -1$) because of the following observations.

Proposition 6. *The entering arc must not be connected to the leaving arc by an even number of basic arcs.*

Algorithm

(0) Let $S = \emptyset$, $T = \{\text{all nodes}\}$.

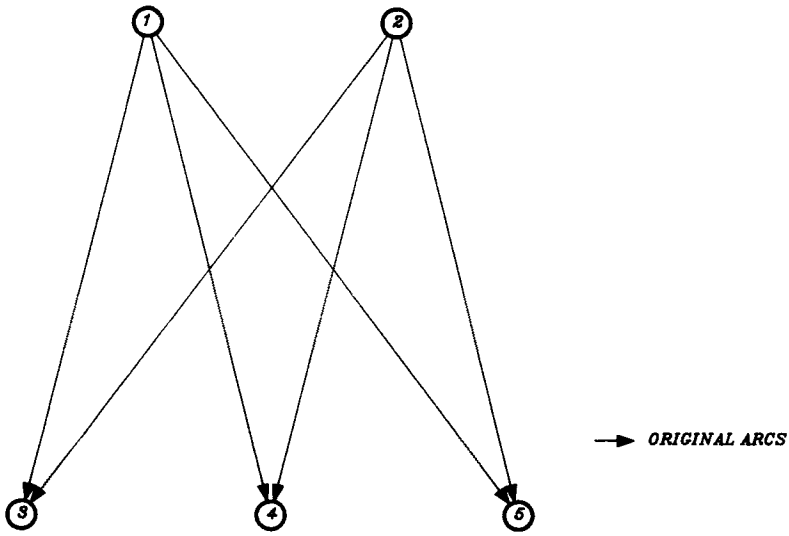


Fig. 2. A transportation network.

(1) Define the head and tail nodes of the leaving arc to be blocked and put their indices into the set S: Let $T = T - S$.
(2) Find all nodes in T connected to nodes in S by basic arcs. Let these be open and redefine S to contain only these open nodes. Let $T = T - S$. If $T = \emptyset$, stop, otherwise go to 3.
(3) Find all nodes in T connected to nodes in S by basic arcs. Define these as being blocked nodes, and redefine S to contain only these blocked nodes. Let $T = T - S$. If $T = \emptyset$, stop, otherwise go to 2.

When this algorithm terminates, a certain set of nodes are open and another set is blocked, see Fig. 3. All arcs that have both their originating and terminating node open, satisfy $a_{k,ij} < 0$. In Fig. 3 these are arcs $(1, 4)$ and $(3, 4)$. Note also that all basic arcs automatically are excluded from this set since one of their end nodes will be blocked. Furthermore, non-basic arcs which create a cycle in the basis not containing arc k will also be excluded since they automatically have a blocked node in one end and an open node in the other. This follows from the fact that all cycles

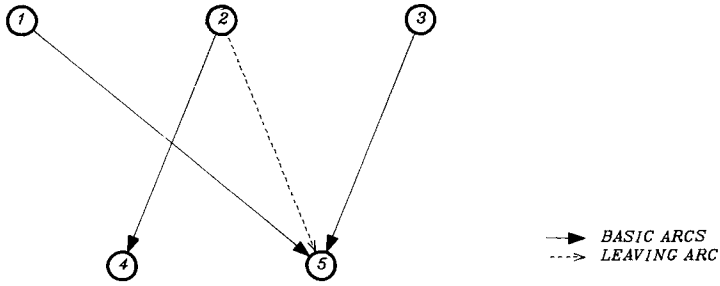


Fig. 3. With arc(2, 5) as leaving arc, nodes 2 and 5 are blocked and nodes 1, 3 and 4 are open.

are of even length and nodes along cycles without arc k will alternate as open and closed nodes.

According to the dual simplex method we are seeking the arc(i, j) such that

$$\frac{\bar{c}_{ij}}{a_{k,ij}} = \max_{a_{k,rs}<0} \frac{\bar{c}_{rs}}{a_{k,rs}}.$$

Since the relevant value of $a_{k,ij}$ is -1, we are looking for the (i, j) such that

$$\bar{c}_{ij} = \min_{rs \in K} \bar{c}_{rs}$$

where K is the set of arcs which have open nodes at both ends. This is the pivoting rule of the dual simplex method for transportation networks.

The duals are easily found by the usual formula

$$\pi_i^S = \pi_j^D + c_{ij}$$

for the arc from i to j basic, and therefore $\bar{c}_{ij}$ is found by

$$\bar{c}_{ij} = \pi_j^D + c_{ij} - \pi_i^S.$$

We have now established the formulas if potential entering arcs exist. What if such an arc does not exist?

First we show that this situation occurs if and only if the leaving arc is connected to a leaf node. A node is a *leaf* if it is an end node for only one basic arc.

This can easily be seen as follows: By deleting the leaving arc, which is not connected to a leaf node, from the basic tree, we are left with two disjoint trees. One of them will have a blocked supply node, the other a blocked demand node. In the tree with one blocked supply node, *all* supply nodes are blocked and all

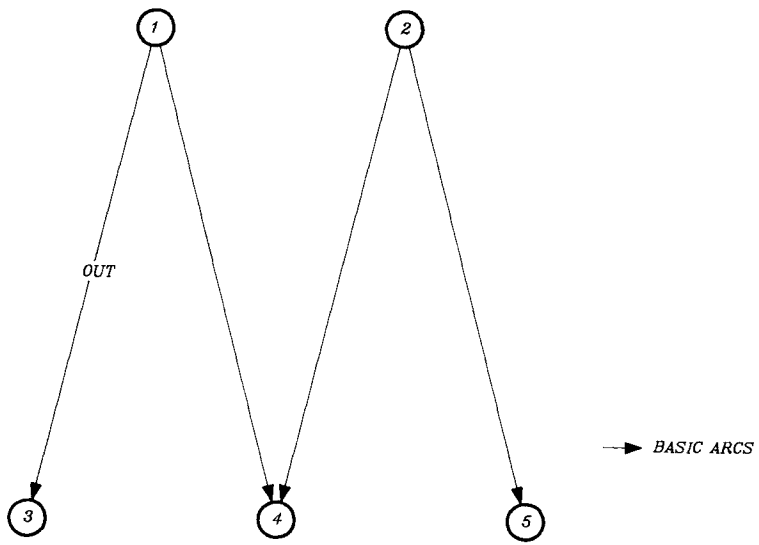


Fig. 4. Leaving arc is connected to a leaf node.

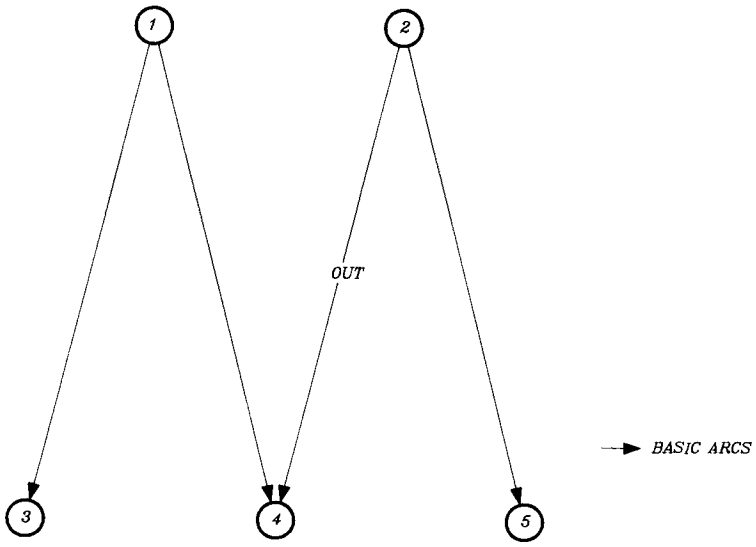


Fig. 5. Leaving arc is not connected to a leaf node.

demand nodes are open. In the other tree all demand nodes are blocked and all supply nodes are open. But if the leaving arc is connected to a leaf, that part of the tree will have no open nodes, and therefore no arc will have open nodes at both ends.

It is therefore clear that the number of leaves in a tree tells us something about how 'extreme' the given basis is, i.e. in how many directions we will end up outside pos N.

Since all trees have at least two leaves, no basis can be said to be 'internal' i.e. they all have at least two faces in common with pos N. The interested reader is referred to [5].

From the definition of the transportation problem, we know that as long as we let supply equal demand, and only supply/demand nonnegative amounts, our problem is feasible. Therefore ending up outside pos N is in this case equivalent to a badly stated problem.

If we make a cut through a three-dimensional decomposed cone, we get a figure like Fig. 6 (the points corresponding to rays generated by the columns). An interesting question is whether or not we can have a situation as in Fig. 7 i.e. a linear dependence among 3 (or in general M) columns. The answer is

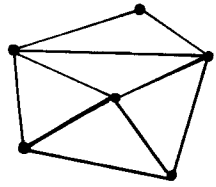


Fig. 6. A cut through a decomposed polyhedral cone in $\mathbb{R}^3$.

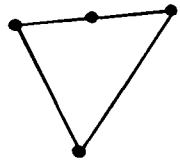


Fig. 7. A type of linear dependence we never experience.

Proposition 7. *We will never experience that a set of weights* $k_{ij} \geq 0$, $\sum k_{ij} \neq 0$ *exists, such that*

$$\sum_{ij=1}^{M} k_{ij} a_{ij} = 0$$

where a_{ij} *is the column of* N *representing* arc(i, j).

Proof. All columns contain nonzero entries and all nonzero entries in one row have the same sign.

This does not rule out degeneracy, but clearly simplifies the nature of degeneracy.

4. Degeneracy

We now consider degeneracy, and *not* for purely theoretical reasons. The following figure illustrates the problem.

Assume that at a certain step of the algorithm we have found both the cells (bases) I and II in Fig. 8 and that all the dots represent columns that are such that the reduced costs are unaffected by which three we pick. The risk is then that from I we pick arc 2 and from II we pick arc 1 thereby creating a situation like Fig. 9.

This situation is clearly nonunique and is caused by dual degeneracy. It is also clear that the nonuniqueness is only within the area of constant reduced costs, and that cycling is no problem if the data is properly organized. It should also be clear that whether or not the nonuniqueness is a problem at all, is a question of algorithms and data structures. Very advanced structures will be needed, however, since the

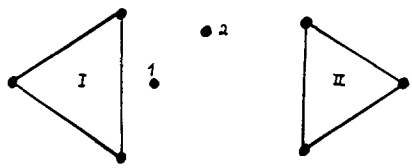


Fig. 8. The situation in an intermediate step of the algorithm. The dots represent columns. The triangles represent optimal bases already found.

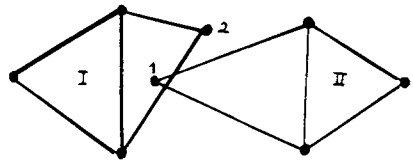


Fig. 9. A possible outcome from Fig. 8 if degeneracy is not considered. This situation represents nonuniqueness of the decomposition.

pointer-structure explained in the next section could be spoilt. We shall not discuss data structures, but rather give an algorithm to prevent degeneracy. For further details the reader is referred to [3]. Let

$$c_{ij}(\theta) = c_{ij} - \theta^{M_2(i-1)+j}.$$

This implies that arcs originating in supply node 1 are numbered 1 through M_2, and those originating in supply node M_1 are numbered from $M_1 * M_2 - M_2$ to $M_1 * M_2$ i.e. from $N - M_2$ to N.

By choosing θ small enough we know that $\theta^i < \theta^j$ if and only if $i > j$. The definition of $\bar{c}_{ij}(\theta)$ is now (arc(i,j) not basic).

$$\begin{aligned}\bar{c}_{ij}(\theta) &= c_{ij} - \theta^{M_2(i-1)+j} + \theta_B B^{-1} a_{ij} - c_B B^{-1} a_{ij} \\ &= \bar{c}_{ij} - \theta^{M_2(i-1)+j} - \theta_B \mu(\gamma_{ij}) = \bar{c}_{ij} - \sum_{k \in \mathrm{FORW}(i,j)} \theta^k + \sum_{k \in \mathrm{REV}(i,j)} \theta^k\end{aligned}$$

where $\bar{c}_{ij}$ is the unperturbed reduced cost, FORW(i,j) is the set of forward arcs in the cycle created by arc(i,j) in the basic trees and REV(i,j) the set of reverse arcs in the cycle. Note that arc(i,j) is included in the set FORW(i,j).

If therefore $\bar{c}_{ij} = \bar{c}_{rs}$ we shall look for the lowest indexed column in each of the cycles γ_{ij} and γ_{rs}. If one has a lower index than the other, the min of $\bar{c}_{ij}(\theta)$ and $\bar{c}_{rs}(\theta)$ depends on whether this is a forward or reverse arc in the cycle.

Therefore, to check for min $\bar{c}_{ij}(\theta)$ when we have a tie on $\bar{c}_{ij}$, we must identify the cycles created by these arcs in the basic tree in order to break the tie.

Since transportation networks are likely to be very degenerate (at least with integer costs), an efficient method must be found to solve the above-mentioned problem, since ties are very likely to occur often.

5. Checking if a right-hand side belongs to a cell

A right-hand side belongs to a given basis (cell) if it can be written as a nonnegative linear combination of the columns in that basis.

Since we know that any basis in a pure network can be written as an upper triangular matrix, we shall be looking for a method that takes advantage of this property.

We shall store the basic arcs in an integer vector as follows (using $k = M_2(i-1)+j$ as column index):

(i) Give each node in the network an index showing how often it appears as an end node of a basic arc (clearly the first M_1 nodes are only tail nodes, the last M_2 nodes are only head nodes).

(ii) Find a node with index 1.

(iii) Put the arc which has this node as an end node in the next position in the arc list, with a positive sign if the node was a tail node, and with negative sign otherwise.

(iv) Reduce the indices of the end nodes of the arc by one.

(v) If all nodes have index 0, stop, otherwise go to (ii).

Example. With the following simple network, an acceptable vector of basic arcs is $(5, -4, 3, -1)$.

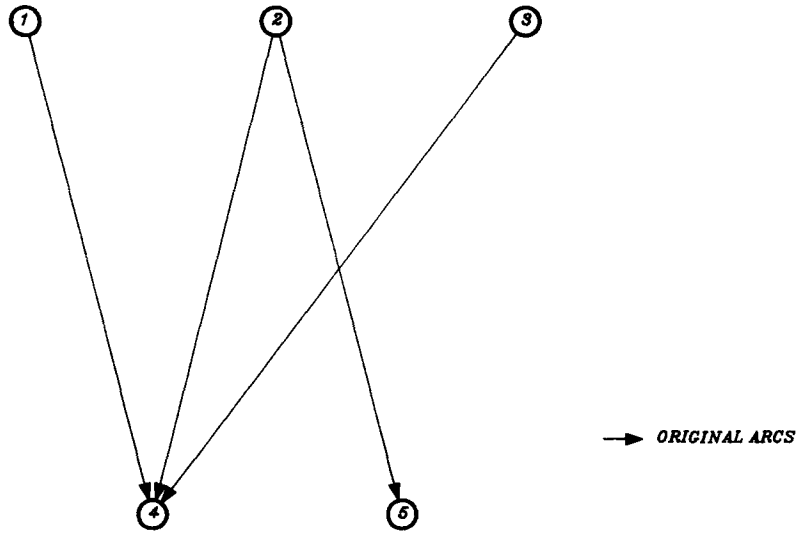


Fig. 10. Basic network for the example.

With this structure for the basic arcs we shall show how to check if a right hand side belongs to this cell by means of an example. Let $b = (2, 3, 4, 5, 4)$ where we have left out the negative signs for the demand. (Note that in the example we have ordered the nodes consecutively, and not restarted the numbering at the first demand-node, as the formulas require).

Since the first element of the basic arc list is 5, we start by letting all supply in node 3 (tail node of arc 5) be sent to node 4, i.e. $y_{34} = 4$. Then b is changed to $b = (2, 3, 0, 1, 4)$.

The next element in the list of basic arcs is -4. Therefore all demand at the head node of arc 4 is to be satisfied, i.e. $y_{25} = 4$. b then becomes $b = (2, -1, 0, 1, 0)$.

The next element of the list is 3, i.e. $y_{24} = -1$ which makes $b = (2, 0, 0, 2, 0)$.

The last element in the list is -1 (could as well have been $+1$) i.e. $y_{14} = 2$.

Therefore the solution is (using $k = M_2(i-1)+j$), $(2, 0, -1, 4, 4, 0)$, and hence b does not belong to this cell.

The next step is therefore to check the neighbour of this cell found by leaving out arc 3 (arc(2, 4)).

As can be seen we only need to go through the M-list of basic arcs once to check if a given right hand side belongs to a given cell. If it does not belong to the cell, that will often be discovered before we have gone through the whole list.

6. Organizing the data

In the previous section we showed how to store the data once the decomposition was done. In this section we show how to organize the data in order to efficiently perform the decomposition. If the decomposition is done as we pass along, i.e. we create new cells as they are needed, an appropriate adjustment must be made.

We shall now construct a graph. The graph consists of a set of nodes, one for each cell and a set of arcs, one for each pair of neighbour nodes/cell/bases. Two bases are neighbours if and only if they differ in exactly one column. The arcs are represented by pointers. We shall also construct a list that consists of the same nodes as in the graph, but here the pointers order the nodes sequentially. The list starts by a dummy node 0.

To each node is attached

α_i: Integer array of basic variables of length M.

β_i: Boolean array of length M.

ρ_i: Array of pointers of length M.

P_i: List of pointers.

V_i: Objective value—real.

The objective value is defined as $\sum_{k=1}^{M} c(\alpha_i(k))$ and is used to simplify a search which will be explained later. The array α_i will be ordered according to increasing values of $k = (i-1)M_2 + j$. If $\beta_i(t) = \text{true}$, it has not yet been checked if a neighbour can be found by taking arc $\alpha_i(t)$ out of basis. As long as $\beta_i(t) = \text{true}$, $\rho_i(t)$ has no meaning. If $\beta_i(t) = \text{false}$, $\rho_i(t)$ will be the null-pointer if the arc $\alpha_i(t)$ is connected to a leaf, and $\rho_i(t)$ will be a pointer to another node in the graph if the arc is not connected to a leaf. This other node represents the basis which we get if we take arc $\alpha_i(t)$ out of the basis. The P_i-pointers order the nodes according to increasing values of V_i. Let $\{x\}$ be the set containing the elements of the vector x, unordered.

The algorithm then goes as follows:

(i) Pick an arbitrary basic tree which has non-negative reduced costs. Construct α_1: its index vector. Find V_1. For all basic arcs k such that the arc is connected to a leaf, let $\rho_1(k) = 0$ and $\beta_1(k) = \text{false}$. Let the rest of β_1 be true. Let $r = 1$. Let $p_0 = 1$ and $p_1 = 0$.

(ii) Remove node r from the list (of p-pinters). For all k such that $\beta_r(k) = \text{true}$, find the entering arc k_1. Let $\text{OBJ} = V_r - c(\alpha_r(k)) + c(k_1)$. Go through the list of nodes, starting with node p_0 to find $T = \{t: \text{OBJ} = V_t\}$. If $T = \emptyset$ go to (iii): We have found a new node. Else check for all $t \in T$ if $\{\alpha_r\} - \{\alpha_r(k)\} + \{k_1\} = \{\alpha_\tau\}$. If no such t exists, go to (iii); a new node is found. Otherwise an already created basis has been reached—go to (iv) with t_1 as the pointer to the node.

(iii) A new node s is to be created. Create α_s from α_r by leaving out $\alpha_r(k)$ and putting k_1 in its appropriate place. Let $V_s = \text{OBJ}$. Define β_s and ρ_s as in (i). Let k_2 be such that $\alpha_s(k_2) = k_1$. Let $\rho_s(k_2) = r$ and $\beta_s(k_2) = \text{false}$. Furthermore, let $\rho_r(k) = s$ and put node s into its correct place in the list by giving p_s its appropriate value. Go to (v).

(iv) The node was already created, namely node t_1. Let k_2 be such that $\alpha_{t_1}(k_2) = k_1$. Let $\beta_{t_1}(k_2) = \text{false}$ and $\rho_{t_1}(k_2) = r$. Let $\rho_r(k) = t_1$. Go to (v).

(v) Let $r = r+1$: If r is greater than the number of nodes created so far, stop. Otherwise go to (ii).

As we can see, the ordering of a_i according to increasing values for k makes it easier to find k_2 such that $\alpha_i(k_2) = k_1$. But this does not fit with what we suggested to check efficiently if a given right hand side belongs to a given cell or not. Therefore, in order to facilitate this, we proceed as follows:

(1) Resort α_i (and remember to resort ρ_i accordingly) as needed to make the above-mentioned check easy, cf. Section 5.

(2) Drop V, P and β for all nodes.

Now it is easy to find the appropriate cell for a given right hand side b. Pick a node. Use the algorithms in Section 5 to check if the cell corresponding to this node contains b. As soon as a negative value shows up, e.g. $y_{\alpha(k)} < 0$ follow pointer $\beta(k)$ to a new node and check that one. Continue until the correct node has been found.

Alternatively one might want to find *all* the negative y's and follow the pointer of the one with largest absolute value. Here the search takes longer, but the number of nodes visited is likely to be smaller.

7. The number of cells

In [1] Balinski showed that the number of cells is independent of the cost structure of the transportation problem. Furthermore he obtained the following theorem.

Theorem 8. *Let $\mathbf{a} = (a_1, a_2, \ldots, a_{M_1})$ be a vector of integers with $a_i \geq 1$ and $\sum a_i = M_1 + M_2 - 1$. There is a one-to-one correspondence between cells in a nondegenerate transportation problem and possible ways of writing $\mathbf{a}$, when M_1 is the number of supply-points and M_2 is the number of demand-points.*

From Theorem 8 we can prove the following theorem.

Theorem 9. *With M_1 supply points and M_2 demand points there are exactly*

$$\binom{M_1+M_2-2}{M_1-1}$$

cells in the decomposition of pos N.

Proof. Define the sequence $\{b_i\}_1^{M_1-1}$ such that

$$1 \leq b_1 < b_2 < \cdots < b_{M_1-1} \leq (M_1+M_2-2).$$

Then let

$$a_1 = b_1, \quad a_2 = b_2 - b_1, \ldots, a_{M_1-1} = b_{M_1-1} - b_{M_1-2}$$

and

$$a_{m_1} = M_1 + M_2 - 1 - (a_1 + \cdots + a_{M_1-1}).$$

This sequence $\{a_i\}_1^{M_1}$ is of the type required in Theorem 8. The problem is therefore reduced to finding the number of ways we can pick $M_1 - 1$ numbers from $M_1 - M_2 - 2$ numbers, i.e.

$$\binom{M_1+M_2-2}{M_1-1}. \qquad \square$$

8. The size of a cell

As shown in the previous section, the number of cells increases rapidly as M_1 and M_2 increase. The largest system that has less than 4000 cells is one with 8 supply points and 8 demand points. Therefore we cannot create all the cells in advance if the problem is of a reasonably large size. The following question therefore arises: Which cells should be generated? What is a large cell?

In order to answer this question, we made the following experiment:

Consider a system with 5 supply points and 5 demand points and let $C_{ij} = 0$ for all i and j. We have chosen this cost structure (which is extremely degenerate) because we want to test the effect of the cells' relative position to one another and not the effects of the cost structure. (It might be fruitful to test the effect of costs too, but at least the two should be separated, if possible). Theorem 9 in the previous section shows that a $5 * 5$ system has 70 cells.

We next created 10 000 random right hand sides in the following way:

(1) For supply point 2 to 5 and demand point 1 to 4, pick a random number from a uniform distribution between 0 and $\frac{1}{2}$.

(2) If total supply exceeds total demand, let the demand at point 5 be the difference and let the supply at point 1 be 0.

(3) If total demand exceeds total supply, let supply point 1 supply the difference and let the demand at point 5 be 0.

As we can see, the expected total supply is one and the expected total demand is one, thereby giving node 1 an expected supply of zero and demand point 5 an expected demand of zero.

We then ran a program that checked how many right hand sides fit into each of the 70 cells.

We then posed two questions:

(1) What is the significance of a cell's position relative to the cell that contains the expected values for the supply and demand?

(2) Is it important how many neighbours a given cell has? I.e. how 'extreme' it is?

To answer question 1 we attached to each cell a number showing how many pivots were necessary to get to it from the central cell (containing the expected values of supply and demand) along the shortest path. The results are shown in Table 1.

Table 1

Statistics for 10 000 random, uniformly distributed right-hand sides on a 5 * 5 system with all costs zero. Criterion: Number of pivots from central cell.

Number of pivots from central cell	Number of cells	Number of cells used	Number of right hand sides in these cells	Average for all cells	Aggregate average for all cells	Average for used cells	Aggregate average for used cells
0	1	1	483	483	483	483	483
1	6	6	1921	320	343	320	343
2	14	13	3136	224	264	241	277
3	18	15	2573	143	208	172	232
4	13	9	1129	87	178	125	210
5	9	3	412	46	158	137	205
6	5	3	273	55	150	91	199
7	4	1	73	18	143	73	196

As can be seen in the table, we have distinguished between the number of cells at a certain distance and the number of cells used at the same distance. The reason is as follows: Certain cells will never be used (the probability is zero) because they are generated by an arc structure illustrated in Fig. 11.

Since the probability that $D_2 = 0$ (i.e. the demand at node 5) is zero and since the probability that $S_3 = 0$ is zero, and since we have that both $S_1 > 0$ and $D_3 > 0$ never happen, one of two things can happen:

(a) $S_1 = 0$, $D_2 > 0$ making it impossible to satisfy the demand at node 5.

(b) $D_3 = 0$ and $S_3 > 0$ making it impossible to send the supply at node 3 anywhere.

The probability of (a) or (b) is one.

Therefore all cells having the property that both supply node 1 and demand node M_2 is connected to at least one leaf node each, will fall into the category of 'cells

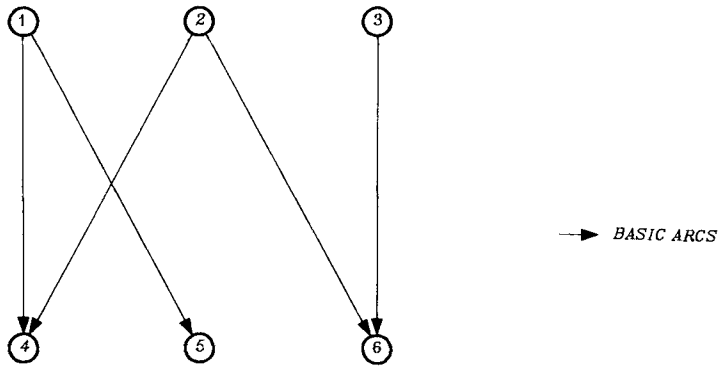


Fig. 11. A cell for which the probability of being used is zero.

never used' (with probability one). But even if the 'cells never used' can be identified in advance through this property, they must still be used in the decomposition since they represent pivot steps.

The result to be read from Table 1 is very clear:

- Being close to the cell containing the expected values (the central cell) is a good measure of size. The first 20 cells (29%) (those with 0, 1 and 2 pivots) contain 55% of the right-hand sides and if we also add those with 3 pivots, we get that 38 cells (54%) contain 81% of the right-hand sides.

This supports the idea of creating cells in circles around the central cell if not all cells can be created. If a right-hand side does not fit any of the cells, use a dual method to calculate the solution from the last cell found. Few steps will probably be needed.

This result is supported by numerous tests with networks other than $5 * 5$.

We mentioned in Section 3 that the number of neighbours for a cell shows how 'extreme' it is. Few neighbours mean that the cell has a high number of faces in common with those of the polyhedral cone pos N. Therefore we checked if the number of neighbours was a good measure of the size of a cell. The result is shown in Table 2. The explanation of the columns in Table 2 is the same as for Table 1.

The lesson to be drawn is also clear:

- The number of neighbours is a good measure of size. The 21 largest cells (among which 6 are never used) (30% of the total) contain 48% of the right hand sides, and if we also add those with 4 neighbours, 48 cells (69%) contain 86% of the right-hand sides.

Although we now have found two different measures of size for a cell, it should be clear that the last one is not very practical from a computational point of view. We therefore conclude (noting that the central cell has 6 neighbours):

If only a limited number of cells can be generated, one should pick those that have the lowest distance from the central cell.

Table 2

Statistics for 10 000 random, uniformly distributed right hand sides on a 5 * 5 system with all costs zero. Criterion: Number of neighbours.

Number of neighbours	Number of cells	Number of cells used	Number of right hand sides in these cells	Average for all cells	Aggregate average for all cells	Average used cells	Aggregate average for used cells
6	5	4	1664	333	333	416	416
5	16	11	3127	195	228	284	319
4	27	19	3765	139	178	198	252
3	17	12	1204	71	150	100	212
2	4	3	154	39	143	51	202
1	1	1	86	86	143	86	200

We have also tested the following problems: When a search for the correct cell is initiated, in which cell should the search start? We have checked 2 possibilities:

- pick the cell where the last search ended,
- pick the central cell.

With no kind of sorting of the right-hand side (which perhaps could favour alternative 1), alternative 2 was significantly better than the other. Hence we suggest:

The search for the correct cell should always start in the central cell (the cell containing the expected values of supply and demand).

To fit 10 000 right-hand sides into the 70 cells of a 5 * 5 system takes about 35 seconds CPU-time on a PRIME 850. We doubt that it will be faster to sort the 10 000 right-hand sides in order to make the *total* process faster by using the above-mentioned alternative 1 for the initial cell. One of the reasons is that it is not at all clear what criterion the sort should be based on. A suggestion might be to sort on the basis of the Euclidean distances between the right-hand sides. Note, however, that the sort must be sequential, such that K right-hand sides will require $(K-1)!$ distance calculations and comparisons.

It is clear that provided pos N is totally decomposed, this method is superior to bunching [7], once the decomposition is done. But if only a limited number of cells have been found, bunching might be useful on those right hand sides that fall outside the decomposed area. We have not compared with sifting [7].

9. Example

Let the network of Fig. 12 be given for $M_1 = 2$, $M_2 = 3$. Furthermore, let

$$c = \begin{pmatrix} 0 & 0 & 0 \\ 3.67 & 6.12 & 3.4 \end{pmatrix}.$$

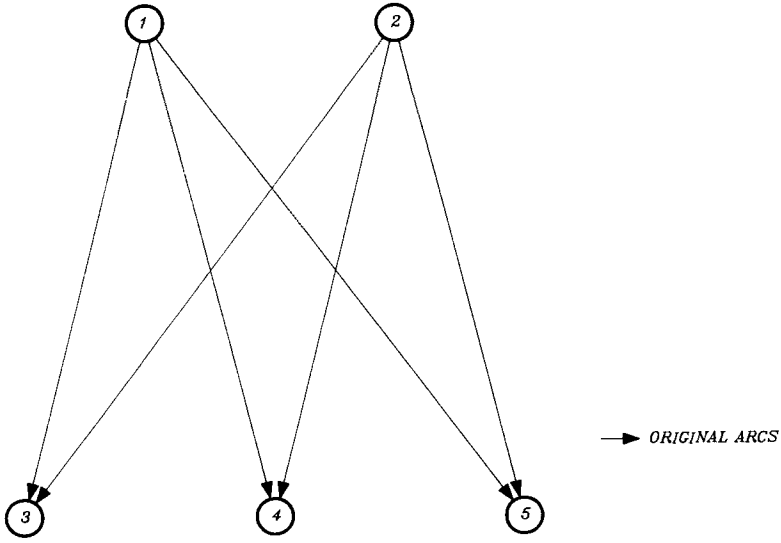
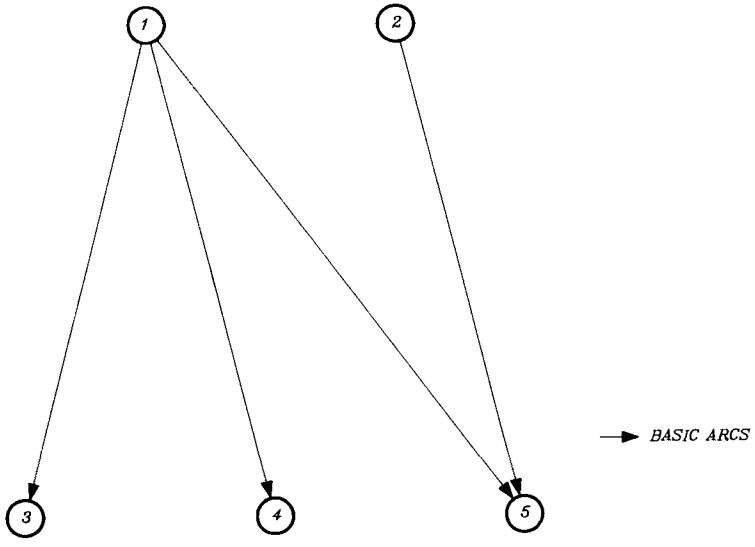


Fig. 12. Original transportation problem.

This problem decomposes into the three cells shown in Figs. 13, 14 and 15. The important variables can be found in Table 3. We therefore know that every consistent combination of supply and demand will have one of these cells as optimal.



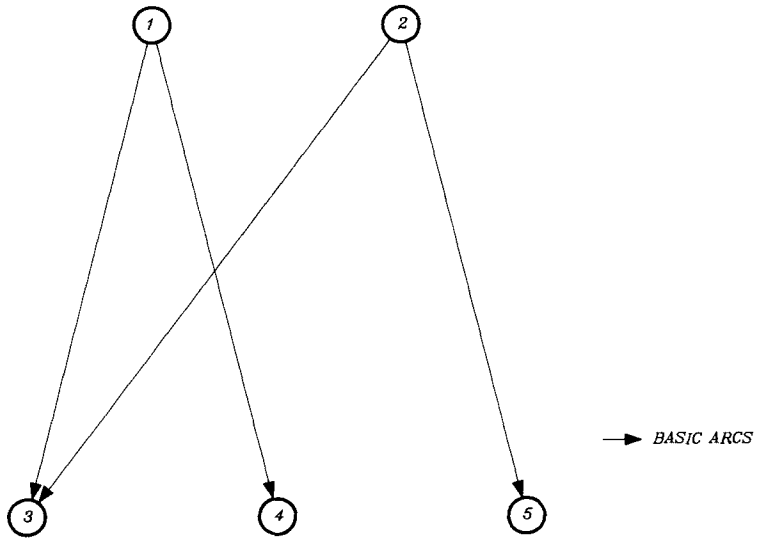


Fig. 14. Cell 2.

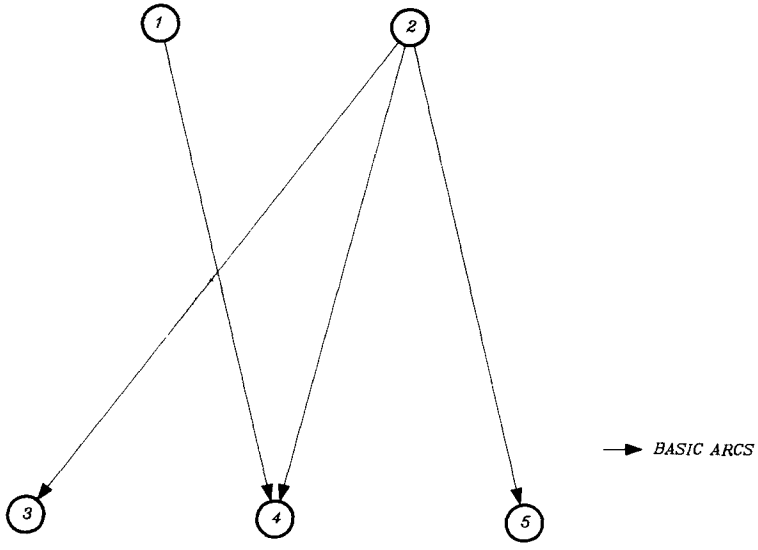


Table 3
The most important variables for the example

Cell	Basis	Reduced cost	Pointer	Duals
1	1, 2, 3, 6	0, 0, 0, 0.27, 2.72, 0	0, 0, 2, 0	0, 3.4, 0, 0, 0
2	1, 2, 4, 6	0, 0, 0.27, 0, 2.45, 0	3, 0, 1, 0	0, 3.67, 0, 0, 0.27
3	2, 4, 5, 6	2.45, 0, 2.72, 0, 0, 0	0, 0, 2, 0	0, 6.12, 2.45, 0, 2.72

References

[1] M.L. Balinski, "The Hirsch conjecture for dual transportation polyhedra", Collaborative paper CP-83-9, IIASA, Laxenburg, Austria (1983).
[2] G. Gallo and C. Sodini, "Extreme points and adjacency relationship in the flow polytope", *Calcolo* 15 (1978) 277–288.
[3] D.W. Walkup and R. J.-B. Wets, "Lifting projections of convex polyhedra", *Pacific Journal of Mathematics* 28 (1969) 465–475.
[4] S.W. Wallace, "A two-stage stochastic facility-location problem with time-dependent supply", Report no. 832555-10, Chr. Michelsen Institute, Bergen, Norway (1984).
[5] R. J-B. Wets and C. Witzgall, "Algorithms for frames and lineality spaces of cones", *Journal of Research of the National Bureau of Standards* 71B (1967) 1–7.
[6] R. J-B. Wets, "Stochastic programs with fixed recourse: The equivalent deterministic program", *SIAM Review*, 16 (1974) 309–339.
[7] R. J-B. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming: The state of the art, Bonn 1982* (Springer-Verlag, Berlin, 1983) pp. 566–603.

Mathematical Programming Study 28 (1986) 48–62
North-Holland

MULTISTAGE STOCHASTIC PROGRAMS WITH BLOCK-SEPARABLE RECOURSE

François V. LOUVEAUX
Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium

Received 25 January 1984
Revised manuscript received 8 October 1984

The concept of block-separable recourse is introduced as a structural property that can usefully be exploited in multistage stochastic programs with recourse. Examples of applications where this property can be found are given. The equivalence of multistage programs with block-separable recourse and two-stage programs is shown and the algorithmic implications of block-separability are studied.

Key words: Multistage Stochastic Programming, Nested Decomposition, Stochastic Integer Programming.

1. Introduction

Much research has been devoted in the field of stochastic programming to the two-stage programs, see e.g. Dempster [4, Introduction] and Wets [18] for recent surveys.

In practice however, many problems have an inherently dynamic structure. Most theoretical results naturally extend from the two-stage situation to the multistage one, see e.g. Olsen [12, 13]. Algorithms that have been proposed for multistage programs can also be considered as extensions of the two-stage L-shaped algorithm of Van Slyke and Wets [16]. They are mainly nested decompositions studied by Birge [2] in the linear case and Louveaux [9] in the quadratic case. Basic definitions of multistage programs are reviewed in Section 1.

The object of this paper is to study a structural property that most multistage programs either naturally possess or can be led to possess: the so-called block-separability. If the multistage problem is viewed as a set of subproblems in each node of some decision tree, block-separability means that the decision vector in any subproblem is formed of two vectors, the aggregate level decision vector and the detailed level decision vector, and that the objective function and the constraint matrices related to those two vectors are separable, in the mathematical sense. The definition of block-separability is given in Section 3. Examples of applications where this property holds are presented in Section 4: capacity expansion, resource exhaustion and hierarchical planning systems from which the terminology aggregate-detailed is taken [7]. In Section 5, we show that any multistage program with block-separable recourse can be transformed into a two-stage stochastic program

where the first-stage is the extensive form of the aggregate level problems and the second-stage is related to the detailed level. We also discuss algorithmic implications of this property, in particular the possibility of solving multistage stochastic integer programs when the integer variables are associated with the aggregate level decisions.

Finally, in Section 6, we study how a nested decomposition approach can still be used to solve the two-stage equivalent form in the linear case but not in the quadratic case.

2. Multistage stochastic programs

Let $x_t = (x_1, \ldots, x_t)$ denote the set of decisions made up to and including time t, and $\xi_t = (\xi_1, \ldots, \xi_t)$ the set of realisations of the random vectors as seen from period t.

Then, a multistage stochastic program with recourse can be defined recursively as follows:

$$\text{(M.S.P.)} \quad \inf \quad \{z_1(x_1) \mid x_1 \in D_1\}$$

where

$$z_t(x_t, \xi_t) = c_t(x_t) + E_{\xi_{t+1}|\xi_t} Q_t(x_t, \xi_{t+1})$$

and

$$\begin{aligned} Q_t(x_t, \xi_{t+1}) = \inf \quad & z_{t+1}(x_{t+1}, \xi_{t+1}) \\ \text{s.t.} \quad & x_{t+1} \in F_{t+1}, \\ & T_{t+1}x_t + W_{t+1}x_{t+1} = h_{t+1} \end{aligned}$$

for

$$1 \leq t \leq N-1, \text{ with } Q_N(\ ) \equiv 0 \text{ and } z_1(x_1) = z_1(x_1, \xi_1).$$

For all stages t, the function $c_t(x_t)$ describes the current objective function at time t and its expression may contain random elements. In particular, if $c_t(x_t) = c_t \cdot x_t$ is linear, the vector c_t is random. In addition, the vector h_{t+1} and the matrices T_{t+1} and W_{t+1} are random. Those vectors have sizes consistent with $x_{t+1} \in \mathbb{R}^n$ and $h_{t+1} \in \mathbb{R}^m$, so W_{t+1} is a $m \times n$ matrix and T_{t+1} is a $m \times (n \times t)$ matrix, e.g.

Properties of the multistage stochastic programs are best expressed in terms of the deterministic equivalent program as seen from period t which is defined as follows.

$$\begin{aligned} \text{(D.E.P.)}_t \quad \text{Find} \quad \inf \quad & z_t(x_t, \xi_t) \\ \text{s.t.} \quad & x_t \in F_t, \\ & T_t x_{t-1} + W_t x_t = h_t. \end{aligned}$$

Properties of the (D.E.P.)$_t$ have been extensively studied when the random vectors $\xi_{t+1}|\xi_t$ have a finite support for all t. Then the (D.E.P.)$_t$'s are known to be piecewise linear if $c_t(x_t)$ is linear and piecewise quadratic when $c_t(x_t)$ is quadratic (see Gartska and Wets [5] and also Wets [17] for results on feasibility sets).

Solution techniques for multistage program are based on dual decomposition schemes and are as such extensions of the 2-stage L-shaped algorithm proposed by Van Slyke and Wets [16], which is directly related to Benders's decomposition [1]. The multistage algorithms solve nested sequences of (D.E.P.)$_t$'s using backward and forward induction, see e.g. Birge [2] in the linear case, and Louveaux [9] in the quadratic case.

3. Multistage stochastic program with block-separable recourse

Definition. A multistage stochastic program is said to have block-separable recourse if for all periods $t = 1, \ldots, N$ and all random events $\xi_{t+1}|\xi_t$ the decision vectors x_t can be separated into two subvectors $x_t = (w_t, y_t)$ in such a way that

(a) The objective function $c_t(x_t)$ is separable in two parts:

$$c_t(x_t) = r_t(w_t) + q_t(y_t).$$

(b) The constraint matrix W_t is block diagonal

$$W_t = \begin{pmatrix} A_t & 0 \\ 0 & B_t \end{pmatrix}$$

where the first block A_t is associated to the vector w_t and the second block B_t to the vector y_t.

(c) The region $x_t \in F_t$ is equivalent to

$$\{x_t \in F_t\} = \{w_t \in D_t\} \times \{y_t \in E_t\}.$$

The w_t vector is called the aggregate level decision vector and the y_t vector the detailed level decision vector.

(d) The T_t and h_t matrices are decomposed to conform with the (w_t, y_t) separation and T_t only possesses nonzero elements in the columns associated with the aggregate level decisions

$$T_t = \begin{pmatrix} R_t & 0 \\ S_t & 0 \end{pmatrix} \quad \text{and} \quad h_t = \begin{pmatrix} b_t \\ d_t \end{pmatrix}.$$

Illustrations of examples where the block-separable recourse property occurs are given in the next section.

Then, a multistage stochastic program with block-separable recourse (in short BS MSP for block-separable multistage stochastic program) is defined as follows:

(BS-MSP) $\inf \{z_1(w_1) | w_1 \in D_1\}$

where

$$z_t(w_t, \xi_t) = r_t(w_t) + E_{\xi_{t+1}|\xi_t}\{Q_t^w(w_t, \xi_t) + Q_t^y(w_t, \xi_t)\}$$

and

$$\begin{aligned} Q_t^w(w_t, \xi_t) = \inf \quad & z_{t+1}(w_{t+1}) \\ \text{s.t.} \quad & w_{t+1} \in D_{t+1}, \\ & R_{t+1} \cdot w_t + A_{t+1} w_{t+1} = b_{t+1} \end{aligned}$$

and

$$\begin{aligned} Q_t^y(w_t, \xi_t) = \inf \quad & q_{t+1}(y_{t+1}) \\ \text{s.t.} \quad & y_{t+1} \in E_{t+1}, \\ & S_{t+1} w_t + B_{t+1} y_{t+1} = d_{t+1} \end{aligned}$$

for $t = 1, \ldots, N-1$, with $z_1(w_1) = z_1(w_1, \xi_1)$ and $Q_{N-1}^w(\cdot, \cdot) = 0$, $Q_N^y(\cdot, \cdot) = 0$.

Comments. 1. The ordinary two-stage stochastic program appears to be a particular case of block-separable multistage program where the first stage decisions only consist of aggregate level decisions and the second stage only consist of detailed level decisions.

2. Multistage programs with simple recourse, as discussed in Kusy and Ziemba [6], naturally possess the block-separability property. In the linear case, using the above notation, they can be defined by taking

$$r_t(w_t) = r_t \cdot w_t, \quad D_t = \mathbb{R}_+^n, \quad R_t \equiv 0, \quad E_t = \mathbb{R}_+^{2m}, \quad q_t(y_t) = q_t^+ \cdot y_t^+ + q_t^- \cdot y_t^-, \quad B_t = (I, -I) \quad \text{and} \quad d_t = \xi_t.$$

The multistage extension of the quadratic problem with simple recourse considered by Rockafellar and Wets [15] can be defined in a similar fashion using a piecewise quadratic function $q(y)$.

3. Block-separability has immediate advantages from the algorithmic point of view: a reduction in size for each of the subproblems to be solved as well as a more efficient description of the polyhedral cells needed in the quadratic case, see Louveaux and Smeers [10] for details. Other interesting properties of block-separability are given in Section 5 and 6.

4. Examples of block-separability

Example 1. *Capacity expansion of power plants.* The basic description and most of the results cited in this example are taken from Louveaux and Smeers [10].

We assume here a given set of commercially available technologies, along with the date when each technology becomes commercially available. So we only consider

the timing and the size of the investments made in the existing technologies (excluding the aspects of financing *R&D* for creating new technologies) so as to meet future electricity demand. We give a description of the objective function and of the constraints for one (D.E.P.)$_t$. Let

$i = 1, \ldots, n$, index the set of technologies,

$j = 1, \ldots, m$, index the set of operating modes, in which we divide the area under the load curve, described in Fig. 1, which represents the number of hours h during which the total electricity demand during one year period reaches the level D.

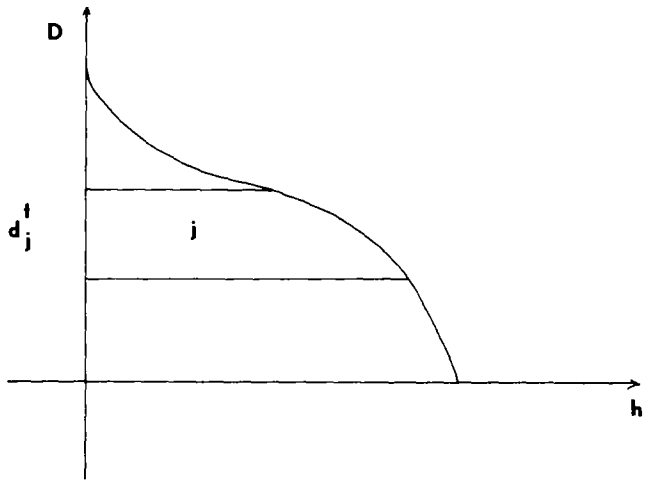


Fig. 1.

Define

x_i^t = new capacity (investment) decided for technology (equipment type) i at time t,

s_i^t = total capacity of technology i available plus in order at time t,

y_{ij}^t = capacity of technology i effectively used at time t in operating mode j,

Δ_i = construction lead time for technology i,

α_i = utilization factor of technology i,

g_i^t = existing capacity of technology i which has been decided before period 1, and is therefore an exogeneous data for the problem,

d_j^t = maximal demand (power) covered in operating mode j.

We also assume there exists one technology, say technology n, with zero construction lead-time ($\Delta_n = 0$) and large operating cost. The (D.E.P.)$_t$ can be defined as

$$\text{(D.E.P.)}_t \quad \min \quad c_t(x_t) + q_t(y_t) + E_{\xi_{t+1}|\xi_t} Q_t(s^t, \xi_{t+1}) \tag{1}$$

$$\text{s.t.} \quad s_i^t = s_i^{t-1} + x_i^t, \quad i = 1, \ldots, n, \tag{2}$$

$$\alpha_n(g_n^t + s_n^{t-1} + x_n^t) \geq \sum_{j=1}^{m} d_j^t - \sum_{i=1}^{n-1} \alpha_i(g_i^t + s_i^{t-\Delta_i}), \tag{3}$$

$$\sum_{i=1}^{n} y_{ij}^t = d_j^t, \quad j = 1, \ldots, m, \tag{4}$$

$$\sum_{j=1}^{m} y_{ij}^t \leq \alpha_i \cdot (g_i^t + s_i^{t-\Delta_i}), \quad i = 1, \ldots, n, \tag{5}$$

$$x, y, s \geq 0.$$

Constraint (3) defines the minimal investment needed in the zero lead-time technology so as to ensure feasibility of the (D.E.P.)$_t$ for any past decisions (relatively complete recourse). Since we only consider new investments over a relatively short time horizon ($N = 5$ periods of 5 years each), such that most equipments will not become obsolete inside the period, all costs are computed in a yearly equivalent basis and new investment decisions taken after period 1 are assumed to be repeated when the equipment becomes obsolete, so an explicit treatment of obsolescence in equation (2) is avoided.

The cost function $c_t(x_t)$ describes the investment cost and is linear in x_t. The function $q_t(y_t)$ describes the cost of meeting demand as a function of the operation of the equipments. It is nonlinear in the y's (since the load curve is nonlinear) and can be approximated by simple functions such as piecewise linear or piecewise quadratic functions.

Given those assumptions, the optimal investment in the technology n is described by

$$x_n^t = \frac{1}{\alpha_n} \max\left\{0; \sum_{j=1}^{m} d_j^t - \sum_{i=1}^{n-1} \alpha_i (g_i^t + s_i^{t-\Delta_i}) - \alpha_n (g_n^t + s_n^{t-1})\right\} \tag{6}$$

hence

$$s_n^t = \max\left\{s_n^{t-1}; \frac{\sum_{j=1}^{m} d_j^t - \sum_{i=1}^{n-1} \alpha_i (g_i^t + s_i^{t-\Delta_i})}{\alpha_n} - g_n^t\right\}. \tag{7}$$

It follows that the (D.E.P.)$_t$ possesses the block-separability property since it can be divided in

- an aggregate level problem involving the investment decisions in the $n-1$ equipment types, plus in the zero lead-time technology as given by (6);
- a detailed level problem involving setting the operations levels for the n equipment types in the m operating modes with available capacities $g_i^t + s_i^{t-\Delta_i}$ for $i = 1, \ldots, n-1$ and $g_n^t + s_n^t$ for $i = n$, where s_n^t is given by (7).

Furthermore, the detailed level problem can now readily be solved through a simple rule: the so-called 'order merit rule' is optimal, namely demand should be met by using the equipment types in order of increasing operating costs up to their capacity limit. It follows that the computation of the recourse function $Q_t^y(w_{t-1})$ and, in the quadratic case the construction of the associated polyhedral cell, can be performed very effectively by using a specific and separate subroutine implementing the order of merit rule. This is another important advantage of block-separability.

Example 2. *Capacity expansion and resource exhaustion.* This example is taken from a larger problem which also involves the optimal allocation of *R&D* budgets to a number of electrical and non electrical technologies. It is described in de Groote, Louveaux, Poncelet and Smeers [3]. Here we consider the part of the problem related to the investments in the commercially available technologies as well as the usage of the exhaustible resources. Let

$i = 1, \ldots, n$, index the set of technologies

$j = 1, \ldots, m$, index the set of resources. Resources from different origins (extracted/or imported coal from various regions, e.g.) are considered distinct. A resource wears the same index as the corresponding extraction/import technology (equipment type) hence $m < n$.

$k = 1, \ldots, f$ index the set of fuels (for one fuel, there are several resources).

Let

x_i^t = investment in technology i decided at time t,
s_i^t = total capacity of i available plus in order at time t,
y_i^t = capacity of i effectively used at time t,
u_j^t = amount of resource j effectively extracted/imported at time t,
e_j^t = total amount of resource j extracted/imported from period 1 to period t,

Let also

a_j = total availability of resource j,
Δ_i = construction lead-time for equipment type i,
L_i = life-time of equipment type i,
α_i = utilisation factor of equipment type i,
J_k = set of resources from which fuel k can be obtained,
T_k = set of technologies using fuel k.

Then, the (D.E.P.)$_t$ can be described as follows:

$$(\text{D.E.P.})_t \quad \min \quad c_t \cdot x_t + q_t(y_t, u_t) + E_{\xi_{t+1}|\xi_t} Q_t(x_t, \xi_t) \tag{8}$$

$$\text{s.t.} \quad s_i^t = s_i^{t-1} + x_i^t - x_i^{t-Li}, \quad i = 1, \ldots, n, \tag{9}$$

$$y_i^t \leq \alpha_i s_i^{t-\Delta_i}, \quad i = m+1, \ldots, n, \tag{10}$$

$$e_j^t = e_j^{t-1} + u_j^t, \quad j = 1, \ldots, m, \tag{11}$$

$$u_j^t \leq \alpha_j \cdot s_j^{t-\Delta_j}, \quad j = 1, \ldots, m, \tag{12}$$

$$e_j^t \leq a_j, \quad j = 1, \ldots, m, \tag{13}$$

$$\sum_{i \in T_k} y_i^t \leq \sum_{j \in J_k} u_j^t, \quad k = 1, \ldots, f, \tag{14}$$

$$(x_t, s_t, e_t) \in D_t, (y_t, u_t) \in E_t.$$

The objective function in (D.E.P.)$_t$ contains one linear term $c_t x_t$ associated with the new investments decided in t and a nonlinear term $q_t(y_t, u_t)$ describing minus

the consumer's surplus one is willing to maximize. Since we consider evaluating the impact of *R&D* in the long run, this approach of using the consumer's surplus avoids the difficult exercise of forecasting demand in the long run.

Constraints (9) and (10) define the cumulative investment in equipment type i and its usage. As seen from those two equations when using a nested decomposition approach, it is necessary to maintain information about the investment decisions over $L_i + \Delta_i$ periods. This is a major difficulty to which approximating (9), see [3], or block-separability, see Section 5, can give some solution.

Constraints (11) to (13) describe the extraction of resources. Constraint (14) relates usages of fuels to effective usage of equipment types. The set D_t contains limitations on the penetration of new electrical technologies, a simplification which avoids an explicit description of the load curve as in Example 1. The set E_t defines the demand in the electrical and non electrical part as a function of the effective usage of fuels and equipment types as well as a relation between production, usage and inventory of nuclear fuels. Two main directions for simplifying the $(\text{D.E.P.})_t$ formulation and the solution method can be proposed.

Method 1. '*Wise planner*'. The first approach has become classical in the field of energy modelling, see Manne, Richels and Weyant [11]. It assumes that new capacities are built in such a way that they are always fully used in all later periods. This means that constraints

$$y_i^t \leq \alpha_i s_i^{t-\Delta_i}, \quad i = m+1, \ldots, n,$$

and

$$u_j^t \leq \alpha_j s_j^{t-\Delta_j}, \quad j = 1, \ldots, n,$$

are all binding.

The advantage of making these assumptions is that the size of the $(\text{D.E.P.})_t$ is strongly reduced. However the $(\text{D.E.P.})_t$ loses the relatively complete recourse property and becomes difficult to handle in a nested decomposition approach. In addition, this approach is somewhat unnatural in a stochastic framework, where by definition, there cannot exists such a thing as a wise planner in the above sense.

Method 2. '*Myopic consumer*'. The approach we propose is to make the $(\text{D.E.P.})_t$ block-separable. Separability between investment and usage of technologies can be obtained as in Example 1, either by assuming a construction lead time greater than or equal to one for all equipments, or that a rule similar to (6) exists for equipment types with zero lead time.

In addition, some assumption is to be made on the usage of resources since they have an impact on later availabilities. To obtain separability, any sort of myopic behaviour should be assumed. One such example is the following.

Replace constraints (11), (12) and (13) by the relations

$$u_j^t = \min(\alpha_j s_j^{t-\Delta_j}, a_j - e_j^{t-1}), \quad j = 1, \ldots, m, \tag{16}$$

hence

$$e_j^t = \min(e_j^{t-1} + \alpha_j s_j^{t-\Delta_j}, a_j), \quad j = 1, \ldots, m. \tag{17}$$

The above assumption, along with a similar assumption on the nuclear fuel, induces the block-separability of the (D.E.P.)$_t$ in two types of decisions:

- *aggregate level decisions*: capacity investments (x_t, s_t) as well as cumulative extraction given by (17);
- *detailed level decisions*: effective use of capacities (u_j, y_j) based on available resources given by (16).

It is interesting to observe that block-separability is obtained with a weaker assumption than in the wise planner method. Moreover, if block-separability holds, it is still possible to consider designing simple (heuristic) rules in order to obtain an effective computation of the recourse function $Q_t^y{}_1(x_{t-1}, \xi_{t-1})$, and in the quadratic case, of the associated polyhedral cell. It might be more difficult however to find a rule which can be proved to be optimal as is the case for the order of merit rule in Example 1.

Example 3. *Stochastic integer programming.* Several examples of block-separability can be found in the literature on hierarchical planning systems from which the terminology (aggregated level-detailed level) has been taken, see, e.g. [7]. To cite only a few, consider the machine scheduling problem (aggregate level = number of machines to buy, detailed level = scheduling of jobs to machines), vehicle routing problem (aggregate level = vehicle capacity, detailed level = routing of vehicle), location-allocation problem (aggregate level = location and capacity of plants, detailed level = allocation of available capacity to the most profitable demand points).

One of the immediate advantage of block-separability is in fact to make it possible to solve multistage stochastic integer programs, as we will see in the next section.

5. Block-separable multistage programs as two-stage programs

In this section, we indicate how any block-separable multistage program can be transformed into a two-stage program.

Assumption. The random variable $\xi_t | \xi_{t-1}$ has a finite support. Let $j = 1, \ldots, K_t$ index the possible realisations of $\xi_t | \xi_{t-1}$. Hence, at time t, the tree of outcomes includes $k_t = \prod_{i=1}^{t} K_i$ nodes. Let p_j^t be the probability associated with node j at time t, and let $\hat{j}$ denote a predecessor of node j (these notations are taken from Birge [2]).

Define

$$\boldsymbol{w} = (w_1, w_2^1, \ldots, w_2^{k_2}, \ldots, w_{T-1}^1, \ldots, w_{T-1}^{k_{T-1}}),$$

the vector of all aggregate-level decisions, and

$$\boldsymbol{y} = (y_2^1, \ldots, y_2^{k_2}, \ldots, y_T^1, \ldots, y_T^{k_T}),$$

the vector of all detailed level decisions. The vector of all aggregate-level decisions up to time t for outcome j is denoted by

$$\boldsymbol{w}_t^j = (w_1, w_2^1, \ldots, w_2^{k_2}, \ldots, w_{t-1}^1, \ldots, w_{t-1}^{k_{t-1}}, w_t^j).$$

In practice, $\boldsymbol{w}_t^j$ could be limited to the decisions vectors associated to all predecessors of j.

Define in a similar fashion the immediate objective

$$r(\boldsymbol{w}) = r_1(w_1) + \sum_{j=1}^{k_2} p_2^j r_2(w_2^j) + \cdots + \sum_{j=1}^{k_{T-1}} p_{T-1}^j r_{T-1}(w_{T-1}^j)$$

and the recourse function

$$Q(\boldsymbol{w}, \xi) = Q_1^y(w_1, j) + \cdots + \sum_{j=1}^{k_T} p_T^j Q_T^y(\boldsymbol{w}_{T-1}^{\hat{j}}, j)$$

where

$$\begin{aligned} Q_t^y(\boldsymbol{w}_{t-1}^{\hat{j}}, j) = \min \quad & q_t(y_t^j) \\ \text{s.t.} \quad & S_{t-1}\boldsymbol{w}_{t-1}^{\hat{j}} + B_t y_t^j = d_t^j, \\ & y_t^j \in E_t^j. \end{aligned}$$

Finally, let

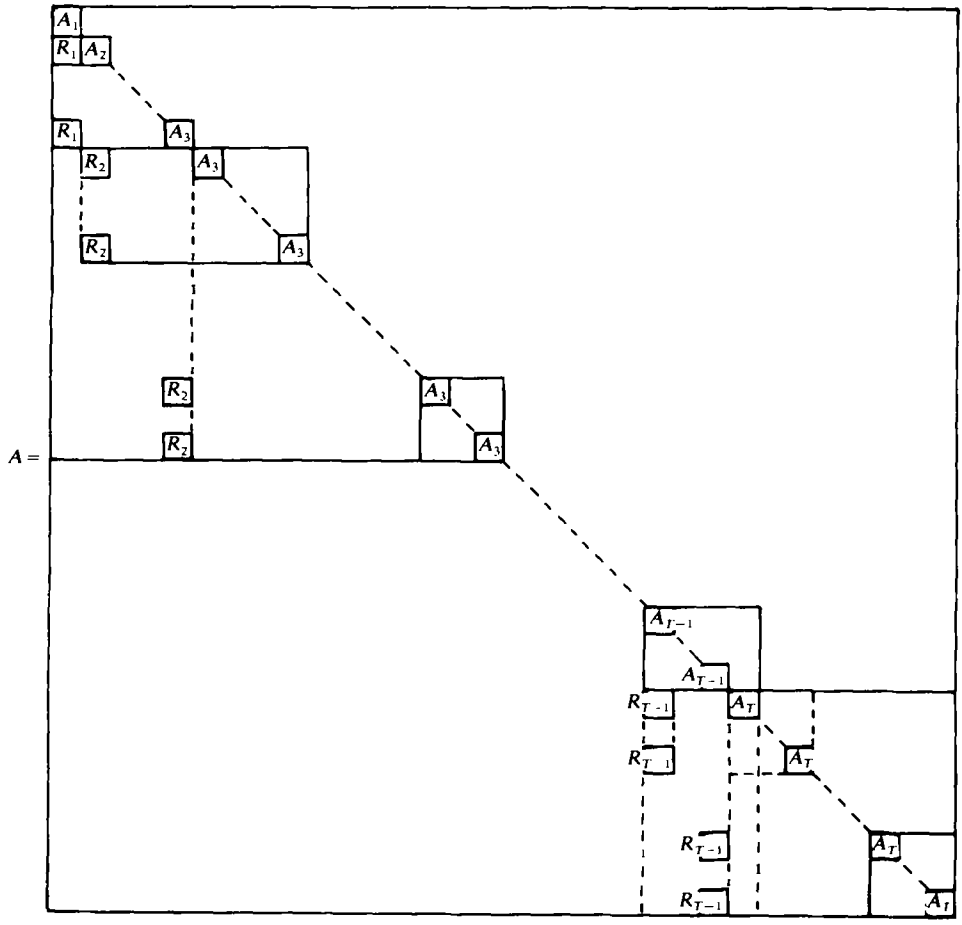
$$\boldsymbol{D} = \{\boldsymbol{w} \mid w_t^j \in \boldsymbol{D}_t^j \text{ for all } t, j\},$$

$$\boldsymbol{A} = \{\boldsymbol{w} \mid R_t \boldsymbol{w}_{t-1}^j + A_t w_t^j = b_t^j \quad \text{for all } t, j\}$$

and

$$\boldsymbol{b}^{\mathrm{T}} = [b_1, b_2^1, \ldots, b_2^{k_2}, \ldots, b_{\mathrm{T}}^1, \ldots, b_{\mathrm{T}}^{k_{\mathrm{T}}}].$$

A block-representation of $\boldsymbol{A}$ is depicted below, where for simplicity, we assume that $R_{t+1} \cdot \boldsymbol{w}_t$ is limited to $R_{t+1}\, w_t$.



It follows from this construction, that the stochastic multistage program with block-separable recourse can be written as

$$\text{(BS-2)} \quad \min \quad r(\boldsymbol{w}) + E_\xi Q(\boldsymbol{w}, \xi)$$

$$\text{s.t.} \quad \boldsymbol{w} \in \boldsymbol{D},$$

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}$$

which is the 2-stage version of a block-separable problem represented as BS-2. We have therefore obtained the following result:

Proposition. *A multistage stochastic program with block-separable recourse is equivalent to a two-stage stochastic program where the first-stage is the extensive form of the aggregate level problems and the value function of the second stage for one realisation of the random vector in the second stage is the sum (weighted by the appropriate probabilities) of the detailed level recourse functions* $Q_t^y(\boldsymbol{w}_t, \xi)$ *for that realisation of the random vector and all its successors.*

Comments

1. As far a numerical methods are concerned, it follows that L-shaped based techniques can again be applied to solve BS-2, with the additional possibility of using large scale LP. techniques for the extensive form of the aggregate level decisions, and sometimes specific routines for the detailed level decisions, as suggested in Example 1.

2. Constraints of the form

$$s_i^t = s_i^{t-1} + x_i^t - x_i^{t-L_i}$$

where L_i is large can be handled explicitly without increasing the number of variables, as is needed in the usual nested decomposition techniques to maintain a staircase structure.

3. The use of nested decomposition for block-separable multistage programs is further discussed in the next section.

4. The use of 2-stage programs to solve multistage programs was already considered by some authors using a rolling horizon approximation schema, see e.g. Kusy and Ziemba [6] for an example in finance, and Prekopa [14] for use with chance-constraints. The above proposition however does not use any approximation.

5. Last but not least, it follows from the above proposition that multistage programs having INTEGER variable in ALL STAGES, (not only the first one) can be solved using available two-stage techniques such as the one proposed by Wollmer [19], provided the integer variables are associated with the aggregate level decisions, which is very often the case as suggested by our discussion in Example 3.

6. Nested decomposition for block-separable multistage programs

Nested decomposition can readily be used to solve block-separable multistage stochastic programs, simply by using the separability of the value function $Q_t(x_t, \xi_t)$ in its two components $Q_t^w(w_t, \xi_t)$ and $Q_t^y(w_t, \xi_t)$ with possibly an efficient simple rule for the latter.

6a. The linear case

It is interesting to observe that nested decomposition can also be applied to solve the 2-stage equivalent program (BS-2).

The L-shaped algorithm, applied to BS-2, solves a sequence of problems of the type

$$\min \quad z = cw + \theta, \tag{18}$$

$$Aw = b, \tag{19}$$

$$\sigma^l \cdot w + \theta \geq \rho^l, \quad l = 1, \ldots, L, \tag{20}$$

where the optimality cuts $l = 1, \ldots, L$ are facets of the recourse function $E_\xi Q(w, \xi)$.

For simplicity, we assume here relatively complete recourse, hence induced constraints need not be introduced.

As such, constraints (20) are dense and destroy the staircase structure present in A. To restore the staircase structure, we consider a different approximation of $E_\xi Q(w, \xi)$ as follows.

$$\min \quad z = cw + \theta_1 + \sum_{j=1}^{k_2} \theta_2^j + \cdots + \sum_{j=1}^{k_{T-1}} \theta_{T-1}^j, \tag{21}$$

$$Aw = b, \tag{22}$$

$$\sigma_{t,j}^l w_t^j + \theta_t^j \geq \rho_{t,j}^l, \quad l = 1, \ldots, L, \; j = 1, \ldots, k_t, \; t = 1, \ldots, N-1. \tag{23}$$

The values of $\sigma_{t,\hat{j}}^l$ and $\rho_{t,\hat{j}}^l$ are obtained through the computation of the optimal dual variables $u(w_t^{\hat{j}}, j)$ of

$$Q(w_t^{\hat{j}}, j) = \min \quad q_{t+1} y_{t+1}^j \tag{24}$$

$$\text{s.t.} \quad S_t \cdot w_t^{\hat{j}} + B_{t+1} \cdot y_{t+1}^j \leq d_{t+1}^j, \tag{25}$$

$$y_{t+1}^l \in E_{t+1}^j$$

as

$$\rho_{t,\hat{j}}^l = E_{\xi_{t+1}|\xi_t} u(w_t^{\hat{j}}, j) \cdot d_{t+1}^l \tag{26}$$

and

$$\sigma_{t,\hat{j}}^l = E_{\xi_{t+1}|\xi_t} u(w_t^{\hat{j}}, j) \cdot S_t \tag{27}$$

where the expectation is taken over all successors j of $\hat{j}$.

If we consider (w_1, θ_1) as the decision vector of the first stage, $(w_2^1, \theta_2^1), \ldots, (w_2^{k_2}, \theta_2^{k_2})$ as the decision vector of the second stage for $j = 1, \ldots, k_2$, and so on, the contraint matrix of any iteration of the L-shaped algorithm used to solve (BS-2) resumes a staircase structure and is thus amenable to a nested decomposition technique. In addition, if specific routines as suggested in Example 1 exist, the coefficients ρ and σ are obtained in a very efficient way.

6b. The quadratic case

The same approach is unfortunately not so useful when the recourse function $E_\xi Q(w, \xi)$ is piecewise quadratic. In order to use a 'L-shaped' based technique, we consider the problem (21), (22), (23) at some iteration and look for efficient cuts (23). This is done by applying the subgradient inequality for convex functions. The functions θ_t^j are used to describe the function $z_t(w_t^j)$ which is piecewise quadratic in w_t^j when $q_{t+1}(\cdot, \cdot)$ is quadratic. This means, see Louveaux [8], that there exists a polyhedral decomposition of the constraint set (related to the constraints (22) and (23) containing w_t^j and θ_t^j) in a finite collection of closed convex sets C_{tk}^j, called

the cells of the decomposition, such that the intersection of two distinct cells has an empty interior and such that either the function $z_t^j(\cdot)$ is identically $-\infty$ or on each cell, the function $z_t^j(\cdot)$ is a positive semi-definite quadratic form, say $z_{tk}^j(\cdot)$. Let

$$v_{tk}^j = \arg\min\{z_{tk}^j(w_t^j) \mid w_t^j \in C_{tk}^j\}. \tag{28}$$

The subgradient inequality applied in v_{tk}^j implies

$$\theta_t^j \geq z_{tk}^j(v_{tk}^j) + \nabla z_{tk}^j(v_{tk}^j) \cdot (w_t^j - v_{tk}^j) \tag{29}$$

for all (w_t^j, θ_t^j) since $\nabla z_{tk}^j(v_{tk}^j) \in \partial z_t^j(v_{tk}^j)$, where ∇ and ∂ denote the gradient and subdifferential operators, respectively.

In the nested decomposition approach, the constraint (29) can be simplified since the relation

$$z_t^j(w_t^j) \leq z_{tk}^j(v_{tk}^j) \tag{30}$$

holds for an optimal vector w_t^j associated to $(\text{D.E.P.})_t^j$, the $(\text{D.E.P.})_t$ obtained at node j. Moreover, the resulting cut is efficient in the sense that either v_{tk}^j is an optimal point of $(\text{D.E.P.})_t^j$ or no interior point of C_{tk}^j satisfies this cut.

In the L-shaped approach (21), (22), (23), the relation (30) no longer holds since $z_t^j(w_t^j)$ or equivalently θ_t^j is only one term in the summation (21).

Therefore (29) cannot be further simplified. Using the L-shaped approach with (29) would mean approximating the piecewise quadratic objective function by linear constraints. The finite convergence property of MQSP no longer carries over and therefore the nested decomposition approach cannot be suggested to solve the extensive form of the aggregate level decision in BS-2 when the recourse functions associated with the detailed level decisions are quadratic.

Acknowledgments

The first two examples in Section 3 where block-separability can be obtained are derived from joint work with Prof. Y. Smeers, see [3, 10]. I gratefully acknowledge his support and suggestions in obtaining the above concepts and in applying them to these examples.

References

[1] J.F. Benders, "Partitioning procedures for solving mixed-variables programming problems", *Numerische Mathematik* 4 (1962) 238-252.

[2] J. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs", Technical Report 82-6, Department of Industrial and Operations Engineering, University of Michigan, 1982.

[3] X. de Groote, F. Louveaux, A.M. Poncelet and Y. Smeers, "A stochastic Energy R&D budget allocation model", DG XII, Commission of the European Communities, Bruxelles-Luxembourg, 1983.

[4] M. Dempster, *Stochastic programming* (Academic Press, London, 1980).
[5] S.J. Gartska and R. Wets, "On decision rules in stochastic programming", *Mathematical Programming* 7 (1974) 117-143.
[6] M.I. Kusy and W.T. Ziemba, "A bank asset and liability management model", IIASA, CP 83-59, December 1983.
[7] J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie, "A framework for the probabilistic analysis of hierarchical planning systems", Report 8311/0, Econometric Institute, Erasmus Universiteit Rotterdam, 1983.
[8] F.V. Louveaux, "Piecewise convex programs", *Mathematical Programming* 15 (1978) 53-62.
[9] F.V. Louveaux, "A solution method for multistage stochastic programs with recourse, with application to an energy investment problem", *Operations Research* 28 (1980) 889-902.
[10] F.V. Louveaux and Y. Smeers, "Stochastic optimization for the introduction of a new energy technology", to appear in *Stochastics*.
[11] A.S. Manne, R.G. Richels and J.P. Weyant, "Energy policy modeling: A survey", *Operations Research* 27 (1979) 1-34.
[12] P. Olsen, "Multistage stochastic program with recourse: The equivalent deterministic problem", *SIAM Journal on Control Optimization* 14 (1976) 495-517.
[13] P. Olsen, "When is a multistage stochastic programming problem well defined?", *SIAM Journal on Control Optimization* 14 (1976) 518-527.
[14] A. Prekopa, "Dynamic type stochastic programming models", in: A. Prékopa ed., *Studies in applied stochastic programming* (Hungarian Academy of Sciences, Budapest, 1978), 179-209.
[15] R.T. Rockafellar and R. Wets, "A dual solution procedure for quadratic stochastic programs with simple recourse", in: V. Pereyra and A. Reinoza, eds., *Numerical methods*, Lecture Notes in Mathematics, Vol. 1005, (Springer-Verlag, Berlin, 1983), pp. 252-265.
[16] R. Van Slyke and R. Wets, "*L*-shaped linear programs with applications control and stochastic programming", *SIAM Journal on Applied Mathematics* 17 (1969) 638-663.
[17] R. Wets, "Characterization theorems for stochastic programs", *Mathematical Programming* 2 (1972) 166-175.
[18] R. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming: The state of the art* (Springer-Verlag, Berlin, 1983) pp. 566-603.
[19] R.D. Wollmer, "Two stage linear programming under uncertainty with 0-1 integer first stage variables, *Mathematical Programming* 19 (1980) 279-288.

Mathematical Programming Study 28 (1986) 63-93
North-Holland

A LAGRANGIAN FINITE GENERATION TECHNIQUE FOR SOLVING LINEAR-QUADRATIC PROBLEMS IN STOCHASTIC PROGRAMMING

R.T. ROCKAFELLAR
Department of Mathematics, University of Washington, Seattle, WA 98115, USA
R.J.-B. WETS
Department of Mathematics, University of California, Davis, CA 95616, USA

Received 1 March 1984
Revised manuscript received 8 December 1984

A new method is proposed for solving two-stage problems in linear and quadratic stochastic programming. Such problems are dualized, and the dual, althought itself of high dimension, is approximated by a sequence of quadratic programming subproblems whose dimensionality can be kept low. These subproblems correspond to maximizing the dual objective over the convex hull of finitely many dual feasible solutions. An optimizing sequence is produced for the primal problem that converges at a linear rate in the strongly quadratic case. An outer algorithm of augmented Lagrangian type can be used to introduce strongly quadratic terms, if desired.

Key words: Stochastic Programming, Large-Scale Quadratic Programming, Lagrangian Methods.

1. Introduction

In the recourse model in stochastic programming, a vector x must be chosen optimally with respect to present costs and constraints as well as certain expected costs and induced constraints that are associated with corrective actions available in the future. Such actions may be taken in response to the observation of the values of various random variables about which there is only statistical information at the time x is selected. The actions involve costs and constraints that depend on these observed values and on x. The theory of this kind of stochastic programming and the numerical methods that have been proposed for it has been surveyed recently by Wets [12].

We aim here at developing a new solution procedure for the case where the first and second stage problems in the recourse model fit the mold of linear or quadratic (convex) programming. We assume for simplicity that the random variables are discretely distributed with only finitely many values. This restriction is not fully necessary in theory, but it reflects the realities of computation and a natural division among the questions that arise. Every continuous distribution must in practice be replaced by a finite discrete one, whether empirically, or through sampling, mathematical approximation, or in connection with the numerical calculation of integrals expressing expectations. The effects of such discretization raise important questions

This work was supported in part by grants from the National Science Foundation.

of convergence and statistical confidence in the solutions that are obtained, but such matters are best left to separate study.

We assume therefore that the probability space is a finite set Ω: the probability associated with an element $\omega \in \Omega$ is p_ω, and the expectation of a quantity u_ω that depends on ω is

$$Eu_\omega := \sum_{\omega \in \Omega} p_\omega u_\omega.$$

The fundamental problem we want to address is

$$\text{minimize} \quad c \cdot x + \tfrac{1}{2} x \cdot Cx + E\psi_\omega(x) \quad \text{over all} \quad x \in X \subset \mathbb{R}^n, \tag{P}$$

where X is a nonempty convex polyhedron, c is a vector in $\mathbb{R}^n$, C is a symmetric matrix in $\mathbb{R}^{n+n}$ that is positive semidefinite, and $\psi_\omega(x)$ is the minimum cost in a certain recourse subproblem that depends on ω and x. (Here $x \cdot y$ denotes the inner product of x and y.) We view this recourse subproblem as one of linear or quadratic programming, but instead of handling it directly we work with its dual. More will be said about this later (see Proposition 1 in Section 2 and the comments after its proof), but what counts in the end is the following: we suppose a representation

$$\psi_\omega(x) = \max_{z_\omega \in Z_\omega} \{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \tag{1.1}$$

is available, where Z_ω is a nonempty convex polyhedron in $\mathbb{R}^m$, T_ω is a matrix in $\mathbb{R}^{m \times n}$, h_ω is a vector in $\mathbb{R}^m$, and H_ω is a symmetric matrix in $\mathbb{R}^{m \times m}$ that is positive semidefinite. Such a formulation also covers important cases where, as will be explained presently, "recourse" is not the key idea and instead $\psi_\omega(x)$ arises when penalty expressions of a certain general type are introduced to restrain the difference vector $h_\omega - T_\omega x$. Note from the subscript ω that all the elements in the representation (1.1) are in principle allowed to be random, although a particular application might not involve quite so much randomness.

Two basic conditions are imposed on the given data. We assume X and C are such that for every $v \in \mathbb{R}^n$ the set

$$\xi(v) := \operatorname*{argmin}_{x \in X} \{v \cdot x + \tfrac{1}{2} x \cdot Cx\} \tag{1.2}$$

is nonempty and bounded. We also assume Z_ω, h_ω, T_ω, and H_ω, are such that for every $x \in X$ the set

$$\zeta_\omega(x) := \operatorname*{argmax}_{z\omega \in Z_\omega} \{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \tag{1.3}$$

is nonempty and bounded. Certainly the first condition holds if X is bounded or C is positive definite, and the second holds if Z_ω is bounded or H_ω is positive definite.

The first condition is quite innocuous, since in practice X can always be taken to be bounded. It implies that the function

$$\phi(v) = \inf_{x \in X} \{v \cdot x + \tfrac{1}{2} x \cdot Cx\}, \tag{1.4}$$

which will have a role in duality, is finite everywhere.

The second condition is more subtle, since it involves dual elements that might not be given directly but derived instead from a primal statement of the recourse subproblem that depends on x and ω. It ensures in particular that for every $x \in X$ and $\omega \in \Omega$, the optimal value $\psi_\omega(x)$ in this subproblem is finite, and an optimal recourse exists. This means that our stochastic programming problem (P) is one of *relatively complete recourse* [10]: there are no induced constraints on x that arise from the need to keep open the possibility of recourse at a later time.

Of course, if our problem were not one of relatively complete recourse, we could make it so by identifying the induced constraints and shrinking the set X until they were all satisfied. The smaller X would still be a convex polyhedron, although its description might be tedious in situations where special approaches such as in [10, Section 1] can't be followed. In this sense our second condition forces no real restriction on the problem either, except in requiring that the induced constraints, if any, be identified thoroughly in advance.

In some of the situations that motivate our model the recourse subproblem is actually trivial and its solution can be given in closed form. Such situations occur when constraints are represented by penalties: the term $E\psi_\omega(x)$ in (P) can then be interpreted as an *expected penalty*. Indeed, using the notation

$$\theta_\omega(u) = \max_{z_\omega \in Z_\omega} \{z_\omega \cdot u - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \tag{1.5}$$

we can write

$$\psi_\omega(x) = \theta_\omega(h_\omega - T_\omega x). \tag{1.6}$$

If $0 \in Z_\omega$, then

$$\theta_\omega(u) \geq 0 \quad \text{for all } u, \qquad \theta_\omega(0) = 0, \tag{1.7}$$

so we can view $\theta_\omega(h_\omega - T_\omega x)$ as a penalty attached to certain degrees or directions of deviation of $T_\omega x$ from the vector h_ω. Many useful penalty functions of linear-quadratic type can be expressed as in (1.5). In particular the case where $\theta_\omega(h_\omega - T_\omega x)$ is a sum of separate terms, one for each scalar component of the deviation vector $h_\omega - T_\omega x$, can be identified with the case where each Z_ω is a product of intervals and H_ω is diagonal. This case underlies a special model we have treated in [9].

The solution procedure that we shall present depends on a Lagrangian representation of problem (P) which leads to the dual problem

$$\begin{aligned} &\text{maximize} \quad \phi(c - ET^*_\omega z_\omega) + E\{z_\omega \cdot h_\omega - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \\ &\text{subject to} \quad z_\omega \in Z_\omega \ \text{ for all } \omega \in \Omega. \end{aligned} \tag{D}$$

Here ϕ is the function in (1.4), for which another representation will later be given (Proposition 2 in Section 2). The asterisk * signals the transpose of a matrix. The maximization in (D) takes place over the convex polyhedron

$$Z = \Pi_{\omega \in \Omega} Z_\omega \subset (\mathbb{R}^m)^\Omega; \tag{1.8}$$

we think of z_ω as the component in Z_ω of a point $z \in Z$. The vector space $(\mathbb{R}^m)^\Omega$ here, which is a product of copies of $\mathbb{R}^m$, one for each $\omega \in \Omega$, is likely to be of very high dimension, since the number of points in Ω may be very large. Despite this formidable dimensionality it is by way of (D), at least in concept, that we propose to solve (P). Properties of expectation, decomposition and quadratic structure, will make this possible. The relationship between (P) and (D) is explored in Section 2 along with other questions of quadratic programming duality that are crucial to in our formulation and our algorithm.

We approach problem (D) by a finite generation technique in which the feasible region Z is approximated from within by polytopes of comparatively low dimension, a *polytope* being a subset generated as the convex hull of finitely many points. This technique is presented in Section 3. It resembles the classical finite-element or Galerkin approach to the unconstrained maximization of a functional defined over an infinite-dimensional space, where one maximizes over finite-dimensional subspaces that grow in size as the approximation is refined. An important difference, however, is that in our case the new element or elements that are introduced at each stage in modifying the polytope over which we maximize are not obtained from some predetermined scheme, as classically, but identified in an 'adaptive' manner. Furthermore, the total number of elements used in generating the polytope does not have to keep increasing; the sequence of polytopes does not have to be nested. We prove in Section 4 that when the matrix C is positive definite these elements can readily be consolidated without threat to ultimate convergence, although the rate of progress may be better if a substantial set of generating elements is maintained. In this way the dimension of the subproblem to be solved in every iteration can be kept as low as seems desirable.

The subproblem of maximizing over a polytope can be represented as a standard type of quadratic programming problem and solved exactly by available codes. It yields as a byproduct an approximate solution vector for (P) along with bounds that provide a test of near optimality. The sequence of such approximate solutions converges to an optimal solution to (P). If not only C but also the matrices H_ω are positive definite, the rate of convergence is linear, in fact with guaranteed progress of a certain sort in every iteration, not just for the tail of the sequence.

In producing a new element to be used in the subrepresentation of Z in terms of a convex polytope, we have a particular x on hand and must carry out the maximization in (1.1) for every $\omega \in \Omega$. In other words, we must solve a large number of closely related linear or quadratic programming problems in $\mathbb{R}^m$. This could be a difficult task in general, but techniques such as have already been developed in connection with other approaches to stochastic programming problems of a more special nature (see Wets [12]) do offer hope. Furthermore, there are cases of definite interest where the maximization in (1.1) is trivial, for instance where Z_ω is a product of intervals and H_ω is diagonal. Such a case has been described in [11].

Not all of the problems we wish to solve have C and H_ω positive definite, but this does not prevent the application of our method and the achievement of a linear

rate of convergence. Augmented Lagrangian techniques [7] can be effective in approximating any problem (P) by a sequence of similar problems that do exhibit positive definiteness. We explain this in Section 5 after having established in Section 4 the results that show the advantages of the strongly quadratic case.

Our algorithm has been implemented successfully by Alan King on a VAX 11/780 at IIASA and at the University of Washington for solving quadratic stochastic programs with simple recourse. We have solved some product-mix test problems, and used it in the analysis of investment strategies to control the eutrophication process of a shallow lake. This last class of problems involved 56 decision variables, most of them with upper and lower bounds; the set X was determined by 35 linear constraints. The matrix $T \in \mathbb{R}^{4\times 56}$ and the vector $h \in R^4$ were random, whereas the (nonstochastic) quadratic term involving H was introduced as a result of the augmentation procedure suggested in Section 5. A report on this implementation and the numerical results that have been obtained has been written by A. King [4].

2. Lagrangian representation and duality

As the *Lagrangian* associated with problem (P) under the representation (1.1) of the recourse costs, we shall mean the function

$$L(x, z) = c \cdot x + \tfrac{1}{2} x \cdot Cx + E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}$$
$$\text{for } x \in X,\ z \in Z, \tag{2.1}$$

where Z is the convex polyhedron in (1.8). Clearly $L(x, z)$ is convex in x and concave in z, since C and H_ω are positive semidefinite. General duality theory [6] associates with L, X, and Z, the primal problem

$$\text{minimize} \quad F \text{ over } X, \quad \text{where } F(x) := \max_{z \in Z} L(x, z), \tag{2.2}$$

and the dual problem

$$\text{maximize} \quad G \text{ over } Z, \quad \text{where } G(z) := \min_{x \in X} L(x, z). \tag{2.3}$$

The functions F and G are convex and concave, respectively. Our assumptions in Section 1 allow us to write 'max' and 'min' in their definitions rather than 'sup' and 'inf'.

These problems turn out to be the ones already introduced. In terms of the notation in (1.2) and (1.3), we have

$$\operatorname*{argmax}_{z \in Z} L(x, z) = \{z \mid z_\omega \in \zeta_\omega(x), \text{ for all } w \in \Omega\}, \tag{2.4}$$

$$\operatorname*{argmin}_{x \in X} L(x, z) = \xi(c - ET_\omega^* z_v). \tag{2.5}$$

Moreover for $x \in X$ and $z \in Z$ we have

$$F(x) = c \cdot x + \tfrac{1}{2} x \cdot Cx + E\psi_\omega(x), \tag{2.6}$$

$$G(z) = \varphi(c - ET^*_\omega z_\omega) + E\{z_\omega \cdot h_\omega - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{2.7}$$

Thus the primal and dual problems (2.2) and (2.3) can be identified with (P) and (D), respectively.

In order to continue with our analysis of these problems, we need to step back briefly for a look at some basic facts about duality in quadratic programming, not only as they might apply to (P) and (D), but also to various subproblems in our schemes. A quadratic programming problem is usually defined as a problem in which a quadratic convex function is minimized (or a quadratic concave function maximized) subject to a system of linear constraints, or in other words, over a convex polyhedron. As is well known, such a problem has an optimal solution whenever its optimal value is finite (see Frank and Wolfe [3, Appendix (i)]); the Kuhn-Tucker conditions are both necessary and sufficient for optimality. For the purpose at hand, it is essential to adopt a more general point of view in which a problem is considered to fall in the category of quadratic programming as long as it can be *represented* in this traditional form, possibly through the introduction of auxiliary variables.

Consider an arbitrary Lagrangian of the form

$$l(u, v) = p \cdot u + q \cdot v + \tfrac{1}{2} u \cdot Pu - \tfrac{1}{2} v \cdot Qv - v \cdot Ru \quad \text{for } u \in U,\ v \in V, \tag{2.8}$$

where U and V are nonempty convex polyhedra, and P and Q are symmetric, positive semidefinite matrices. Let

$$f(u) = \sup_{v \in V} \{v \cdot [q - Ru] - \tfrac{1}{2} v \cdot Qv\}, \tag{2.9}$$

$$U_0 = \{u \,|\, f(u) \text{ finite}\} = \{u \,|\, \sup \text{ in (2.9) attained}\}, \tag{2.10}$$

$$g(v) = \inf_{u \in U} \{u \cdot [p - R^* v] - \tfrac{1}{2} u \cdot Pu\}, \tag{2.11}$$

$$V_0 = \{v \,|\, g(v) \text{ finite}\} = \{v \,|\, \inf \text{ in (2.11) attained}\}. \tag{2.12}$$

The primal and dual problems associated with l, U, and V by general duality theory can then we written as:

$$\text{minimize } p \cdot u + \tfrac{1}{2} u \cdot Pu + f(u) \text{ over } u \in U \cap U_0, \tag{P_0}$$

$$\text{maximize } q \cdot v - \tfrac{1}{2} v \cdot Qv + g(v) \text{ over } v \in V \cap V_0. \tag{D_0}$$

The following duality theorem for (P_0) and (D_0) extends the standard results in quadratic programming that were achieved by Dorn [2] and Cottle [1]. Those authors concentrated in effect on the case where U and V are orthants. The proof that we furnish is directed not only at an extension of theory, however. It explains how the optimal solutions to problems in the general framework of (P_0) and (D_0) can be identified in terms of the input and output of standard algorithms in quadratic

programming after a reformulation. This observation is essential in dealing with the variouus subproblems that will play a role in Section 3.

Theorem 1. *Problems* (P_0) *and* (D_0) *are representable as quadratic programming in the traditional sense. If* (P_0) *and* (D_0) *both have feasible solutions, or if either* (P_0) *or* (D_0) *has finite optimal value, then both have optimal solutions, and*

$$\min(P_0) = \max(D_0).$$

This occurs if and only if the Lagrangian l *has a saddle point* $(\bar{u}, \bar{v})$ *relative to* $U \times V$, *in which case the saddle value* $l(\bar{u}, \bar{v})$ *coincides with the common optimal value in* (P_0) *and* (D_0), *and the saddle points are the pairs* $(\bar{u}, \bar{v})$ such that $\bar{u}$ *is an optimal solution to* (P_0) *and* $\bar{v}$ *is an optimal solution to* (D_0).

Proof. General duality theory [8] assures us that $\inf(P_0) \geq \sup(D_0)$ and in particular that both (P_0) and (D_0) have finite optimal value if both have feasible solutions. It also informs us that $(\bar{u}, \bar{v})$ is a saddle point of l on $U \times V$ if and only if $\bar{u}$ is an optimal solution to (P_0), $\bar{v}$ is an optimal solution to (D_0), and $\min(P_0) = \max(D_0)$, this common optimal value then being equal to $l(\bar{u}, \bar{v})$. We know further that a quadratic programming problem in the traditional sense has an optimal solution if it has finite optimal value [3, Appendix (i)]. The Kuhn-Tucker conditions are both necessary and sufficient for optimality in such a problem, because the constraint system is linear. The proof of the theorem can be reduced therefore to demonstrating that (P_0) and (D_0) are representable as quadratic programming in the traditional sense and in such a manner that the Kuhn-Tucker conditions for either problem correspond to the saddle point condition for l on $U \times V$.

The sets U and V are associated with systems of linear constraints that can be expressed in various ways, but to be specific we can suppose that

$$U = \{u \in \mathbb{R}^n \mid Au \geq a\} \neq \emptyset \quad \text{and} \quad V = \{v \in \mathbb{R}^m \mid B^* v \leq b\} \neq \emptyset, \tag{2.13}$$

where A is $m' \times n$ and B is $m \times n'$. Let $u' \in \mathbb{R}^{n'}$ and $v' \in \mathbb{R}^{m'}$ be Lagrange multiplier vectors paired with the conditions $B^* v \leq b$ and $Au \geq a$, respectively.

Formula (2.9) gives $f(u)$ as the optimal value in a classical quadratic programming problem in v. The optimal solutions to this problem are vectors that satisfy the usual Kuhn-Tucker conditions, or in other words, correspond to saddle points of the Lagrangian

$$v \cdot [q - Ru] - \tfrac{1}{2} v \cdot Qv + u' \cdot [b - B^* v] = b \cdot u' + v \cdot [q - Ru - Bu'] - \tfrac{1}{2} v \cdot Qv \tag{2.14}$$

relative to $u' \in \mathbb{R}^{n'}_+$ and $v \in \mathbb{R}^m$. In partcular, then, we have

$$f(u) = \inf_{u' \in \mathbb{R}^{n'}_+} \sup_{v \in \mathbb{R}^m} \{b \cdot u' + v \cdot [q - Ru - Bu'] - \tfrac{1}{2} v \cdot Qv\}. \tag{2.15}$$

The inner supremum here is attained whenever finite, and it is attained at a point $v = u''$. Thus it equals inf unless there exists a vector $u'' \in \mathbb{R}^m$ such that $[q - Ru - Bu'] -$

$Qu'' = 0$, in which case it equals $b \cdot u' + \frac{1}{2}u'' \cdot Qu''$, a value that actually depends only on u and u'. We may conclude that

$$U_0 = \{u \in \mathbb{R}^n \mid \exists u' \in \mathbb{R}^{n'}_+, \text{ with } Ru + Bu' + Qu'' = q\}, \tag{2.16}$$

$$f(u) = \text{minimum of } b \cdot u' + \tfrac{1}{2}u'' \cdot Qu''$$
$$\text{subject to } u' \in \mathbb{R}^{n'}_+,\ u'' \in \mathbb{R}^m,\ Ru + Bu' + Qu'' = q. \tag{2.17}$$

We can therefore represent (P_0) as

$$\begin{aligned} &\text{minimize} \quad p \cdot u + \tfrac{1}{2}u \cdot Pu + b \cdot u' + \tfrac{1}{2}u'' \cdot Qu'' \\ &\text{subject to} \quad Au \geq a,\ u' \geq 0,\ Ru + Bu' + Qu'' = q, \end{aligned} \tag{$\tilde{P}_0$}$$

where the value of $u'' \cdot Qu''$ does not depend on the particular choice of the vector u'' satisfying $Ru + Bu' + Qu'' = q$ but only on u and u'. This is a quadratic programming problem in the usual sense, but in which u'' is a sort of vector of dummy variables that can be eliminated, if desired. In any case it follows that (P_0) has an optimal solution if its optimal value is finite, inasmuch as this property holds for $(\tilde{P}_0)$.

The optimal solutions $(\bar{u}, \bar{u}', \bar{u}'')$ to $(\tilde{P}_0)$ are characterized by the Kuhn-Tucker conditions that involve multiplier vectors $\bar{v}$ for the constraint $Ru + Bu' + Qu'' = q$ and $\bar{v}'$ for the constraint $Au \geq a$. These conditions take the form:

$$A\bar{u} \geq a, \quad \bar{v}' \geq 0, \quad \bar{v}' \cdot [A\bar{u}' - a] = 0,$$
$$\bar{u}' \geq 0, \quad B^*\bar{v} \leq b, \quad \bar{u}' \cdot [B^*\bar{v} - b] = 0,$$
$$R^*\bar{v} + A^*\bar{v}' - P\bar{u} = p, \quad R\bar{u} + B\bar{u}' + Q\bar{u}'' = q, \quad Q\bar{u}'' = Q\bar{v}.$$

Because of the final condition we can write the next-to-last condition instead as $R\bar{u} + B\bar{u}' + Q\bar{v} = q$. Note that there is no restriction then on $\bar{u}''$, except that $Q\bar{u}'' = Q\bar{v}$; we could always take $\bar{u}'' = \bar{v}$ in particular. This is in keeping with our observation that $(\tilde{P}_0)$ is really just a problem in u and u'. We see in fact that the pairs $(\bar{u}, \bar{u}')$ which are optimal for $(\tilde{P}_0)$ are the ones which, for some pair $(\bar{v}, \bar{v}')$, satisfy the conditions

$$\begin{aligned} &A\bar{u} \geq 0, \quad \bar{v}' \geq 0, \quad \bar{v}' \cdot [A\bar{u}' - a] = 0, \\ &\bar{u}' \geq 0, \quad B^*\bar{v} \leq b, \quad \bar{u}' \cdot [B^*\bar{v} - b] = 0, \\ &R\bar{u} + B\bar{u}' + Q\bar{v} = q, \quad R^*\bar{v} + A^*\bar{v}' - P\bar{u} = p. \end{aligned} \tag{2.18}$$

Problem (D_0) can be understood in the same way. From the formula (2.11) for $g(v)$ we deduce that

$$V_0 = \{v \in \mathbb{R}^m \mid \exists v' \in \mathbb{R}^{m'}_+,\ v'' \in \mathbb{R}^N, \text{ with } R^*v + B^*v' - Pv'' = p\}, \tag{2.19}$$

$$g(v) = \text{maximum of } a \cdot v' - \tfrac{1}{2}v'' \cdot Pv''$$
$$\text{subject to } v' \in \mathbb{R}^{m'}_+,\ v'' \in \mathbb{R}^n,\ R^*v + A^*v' - Pv'' = p. \tag{2.20}$$

These formulas yield for (D_0) the representation

$$\begin{aligned} &\text{maximize} \quad q \cdot v - \tfrac{1}{2} v \cdot Qv + a \cdot v' - \tfrac{1}{2} v'' \cdot Pv'' \\ &\text{subject to} \quad B^* v \leq b, \; v' \geq 0, \; R^* v + A^* v' - Pv'' = p, \end{aligned} \tag{$\tilde{D}_0$}$$

where the value of $v'' \cdot Pv''$ does not depend on the particular v'' satisfying $R^*v + A^*v' - Pv'' = p$ but only on v and v'. This is really a problem in v and v', and the Kuhn-Tucker conditions characterize $\bar{v}$ and $\bar{v}'$ as optimal if and only if there exist $\bar{u}$ and $\bar{u}'$ such that (2.18) holds, the same conditions as before. Since $(\tilde{D}_0)$ is a quadratic programming problem in the usual sense, it has an optimal solution whenever its optimal value is finite, and (D_0) therefore has this property too.

Our argument demonstrates that if either (P_0) or (D_0) has finite optimal value, then both problems have optimal solutions. The optimal solutions in both cases are characterized by the existence of auxiliary vectors such that (2.18) holds. But (2.18) can also be seen as the Kuhn-Tucker conditions for $(\bar{u}, \bar{v})$ to be a saddle point of the Lagrangian (2.8), when U and V are given by (2.13). Thus for $\bar{u}$ and $\bar{v}$ to be optimal solutions to (P_0) and (D_0) respectively, it is necessary and sufficient that $(\bar{u}, \bar{v})$ be a saddle point in (2.18). Following on the remarks at the beginning of the proof, this establishes the theorem. □

Corollary. *Any standard quadratic programming method can in principle be used to solve problems of the form* (P_0) *or* (D_0), *in fact both simultaneously, thereby determining a saddle point of the corresponding Lagrangian l on* $U \times V$, *if such a saddle point exists.*

Proof. The representations in the proof of the theorem show more specifically that if an algorithm is applied to $(\tilde{P}_0)$, the optimal solution vectors $\bar{u}$, $\bar{u}'$ and multiplier vectors $\bar{v}$ and $\bar{v}'$ which it produces yield optimal solutions $\bar{u}$ to (P_0) and $\bar{v}$ to (D_0), and $(\bar{u}, \bar{v})$ is a saddle point in (2.8). The same holds if an algorithm is applied to $(\tilde{D}_0)$, except that then $\bar{v}$ and $\bar{v}'$ are the optimal solution vectors, whereas $\bar{u}$ and $\bar{u}'$ are the multiplier vectors. □

Theorem 2. *The stochastic programming problems* (P) *and* (D) *are representable as quadratic programming problems in the traditional sense, although with potentially very high dimensionality. Both problems have optimal solutions, and*

$$\min(P) = \max(D).$$

A pair $(\bar{x}, \bar{z})$ *is a saddle point of the Lagrangian L relative to* $X \times Z$ *if and only if* $\bar{x}$ *is an optimal solution to* (P) *and* $\bar{z}$ *is an optimal solution to* (D). *The set of such pairs* $(\bar{x}, \bar{z})$ *is bounded.*

Proof. We need only observe that the triple L, X, Z, can be construed as a special case of the triple l, U, V, in Theorem 1. A term like $Ez_\omega \cdot H_\omega z_\omega$ can be expressed as $z \cdot Qz$ for certain matrix Q, and so forth. Our assumption that the extremal sets $\xi(v)$

in (1.2) and $\zeta_\omega(x)$ in (1.3) are nonempty for all $v \in \mathbb{R}^n$, $x \in X$ and $\omega \in \Omega$, guarantees that every $x \in X$ is feasible for (P), and every $z \in Z$ is feasible for (D). Therefore we are in the case of Theorem 1 where both problems have feasible solutions.

As for the boundedness of the set of saddle points $(\bar{x}, \bar{z})$, consider a particular pair of optimal solutions $\bar{x}^*$ and $\bar{z}^*$ to (P) and (D). Observe that for every optimal solution $\bar{x}$ to (P), $(\bar{x}, \bar{z}^*)$ is a saddle point and therefore satisfies

$$\bar{x} \in \operatorname*{argmin}_{x \in X} L(x, \bar{z}^*) = \xi(c - ET^*_\omega \bar{z}^*_\omega)$$

(cf. (2.5)). But the set on the right is bounded (one of our basic assumptions in Section 1). Likewise for every optimal solution $\bar{z}$ to (D), $(\bar{x}^*, \bar{z})$ is a saddle point and therefore satisfies

$$\bar{z} \in \operatorname*{argmax}_{z \in Z} L(\bar{x}^*, z), \quad \text{so } \bar{z}_\omega \in \zeta_\omega(\bar{x}^*) \text{ for all } w \in \Omega.$$

(cf. (2.4)). The sets $\zeta_\omega(\bar{x}^*)$ are all bounded (again by one of our basic assumptions in Section 1), so $\bar{z}$ belongs to a certain bounded set. The pairs $(\bar{x}, \bar{z})$ thus all belong to a product of bounded sets dependent only on $\bar{x}^*$ and $\bar{z}^*$. □

The following pair of results will help to clarify the quadratic programming nature of problems (P) and (D).

Proposition 1. *For the function ψ_ω given by* (1.1), *if the polytope Z_ω has a representation*

$$Z_\omega = \{z_\omega \in \mathbb{R}^m \mid B^*_\omega z_\omega \leq b_\omega\} \tag{2.21}$$

for some vector $b_\omega \in \mathbb{R}^s$ and matrix $B_\omega \in \mathbb{R}^{m \times s}$ (with s independent of ω), then ψ_ω has an alternative expression of the form

$$\begin{aligned} \psi_\omega(x) = \text{minimum of } & d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega \\ \text{subject to } & y_\omega \in Y_\omega, \ \bar{T}_\omega x + W_\omega y_\omega = \bar{h}_\omega, \end{aligned} \tag{2.22}$$

for certain vectors $d_\omega \in \mathbb{R}^s$, $\bar{h}_\omega \in \mathbb{R}^q$, and matrices $\bar{T}_\omega \in \mathbb{R}^{q \times n}$, $W_\omega \in \mathbb{R}^{q \times s}$, and $D_\omega \in \mathbb{R}^{s \times s}$ with D_ω symmetric and positive semidefinite, and where

$$Y_\omega = \{y_\omega \in \mathbb{R}^s \mid A_\omega y_\omega \geq a_\omega\} \tag{2.23}$$

for some $a_\omega \in \mathbb{R}^p$ and $A_\omega \in \mathbb{R}^{p \times s}$.

Conversely, any function ψ_ω having a representation (2.22) *as just described (with $\psi_\omega(x)$ finite for all $x \in X$) also has a representation* (1.1) *with Z_ω of the form* (2.21).

Proof. Starting with the representation (1.1) and Z_ω of the form (2.21), view the maximization problem in (1.1) as the dual problem associated with the Lagrangian

$$l_{x,\omega}(u_\omega, z_\omega) = u_\omega \cdot [b_\omega - B^*_\omega z_\omega] + z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega$$

$$\text{for } u_\omega \in \mathbb{R}^s_+ \text{ and } z_\omega \in \mathbb{R}^m.$$

The corresponding primal problem, whose optimal value is also equal to $\psi_\omega(x)$ by Theorem 1 (as long as $x \in X$, so that $\psi_\omega(x)$ is finite by assumption) is

$$\text{minimize} \quad b_\omega \cdot u_\omega + f_\omega(u_\omega) \quad \text{over} \quad u_\omega \in \mathbb{R}^s_+, \quad \text{where}$$

$$f_\omega(u_\omega) = \sup_{z_\omega \in \mathbb{R}^m} \{z_\omega \cdot [h_\omega - T_\omega x - B_\omega u_\omega] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}.$$

Using the trick in the proof of Theorem 1, we can reformulate the latter as

$$\text{minimize} \quad b_\omega \cdot u_\omega + \tfrac{1}{2} u''_\omega \cdot H_\omega u''_\omega$$

$$\text{subject to} \quad u_\omega \in \mathbb{R}^s_+, \; u''_\omega \in \mathbb{R}^m, \; B_\omega u_\omega + H_\omega u''_\omega = h_\omega - T_\omega x.$$

We can then pass to form (2.22) in terms of $y_\omega = (u_\omega, u''_\omega)$ (or by setting $y_\omega = u_\omega$ after algebraic elimination of u''_ω, if the rank of H_ω is the same for all $\omega \in \Omega$).

Starting with the representation (2.22) and Y_ω of the form (2.23), on the other hand, we can view $\psi_\omega(x)$ as the optimal value for the primal problem associated with the Lagrangian

$$l_{x,\omega}(y_\omega, v_\omega) = d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega + v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x - W_\omega y_\omega]$$

$$\text{for } y_\omega \in Y_\omega \text{ and } v_\omega \in \mathbb{R}^q.$$

Then $\psi_\omega(x)$ (when finite) is also the optimal value in the corresponding dual problem

$$\text{maximize} \quad v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x] + g_\omega(v_\omega) \quad \text{over} \quad v_\omega \in R^q, \quad \text{where}$$

$$g_\omega(v_\omega) = \inf_{y_\omega \in Y_\omega} \{y_\omega \cdot [d_\omega - W^*_\omega v_w] + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega\}.$$

As we saw in the proof of Theorem 1, this problem can also be written as

$$\text{maximize} \quad v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x] + v'_\omega \cdot a_\omega - \tfrac{1}{2} v''_\omega \cdot D_\omega v''_\omega$$

$$\text{subject to} \quad v_\omega \in \mathbb{R}^q, \; v'_\omega \in \mathbb{R}^p_+, \; W^*_\omega v_\omega + A^*_\omega v'_\omega + D_\omega v''_\omega = d_\omega.$$

With $z_\omega = (v_\omega, v'_\omega, v''_\omega)$, this can be brought into the form (1.1) with Z_ω as in (2.21). (Alternatively one could take $z_\omega = (v_\omega, v'_\omega)$ and eliminate v''_ω algebraically, provided that the rank of D_ω is independent of ω. If also the rank of the matrix W_ω is independent of ω, one could even eliminate v_ω from the problem and just take $z_\omega = v'_\omega$ to get a representation (1.1) in fewer variables.) □

Proposition 2. *The function φ in* (1.4) *also has a representation*

$$\varphi(v) = \text{maximum of } q \cdot u - \tfrac{1}{2} u \cdot Qu \text{ over all } u \in U \text{ satisfying } Bu = v$$

for some choice of vectors b and q and matrices B and Q with Q symmetric and positive semidefinite, where U is a convex polyhedron.

Proof. Recall that $\varphi(v)$ is finite for all v by assumption. Express X as $\{x \in \mathbb{R}^n \mid Ax \geq a\}$ for some $a \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times n}$, and consider the Lagrangian

$$l_v(x, u') = v \cdot x + \tfrac{1}{2} x \cdot Cx + u' \cdot [a - Ax] \quad \text{for } x \in \mathbb{R}^n \text{ and } u' \in \mathbb{R}^p_+.$$

The primal problem associated with this Lagrangian is the minimization problem in (1.4), whereas the dual problem, which also has $\varphi(v)$ as its optimal value, is

$$\text{maximize} \quad a \cdot u' + g(u') \quad \text{over} \quad u' \in \mathbb{R}^p_+,$$
$$\text{where } g(u') = \inf_{x \in R^n} \{x \cdot [v - A^* u'] + \tfrac{1}{2} x \cdot Cx\}.$$

The reformulation trick in Theorem 1 translates this into

$$\text{maximize} \quad a \cdot u' - \tfrac{1}{2} u'' \cdot Cu''$$
$$\text{subject to} \quad u' \in \mathbb{R}^p_+, \ u'' \in \mathbb{R}^n, \ A^* u' - Cu'' = v.$$

We can then get a representation (2.24) in terms of $u = (u', u'')$. □

Propositions 1 and 2 make possible a more complete description of the quadratic programming representation of problems (P) and (D) indicated in Theorem 2. When $\psi_\omega(x)$ is expressed in terms of a recourse subproblem in y_ω as in Proposition 1, we can identify (P) with the problem

$$\begin{aligned} &\text{minimize} \quad c \cdot x + \tfrac{1}{2} x \cdot Cx + E\{d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega\} \\ &\text{subject to} \quad x \in X, \ y_\omega \in Y_\omega, \ \bar{T}_\omega x + W_\omega \cdot y_\omega = \bar{h}_\omega \ \text{ for all } \omega \in \Omega. \end{aligned} \tag{2.25}$$

Similarly, when φ is expressed as in Proposition 2 we can pose (D) as

$$\begin{aligned} &\text{maximize} \quad q \cdot u - \tfrac{1}{2} u \cdot Qu + E\{z_\omega \cdot h_\omega - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \\ &\text{subject to} \quad u \in U, \ z_\omega \in Z_\omega, \ \text{and} \ \ Bu + E\{T^*_\omega z_\omega\} = c. \end{aligned} \tag{2.26}$$

In the latter, our assumption that $\varphi(v)$ is finite for all $v \in \mathbb{R}^n$ implies that no matter what the choice of vectors $z_\omega \in Z_\omega$, there does exist a $u \in U$ such that the constraint $Bu + E\{T^*_\omega z_w\} = c$ is satisfied.

3. Finite generation method

Our aim is to solve problem (P) by way of (D) according to the following scheme. We replace (D) by a sequence of subproblems

$$\text{maximize} \quad G(z) \quad \text{over all} \quad z \in Z^\nu \subset Z \tag{D$^\nu$}$$

for $\nu = 1, 2, \ldots$, where G is the dual objective function in (2.3) and (2.7), and Z^ν is a polytope of relatively low dimension generated as the convex hull of finitely many points in Z. Obviously (D$^\nu$) is the dual of the problem

$$\text{minimize} \quad F^\nu(x) \quad \text{over all} \quad x \in X, \tag{P$^\nu$}$$

where F^ν is obtained by substituting Z^ν for Z in the formula (2.2) for the primal

objective function F:

$$F^\nu(x) = \max_{z \in Z^\nu} L(x, z)$$

$$= c \cdot x + \tfrac{1}{2} x \cdot Cx + \max_{z \in Z^\nu} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{3.1}$$

Indeed, (P^ν) and (D^ν) are the primal and dual problems that correspond to L on $X \times Z^\nu$ rather than $X \times Z$. In calculating a solution $\bar{z}^\nu$ to (D^ν) we obtain also a solution $\bar{x}^\nu$ to (P^ν) that can be viewed as an approximately optimal solution to (P). From $\bar{z}^\nu$ and $\bar{x}^\nu$ we gain information that helps in determining the polytope $Z^{\nu+1}$ to be used in the next iteration. The new polytope $Z^{\nu+1}$ is not necessarily 'larger' than Z^ν.

Problems (P^ν) and (D^ν) belong to the realm of 'generalized' quadratic programming as demarcated in Section 2. Clearly

$$F(x) \geq F^\nu(x) \quad \text{for all } x, \tag{3.2}$$

where F is the primal objective function in (2.2) and (2.6), so (P^ν) can be regarded as a 'lower envelope approximation' to (P). The feasible sets in (P^ν) and (D^ν) are X and Z^ν, respectively, whereas the ones in (P) and (D), are X and Z. From Theorem 1, therefore, we know that optimal solutions $\bar{x}^\nu$ and $\bar{z}^\nu$ to (P^ν) and (D^ν) exist and satisfy

$$F^\nu(\bar{x}^\nu) = G(\bar{z}^\nu), \tag{3.3}$$

$$\bar{x}^\nu \in \operatorname*{argmin}_{x \in X} F^\nu(x) \subset \operatorname*{argmin}_{x \in X} L(x, \bar{z}^\nu), \tag{3.4}$$

$$\bar{z}^\nu \in \operatorname*{argmax}_{z \in Z^\nu} G(z) \subset \operatorname*{argmax}_{z \in Z^\nu} L(\bar{x}^\nu, z). \tag{3.5}$$

Having determined a pair $(\bar{x}^\nu, \bar{z}^\nu)$ of this type, which is a saddle point of L relative to $X \times Z^\nu$, we can test whether it is actually a saddle point of L relative to $X \times Z$. This amounts to checking the maximum of $L(\bar{x}^\nu, z)$ over all $z \in Z$ to see if it occurs at $z = \bar{z}^\nu$. If yes, $\bar{x}^\nu$ and $\bar{z}^\nu$ are optimal solutions to (P) and (D), and we are done. If no, we obtain from the test an element

$$z^\nu \in \operatorname*{argmax}_{z \in Z} L(\bar{x}^\nu, z) \tag{3.6}$$

and have

$$L(\bar{x}^\nu, \bar{z}^\nu) < L(\bar{x}^\nu, z^\nu) = F(\bar{x}^\nu). \tag{3.7}$$

The crucial feature that makes the test possible is the decomposition in (2.4): maximizing $L(\bar{x}^\nu, z)$ in $z \in Z$ reduces a solving a separate quadratic programming problem (perhaps trivial) in $z_\omega \in Z_\omega$ for each $\omega \in \Omega$. Anyway, with such a z^ν we have

$$F(x) \geq L(x, z^\nu) \quad \text{for all } x, \text{ with equality when } x = \bar{x}^\nu. \tag{3.8}$$

We can use this in conjunction with (3.3) in constructing a new lower envelope approximation $F^{\nu+1}$ for F, which in primal terms is what is involved in constructing a new set $Z^{\nu+1}$ to replace Z^ν. More will be said about this later.

Of course the optimality test also furnishes a criterion for termination with suboptimal solutions, if desired. Since $\bar{x}^\nu$ and $\bar{z}^\nu$ are feasible solutions to (P) and (D) and satisfy (by Theorem 2)

$$F(\bar{x}^\nu) \geqslant \min(\mathrm{P}) = \max(\mathrm{D}) \geqslant G(\bar{z}^\nu), \tag{3.9}$$

we know that for $\varepsilon_\nu = F(\bar{x}^\nu) - G(\bar{z}^\nu)$, both $\bar{x}^\nu$ and $\bar{z}^\nu$ are ε_ν-optimal:

$$|F(\bar{x}^\nu) - \min(\mathrm{P})| \leqslant \varepsilon_\nu \quad \text{and} \quad |G(\bar{x}^\nu) - \max(\mathrm{D})| \leqslant \varepsilon_\nu. \tag{3.10}$$

Our basic procedure can be summarized now as follows.

Algorithm

Step 0 (Initialization). Choose the optimality test parameter $\bar{\varepsilon} \geqslant 0$ and the initial convex polytope $Z^1 \subset Z$. Set $\nu = 1$.

Step 1 (Approximate Solution). Determine a saddle point $(\bar{x}^\nu, \bar{z}^\nu)$ of L relative to $X \times Z^\nu$ and the value $\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu)$.

Step 2 (Decomposition). For each $\omega \in \Omega$, determine an optimal solution z_ω^ν to the problem

$$\text{maximize} \quad z_\omega \cdot [h_\omega - T_\omega \bar{x}^\nu] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega \text{ over } z_\omega \in Z_\omega \tag{3.11}$$

and the optimal value α_ω^ν. Let z^ν be the element of Z having component z_ω^ν in Z_ω, and let

$$\alpha_\nu = c \cdot \bar{x}^\nu + \tfrac{1}{2} \bar{x}^\nu \cdot C\bar{x}^\nu + E\alpha_\omega^\nu = L(\bar{x}^\nu, z^\nu). \tag{3.12}$$

Step 3 (Optimality Test). Let $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu$. Then $\bar{x}^\nu$ is an ε_ν-optimal solution to (P), $\bar{z}^\nu$ is an ε_ν-optimal solution to (D), and

$$\alpha_\nu \geqslant \min(\mathrm{P}) = \max(\mathrm{D}) \geqslant \bar{\alpha}_\nu. \tag{3.13}$$

If $\varepsilon_\nu \leqslant \bar{\varepsilon}$ terminate.

Step 4 (Polytope Modification). Choose a new convex polytope $Z^{\nu+1}$ that contains both $\bar{z}^\nu$ and z^ν, although not necessarily all of Z^ν. Replace ν by $\nu+1$; return to Step 1.

We proceed to comment on these algorithmic steps individually in more detail, one by one. Properties of the algorithm as a whole will be developed in Section 4 and Section 5.

The most important observation concerns the quadratic programming nature of the subproblem solved in Step 1. Suppose that Z^ν is generated from certain elements $\tilde{z}_k^\nu \in Z$:

$$Z^\nu = \mathrm{co}\{\tilde{z}_k^\nu | k = 1, \ldots, m_\nu\} = \left\{ \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu \,|\, \lambda_k \geqslant 0, \sum_{k=1}^{m_\nu} \lambda_k = 1 \right\}. \tag{3.14}$$

Finding a saddle point $(\bar{x}^\nu, \bar{z}^\nu)$ of $L(x, z)$ relative to $x \in X$ and $z \in Z^\nu$ is equivalent

to finding a saddle point $(\bar{x}^\nu, \bar{\lambda}^\nu)$ of

$$L^\nu(x, \lambda) = L\left(x, \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu\right) \tag{3.15}$$

relative to $x \in X$ and $\lambda \in \Lambda^\nu$, where Λ^ν is the unit simplex in $\mathbb{R}^{m_\nu}$,

$$\Lambda^\nu := \left\{\lambda = (\lambda_1, \ldots, \lambda_{m_\nu}) \,\middle|\, \lambda_k \geq 0, \sum_{k=1}^{m_\nu} \lambda_k = 1\right\}, \tag{3.16}$$

and then setting

$$\bar{z}^\nu = \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu. \tag{3.17}$$

But from the definition (2.1) of $L(x, z)$ we have

$$\begin{aligned} L^\nu(x, \lambda) &= c \cdot x + \tfrac{1}{2} x \cdot Cx + \sum_{k=1}^{m_\nu} \lambda_k E\{\tilde{z}_{k\omega}^\nu \cdot [h_\omega - T_\omega x]\} \\ &\quad - \tfrac{1}{2} \sum_{j=1}^{m_\nu} \sum_{k=1}^{m_\nu} \lambda_j \lambda_k E\{\tilde{z}_{k\omega}^\nu \cdot H_\omega \tilde{z}_{k\omega}^\nu\} \\ &= c \cdot x + \tfrac{1}{2} x \cdot Cx + \lambda \cdot [\tilde{h}^\nu - \tilde{T}^\nu x] - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda, \end{aligned} \tag{3.18}$$

where

$$\tilde{h}^\nu \in \mathbb{R}^{m_\nu} \quad \text{with components } \tilde{h}_k^\nu = E\{\tilde{z}_{k\omega}^\nu \cdot h_\omega\}, \tag{3.19}$$

$$\tilde{H}^\nu \in \mathbb{R}^{m_\nu \times m_\nu} \quad \text{with entries } \tilde{H}_{jk}^\nu = E\{\tilde{z}_{j\omega}^\nu \cdot H_\omega \tilde{z}_{k\omega}^\nu\}, \tag{3.20}$$

$$\tilde{T}^\nu \in \mathbb{R}^{m_\nu \times n} \quad \text{with entries } \tilde{T}_{ki}^\nu = E\{\tilde{z}_{k\omega}^\nu \cdot T_\omega^i\}, \tag{3.21}$$

T_ω^i being the ith column of the matrix $T_\omega \in \mathbb{R}^{m \times m}$. Problem (D^ν) thus reduces to a *deterministic quadratic programming problem in which the coefficients are certain expectations*:

$$\text{maximize} \quad \phi(c - \tilde{T}^{\nu *} \lambda) - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda \quad \text{over all } \lambda \in \Lambda^\nu. \tag{$\tilde{D}^\nu$}$$

Here φ is the function in (1.4), which has alternative representations such as in Proposition 2 that can be used to place $(\tilde{D}^\nu)$ in a more traditional quadratic programming format. Regardless of such reformulation, the dimensionality of this quadratic programming problem will be relatively low as long as m_ν, the number of elements $\tilde{z}_k^\nu$ used in generating Z^ν, is kept modest.

The translation of (D^ν) into $(\tilde{D}^\nu)$ also sheds light on the lower envelope function F^ν in the approximate primal subproblem (P^ν):

$$F^\nu(x) = \max_{\lambda \in \Lambda^\nu} L(x, \lambda) = c \cdot x + \tfrac{1}{2} x \cdot Cx + \Psi^\nu(x), \tag{3.22}$$

where

$$\Psi^\nu(x) := \max_{\lambda \in \Lambda^\nu} \{\lambda \cdot [\tilde{h}^\nu - \tilde{T}^\nu x] - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda\} = \max_{z \in Z^\nu} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{3.23}$$

Clearly $\Psi^\nu(x)$ is a lower envelope approximation to the recourse cost function

$$\Psi(x) := \max_{z \in Z} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} = E\psi_\omega(x). \tag{3.24}$$

Especially worth noting in (3.23) is the case where there are no quadratic terms $z_\omega \cdot H_\omega z_\omega$, i.e. where $H_\omega = 0$ for all $\omega \in \Omega$ and consequently $\tilde{H}^\nu = 0$. Then

$$\Psi^\nu(x) = \max_{k=1,\dots,m_\nu} \{\tilde{h}^\nu_k - \tilde{T}^\nu_k \cdot x\}, \tag{3.25}$$

where $\tilde{T}^\nu_k$ is the vector in $\mathbb{R}^n$ given by the kth row of the matrix $\tilde{T}^\nu$ in (3.21):

$$\tilde{T}^\nu_k = E\{T^*_\omega z^\nu_{k\omega}\}. \tag{3.26}$$

In this case Ψ^ν is a *polyhedral* convex envelope representation of Ψ, the pointwise maximum of a collection of affine functions

$$l_k(x) = \tilde{h}^\nu_k - \tilde{T}^\nu_k \cdot x \quad \text{for } k = 1, \dots, m_\nu.$$

Our technique then resembles a cutting-plane method, at least as far as the function Ψ is concerned.

Indeed, if not only $H_\omega = 0$ but $C = 0$, so that there are no quadratic cost terms at all and (P) is a purely linear stochastic programming problem, we can regard F^ν as a polyhedral convex subrepresentation of F. Then the subproblems (P^ν) and ($\tilde{D}^\nu$) can be solved by linear rather than quadratic programming algorithms. Furthermore the function $L(x, z^\nu)$ determined in (3.8) is then affine in x. If in fact we were to take $Z^{\nu+1} = \text{co}\{Z^\nu, z^\nu\}$, we would get

$$F^{\nu+1}(x) = \max\{F^\nu(x), L(x, z^\nu)\},$$

and this would truly be a cutting-plane method applied to problem (P).

It must be remembered, though, that in such a cutting-plane approach it might generally be necessary to retain more and more affine functions in the polyhedral approximation to F, since the conditions that theoretically validate the dropping of earlier cutting-planes might not be met. The dimension of the linear programming subproblem to be solved in each iteration would become progressively larger. In contrast, by taking advantage of the quadratic structure even to the extent of introducing it when it is not already at hand (as proposed in Section 5), one can avoid the escalation of dimensionality and at the same time get convergence results of a superior character (as presented in Section 4).

Note that with a nonvanishing quadratic term $\lambda \cdot \tilde{H}^\nu \lambda$ in (3.23) (the matrix $\tilde{H}^\nu$ being positive semidefinite, of course) the lower envelope approximation Ψ^ν to Ψ will generally *not* be polyhedral but have 'rounded corners'. As a matter of fact, if $\tilde{H}^\nu$ is nonsingular, then Ψ^ν is a smooth convex function with Lipschitz continuous derivatives.

In Step 2 of the algorithm, we need to solve a potentially large number of quadratic programming problems (3.11) in the vectors z_ω. This could be a trouble spot. If the problems are complicated and require full application of some quadratic program-

ming routine, the secret to success would have to lie in taking advantage of the similarities between neighboring problems. Techniques of parametric programming and 'bunching' might be useful. Not to be overlooked, however, are the situations in which each problem (3.11) decomposes further into something simpler.

Especially important is the case where

$$Z_\omega = Z_{\omega 1} \times Z_{\omega 2} \times \cdots \times Z_{\omega r} \tag{3.27}$$

and H_ω does not involve cross terms between the sets in this product:

$$H_\omega = \operatorname{diag}[H_{\omega 1}, H_{\omega 2}, \ldots, H_{\omega r}]. \tag{3.28}$$

Then (3.11) reduces to a separate problem over each of the sets $Z_{\omega 1}, \ldots, Z_{\omega r}$. If these sets are actually intervals (bounded or unbounded), then the separate problems are one-dimensional, and their solutions can be given *in closed form*. Such is indeed what happens when the costs $\psi_\omega(x)$ in (P) are penalties $\theta_\omega(h_\omega - T_\omega x)$ as in (1.5), (1.6), (1.7), and $\theta_\omega(h_\omega - T_\omega x)$ is a sum of separate terms, one for each real component of the vector $h_\omega - T_\omega x$. The special model we have treated in [11] makes use of this simplification. In such a setting the vector z^ν is readily computed as a simple function of $\bar{x}^\nu$, and indeed one can get away with storing only $\bar{x}^\nu$, which has only a small number of components compared to z^ν; cf. [4].

The product form (3.27) for Z_ω, if it is present, also raises further possibilities for structuring the subproblems introduced in Step 1, by the way. One could write

$$Z = Z_1 \times \cdots \times Z_r \quad \text{with } Z_j = \Pi_{\omega \in \Omega} Z_{\omega j} \tag{3.29}$$

and work with polytopes of the form

$$Z^\nu = Z_1^\nu \times \cdots \times Z_r^\nu \quad \text{with } Z_j^\nu \subset Z_j, \tag{3.30}$$

for instance. This could be advantageous in holding the dimensionality down. If each Z_j^ν is generated as the convex hull of a finite subset of Z_j consisting of n_ν elements, one can get away with describing the points of Z^ν by rn_ν parameters λ_{jk}. On the other hand, if Z^ν is regarded as the convex hull of the product of these finite subsets of $Z_1, \ldots, Z_r$, one would need $(n_\nu)^r$ parameters.

The procedure invoked in Step 4 of the algorithm has been left open to various possibilities, which could be influenced too by such considerations as the foregoing. Two basic possibilities that immediately come to mind are:

$$Z^{\nu+1} = \operatorname{co}\{\bar{z}^\nu, z^\nu\} \quad \text{(generalized Frank-Wolfe rule)} \tag{3.31}$$

and

$$Z^{\nu+1} = \operatorname{co}\{Z^\nu, z^\nu\} \quad \text{(generalized cutting-plane rule).} \tag{3.32}$$

The first of these is adequate for convergence if the matrix C is positive definite, as we shall see in Section 4. It is certainly the simplest but might suffer from too much information being thrown away between one iteration of Step 1 and the next. It gets its name from the interpretation in terms of problem (D) that will underly the proof of Theorem 5.

The second formula goes to the opposite extreme. It achieves better and better representations of the primal objective F, in the sense that

$$F(x) \geq F^{\nu+1}(x) \geq \max\{F^{\nu}(x), L(x, z^{\nu})\} \quad \text{for all } x, \quad \text{with}$$

$$F(\bar{x}^{\nu}) = F^{\nu+1}(\bar{x}^{\nu}) = L(\bar{x}^{\nu}, z^{\nu}), \tag{3.33}$$

but this is at the expense of keeping all information and continually enlarging the size of the quadratic programming subproblem. A good compromise possibility might be

$$Z^{\nu+1} = \operatorname{co}\{Z^1, \bar{z}^{\nu}, z^{\nu}\}, \tag{3.34}$$

where Z^1 is the fixed initial polytope.

This brings us to the choice of Z^1 in Step 0, which in determining the first approximate solutions $\bar{x}^1$ and $\bar{z}^1$ could have a big effect on the progress of the computations. We can, of course, start with $Z^1 = \{\hat{z}\}$, where $\hat{z}$ is an element of Z that may be regarded as an estimate for an optimal solution to (D). For example, if an initial guess $\hat{x}$ is available for an optimal solution to (P), one might take $\hat{z}$ to be a vector constructed by calculating an element $\hat{z}_{\omega} \in \zeta_{\omega}(\hat{x})$ for each ω. This approach makes sense especially in situations where $\zeta_{\omega}(\hat{x})$ is a singleton for each $\omega \in \Omega$, so that $\hat{z}$ is uniquely determined by the estimate $\hat{x}$.

Another approach to the initial Z^1 requires no guesses or prior information about solutions. A fixed number of elements $a_{k\omega}$ $(k = 1, \ldots, p)$ is chosen from each Z_{ω}, such as the set of extreme points of Z_{ω} augmented by some selected internal points. These yield p elements a_k of Z, where a_k has component $a_{k\omega}$ in Z_{ω}. The convex hull of these a_k's can be taken as Z^1. Such an approach to initialization has turned out to be very effective in the case of our special model in [11] when adapted to a product structure (3.27); see King [4].

In summary, there are many possibilities for choosing the initial polytope Z^1 in Step 0 and modifying it iteratively in Step 4. They can be tailored to the structure of the problem. Various product representations of Z and Z^{ν} could be helpful in particular. Versions of rules (3.31), (3.32), and (3.34), which maintain the product form, can be developed.

See the end of Section 4 for other comments on forming $Z^{\nu+1}$ from Z^{ν}.

4. Convergence results

Properties of the sequences produced by the finite generation algorithm in Section 3 will now be derived. For this purpose we ignore the optimality test in Step 3 of the algorithm, since our interest is centered on what happens when the procedure is iterated indefinitely. Unless otherwise indicated, our assumptions are merely the basic ones in Section 1. The initial polytope Z^1 is arbitrary, and $Z^{\nu+1}$ is not subjected to any requirement stricter than the one in Step 4, namely that $Z^{\nu+1} \supset \{\bar{z}^{\nu}, z^{\nu}\}$. In addition to the symbols already introduced in the statement of the algorithm in

Section 3 we use the supplementary notation

$$\bar{\alpha} = \min(\mathrm{P}) = \max(\mathrm{D}), \tag{4.1}$$

$$\bar{\varepsilon}_\nu = \bar{\alpha} - \bar{\alpha}_\nu, \tag{4.2}$$

$$\bar{w}^\nu = \nabla_x L(\bar{x}^\nu, \bar{z}^\nu) = c + C\bar{x}^\nu - ET^*_\omega \bar{z}^\nu_\omega, \tag{4.3}$$

$$\|x\|_C = [x \cdot Cx]^{1/2}. \tag{4.4}$$

Of course $\|x\|_C$ is a norm on $\mathbb{R}^n$ if C is positive definite. If C is only positive semidefinite, then $\|x\|_C$ vanishes on the subspace $\{x \in \mathbb{R}^n \mid Cx = 0\}$ but is positive elsewhere.

Theorem 3. *The sequences* $\{\bar{x}^\nu\}$, $\{\bar{z}^\nu\}$, *and* $\{z^\nu\}$ *are bounded and satisfy*

$$F(\bar{x}^\nu) = \alpha_\nu \geq \bar{\alpha} \geq \cdots \geq \bar{\alpha}_{\nu+1} \geq \bar{\alpha}_\nu = G(\bar{z}^\nu). \tag{4.5}$$

Furthermore one has the estimate.

$$\tfrac{1}{2}\|\bar{x} - \bar{x}^\nu\|_C^2 \leq \bar{\varepsilon}_\nu - \bar{w}^\nu \cdot (\bar{x} - \bar{x}^\nu) \leq \bar{\varepsilon}_\nu \leq \bar{\varepsilon}_\nu \tag{4.6}$$

for every optimal solution $\bar{x}$ *to* (P), *where*

$$\bar{w}^\nu \cdot (x - \bar{x}^\nu) \geq 0 \quad \text{for every } x \in X. \tag{4.7}$$

If $\varepsilon_\nu \to 0$, *then every cluster point of* $\{\bar{x}^\nu\}$ *is an optmal solution to* (P), *and every cluster point of* $\{\bar{z}^\nu\}$ *is an optimal solution to* (D).

Proof. We have $\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu)$ and $\alpha_\nu = L(\bar{x}^\nu, z^\nu)$ by definition, so $F(\bar{x}^\nu) = \alpha_\nu$ by (3.7). Then $\alpha_\nu \geq \bar{\alpha} \geq \bar{\alpha}_\nu$ by (3.9). By the same token, $G(\bar{z}^{\nu+1}) = \bar{\alpha}_{\nu+1}$ and $\bar{\alpha} \geq \bar{\alpha}_{\nu+1}$. But also

$$G(\bar{z}^{\nu+1}) = \max_{z \in Z^{\nu+1}} G(z) \geq G(\bar{z}^\nu)$$

because $\bar{z}^\nu \in Z^{\nu+1}$. All the relations in (4.5) are therefore correct.

Next we verify that the sequence $\{\bar{z}^\nu\}$ is bounded. Recall that G is a continuous concave function on Z, since G is given by (2.7), where φ is the concave function defined by (1.4); our basic assumption about the sets $\xi(v)$ being bounded implies φ is finite everywhere. (As is well known, a concave function is continuous at a point if it is finite on a neighborhood of the point [9, Theorem 10.1].) We know from (4.5) that the sequence $\{G(\bar{z}^\nu)\}$ us nondecreasing, so the boundedness of $\{\bar{z}^\nu\}$ can be established by showing that the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is bounded. Consider the closed concave function

$$g(z) = \begin{cases} G(z) & \text{if } z \in Z, \\ \infty & \text{if } z \notin Z. \end{cases}$$

We wish to show that a certain level set $\{z \mid g(z) \geq \bar{\alpha}_1\}$ is bounded. But the level sets $\{z \mid g(z) \geq \alpha\}$, $\alpha \in R$, are all bounded if merely one of them is bounded and nonempty

(see [9, Corollary 8.7.1]). In the present case we know that the level set

$$\{z \mid g(z) \geq \bar{\alpha}\} = [\text{set of all optimal solutions to (D)}]$$

is bounded and nonempty (Theorem 2). Therefore the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is indeed bounded, and the sequence $\{\bar{z}^\nu\}$ is bounded as claimed.

We invoke now the fact that

$$\bar{x}^\nu \in \xi(c - ET_\omega^* \bar{z}_\omega^\nu) \quad \text{for all } \nu, \tag{4.8}$$

which is true by (2.5) because $(\bar{x}^\nu, \bar{z}^\nu)$ is a saddle point of L relative to $X \times Z^\nu$. In terms of the finite concave function φ we have

$$\xi(v) = \partial\varphi(v) \quad \text{for all } v \in \mathbb{R}^n. \tag{4.9}$$

Indeed, (1.4) defines φ as the conjugate of the closed proper concave function

$$\gamma(x) = \begin{cases} -\frac{1}{2} x \cdot Cx & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

so $\partial\varphi(v)$ consists of the points x which minimize $v \cdot x - \gamma(x)$ over $\mathbb{R}^n$ (see [9, Theorem 23.5]). These are the points that make up the set $\xi(v)$ in (1.2). Thus

$$\bar{x}^\nu \in \partial\varphi(\bar{v}^\nu) \quad \text{for all } \nu, \quad \text{where } \bar{v}^\nu = c - ET_\omega^* \bar{z}_\omega^\nu. \tag{4.10}$$

The sequence $\{\bar{v}^\nu\}$ is bounded, since $\{\bar{z}^\nu\}$ is. Moreover the multifunction $\partial\varphi$ is locally bounded: for every $\bar{v} \in \mathbb{R}^n$ there is a $\delta > 0$ such that the set $\bigcup \{\partial\varphi(v) \mid |v - \bar{v}| \leq \delta\}$ is bounded (see [9, Corollary 24.5.1]). It follows by a simple compactness argument that $\partial\varphi$ carries bounded sets into bounded sets: if $V \subset \mathbb{R}^n$ is bounded, then $\bigcup \{\partial\varphi(v) \mid v \in V\}$ is bounded. Taking $V = \{\bar{v}^\nu\}$, we conclude that the sequence $\{\bar{x}^\nu\}$ is bounded.

The argument establishing that $\{z^\nu\}$ is bounded is similar. We have $z_\omega^\nu \in \zeta_\omega(\bar{x}^\nu)$, where ζ_ω is the multifunction defined in (1.3). Since the sequence $\{\bar{x}^\nu\}$ is now known to be bounded, we need only show that ζ_ω is locally bounded at every $\bar{x}^\nu$ in order to conclude that each of the sequences $\{z_\omega^\nu\}$ is bounded and consequently that $\{z^\nu\}$ is bounded.

In terms of the convex function θ_ω defined in (1.5) we have

$$\zeta_\omega(x) = \partial\theta_\omega(h_\omega - T_\omega x) \quad \text{for all } x \in X. \tag{4.11}$$

This holds because (1.5) expresses θ_ω as the conjugate of the closed proper convex function

$$f_\omega(z_\omega) = \begin{cases} \frac{1}{2} z_\omega \cdot H_\omega z_\omega & \text{if } z_\omega \in Z_\omega, \\ \infty & \text{if } z_\omega \notin Z_\omega. \end{cases}$$

The vectors $z_w \in \partial\theta_\omega(u)$ are therefore the ones that maximize $u \cdot z_\omega - f_\omega(z_\omega)$ (see [9, Theorem 23.5]). Our assumption that $\zeta_\omega(x)$ is nonempty and bounded for every $x \in X$ means that $\partial\theta_\omega(u)$ is nonempty and bounded for every u of the form $h_\omega - T_\omega x$ for some $x \in X$. Every such $u = h_\omega - T_\omega x$ therefore belongs to $\operatorname{int}(\operatorname{dom} \theta_\omega)$ (cf.

[9, Theorem 23.4]). It follows then that $\partial\theta_\omega$ is locally bounded at u (cf. [9, Corollary 24.5.1]). The mapping $x \to h_\omega - T_\omega x$ is continuous, so this implies ζ_ω is locally bounded at x for every $x \in X$, as we needed to prove.

The argument just given shows also that the convex function θ_ω is continuous at $h_\omega - T_\omega x$ for every $x \in X$ (since θ_ω is continuous on $\operatorname{int}(\operatorname{dom}\theta_\omega)$ [9, Theorem 10.1]). Therefore F is continuous on X by (1.6) and (2.6). We observed earlier in the proof that G is also continuous on Z. Of course X and Z, being convex polyhedra, are closed sets. Hence if $\varepsilon_\nu \to 0$, so that $F(\bar{x}^\nu) \to \bar{\alpha}$ and $G(\bar{z}^\nu) \to \bar{\alpha}$, any cluster points $\bar{x}^\infty$ of $\{\bar{x}^\nu\}$ and $\bar{z}^\infty$ of $\{\bar{z}^\nu\}$ must satisfy $F(\bar{x}^\infty) = \bar{\alpha} = G(\bar{z}^\infty)$ and be optimal solutions to (P) and (D).

We turn finally to the estimate (4.6). The saddle point condition on $(\bar{x}^\nu, \bar{z}^\nu)$ entails

$$\bar{x}^\nu \in \operatorname*{argmin}_{x \in X} L(x, \bar{z}^\nu).$$

Since X is a closed convex set and $L(x, \bar{z}^\nu)$ is a differentiable convex function of x, this condition implies that the vector $-\bar{w}^\nu = -\nabla_x L(\bar{x}^\nu, \bar{z}^\nu)$ belongs to the normal cone to X at $\bar{x}^\nu$ (cf. [9, Theorem 27.4]), which is exactly the assertion of (4.7). We have

$$\begin{aligned} L(x, \bar{z}^\nu) &= L(\bar{x}^\nu, \bar{z}^\nu) + \nabla_x L(\bar{x}^\nu, \bar{z}^\nu) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu) \\ &= \bar{\alpha}_\nu + \bar{w}^\nu \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}\|x - \bar{x}^\nu\|_C^2 \text{ for all } x \end{aligned} \tag{4.12}$$

from the quadratic nature of L, and also

$$L(x, \bar{z}^\nu) \leq F(x) \quad \text{for all } x \in X$$

by (2.2). For any optimal solution $\bar{x}$ to (P), then, we have

$$\bar{\alpha}_\nu + \bar{w}^\nu \cdot (\bar{x} - \bar{x}^\nu) + \tfrac{1}{2}\|\bar{x} - \bar{x}^\nu\|_C^2 \leq F(\bar{x}) = \bar{\alpha}.$$

In terms of $\tilde{\varepsilon}_\nu = \bar{\alpha} - \bar{\alpha}_\nu$, this can be written as the first inequality in (4.6). The rest of (4.6) then follows from (4.7), inasmuch as $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu = \tilde{\varepsilon}_\nu + \alpha_\nu - \bar{\alpha} \geq \tilde{\varepsilon}_\nu$. □

Theorem 3 focuses our attention on finding conditions that guarantee $\varepsilon_\nu \to 0$. Our first result in this direction makes no additional assumptions on the data in the problem and therefore serves as a baseline. It relies on an increasing sequence of polytopes in Step 4, however. The generalized cutting-plane rule in (3.32) is covered as a special case.

Theorem 4. *If $Z^{\nu+1} \supset Z^\nu \cup \{z^\nu\}$ in Step 4 of the algorithm, then $\varepsilon_\nu \to 0$.*

Proof. Let $\bar{\alpha}_\infty = \lim_\nu \bar{\alpha}_\nu$ and $\alpha_\infty = \limsup_\nu \alpha_\nu$. (The first limit exists because $\{\bar{\alpha}_\nu\}$ is nondecreasing in (4.5).) Since $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu \geq 0$ for all ν, we need only demonstrate that $\alpha_\infty \leq \bar{\alpha}_\infty$. The sequences $\{\bar{x}^\nu\}$, $\{\bar{z}^\nu\}$, and $\{z^\nu\}$, are bounded by Theorem 3, so we can extract convergent subsequences with a common index set $N \subset \{1, 2, \ldots\}$ such that

$$\bar{x}^\nu \xrightarrow[\nu \in N]{} \bar{x}^\infty, \qquad \bar{z}^\nu \xrightarrow[\nu \in N]{} \bar{z}^\infty, \qquad z^\nu \xrightarrow[\nu \in N]{} z^\infty, \qquad \alpha_\nu \xrightarrow[\nu \in N]{} \alpha_\infty.$$

Then since

$$\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu) \xrightarrow[\nu \in N]{} L(\bar{x}^\infty, \bar{z}^\infty), \qquad \alpha_\nu = L(\bar{x}^\nu, z^\nu) \xrightarrow[\nu \in N]{} L(\bar{x}^\infty, z^\infty),$$

we have $\bar{\alpha}_\infty = L(\bar{x}^\infty, \bar{z}^\infty)$ and $\alpha_\infty = L(\bar{x}^\infty, z^\infty)$. Our task now is to prove that $L(\bar{x}^\infty, z^\infty) \leq L(\bar{x}^\infty, \bar{z}^\infty)$.

From the saddle point condition on $(\bar{x}^\nu, \bar{z}^\nu)$ we have

$$L(\bar{x}^\nu, z) \leq L(\bar{x}^\nu, \bar{z}^\nu) \quad \text{for all } z \in Z^\nu.$$

Let $Z^\infty = \bigcup_{\nu=1}^\infty Z^\nu$. Since $Z^\nu \subset Z^{\nu+1} \subset \cdots$ we know that for any fixed $z \in Z^\infty$ the inequality $L(\bar{x}^\nu, z) \leq L(\bar{x}^\nu, \bar{z}^\nu)$ holds for all ν sufficiently high. Taking the limit as $\nu \to \infty$, $\nu \in N$, we obtain $L(\bar{x}^\infty, z) \leq L(\bar{x}^\infty, \bar{z}^\infty)$. This holds for arbitrary $z \in Z^\infty$, so

$$L(\bar{x}^\infty, z) \leq L(\bar{x}^\infty, \bar{z}^\infty) \quad \text{for all } z \in \operatorname{cl} Z^\infty.$$

But z^∞ is one of the elements of $\operatorname{cl} Z^\infty$, since $z_\nu \in Z^{\nu+1}$ for all ν. Therefore $L(\bar{x}^\infty, z^\infty) \leq L(\bar{x}^\infty, \bar{z}^\infty)$ in particular, and the proof is complete. □

Our main result comes next. It assures us that when C is positive definite, we do *not* have to keep increasing the size of the polytope Z^ν in order to have convergence. The number of elements used to generate Z^ν can be kept at whatever level seems adequate in maintaining a robust representation of F and G.

Theorem 5. *Suppose the matrix C in* (P) *is positive definite. Then under the minimal requirement $Z^{\nu+1} \supset \{\bar{z}^\nu, z^\nu\}$ in Step 4 of the algorithm, one has $\varepsilon_\nu \to 0$ and also $\bar{x}^\nu \to \bar{x}$, where $\bar{x}$ is the unique optimal solution to* (P).

If in addition there exists $\rho \geq 0$ such that

$$z_\omega \cdot T_\omega C^{-1} T_\omega^* z_\omega \leq \rho z_\omega \cdot H_\omega z_\omega \quad \text{for all } z_\omega \in \mathbb{R}^m, \ \omega \in \Omega, \tag{4.13}$$

(*as is true in particular if every H_ω is positive definite*), *then in the estimate* (4.6) *one has*

$$\bar{\varepsilon}_{\nu+1} \leq \tau \bar{\varepsilon}_\nu \quad \text{for } \nu = 1, 2, \ldots \tag{4.14}$$

where the factor $\tau \in [0, 1)$ is given by

$$\tau = \begin{cases} \rho & \text{if } 0 \leq \rho \leq \frac{1}{2}, \\ 1 - \frac{1}{4}\rho^{-1} & \text{if } \rho \geq \frac{1}{2}. \end{cases} \tag{4.15}$$

Thus

$$\bar{\varepsilon}_{\nu+\mu} \leq \tau^\mu \bar{\varepsilon}_\nu \leq \tau^\mu \varepsilon_\nu \quad \text{for } \nu = 1, 2, \ldots, \text{ and } \mu = 1, 2, \ldots. \tag{4.16}$$

Note that Theorem 5 asserts in (4.14) a linear rate of convergence of $\bar{\alpha}_\nu$ to $\bar{\alpha}$ with modulus τ, and the estimate (4.6) effectively translates this into a linear rate of convergence of $\bar{x}^\nu$ to $\bar{x}$ with modulus $\tau^{1/2}$. Indeed, from (4.6) and (4.16) we have

$$\|\bar{x} - \bar{x}^{\nu+\mu}\|_C \leq [2\tau^\mu \varepsilon_\nu]^{1/2} \quad \text{for } \nu = 1, 2, \ldots \text{ and } \mu = 1, 2, \ldots.$$

This is an unusual sort of result, because it applies not just to the tail of the sequence $\{\bar{x}^\nu\}$ but right from the beginning. Moreover the value of ε_ν is known in each iteration, and the value of $\tau \in [0, 1)$ can be estimated in advance.

Theorem 5 makes no assertion about the convergence of $\{\bar{z}^\nu\}$ beyond the one in Theorem 3. Of course if there is a *unique* optimal solution $\bar{z}$ to (D), then by Theorem 3 we have $\bar{z}^\nu \to \bar{z}$ whenever $\varepsilon_\nu \to 0$, as is the case here. In particular (D) has a unique optimal solution if the matrices H_ω are all positive definite.

The proof of Theorem 5 depends on further analysis of the dual objective function G. Essentially what we must provide is a lower estimate of G that ensures that the direction $z^\nu - \bar{z}^\nu$ determined in Step 2 of the algorithm is always a direction of ascent for G.

Proposition 3. *Let*

$$f^\nu(w) = \max_{x \in X} \{(w - \bar{w}^\nu) \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \quad \text{for } w \in \mathbb{R}^n. \tag{4.17}$$

Then f^ν is a finite convex function on $\mathbb{R}^n$ with $0 = f^\nu(0) \leq f^\nu(w)$ for all w, and

$$\begin{aligned} 0 &\leq L(\bar{x}^\nu, z) - G(z) - f^\nu(ET^*_\omega(z_\omega - \bar{z}^\nu_\omega)) \\ &\leq Ef^\nu(T^*_\omega(z_\omega - \bar{z}^\nu_\omega)) \quad \text{for all } z \in Z. \end{aligned} \tag{4.18}$$

If C is positive definite, then

$$f^\nu(w) \leq \tfrac{1}{2}[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot C^{-1}[(w - \bar{w}^\nu) + s\bar{w}^\nu] \quad \text{for all } s \geq 0, \tag{4.19}$$

so that in particular (for $s = 1$)

$$G(z) \geq L(\bar{x}^\nu, z) - \tfrac{1}{2}E\{(z_\omega - \bar{z}^\nu_\omega) \cdot T_\omega C^{-1} T^*_\omega(z_\omega - \bar{z}^\nu_\omega)\} \quad \text{for all } z \in Z. \tag{4.20}$$

Proof. First re-express f^ν in terms of the finite concave function φ in (1.4), so as to verify that f^ν is a finite convex function and that 'max' rather than 'sup' is appropriate in (4.17):

$$\begin{aligned} -f^\nu(w) &= \min_{x \in X} \{(\bar{w}^\nu - w) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \\ &= (w - \bar{w}^\nu) \cdot \bar{x}^\nu + \tfrac{1}{2}\bar{x}^\nu \cdot C\bar{x}^\nu + \max_{x \in X} \{[\bar{w}^\nu - C\bar{x}^\nu - w] \cdot x + \tfrac{1}{2}x \cdot Cx\} \\ &= (w - \bar{w}^\nu) \cdot \bar{x}^\nu + \tfrac{1}{2}\bar{x}^\nu \cdot C\bar{x}^\nu + \varphi(\bar{w}^\nu - C\bar{x}^\nu - w). \end{aligned}$$

Clearly $f^\nu(w) \geq 0$ for all w, because $x = \bar{x}^\nu$ is one of the points considered in taking the maximum in (4.17). Furthermore

$$-f^\nu(0) = \min_{x \in X} \{\bar{w}^\nu \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}.$$

Recalling the expansion (4.12) of $L(x, \bar{z}^\nu)$ around $\bar{x}^\nu$ and the fact that $\bar{x}^\nu$ minimizes $L(x, \bar{z}^\nu)$ over X (since $(\bar{x}^\nu, \bar{z}^\nu)$ is a saddle point of L on $X \times Z^\nu$), we see that $f^\nu(0) = 0$.

To get the equation in (4.18), from which the two inequalities in (4.18) immediately follows (the first because $f^\nu(w) \geq 0$ and the second by Jensen's inequality, because f^ν is convex), we look at the expansion

$$L(x, z) = L(\bar{x}^\nu, z) + \nabla_x L(\bar{x}^\nu, z) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu),$$

where

$$\nabla_x L(\bar{x}^\nu, z) = c + C\bar{x}^\nu - ET^*_\omega z_\omega = \bar{w}^\nu - ET^*_\omega(z_\omega - \bar{z}^\nu_\omega).$$

From this we calculate

$$\begin{aligned} L(\bar{x}^\nu, z) - G(z) &= L(\bar{x}^\nu, z) - \min_{x \in X} L(x, z) = \max_{x \in X} \{L(\bar{x}^\nu, z) - L(x, z)\} \\ &= \max_{x \in X} \{[ET^*_\omega(z_\omega - \bar{z}^\nu_\omega) - \bar{\omega}^\nu] \cdot (x - \bar{x}^\nu) \\ &\quad - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \\ &= f^\nu(ET^*_\omega(z_\omega - \bar{z}^\nu_\omega)). \end{aligned}$$

This establishes (4.18).

Finally we use property (4.7) in Theorem 3 to estimate for arbitrary $s \geq 0$:

$$\begin{aligned} f^\nu(w) &\leq \sup_{x \in X} \{[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \\ &\leq \sup_{x \in \mathbf{R}^n} \{[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}. \end{aligned}$$

When C is positive definite, this last supremum equals the quadratic expression on the right side of (4.19).

Proof of Theorem 5. Since $(\bar{x}^{\nu+1}, \bar{z}^{\nu+1})$ is a saddle point of L relative to $X \times Z^\nu$, we have

$$\bar{\alpha}_{\nu+1} = G(\bar{z}^{\nu+1}) = \max_{z \in Z^{\nu+1}} G(z).$$

But $Z^{\nu+1}$ includes the line segment joining $\bar{z}^\nu$ and z^ν. Therefore

$$\bar{\alpha}_{\nu+1} \geq \max_{0 \leq t \leq 1} G(\bar{z}^\nu + t(z^\nu - \bar{z}^\nu)). \tag{4.21}$$

To see what this implies, we substitute $z = \bar{z}^\nu + t(z^\nu - \bar{z}^\nu)$ into the estimate (4.20) of Proposition 3 and make use of the fact that, for $0 \leq t \leq 1$,

$$\begin{aligned} L(\bar{x}^\nu, \bar{z}^\nu + t(z^\mu - \bar{z}^\nu)) &= L(\bar{x}^\nu, (1-t)\bar{z}^\nu + tz^\nu) \\ &\geq (1-t)L(\bar{x}^\nu, \bar{z}^\nu) + tL(\bar{x}^\nu, z^\nu) = (1-t)\bar{\alpha}_\nu + t\alpha_\nu - \bar{\alpha}_\nu + t\varepsilon_\nu. \end{aligned} \tag{4.22}$$

This yields

$$G(\bar{z}^\nu + t(z^\nu)) \geq \bar{\alpha}_\nu + t\varepsilon_\nu - \tfrac{1}{2}t^2\delta_\nu \quad \text{for } 0 \leq t \leq 1 \tag{4.23}$$

where

$$\delta_\nu := E\{(z_\omega^\nu - \bar{z}_\omega^\nu) \cdot T_\omega C^{-1} T_\omega^*(z_\omega^\nu - \bar{z}_\omega^\nu)\}. \tag{4.24}$$

Combining (4.23) with (4.21), we get

$$\bar{\alpha}_{\nu+1} \geq \bar{\alpha}_\nu + \sigma(\varepsilon_\nu, \delta_\nu), \tag{4.25}$$

where

$$\sigma(\varepsilon, \delta) := \max_{0 \leq t \leq 1} \{t\varepsilon - \tfrac{1}{2}t^2\delta\} = \begin{cases} \varepsilon - \frac{1}{2}\delta & \text{if } 0 \leq \delta \leq \varepsilon, \\ \frac{1}{2}\varepsilon^2\delta^{-1} & \text{if } \delta > \varepsilon. \end{cases} \tag{4.26}$$

Note that σ is a continuous function of $(\varepsilon, \delta) \in \mathbb{R}_+^2$ with $\sigma(\varepsilon, \delta) = 0$ if $\varepsilon = 0$, but $\sigma(\varepsilon, \delta) > 0$ if $\varepsilon > 0$. The sequence $\{\bar{\alpha}_\nu\}$ is nondecreasing and bounded above by $\bar{\alpha}$ (cf. (4.5)), so $\sigma(\varepsilon_\nu, \delta_\nu) \to 0$. The sequence $\{\delta_\nu\}$ is bounded, because the sequences $\{\bar{z}^\nu\}$ and $\{z^\nu\}$ are bounded (Theorem 3). From the cited properties of σ, it follows then that $\varepsilon_\nu \to 0$. This implies $x^\nu \to \bar{x}$ by property (4.6) in Theorem 1.

We can also write (4.25) as

$$\bar{\varepsilon}_{\nu+1} \leq \bar{\varepsilon}_\nu - \sigma(\varepsilon_\nu, \delta_\nu). \tag{4.27}$$

Under the additional assumption in Theorem 5 that (4.13) holds, we have

$$\delta_\nu \leq \rho\beta_\nu, \quad \text{where } \beta_\nu := E\{(z_\omega^\nu - \bar{z}_\omega^\nu) \cdot H_\omega(z_\omega^\nu - \bar{z}_\omega^\nu)\}. \tag{4.28}$$

Consider now the quadratic function

$$q(t) = L(\bar{x}^\nu, \bar{z}^\nu + t(z^\nu - \bar{z}^\nu)) \quad \text{for } 0 \leq t \leq 1.$$

This has $q(0) = L(\bar{x}^\nu, \bar{z}^\nu) = \bar{\alpha}_\nu$, $q(1) = L(\bar{x}^\nu, z^\nu) = \alpha_\nu$, $q''(t) \equiv -\beta_\nu$, so q must be of the form

$$q(t) = (1-t)\bar{\alpha}_\nu + t\alpha_\nu + \tfrac{1}{2}t(1-t)\beta_\nu.$$

Moreover the maximum of $q(t)$ over $0 \leq t \leq 1$ is attained at $t = 1$, since the maximum of $L(\bar{x}^\nu, z)$ over $z \in Z$ is attained at $z = z^\nu$. Therefore

$$(1-t)\bar{\alpha}_\nu + t\alpha_\nu + \tfrac{1}{2}t(1-t)\beta_\nu \leq \alpha_\nu \quad \text{for } 0 \leq t \leq 1,$$

or in other words,

$$t(1-t)\beta_\nu \leq 2(1-t)(\alpha_\nu - \bar{\alpha}_\nu) = 2(1-t)\varepsilon_\nu \quad \text{for } 0 \leq t \leq 1.$$

This implies $\beta_\nu \leq 2\varepsilon_\nu$, and then (4.28) yields

$$\delta_\nu \leq 2\rho\varepsilon_\nu. \tag{4.29}$$

Formula (4.26) now gives us

$$\sigma(\varepsilon_\nu, \delta_\nu) \geq \sigma(\varepsilon_\nu, 2\rho\varepsilon_\nu) = \varepsilon_\nu\sigma(1, 2\rho) \geq \bar{\varepsilon}_\nu\sigma(1, 2\rho).$$

Substituting in (4.27) we get

$$\bar{\varepsilon}_{\nu+1} \leq \bar{\varepsilon}_\nu - \bar{\varepsilon}_\nu\sigma(1, 2\rho) = [1 - \sigma(1, 2\rho)]\bar{\varepsilon}_\nu, \tag{4.30}$$

where

$$\sigma(1, 2\rho) = \begin{cases} 1-\rho & \text{if } 0 \leq \rho \leq \frac{1}{2}, \\ \frac{1}{4}\rho^{-1} & \text{if } \rho \leq \frac{1}{2}. \end{cases}$$

The factor $1-\sigma(1, 2\rho)$ is the number τ defined in (4.15), and (4.30) is thus the desired condition (4.14). □

Remark. Proposition 3 provides additional information that could be used in the direction search and polytope modification steps in the algorithm. Inequality (4.18) asserts that

$$L(\bar{x}^\nu, z) \geq G(z) \geq L(\bar{x}^\nu, z) - Ef^\nu(T^*_\omega(z_\omega - \bar{z}^\nu_\omega))$$

$$\text{for all } z \in Z, \text{ with equality when } z = \bar{z}^\nu. \tag{4.31}$$

The vector z^ν maximizes $L(\bar{x}^\nu, z)$ over all $z \in Z$ and thus provides not only the needed value $L(\bar{x}^\nu, \bar{z}^\nu) = F(\bar{x}^\nu)$ but also a clue as to where we might look to move next in trying to improve on the current value $G(\bar{z}^\nu)$ of G. A further clue can be found by maximizing the right side of (4.31) over Z to get a vector $\hat{z}^\nu$. This is possible because the right side decomposes into separate terms for each ω. Indeed, the components $\hat{z}^\nu_\omega$ of $\hat{z}^\nu$ can be determined by

$$\hat{z}^\nu_\omega \in \operatorname*{argmax}_{z_\omega \in Z_\omega} \{f^\nu(T^*_\omega(z_\omega - \bar{z}^\nu_\omega)) + z_\omega \cdot [h_\omega - T_\omega \bar{x}^\nu] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{4.32}$$

In view of the form of f^ν in (4.17), this amounts to solving a special quadratic programming problem for each $\omega \in \Omega$.

If $\hat{z}^\nu$ is calculated in this way along with z^ν in Step 2, it can also be incorporated in the new polytope $Z^{\nu+1}$ in Step 4 in order to enrich the representation of G.

5. Adding strongly quadratic terms

The theoretical convergence properties of the finite generation algorithm are markedly superior when the quadratic forms that are involved are positive definite. But many problems lack this positive definiteness. Stochastic linear programming problems, for instance, have no quadratic terms at all. Such problems can be handled by a procedure which combines the finite generation algorithm with an augmented Lagrangian technique that introduces the desired property.

The technique in question was developed by Rockafellar [7] in a general context of minimax problems and variational inequalities. As applied to the present situation, it concerns the replacement of the saddle point problem for L on $X \times Z$ by a sequence of saddle point problems for augmented Lagrangians of the form

$$L_\mu(x, z) = L(x, z) + \frac{\eta}{2}(x - \bar{x}^\mu_*) \cdot \bar{C}(x - \bar{x}^\mu_*) - \frac{\eta}{2} E\{(z_\omega - \bar{z}^\mu_{*\omega}) \cdot \bar{H}_\omega(z_\omega - \bar{z}^\mu_{*\omega})\}$$

$$\text{on } X \times Z \text{ for } \mu = 1, 2, \ldots. \tag{5.1}$$

Here $\bar{C}$ and $\bar{H}_\omega$ are fixed positive definite matrices, η is a penalty parameter value that helps to control the rate of convergence, and $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ is a current 'estimate' for a saddle point of L itself on $X \times Z$, i.e. for an optimal solution pair for problems (P) and (D).

When the augmenting terms in L_μ are expanded and combined with those in L, the expression (5.1) turns into

$$L_\nu(x, z) = c_*^\mu \cdot x + \tfrac{1}{2} x \cdot C_* x + E\{z_\omega \cdot [h_{*\omega} - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_{*\omega} z_\omega\} + \text{const.} \tag{5.2}$$

where

$$C_* = C + \eta \bar{C}, \qquad H_{*\omega} = H_\omega + \eta \bar{H}_\omega, \tag{5.3}$$

$$c_*^\mu = c - \eta \bar{C} \bar{x}_*^\mu, \qquad h_{*\omega}^\mu = h_\omega - \eta \bar{H}_\omega \bar{z}_{*\omega}^\mu. \tag{5.4}$$

Note that the vectors c_*^μ and $h_{*\omega}^\mu$ giving the linear terms in L_μ depend on the μth solution estimates, but the matrices C_* and $H_{*\omega}$ giving the quadratic terms remain fixed as long as the value of η is not varied. Since $\eta > 0$, these matrices are positive definite. Therefore the saddle point problem for L_μ on $X \times Z$ can be solved by the finite generation algorithm with an essentially linear rate of convergence (cf. Theorem 5).

We make use of this as follows.

Master Algorithm

Step 0 (Initialization). Fix the matrices $\bar{C}$, $\bar{H}_\omega$, and the parameter value $\eta > 0$. Choose initial points $\bar{x}_*^1 \in X$ and $\bar{z}_*^1 \in Z$. Set $\mu = 1$.

Step 1 (Finite Generation Method). Use the finite generation algorithm to determine an approximate saddle point $(\bar{x}_*, \bar{z}_*)$ of L_μ on $X \times Z$ (according to a stopping criterion given below).

Step 2 (Update). Set $(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) = (\bar{x}_*, \bar{z}_*)$. Replace μ by $\mu + 1$ and return to Step 1 (with the same value of η).

The finite generation method in Step 1 generates for the function L_μ a sequence of pairs $(\bar{x}^\nu, \bar{z}^\nu)$ and test values ε_ν. To get an approximate saddle point we take

$$(\bar{x}_*, \bar{z}_*) = (\bar{x}^\nu, \bar{z}^\nu) \quad \text{when } \varepsilon_\nu \leq \bar{\varepsilon}_*^\mu(\bar{x}^\nu, \bar{z}^\nu), \tag{5.5}$$

where the function $\bar{\varepsilon}_*^\mu$ in the stopping criterion is defined as follows. In terms of the norms

$$\|x\|_* = [x \cdot \bar{C} x]^{1/2} \text{ for } x \in \mathbb{R}^n, \qquad \|z\|_* = [E\{z_\omega \cdot \bar{H}_\omega z_\omega\}]^{1/2} \text{ for } z \in (\mathbb{R}^m)^\Omega,$$

$$\|(x, z)\|_* = [\|x\|_*^2 + \|z\|_*^2]^{1/2} \tag{5.6}$$

we set

$$\bar{\varepsilon}_*^\mu(x, z) = \theta_\mu^2 \min\{1, (\eta/2)\|(x, z) - (\bar{x}^\mu, \bar{z}^\mu)\|_*^2\} \quad \text{with } \theta_\mu > 0, \sum_{\mu=1}^{\infty} \theta_\mu < \infty. \tag{5.7}$$

Obviously $\bar{\varepsilon}_*^\mu(x, z) > 0$ unless $(x, z) = (\bar{x}^\mu, \bar{z}^\mu)$. The sequence $\{(\bar{x}^\nu, \bar{z}^\nu)\}$ converges to the unique saddle point of L_μ on $X \times Z$, so except in the lucky, degenerate case where $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ is already that saddle point, the values $\varepsilon_*^\mu(\bar{x}^\nu, \bar{z}^\nu)$ will be bounded away from zero, and the stopping criterion in (5.5) will eventually be satisfied. (In the degenerate case, $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ must in fact be a saddle point of L itself and there is no need to leave Step 1: the sequence $\{(\bar{x}^\nu, \bar{z}^\nu)\}$ converges to this saddle point at a linear rate.)

Theorem 6. *The sequences $\{\bar{x}_*^\mu\}$ and $\{\bar{z}_*^\mu\}$ generated by the master algorithm converge to particular optimal solutions $\bar{x}$ and $\bar{z}$ to problems* (P) *and* (D), *respectively. If $\bar{x}$ and $\bar{z}$ are the unique optimal solution to* (P) *and* (D), *then there is a number $\beta(\eta) \in [0, 1)$ such that $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ converges to $(\bar{x}, \bar{z})$ at a linear rate with modulus $\beta(\eta)$. Moreover $\beta(\eta) \to 0$ as $\eta \to 0$.*

Proof. We shall deduce this from [7, Theorems 1 and 2], which are general results applicable to the calculation of a saddle point of a convex-concave function on a product of Hilbert spaces. The Hilbert spaces in this case are $\mathbb{R}^n$ and $(\mathbb{R}^m)^\Omega$ under the norms in (5.5). The convex-concave function in question is

$$\bar{L}(x, z) = \begin{cases} L(x, z) & \text{if } x \in X \text{ and } z \in Z, \\ -\infty & \text{if } x \in X \text{ but } z \notin Z, \\ \infty & \text{if } x \notin X. \end{cases}$$

The saddle points of $\bar{L}$ on $\mathbb{R}^n \times (\mathbb{R}^m)^\Omega$ are the same as those of L on $X \times Z$. The problem of finding a saddle point of $\bar{L}_\mu$ on $X \times Z$ reduces to the one for

$$\bar{L}_\mu(x, z) = \bar{L}(x, z) + \frac{\eta}{2}\|x - \bar{x}_*^\mu\|_*^2 - \frac{\eta}{2}\|z - \bar{z}_*^\mu\|_*^2$$

on $\mathbb{R}^n \times (\mathbb{R}^m)^\Omega$.

Denote by $P(\bar{x}^\mu, \bar{z}^\mu)$ the unique saddle point of $\bar{L}_\mu$ on $R^n \times (R^m)^\Omega$, which is also the unique saddle point of L_μ on $X \times Z$. The mapping P is the 'proximal mapping' associated with the maximal monotone multifunction T that corresponds to $\eta^{-1}\bar{L}$ in the sense of [7, Section 1 and Section 5]. In consequence of [7, Theorem 1], the sequence $\{(\bar{x}_*^\mu, \bar{z}_*^\mu)\}$ generated by the master algorithm will converge to a particular saddle point $(\bar{x}, \bar{z})$ of $\bar{L}$ on $\mathbb{R}^n \times (\mathbb{R}^m)^\omega$ (the same as a saddle point of L on $X \times Z$) if

$$\|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - P(\bar{x}^\mu, \bar{z}^\mu)\|_* \leq \gamma_\mu \quad \text{with } \gamma_\mu > 0, \sum_{\mu=1}^{\infty} \gamma_\mu < \infty. \tag{5.8}$$

Under the more stringent condition

$$\|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - P(\bar{x}^{\mu}, \bar{z}^{\mu})\|_* \leq \theta_\mu \|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - (\bar{x}^{\mu}, \bar{z}^{\mu})\|_*$$

$$\text{with } \theta_\mu > 0, \ \sum_{\mu=1}^{\infty} \theta_\mu < \infty, \tag{5.9}$$

we know from [7, Theorem 2] that if $(\bar{x}, \bar{z})$ is the unique saddle point of $\bar{L}$ and a certain Lipschitz property holds in terms of a constant $\sigma \geq 0$, we will have

$$\limsup_{\mu \to \infty} \|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - (\bar{x}, \bar{z})\|_* / \|(\bar{x}_*^{\mu}) - (\bar{x}, \bar{z})\|_* = \beta(\eta),$$

where

$$\beta(\eta) = \sigma\eta/(1+\sigma^2\eta^2)^{1/2} < 1. \tag{5.10}$$

The Lipschitz property in question is the following: for $\sigma \geq 0$ and some $\varepsilon > 0$, all the saddle points $(\tilde{x}, \tilde{z})$ of any perturbed Lagrangian of the form

$$\tilde{L}(x, y) = L(x, y) + \tilde{c} \cdot x + E\{\tilde{h}_\omega \cdot z_\omega\} \quad \text{on } X \times Z,$$

with $\tilde{c} \in \mathbb{R}^n$ and $\tilde{h} = (\ldots, \tilde{h}_\omega, \ldots) \in (\mathbb{R}^m)^\Omega$, will satisfy

$$\|(\tilde{x}, \tilde{z}) - (\bar{x}, \bar{z})\|_* \leq \sigma \|(\tilde{c}, \tilde{h})\|_{**} \quad \text{when } \|(\tilde{c}, \tilde{h})\|_{**} \leq \varepsilon.$$

(Here $\|\cdot\|_{**}$ is the norm dual to $\|\cdot\|_*$.) This needed property does hold, because of the quadratic nature of our problem. The optimality conditions that characterize $(\tilde{x}, \tilde{z})$ as a saddle point of $\tilde{L}$ on $X \times Z$ are all linear; the multifunction that associates with each $(\tilde{c}, \tilde{h})$ this corresponding set of saddle points is in fact a *polyhedral* multifunction in the sense of Robinson, i.e. its graph is the union of finitely many convex polyhedra. Any such multifunction has the Lipschitz property in question; see Robinson [6].

We shall show now that our stopping criterion (5.5), (5.7), does imply (5.8) and (5.9) with $\gamma_\nu = \theta_\mu[2/\eta]^{1/2}$. Consider the primal and dual objective functions associated with L_μ namely

$$F_\mu(x) = \max_{z \in Z} L_\mu(x, z), \qquad G_\mu(z) = \min_{x \in X} L_\mu(x, z). \tag{5.11}$$

The approximate saddle point $(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) = (\bar{x}^\nu, \bar{z}^\nu)$ satisfies

$$F_\mu(\bar{x}_*^{\mu+1}) - G_\mu(\bar{z}_*^{\mu+1}) \leq \varepsilon_\nu \leq \varepsilon_*^\mu(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) \tag{5.12}$$

by Theorem 3 (as applied to L_μ) and (5.5). The true saddle point $(\hat{x}_*^\mu, \hat{z}_*^\mu) = P(\bar{x}_*^\mu, \bar{z}_*^\mu)$ satisfies

$$\min_{x \in X} L_\mu(x, \bar{z}_*^\mu) = L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) = \max_{z \in Z} L_\mu(\hat{x}_*^\mu, z),$$

and because L_μ is strongly quadratic by virtue of the terms added to form it from L, this must actually hold in the strong sense that

$$L_\mu(x, \hat{z}_*^\mu) \geq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) + (\eta/2)\|x - \hat{x}_*^\mu\|_*^2 \quad \text{for all } z \in Z,$$

$$L_\mu(\hat{x}_*^\mu, z) \leq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) - (\eta/2)\|z - \hat{z}_*^\mu\|_*^2 \quad \text{for all } z \in Z.$$

Taking $x = \bar{x}_*^{\mu+1}$ and $z = \bar{x}_*^{\mu+1}$ in these inequalities and observing from definition (5.11) that

$$F_\mu(\bar{x}_*^\mu) \geq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) \geq G_\mu(\bar{z}_*^\mu),$$

we obtain

$$\begin{aligned} F_\mu(\bar{x}_*^{\mu+1}) - G_\mu(\bar{z}_*^{\mu+1}) &\geq (\eta/2)\|\bar{x}_*^{\mu+1} - \hat{x}_*^\mu\|_*^2 + (\eta/2)\|\bar{z}_*^{\mu+1} - \hat{z}_*^\mu\|_*^2 \\ &= (\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - P(\bar{x}_*^\mu, \bar{z}_*^\mu)\|_*^2. \end{aligned}$$

This, combined with (5.12) and (5.7), yields

$$\begin{aligned} &(\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - P(\bar{x}_*^\mu, \bar{z}_*^\mu)\|_*^2 \\ &\quad \leq \theta_\mu^2 \min\{1, (\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - (\bar{x}_*^\mu, \bar{z}_*^{\mu+1})\|_*^2\}. \end{aligned}$$

Then (5.8) and (5.9) hold as claimed, with $\gamma_\mu = \theta_\mu[2/\gamma]^{1/2}$. □

We conclude by connecting the choice of the matrices $\bar{C}$ and $\bar{H}_\omega$ in (5.1) with the convergence rate of the finite generation algorithm in Step 1 of the master algorithm.

Proposition 4. *Suppose $\bar{C}$ and $\bar{H}_\omega$ are selected so that for a certain $\bar{\rho} > 0$,*

$$z_\omega \cdot [T_\omega \bar{C}^{-1} T_\omega^*] z_\omega \leq \bar{\rho}[z_\omega \cdot \bar{H}_\omega z_\omega] \quad \text{for all } z_\omega \in \mathbb{R}^m. \tag{5.13}$$

Then the matrices C_* and $H_{*\omega}$ in (5.3) have

$$z_\omega \cdot [T_\omega C_*^{-1} T_\omega^*] z_\omega \leq (\bar{\rho}/\eta^2)[z_\omega \cdot H_{*\omega} z_\omega] \quad \text{for all } z_\omega \in \mathbb{R}^m, \tag{5.14}$$

so that when the finite generation algorithm is applied to finding a saddle point of L_μ, the convergence results in Theorem 3 will be valid for $\rho = \bar{\rho}/\eta^2$.

Proof. Let us simplify notation by writing $A \leq B$ for positive definite symmetric matrices A and B to mean that $B - A$ is positive semidefinite. Since A and B can be diagonalized simultaneously, this relation can be interpreted also as a coordinate-wise inequality on the corresponding vectors of eigenvalues. In this notation, our assumption (5.13) is that $T_\omega \bar{C}^{-1} T_\omega^* \leq \bar{\rho}\bar{H}_\omega$. Since $C_* = C + \eta\bar{C}$ we know $C_* \geq \eta\bar{C}$ and therefore $C_*^{-1} \leq \eta^{-1}\bar{C}^{-1}$. But also, from $H_{*\omega} = H_\omega + \eta\bar{H}_\omega$ we have $\eta\bar{H}_\omega \leq H_{*\omega}$, or in other words $\bar{H}_\omega \leq \eta^{-1} H_{*\omega}$. It follows that

$$T_\omega C_*^{-1} T_\omega^* \leq \eta^{-1} T_\omega \bar{C}^{-1} T_\omega^* \leq \eta^{-1}\bar{\rho}\bar{H}_\omega \leq \eta^{-2}\bar{\rho} H_{*\omega}$$

as claimed in (5.14). □

This result reveals a trade-off between the rates of linear convergence that can be achieved in the finite generation algortihm and in the master algorithm. The modulus $\beta(\eta)$ for the latter can be improved by making η smaller. But one cannot

at the same time make ρ smaller, as would be desirable for the finite generation algorithm in the light of Theorem 5.

References

[1] R.W. Cottle, "Symmetric dual quadratic programs", *Quarterly of Applied Mathematics* 21 (1983) 237-243.

[2] W.S. Dorn, "Duality in quadratic programming", *Quarterly of Applied Mathematics* 18 (1960) 155-162.

[3] M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3 (1956) 95-110.

[4] A. King, "An implementation of the Lagrangian finite-generation method", in: Y. Ermoliev, R. Wets, eds., *Numerical techniques for stochastic optimaization problems.*

[5] H.W. Kuhn and A.W. Tucker, "Nonlinear programming", in: J. Neyman, ed., *Second Berkeley symposium on mathematical statistics* (University of California Press, Berkeley, 1951) pp. 481-492.

[6] S.M. Robinson, "Some continuity properties of polyhedral multifunctions", *Mathematical Programming Study* 14 (1981) 206-214.

[7] R.T. Rockafellar, "Monotone operators and the proximal point algorithm", *SIAM Journal on Control and Optimization* 14 (1976) 877-898.

[8] R.T. Rockafellar, *Conjugate duality and optimization*, No. 16 in Conference Board of the Mathematical Sciences Series, (SIAM Publications, Philadelphia, 1974).

[9] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).

[10] R.T. Rockafellar and R.J.-B. Wets, "Stochastic convex programming: Relatively complete recourse and induced feasibility", *SIAM Journal on Control and Control and Optimization* 14 (1976) 574-589.

[11] R.T. Rockafellar and R.J.-B. Wets, "A dual solution procedure for quadratic stochastic programs with simple recourse", in: V. Pereyra and A. Reinoza, eds., *Numerical methods*, Lecture notes in mathematics 1005 (Springer-Verlag, Berlin, 1938) pp. 252-265.

[12] R.J.-B. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem et al., eds., *Mathematical programming: The state-of-the-art* (Springer-Verlag, Berlin, 1983) pp. 566-603.

Mathematical Programming Study 28 (1986) 94–112
North-Holland

A DUAL METHOD FOR PROBABILISTIC CONSTRAINED PROBLEMS

É. KOMÁROMI
National Management Development Centre, Budapest, Hungary

Received 20 January 1984
Revised manuscript received 2 May 1984

This paper introduces a problem with stochastic objective function and linear constraints as a dual to the probabilistic constrained problem with stochastic right-hand side. The random vector involved in both problems is supposed to have logarithmic concave joint probability distribution. A dual type algorithm is presented for solving both problems simultaneously.

Key words: Probabilistic Constrained Problems, Stochastic Programming, Duality, Optimization.

Introduction

In this paper we are concerned with the following probabilistic constrained problem:

(1) Find $\hat{x}$, if it exists, such that

$$c\hat{x} = \min_{x \in X} cx,$$

$$\hat{x} \in X = \{x \in \mathbb{R}^n | \mathcal{P}(A_1 x \geq \beta) \geq p, A_2 x \geq b\}$$

where A_1 and A_2 are deterministic matrices of dimension $m \times n$ and $r \times n$, b and c are deterministic vectors of dimension r and n, respectively; β is an m-dimensional random vector with associated continuous joint probability distribution function F; $\mathcal{P}$ denotes probability, p is a constant reliability level $(0 < p < 1)$; x is an n-dimensional decision vector.

The original statement of probabilistic constrained programming is due to Charnes and Cooper [1] who prescribed reliability levels for each constraint separately; this problem was shown to be equivalent deterministically to a linear program. The extension of this concept to joint probabilistic constraints was first developed by Miller and Wagner [6], who assumed uncorrelated components in β and later by Prékopa [7] who permitted multivariate distribution F for β.

Prékopa's results on logarithmic concave probability distributions [8] guarantee the convexity of the set of solutions of (1) for a rather broad class of distributions. But even if convexity can be asserted, there are major numerical difficulties to overcome in order to solve it. This is partly due to the fact that handling nonlinear

constraints is usually more complicated than handling nonlinear objective function. In this paper we suggest the other way: we will introduce a dual maximization problem with linear constraints and concave objective function, and develop a dual type algorithm which solves both problems simultaneously under reasonable convexity assumptions.

We begin by formulating a deterministic equivalent to problem (1). Let us observe that $x \in X$ if and only if there exists $y \in \mathbb{R}^m$ such that x and y satisfy the constraints

$$A_1 x \geq y, \qquad A_2 x \geq b, \qquad \mathcal{P}(\beta \leq y) \geq p.$$

Let us also observe that $A_1 x \geq y$ implies that $A_1 x \geq y'$ for all $y' \leq y$. Therefore y can be chosen in the support set of the probability distribution function F: $y \in \operatorname{supp} F \subset \mathbb{R}^m$. (The support of F is the smallest closed subset of $\mathbb{R}^m$ with corresponding probability measure equal to 1. Of course, the equality $\operatorname{supp} F = \mathbb{R}^m$ is not excluded.) Taking into account these observations and the fact that $\mathcal{P}(\beta \leq y) = F(y)$ by the continuity of F, we redefine the set X so as

$$X = \{x \mid \exists y \in \operatorname{supp} F: A_1 x \geq y, A_2 x \geq b, F(y) \geq p\}.$$

Now divide the constraints into two parts: one for the linear constraints and the other for the nonlinear constraint. Accordingly, define the sets

$$Y = \{y \in \operatorname{supp} F \mid F(y) \geq p\}, \qquad X(y) = \{x \mid A_1 x \geq y, A_2 x \geq b\}.$$

Furthermore, for a given $y \in \mathbb{R}^m$ define the convex function $g(y)$ to be

$$g(y) = \begin{cases} \inf\limits_{x \in X(y)} cx & \text{if } X(y) \neq \emptyset, \\ +\infty & \text{otherwise.} \end{cases}$$

Then our original stochastic problem is equivalent to the following deterministic problem:

(2) Find $\hat{y}$, if it exists, such that

$$g(\hat{y}) = \min_{y \in Y} g(y), \quad \hat{y} \in Y.$$

These two problems are equivalent in the sense that if $\hat{y}$ is optimal to (2) then $\hat{x} \in \operatorname{Arg\,min}_{x \in X(\hat{y})} cx$ is optimal to (1), and conversely, if $\hat{x}$ is optimal to (1) then any $\hat{y} \in \{y \in Y \mid A_1 \hat{x} \geq y\}$ is optimal to (2).

Let us now summarize some properties of problem (2):

(a) Y is closed because F is assumed to be continuous.

(b) There exists a $w \in \mathbb{R}^m$, such that $y \in Y$ implies that $y \geq w$. This holds by the fact that the probability distribution is a monotonic function of each of its variables and $F(y)$ tends to 0 if one component y_i of y tends to $-\infty$; for example, for $i = 1, \ldots, r$ set $F_i(w_i) = p$ where F_i is the ith marginal.

(c) If $\sup_{y \in \operatorname{supp} F} y_i < +\infty$ for $i = 1, \ldots, m$, then Y is bounded by (b) and thus compact by (a).

(d) $g(y) > -\infty$ if and only if the dual feasibility set of the linear programming problem $\min_{x \in X(y)} cx$ (see linear programming results in Dantzig's book [2])

$$V = \{(u, v) \geq 0 \mid uA_1 + vA_2 = c\}$$

is nonempty, by the duality theorem of linear programming.

(e) $\inf_{y \in Y} g(y) > -\infty$ if and only if the set V is nonempty, by (b) and (d).

Suppose that the set V is nonempty. Then

$$g(y) \equiv \begin{cases} \max_{(u,v) \in V} (uy + vb) & \text{if } X(y) \neq \emptyset, \\ +\infty & \text{otherwise,} \end{cases}$$

by the duality theorem of linear programming. Thus, problem (2) is to find $\hat{y} \in Y$ such that

$$g(\hat{y}) = \min_{y \in Y} \left[\max_{(u,v) \in V} (uy + vb) \right].$$

Obviously,

$$\sup_{(u,v) \in V} \inf_{y \in Y} (uy + vb) \leq \inf_{y \in Y} \sup_{(u,v) \in V} (uy + vb).$$

If the function $uy + vb$ did have a saddle-point on $V \times Y$ then the original stochastic problem could be solved via maximizing the function $(\min_{y \in Y} uy + vb)$ over the polyhedral set V.

We are ready now to derive the dual problem by examining the existence of a saddlepoint of the function $uy + vb$ (with respect to maximizing over V and minimizing over Y). We begin Section 1 by establishing a theorem about necessary and sufficient optimality criteria for problem (2) in relation with the existence of a saddle-point. Then we obtain the dual problem and investigate its properties when F is continuously differentiable and strictly logarithmic concave. In Section 2, under the same assumption for F, we describe an algorithm for solving problem (2) and the dual simultaneously. The convergence of the algorithm is proved for compact V. Finally, in Section 3 we describe the computational experience we have with this algorithm.

1. The dual problem

The formulation of the dual problem is motivated by the following theorem.

Theorem 1. (1a) *If the function $uy + vb$ has a saddle-point $((\hat{u}, \hat{v}), \hat{y})$ with respect to maximizing over V and minimizing over Y then $\hat{y}$ is an optimal solution of problem* (2) *and $((\hat{u}, \hat{v}), y)$ is a saddle-point for any optimal y.*

(1b) *Let the probability distribution function F be quasiconcave. Let $\hat{y}$ be an optimal solution of problem* (2).

- *If the interior of the set* $\{y | y \in Y, X(y) \neq \emptyset\}$ *is nonempty then there exists* $(\hat{u}, \hat{v}) \in V$ *such that* $((\hat{u}, \hat{v}), \hat{y})$ *is the saddle-point of* $uy + vb$ *with respect to maximizing over* V *and minimizing over* Y.
- *If the interior of the set* $\{y | y \in Y, X(y) \neq \emptyset\}$ *is empty then there exists*

$$(u', v') \in V' = \left\{ (u, v) \geq 0 \,|\, uA_1 + vA_2 = 0, \sum_{i=1}^{m} u_i = 1 \right\}$$

such that $((u', v'), \hat{y})$ *is the saddle-point of* $uy + vb$ *with respect to maximizing over* V' *and minimizing over* Y, $u'\hat{y} + v'b = 0$.

Proof. (1a) follows from the definition of the saddle-point. (1b) The proof is based on the Kuhn–Tucker saddle-point necessary optimality theorem [5, Chapter 5]. According to this theorem,

if $Z^0 = \{(x, y) | y \in Y, A_2 x \geq b\}$ is convex (it is, if F is quasiconcave) and

if there exists $(x', y') \in Z^0$ such that $A_1 x' > y'$ (that is, if $\operatorname{int}\{y \in Y \,|\, X(y) \neq \emptyset\} \neq \emptyset$) and

if $(\hat{x}, \hat{y}) \in Z = \{(x, y) | (x, y) \in Z^0, A_1 x \geq y\}$ is minimizing cx over the set Z (that is, if $\hat{y}$ is optimal for problem (2) and $\hat{x} \in \operatorname{Arg\,min}_{x \in X(\hat{y})} cx$)

then there exists $\hat{u} \geq 0$ such that $\hat{u}(\hat{y} - A_1\hat{x}) = 0$ and

$$c\hat{x} + u(\hat{y} - A_1\hat{x}) \leq c\hat{x} + \hat{u}(\hat{y} - A_1\hat{x}) \leq cx + \hat{u}(y - A_1 x)$$

for any $u \geq 0$, $(x, y) \in Z^0$. The second inequality implies that

$$c\hat{x} = c\hat{x} + \hat{u}(\hat{y} - A_1\hat{x}) = \min_{x \in \{x | A_2 x \geq b\}} (c - \hat{u}A_1)x + \min_{y \in Y} \hat{u}y,$$

so

$$c\hat{x} - \hat{u}A_1\hat{x} = \max_{v \in \{v \geq 0 | vA_2 = c - \hat{u}A_1\}} vb = \hat{v}b \quad \text{for a } \hat{v} \in \{v \geq 0 | vA_2 = c - \hat{u}A_1\}.$$

Hence $(\hat{u}, \hat{v}) \in V$ and

$$u\hat{y} + vb \leq \hat{u}\hat{y} + \hat{v}b = c\hat{x} \leq \hat{u}y + \hat{v}b$$

for any $(u, v) \in V$ and $y \in Y$, which is the first statement of (1b).

To prove the second statement of (1b) we apply the Kuhn–Tucker saddle-point theorem for the following problem:

$$x_0 \to \min,$$

$$A_{1i}x + x_0 \geq y_i, \quad i = 1, \ldots, m,$$

$$(x, y) \in Z^0,$$

where A_{1i} is the ith row of A_1.

This problem satisfies the assumptions of the Kuhn–Tucker theorem:
Z^0 is convex;
for $(\hat{x}, \hat{y})$ and $x_0 > 0$, $A_{1i}\hat{x} + x_0 > \hat{y}_i$ $(i = 1, \ldots, m)$;
$(\hat{x}, \hat{y})$ and $\hat{x}_0 = 0$ is minimizing x_0 because

$$\operatorname{int}\{y \in Y \mid X(y) \neq \emptyset\} = \emptyset.$$

Hence, by the above argumentation, there exists

$$(u', v') \in V' = \left\{ (u, v) \geq 0 \mid uA_1 + vA_2 = 0, \sum_{i=1}^{m} u_i = 1 \right\}$$

such that

$$u\hat{y} + vb \leq u'\hat{y} + v'b = 0 \leq u'y + v'b$$

for any $(u, v) \in V'$ and $y \in Y$. Thus the proof is complete. □

This theorem offers the basis to derive the dual in the form of the following problem:

(3) Find $(\hat{u}, \hat{v})$, if it exists, such that

$$h(\hat{u}, \hat{v}) = \max_{(u,v) \in V} h(u, v),$$

$$(\hat{u}, \hat{v}) \in V = \{(u, v) \geq 0 \mid uA_1 + vA_2 = c\},$$

$$h(u, v) = \begin{cases} \inf\limits_{y \in Y} uy + vb & \text{if } u \geq 0,\ u \neq 0, \\ vb & \text{if } u = 0, \\ -\infty & \text{otherwise,} \end{cases}$$

$$Y = \{y \in \operatorname{supp} F \mid F(y) \geq p\}.$$

It is easy to see that $\inf uy$ is a concave function of u. Hence this problem is always convex, even if the function F is not quasiconcave (the set Y is not convex). However, in the absence of convexity the equality of the optimal values of problems (2) and (3) cannot be guaranteed.

We shall use the notation $u \geq 0$ for *semipositive* vectors u, that is, for u: $u \geq 0$, $u \neq 0$.

Suppose now that the probability distribution function F is *strictly logarithmic concave and continuously differentiable on the interior of* supp F. Since $F(y) \geq p \Leftrightarrow \ln F(y) \geq \ln p$, Y is convex, thus, this assumption implies that $\inf_{y \in Y} uy$ is attained for $u > 0$ at unique point (see Rockafellar's results in [10, Section 27]). Denote this minimum point by $y(u)$:

$$y(u) = \operatorname{Arg} \min_{y \in Y} uy \quad (u > 0).$$

Since $p < 1$ hence the constraint $\ln F(y) \geq \ln p$ fulfills the Slater condition: there exists $y' \in \mathbb{R}^m$ such that $\ln F(y') > \ln p$. Thus there exists $\delta \in \mathbb{R}$ such that δ and $y(u)$

satisfy the following Kuhn–Tucker conditions [4]:

$$u_i = \delta \frac{1}{F(y)} \frac{\partial F(y)}{\partial y_i} \quad (i = 1, \ldots, m),$$
$$\delta \geq 0, \qquad F(y) \geq p, \qquad \delta(\ln p - \ln F(y)) = 0. \tag{K–T}$$

It is known from the probability theory that between the partial derivate $\partial F(y)/\partial y_i$ of F and the conditional probability distribution function $F(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m | y_i)$ of $\beta_1, \ldots, \beta_m$ at $\beta_i = y_i$ the following relation holds:

$$\frac{\partial F(y_1, \ldots, y_m)}{\partial y_i} = F(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m | y_i) \cdot f_i(y_i)$$

where f_i is the probability density of the random variable β_i. This relation implies that

$$y_i(u) \to \sup_{y \in \operatorname{supp} F} y_i \quad \text{if } u_i \to 0$$

where $\sup y_i$ may take the value $+\infty$ when supp F is not bounded.

Let us define $y(u)$ for any $u \geq 0$ by taking this limit. Then the objective function of (3) can be reformulated for $u \geq 0$ so as

$$h(u, v) = uy(u) + vb \quad (u \geq 0).$$

The function $uy(u)$ and hence $h(u, v)$ has remarkably nice properties. These are summarized in the following theorem.

Theorem 2. *Suppose that the probability distribution F is continuously differentiable and strictly logarithmic concave on* int supp F. *Then*

(2a) $y(u)$ *is continuous on* $\{u | u > 0\}$. *If* $\lim_{k \to \infty} u^k = \hat{u}$ *for a sequence* $u^1, u^2, \ldots, (\hat{u} \geq 0,\ u^k \geq 0$ *for all* $k)$ *then* $\lim_{k \to \infty} y(u^k) = y(\hat{u})$.

(2b) $uy(u)$ *is continuous on* $\{u | u \geq 0\}$.

(2c) $uy(u)$ *is strictly concave on* $\{u | u \geq 0\}$ *in the sense that*

$$[\lambda u^1 + (1-\lambda)u^2]y(\lambda u^1 + (1-\lambda)u^2) > \lambda u^1 y(u^1) + (1-\lambda)u^2 y(u^2)$$

for any $u^1 \geq 0,\ u^2 \geq 0,\ u^1/|u^1| \neq u^2/|u^2|,\ 0 < \lambda < 1$.

(2d) $uy(u)$ *is differentiable,* $\nabla uy(u) = y(u)$ *on* $\{u | u > 0\}$.

Proof. (2a) For $u > 0$, $y(u)$ satisfies the above (K–T) conditions where necessarily $F(y(u)) = p$, $\delta > 0$, in fact

$$\delta = p \frac{\sum_{i=1}^m u_i}{\sum_{i=1}^m \partial F(y(u))/\partial y_i}.$$

Therefore the (K–T) conditions represent a continuous one-to-one correspondence between $\{y \in \operatorname{supp} F | F(y) = p\}$ and $\{u > 0 | \sum_{i=1}^m u_i = 1\}$ under the assumption made for F. Hence $y(u)$ is continuous on $\{u | u > 0\}$ which implies that $\lim_{u^k \to \hat{u}} y(u^k) = y(\hat{u})$ if $\hat{u} > 0$. For $\hat{u} \geq 0$, $\hat{u} \not> 0$ the statement follows from the definition of $y(u)$.

(2b) Define $uy(u)$ to be 0 for $u=0$ and $-\infty$ for $u\ngeq 0$. Then $-uy(u)$ is the support function of the convex set Y therefore it is continuous on $\{u \mid u \geqq 0\}$. (For details and results about support functions see [10, Section 13].)

(2c) The statement is immediate from the definition of $y(u)$: $uy' > uy(u)$ for any $u \geqq 0$, $y' \in Y$, $y' \neq y(u)$.

(2d) By definition a $t \in \mathbb{R}^m$ is a subgradient of the concave function $uy(u)$ at $\hat{u}$ if

$$zy(z) \leqq \hat{u}y(\hat{u}) + t(z-\hat{u}) \quad \text{for all } z \geqq 0.$$

By Rockafellar's theorem [10, Section 25] $uy(u)$ is differentiable at $\hat{u}$ and $\nabla \hat{u}y(\hat{u}) = y(\hat{u})$ if $y(\hat{u})$ is the unique subgradient of $uy(u)$ at $\hat{u}$. $t = y(\hat{u})$ is finite at $\hat{u} > 0$ and it satisfies the above inequality because, for any $z \geqq 0$,

$$zy(z) - zt \leqq \hat{u}y(\hat{u}) - \hat{u}t = 0$$

by the definition of $y(z)$.

To prove (2d) we show that for $t \neq y(\hat{u})$ there exists a $z > 0$ such that

$$zy(z) - zt > \hat{u}y(\hat{u}) - \hat{u}t.$$

Suppose first that $\hat{u}y(\hat{u}) - \hat{u}t = \rho \neq 0$. Choose $z = \lambda\hat{u}$, where $\lambda > 1$ if $\rho > 0$, $0 < \lambda < 1$ if $\rho < 0$. Then, taking into account the fact that $y(\lambda\hat{u}) = y(\hat{u})$ for $\lambda > 0$ one has

$$zy(z) - zt = \lambda[\hat{u}y(\hat{u}) - \hat{u}t] > \hat{u}y(\hat{u}) - \hat{u}t.$$

Suppose next that $\hat{u}y(\hat{u}) - \hat{u}t = 0$. Let $I = \{i \mid y_i(\hat{u}) - t_i > 0\} \subset \{1, \ldots, m\}$. $I \neq \emptyset$ since $t \neq y(\hat{u})$ and $\hat{u} > 0$. Choose $z > 0$ and $0 < \lambda < 1$ such that

$$z_i = (1-\lambda)\hat{u}_i \quad \text{if } y_i(\hat{u}) - t_i \leqq 0,$$

$$z_i = \hat{u}_i \qquad \text{if } y_i(\hat{u}) - t_i > 0$$

and $y_i(z) - t_i$ is still nonnegative for $i \in I$. Such a λ exists by (2a). Then

$$zy(z) - zt = (1-\lambda)[\hat{u}y(z) - \hat{u}t] + \lambda \sum_{i \in I} \hat{u}_i[y_i(z) - t_i].$$

But $\hat{u}y(z) > \hat{u}y(\hat{u})$ and $\sum_{i \in I} \hat{u}_i[y_i(z) - t_i] \geqq 0$ by the choice of λ; therefore

$$zy(z) - zt > \hat{u}y(\hat{u}) - \hat{u}t = 0$$

Thus the proof is complete. □

The next two properties are most important from the point of view of the algorithm in Section 2.

Corollary 1. *$u^1y(u+\lambda u^1)$ is a continuous, and strictly decreasing function of λ, $\lim_{\lambda \to \infty} u^1y(u+\lambda u^1) = u^1y(u^1)$ for $\lambda > 0$, $u^1 \geqq 0$, $u \geqq 0$, $u^1/|u^1| \neq u/|u|$, under the assumption of Theorem 2.*

Proof. The fact that $u^1 y(u+\lambda u^1)$ is a continuous function of λ ($\lambda > 0$) with

$$\lim_{\lambda \to \infty} u^1 y(u+\lambda u^1) = \lim_{\lambda \to \infty} u^1 y\left(\frac{1}{\lambda} u + u^1\right) = u^1 y(u^1)$$

is immediate from (2a). To prove that it is strictly decreasing we show that $u^1 y(u+\lambda_1 u^1) < u^1 y(u+\lambda_2 u^1)$ if $0 < \lambda_2 < \lambda_1$. One has

$$\begin{aligned}(u+\lambda_1 u^1) y(u+\lambda_1 u^1) &< (u+\lambda_1 u^1) y(u+\lambda_2 u^1) \\ &= (u+\lambda_2 u^1 + (\lambda_1 - \lambda_2) u^1) y(u+\lambda_2 u^1) \\ &< (u+\lambda_2 u^1) y(u+\lambda_1 u^1) + (\lambda_1 - \lambda_2) u^1 y(u+\lambda_2 u^1)\end{aligned}$$

which implies the statement. □

Corollary 2. *If* (u^1, v^1) *and* (u^2, v^2) *are optimal solutions of problem* (3) *and* $u^1 \geq 0$ *then either* $u^2 = 0$ *or* $u^1/|u^1| = u^2/|u^2|$, *under the assumption of Theorem* 2.

Proof. If, on the contrary, $u^2 \geq 0$ and $u^1/|u^1| \neq u^2/|u^2|$ then

$$\begin{aligned}h(\lambda u^1 + (1-\lambda) u^2, \lambda v^1 + (1-\lambda) v^2) &> \lambda h(u^1, v^1) + (1-\lambda) h(u^2, v^2) \\ &= h(u^1, v^1)\end{aligned}$$

by (2c). Since $(\lambda u^1 + (1-\lambda) u^2, \lambda v^1 + (1-\lambda) v^2) \in V$, it contradicts the assumption that (u^1, v^1) and (u^2, v^2) are optimal. □

We end this section by characterizing the situation in which an optimal solution $(\hat{u}, \hat{v})$ of problem (3) offers an optimal solution for problem (2).

Corollary 3. *Suppose that the assumption of Theorem* 2 *is fulfilled. If there exists* $(\hat{u}, \hat{v}) \in V$, $\hat{u} \geq 0$, $y(\hat{u})$ *finite and* $(u^0, v^0) \in V$, $u^0 = 0$ *such that*

$$v^0 b = \max_{(u,v) \in V} uy(\hat{u}) + vb$$

then $y(\hat{u})$ *is an optimal solution of problem* (2). *If* $V \neq \emptyset$ *and* $(u, v) \in V$ *implies that* $u = 0$ *then any* $y \in Y$ *is an optimal solution of problem* (2).

Proof. The first statement is immediate from the fact that

$$\hat{v} b = \max_{v \in \{v \geq 0 \mid vA_2 \leq c\}} vb = \min_{x \in \{x \mid A_2 x \geq b\}} cx \leq \min_{x \in X(y)} cx$$

for any $y \in \mathbb{R}^m$, $X(y) \neq \emptyset$. The second statement follows from the fact that $(u, v) \in V \Rightarrow u = 0$ is equivalent, by Motzkin's theorem of the alternative (see in [5, Chapter 2]), to: that $\{x \mid A_1 x > 0, A_2 x \geq 0, cx = 0\}$ is not empty.

Theorem 3. *Suppose the assumption of Theorem* 2 *is fulfilled. Suppose that* $(\hat{u}, \hat{v}) \in V$

is an optimal solution of problem (3), $\hat{u} \geq 0$. *Then either* $y(\hat{u})$ *is finite in which case it is the unique optimal solution of problem* (2), *or problem* (2) *has no optimal solution.*

Proof. First we show that no y other than $y(\hat{u})$ is optimal for problem (2). Suppose, on the contrary, that $y' \neq y(\hat{u})$ is optimal. We distinguish two cases: (I) y' is the unique element of Y such that $X(y') \neq \emptyset$; (II) there exists $y'' \neq y'$ such that $y'' \in Y$ and $X(y'') \neq \emptyset$.

In *case* (*I*) $\operatorname{int}\{y \in Y \mid X(y) \neq \emptyset\} = \emptyset$. Hence, by Theorem 1, there exists

$$(u'v') \in \left\{ (u, v) \geq 0 \mid uA_1 + vA_2 = 0, \sum_{i=1}^{m} u_i = 1 \right\}$$

such that

$$y(u') = y' \quad \text{and} \quad u'y' + v'b = 0.$$

By picking $\lambda > 0$ arbitrarily, we get

$$(\hat{u} + \lambda u', \hat{v} + \lambda v') \in V; \qquad \hat{u}y(\hat{u} + \lambda v') > \hat{u}y(\hat{u})$$

$$u'y(\hat{u} + \lambda v') + v'b > u'y(u') + v'b = 0.$$

Hence $h(\hat{u} + \lambda v', \hat{v} + \lambda v') > h(\hat{u}, \hat{v})$, which contradicts the fact that $(\hat{u}, \hat{v})$ is optimal.

In *case* (*II*) $F(\lambda y' + (1-\lambda)y'') > p$ and $X(\lambda y' + (1-\lambda)y'') \neq \emptyset$ for any $0 < \lambda < 1$; therefore $\operatorname{int}\{y \in Y \mid X(y) \neq \emptyset\} \neq \emptyset$. By Theorem 1 there exists $(\hat{\hat{u}}, \hat{\hat{v}}) \in V$ such that

$$y(\hat{\hat{u}}) = y' \text{ if } \hat{\hat{u}} \neq 0 \quad \text{and} \quad \hat{\hat{u}}y' + \hat{\hat{v}}b = \max_{(u,v) \in V} (uy' + vb).$$

Obviously, $\hat{u}y(\hat{u}) + \hat{v}b < \hat{\hat{u}}y' + \hat{\hat{v}}b$ because $y' \neq y(\hat{u})$ and $\hat{u}y' + \hat{v}b \leq \hat{\hat{u}}y' + \hat{\hat{v}}b$ by the duality theorem of linear programming. Thus,

$$h(\hat{u}, \hat{v}) < h(\hat{\hat{u}}, \hat{\hat{v}})$$

which again contradicts the fact that $(\hat{u}, \hat{v})$ is optimal.

Next we show that if $y(\hat{u})$ is finite then $((\hat{u}, \hat{v}), y(\hat{u}))$ is the saddle-point of the function $uy + vb$ with respect to minimizing over Y and maximizing over V:

$$\hat{u}y(\hat{u}) + \hat{v}b = \max_{(u,v) \in V} (uy(\hat{u}) + vb).$$

Suppose, on the contrary, that either (III) $X(y(\hat{u})) = \emptyset$ or (IV) $X(y(\hat{u})) \neq \emptyset$ but $\max_{(u,v) \in V} (uy(\hat{u}) + vb) > h(\hat{u}, \hat{v})$.

In *case* (*III*) the optimal value of the following linear programming problem is positive:

$$x_0 \to \min,$$

$$A_{1i}x + x_0 \geq y_i(\hat{u}), \quad i = 1, \ldots, m,$$

$$A_{2j}x + x_0 \geq b_j, \qquad j = 1, \ldots, r.$$

Hence there exists

$$(u', v') \in \left\{ (u, v) \geq 0 \mid uA_1 + vA_2 = 0, \sum_{i=1}^{m} u_i + \sum_{j=1}^{r} v_j = 1 \right\}$$

such that $u'y(\hat{u}) + v'b > 0$. By picking $\lambda > 0$ such that

$$u'y(\hat{u} + \lambda u') + v'b > 0$$

is still satisfied (such a λ exists by Corollary 1) we get

$$(\hat{u} + \lambda u', \hat{v} + \lambda v') \in V, \qquad \hat{u}y(\hat{u} + \lambda u') \geq \hat{u}y(\hat{u}).$$

Hence $h(\hat{u} + \lambda u', \hat{v} + \lambda v') > h(\hat{u}, \hat{v})$ which contradicts the fact that $(\hat{u}, \hat{v})$ is optimal.

In *case* (*IV*) there exists $(\hat{\hat{u}}, \hat{\hat{v}}) \in V$ such that

$$\hat{u}y(\hat{u}) + \hat{v}b < \hat{\hat{u}}y(\hat{u}) + \hat{\hat{v}}b.$$

By picking $\rho > 0$ such that

$$\hat{\hat{u}}y(\hat{u} + \rho\hat{\hat{u}}) + \hat{\hat{v}}b > \hat{u}y(\hat{u}) + \hat{v}b$$

is still satisfied (such a ρ exists by Corollary 1) and by letting $\lambda = 1/(1+\rho)$ we get

$$(\lambda\hat{u} + (1-\lambda)\hat{\hat{u}}, \lambda\hat{v} + (1-\lambda)\hat{\hat{v}}) \in V,$$

$$\hat{u}y(\hat{u} + \rho\hat{\hat{u}}) = \hat{u}y(\lambda\hat{u} + (1-\lambda)\hat{\hat{u}}) \geq \hat{u}y(\hat{u}).$$

Hence $h(\lambda\hat{u} + (1-\lambda)\hat{\hat{u}}, \lambda\hat{v} + (1-\lambda)\hat{\hat{v}}) > h(\hat{u}, \hat{v})$ which contradicts the fact that $(\hat{u}, \hat{v})$ is optimal. □

2. The algorithm

In this section we introduce an algorithm for solving problem (3) as well as problem (2). It belongs to the family of feasible directions methods. Its concept is similar to that of the Frank–Wolfe linear approximation method [3] which is intended to solve nonlinear problems with linear constraints. The Frank–Wolfe algorithm requires continuously differentiable objective function and uses the gradient at given feasible points as objective function coefficients for linear programming subproblems. The similarity stands in that we use $(y(u), b)$ in the same role, the difference partly stands in that we use it even at the boundary of the feasibility region although $\nabla h(u, v)$ exists and equals $(y(u), b)$ only for $u > 0$ (see (2d) of Theorem 2), and partly in details concerning the choice of stepsize. Moreover, we have to take care of the case when Y is not compact (that is, the support set of F is not 'bounded above' as it is interpreted in (c) in Introduction) hence some components of a $y(u)$ may even have the value $+\infty$, by definition. In practical problems the support of the probability distribution of the random variables in question is necessarily bounded. However, the support of the approximating theoretical distribution func-

tion is often $\mathbb{R}^m$, like for the multivariate normal distribution, furthermore, an unbounded Y does not exclude the existence of a finite optimal solution for problem (2). Thus, we must be prepared for solving linear programming problems with infinite objective function coefficients. The way of doing it is described and motivated in the Appendix.

The algorithm assumes that F is continuously differentiable and strictly logarithmic concave on int supp F. Although it is intended to solve problem (2) it may, in case of unbounded Y, only solve problem (3). However, by getting an optimal $(\hat{u}, \hat{v})$ either an optimal solution for problem (2) is also obtained in the form of $y(\hat{u})$, or it comes to light that problem (2) has no optimal solution (see Theorem 3 and Corollary 3), or a postanalysis may help when $\hat{u} = 0$. The role of the postanalysis is illustrated by the following example

$$x_2 \to \min,$$

$$F(y_1, y_2) \geq p, \quad x_1 \geq y_1, \quad x_2 \geq y_2, x_2 \geq b,$$

where F is a strictly increasing function of both variables, $F(+\infty, a) = p$, $a < b$. The dual problem is

$$\min_{F(y_1,y_2)\geq p} u_2 y_2 + vb \to \max,$$

$$u_1 = 0, \quad u_2 + v = 1, \quad u_2 \geq 0, \quad v \geq 0.$$

For $u_2 > 0$, $v = 1 - u_2$ we get $y_1 = +\infty$, $y_2 = a$. Furthermore, $\hat{u}_2 = 0$, $\hat{v} = 1$ maximizes $u_2 a + vb$ over $u_2 + v = 1$, $u_2, v \geq 0$, hence it is optimal. It, in turn, gives $x_2 \geq a$, $x_2 \geq b$ which implies that all (y_1, y_2) such that $a < y_2 \leq b$, $y_1 \geq \mathrm{Arg}_y[F(y, y_2) = p]$ are optimal for the primal program.

The algorithm starts with a point $(u^1, v^1) \in V$ in which $u^1 \geq 0$, and constructs a sequence $\{(u^k, v^k)\}$ such that $u^k \geq 0$, $(u^k, v^k) \in V$ and $h(u^k, v^k)$ is strictly increasing. It terminates after a finite number of iterations if it

(1) finds an optimal solution for problem (3),

(2) finds unique feasible solution for problem (2), or

(3) finds that problem (2) is not feasible.

Since the optimal value α, if it exists, satisfies

$$u^k y(u^k) + v^k b \leq \alpha \leq \max_{(u,v)\in V} (uy(u^k) + vb)$$

when $X(y(u^k)) \neq \emptyset$, the relative error

$$\frac{[\max_{(u,v)\in V}(uy(u^k) + vb) - u^k y(u^k) - v^k b]}{|\max_{(u,v)\in V}(uy(u^k) + vb)|}$$

quite correctly evaluates the closeness of the objective function value at the given point (u^k, v^k) to the optimum.

Let us come now to the description of the algorithm. Having (u^k, v^k) the kth iteration consists of the following steps:

Step 1. Find the optimal vertex $(\bar{u}^k, \bar{v}^k)$ of V for the following linear programming problem

$$uy(u^k) + vb \to \max, \quad (u, v) \in V,$$

to obtain the direction of ascent $d^k = (\bar{u}^k - u^k, \bar{v}^k - v^k)$.

If failing then continue at Step 3, otherwise continue at Step 2.

Step 2. If $\bar{u}^k = 0$ or $\bar{u}^k/|\bar{u}^k| = u^k/|u^k|$ then set $y(\bar{u}^k) = y(u^k)$ and stop: $(\bar{u}^k, \bar{v}^k)$ is optimal. Otherwise:

Choose the stepsize λ_k such that

$$\lambda_k = 1 \qquad \text{if } h(\bar{u}^k, \bar{v}^k) > h(u^k, v^k),$$

$$0 < \lambda_k < 1 \text{ arbitrary} \qquad \text{if } h(\bar{u}^k, \bar{v}^k) = h(u^k, v^k),$$

$$\lambda_k = \frac{s_k}{1 + s_k} \qquad \text{if } h(\bar{u}^k, \bar{v}^k) < h(u^k, v^k)$$

where $0 < s_k$ satisfies the following equality

$$\bar{u}^k y(u^k + s_k \bar{u}^k) + \bar{v}^k b = h(u^k, v^k).$$

(Such a s_k exists by Corollary 1.) Set

$$(u^{k+1}, v^{k+1}) = (u^k, v^k) + \lambda_k d^k.$$

Step 3. Find an optimal $(\bar{\bar{u}}^k, \bar{\bar{v}}^k) \in V'$ for the following linear programming problem

$$uy(u^k) + vb \to \max,$$

$$(u, v) \in V' = \left\{ (u, v) \geqq 0 \mid uA_1 + vA_2 = 0, \sum_{i=1}^{m} u_i = 1 \right\}$$

to obtain the direction of ascent $d^k = (\bar{\bar{u}}^k, \bar{\bar{v}}^k)$.

If failing or $V' = \emptyset$ then stop: problem (2) is not feasible because $\{x \mid A_2 x \geqq b\} = \emptyset$.

If $h(\bar{\bar{u}}^k, \bar{\bar{v}}^k) > 0$ then stop: problem (2) is not feasible because $\{y \in Y \mid X(y) \neq \emptyset\} = \emptyset$.

If $h(\bar{\bar{u}}^k, \bar{\bar{v}}^k) = 0$ then check whether $X(y(\bar{\bar{u}}^k)) = \emptyset$ and stop: either problem (2) is not feasible or its unique feasible solution is $y(\bar{\bar{u}}^k)$.

If $h(\bar{\bar{u}}^k, \bar{\bar{v}}^k) < 0$ then choose the stepsize $\tau_k > 0$ such that it satisfies the following inequalities

$$\bar{\bar{u}}^k y(u^k + \tau_k \bar{\bar{u}}^k) + \bar{\bar{v}}^k \leq 0,$$

$$h(u^k + \tau_k \bar{\bar{u}}^k, v^k + \tau_k \bar{\bar{v}}^k) > h(u^k, v^k).$$

(Such a τ_k exists by Corollary 1.) Set

$$(u^{k+1}, v^{k+1}) = (u^k, v^k) + \tau_k d^k.$$

We prove the convergence of the algorithm for a compact feasibility region V. In this case the procedure only includes Step 1 and Step 2.

Theorem 4. *Suppose that the assumption for F of Theorem 2 is fulfilled. Suppose that the set V is nonempty and bounded. If the sequence* $\{(u^k, v^k)\}$ *is infinite then it converges for an optimal solution of problem* (3).

Proof. The sequence $\{(u^k, v^k)\}$ has a cluster point $(\hat{u}, \hat{v}) \in V$ because V is compact. Obviously, $\hat{u} \geq 0$, since $\bar{u}^k \geq 0$ for all k in Step 1. Let us characterize the subsequence of $\{(u^k, v^k)\}$ which converges for $(\hat{u}, \hat{v})$ by the increasing subsequence $\mathcal{K}$ of indices $1, 2, \ldots$ and denote it by $\{(u^k, v^k)\}_{\mathcal{K}}$:

$$\lim_{k \in \mathcal{K}} (u^k, v^k) = (\hat{u}, \hat{v}).$$

(Here we follow the notation used by Zangwill in [11].) It is clear that

$$\lim_{k \to \infty} h(u^k, v^k) = \lim_{k \in \mathcal{K}} h(u^k, v^k) = h(\hat{u}, \hat{v})$$

because $h(u^k, v^k)$ is strictly increasing. Since $\{(u^k, v^k)\}_{\mathcal{K}}$ is infinite there exists a vertex (u^*, v^*) of the polyhedron V which is obtained as $(\bar{u}^k, \bar{v}^k)$ in Step 1 for infinite number of the k's, $k \in \mathcal{K}$. Choose $\mathcal{K}^1 \subset \mathcal{K}$ such that

$$(\bar{u}^k, \bar{v}^k) = (u^*, v^*) \quad \text{for all } k \in \mathcal{K}^1.$$

Due to the choice of λ_k in Step 2

$$h(u^*, v^*) < h(u^k, v^k) \quad \text{for all } k \in \mathcal{K}^1 \tag{i}$$

holds.

First, to prove that $\{(u^k, v^k)\}$ itself converges for $(\hat{u}, \hat{v})$ we shall show that $\lim_{k \in \mathcal{K}^1} \lambda_k = 0$. The inequalities $h(u^k, v^k) < h(u^{k+1}, v^{k+1}) < h(\hat{u}, \hat{v})$ imply that

$$\lim_{k \in \mathcal{K}^1} [h(u^{k+1}, v^{k+1}) - h(u^k, v^k)] = 0.$$

Furthermore, it says that

$$\lim_{k \in \mathcal{K}^1} (1 - \lambda_k)\{u^k y((1 - \lambda_k)u^k + \lambda_k u^*) - u^k y(u^k)\} = 0 \tag{ii}$$

while, by the choice of λ_k in Step 2,

$$u^* y((1 - \lambda_k)u^k + \lambda_k u^*) + v^* b - u^k y(u^k) - v^k b = 0. \tag{iii}$$

Since $\{\lambda_k\}_{\mathcal{K}^1}$ is infinite and λ_k is chosen in the interval $[0, 1]$ thus it has a cluster point λ. Let $\mathcal{K}^2 \subset \mathcal{K}^1$ be such that $\{\lambda_k\}_{\mathcal{K}^2}$ is convergent and

$$\lim_{k \in \mathcal{K}^2} \lambda_k = \lambda.$$

Let $\hat{\hat{u}} = (1 - \lambda)\hat{u} + \lambda u^*$.

In case that $\lambda = 1$, $u^* y(u^*) + v^* b = \hat{u} y(\hat{u}) + \hat{v} b$ from (iii) which contradicts (i).

In case that $0<\lambda<1$, $\hat{u}y(\hat{\hat{u}}) = \hat{u}y(\hat{u})$ from (ii) thus $\hat{\hat{u}}/|\hat{\hat{u}}| = \hat{u}/|\hat{u}|$ by definition of $y(u)$. It implies that $u^*/|u^*| = \hat{u}/|\hat{u}|$ and $u^*y(\hat{\hat{u}}) = u^*y(u^*)$. Thereby, $u^*y(u^*) + v^*b = \hat{u}y(\hat{u}) + \hat{v}b$ from (iii) which contradicts (i).

Thus $\lambda = 0$,

$$u^*y(\hat{u}) + v^*b = \hat{u}y(\hat{u}) + \hat{v}b \tag{iv}$$

by (iii), $\lim_{k\in\mathscr{K}^1}\lambda_k = 0$, $\lim_{k\in\mathscr{K}^1}(u^{k+1}, v^{k+1}) = (\hat{u}, \hat{v})$. Since this argumentation is valid for any vertex of V which is obtained as $(\bar{u}^k, \bar{v}^k)$ in Step 1 for infinite number of the k's, $k\in\mathscr{K}$, hence $\lim_{k\in\mathscr{K}}(u^{k+1}, v^{k+1}) = (\hat{u}, \hat{v})$ which proves that $\{(u^k, v^k)\}$ itself converges for $(\hat{u}, \hat{v})$.

Next, we prove that $(\hat{u}, \hat{v})$ is optimal. If, on the contrary, for a vertex $(u', v')\in V$

$$u'y(\hat{u}) + v'b > \hat{u}y(\hat{u}) + \hat{v}b$$

then, by (2a) of Theorem 2, there exists an index N such that

$$u'y(u^k) + v'b > \hat{u}y(\hat{u}) + \hat{v}b \quad \text{for all } k \geqslant N.$$

But

$$u^*y(u^k) + v^*b \geqslant u'y(u^k) + v'b \quad \text{for all } k\in\mathscr{K}^1$$

due to the choice of $(\bar{u}^k, \bar{v}^k) = (u^*, v^*)$ in Step 1 and $u^*y(\hat{u}) + v^*b = \hat{u}y(\hat{u}) + \hat{v}b$ by (iv) hence $u'y(\hat{u}) + v'b \leqslant \hat{u}y(\hat{u}) + \hat{v}b$. Thus the proof is complete. □

3. Computational experiences

We investigated the numerical features of the algorithm on two types of sample problems. In the first one F was assumed to be the joint distribution of normally distributed independent random variables, in the second one F was a two-variable normal distribution. In the first type of the sample problems the data were composed from the data of the STABIL model [9]. The size of the matrix (A_1', A_2') in the examples was 17×17, 21×31, 28×36 and 28×52, respectively, the number of the random variables varied from 3 to 8. Eleven variants were formed concerning the size of matrix, the number of stochastic constraints and the cost coefficients. The sample problems of the second type together with the necessary data were provided by A. Prékopa, they arose from reservoir system design. The problem included a matrix (A_1', A_2') of the size 2×4. Eight variants were formed concerning the elements of the matrix, cost coefficients, reliability level, the expectations, variances and correlation coefficient of the two-variable normal distribution function F. The computations were executed on the IBM 3031 computer of the Hungarian Academy of Sciences. I worked out the computer code in PL 1, for solving the associated linear programming subproblems I used the MPSX code.

In formulating the models we followed the argumentation used by Prékopa in [9]. We replaced the random vector $(\beta_1, \ldots, \beta_m)$ by $(e_1+\sigma_1\eta_1, \ldots, e_m+\sigma_m\eta_m)$

where e_i, σ_i are the expectation and variance of β_i, the random variable η_i was supposed to have expectation 0 and variance 1 ($i = 1, \ldots, m$). Accordingly, the pair of problems to be solved had the following form:

$$\begin{aligned} &cx \to \min, \\ &A_{1i}x \geq e_i + \sigma_i y_i \quad (i = 1, \ldots, m), \\ &A_2 x \geq b, \\ &F(y) \geq p \end{aligned} \tag{A}$$

and

$$\begin{aligned} &\min_{F(y) \geq p} \left(\sum_{i=1}^{m} u_i \sigma_i y_i \right) + \sum_{i=1}^{m} u_i e_i + \sum_{i=1}^{r} v_i b_i \to \max, \\ &uA_1 + vA_2 = c, \quad u \geq 0, \quad v \geq 0, \end{aligned} \tag{B}$$

where F is the joint probability distribution function of $\eta_1, \ldots, \eta_m$, it is normal in which the correlation coefficients equal the correlation coefficients of $\beta_1, \ldots, \beta_m$.

Thus, in the first type of problems

$$F(y) = \prod_{i=1}^{m} G(y_i) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-(1/2)x^2} \, dx$$

and in the second type

$$F(y_1, y_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} e^{-(1/2)(x_1, x_2) C^{-1} (x_1, x_2)'} \, dx_1 \, dx_2$$

where ρ is the correlation coefficient of β_1 and β_2, C^{-1} is the inverse of the correlation matrix:

$$C^{-1} = \begin{bmatrix} \frac{1}{1-\rho^2} & -\frac{\rho}{1-\rho^2} \\ -\frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix}.$$

It is easy to verify that

– all the assertions made for problem (2) and problem (3) hold for (A) and (B), too;

– the algorithm for solving (A) and (B) only differs from the algorithm presented in Section 2 in insignificant details.

We summarize here a few technical details concerning the computations:

– In Step 2 of the algorithm, instead of the stepsize λ_k specified in Section 2, I used a stepsize $0 < \lambda < 1$ fixed in advance. Due to this fact for some of the sample problems the value of the objective function $h(u, v)$ temporarily decreased. I tried 3 values for λ; these were: 0.25, 0.5, 0.75. As to the number of the necessary iterations I did not find considerable differences, thus I accepted the value 0.75.

– In Step 3 of the algorithm, in the choice of the stepsize τ_k, I also tried to neglect the second condition specified in the description of the algorithm in Section 2. However, I found the procedure faster when I stayed with the specifications.

– In Step 1 and Step 3 the value 10^6 was given to $y_i(u^k)$ if u_i^k happened to be 0.

– As to the 'optimality criterion' the procedure was terminated upon attaining a relative error

$$[\max_{(u,v)\in V}(uy(u^k)+vb)-h(u^k, v^k)]/|\max_{(u,v)\in V}(uy(u^k)+vb)|$$

less than 2%.

– In all the sample problems the feasibility region V of (B) was unbounded. As to the reliability level p, I used 3 values; these were 0.85, 0.9, 0.95.

– In case of both types of F the numerical determination of $y(u)$ was carried out by solving the (K–T) problem described in Section 1.

(i) For $F(y)=\prod_{i=1}^{m} G(y_i)$ the (K–T) problem is the following:

$$u_i=\delta\, e^{-(1/2)y_i^2}\Big/\int_{\infty}^{y_i} e^{-(1/2)x^2}\,dx, \quad i\in I(u)=\{i|u_i>0\}\subset\{1,\ldots,m\},$$

$$\delta=p\cdot e^{(1/2)\sum_{k\in I(u)}y_k^2}\prod_{k\in I(u)}(u_k\cdot\sqrt{2\pi}).$$

For solving this problem we used the following procedure: choose y_i^1, $i\in I(u)$ arbitrarily, set $k=1$. Then

(1) If $|p-\prod_{i\in I(u)}G(y_i^k)|<\varepsilon$ then stop. Otherwise:

(2) Compute

$$\delta^k=p\cdot e^{(1/2)\sum_{i\in I(u)}(y_i^k)^2}\prod_{i\in I(u)}(u_i\cdot\sqrt{2\pi}).$$

(3) Compute

$$y_i^{k+1}=\operatorname*{Arg}_z\left[\frac{u_i}{\delta^k}=e^{-(1/2)z^2}\Big/\int_{-\infty}^{z}e^{-(1/2)x^2}\,dx\right]$$

for $i\in I(u)$.

(4) Set $k=k+1$, go to (1).

This procedure works because the (K–T) problem has unique solution and because $\ln G(z)$ is concave and hence

$$\frac{d\ln G(z)}{dz}=e^{-(1/2)z^2}\Big/\int_{-\infty}^{z}e^{-(1/2)x^2}\,dx$$

is a decreasing function of z. For carrying out this procedure we tabulated the values of $G(z)$ and $d\ln G(z)/dz$ at the beginning of the computation.

(ii) For

$$F(y_1,y_2)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}e^{-(1/2)(x_1,x_2)C^{-1}(x_1,x_2)'}\,dx_1\,dx_2$$

we used the replacement $z = (x_1 - \rho x_2)/\sqrt{1-\rho^2}$ and got

$$F(y_1, y_2) = \frac{1}{2\pi} \int_{-\infty}^{y_2} e^{-(1/2)x_2^2} \int_{-\infty}^{(y_1 - \rho x_2)/\sqrt{1-\rho^2}} e^{-(1/2)z^2} \, dz \, dx_2$$

and $F(y_2, y_1) = F(y_1, y_2)$. Then the (K–T) problem is the following:

$$u_1\sigma_1 = \delta \frac{1}{F(y_1, y_2)} \frac{1}{2\pi} e^{-(1/2)y_1^2} \int_{-\infty}^{(y_2 - \rho y_1)/\sqrt{1-\rho^2}} e^{-(1/2)z^2} \, dz,$$

$$u_2\sigma_2 = \delta \frac{1}{F(y_1, y_2)} \frac{1}{2\pi} e^{-(1/2)y_2^2} \int_{-\infty}^{(y_1 - \rho y_2)/\sqrt{1-\rho^2}} e^{-(1/2)z^2} \, dz,$$

$$\delta \geqslant 0, \quad F(y_1, y_2) = p.$$

This is equivalent to finding a solution for

$$\frac{1}{u_1\sigma_1} e^{-(1/2)y_1^2} \int_{-\infty}^{(y_2 - \rho y_1)/\sqrt{1-\rho^2}} e^{-(1/2)z^2} \, dz$$

$$- \frac{1}{u_2\sigma_2} e^{-(1/2)y_2^2} \int_{-\infty}^{(y_1 - \rho y_2)/\sqrt{1-\rho^2}} e^{-(1/2)z^2} \, dz = 0, \quad F(y_1, y_2) = p.$$

For the solution of this problem we prepared a tabulation of $(y_1, y_2) = \mathrm{Arg}_{(z_1, z_2)}[F(z_1, z_2) = p]$ and of $G(z)$ at the beginning of the computation. Then for $y(u_1\sigma_1, u_2\sigma_2)$ we accepted $(\hat{y}_1, \hat{y}_2)$ for which the absolute value of the lefthand side of the first equality was minimum.

The computational experiences concerning the speed of convergence are encouraging: after not more than five iterations each of the procedures terminated upon attaining a relative error less than 2%. At the worst the computation included the solution of 7 linear programs and the determination of the Kuhn–Tucker stationary point $y(u)$ at 15 turns. For the first type of sample problems we found that the number of subproblems to be solved (the linear programs as well as the (K–T) problems) mainly depended on the number of random variables, less on the size of the problem.

Appendix

It remains to interpret a linear programming problem with infinite objective function coefficients, and to show how to find an optimal solution for it. Let the M-dimensional vector s be such that $s_i = +\infty$ for $1 \leqslant i \leqslant M_1$, s_i is a finite constant for $M_1 + 1 \leqslant i \leqslant M$, $M_1 < M$. Let D be a matrix of dimension $M \times N$, d be a vector of dimension N both having fixed finite elements. Suppose that $W = \{w \in \mathbb{R}^M \mid wD = d, w \geqslant 0\} \neq \emptyset$. Then we say the problem

$$sw \to \max, \quad w \in W, \tag{A1}$$

is solvable if the linear program

$$\left(\sum_{i=1}^{M_1} w_i\right)\gamma + \sum_{i=M_1+1}^{M} s_i w_i \to \max, \quad w \in W, \tag{A2}$$

is solvable for all $\gamma \in \mathbb{R}$.

We will show that if (A1) is solvable then there exists a vertex w_0 of W and $\gamma_0 \in \mathbb{R}$ such that w_0 is an optimal solution of (A2) for all $\gamma \geq \gamma_0$. By the linear programming duality theorem (A1) is solvable if and only if

$$X_\gamma = \{x \in \mathbb{R}^N \mid D_i x \geq \gamma (i = 1, \ldots, M), D_i x \geq s_i (i = M_1 + 1, \ldots, M)\} \neq \emptyset$$

for all $\gamma \in \mathbb{R}$. (Here D_i is the i-th row of D.) Furthermore, it is easy to verify that $X_\gamma \neq \emptyset$ for all $\gamma \in \mathbb{R}$ if and only if both

$$X^s = \{x \mid D_i x \geq 0 \ (i = 1, \ldots, M_1), D_i x \geq s_i \ (i = M_1 + 1, \ldots, M)\} \neq \emptyset$$

and

$$\tilde{X} = \{x \mid D_i x \geq 1 \ (i = 1, \ldots, M_1), D_i x \geq 0 \ (i = M_1 + 1, \ldots, M)\} \neq \emptyset.$$

In order to construct w_0 and γ_0 suppose $\tilde{X} \neq \emptyset$ and $X^s \neq \emptyset$. Let $\tilde{x} \in \operatorname{Arg\,min}_{x \in \tilde{X}} \mathrm{d}x$ and $d_0 = \mathrm{d}\tilde{x}$. Let

$$X_0 = \{(x, x_0) \mid D_i x - x_0 \geq 0 \ (i = 1, \ldots, M_1), D_i x \geq s_i \ (i = M_1 + 1, \ldots, M)\}$$

and

$$W_0 = \left\{ w \mid w \in W, \sum_{i=1}^{M_1} w_i = d_0 \right\}.$$

$X_0 \neq \emptyset$ because $X^s \neq \emptyset$ and $W_0 \neq \emptyset$ because $d_0 = \max_{w \in W} \sum_{i=1}^{M_1} w_i$. Let

$$(\hat{x}, \hat{x}_0) \in \operatorname{Arg} \min_{(x, x_0) \in X_0} (cx - d_0 x_0) \quad \text{and} \quad \hat{w} \in \operatorname{Arg} \max_{w \in W_0} \sum_{i=M_1+1}^{M} s_i w_i.$$

Then $\mathrm{d}\hat{x} - d_0 \hat{x}_0 = \sum_{i=M_1+1}^{M} s_i \hat{w}_i$ which says

$$\mathrm{d}[\hat{x} + (\gamma - \hat{x}_0)\tilde{x}] = \left(\sum_{i=1}^{M_1} \hat{w}_i\right)\gamma + \sum_{i=M_1+1}^{M} s_i \hat{w}_i$$

for any $\gamma \in \mathbb{R}$. Since $\hat{x} + (\gamma - \hat{x}_0)\tilde{x} \in X_\gamma$ if $\gamma - \hat{x}_0 \geq 0$ hence $w_0 = \hat{w}$ is optimal for (A2) for any $\gamma \geq \gamma_0 = \hat{x}_0$. Furthermore, w_0 is a vertex of W if it was chosen to be a vertex of W_0 because W_0 is a face of W. Thus, we have shown that in order to compute an optimal vertex of W for (A1) it is enough to compute an optimal vertex for (A2) for a 'sufficiently big' γ—namely, for $\gamma \geq x_0$.

Acknowledgement

I am indebted to I. Dancs whose suggestions initiated this work and helped me in elaborating it.

References

[1] A. Charnes and W.W. Cooper, "Chance constrained programming", *Management Science* 6 (1959) 73–89.
[2] G.B. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, NJ, 1963).
[3] M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3 (1956) 95–110.
[4] H.W. Kuhn and A.W. Tucker, "Nonlinear programming", in: J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, CA, 1951) pp. 481–492.
[5] O.L. Mangasarian, *Nonlinear programming* (McGraw-Hill, New York, 1969).
[6] B.L. Miller and H.M. Wagner, "Chance constrained programming with joint constraints", *Operations Research* 13 (1965) 930–945.
[7] A. Prékopa, "Contributions to the theory of stochastic programming", *Mathematical Programming* 4 (1973) 202–221.
[8] A. Prékopa, "On logarithmic concave measures and functions", *Acta Scientiarum Mathematicarum* 34 (1973) 335–343.
[9] A. Prékopa, S. Ganczer, J. Deák and K. Patyi, "The STABIL stochastic programming model and its experimental application to the electrical energy sector of the Hungarian economy", in: M.A.H. Dempster, ed., *Stochastic programming* (Academic Press, London, 1980) pp. 369–385.
[10] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).
[11] W.I. Zangwill, *Nonlinear programming* (Prentice-Hall, Englewood Cliffs, NJ, 1969).

Mathematical Programming Study 28 (1986) 113 131
North-Holland

A METHOD OF AGGREGATE STOCHASTIC SUBGRADIENTS WITH ON-LINE STEPSIZE RULES FOR CONVEX STOCHASTIC PROGRAMMING PROBLEMS

Andrzej RUSZCZYŃSKI and Wojciech SYSKI
Instytut Automatyki, Politechnika Warszawska, 00-665 Warszawa, Poland

Received 30 November 1983
Revised manuscript received 3 September 1984

A new stochastic subgradient algorithm for solving convex stochastic programming problems is described. It uses an auxiliary filter to average stochastic subgradients observed and is provided with on-line rules for determining stepsizes and filter gains in the course of computation. Convergence with probability one is proved and numerical examples are described.

Key words: Stochastic Programming, Nondifferentiable Programming, Subgradient Optimization.

1. Introduction

In this paper we define and analyse a stochastic algorithm for solving the problem

$$\text{minimize} \quad F(x) \quad \text{over } x \in X, \tag{1}$$

where X is a convex compact subset of $\mathbb{R}^n$ and F is a real-valued convex function on $\mathbb{R}^n$. We assume that neither the values of F nor its subgradients are available. Instead of those, at any point x^k one can only obtain a random vector $\xi^k = g^k + r^k$, where $g^k \in \partial F(x^k)$ and r^k is a random noise of zero mean. We shall call ξ^k a *stochastic subgradient* of F at x^k.

Example 1. Consider a convex *two-stage problem of stochastic programming* (see e.g. [6, 14]), in which one has

$$F(x) = E\left\{ \min_{y \in Y(x,\omega)} f_0(x, y, \omega) \right\},$$

$$Y(x, \omega) = \{y \in \mathbb{R}^m : f_i(x, y, \omega) \leq 0,\ i = 1, \ldots, p\}.$$

Here ω is an elementary event in a sample space (Ω, Σ, P), E denotes mathematical expectation and the functions $f_i : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}^1$, $i = 0, \ldots, p$, are convex in (x, y). In such a problem, if the mathematical expectation exists for all $x \in X$ then F is convex on X, but its values and subgradients are rather difficult to calculate. Still, for a given x^k one can sample ω^k, minimize $f_0(x^k, y, \omega^k)$ over $y \in Y(x^k, \omega^k)$ and thus obtain a solution y^k and Lagrange multipliers λ_i^k, $i = 1, \ldots, p$. Then the vector

$\xi^k = g_0^k + \sum_{i=1}^p \lambda_i^k g_i^k$, where $g_i^k \in \partial_x f_i(x^k, y^k, \omega^k)$, is a stochastic subgradient of F at x^k, i.e. $E\{\xi^k / x^k\} \in \partial F(x^k)$, provided F is finite around x^k (see [2, 3]).

Example 2. In a convex *stochastic min-max problem* one has

$$F(x) = E\left\{ \max_{1 \leq i \leq m} f_i(x, \omega) \right\}$$

where $f_i: \mathbb{R}^n \times \Omega \to \mathbb{R}^1$ are convex in x. Although it is again hard to evaluate $F(x)$ or its subgradients, for a given x^k one can draw ω^k, find i_k such that $f_{i_k}(x^k, \omega^k) \geq f_i(x^k, \omega^k)$, $i = 1, \ldots, m$, and choose $\xi_k \in \partial_x f_{i_k}(x^k, \omega^k)$. Under simple conditions ξ^k is a stochastic subgradient of F at x^k (see e.g. [2, 13]).

In [1] and [3] a *stochastic subgradient method* for solving (1) was suggested, which consists in the following iterations

$$x^{k+1} = \Pi_X(x^k - \tau_k \xi^k), \quad k = 0, 1, 2, \ldots, \tag{2}$$

where τ_k is a nonnegative stepsize coefficient and Π_X is the orthogonal projection onto X. This method was further extended to nonconvex problems in [9, 13]. Serious difficulties however, are connected with an application of (2) as well as many other stochastic algorithms. First, the direction negative to a subgradient is a poor search direction even in the deterministic case, and in (2) it is additionally disturbed with a random noise r^k. Secondly, it is not easy to determine the sequence of stepsizes $\{\tau_k\}$, since classical rules ($\sum_{k=0}^{\infty} \tau_k = \infty$, $\sum_{k=0}^{\infty} E\tau_k^2 < \infty$) are insufficient for practical computation.

In this paper we aim at constructing (at least for the simplest convex case), a new practical stochastic subgradient algorithm for solving (1). Its idea consists in the application of an auxiliary filter which averages stochastic subgradients observed and produces directions for iterating x^k. This, when applied far from the solution, helps to filter out the direction towards minimum. Similar ideas were used in many works (cf. [4, 5, 7, 10, 15]) and proved efficient for smooth stochastic problems (see [15]) and nonsmooth deterministic problems (see [8]). A novel feature of the new method is the rule for determining stepsizes and coefficients appearing in the auxiliary filter. They are controlled on-line on the basis of the information gathered in the course of calculation. Our approach combines, extends and modifies the methods introduced in [11, 15 and 18]. Our hope is that such stepsize rules will enhance convergence far from the solution, while retaining the local properties of classical methods.

In Section 2 we describe the method and formulate assumptions used in further analysis. In Section 3 we derive some important properties of our stepsize rules and in Section 4 we establish convergence of the method. Section 5 concerns some asymptotic properties of the method. Finally Section 6 is devoted to the description of the practical code, and section 7 provides some numerical evidence.

We use $\langle \cdot, \cdot \rangle$ and $|\cdot|$ to denote the usual inner product and norm in n-dimensional Euclidean space $\mathbb{R}^n$. For a convex function $F: \mathbb{R}^n \to \mathbb{R}^1$ we denote by $\partial F(x)$ the

subdifferential of F at x, i.e. $\partial F(x) = \{g \in \mathbb{R}^n\colon F(y) \geq F(x) + \langle g, y - x\rangle \text{ for all } y \in \mathbb{R}^n\}$. A sequence $x^0, x^1, x^2, \ldots$ is written $\{x^k\}$ and $\{x^k\}_{k\in\mathcal{K}}$ denotes its subsequence associated with an infinite set of indices $\mathcal{K}$. We denote by $E\xi$ the mathematical expectation of a random variable ξ and by $E\{\xi/\mathcal{F}_k\}$ or simply $E_k\xi$ its conditional expectation with respect to a σ-subfield $\mathcal{F}_k$. Abbreviation a.s. is used for 'almost surely'.

2. The algorithm and assumptions

The algorithm generates sequences of random directions d^k and points x^k in $\mathbb{R}^n$, $k = 0, 1, 2, \ldots$, according to the formulae

$$d^k = (\xi^k + i_k \gamma_k d^{k-1})/(1 + \gamma_k), \tag{3}$$

$$x^{k+1} = \Pi_X(x^k - \min(\tau_k(1+\gamma_k),\ t/|d^k|)d^k), \tag{4}$$

where ξ^k is a stochastic subgradient of F at x^k, i.e.

$$\xi^k = g^k + r^k, \quad g^k \in \partial F(x^k), \tag{5}$$

and r^k is a random noise. In (3) and (4), τ_k is a positive *stepsize coefficient*, γ_k is a positive *aggregation coefficient* $i_k \in \{0, 1\}$ is a *reset coefficient*, and $t \in (0; +\infty]$. At the starting point $x^0 \in X$ we set $d^{-1} = 0$ and thus it follows from (3) that the direction d^k is a convex combination of the null vector and the previous stochastic subgradients ξ^i, $i = 0, \ldots, k$. We shall call it the *aggregate stochastic subgradient*. It is used to iterate x^k as in the common projection method, with an optional (for $t < +\infty$) truncation of steplengths.

The stepsizes $\{\tau_k\}$ are computed recursively as follows:

$$\tau_0 > 0, \qquad \tau_k = \min\{\bar{\tau},\ \tau_{k-1} \exp[\min(\eta, -\alpha u_k - j_k \delta \tau_{k-1})]\},\ k = 1, 2, \ldots, \tag{6}$$

where

$$u_k = \langle \xi^k, \Delta x^k\rangle + \lambda |\Delta x^k|^2, \tag{7}$$

$\Delta x^k = x^k - x^{k-1}$, and $\bar{\tau} > 0$, $\eta > 0$, $\alpha > 0$, $\delta > 0$ and $\lambda \geq 0$ are fixed parameters. The coefficient j_k in (6) is a binary multiplier satisfying the relations

$$\begin{aligned} &j_k \in \{0, 1\} \quad \text{if } |\Delta x^k| \geq \Delta_{\min}, \\ &j_k = 1 \qquad \text{if } |\Delta x^k| < \Delta_{\min}, \end{aligned} \tag{8}$$

where $\Delta_{\min}$ is a small positive constant.

Similar rules are used for determining the aggregation coefficients $\{\gamma_k\}$:

$$\gamma_0 = \gamma_1 > 0,$$

$$\gamma_k = \min\{\bar{\gamma},\ \gamma_{k-1} \exp(-\beta v_k - j_k \kappa \gamma_{k-1})\}, \quad k = 2, 3, \ldots \tag{9}$$

with

$$v_k = i_{k-1}(\langle \xi^k, \Delta x^{k-1}\rangle + \lambda\langle \Delta x^k, \Delta x^{k-1}\rangle), \tag{10}$$

and some parameters $\bar{\gamma} > 0$, $\kappa > 0$.

Finally, the reset coefficients $\{i_k\}$ are defined as follows:

$$\begin{aligned} &i_k \in \{0, 1\} \quad \text{if } |\xi^{k-1}| \leq \sigma, \\ &i_k = 0 \qquad \text{if } |\xi^{k-1}| > \sigma, \end{aligned} \tag{11}$$

where $\sigma > 0$ is a fixed threshold.

In further considerations we denote by $\mathcal{F}_k$ the σ-subfield generated by $\{x^0, \xi^0, \ldots, x^{k-1}, \xi^{k-1}, x^k\}$ and by E_k the conditional expectation with respect to $\mathcal{F}_k$.

Remark 1. To motivate the rules (6) and (9) suppose that the algorithm operates in the interior of X and uses $t = +\infty$. For given x^{k-1} and d^{k-2} consider the *regularized improvement function*

$$\varphi_k(\tau, \gamma) = F(x^k(\tau, \gamma, \xi^{k-1})) - F(x^{k-1}) + \tfrac{1}{2}\lambda |x^k(\tau, \gamma, \xi^{k-1}) - x^{k-1}|^2, \tag{12}$$

where $x^k(\tau, \gamma, \xi^{k-1})$ is defined by (3) and (4), i.e.

$$x^k(\tau, \gamma, \xi^{k-1}) = x^{k-1} - \tau(\xi^{k-1} + i_{k-1}\gamma d^{k-2}).$$

A natural and most convenient solution would be to choose τ_{k-1} and γ_{k-1} so as to minimize $\Phi_k(\tau, \gamma) = E_{k-1}\varphi_k(\tau, \gamma)$. This, however, is extremely difficult to realize. Therefore, let us use some values of τ_{k-1} and γ_{k-1} and try to verify a posteriori, i.e. at x^k, their optimality for (12). After simple calculations one obtains

$$\partial\varphi_k(\tau_{k-1}, \gamma_{k-1}) = \left\{(\tilde{u}, \tilde{v}): \tilde{u} = \frac{1}{\tau_{k-1}}[\langle g^k, \Delta x^k\rangle + \lambda|\Delta x^k|^2],\right.$$

$$\left.\tilde{v} = \frac{\tau_{k-1} i_{k-1}}{\tau_{k-2}(1+\gamma_{k-2})}[\langle g^k, \Delta x^{k-1}\rangle + \lambda\langle \Delta x^k, \Delta x^{k-1}\rangle],\ g^k \in \partial F(x^k)\right\}.$$

Recalling (7) and (10) we get

$$E_{k-1}\begin{bmatrix} \dfrac{1}{\tau_{k-1}} & u_k \\ \dfrac{\tau_{k-1}}{\tau_{k-2}(1+\gamma_{k-2})} & v_k \end{bmatrix} \in \partial\Phi_k(\tau_{k-1}, \gamma_{k-1}).$$

Thus the vector (u_k, v_k) may be interpreted as a stochastic subgradient of Φ_k at $(\tau_{k-1}, \gamma_{k-1})$. It is used in (6) and (9) to correct the coefficients τ_{k-1} and γ_{k-1} for the next iteration.

The additional terms $j_k\delta\tau_{k-1}$ and $j_k\kappa\gamma_{k-1}$ in (6) and (9) are to force a slow decrease of $\{\tau_k\}$ and $\{\gamma_k\}$ in case of u_k and v_k being close to zero.

Let us formulate assumptions which we shall use in the analysis of convergence of our method.

Assumptions

(H1) $\lambda + \nu > 0$, where

$$\nu = \sup\{\mu\colon F(y) \geq F(x) + \langle g, y - x\rangle + \mu|y - x|^2$$

for all $x, y \in X$ and all $g \in \partial F(x)\}$.

(H2) $E_k r^k = 0$ for every $k \geq 0$.

(H3) There exist constants $s_0 > 0$ and S such that for any $z \in \mathbb{R}^n$ with $|z| \leq s_0$ one has

$$E_k \exp\langle z, r^k\rangle \leq S \text{ a.s. for all } k \geq 0.$$

Remark 2. It follows from (H1) that in case of a strongly convex objective (with $\nu > 0$) one can remove the regularizing term from (12), i.e. use $\lambda = 0$ in the algorithm.

Remark 3. Condition (H3) is closely related to the stepsize rules (6) and (9), since the noise appears there in the exponents. (H3) is a kind of Cramér's condition for vector-valued variables and holds for each bounded distribution of r^k as well as for many unbounded distributions.

3. Properties of stepsizes

In this section we prove that the sequences $\{\tau_k\}$ and $\{\gamma_k\}$, although determined on-line in a sophisticated way, possess some of the properties usually required from the coefficients in recursive stochastic algorithms. We start from two simple auxiliary results.

Lemma 1. *There exist $\bar{\delta} > 0$ and $\varepsilon > 0$ such that, for all $k \geq 0$,*

$$\tau_k \leq \tau_{k-1} \exp[-\alpha(F(x^k) - F(x^{k-1}) + \langle r^k, \Delta x^k\rangle + \varepsilon|\Delta x^k|^2) - \bar{\delta}\tau_{k-1}].$$

Proof. By (5), (7) and (H1),

$$u_k \geq \langle r_k, \Delta x^k\rangle + F(x^k) - F(x^{k-1}) + (\lambda + \nu)|\Delta x^k|^2.$$

Choosing $0 < \varepsilon < \lambda + \nu$ we see from (6) and the definition of j_k that one can find $\bar{\delta} > 0$ such that

$$\alpha(\lambda + \nu - \varepsilon)|\Delta x^k|^2 + j_k \delta \tau_{k-1} \geq \bar{\delta}\tau_{k-1}$$

for each k, which completes the proof.

Lemma 2. *For each $\varepsilon > 0$ one can find $s_1 > 0$ such that for any $0 < s < s_1$ and every $k \geq 1$ one has*

$$E_k \exp[-\alpha s(\langle r^k, \Delta x^k\rangle + \varepsilon|\Delta x^k|^2)] \leq 1.$$

Proof. By the boundedness of X, $|\Delta x^k| \leq T$ for some $T > 0$ and each k. Hence it follows from (H3) that $E_k \exp(-\alpha s\langle r^k, \Delta x^k\rangle)$ exists for all $0 \leq s \leq s_0/\alpha T$. Let us now use the inequality

$$\exp(-py) + \exp(py) \leq 2 + p^2[\exp(-y) + \exp(y)]$$

which holds for every $0 \leq p \leq 1$ and each $y \in \mathbb{R}^1$. Setting $p = \alpha s|\Delta x^k|/s_0$ and $y = s_0\langle r^k, \Delta x^k\rangle/|\Delta x^k|$ we obtain for $0 \leq s \leq s_0/\alpha T$ the relation

$$\begin{aligned}&\exp(-\alpha s\langle r^k, \Delta x^k\rangle) + \exp(\alpha s\langle r^k, \Delta x^k\rangle)\\ &\quad \leq 2 + (\alpha s|\Delta x^k|/s_0)^2[\exp(-s_0\langle r^k, \Delta x^k\rangle/|\Delta x^k|)\\ &\qquad + \exp(s_0\langle r^k, \Delta x^k\rangle/|\Delta x^k|)].\end{aligned}$$

Let us apply the operator E_k to both sides of the above inequality. By (H3), the conditional expectation of the right-hand side does not exceed $2 + Cs^2|\Delta x^k|^2$, where $C = 2S(\alpha/s_0)^2$. Observe also that by (H2) and the Jensen's inequality one has $E_k \exp(\alpha s\langle r^k, \Delta x^k\rangle) \geq 1$. Therefore

$$E_k \exp(-\alpha s\langle r^k, \Delta x^k\rangle) \leq 1 + Cs^2|\Delta x^k|^2 \leq \exp(Cs^2|\Delta x^k|^2)$$

for every $0 \leq s \leq s_0/\alpha T$ and each k. If $s \leq \alpha\varepsilon/C$ we obtain

$$E_k \exp(-\alpha s\langle r^k, \Delta x^k\rangle) \leq \exp(\alpha s\varepsilon|\Delta x^k|^2),$$

as required.

We are now ready to derive the first important property of stepsizes $\{\tau_k\}$.

Lemma 3. *For any $s > 0$ one has*

$$\sum_{k=0}^{\infty} E\{\tau_k^{1+s}\} < \infty.$$

Proof. By Lemmas 1 and 2, for all sufficiently small $s > 0$ we have

$$E_k\{\tau_k^s \exp[\alpha s F(x^k)]\} \leq \tau_{k-1}^s \exp[\alpha s F(x^{k-1})] \exp(-s\bar{\delta}\tau_{k-1}).$$

Since $0 < \tau_{k-1} \leq \bar{\tau}$, one has $\exp(-s\bar{\delta}\tau_{k-1}) \leq 1 - C\tau_{k-1}$, where $C = [1 - \exp(-s\bar{\delta}\bar{\tau})]/\bar{\tau}$. Define $p_k = \tau_k^s \exp[\alpha s F(x^k)]$. We obtain

$$E_k p_k \leq p_{k-1} \exp(-s\bar{\delta}\tau_{k-1}) \leq p_{k-1} - Csp_{k-1}\tau_{k-1}, \quad k = 1, 2, \ldots.$$

Taking the expectation of both sides of the above inequality and noting that $p_k \geq 0$ for all $k \geq 0$, we conclude that $\sum_{k=0}^{\infty} E\{p_k\tau_k\} < \infty$. Thus $\sum_{k=0}^{\infty} E\{\tau_k^{1+s} \exp[\alpha s F(x^k)]\} < \infty$ for all sufficiently small $s > 0$. Since X is compact, $\exp[\alpha s F(x^k)]$ is bounded from

below by some positive constant, and so $\sum_{k=0}^{\infty} E\{\tau_k^{1+s}\} < \infty$ for all sufficiently small $s > 0$. But $0 < \tau_k \leq \bar{\tau}$ and hence s may be an arbitrary positive number, which completes the proof.

Remark 4. It is clear from the proof of the above lemma why the additional term $j_k \delta \tau_{k-1}$ has been inserted into the exponent in (6). Without it (with $\Delta_{\min} = 0$) one can only show that $\sum_{k=1}^{\infty} E\{\tau_{k-1}^s |\Delta x^k|^2\} < \infty$ for $s > 0$, but this is insufficient for convergence. One has to force faster convergence of $\{\tau_k\}$ towards zero in order to suppress oscillations around the minimum with small $|\Delta x^k|$.

From Lemma 3 we easily deduce the following results.

Lemma 4. $\sum_{k=0}^{\infty} E\{\tau_k^2 |\xi^k|^2\} < \infty$.

Proof. We have $|\xi^k|^2 \leq 2|g^k|^2 + 2|r^k|^2$. The series $\sum_{k=0}^{\infty} E\{\tau_k^2 |g^k|^2\}$ is convergent by Lemma 3 ($s = 1$) and by the boundedness of subgradients $\{g^k\}$ in the bounded set X. Next, it follows from (6) that $\tau_k \leq \tau_{k-1} \exp(\eta)$ and so

$$E\{\tau_k^2 |r^k|^2\} \leq \exp(2\eta) E\{\tau_{k-1}^2 |r^k|^2\}.$$

By (H3) there exists C such that $E_k |r^k|^2 \leq C$ for all k. Since τ_{k-1} is $\mathcal{F}_k$-measurable we get $E\{\tau_{k-1}^2 |r^k|^2\} \leq C E\{\tau_{k-1}^2\}$ and using again Lemma 3 obtain the required result.

Lemma 5. $\sum_{k=0}^{\infty} E\{\tau_k^2 |d^k|^2\} < \infty$ *and* $\sum_{k=1}^{\infty} E|\Delta x^k|^2 < \infty$.

Proof. It follows from (3) that for $k = 1, 2, \ldots$ one has

$$d^k - \frac{1}{1+\gamma_k}\xi^k = \frac{i_k \gamma_k}{1+\gamma_k}\left(d^{k-1} - \frac{1}{1+\gamma_{k-1}}\xi^{k-1}\right) + \frac{\gamma_k}{(1+\gamma_k)(1+\gamma_{k-1})} i_k \xi^{k-1}. \quad (13)$$

By (11), $i_k |\xi^{k-1}| \leq \sigma$ and thus for every k we have

$$\left| d^k - \frac{1}{1+\gamma_k}\xi^k \right| \leq \frac{\bar{\gamma}}{1+\bar{\gamma}} \left| d^{k-1} - \frac{1}{1+\gamma_{k-1}}\xi^{k-1} \right| + \frac{\bar{\gamma}\sigma}{1+\bar{\gamma}}.$$

Since $(1+\gamma_0)d^0 = \xi^0$, we get by induction

$$|d^k - \xi^k/(1+\gamma_k)| \leq \bar{\gamma}\sigma \quad (14)$$

for each k. Because $0 \leq \gamma_k \leq \bar{\gamma}$, our first assertion follows from Lemma 4. The second assertion is a simple corollary from the first one.

In the following two lemmas we prove that the rule (6) does not reduce stepsizes too rapidly.

Lemma 6. $\lim_{k\to\infty} \tau_{k-1}/\tau_k = 1$ *a.s. and* $\lim_{k\to\infty} (1 - \tau_{k-1}/\tau_k) r^k = 0$ *a.s.*

Proof. Consider the exponent in (6). From Lemmas 3 and 5 we see that $\tau_k \to 0$ a.s. and $\Delta x^k \to 0$ a.s. We shall prove that $\langle \xi^k, \Delta x^k \rangle \to 0$ a.s. We have $\langle \xi^k, \Delta x^k \rangle = \langle g^k, \Delta x^k \rangle + \langle r^k, \Delta x^k \rangle$. The sequence $\{g^k\}$ is bounded, hence $\langle g^k, \Delta x^k \rangle \to 0$ a.s. Next, by (H2), (H3) and Lemma 5 the series $\sum_{k=1}^{\infty} \langle r^k, \Delta x^k \rangle$ is a convergent martingale, and thus $\langle r^k, \Delta x^k \rangle \to 0$ a.s. Consequently, the exponent in (6) tends to 0 a.s. and $\tau_{k-1}/\tau_k \to 0$ a.s., as required. Moreover, we also see that there exists a random index m ($m < \infty$ a.s.) such that for all $k \geq m$ one has both $\tau_{k-1}/\tau_k = \exp(\alpha u_k + \delta \tau_{k-1})$ and $\alpha u_k + \delta \tau_{k-1} \leq 1$. Since $\exp(\cdot)$ is convex and increasing, the two preceding relations imply that for $k \geq m$ we have

$$|1 - \tau_{k-1}/\tau_k| \leq e(\alpha |u_k| + \delta \tau_{k-1}),$$

and thus

$$|1 - \tau_{k-1}/\tau_k| \, |r^k| \leq e(\alpha |\xi^k| \, |\Delta x^k| \, |r^k| + \alpha \lambda |\Delta x^k|^2 |r^k| + \delta \tau_{k-1} |r^k|). \tag{15}$$

We have $|\xi^k| \, |\Delta x^k| \, |r^k| \leq |\Delta x^k| \, |r^k|^2 + |g^k| \, |\Delta x^k| \, |r^k|$. By (H3) there is C such that $E_k |r^k|^4 \leq C$ for all k. Therefore $E\{|\Delta x^k|^2 |r^k|^4\} \leq CE|\Delta x^k|^2$. From Lemma 5 we deduce that $\sum_{k=1}^{\infty} E\{|\Delta x^k|^2 |r^k|^4\} < \infty$ which implies that $|\Delta x^k| \, |r^k|^2 \to 0$ a.s. In a similar fashion we treat the other components of the right-hand side of (15) and obtain the second assertion of the lemma. The proof is complete.

Lemma 7. $\sum_{k=0}^{\infty} \tau_k = \infty$ *a.s.*

Proof. From Lemma 3 we deduce that $\tau_k \to 0$ a.s. By Lemma 6, $\tau_k/\tau_{k-1} = \exp(-\alpha u_k - \delta \tau_{k-1})$ for $k \geq m$, where $m < \infty$ a.s. Therefore one must have $\sum_{k=1}^{\infty} (\alpha u_k + \delta \tau_{k-1}) = +\infty$ a.s. Consider the series $\sum_{k=1}^{\infty} u_k = \sum_{k=1}^{\infty} (\langle \xi^k, \Delta x^k \rangle + \lambda |\Delta x^k|^2)$. Since $\sum_{k=1}^{\infty} |\Delta x^k|^2 < \infty$ a.s. by Lemma 5, these components may be left out of account. Next, $\sum_{k=1}^{\infty} \langle \xi^k, \Delta x^k \rangle = \sum_{k=1}^{\infty} [\langle g^k, \Delta x^k \rangle + \langle r^k, \Delta x^k \rangle]$. The series $\sum_{k=1}^{\infty} \langle r^k, \Delta x^k \rangle$ is, by (H2), (H3) and Lemma 5, a convergent martingale and hence does not matter for $\sum_{k=1}^{\infty} (\alpha u_k + \delta \tau_{k-1})$ being infinite. Therefore

$$\sum_{k=1}^{\infty} (\alpha \langle g^k, \Delta x^k \rangle + \delta \tau_{k-1}) = +\infty \quad \text{a.s.} \tag{16}$$

By the compactness of X, there is $C_1 > 0$ such that $\langle g^k, \Delta x^k \rangle \leq C_1 |\Delta x^k|$ for all k. Hence, in view of (6) and (9),

$$\langle g^k, \Delta x^k \rangle \leq C_1 (1 + \bar{\gamma}) \tau_{k-1} |d^{k-1}| \leq C_1 (1 + \bar{\gamma}) \exp(\eta) \tau_{k-2} |d^{k-1}|.$$

Define the variables $\bar{d}_k = E_k |d^k|$. In view of (14), (H3) and the boundedness of X there exists C_3 such that $\bar{d}_k \leq C_3$ for all k. Thus we obtain the inequality

$$\langle g^k, \Delta x^k \rangle \leq C_4 \tau_{k-2} + C_5 \tau_{k-2} (|d^{k-1}| - \bar{d}_{k-1}),$$

where C_4 and C_5 are some constants. Since $E_{k-1}(|d^{k-1}| - \bar{d}_{k-1})^2 \leq$ const by (14) and (H3), the series $\sum_{k=2}^{\infty} \tau_{k-2} (|d^{k-1}| - \bar{d}_{k-1})$ is, in view of Lemma 3 ($s = 1$), a convergent

martingale. Therefore (16) implies that

$$\sum_{k=2}^{\infty} (\alpha C_4 \tau_{k-2} + \delta \tau_{k-1}) = +\infty \quad \text{a.s.},$$

which yields the required result.

Let us now pass on to the analysis of the aggregation coefficients $\{\gamma_k\}$ and the directions $\{d^k\}$.

Lemma 8. $\lim_{k\to\infty} \gamma_k = 0$ *a.s.*

Proof. Proceeding as in the proof of Lemma 6 we deduce that $v_k \to 0$ a.s. By Lemma 5, $|\Delta x^k| \to 0$ a.s., and thus $j_k = 1$ for all sufficiently large k. Suppose that $\limsup \gamma_k = \varepsilon > 0$. Let $\mathcal{K} \subset \mathcal{N}$ be such that $\gamma_k \to \varepsilon$ for $k \in \mathcal{K}$, $k \to \infty$. From (9) we then get

$$\varepsilon = \lim_{k\in\mathcal{K}} \gamma_k \leq \liminf_{k\in\mathcal{K}} \gamma_{k-1} \exp(-\kappa\gamma_{k-1}) \leq \limsup_{k\to\infty} \gamma_k \exp(-\kappa\gamma_k) < \varepsilon.$$

We have arrived at a contradiction which completes the proof.

Lemma 9. $\lim_{k\to\infty} [(1+\gamma_k)d^k - \xi_k] = 0$ *a.s.*

Proof. From (13) and (11) we obtain

$$\left| d^k - \frac{\xi^k}{1+\gamma_k} \right| \leq \gamma_k \left| d^{k-1} - \frac{\xi^{k-1}}{1+\gamma_{k-1}} \right| + \gamma_k \sigma, \quad k = 1, 2, \ldots .$$

This combined with Lemma 8 immediately yields the required result.

Before proceeding to the convergence analysis let us note that all the results of the above section remain valid for a nonconvex function F, too, provided an appropriate definition of a subdifferential may be introduced for them and the constant ν in (H1) is well defined, i.e. $\nu > -\infty$.

4. Convergence

Having established useful properties of stepsizes and aggregation coefficients we shall now prove that our method is convergent a.s.

Let us define

$$\mu = \inf_{x\in X} F(x), \qquad X^* = \{x \in X\colon F(x) = \mu\}.$$

We shall prove convergence in a rather standard way: by recursive inequality which we derive in the following lemma.

Lemma 10. *There exist sequences of scalar random variables* $\{S_k\}$ *and* $\{W_k\}$ *such that*

for any $x^* \in X^*$ *and any* $k \geq 1$ *one has*

$$|x^{k+1} - x^*|^2 \leq |x^k - x^*|^2 + 2\tau_k(\mu - F(x^k) + S_k) + W_k, \tag{17}$$

where

$$\lim_{k\to\infty} S_k = 0 \quad a.s., \tag{18}$$

and

$$\sum_{k=1}^{\infty} W_k < \infty \quad a.s., \tag{19}$$

Proof. Define the coefficients

$$t_k = \begin{cases} \tau_k(1+\gamma_k) & \text{if } \tau_k(1+\gamma_k)|d^k| \leq t, \\ t/|d^k| & \text{if } \tau_k(1+\gamma_k)|d^k| > t. \end{cases}$$

From (4) one obtains

$$|x^{k+1} - x^*|^2 \leq |x^k - x^*|^2 + 2t_k\langle d^k, x^* - x^k\rangle + t_k^2|d^k|^2. \tag{20}$$

We have

$$\begin{aligned} t_k d^k &= \tau_k(1+\gamma_k)d^k + (t_k - \tau_k(1+\gamma_k))d^k \\ &= \tau_k g^k + \tau_k r^k + \tau_k((1+\gamma_k)d^k - \xi^k) + (t_k - \tau_k(1+\gamma_k))d^k \\ &= \tau_k g^k + \tau_{k-1} r^k + \tau_k\left((1+\gamma_k)d^k - \xi^k + \left(1 - \frac{\tau_{k-1}}{\tau_k}\right)r^k\right) \\ &\quad + (t_k - \tau_k(1+\gamma_k))d^k. \end{aligned}$$

Using this identity in (20) we get

$$|x^{k+1} - x^*|^2 \leq |x^k - x^*|^2 + 2\tau_k(\langle g^k, x^* - x^k\rangle + S_k) + W_k, \tag{21}$$

where

$$S_k = \left\langle (1+\gamma_k)d^k - \xi^k + \left(1 - \frac{\tau_{k-1}}{\tau_k}\right)r^k, x^* - x^k \right\rangle$$

and

$$W_k = \tau_{k-1}\langle r^k, x^* - x^k\rangle + (t_k - \tau_k(1+\gamma_k))\langle d^k, x^* - x^k\rangle + t_k^2|d^k|^2.$$

In view of the subgradient inequality $\langle g^k, x^* - x^k\rangle \leq F(x^*) - F(x^k)$, (21) may be rewritten in the required form (17). Directly from Lemmas 6 and 9 we see that $S_k \to 0$, i.e. (18) is true. Consider the series $\sum_{k=1}^{\infty} W_k$. By (H2), (H3) and Lemma 3 the series $\sum_{k=1}^{\infty} \tau_{k-1}\langle r^k, x^* - x^k\rangle$ is a convergent martingale. Next, the series $\sum_{k=1}^{\infty}(t_k - \tau_k(1+\gamma_k))\langle d^k, x^* - x^k\rangle$ is convergent too, since $t_k = \tau(1+\gamma_k)$ for all

sufficiently large k. Finally, the series $\sum_{k=1}^{\infty} t_k^2 |d^k|^2$ is convergent a.s. by Lemma 5, and thus (19) holds. The proof is complete.

We are now ready to prove our main result.

Theorem 1. *Assume* (H1) *to* (H3). *Then almost surely all accumulation points of the sequence* $\{x^k\}$ *generated by the method* (3)–(11) *belong to* X^*.

Proof. Let Ω be the sample space on which the process $\{x^k\}$ is defined and let Ω_0 be the union of null sets excluded in Lemmas 7 and 10. Let us fix an event $\omega \notin \Omega_0$ and consider the path $\{x^k(\omega)\}$. Henceforth we shall for brevity omit the argument ω.

Let us use Lemma 10 and consider two cases.

Case 1: there exists $k_0 \geq 0$ such that $F(x^k) > \mu + S_k$ for all $k \geq k_0$.

Suppose that there exist $\varepsilon > 0$ and $k_1 \geq 0$ such that $F(x^k) \geq \mu + S_k + \varepsilon$ for all $k \geq k_1$. Then for every $x^* \in X^*$ and all $l \geq k \geq k_1$ inequality (17) yields

$$|x^l - x^*|^2 \leq |x^k - x^*|^2 - 2\varepsilon \sum_{i=k}^{l-1} \tau_i + \sum_{i=k}^{l-1} W_i.$$

Letting l approach infinity we obtain a contradiction with (19) and Lemma 7. Therefore, in view of (18), one can extract from $\{x^k\}$ and infinite subsequence $\{x^k\}_{k \in \mathcal{K}}$ convergent to some $\hat{x} \in X^*$. Define for every sufficiently large l an index $k(l) = \max\{k \in \mathcal{K}: k < l\}$. Since $\mathcal{K}$ is infinite, $k(l) \geq k_0$ for all sufficiently large l. Then (17) with $x^* = \hat{x}$ gives

$$|x^l - \hat{x}|^2 \leq |x^{k(l)} - \hat{x}|^2 + \sum_{i=k(l)}^{l-1} W_i.$$

If $l \to \infty$ then $k(l) \to \infty$, hence $x^{k(l)} \to \hat{x}$ and, in view of (19), $\sum_{i=k(l)}^{l-1} W_i \to 0$. Therefore $\hat{x}$ is the only accumulation point of $\{x^k\}$.

Case 2: inequality $F(x^k) \leq \mu + S_k$ is fulfilled infinitely often.

Suppose that there exists a subsequence $\{x^l\}_{l \in \mathcal{L}}$ convergent to some $x' \notin X^*$. Define $k(l) = \max\{k: k < l \text{ and } F(x^k) \leq \mu + S_k\}$. By hypothesis, $k(l)$ is well-defined for all sufficiently large l, and $k(l) \to \infty$ as $l \to \infty$. Extracting a sub-subsequence of $\{x^l\}_{l \in \mathcal{L}}$ if necessary, without loss of generality we may suppose that $\{x^{k(l)}\}_{l \in \mathcal{L}}$ converges to some $\hat{x} \in X$. Owing to the definition of $k(l)$, one has $F(\hat{x}) = \mu$, i.e. $\hat{x} \in X^*$. From (17) for $x^* = \hat{x}$ we obtain

$$|x^l - \hat{x}|^2 \leq |x^{k(l)} - \hat{x}|^2 + 2 \sum_{i=k(l)}^{l-1} \tau_i (\mu - F(x^i) + S_i) + \sum_{i=k(l)}^{l-1} W_i.$$

By construction, $F(x^i) > \mu + S_i$ for $i = k(l)+1, \ldots, l-1$, and so

$$|x^l - \hat{x}|^2 \leq |x^{k(l)} - \hat{x}|^2 + 2\tau_{k(l)} (\mu - F(x^{k(l)}) + S_{k(l)}) + \sum_{i=k(l)}^{l-1} W_i.$$

The right-hand side of the above inequality tends to zero as $l \to \infty$, whereas the left-hand side converges to $|x' - \hat{x}|^2 > 0$. We have arrived at a contradiction, hence there are no cluster points outside X^*. The proof is complete.

In a similar way one can obtain convergence results for many other classes of problems, e.g. for smooth, unconstrained, but not necessarily convex problems [16].

5. Some asymptotic properties

Although our aim is to speed up convergence far from the solution, it could be interesting to verify whether our stepsize rule changes asymptotic properties of the method, when compared with classical approaches (cf. [9, 12]). Clearly, the crucial question here is the asymptotic behavior of the sequence $\{\tau_k\}$. We pursue this question in the following theorem.

Theorem 2. *Assume* (H1) *to* (H3). *Additionally suppose that F is twice continuously differentiable in an open set containing X. Then*

$$\lim_{k \to \infty} (k+1)\tau_k = \frac{1}{\delta} \quad a.s.$$

Proof. Define the variables $z_k = \ln[(k+1)\tau_k]$, $k = 1, 2, \ldots$. By Lemma 6 for sufficiently large k one has

$$z_k = z_{k-1} + \ln\left(1 + \frac{1}{k}\right) - \alpha u_k - \delta \tau_{k-1}.$$

This combined with (7) and (5) yields

$$z_k = z_{k-1} + \frac{1}{k}(1 - \delta \exp(z_{k-1})) + \ln\left(1 + \frac{1}{k}\right) - \frac{1}{k}$$
$$- \alpha(\langle r^k, \Delta x^k \rangle + \langle g^k, \Delta x^k \rangle + \lambda |\Delta x^k|^2). \tag{22}$$

Obviously, $\sum_{k=1}^{\infty} [\ln(1+1/k) - 1/k]$ is convergent. From (H2), (H3) and Lemma 5, via the martingale convergence theorem, we deduce that the series $\sum_{k=1}^{\infty} \alpha \langle r^k, \Delta x^k \rangle$ is convergent a.s. Directly from Lemma 5 we see that $\sum_{k=1}^{\infty} \alpha\lambda |\Delta x^k|^2 < \infty$ a.s. Next, expansion of F gives $\langle g^k, \Delta x^k \rangle = F(x^k) - F(x^{k-1}) + \frac{1}{2}\theta_k |\Delta x^k|^2$ where θ_k is uniformly bounded for all k. Since $F(x^k) \to \mu$ a.s. by Theorem 1 and $\sum_{k=1}^{\infty} \theta_k |\Delta x^k|^2$ is convergent a.s. by Lemma 5, the series $\sum_{k=1}^{\infty} \langle g^k, \Delta x^k \rangle$ is convergent, too. Therefore one may rewrite (22) as follows

$$z_k = z_{k-1} + \frac{1}{k}(1 - \delta \exp(z_{k-1})) + W_k,$$

where the series $\sum_{k=1}^{\infty} W_k$ is convergent a.s. The above formula is a simple Robbins-

Monro algorithm for solving the equation $1-\delta\exp(z)=0$. By virtue of standard results on convergence of stochastic approximation algorithms (cf. [9, 12]) we conclude that $z_k \to -\ln\delta$ a.s., as required.

It follows from the above theorem that for sufficiently large k one can write

$$\tau_k = \frac{1}{\delta(k+1)} + o\left(\frac{1}{k+1}\right).$$

Hence one can apply (with some modifications) the techniques of [9] or [12] to obtain some results on the rate of convergence of our method in the smooth case. Since this requires a number of additional assumptions and the way of analysis is well-known, we shall go no further into the matter confining ourselves within the observation that $\{\tau_k\}$ behaves in the neighborhood of the solution as the sequence acknowledged to be the best one (for the smooth unconstrained case) among sequences of the form $\{1/(k+1)^\sigma\}$, $\sigma \leq 1$.

It should be stressed, however, that even in the differentiable case the asymptotic behavior of $\{\tau_k\}$ as $\{1/\delta(k+1)\}$ is not a prepared cause of $\{x^k\}$ approaching a neighborhood of the minimum, since we see from (2) that $\{z_k\}$ converges to $-\ln\delta$ owing to $u_k \to 0$. Besides, the term $\delta\tau_{k-1}$ activates only for small $|\Delta x^k|$ (see (6), (8)). So, the asymptotic properties of $\{\tau_k\}$ and the convergence of $\{x^k\}$ are mutually related.

In the nonsmooth case, which is our main concern in this paper, the assertion of Theorem 2 is doubtful, because the terms $\langle g^k, \Delta x^k\rangle$ in (22) need not be summable and may be significant as compared with $\delta\tau_{k-1}$, thus causing irregular fluctuations of $\{z_k\}$.

6. Modifications of the method

The basic model (3)-(11) may be modified in various ways so as to improve its practical efficiency and preserve theoretical convergence properties.

Crucial from the practical point of view are the values of parameters α, β, δ and κ in (6) and (9). With constant values of these parameters, there is a danger of rapid changes of stepsizes and aggregation coefficients due to a wide range of changes of stochastic subgradients ξ^k. It would be convenient in practice to ensure for $k=1, 2, \ldots$ the fulfillment of the conditions

$$\left|\ln\frac{\tau_k}{\tau_{k-1}}\right| \leq \eta \quad \text{and} \quad \left|\ln\frac{\gamma_k}{\gamma_{k-1}}\right| \leq \eta \tag{23}$$

with a fixed $\eta > 0$. To this end one can replace α, β, δ and κ in (6) and (9) with varying coefficients α_k, β_k, δ_k and κ_k, provided that the following conditions are satisfied:

(i) for $k=0, 1, 2, \ldots$ the coefficients α_k, β_k, δ_k and κ_k are $\mathcal{F}_{k+1}$-measurable;

(ii) there exist $0<\underline{\alpha}\leq\bar{\alpha}$ and $0<\underline{\delta}\leq\bar{\delta}$ such that, for $k=0,1,2,\ldots$, $\underline{\alpha}\leq\alpha_k\leq\bar{\alpha}$, $\underline{\alpha}\leq\beta_k\leq\bar{\alpha}$, $\underline{\delta}\leq\delta_k\leq\bar{\delta}$ and $\underline{\delta}\leq\kappa_k\leq\bar{\delta}$;

(iii) there exist positive constants b, C_1 and C_2 such that $b<\lambda+\nu$ and for each $k\geq 2$ one has

$$\sum_{i=2}^{k}\left[\left(\frac{1}{\alpha_{i-1}}-\frac{1}{\alpha_i}\right)\ln\tau_{i-1}+b|\Delta x^i|^2\right]\geq C_1+C_2\ln\tau_k.$$

Under these conditions and (H1)-(H3) the algorithm (3)-(11) remains convergent, i.e. Theorem 1 is still true. The way of proving it is in general similar to that of Sections 3 and 4. The only major difference is in the proof that $\sum_{k=0}^{\infty}E\{\tau_k^{1+s}\}<\infty$ for all $s>0$ (Lemma 3). Let us observe that (5), (7) and (H1) imply that for each $k\geq 1$

$$\tau_k\leq\tau_{k-1}\exp[-\alpha_k(F(x^k)-F(x^{k-1})+(\lambda+\nu)|\Delta x^k|^2+\langle r^k,\Delta x^k\rangle)-j_k\delta_k\tau_{k-1}].$$

Let us denote

$$p_k=e^{F(x^k)}\tau_k^{1/\alpha_k}\prod_{i=2}^{k}\tau_{i-1}^{1/\alpha_{i-1}-1/\alpha_i}\,e^{b|\Delta x^i|^2},\quad k=2,3,\ldots.$$

Proceeding exactly like in Lemmas 1-3 we demonstrate that

$$p_k\leq p_{k-1}\exp[-(\lambda+\nu-b)|\Delta x^k|^2-\langle r^k,\Delta x^k\rangle-j_{k\underline{\delta}}\tau_{k-1}/\bar{\alpha}]$$

and then

$$\sum_{k=2}^{\infty}E\{\tau_k p_k^s\}<\infty$$

for sufficiently small $s>0$. From condition (iii) we see that

$$p_k\geq\tau_k^{C_2+1/\alpha_k}\,e^{F(x^k)+C_1}\geq C_3\tau_k^{C_4}$$

for some $C_3>0$ and $C_4>0$. Hence

$$\sum_{k=0}^{\infty}E\{\tau_k^{1+s}\}<\infty,$$

as required. Differences in other proofs are only of technical nature.

Let us return to the construction of a flexible practical algorithm. Condition (iii), although sufficient for convergence, cannot be used directly for determining gains $\{\alpha_k\}$, since the right-hand side of this inequality does not necessarily decrease, if k increases. However, if we ensure that $\tau_k\leq a\min_{0\leq i\leq k-1}\tau_i$ for some $a>1$, a more convenient sufficient condition for (iii) may be derived. To this end define $\underline{\tau}_k=\min_{0\leq i\leq k}\tau_i$, $\bar{\tau}_k=\max_{0\leq i\leq k}\tau_i$, replace $\bar{\tau}$ in (6) with $a\underline{\tau}_{k-1}$ and assume that the gains $\{\alpha_k\}$ satisfy the condition

$$\sum_{i=2}^{k}\left[\left(\frac{1}{\alpha_{i-1}}-\frac{1}{\alpha_i}\right)(\ln\tau_{i-1}-\ln\bar{\tau}_{i-1})+b|\Delta x^i|^2\right]\geq\left(\frac{1}{\underline{\alpha}_k}-\frac{1}{\bar{\alpha}_k}\right)(\ln\underline{\tau}_{k-1}-\ln\bar{\tau}_{k-1}),\tag{24}$$

where $\underline{\alpha}_k = \min_{0 \leq i \leq k} \alpha_i$ and $\bar{\alpha}_k = \max_{0 \leq i \leq k} \alpha_i$. By Abel's criterion and the fact that $\tau_k / \tau_{k-1} \leq \exp(\eta)$, condition (24) implies (iii).

The above ideas are the basis of the following useful modification of the rules (6) and (9):

$$\tau_k = \min\{a\underline{\tau}_{k-1}, \tau_{k-1} \exp(-\alpha_k u_k - j_k \delta_k \tau_{k-1}), \quad k = 1, 2, \ldots, \tag{25}$$

$$\gamma_k = \min\{a\underline{\gamma}_{k-1}, \gamma_{k-1} \exp(-\beta_k v_k - j_k \kappa_k \gamma_{k-1}), \quad k = 2, 3, \ldots, \tag{26}$$

where $\delta_k = \min\{\delta, \eta/2\tau_{k-1}\}$, $\kappa_k = \min\{\kappa, \eta/2\gamma_{k-1}\}$, $\underline{\gamma}_{k-1} = \min_{0 \leq i \leq k-1} \gamma$, $a > 1$, $\delta > 0$, $\kappa > 0$ and $\eta > 0$. The gains $\{\alpha_k\}$ and $\{\beta_k\}$ are determined according to the following procedure. For each k we define auxiliary quantities

$$L_k = \left(|\xi^k|^2 - \frac{\langle \xi^k, x^k - y^k \rangle^2}{|x^k - y^k|^2} \right)^{1/2}$$

where

$$y^k = x^{k-1} - \min\left\{ \tau_{k-1}(1 + \gamma_{k-1}), \frac{t}{|d^{k-1}|} \right\} d^{k-1}.$$

If $y^k \neq x^k$ then L_k is the length of the projection of ξ^k on the hyperplane supporting X at x^k; if $y^k = x^k$ then $L_k = |\xi^k|$. We define coefficients $\{q_k\}$ and $\{s_k\}$ recursively by:

$$q_{k+1} = q_k/(1 - q + q_k), \quad k = 2, 3, \ldots, \quad q_2 = 1,$$

$$s_{k+1} = s_k/[1 - i_{k-1}(q - s_k)], \quad k = 3, 4, \ldots, \quad s_3 = 1,$$

where $0 < q \leq 1$. They are used to calculate scaling factors $\{U_k\}$ and $\{V_k\}$ according to the formulae

$$U_{k+1} = U_k + q_k(L_k|\Delta x^k| + \lambda|\Delta x^k|^2 - U_k), \quad k = 2, 3, \ldots, \tag{27}$$

$$V_{k+1} = V_k + s_k i_{k-1}(L_k|\Delta x^{k-1}| + \lambda|\Delta x^k|\,|\Delta x^{k-1}| - V_k), \quad k = 2, 3, \ldots, \tag{28}$$

where $U_1 = U_2 = L_1|\Delta x^1| + \lambda|\Delta x^1|^2$, $V_2 = V_3 = L_2|\Delta x^1| + \lambda|\Delta x^1|\,|\Delta x^2|$. We also compute recursively auxiliary quantities $\{W_k\}$ and $\{c_k\}$ by

$$W_{k+1} = W_k + b|\Delta x^k|^2 + \left(\frac{1}{\alpha_{k-1}} - \frac{1}{\alpha_k} \right)(\ln \tau_{k-1} - \ln \bar{\tau}_{k-1}), \quad k = 2, 3, \ldots$$

where $W_1 = W_2 = 0$, and

$$c_k = \max\{1/\bar{\alpha}, U_k/h, |u_k|/(\eta - j_k \delta_k \tau_{k-1})\}, \tag{29}$$

with $h>0$. Finally, we define

$$\frac{1}{\alpha_k}=\begin{cases} c_k \quad \text{if } W_k+b|\Delta x^k|^2+(1/\alpha_{k-1}-c_k)(\ln\tau_{k-1}-\ln\bar{\tau}_{k-1}) \\ \qquad \geq (1/\underline{\alpha}_{k-1}-c_k)(\ln\underline{\tau}_{k-1}-\ln\bar{\tau}_{k-1}), \\ \max\left\{c_k, \dfrac{1}{\alpha_{k-1}}\right. \\ \qquad \left. +\dfrac{W_k+b|\Delta x^k|^2+(1/\underline{\alpha}_{k-1}-1/\bar{\alpha}_{k-1})(\ln\bar{\tau}_{k-1}-\ln\underline{\tau}_{k-1})}{\ln\tau_{k-1}-\ln\bar{\tau}_{k-1}}\right\} \\ \qquad \text{otherwise,} \end{cases} \tag{30}$$

and

$$\frac{1}{\beta_k}=\max\left\{\frac{1}{\bar{\alpha}}, \frac{V_k}{h}, \frac{|v_k|}{\eta-j_k\kappa_k\gamma_{k-1}}\right\}. \tag{31}$$

Typically, we shall have in (29)-(31) $c_k=U_k/h$ and $\alpha_k=h/U_k$, $\beta_k=h/V_k$, where h is a certain gain and U_k and V_k are normalizing factors for u_k and v_k, calculated by (27), (28). However, if these values of α_k and β_k violate (23), (i), (ii), or (24), other parts of (29)-(31) are used, which ensure that all these conditions are satisfied.

Following the main line of argument presented in this paper one can still prove that for a bounded noise this modified version of the method is convergent a.s.

In the practical algorithm the reset rule is specified as follows: if $\alpha_k<\bar{\alpha}$, $\beta_k<\bar{\alpha}$, $v_k>0$ and the new direction $\tilde{d}^k$ calculated by (3) with $i_k=1$ satisfies $\langle\tilde{d}^k, \xi^k\rangle>0$ (i.e. ξ^k would disagree with two successive directions), we set $i_k=0$, $d_k=\xi^k/(1+\gamma_k)$; otherwise, we keep $i_k=1$ and continue averaging.

7. Examples

The algorithm described in this paper was implemented in FORTRAN and tested successfully on manifold stochastic programming problems. Owing to the on-line rules for adjusting stepsizes and gains, the final version of sec. 6 turned out to be insensitive to many parameters, and worked properly for the following values: $t=a=\sigma=\bar{\alpha}=10^{10}$ (these can be arbitrary large numbers), $\lambda=0$, $\delta=\underline{\alpha}=10^{-10}$ (these can be arbitrary sufficiently small numbers), $q=0.1\div 0.3$, $h=0.05$, $\eta=0.3$.

To gain an insight into the numerical properties of the method let us consider two simple computational examples.

Example 3 (a two-stage problem). Consider the problem of Example 1 with

$$f_0(x, y, \omega)=(x_1)^2+3(x_2)^2+10y_1+10y_2,$$

$$Y(x, \omega)=\{y\in\mathbb{R}^2: y\geq 0, 2y_1+y_2\geq x_1+x_2+\omega\}, \qquad X=\mathbb{R}^2.$$

Here ω is a discrete random variable with the following distribution: $P\{\omega = 1\} = P\{\omega = -1\} = 0.5$. After simple calculations we see that

$$F(x) = E\left\{\min_{y \in Y(x,\omega)} f_0(x, y, \omega)\right\} = (x_1)^2 + 3(x_2)^2 + 5E\{\max(0, x_1 + x_2 + \omega)\}$$

and the minimum of F is attained at $\hat{x} = (-0.75, -0.25)$. At this point F is nondifferentiable and

$$\partial F(\hat{x}) = \text{co}\{(-1.5, -1.5), (1, 1)\},$$

i.e. F is 'essentially gully' and is difficult to minimize even for deterministic methods.

For the purpose of testing our method, the analytical form of F was assumed unknown and only stochastic subgradients ξ^k, calculated in a way indicated in Example 1, were used. The results of computations are shown in Table 1. The differences $F(x^k) - F(\hat{x})$, not used in the computation, were calculated analytically 'a posteriori'.

Table 1
Results for Example 3

k	τ_k	γ_k	x_1^k	x_2^k	$F(x^k) - F(\hat{x})$
0	2.0×10^{-1}	10	5.000	5.000	149.2
10	1.5×10^{-1}	8.5	0.5260	−0.0566	3.210
20	1.4×10^{-1}	7.8	−0.9657	0.2543	1.098
30	1.0×10^{-1}	6.3	−0.5095	−0.2866	0.2657
40	7.7×10^{-2}	5.3	−0.5804	−0.0011	0.6386
50	5.5×10^{-2}	3.2	−0.1970	−0.2610	0.8482
60	4.4×10^{-2}	3.1	−0.6675	−0.1660	0.1944
70	3.4×10^{-2}	2.7	−0.6427	−0.1188	0.3017
80	4.1×10^{-2}	2.0	−0.6674	−0.3282	0.2953×10^{-1}
90	3.6×10^{-2}	1.7	−0.3741	−0.1883	0.5903
100	2.5×10^{-2}	2.0	−0.7747	−0.2861	0.9576×10^{-1}
150	1.3×10^{-2}	9.9×10^{-1}	−0.6008	−0.1371	0.3226
200	6.1×10^{-3}	6.5×10^{-1}	−0.7226	−0.2424	0.3589×10^{-1}
300	1.1×10^{-3}	1.6×10^{-1}	−0.7433	−0.2487	0.8026×10^{-2}
400	5.6×10^{-4}	4.6×10^{-2}	−0.7457	−0.2496	0.4712×10^{-2}
500	1.7×10^{-4}	3.8×10^{-2}	−0.7484	−0.2519	0.5288×10^{-3}
1000	9.3×10^{-8}	4.1×10^{-5}	−0.7483	−0.2517	0.1146×10^{-4}

Example 4 (production yield optimization). Let $x \in \mathbb{R}^n$ be the nominal values of elements in an electronic circuit and let ω represent random deviations of real values from the nominal ones. The problem consists in maximizing the production yield

$$Y(x) = P\{x - \omega \in A\}$$

where $A \subset \mathbb{R}^n$ is a set of acceptable values of elements. Defining

$$f(x - \omega) = \begin{cases} 1 & \text{if } x - \omega \notin A, \\ 0 & \text{if } x - \omega \in A \end{cases}$$

Table 2
Results for Example 4

k	τ_k	γ_k	x_1^k	x_2^k	$Y(x^k)$
0	1.5×10^{-3}	1.0	1.300	1.300	0.12
10	2.0×10^{-3}	1.3	1.153	1.387	0.27
20	2.3×10^{-3}	1.2	1.193	1.338	0.35
30	2.3×10^{-3}	1.1	1.135	1.325	0.41
40	2.4×10^{-3}	1.0	1.113	1.230	0.62
50	2.7×10^{-3}	1.1	0.9968	1.159	0.65
60	3.6×10^{-3}	1.7	1.004	0.9595	0.84
70	2.7×10^{-3}	1.6	0.9649	0.8904	0.79
80	2.6×10^{-3}	1.5	1.049	0.9177	0.63
90	1.6×10^{-3}	1.3	1.067	0.9591	0.67
100	1.5×10^{-3}	1.2	0.9680	0.9156	0.83
200	2.2×10^{-3}	0.67	1.028	1.068	0.86
300	1.3×10^{-3}	0.43	0.9916	1.018	0.89
400	4.4×10^{-4}	0.20	0.9920	1.033	0.88
500	3.6×10^{-4}	0.31	0.9970	1.033	0.88
1000	5.7×10^{-5}	0.05	0.9932	1.013	0.89

one can rewrite our problem as

$$\min_{x\in\mathbf{R}^n}\left[F(x)=Ef(x-\omega)=\int f(x-\omega)p(\omega)\,\mathrm{d}\omega\right],$$

where $p(\omega)$ is the density of the distribution of ω. Using rules for differentiating convolutions one immediately gets the formula for generating a stochastic gradient of F at a point x^k:

$$\xi^k=\frac{f(x^k-\omega^k)p'(\omega^k)}{p(\omega^k)}\quad(\omega^k-\text{sampled}),$$

which in case of a normal distribution takes on the form

$$\xi^k=-D^{-1}\omega^k f(x^k-\omega^k),$$

where D is the covariance matrix of ω.

As a simple computational example consider a voltage divider, in which $x\in\mathbb{R}^2$, $A=\{y\in\mathbb{R}^2\colon 0\leq y_1\leq 1.2,\ 0\leq y_2\leq 1.3,\ 0.45\leq y_2/(y_1+y_2)\leq 0.55\}$, and ω is a two-dimensional Gaussian variable with $E\omega=0$, $E\omega_1^2=0.01$, $E\omega_2^2=0.04$ and $E\omega_1\omega_2=0.014$. The results of computations are shown in Table 2. The yield $Y(x^k)$ was computed 'a posteriori', on the basis of 10 000 simulations, to gain an insight into the progress of the algorithm.

8. Conclusions

The algorithm described in this paper differs from the basic stochastic subgradient method in two ways.

First, averaging of stochastic subgradients observed at successive points is used. It is organized in a different manner than in earlier works (cf. [4, 5, 7, 15]), since

we use adaptive aggregation coefficients and reset tests to avoid excessive inertia of the method.

Secondly, on-line rules for determining stepsizes and aggregation coefficients are introduced. The rules are derived from the concept of regularized improvement functions and increase adaptive properties of the method, while retaining its theoretical convergence properties. It seems that in a similar way one can modify other theoretical models of stochastic algorithms thus broadening the area of their practical applications.

Finally, it is worth noting that without the term $\delta\tau_{k-1}$ in (6) (see Remark 4) one can still prove convergence a.s. of the sequence $\bar{x}^k = (\sum_{i=0}^{k} \tau_i x^i)/\sum_{i=0}^{k} \tau_i$ (see [17]).

References

[1] Yu.M. Ermol'ev, "On a method of generalized stochastic gradients and stochastic quasi-Fejer sequences", *Kibernetika* 2 (1969) 73-83 [in Russian]. *Cybernetics* 5 (1969) 208-220.

[2] Yu.M. Ermol'ev, *Stochastic programming methods* (Nauka, Moscow, 1976) [in Russian].

[3] Yu.M. Ermol'ev and N.Z. Shor, "Method of random walk for the two-stage problem of stochastic programming and its generalization", *Kibernetika* 1 (1968) 90-92 [in Russian]. *Cybernetics* 4 (1968) 59-60.

[4] A.M. Gupal, *Stochastic methods for solving nonsmooth extremal problems* (Naukova Dumka, Kiev, 1979) [in Russian].

[5] A.M. Gupal and L.T. Bazhenov, "A stochastic analog of the conjugate gradient method", *Kibernetika* 1 (1972) 124-126 [in Russian]. *Cybernetics* 8 (1972) 138-140.

[6] J.B. Hiriart-Urruty, "Conditions nécessaires d'optimalité pour un programme stochastique avec recours", *SIAM Journal on Control and Optimization* 16 (1978) 317-329.

[7] A.P. Korostelev, "On multi-step procedures of stochastic optimization" *Avtomatika i Telemekhanika* 5 (1981) 82-90 [in Russian]. *Automation and Remote Control* 42 (1981) 621-628.

[8] K. Kiwiel, "An aggregate subgradient method for nonsmooth convex minimization", *Mathematical Programming* 27 (1983) 320-341.

[9] H.J. Kushner and D.S. Clark, *Stochastic approximation methods for constrained and unconstrained systems* (Springer-Verlag, Berlin, New York, 1978).

[10] H.J. Kushner and Hai-Huang, "Asymptotic properties of stochastic approximations with constant coefficients", *SIAM Journal on Control and Optimization* 19 (1981) 87-105.

[11] F. Mirzoakhmedov and S.P. Uryas'ev, "Adaptive stepsize regulation for a stochastic optimization algorithm", *Žurnal Vičislitelnoi Matematiki i Matematičeskoi Fiziki* 23 (1983) 1314-1325 [in Russian].

[12] M.B. Nevelson and B.Z. Khasminski, *Stochastic approximation and recursive estimation* (Nauka, Moscow, 1972) [in Russian].

[13] E.A. Nurminski, *Numerical methods for solving deterministic and stochastic minimax problems* (Naukova Dumka, Kiev, 1979) [in Russian].

[14] R.T. Rockafellar and R.J.B. Wets, "Stochastic convex programming: Relatively complete recourse and induced feasibility", *SIAM Journal on Control and Optimization* 14 (1976) 574-589.

[15] A. Ruszczyński and W. Syski, "Stochastic approximation method with gradient averaging for unconstrained problems", *IEEE Transactions on Automatic Control* AC-28 (1983) 1097-1105.

[16] A. Ruszczyński and W. Syski, "Stochastic approximation algorithm with gradient averaging and on-line stepsize rules", contributed paper, Eighth World Congress of the IFAC, Budapest, 1984.

[17] A. Ruszczyński and W. Syski, "On convergence of the stochastic subgradient method with on-line stepsize rules", Technical Report R.I.21/84, Instytut Automatyki, Politechnika Warszawska, Warszawa, 1984, to appear in *Journal on Mathematical Analysis and Applications.*

[18] S.P. Uryas'ev, "Step control for direct stochastic programming methods", *Kibernetika* 6 (1980) 85-87 [in Russian]. *Cybernetics* 16 (1980) 886-890.

Mathematical Programming Study 28 (1986) 132-156
North-Holland

COMPUTATION OF DESCENT DIRECTIONS AND EFFICIENT POINTS IN STOCHASTIC OPTIMIZATION PROBLEMS WITHOUT USING DERIVATIVES

K. MARTI and E. FUCHS

Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Werner-Heisenberg-Weg 39, D-8014 Neubiberg/München, FR Germany

Received 3 January 1984
Revised 23 April 1985

Searching for descent directions in stochastic optimization problems having an objective function F of the type $F(x) = Eu(A(\omega)x - b(\omega))$, $x \in \mathbb{R}^n$, where u is a convex loss function, $(A(\omega), b(\omega))$ is a random $m \times (n+1)$ matrix and E denotes the expectation operator, the main difficulty is the fact that the mean value function F is represented in general by a multiple integral. Hence, the derivatives of F may be obtained only with a big computational effort.

Under the weak assumption that F is not constant on line segments of the feasible domain D, for finding feasible descent directions $h = y - x$, $y \in D$, for F at $x \in D$, a derivative-free method is presented which is based on certain relations between the random m-vectors $A(\omega)x - b(\omega)$ and $A(\omega)y - b(\omega)$. Furthermore, introducing the concept of efficient points for stochastic optimization problems, necessary optimality conditions not involving any derivatives of F are obtained.

Key words: Stochastic Programming, Feasible Descent Directions, Efficient Points, Necessary Optimality Conditions, Stochastic Dominance, Derivative-Free Methods.

1. Introduction

Let $(\Omega, \mathfrak{A}, \mu)$ be a probability space and $(A(\omega), b(\omega))$ a random $m \times (n+1)$ matrix. By P we denote the distribution of $(A(\omega), b(\omega))$ and $P_{A(\cdot)x - b(\cdot)}$ is the distribution of the random m-vector $A(\omega)x - b(\omega)$, where x is a vector of $\mathbb{R}^n$.

Let $u: \mathbb{R}^m \to \mathbb{R}$ be a given (loss) function such that the expectation $Eu(A(\omega)x - b(\omega))$ exists for every $x \in \mathbb{R}^n$ and define the mean value function $F: \mathbb{R}^n \to \mathbb{R}$ by

$$F(x) = Eu(A(\omega)x - b(\omega)), \quad x \in \mathbb{R}^n. \tag{1}$$

If $\hat{x} = \binom{x}{-1}$ for $x \in \mathbb{R}^n$ and $H_x: \mathbb{R}^{m(n+1)} \to \mathbb{R}^m$ is defined by

$$H_x(M) = M\hat{x}, \; M \in \mathbb{R}^{m(n+1)}, \quad x \in \mathbb{R}^n,$$

then $F(x)$ may also be written in the form

$$F(x) = \int u(A(\omega)x - b(\omega))\mu(\mathrm{d}\omega) = \int_{\cdot} (u \circ H_x)(A, b) P(\mathrm{d}A, \mathrm{d}b)$$

$$= \int u(z) P_{A(\cdot)x - b(\cdot)}(\mathrm{d}z). \tag{2}$$

Supported by the Deutsche Forschungsgemeinschaft.

Remark. In this paper the elements M of $\mathbb{R}^{m(n+1)}$ are $m \times (n+1)$ matrices. Furthermore, vec M is defined by vec $M = (M_1, M_2, \ldots, M_m)$, where M_i is the ith row of M.

The problem to be considered in the following is

$$\text{minimize } F(x) \quad \text{s.t. } x \in D, \tag{3}$$

where D is a convex subset of $\mathbb{R}^n$.

Problems of the form (3) arise in many different connections:

(a) In stochastic linear programming with recourse [11, 19] the objective function F has the special form

$$F(x) = \tilde{c}'x + Ep(\tilde{A}(\omega)x - \tilde{b}(\omega)), \quad x \in \mathbb{R}^n, \tag{4.1}$$

where $\tilde{c}$ is a fixed n-vector, $(\tilde{A}(\omega), \tilde{b}(\omega))$ is an $m \times (n+1)$ random matrix and p is a sublinear function [11] on $\mathbb{R}^m$. If $(A(\omega), b(\omega))$ is defined by

$$(A(\omega), b(\omega)) = \begin{pmatrix} \tilde{c}' & 0 \\ \tilde{A}(\omega) & b(\omega) \end{pmatrix}, \tag{4.2}$$

where m is replaced by $m+1$, and $u: \mathbb{R}^{1+m} \to \mathbb{R}$ is given by $u\binom{t}{z} = t + p(z)$, $\binom{t}{z} \in \mathbb{R}^{1+m}$, then obviously (4.1) is a mean value function of the type (1).

(b) Portfolio optimization [14, 17] is characterized within our framework by $m = 1$ and $b(\omega) \equiv 0$, where $u(z)$ is the negative utility of the return $z = A(\omega)x$ of the portfolio $x = (x_1, \ldots, x_n)'$.

(c) In error minimization and optimal design [1, 15] the loss function $u(z)$ measures the deviation $z = A(\omega)x - b(\omega)$ between the output $A(\omega)x$ of a stochastic linear system $x \to A(\omega)x$ with input x and the target $b(\omega)$. Solving (1), one wants to minimize the mean deviation between output and target.

Solving (3) by one of the standard mathematical programming routines [5] based mostly on certain derivatives of F, one meets the following difficulties:

(A) Under weak assumptions on the loss function u and the distribution P of the random matrix $(A(\omega), b(\omega))$ the gradient (or subgradient) of F exists and has the form

$$\nabla F(x) = EA(\omega)' \nabla u(A(\omega)x - b(\omega)), \quad x \in \mathbb{R}^n, \tag{5}$$

where ∇u is the gradient (or subgradient) of u and A' is the transpose of A. Corresponding formulas can also be obtained for the higher derivatives of F. However, since ∇F and the higher derivatives of F are defined in our case by certain multiple integrals, any programming routine involving derivatives of F is not very useful in practice since multiple integrals can be computed only with a big computational effort. In order to omit the expensive computation of the gradient ∇F of F in every iteration, we may replace $\nabla F(x)$ by a stochastic gradient (or subgradient) [6, 8, 12], as e.g.

$$Y(x) = A' \nabla u(Ax - b),$$

where (A, b) is a realization of the random matrix $(A(\omega), b(\omega))$ obtained e.g. by means of a pseudo random generator [12]. Unfortunately, since the deviation $\nabla F(x) - Y(x)$ between the gradient and its stochastic approximate is equal to zero

only in the mean, there is always a nonzero probability that the negative stochastic gradient is not a direction of decrease for F at x. In fact, in practice one observes [7] that stochastic approximation algorithms only have a poor convergence behaviour though under some weak conditions [6, 8] they converge with probability one to an optimal solution x^* of (3).

(B) Besides the above mentioned difficulties in the computation of multiple integrals, in the calculation of ∇F there is still a second main source of difficulties: The objective function F of (3) depends on the loss function u, hence $F = F_u$. Since in practice the penalty costs involved in u hardly can be specified exactly, there is always a certain uncertainty about the true loss function $u = u_0$. Hence, if $u \neq u_0$, then the direction given by formula (5) may be far away from the true gradient $\nabla F_{u_0}(x)$, see e.g. [11].

Consequently, in this paper we want to find feasible descent directions h of F at $x \in D$ such that

(I) h can be obtained without using any derivatives of F and

(II) h is stable with respect to variations of the loss function u in a large class U of loss functions u coming into question.

Note that safeguarding against variations of u is closely related to the stochastic dominance considerations in [9, 10] as well as in many other areas of decision theory, see e.g. [4, 16, 18].

In practice engineers often have already a certain approximation x_0 of an optimal solution x^* of (1). Having at x_0 a direction h of the above type, then we can compute at least an improvement x_1 of x_0, i.e. a vector $x_1 \in D$ such that $F(x_1) < F(x_0)$. Hence, it is certainly not reasonable to throw away descent directions h obtained in this way.

Moreover, introducing the set of 'efficient points' of (3), which may be computed also without using any derivatives of F, we get some information about the location of the optimal solutions $x^* = x_u^*$ of (3) for an arbitrary $u \in U$.

It turns out that the mentioned construction of derivative-free descent directions and efficient points depends on the weak assumption that $F = F_u$ is a quasiconvex function being not constant on the line segments under consideration, i.e. the class U of loss functions u is defined by

$$U = \{u\colon F = F_u \text{ is quasiconvex and } F_u \text{ is not constant on line segments } xy \text{ joining arbitrary points } x \in D \text{ and } y \in \mathbb{R}^n\}.$$

We want to add here that the proposed method for finding descent directions fails at every efficient point x^0. Consequently, in order to have a 'complete' algorithm for solving (3), one has to combine the proposed method with some other procedure, e.g. a stochastic approximation algorithm.

2. Some preliminary discussions

Denote by $K_x(D)$ the convex cone of feasible directions for D at x and let $xy = \{\lambda x + (1-\lambda)y\colon 0 \leq \lambda \leq 1\}$ be the line segment joining x and y. The simple basic

idea how feasible descent directions h of F at certain points $x \in D$ can be constructed is contained in the following theorem.

Theorem 2.1. *Assume that F is quasiconvex on D. If $x \in D$ is given and $y \in \mathbb{R}^n$ is such that $F(y) \leq F(x)$, then $h = y - x$ is a feasible descent direction of F at x provided that F is not constant on xy and $y - x \in K_x(D) \setminus \{0\}$.*

Proof. Obvious.

Before we are going to characterize points y such that $F(y) = F(x)$ in Section 3 and more generally $F(y) \leq F(x)$ in Section 4, we want to discuss the hypotheses on F.

Lemma 2.1. *Let $v: \Omega \times D \to \mathbb{R}$ be defined by $v(\omega, x) := u(A(\omega)x - b(\omega))$. If $v(\omega, \cdot)$ is convex a.s., i.e. for all ω up to a μ-null set, then F is convex on D, hence also quasiconvex. From the convexity of u the convexity of $v(\omega, \cdot)$ follows, but not vice versa.*

Lemma 2.2. *Assume that $v(\omega, \cdot)$ is convex a.s. Then F is constant on xy if and only if $F(x) = F(y)$ and $v(\omega, \cdot)$ is affine on xy a.s., i.e.*

$$v(\omega, \lambda x + (1-\lambda)y) = \lambda v(\omega, x) + (1-\lambda)v(\omega, y)$$

for all $\lambda \in [0, 1]$ and almost all $\omega \in \Omega$.

Proof. The 'if' part is clear. For the converse assume that F is constant on xy. Then for $0 \leq \lambda \leq 1$ it is

$$\begin{aligned} 0 &= \lambda F(x) + (1-\lambda)F(y) - F(\lambda x + (1-\lambda)y) \\ &= \int [\lambda v(\omega, x) + (1-\lambda)v(\omega, y) - v(\omega, \lambda x + (1-\lambda)y)]\mu(\mathrm{d}\omega). \end{aligned}$$

Since $v(\omega, \cdot)$ is convex a.s., the integrand is nonnegative a.s. Therefore the integral can be 0 only if the integrand is 0 a.s. This means that $v(\omega, \cdot)$ is affine on xy a.s.

Corollary 2.1. *Assume that $v(\omega, \cdot)$ is convex a.s. Then F is affine on xy if and only if $v(\omega, \cdot)$ is affine on xy a.s.*

Remark. If $v(\omega, \cdot)$ is convex a.s. and if there is a $\lambda_0 \in (0, 1)$ such that $v(\omega, \lambda_0 x + (1-\lambda_0)y) = \lambda_0 v(\omega, x) + (1-\lambda_0)v(\omega, y)$ a.s., then $v(\omega, \cdot)$ is affine on xy a.s.

Corollary 2.2. *If there is a vector $\eta \in \mathbb{R}^n$, $\eta \neq 0$, such that*

$$A(\omega)\eta = 0 \quad \text{a.s.}, \tag{4}$$

then F is constant on any line segment with direction η.

Proof. $F(x + \lambda\eta) = \int u(A(\omega)(x + \lambda\eta) - b(\omega))\mu(\mathrm{d}\omega) = \int u(A(\omega)x - b(\omega))\mu(\mathrm{d}\omega)$ $= F(x)$.

Remark. If $m=1$, i.e. if $A(\omega)$ is a row vector, and if some (≥ 2) components of $A(\cdot)$ are nonstochastic, then there exists an $\eta \neq 0$ such that (4) is satisfied. More generally assume that there are some (≥ 2) column vectors of $A(\cdot)$ which are linearly dependent. Then again there is a nontrivial solution η of (4). In [13] sufficient conditions are given for F being not constant on line segments xy.

In the situation of Corollary 2.2 it is easy to conjecture that the dimension n of problem (3) may be reduced.

Indeed, assume that $\eta=(\eta_1,\ldots,\eta_n)\neq 0$ is such that (4) is satisfied and without loss of generality let $\eta_n \neq 0$. Define $T:\mathbb{R}^n \to \mathbb{R}^{n-1}$ and $\tilde{D}$ by

$$(T(x_1,\ldots,x_n))_j := x_j - \frac{x_n}{\eta_n}\eta_j, \quad j=1,\ldots,n-1, \qquad \tilde{D} := T(D). \tag{5}$$

If $A(\omega)=(a_1(\omega),\ldots,a_n(\omega))$, then let $\tilde{A}(\omega)=(a_1(\omega),\ldots,a_{n-1}(\omega))$. Furthermore, define $\tilde{F}:\mathbb{R}^{n-1}\to\mathbb{R}$ by

$$\tilde{F}(\tilde{x}) = Eu(\tilde{A}(\omega)\tilde{x} - b(\omega)) \quad \text{for } \tilde{x}\in\mathbb{R}^{n-1}. \tag{6}$$

Then for $x\in D$ we have $F(x)=\tilde{F}(Tx)$. Hence $F(D)=\tilde{F}(\tilde{D})$ and therefore

$$\inf_{x\in D} F(x) = \inf_{\tilde{x}\in\tilde{D}} \tilde{F}(\tilde{x}).$$

If $x_0\in D$ is such that

$$F(x_0) = \inf_{x\in D} F(x), \tag{7}$$

then $Tx_0\in\tilde{D}$ is such that

$$\tilde{F}(Tx_0) = \inf_{\tilde{x}\in\tilde{D}} \tilde{F}(\tilde{x}). \tag{8}$$

Conversely, assume that $\tilde{x}_0\in\tilde{D}$ is such that

$$\tilde{F}(\tilde{x}_0) = \inf_{\tilde{x}\in\tilde{D}} \tilde{F}(\tilde{x}). \tag{9}$$

Then for any

$$x_0 \in T^{-1}\{\tilde{x}_0\} \cap D \tag{10}$$

we have

$$F(x_0) = \inf_{x\in D} F(x).$$

So we have proved the following

Theorem 2.2. *If $\eta=(\eta_1,\ldots,\eta_n)\neq 0$ satisfies* (4), *then the n-dimensional problem* (3) *may be transformed into a $n-1$-dimensional one. If $\eta_n\neq 0$, then the latter is given by*

$$\min_{\tilde{x}\in\tilde{D}} \tilde{F}(\tilde{x}), \tag{11}$$

where $\tilde{D}$ and $\tilde{F}$ are defined by (5) *and* (6).

The minimum in (3) *is attained if and only if the corresponding one in* (11) *is attained. The minima of problems* (3) *and* (11) *are equal and the points where they are attained are related by* (7)-(10).

More generally than above one may ask the consequences of

$$A(\omega)\eta = c \quad \text{a.s.} \tag{12}$$

for some $\eta \neq 0$. In a special case we can give the following answer:

Theorem 2.3. *If $m = 1$ and $b(\omega) = 0$ a.s. and if there is an $\eta \neq 0$ such that $\{\lambda\eta: \lambda \in \mathbb{R}\} \subset D$ and (12) is satisfied with a $c \neq 0$, then*

$$\inf_{x \in D} F(x) = \inf_{\lambda \in \mathbb{R}} F(\lambda\eta) = \inf_{z \in \mathbb{R}} u(z).$$

Proof. $F(x) = \int u(z) P_{A(\cdot)x}(\mathrm{d}z) \geq \inf_{z \in \mathbb{R}} u(z)$,

hence

$$\inf_{x \in D} F(x) \geq \inf_{z \in \mathbb{R}} u(z).$$

Furthermore

$$F(\lambda\eta) = \int u(A(\omega)\lambda\eta)\mu(\mathrm{d}\omega) = \int u(\lambda c)\mu(\mathrm{d}\omega) = u(\lambda c).$$

Therefore

$$\inf_{\lambda \in \mathbb{R}} F(\lambda\eta) = \inf_{\lambda \in \mathbb{R}} u(\lambda c) = \inf_{z \in \mathbb{R}} u(z).$$

Remark. Problems (3) arising from portfolio analysis are characterized by $m = 1$ and $b(\omega) = 0$.

As a generalization of the above Theorem 2.3 we formulate the

Remark. Assume $b(\omega) = 0$ a.s. and let $\eta_1, \ldots, \eta_m \in \mathbb{R}^n$ be such that $A(\omega)\eta_j = c_j$ a.s., $j = 1, \ldots, m$, with linearly independent vectors $c_1, \ldots, c_m \in \mathbb{R}^m$. If $\mathcal{L}\{\eta_1, \ldots, \eta_m\} \subset D$ ($\mathcal{L}$ is the linear hull), then

$$\inf_{x \in D} F(x) = \inf_{(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m} F(\lambda_1\eta_1 + \cdots + \lambda_m\eta_m) = \inf_{z \in \mathbb{R}^m} u(z).$$

We can also give a generalization of Corollary 2.2.

Corollary 2.3. *If there is a vector $\eta \in \mathbb{R}^n$, $\eta \neq 0$, which satisfies (12) with a vector c such that, for all $\lambda \in \mathbb{R}$ and $x \in D$,*

$$\int u(A(\omega)x - \lambda c - b(\omega))\mu(\mathrm{d}\omega) = \int u(A(\omega)x - b(\omega))\mu(\mathrm{d}\omega), \tag{13}$$

then F is constant on any line segment with direction η.

Remark. (13) is satisfied if

(i) $c = 0$ or if

(ii) $u(z + \lambda c) = u(z)$ for all $\lambda \in \mathbb{R}$,

i.e. u is invariant with respect to translations parallel to c.

Obviously, Theorem 2.2 remains valid in the more general situation of Corollary 2.3.

3. How to find points y with $F(y) = F(x)$

In this section we want to restrict our attention to the problem of finding a $y \in \mathbb{R}^n$ with $F(x) = F(y)$ for a given $x \in D$. If not otherwise stated we want to assume in what follows that F is quasiconvex and not constant on the line segments under consideration. Then from $x \in D$, $y \in \mathbb{R}^n$, $F(x) = F(y)$ and $y - x \in K_x(D)\backslash\{0\}$ we can conclude that $y - x$ is a feasible descent direction of F at x. A vector $h \in K_x(D)\backslash\{0\}$ is called a nontrivial feasible direction for D at x. In the next subsection we give a lemma which is fundamental for the rest of this paragraph.

(3a) A basic lemma

Lemma 3.1. *Let $x \in D$ be given. If $y \in \mathbb{R}^n$ is such that*

$$u(A(\omega)x - b(\omega)) = u(A(\omega)y - b(\omega)) \quad a.s. \tag{14}$$

or

$$P_{A(\cdot)x - b(\cdot)} = P_{A(\cdot)y - b(\cdot)}, \tag{15}$$

then $F(x) = F(y)$.

In the situation of (15) $F(x) = F(y)$ holds for all loss functions u for which F exists.

Proof. The assertions follow immediately from the representation (2) of F. Of course, (14) is of minor importance. We only give an example when (14) is satisfied.

Remark. Assume that A is a $n \times n$ matrix and let S be a constant matrix of the same dimension satisfying

$$u(Sz) = u(z) \qquad \text{for all } z \in \mathbb{R}^n,$$
$$SA(\omega) = A(\omega)S \quad \text{a.s.},$$
$$Sb(\omega) = b(\omega) \qquad \text{a.s.}$$

Let $x \in D$ be given and define $y = Sx$. Then (14) is satisfied.

Before we are going to discuss the decisive equation (15), we want to introduce first the concept of an efficient point.

(3b) Efficient points

Definition 3.1a. A point $x \in D$ is called efficient for (3) if there is no point $y \in \mathbb{R}^n$ such that (15) is satisfied and $y - x \in K_x(D)\backslash\{0\}$.

If $x \in \mathring{D}$, where $\mathring{D}$ denotes the topological interior of D, then $K_x(D) = \mathbb{R}^n$ and consequently x is efficient in this case if and only if there is no solution y of (15) such that $y \neq x$.

The significance of efficient points for (3) is easy to see. Let $x^* \in D$ be an optimal solution of (3). Then there can be no y, $y - x^* \in K_{x^*}(D) \setminus \{0\}$ such that (15) with $x = x^*$ holds. Hence, the optimal solutions of (3) are efficient points or to state it otherwise, the efficient points of (3) are possible candidates for optimal solution of (3).

Let E denote the set of efficient points for (3).

Remark. (a) If besides the weak condition $u \in U$, i.e. that the objective function F_u of (3) is quasiconvex and is not constant on the line segments under consideration, there is some additional knowledge about the loss function u, then a stronger definition of an efficient point can be given. We mention two examples.

(1) Stochastic linear programs with recourse

According to the special form (4) of the objective function F, Definition 3.1a can be modified here as follows.

Definition 3.1b. For a stochastic linear program with recourse a point $x \in D$ is called efficient if there is no point $y \in \mathbb{R}^n$ such that $y - x \in K_x(D) \setminus \{0\}$ and

$$\tilde{c}'y \leqq \tilde{c}'x, \tag{16.1}$$

$$P_{\tilde{A}(\cdot)y - \tilde{b}(\cdot)} = P_{\tilde{A}(\cdot)x - \tilde{b}(\cdot)}. \tag{16.2}$$

(2) Problems with a monotone loss function u

In some applications, e.g. in portfolio optimization [14], the loss function u has a monotonicity property of the type

$$u(z) \leqq u(w) \quad \text{if } z \leqq (\geqq) w, \tag{17}$$

where the partial ordering $z \leqq (\geqq) w$ is defined componentwise. Let $(\bar{A}, \bar{b})$ denote the mean of $(A(\omega), b(\omega))$ and define $(A_0(\omega), b_0(\omega)) = (A(\omega) - \bar{A}, b(\omega) - \bar{b})$. Then Definition 3.1a can be modified in this way.

Definition 3.1c. If the loss function u is monotone according to (17), then a point $x \in D$ is called efficient if there is no point $y \in \mathbb{R}^n$ such that $y - x \in K_x(D) \setminus \{0\}$ and

$$\bar{A}y \leqq (\geqq) \bar{A}x, \tag{18.1}$$

$$P_{A_0(\cdot)y - b_0(\cdot)} = P_{A_0(\cdot)x - b(\cdot)}. \tag{18.2}$$

We observe that the notion of efficiency in optimal portfolio theory is closely related to the efficiency concept given by the above Definition 3.1c.

Notes. (a) Equations (15), (16.2) and (18.2) are of the same type!

(b) It may happen, e.g. if $n=1$, that the set E of efficient points described by Definitions 3.1a–c is not much different from the set D of all admissible points of (3) or is even equal to D. Clearly, if this occurs, then the notion of efficiency yields only few or even trivial information about the location of an optimal solution x^* of (3). On the other hand, if this happens, then we know at least that the corresponding problem is very sensitive with respect to variations of u. Hence, in this case one must be very careful in selecting an appropriate loss function u for (3).

In any case, an important problem is therefore to compute the set E of efficient points or if not possible to give inclusions for this set.

Theorem 3.1. *For given $x \in \mathbb{R}^n$ let A_x and B_x be two systems of conditions for an n-vector y. Assume that whenever a vector y satisfies A_x for a given x, then y is also a solution of* (15) (*relations* (16), (18) *rsp.*). *Conversely, assume that whenever a vector y satisfies* (15) (*relations* (16), (18) *rsp.*) *for a given x, then y is also a solution of B_x. Let A (B rsp.) be the set of points $x \in D$ such that no solution y of A_x (B_x rsp.) with $y - x \in K_x(D)\backslash\{0\}$ exists. Then*

$$B \subset E \subset A,$$

where E is the set of efficient points according to Definition 3.1a (3.1b, 3.1c *rsp.*).

Proof. If $x \in B$, then B_x has no solution y with $y - x \in K_x(D)\backslash\{0\}$. Consequently, also (15) (relations (16), (18) rsp.) has no solution y such that $y - x \in K_x(D)\backslash\{0\}$. For if y satisfies (15) (relations (16), (18) rsp.), then it satisfies B_x, too. Hence x is efficient.

The other inclusion is proved similarly.

By means of Theorem 3.1 it is possible now to compare the notion of efficiency given by Definition 3.1a and Definition 3.1b, 3.1c rsp.

(1) Stochastic linear programs with recourse. Since $(A(\omega), b(\omega))$ is defined here by (4.2), $\tilde{c}$ being a fixed vector, (15) holds if and only if $\tilde{c}'x = \tilde{c}'y$ and $P_{\tilde{A}(\cdot)x-\tilde{b}(\cdot)} = P_{\tilde{A}(\cdot)y-\tilde{b}(\cdot)}$. Hence, (15) implies relation (16). From Theorem 3.1 follows now that the set E_{SLP} of efficient points described by Definition 3.1b is a subset of the set E of efficient points according to Definition 3.1a.

(2) Problems with a monotone loss function u. According to the definition of $(\bar{A}, \bar{b})$ and $(A_0(\omega), b_0(\omega))$, (15) holds if and only if $\bar{A}x = \bar{A}y$ and $P_{A_0(\cdot)x-b(\cdot)} = P_{A_0(\cdot)y-b_0(\cdot)}$. Hence, (15) implies relation (18). Again from Theorem 3.1 follows that the set E_M of efficient points according to Definition 3.1c is a subset of E.

Since in the above two examples we have additional information about the loss function u, we obtain a notion of efficiency which is sharper than that given by the general Definition 3.1a working also in the case with maximal uncertainty about u. Clearly, the set of efficient points for (3) depends on D. We give the following

Lemma 3.2. *Let $E(D)$ denote the set of efficient points (according to Definition* 3.1a, 3.1b *or* 3.1c) *for* (3). *If $D_1 \subset D_2$ are two convex subset of $\mathbb{R}^n$, then $E(D_1) \supset D_1 \cap E(D_2)$.*

Proof. Because of $D_1 \subset D_2$ we also have that $K_x(D_1) \subset K_x(D_2)$ for every x. Let $x \in D_1 \cap E(D_2)$ and suppose that $x \notin E(D_1)$. Then there is a $y \in \mathbb{R}^n$ with $y - x \in K_x(D_1)\backslash\{0\}$ such that (15) (relation (16), (18) rsp.) is satisfied. Since $K_x(D_1) \subset K_x(D_2)$, we also have that $y - x \in K_x(D_2)\backslash\{0\}$. Hence, there is a $y \in \mathbb{R}^n$ with $y - x \in K_x(D_2)\backslash\{0\}$ such that (15) (relation (16), (18) rsp.) is satisfied, which is a contradiction to $x \in E(D_2)$.

Proving the inclusions $E_{SLP} \subset E$ and $E_M \subset E$, we already have worked with a system of relations B_x. Further examples for our systems A_x and B_x will be given in the next subsections.

(3c) Conditions necessary for equation (15) ((16.2), (18.2) rsp.)

Since (15), (16.2) and (18.2) are of the same type, it is sufficient to consider (15) only. Note that these equations may also be described by means of the characteristic functions or the densities (if they exist) of the corresponding probability distributions, see e.g. [13].

Theorem 3.2. *If two n-vectors x, y are related by* (15), *then for every measurable function $f: \mathbb{R}^m \to \mathbb{R}^r$ we have*

$$P_{f(A(\cdot)x - b(\cdot))} = P_{f(A(\cdot)y - b(\cdot))},$$

and, provided these expectations exist,

$$Ef(A(\omega)x - b(\omega)) = Ef(A(\omega)y - b(\omega)).$$

Especially, for each $r \times m$ matrix H the equation

$$P_{H(A(\cdot)x - b(\cdot))} = P_{H(A(\cdot)y - b(\cdot))}$$

must hold and hence also

$$P_{A_i(\cdot)x - b_i(\cdot)} = P_{A_i(\cdot)y - b_i(\cdot)} \quad \text{for } i = 1, \ldots, m, \tag{19}$$

where (A_i, b_i) is the ith row of (A, b). Also all moments of $A(\omega)x - b(\omega)$ and $A(\omega)y - b(\omega)$ must be equal, provided they exist. Let us note especially

$$\bar{A}x = \bar{A}y, \tag{20}$$

$$\operatorname{var}(A(\cdot)x - b(\cdot)) = \operatorname{var}(A(\cdot)y - b(\cdot)), \tag{21}$$

where $\bar{A} = EA(\omega)$ is the mean of $A(\omega)$ and $\operatorname{var}(A(\cdot)x - b(\cdot))$ denotes the covariance matrix of $A(\omega)x - b(\omega)$, provided these moments exist.

Remark. The equations of Theorem 3.2 are examples for relation systems B_x. Hence, they are suitable for searching efficient points.

In some cases the conditions of Theorem 3.2 are also sufficient for (15) as demonstrated by this

Lemma 3.3. (a) *If the rows of* $(A(\omega), b(\omega))$ *are stochastically independent, then equations* (19) *are also sufficient for* (15).

(b) *If the random matrix* $(A(\omega), b(\omega))$ *has a normal distribution, then* (20) *and* (21) *are also sufficient for* (15).

(c) *If* $(A(\omega), b(\omega))$ *has an elliptically contoured distribution (see e.g.* [13]), *then* (15) *is equivalent to a system of linear and quadratic equations which have the same functional form as the equations given by* (20) *and* (21).

Among many other distributions, the multivariate normal, Cauchy, stable, Student t-, inverted Student t-distribution as also their truncations to an elliptically contoured set in $\mathbb{R}^{m(n+1)}$ and the uniform distributions on elliptically contoured sets in $\mathbb{R}^{m(n+1)}$ are members of the family of elliptically contoured distributions, see e.g. [13].

We want to discuss (20) and (21) in some detail. First we will consider only the row means

$$\Theta = (\bar{A}_i)_{i \in J}$$

for some $J \subset \{1, \ldots, m\}$, i.e. (20) implies that

$$\Theta(y - x) = 0. \tag{22}$$

The general solution of (22) is given by

$$y - x = B_J h,$$

where $h \in \mathbb{R}^r$, B_J is a $n \times r$ matrix with rank $B_J = r$ and $r = n - \text{rank } \Theta$. If $x \in \mathring{D}$ and if rank $\Theta < n$ (the latter is true in practical relevant cases since $m < n$), then there is always a solution $y - x \in K_x(D) \setminus \{0\}$ of (22). So assume $x \in \partial D$, ∂D being the boundary of D. Let D be given e.g. by

$$D = \{x \in \mathbb{R}^n : g(x) \leqslant g_0\},$$

where g is a differentiable function and $g_0 \in \mathbb{R}$. If $g(x) = g_0$ and $\nabla g(x) \neq 0$, then the topological closure $\overline{K_x(D)}$ of $K_x(D)$ is given by

$$\overline{K_x(D)} = \{\eta \in \mathbb{R}^n : \nabla g(x)' \eta \leqslant 0\}.$$

We can give this

Lemma 3.4. *Let* $x \in \partial D$ *be given. Assume that* $g(x) = g_0$, $\nabla g(x) \neq 0$ *and* $K_x(D) = \{\eta \in \mathbb{R}^n : \nabla g(x)' \eta < 0\}$.

If $B_J' \nabla g(x) = 0$, *then* x *is efficient.*

Proof. By the above an $x \in \partial D$ with $\nabla g(x) \neq 0$ is efficient if

$$\Theta \eta = 0, \qquad \nabla g(x)' \eta < 0$$

have no solution $\eta \in \mathbb{R}^n$ or equivalently if

$$\nabla g(x)' B_J h < 0$$

has no solution $h \in \mathbb{R}^r$.

Of course this is the case if and only if $B_J' \nabla g(x) = 0$.

Example 1. If $|J|=1$ and $\Theta \neq 0$, then rank $B_J = n-1$. So, in order that $B_J' \nabla g(x) = 0$ there must exist an $\alpha \in \mathbb{R}$ such that

$$\nabla g(x) = \alpha \Theta'.$$

That is, if we can find a row vector $EA_i(\omega)$ such that for an $x \in \partial D$ the relation

$$\nabla g(x)' = \alpha E A_i(\omega) \neq 0$$

holds, then x is efficient.

Example 2. Let D be an ellipsoid, that is $g_0 = 0$ and $g(x) = x'Cx + c'x + \gamma$, where C is a positive definite matrix. From Lemma 3.4 it follows that if

$$x'Cx + c'x + \gamma = 0, \qquad 2Cx + c \neq 0, \qquad 2B_J'Cx = -B_J'c,$$

then x is efficient.

Let us now take into consideration only the covariances, i.e., (21) alone.

Denote by Q_{ij} the $(n+1)\times(n+1)$-matrix

$$Q_{ij} = \operatorname{cov}((A_i(\omega), b_i(\omega)), (A_j(\omega), b_j(\omega))).$$

Then it is easily verified that

$$\operatorname{var}(A(\omega)x - b(\omega)) = (\hat{x}' Q_{ij} \hat{x})_{i,j=1,\ldots,m},$$

where again $\hat{x} = \binom{x}{-1}$.

Hence, (21) is equivalent to

$$\hat{x}' Q_{ij} \hat{x} = \hat{y}' Q_{ij} \hat{y}, \quad i,j = 1, \ldots, m,$$

and since $Q_{ji} = Q_{ij}'$ this can be reduced to

$$\hat{x}' \frac{Q_{ij} + Q_{ji}}{2} \hat{x} = \hat{y}' \frac{Q_{ij} + Q_{ji}}{2} \hat{y} \quad \text{for } i \leq j,$$

where now $\frac{1}{2}(Q_{ij} + Q_{ji})$ is a symmetric matrix. So we can write

$$\tfrac{1}{2}(Q_{ij} + Q_{ji}) = \begin{pmatrix} R_{ij} & d_{ij} \\ d_{ij}' & q_{ij} \end{pmatrix} \quad \text{for } i \leq j,$$

where the R_{ij} are symmetric $n \times n$-matrices and $d_{ij} \in \mathbb{R}^n$, $q_{ij} \in \mathbb{R}$. Hence,

$$\hat{y}' \frac{Q_{ij} + Q_{ji}}{2} \hat{y} = y' R_{ij} y - 2 d_{ij}' y + q_{ij} \quad \text{for } i \leq j.$$

So (21) is equivalent to

$$x' R_{ij} x - 2 d_{ij}' x = y' R_{ij} y - 2 d_{ij}' y \tag{23}$$

for $i \leq j$ and x is efficient if there is no y, $y - x \in K_x(D) \backslash \{0\}$, which satisfies (23) simultaneously for all $i \leq j$.

Denote by K_1 (K_2 rsp.) those pairs (i,j), $i \leq j$, for which R_{ij} is positive (negative rsp.) definite. R_{ii} is positive semidefinite, so chances are good that K_1 is not empty.

For $(i,j) \in K_1$ and $y \neq x$ we have

$$y'R_{ij}y - 2d'_{ij}y > x'R_{ij}x - 2d'_{ij}x + 2(R_{ij}x - d_{ij})'(y-x)$$

and hence, from (23),

$$0 > (R_{ij}x - d_{ij})'(y-x) \quad \text{for } (i,j) \in K_1,\ y \neq x.$$

Analogously, we must have

$$0 < (R_{ij}x - d_{ij})'(y-x) \quad \text{for } (i,j) \in K_2,\ y \neq x.$$

Hence, we can prove

Theorem 3.3. *If there is no solution $\eta \in K_x(D)$ of*

$$\begin{aligned} 0 &> (R_{ij}x - d_{ij})'\eta \quad \text{for } (i,j) \in K_1, \\ 0 &< (R_{ij}x - d_{ij})'\eta \quad \text{for } (i,j) \in K_2, \end{aligned} \tag{24}$$

where $x \in D$, then x is efficient.

Proof. If x is not efficient, then there is a solution y of (21) with $y - x \in K_x(D)\backslash\{0\}$. By the above considerations we know that in this case $\eta = y - x \in K_x(D)$ is a solution of (24).

Finally we want to take into consideration both mean and covariance.

Denote by B the matrix B_J corresponding to $J = \{1, \ldots, m\}$. Substituting $y = x - Bh$ into (23) we can write

$$x'R_{ij}x - 2d'_{ij}x = x'R_{ij}x + h'B'R_{ij}Bh + 2h'B'R_{ij}x - 2d'_{ij}x - 2d'_{ij}Bh.$$

So we have

Theorem 3.4. *Equations* (20) *and* (21) *have no solution y with $y - x \in K_x(D)\backslash\{0\}$ iff there is no $h \in \mathbb{R}^r$ with $Bh \in K_x(D)\backslash\{0\}$ such that*

$$g_{ij}(h) = h'B'R_{ij}Bh + 2(B'R_{ij}x - B'd_{ij})'h = 0 \tag{25}$$

simultaneously for all pairs (i,j) with $i \leq j$. If there is no such $h \in \mathbb{R}^r$, then x is efficient.

Let us note an interesting consequence:

Corollary 3.1. *Denote by K those pairs (i,j) with $i \leq j$ such that R_{ij} is definite. If $x \in D$ is such that*

$$B'R_{ij}x = B'd_{ij}, \tag{26}$$

for some $(i,j) \in K$, then x is efficient.

Proof. (26) implies $g_{ij}(h) = h'B'R_{ij}Bh$. By the definition of K we conclude that rank $B'R_{ij}B = r$ for $(i,j) \in K$. So $B'R_{ij}B$ is a definite matrix and $g_{ij}(h) = 0$ is true if and only if $h = 0$. But this means that only $h = 0$ satisfies (25) for all $i \leq j$. Hence x is efficient by Theorem 3.4.

Corollary 3.1 can also be derived from Theorem 3.3. If we take into consideration that $\eta = Bh$ is the general solution of (20), then we can tell that $x \in D$ is efficient if there is no solution h of

$$0 > (R_{ij}x - d_{ij})'Bh \quad \text{for } (i,j) \in K_1,$$

$$0 < (R_{ij}x - d_{ij})'Bh \quad \text{for } (i,j) \in K_2.$$

So if $(R_{ij}x - d_{ij})'B = 0$ for a $(i,j) \in K = K_1 \cup K_2$, then x is efficient. Since $K_1 \cup K_2$ is the set of indices (i,j) such that R_{ij} is definite we have again proved Corollary 3.1.

Corollary 3.2. *For each $(i,j) \in K$ all points $x \in D$ contained in the $n-r$ dimensional linear manifold*

$$S_{ij} = \{x \in D: x = R_{ij}^{-1}(d_{ij} + \lambda_1 EA_1'(\omega) + \cdots + \lambda_m EA_m'(\omega)), (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m\}$$

are efficient. Here $A_j(\omega)$ denotes the jth row of $A(\omega)$.

Proof. (26) means that $R_{ij}x - d_{ij}$ is in the kernel of B'. By the very definition of B the kernel of B' is the linear hull of the rows of $EA(\omega)$. This proves the corollary.

All $x \in D$ in the union of the $n-r$ dimensional linear manifolds S_{ij} are efficient. But in general there are still more efficient points. For $(i,j) \in K$ the set $\{x \in \mathbb{R}^n: g_{ij}(x) = 0\}$ is an ellipsoid. Now, $x \in D$ is efficient if any two such ellipsoids, say those pertaining to (i,j) and (k,l), are tangent to each other from outside in 0. This is true if for a $\lambda < 0$

$$\nabla g_{ij}(0) = \lambda \nabla g_{kl}(0),$$

which is equivalent to

$$B'R_{ij}x - B'd_{ij} = \lambda(B'R_{kl}x - B'd_{kl}).$$

So we have proved

Corollary 3.3. *Let $(i,j) \in K$ and $(k,l) \in K$ be given. If $\lambda < 0$ and $x \in D$ are such that*

$$B'(R_{ij} - \lambda R_{kl})x = B'(d_{ij} - \lambda d_{kl}),$$

then x is efficient.

Note that Corollary 3.1 is a special case of Corollary 3.3 if we allow $\lambda = 0$ in the latter.

Corollary 3.4. *Suppose that $m = 1$ and that $(A(\omega), b(\omega))$ has a normal distribution. Let R, where* det $R \neq 0$, *d and q be such that*

$$\operatorname{var}(A(\omega), b(\omega)) = \begin{pmatrix} R & d \\ d' & q \end{pmatrix}$$

and let B be an $n\times(n-1)$ matrix with rank $B=n-1$ *and* $EA(\omega)B=0$. *If $x\in D$ is efficient then*

$$x\in\partial D\cup(\mathring{D}\cap G),$$

where G is the straight line

$$G=\{x\colon x=R^{-1}(d+\lambda EA(\omega)'),\ \lambda\in\mathbb{R}\}.$$

If $D=\mathbb{R}^n$, then the set of efficient points is equal to G.

Proof. In our situation (25) reads

$$g(h)=h'B'RBh+2(B'Rx-B'd)'h=0. \tag{27}$$

By Lemma 3.2b and Theorem 3.5, $x\in D$ is efficient if and only if (27) has no solution $h\in\mathbb{R}^{n-1}$ with $Bh\in K_x(D)\backslash\{0\}$. Let $x\in\mathring{D}$ be efficient. Then $K_x(D)=\mathbb{R}^n$ and (27) has no solution $h\neq 0$. g being continuous this can be true only if $g(h)<0$ for all $h\neq 0$ or $g(h)>0$ for all $h\neq 0$. In both cases $h=0$ is an extremum of $g(\cdot)$. So necessarily $\nabla g(0)=0$ which gives $B'(Rx-d)=0$. But this means that $Rx-d$ is in the kernel of B' and therefore $Rx-d=\lambda EA(\omega)'$ for a $\lambda\in\mathbb{R}$. This proves the first part of our corollary.

If $D=\mathbb{R}^n$ then we have proved that E, the set of efficient points, is a subset of G. By Corollary 3.1 a point x with $B'(Rx-d)=0$ is efficient. So the points of G are efficient which proves the second part of our corollary.

(3d) Conditions sufficient for (15)

Let us define $E:=\mathbb{R}^{m(n+1)}$, where again the elements of E are $m\times(n+1)$ matrices. Then P, the distribution of $(A(\cdot),b(\cdot))$, is a probability measure on E.

Theorem 3.5. *Assume that there are two measurable mappings*

$$\Lambda_k\colon E\to E \quad \text{for } k=1,2,$$

such that P is invariant with respect to Λ_k, i.e.

$$\Lambda_k(P)=P \quad \text{for } k=1,2.$$

Let $x\in\mathbb{R}^n$ be given. If $y\in\mathbb{R}^n$ is such that

$$H_x\circ\Lambda_1=H_y\circ\Lambda_2 \quad P\text{-a.s.}, \tag{28}$$

then (15) *is satisfied.*

Proof. By our hypotheses

$$P_{A(\cdot)x-b(\cdot)}=H_x(P)=H_x(\Lambda_1(P))=H_y(\Lambda_2(P))=H_y(P)=P_{A(\cdot)y-b(\cdot)}.$$

Remark 1. If Λ_2 is a 1-1 mapping, then (28) reduces to the condition $H_x\circ\Lambda=H_y$ with $\Lambda\colon E\to E$ known from [12].

Remark 2. For an invariance Λ of P clearly Λ^r is also an invariance of P for all $r \in \mathbb{N}$. So (15) is satisfied if $H_x \circ \Lambda_1^r = H_y \circ \Lambda_2^s$ for some $r \in \mathbb{N}$, $s \in \mathbb{N}$ and invariances Λ_1, Λ_2 of P.

Remark 3. Equation (28) is an example for a relation system A_x (see Theorem 3.2).

So an immediate consequence is

Corollary 3.5. *If for a given $x \in D$ there is a y with $y - x \in K_x(D) \setminus \{0\}$ which satisfies* (28) *for some invariances Λ_1, Λ_2, then x is not efficient.*

Let us restrict our attention to affine transformations $\Lambda_k : E \to E$ given by

$$\Lambda_k(M) = T_k M + \Delta_k \quad \text{for } k = 1, 2,$$

where $T_k : E \to E$ are linear transformations and Δ_k are fixed elements of E for $k = 1, 2$.

Then (28) is equivalent to

$$(T_1 M + \Delta_1)\hat{x} = (T_2 M + \Delta_2)\hat{y} \quad P\text{-a.s.} \tag{29}$$

Let $\tau_{ij}^{(k)}$ be $(n+1) \times (n+1)$ matrices for $i, j = 1, \ldots, m$ and $k = 1, 2$, $\tau_k = (\tau_{ij}^{(k)})_{i,j=1,\ldots,m}$, $k = 1, 2$, and assume that T_k is given by $\operatorname{vec}(T_k M) = (\operatorname{vec} M)\tau_k$ for $k = 1, 2$.

Introducing the notation

$$S_x = \operatorname{diag}(\hat{x}, \ldots, \hat{x}),$$

S_x being an $m(n+1) \times m$ matrix, we see that (29) can be written as

$$(\Delta_1 \hat{x})' + (\operatorname{vec} M)\tau_1 S_x = (\Delta_2 \hat{y})' + (\operatorname{vec} M)\tau_2 S_y \quad P\text{-a.s.} \tag{30}$$

(30) can be further reduced if we have some information about the support C_P of P (C_P is the minimal closed subset of E such that $P(C_P) = 1$).

Lemma 3.5. *If $\Lambda(P) = P$ and $\Lambda(E)$ is a closed set, then*

$$C_P \subset \Lambda(E).$$

Proof. Of course $E = \Lambda^{-1}(\Lambda(E))$ and hence, also using $\Lambda(P) = P$, we have

$$P(E) = P(\Lambda^{-1}(\Lambda(E))) = P(\Lambda(E)),$$

hence $P(E) = 1$. Since $\Lambda(E)$ is closed by our assumption, according to the definition of support we conclude that $C_P \subset \Lambda(E)$.

If Λ is an affine transformation then $\Lambda(E)$ is a linear manifold and therefore closed. So Lemma 3.5 applies in this case. We want to give now examples for supports C_P and the corresponding simplifications of (30).

Example 1. *Full support.* Let 0 be an inner point of C_P. Then C_P is not a subset of any linear manifold other than E. So by Lemma 3.5 we have $\Lambda_2(E) = E$ and since E is finite dimensional Λ_2 must be 1-1. By Remark 1, (30) can be reduced to

$$(\Delta\hat{x})' + (\text{vec } M)\tau S_x = (\text{vec } M) S_y \quad \text{for all } M \in C_P.$$

Since 0 is an inner point of C_P this is satisfied if and only if

$$\Delta\hat{x} = 0, \qquad \tau S_x = S_y.$$

These conditions were obtained in [13].

Example 2. *Affine support.* Let $M_0 \in E$ and $H = (H_1, \ldots, H_m)$, H_j being a $r \times (n+1)$ matrix, be given. Assume that C_P is the linear manifold given by

$$C_P = \{M + M_0, \text{vec } M = \lambda' H, \lambda \in \mathbb{R}^r\}.$$

Now (30) is equivalent to

$$(\Delta_1\hat{x})' + (\text{vec } M_0 + \lambda' H)\tau_1 S_x = (\Delta_2\hat{y})' + (\text{vec } M_0 + \lambda' H)\tau_2 S_y \quad \text{for all } \lambda \in \mathbb{R}^r.$$

This is true if and only if

$$(\Delta_1\hat{x})' + (\text{vec } M_0)\tau_1 S_x = (\Delta_2\hat{y})' + (\text{vec } M_0)\tau_2 S_y, \qquad H\tau_1 S_x = H\tau_2 S_y.$$

Example 3. *Polyhedral support.* Let C_P be given by

$$C_P = \text{conv}\{M^{(1)}, \ldots, M^{(r)}\},$$

i.e. vec $M = \sum_{j=1}^r \alpha_j \text{vec } M^{(j)}$, $\alpha_j \geq 0$, $\sum_{j=1}^r \alpha_j = 1$. If $0 \in C_P$ then (30) reduces to

$$\Delta_1\hat{x} = \Delta_2\hat{y}, \qquad \sum_{j=1}^r \alpha_j \text{vec } M^{(j)}\tau_1 S_x = \sum_{j=1}^r \alpha_j \text{vec } M^{(j)}\tau_s S_y.$$

Of course the latter equation is valid if

$$\text{vec } M^{(j)}\tau_1 S_x = \text{vec } M^{(j)}\tau_2 S_y \quad \text{for } j = 1, \ldots, r.$$

In [13] methods for finding invariances for concrete distributions are discussed.

4. How to find points y with $F(y) \leqq F(x)$

Let $K: \mathbb{R}^m \times \mathcal{B}^m \to \bar{\mathbb{R}}$ be a Markov kernel from $\mathbb{R}^m$ to $\mathbb{R}^m$, i.e., assume that $K(z, \cdot)$ is a measure on $(\mathbb{R}^m, \mathcal{B}^m)$ for each fixed $z \in \mathbb{R}^m$ and that $K(\cdot, B)$ is a measurable function on $\mathbb{R}^m$ for each fixed $B \in \mathcal{B}^m$. Assume further that $K(z, B)$ is a dilatation, i.e.,

$$\int wK(z, \mathrm{d}w) = z \quad \text{for } z \in \mathbb{R}^m.$$

If ν is any measure on $(\mathbb{R}^m, \mathcal{B}^m)$, then by $K\nu$ we designate the measure

$$(K\nu)(B) = \int K(z, B)\nu(\mathrm{d}z).$$

Now we have this

Theorem 4.1. *Assume that*

$$P_{A(\cdot)x-b(\cdot)} = KP_{A(\cdot)y-b(\cdot)} \tag{31}$$

with a dilatation K. Then

(a) $EA(\omega)x = EA(\omega)y$,

(b) $F(y) \leq F(x)$

for all convex loss functions u.

Proof. (a)

$$E(A(\omega)x - b(\omega)) = \int wP_{A(\cdot)x-b(\cdot)}(\mathrm{d}w) = \int wKP_{A(\cdot)y-b(\cdot)}(\mathrm{d}w)$$

$$= \int \left(\int wK(z, \mathrm{d}w) \right) P_{A(\cdot)y-b(\cdot)}(\mathrm{d}z)$$

$$= \int zP_{A(\cdot)y-b(\cdot)}(\mathrm{d}z) = E(A(\omega)y - b(\omega)).$$

(b) By Jensen's inequality we obtain

$$F(x) = \int u(z)P_{A(\cdot)x-b(\cdot)}(dz) = \int u(z)KP_{A(\cdot)y-b(\cdot)}(dz)$$

$$= \int \left(\int u(z)K(w, dz) \right) P_{A(\cdot)y-b(\cdot)}(dw)$$

$$\geq \int u\left(\int zK(w, dz) \right) P_{A(\cdot)y-b(\cdot)}(dw)$$

$$= \int u(w)P_{A(\cdot)y-b(\cdot)}(dw) = F(y).$$

Remark. It is easily seen that the second statement in Lemma 3.1 is a special case of the above theorem.

For this observe that the Markov kernel $K(z, \cdot) = \varepsilon_z$, where ε_z is the one-point measure defined by

$$\varepsilon_z(B) = \begin{cases} 1 & \text{if } z \in B, \\ 0 & \text{else} \end{cases}$$

is a dilatation and that

$$P_{A(\cdot)x-b(\cdot)} = KP_{A(\cdot)y-b(\cdot)}$$

is equivalent to

$$P_{A(\cdot)x-b(\cdot)} = P_{A(\cdot)y-b(\cdot)}.$$

Let for fixed x and y the stochastic variables X and Y be defined by $X(\omega) = A(\omega)x - b(\omega)$ and $Y(\omega) = A(\omega)y - b(\omega)$. Then the conditional distribution $P_{X/Y=z}(B)$ of X under the hypothesis that $Y = z$ has the property

$$\int P_{X/Y=z}(B) P_Y(\mathrm{d}z) = P_X(B) \quad \text{for all } B \in \mathcal{B}^m$$

(see e.g. [2]). Furthermore,

$$E(X/Y=z) = \int x P_{X/Y=z}(\mathrm{d}x) \quad P_Y\text{-a.s.},$$

where $E(X/Y=z)$ is the conditional expectation of X under the hypothesis that $Y = z$.

So $P_{X/Y=z}(B)$ is a dilatation if and only if

$$E(X/Y=z) = z \quad P_Y\text{-a.s.}$$

We have proved this

Corollary 4.1. *If*

$$E(A(\omega)x - b(\omega)/A(\omega)y - b(\omega)) = A(\omega)y - b(\omega) \quad \mu\text{-a.s.},$$

then the statements of Theorem 4.1 *are valid.*

In a forthcoming paper the situation of Theorem 4.1 is discussed in the case of a discrete distribution μ.

For another application assume that the kernel $K(z, \cdot)$ and the measures $P_{A(\cdot)x-b(\cdot)}$, $P_{A(\cdot)y-b(\cdot)}$ have densities $k(z, w)$, $p_x(z)$ and $p_y(z)$ with respect to the Lebesgue-Borel-measure on $\mathbb{R}^m$. Then the following conditions must be satisfied:

$$p_x(w) = \int k(z, w) p_y(z)\, \mathrm{d}z \quad \text{a.s.}, \tag{32}$$

$$\int k(z, w)\, \mathrm{d}w = 1 \quad \text{for all } z \in \mathbb{R}^m,$$

$$k(z, w) \geq 0 \quad \text{for all } (z, w) \in \mathbb{R}^m \times \mathbb{R}^m, \tag{33}$$

$$\int w k(z, w)\, \mathrm{d}w = z \quad \text{for all } z \in \mathbb{R}^m. \tag{34}$$

Here conditions (33) guarantee that $k(z, \cdot)$ is a density for each fixed $z \in \mathbb{R}^m$, (32) is equivalent to (31) and (34) means that $K(z, \cdot)$ is a dilatation.

Let us further simplify the situation by assuming that we have a translation model, i.e.,

$$k(z, w) = \phi(w - z) \quad \text{for } w, z \in \mathbb{R}^m,$$

where ϕ is a probability density on $\mathbb{R}^m$. Then (32) has the form

$$p_x(w) = \int \phi(w - z) p_y(z)\, dz = (\phi * p_y)(w) \quad \text{a.s.},$$

hence

$$p_x = \phi * p_y \quad \text{a.s.}, \tag{35}$$

where $\phi * p_y$ denotes the convolution of the densities ϕ and p_y. Now we have this

Corollary 4.2. *Assume that ϕ is a probability density on $\mathbb{R}^m$ with*

$$\int z\phi(z)\, dz = 0.$$

If for the Fourier-transforms $\hat{p}_x$, $\hat{p}_y$, $\hat{\phi}$ of p_x, p_y and ϕ, respectively, the equation

$$\hat{p}_x = \hat{\phi}\hat{p}_y \tag{36}$$

holds, then $F(y) \leqslant F(x)$.

Proof. We show that K, given by $K(z, B) = \int_B \phi(w - z)\, dw$, is a dilatation, which satisfies (31). Then the assertion of the Corollary follows from Theorem 4.1. It is sufficient to show that $k(z, w) = \phi(w - z)$ satisfies (32)-(34).

By the well known relation between the convolution of two functions and their Fourier-transforms, (36) is equivalent to (35) and therefore also to (32). Furthermore, (33) is true because ϕ is a probability density.

Finally,

$$\int wk(z, w)\, dw = \int w\phi(w - z)\, dw = \int (w - z)\phi(w - z)\, dw + z \int \phi(w - z)\, dw = z,$$

so (34) holds, too.

The above corollary suggests, that if x is given, then one has to choose y such that $\hat{p}_x / \hat{p}_y$ is the characteristic function $\hat{\phi}$ of a probability density ϕ, which has mean zero.

Example 1. Assume that P is normal distribution. Then also $P_{A(\cdot)x - b(\cdot)}$ and $P_{A(\cdot)y - b(\cdot)}$ are normal distributions with e.g.

$$\hat{p}_x(z) = \exp(i(\bar{A}x - \bar{b})'z - \tfrac{1}{2} z' Q_x z),$$

where $(\bar{A}, \bar{b}) = E(A(\omega), b(\omega))$ and $Q_x = \operatorname{var}(A(\cdot)x - b(\cdot))$.

In this situation we can give a theorem which generalizes some discussions in Section 3.

Theorem 4.2. *Let x and y be related by*

$$\bar{A}x = \bar{A}y, \tag{37}$$

$$Q_x \geq Q_y. \tag{38}$$

Then $F(x) \geq F(y)$. Here $Q_x \geq Q_y$ means that $Q_x - Q_y$ is positive semidefinite.

Proof. By the above considerations,

$$\frac{\hat{p}_x(z)}{\hat{p}_y(z)} = \exp[-\tfrac{1}{2}(z'(Q_x - Q_y)z)] = \hat{\phi}(z).$$

So $\hat{\phi}(z)$ is the Fourier-transform of a density of a normal distribution with mean zero and covariance matrix $\Lambda := Q_x - Q_y$.

When we want to apply Theorem 4.2 the problem is to find for a given x a y such that conditions (37) and (38) are satisfied. In the special case when the rows of $(A(\omega), b(\omega))$ are stochastically independent, then $Q_x - Q_y$ is the diagonal matrix

$$Q_x - Q_y = ((\hat{x}'Q_{ii}\hat{x} - \hat{y}'Q_{ii}\hat{y})\delta_{ij})_{i,j=1,\ldots,m}$$

for any choice of x and y. Here again $\hat{x} = \binom{x}{-1}$ and $\delta_{ij} = 0$, $i \neq j$, $\delta_{ii} = 1$. Hence, in this case we have to choose y such that

$$\hat{x}'Q_{ii}\hat{x} \geq \hat{y}'Q_{ii}\hat{y} \quad \text{for } i = 1, \ldots, m,$$

subject to $\bar{A}x = \bar{A}y$.

In the general case denote by $\underline{\lambda}(R)$ and $\bar{\lambda}(R)$ the smallest and the greatest eigenvalue of a symmetric matrix R, respectively. Then y is a solution of $Q_x \geq Q_y$ if and only if

$$\underline{\lambda}(Q_x - Q_y) \geq 0. \tag{39}$$

So we have the problem

$$\max_{y \in \mathbf{R}^n} \underline{\lambda}(Q_x - Q_y) \quad \text{subject to } \bar{A}x = \bar{A}y. \tag{40}$$

Because $y = x$ is a feasible point for problem (40) we have that for any solution y of (40) also conditions (37), (38) are satisfied.

If $\underline{\lambda}(Q_x) - \bar{\lambda}(Q_y) \geq 0$ for a point y with $\bar{A}x = \bar{A}y$, then y is a solution of (37), (38).

Remark. The optimization problem (40) may be very complicated in practice. However, we observe the following:

(A) The smallest eigenvalue $\underline{\lambda}(Q_x - Q_y)$ of $Q_x - Q_y$ can be computed numerically by one of the very fast standard eigenvalue routines!

(B) In many important special cases $\underline{\lambda}(Q_x - Q_y)$ can be given explicitly, as e.g. in the following examples.

(i) If $(A(\omega), b(\omega))$ has stochastically independent rows (including the important special case $m = 1$), then

$$\underline{\lambda}(Q_x - Q_y) = \min_{1 \le i \le m} \hat{x}' Q_{ii} \hat{x} - \hat{y}' Q_{ii} \hat{y}.$$

Hence, in this case (40) turns out to be a convex programming problem!

(ii) If the covariance matrix Q of $(A(\omega), b(\omega))$ is a block-tridiagonal matrix, i.e. if $Q_{ij} = 0$ for $|j - i| \ge 2$, and furthermore

$$Q_{11} = Q_{22} = \cdots = Q_{mm} \quad \text{and} \quad Q_{12} = Q_{23} = \cdots = Q_{m-1,m},$$

then (cf. [3])

$$\underline{\lambda}(Q_x - Q_y) = (\hat{x}' Q_{11} \hat{x} - \hat{y}' Q_{11} \hat{y}) - \left(2 \cos \frac{\pi}{m+1}\right) |\hat{x}' Q_{12} \hat{x} - \hat{y}' Q_{12} \hat{y}|.$$

Furthermore, several sufficient conditions for the decisive inequality (39) can be given, as e.g.:

(C) Let $\Delta = (\eta_{ij}) = Q_x - Q_y$ and denote by $D_k = D_k(x, y)$, $k = 1, \ldots, m$, the determinant of the $k \times k$ submatrix $\Delta_{(k)} = (\eta_{ij})_{i,j=1,\ldots,k}$ of Δ. Then (39) is implied by the m inequalities

$$D_k(x, y) \ge 0, \quad k = 1, \ldots, m.$$

(D) According to Gersgorin's circle theorem, inequality (39) is implied also by the m inequalities

$$\sum_{j \ne i}^{m} |\hat{x}' Q_{ij} \hat{x} - \hat{y}' Q_{ij} \hat{y}| \le \hat{x}' Q_{ii} \hat{x} - \hat{y}' Q_{ii} \hat{y}, \quad i = 1, 2, \ldots, m.$$

Example 2. Assume that the distribution P of $(A(\omega), b(\omega))$ is a stable distribution of order s, given by its characteristic function

$$\hat{P}(M) = \exp\left(\mathrm{i} \cdot \operatorname{tr} M \Xi' - \tfrac{1}{2} \sum_{\sigma=1}^{s} ((\operatorname{vec} M) Q(\sigma) (\operatorname{vec} M)')^{\alpha(\sigma)/2}\right), \tag{41}$$

see [12]. Here vec M means the row vector $(M_1, M_2, \ldots, M_m)$ having $m(n+1)$ components and 'tr' denotes the trace of a matrix.

If the $(n+1) \times (n+1)$ blocks of $Q(\sigma)$ are designated by $Q_{ij}(\sigma)$, $i, j = 1, \ldots, m$, i.e.

$$Q(\sigma) = (Q_{ij}(\sigma))_{i,j=1,\ldots,m},$$

then define

$$Q_x(\sigma) = (\hat{x}' Q_{ij}(\sigma) \hat{x})_{i,j=1,\ldots,m}.$$

Now the characteristic function of $P_{A(\cdot)x - b(\cdot)}$ may be written in the form

$$\hat{P}_{A(\cdot)x - b(\cdot)}(z) = \exp\left(\mathrm{i} z' \Xi \hat{x} - \tfrac{1}{2} \sum_{\sigma=1}^{s} (z' Q_x(\sigma) z)^{\alpha(\sigma)/2}\right).$$

Hence, if x and y are related by $\Xi\hat{x} = \Xi\hat{y}$, where Ξ is a given fixed $m \times (n+1)$ matrix, then

$$\frac{\hat{P}_{A(\cdot)x - b(\cdot)}}{\hat{P}_{A(\cdot)y - b(\cdot)}} = \exp\left(-\tfrac{1}{2} \sum_{\sigma=1}^{s} ((z'Q_x(\sigma)z)^{\alpha(\sigma)/2} - (z'Q_y(\sigma)z)^{\alpha(\sigma)/2})\right). \tag{42}$$

By Corollary 4.2 we have this

Theorem 4.3a. *Let the distribution P of $(A(\omega), b(\omega))$ be given by* (41), *where Ξ is a fixed $m \times (n+1)$ matrix and $1 < \alpha(\sigma) \leq 2$, $\sigma = 1, \ldots, s$. Furthermore, assume that $Q(\sigma)$ is a fixed $m(n+1) \times m(n+1)$ positive semidefinite matrix such that $Q_{ij}(\sigma) = 0$ for all $i, j = 1, \ldots, m$, $i \neq j$ and $\sigma = 1, \ldots, s$. If, for a given n-vector x, y is chosen such that*

$$\begin{aligned} &\Theta x = \Theta y \quad (\text{where } \Xi = (\Theta, \theta)), \\ &\hat{x}'Q_{ii}(\sigma) \geq \hat{y}'Q_{ii}(\sigma)\hat{y} \quad \text{for } i = 1, \ldots, m, \ \sigma = 1, \ldots, s, \end{aligned} \tag{43}$$

Proof. In our case the right hand side of (42) is given by

$$\begin{aligned} &\exp\left(-\tfrac{1}{2} \sum_{\sigma=1}^{s} \sum_{i=1}^{m} ((z_i^2\hat{x}'Q_{ii}(\sigma)\hat{x})^{\alpha(\sigma)/2} - (z_i^2\hat{y}'Q_{ii}(\sigma)\hat{y})^{\alpha(\sigma)/2})\right) \\ &= \exp\left(-\tfrac{1}{2} \sum_{\sigma=1}^{s} \sum_{i=1}^{m} |z_i|^{\alpha(\sigma)}((\hat{x}'Q_{ii}(\sigma)\hat{x})^{\alpha(\sigma)/2} - (\hat{y}'Q_{ii}(\sigma)\hat{y})^{\alpha(\sigma)/2})\right) \\ &= \exp\left(-\tfrac{1}{2} \sum_{\sigma=1}^{s} \sum_{i=1}^{m} (z_i^2((\hat{x}'Q_{ii}(\sigma)\hat{x})^{\alpha(\sigma)/2} - (\hat{y}'Q_{ii}(\sigma)\hat{y})^{\alpha(\sigma)/2})^{2/\alpha(\sigma)})^{\alpha(\sigma)/2}\right) \\ &= \exp\left(-\tfrac{1}{2} \sum_{\sigma=1}^{s} (z'\Delta(\sigma)z)^{\alpha(\sigma)/2}\right) \\ &= \prod_{\sigma=1}^{s} \exp(-\tfrac{1}{2}(z'\Delta(\sigma)z)^{\alpha(\sigma)/2}), \end{aligned}$$

where $\Delta(\sigma)$ is the matrix

$$\Delta(\sigma) = (((\hat{x}'Q_{ii}(\sigma)\hat{x})^{\alpha(\sigma)/2} - (\hat{y}'Q_{ii}(\sigma)\hat{y})^{\alpha(\sigma)/2})^{2/\alpha(\sigma)}\delta_{ij}).$$

From (43) follows that $\Delta(\sigma)$ is a positive semidefinite matrix for each $\sigma = 1, \ldots, s$. So $z \to \exp(-\frac{1}{2}(z'\Delta(\sigma)z)^{\alpha(\sigma)/2})$ is the characteristic function of a probability distribution. Since $1 < \alpha(\sigma) \leq 2$, this distribution has mean zero. Hence, the right hand side of (44) is the Fourier-transform of a convolution of s probability distributions all having mean zero. This means that it is the characteristic function of a distribution with mean zero. Now the assertion follows by Corollary 4.2.

Theorem 4.3b. *Suppose that $Q_x(\sigma) = q_x(\sigma)R(\sigma)$, where $q_x(\sigma) \geq 0$ and $R(\sigma)$ is an $m \times m$ matrix. If n-vectors x, y are related such that*

$$\Theta x = \Theta y, \qquad q_x(\sigma) \geq q_y(\sigma), \quad \sigma = 1, \ldots, s,$$

then $F(x) \geq F(y)$.

Proof. In the present case the right hand side of (42) is given by

$$\exp\left(-\tfrac{1}{2}\sum_{\sigma=1}^{s}\left((q_x(\sigma)z'R(\sigma)z)^{\alpha(\sigma)/2}-(q_y(\sigma)z'R(\sigma)z)^{\alpha(\sigma)/2}\right)\right)$$

$$=\exp\left(-\tfrac{1}{2}\sum_{\sigma=1}^{s}(z'\Delta(\sigma)z)^{\alpha(\sigma)/2}\right),$$

where the matrix $\Delta(\sigma)$ is defined by

$$\Delta(\sigma)=(q_x(\sigma)^{\alpha(\sigma)/2}-q_y(\sigma)^{\alpha(\sigma)/2})^{2/\alpha(\sigma)}R(\sigma).$$

Now the assertion follows as in the first part of this theorem.

Acknowledgment

The authors are indebted to the referees for their valuable comments and helpful suggestions.

References

[1] K.J. Aström, *Introduction to stochastic control theory* (Academic Press, New York, London, 1975).
[2] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie* (De Gruyter, Berlin, 1968).
[3] R. Bellman, *Introduction to matrix analysis* (McGraw-Hill, New York, London, Toronto, 1960).
[4] D. Blackwell, "Equivalent comparison of experiments", *Annals of Mathematical Statistics* 24 (1953) 265–272.
[5] E. Blum and W. Oettli, *Mathematische Optimierung* (Springer-Verlag, Berlin, Heidelberg, New York, 1975).
[6] Y. Ermoliev, "Stochastic quasigradient methods and their application to system optimization", *Stochastics* 9 (1983) 1–36.
[7] Y. Ermoliev and A. Gaivoronski, "Stochastic quasigradient methods and their implementation", Working Paper, IIASA Laxenburg (1983).
[8] J.B. Hiriart-Urruty, "Contributions à la programmation mathématique: Cas déterministe et stochastique", Dissertation, University of Clermont-Ferrand II (1977).
[9] K. Marti, "Stochastische Dominanz und Stochastische lineare Programme", *Methods of Operations Research* 23 (1977) 141–160.
[10] K. Marti, "Stochastic linear programs with stable distributed random variables", in: J. Stoer, ed., *Optimization techniques, Part 2*, Lecture Notes in Control and Information Sciences 7 (Springer-Verlag, Berlin, Heidelberg, New York, 1978) pp. 76–86.
[11] K. Marti, "Approximationen stochastischer Optimierungsprobleme", *Mathematical Systems in Economics* 43 (A. Hain, Königstein/Ts., 1979).
[12] K. Marti, "Solving stochastic linear programs by semi-stochastic approximation algorithms", in: P. Kall and A. Prekopa, eds., *Recent results in stochastic programming*, Lecture Notes in Economics and Mathematical Systems 179 (Springer-Verlag, Berlin, Heidelberg, New York, 1980) pp. 191–213.
[13] K. Marti, "Computation of descent directions in stochastic optimization problems with invariant distributions", to appear in *Zeitschrift für angewandte Mathematik und Mechanik* (1985).
[14] K. Marti and R.J. Riepl, "Optimale Portefeuilles mit stabil verteilten Renditen", *Zeitschrift für angewandte Mathematik und Mechanik* 57 (1977) T337–T339.
[15] H.W. Sorenson, *Parameter estimation* (Marcel Dekker, New York, Basel, 1980).

[16] D. Stoyan, *Qualitative Eigenschaften und Abschätzungen stochastischer Modelle* (R. Oldenbourg, München, Wien, 1977).

[17] R.G. Vickson and W.T. Ziemba, eds., *Stochastic optimization models in finance* (Academic Press, New York, London, 1970).

[18] R.G. Vickson, "Stochastic orderings from partially known utility functions", *Mathematics of Operations Research* 2 (1977) 244–252.

[19] R. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming - The state of the art* (Springer-Verlag, Berlin, Heidelberg, New York, 1983) pp. 566–603.

Mathematical Programming Study 28 (1986) 157-181
North-Holland

LINEARIZATION METHODS FOR OPTIMIZATION OF FUNCTIONALS WHICH DEPEND ON PROBABILITY MEASURES

A. GAIVORONSKI

International Institute for Applied Systems Analysis, Laxenburg, Austria, and Glushkov Institute of Cybernetics, Kiev, USSR

Received 2 July 1984
Revised manuscript received 14 August 1984

The main purpose of this paper is to discuss numerical optimization procedures for problems in which both the objective function and the constraints depend on distribution functions. The objective function and constraints are assumed to be nonlinear and to have directional derivatives. The proposed algorithm is based on duality relations between the linearized problem and some special finite-dimensional minimax problem and is of the feasible-direction type. The resulting minimax problem is solved using the cutting-plane technique.

Key words: Optimization in the Space of Probability Measures, Generalized Linear Programming, Minimax Problems.

1. Introduction

The purpose of this paper is to present numerical methods for solving optimization problems in which both the objective function and the constraints depend on distribution functions. Such problems occur in stochastic programming [8, 9], reliability theory and various branches of operations research (surveyed in [15]), and robust statistics [17], among others.

In what follows we discuss two examples of this type of problem in more detail.

Stochastic programming problems with partially-known distribution functions

Stochastic programming problems may be formulated quite generally as:

$$\max_{x \in X} E_y F(x, y) = \max_{x \in X} \int_Y F(x, y)\, \mathrm{d}H(y)$$

where the x are controlled variables which belong to feasible set X and the y are random parameters with distribution function $H(y)$. It is usually assumed that the distribution function $H(y)$ is known. We obtain this information from some set of observations $y_1, \ldots, y_k, \ldots$ of the random parameters y, using statistical techniques. However, in most cases this information is not sufficient to define a unique distribution; we estimate instead some statistical properties such as the mean or the variance of the distribution $H(y)$. In this case we can only be sure that $H(y) \in G$ where the

set of distributions G is defined by

$$G = \left\{ H: \int_Y q^i(y)\, dH(y) \leq 0,\ i = \overline{1, m} \right\},$$

and $q^i(y)$ are known functions. In this case it is natural to use the minimax approach [5, 6, 25] and instead of solving the original problem to consider:

$$\max_{x \in X} \min_{H \in G} \int_Y F(x, y)\, dH(y)$$

where the inner problem is one of minimizing an integral functional over a set of distribution functions (for details see [8, 9]). The other problem is to find lower and/or upper bounds on the probability that a random variable with distribution function $H(y)$ takes values in a specified set Z when $H(y)$ is not known exactly. In this problem both the objective function and the constraints depend linearly on the distribution function:

$$\min \int_Y q^0(y)\, dH(y), \tag{1}$$

$$\int_Y q^i(y)\, dH(y) \leq 0, \quad i = \overline{1, m}, \tag{2}$$

$$\int_Y dH(y) = 1, \tag{3}$$

where Y is a set in Euclidean space $\mathbb{R}^n$. In this particular case $q^0(y)$ is the indicator function of the set Z. Some other applications of this approach to stochastic programming problems have been studied in [1, 2, 5, 6, 25]. Some specific cases of problem (1)-(3) arise in the moment theory and can be solved analytically [18, 21]. However, the success of analytical methods is limited even in the linear case (1)-(3) and therefore numerical algorithms are necessary.

Design of optimal experiments

Much attention is paid in regression analysis to the following problem of parameter estimation:

$$z^i = \sum_{j=1}^{l} \varphi^j(y^i) a_j + \xi^i, \quad i = \overline{1, N},$$

where the z^i are observations, the y^i are controlled variables whose value can be chosen by the experimenter, the ξ^i are (in the simplest case) identically distributed independent random variables with zero mean, the φ^j are known functions and the a_j are the parameters to be estimated. The problem of optimal experiment design [11, 20, 24] is to choose the points y^i in some optimal way, that is, to minimize some characteristics of the variance matrix of the best linear estimate of parameters a_j.

This matrix equals

$$D(\hat{a}) = E[\hat{a}-a)(\hat{a}-a)^{\mathrm{T}}] = \left[\sum_{i=1}^{N} \varphi(y_i)\varphi^{\mathrm{T}}(y_i)\right]^{-1} E(\xi^i)^2 = M^{-1}E(\xi^i)^2,$$

where $\hat{a} = M^{-1}\sum_{i=1}^{N} \varphi(y_i)z^i$ is the best linear estimate, M is called information matrix, $\varphi(y) = (\varphi^1(y), \ldots, \varphi^l(y))$ and the superscript T indicates transposition. In general, the probability measure $H(y)$ is considered instead of the set of points y^i; these points are chosen later on the basis of this measure. The information matrix $M(H)$ then becomes:

$$M(H) = \int_Y \varphi(y)\varphi(y)^{\mathrm{T}}\,\mathrm{d}H(y)$$

where $\varphi(y) = (\varphi^1(y), \ldots, \varphi^l(y))$ and the superscript T indicates transposition. The problem is to minimize some function $\Psi(M)$, which may be the determinant of M^{-1}, the trace of M^{-1} or some other characteristics. Very often it is desirable to impose additional constraints on the measure $H(y)$, such as bounds on its covariance matrix:

$$\int_Y yy^{\mathrm{T}}\,\mathrm{d}H(y) \leq Q.$$

Once again we have a problem in which both the objective function and the constraints depend on a distribution function; in this particular case the objective function is nonlinear.

Numerical methods for solving such problems began to appear in the middle of the last decade. One approach is to approximate set Y by a sequence of finite subsets $Y_1, Y_2, \ldots, Y_s, \ldots$, where

$$Y_s = (y_s^1, \ldots, y_s^{n_s}).$$

This sequence of sets has the property that

$$\sup_{z\in Y} \min_{y\in Y_s} \|y-z\| \to 0$$

as s tends to infinity, i.e., the greatest distance between points in set Y and set Y_s tends to zero. Let G_s be the set of all distribution functions which correspond to probability measures with support Y_s:

$$G_s = \left\{(y_s^1, p_1), \ldots, (y_s^{n_s}, p_{n_s}), \sum_{i=1}^{n_s} p_i = 1, p_i \geq 0\right\}.$$

If we include one more constraint,

$$H \in G_s, \tag{4}$$

problem (1)-(4) becomes a finite-dimensional linear programming problem; this raises the possibility of approximating the original problem (1)-(3) by a sequence

of finite-dimensional problems (1)-(4). The dual simplex method is then applied to this sequence in such a way that the solution of problem number s becomes the initial approximation to the solution of problem number $s+1$. This idea was explored in [15, 16], where it was also used to solve nonlinear and minimax problems involving distribution functions. However, this approach can be used only for sets Y of small dimension (in fact not exceeding three) due to the high dimensionality of the associated linear programming problems.

Another possible way of solving (1)-(3) is based on the duality relations between problem (1)-(3) and some finite-dimensional minimax problem which may be solved using convex programming methods. First proposed in [8], this idea was extended in [9, 10] using generalized linear programming techniques [3, 23].

The purpose of this paper is to use this last approach to develop solution techniques for nonlinear problems of the kind:

$$\min \Psi^0(H) \tag{5}$$

subject to

$$H \in G \tag{6}$$

where G is defined by:

$$G = \left\{ H: \Psi^i(H) \leq 0,\ i = \overline{1, m};\ \int_Y \mathrm{d}H(y) = 1 \right\}. \tag{7}$$

We propose an analogue of the linearization (Frank-Wolfe) method in which we are required to solve subproblems of type (1)-(3). It appears that it is not necessary to solve these subproblems precisely: duality relations make it possible to utilize rough solutions of (1)-(3) so that only a limited number of calculations are needed at each iteration. The case in which the constraints are linear, i.e., of type (2), was considered in [14].

2. Characterization of optimal distributions

We shall use the same letter, say H, to denote both the distribution function and the underlying probability measure, where this will not cause confusion. For a given probability measure H we shall denote by $B^+(H)$ the collection of all closed subsets A of Y such that $\int_A \mathrm{d}H(y) = 1$, and by supp H the support set of H, i.e.,

$$\text{supp } H = \bigcap_{A \in B^+} A.$$

Let us first impose some conditions on the functionals under consideration. What we actually need is some analogue of directional differentiability. Suppose that

$$\Psi^i(H_1 + \alpha(H_2 - H_1)) = \Psi^i(H_1) + \alpha \int_Y g^i(y, H_1)\, \mathrm{d}(H_2(y) - H_1(y)) + \tau^i(\alpha, H_1, H_2) \tag{8}$$

for $i=\overline{0, m}$ and all $H_1, H_2 \in G(Y)$, where

$$\tau(\alpha, H_1, H_2)/\alpha \to 0 \quad \text{as } \alpha \to 0$$

and

$$G(Y)=\{H: \operatorname{supp} H \subset Y\}.$$

In what follows we assume that functions $f^i(y, H)$ are such that expression (8) is meaningful.

The following simple conditions are necessary, and in the convex case also sufficient, for distribution H to be a solution of problem (5)-(7):

Lemma 1. *Suppose that $\Psi^0(H^*) \leq \Psi^0(H)$ for some $H^* \in G$ and all $H \in G$ and the following conditions are satisfied:*

1. *$f^i(y, H)$, $i=\overline{0, m}$ are bounded on Y for all $H \in G(Y)$.*
2. *For $\tau^0(\alpha, H_1, H_2)$ from (8) we have:*

$$|\tau^0(\alpha, H_1, H_2)| \leq \bar{\tau}(\alpha), \qquad 0 \leq \alpha \leq 1,$$

and $\bar{\tau}(\alpha)/\alpha \to 0$ as $\alpha \to 0$.

3. *$\Psi^i(H)$, $i=\overline{1, m}$ are convex, i.e.,*

$$\Psi^i(\alpha H_1+(1-\alpha)H_2) \leq \alpha \Psi^i(H_1)+(1-\alpha)\Psi^i(H_2)$$

for $H_1, H_2 \in G(Y)$ and $0 \leq \alpha \leq 1$. Also, there exists an $\tilde{H} \in G$ such that $\Psi^i(\tilde{H}) < -\sigma < 0$ for $i=\overline{1, m}$.

Then

$$z^* = \inf_{H \in G^*} \int_Y f^0(y, H^*)\, dH(y) = \int_Y f^0(y, H^*)\, dH^*(y), \tag{9}$$

where

$$G^* = \left\{ H: \int_Y f^i(y, H^*)\, dH(y) \leq \int_Y f^i(y, H^*)\, dH^*(y), \int_Y dH(y) = 1, i \in I^0 \right\},$$

$$I^0 = \{i: \Psi^i(H^*) = 0\}.$$

If, additionally, $\Psi^0(H)$ is convex and the distribution H^* satisfies (9) then H^* is the solution of problem (5)-(7).

Proof. The proof is of the traditional type for necessary conditions. Note that

$$z^* \leq \int_Y f^0(y, H^*)\, dH^*(y)$$

is always true. Suppose, arguing by contradiction, that there exists a $\gamma > 0$ such that

$$z^* - \int_Y f^0(y, H^*)\, dH^*(y) \leq -2\gamma.$$

Then there exists an $\bar{H} \in G^*$ such that

$$\int_Y f^0(y, H^*)\, d\bar{H}(y) - \int_Y f^0(y, H^*)\, dH^*(y) < -\gamma.$$

Now take $\tau(\alpha, H^*, \bar{H}) = \max\limits_{1 \leq i \leq m} |\tau^i(\alpha, H^*, \bar{H})|$ and consider distributions H_α:

$$H_\alpha = \alpha \bar{H} + (1-\alpha) H^*.$$

Take α such that $\tau(\alpha, H^*, \bar{H})/\sigma < 1$ and

$$\tilde{H}_\alpha = \left(1 - \frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) H_\alpha + \left(\frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) \tilde{H}.$$

Then for $i = \overline{1, m}$ we have

$$\begin{aligned}
\Psi^i(\tilde{H}_\alpha) &\leq \left(1 - \frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) \Psi^i(H_\alpha) - \tau(\alpha, H^*, \bar{H}) \\
&\leq \left(1 - \frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) [\Psi^i(H^*) \\
&\quad + \alpha \int_Y f^i(y, H^*)\, d(\bar{H} - H^*) + \tau^i(\alpha, H^*, \bar{H})] - \tau(\alpha, H^*, \bar{H}) \\
&\leq \left(1 - \frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) \left[\Psi^i(H^*) + \alpha \int_Y f^i(y, H^*)\, d(\bar{H} - H^*) \right].
\end{aligned}$$

Now we have $\Psi^i(\tilde{H}_\alpha) \leq 0$, $i = \overline{1, m}$, for sufficiently small α because $\int_Y f^i(y, H^*)\, d(\bar{H} - H^*) \leq 0$ for $i \in I^0$ and

$$\alpha \int_Y f^i(y, H^*)\, d(\bar{H} - H^*) \leq K\alpha, \quad K < \infty,$$

for $i \notin I^0$. Therefore $\tilde{H}_\alpha \in G$ for $\alpha \leq \bar{\alpha}$, $\bar{\alpha} > 0$. We shall now estimate the value of $\Psi^0(\tilde{H}_\alpha)$. Assumption 2 implies for sufficiently small α and some $C > 0$:

$$\begin{aligned}
\Psi^0(\tilde{H}_\alpha) &\leq \Psi^0(H_\alpha) + \frac{\tau(\alpha, H^*, \bar{H})}{\sigma} \int f^0(y, H_\alpha)\, d(\tilde{H} - H_\alpha) + \bar{\tau}\left(\frac{\tau(\alpha, H^*, \bar{H})}{\sigma}\right) \\
&\leq \Psi^0(H_\alpha) + C \frac{\tau(\alpha, H^*, \bar{H})}{\sigma} \\
&\leq \Psi^0(H^*) - \alpha\gamma + \tau^0(\alpha, H^*, \bar{H}) + C \frac{\tau(\alpha, H^*, \bar{H})}{\sigma}.
\end{aligned}$$

This last inequality gives $\Psi^0(\tilde{H}_\alpha) < \Psi^0(H^*)$ when α is sufficiently small, which contradicts the initial assumption and proves the first part of the lemma.

Now suppose that $\Psi^0(H)$ is convex and that (9) is satisfied for some distribution H^*, but that there exists an $\bar{H} \in G$ such that $\Psi^0(H^*) - \Psi^0(\bar{H}) \geq \gamma \geq 0$. The function

$F(\alpha) = \Psi^0(H^* + \alpha(\bar{H} - H^*))$ is convex for $0 \leq \alpha \leq 1$ and $F_\alpha(\alpha)$ exists and equals

$$F_\alpha(\alpha) = \int_Y f^0(y, H^* + \alpha(\bar{H} - H^*))\, \mathrm{d}(\bar{H} - H^*).$$

This gives

$$\int_Y f^0(y, H^*)\, \mathrm{d}\bar{H}(y) - \int_Y f^0(y, H^*)\, \mathrm{d}H^*(y) = F_\alpha(0)(1-0) \leq F(1) - F(0) \leq -\gamma.$$

which, taken together with the fact that $\bar{H} \in G^*$, contradicts the assumption that H^* satisfies (9). This completes the proof.

Lemma 1 implies that to check the necessary conditions for problem (5)-(7) at point H^* requires solution of a linear problem of the form (1)-(3), where $q^0(y) \equiv f^i(y, H^*)$ and $q^i(y) \equiv f^i(y, H^*) - c_i$, $c_i = \int_Y f^0(y, H^*)\, \mathrm{d}H^*(y)$. The solution of problem (1)-(3) can be characterized through the duality relations summarized in the following theorem, which was proved in [10].

Theorem 1. *Suppose that the following assumptions are satisfied:*

(a) *Set Y is compact and functions $q^i(y)$, $i = \overline{0, m}$, continuous on Y.*

(b) *There exists a distribution $\tilde{H}$ such that*

$$\int_Y q^i(y)\, \mathrm{d}\tilde{H}(y) < 0, \quad i = \overline{1, m}.$$

Then

1. *A solution of problem* (1)-(3) *exists and the optimal value of* $\int q^0(y)\, \mathrm{d}H(y)$ *is equal to the optimal value of the following minimax problem:*

$$\max_{u \in U^+} \bar{\varphi}(u), \quad \bar{\varphi}(u) = \min_{y \in Y} \left(q^0(y) + \sum_{i=1}^m u_i q^i(y) \right)$$

where $U^+ = \{u: u \in R^m, u_i \geq 0\}$.

2. *For any solution H^* of problem* (1)-(3) *there exists a $u^* \in U^+$ such that*

$$\bar{\varphi}(u^*) = \max_{u \in U^+} \bar{\varphi}(u), \qquad \operatorname{supp} H^* \subseteq Y^*(u^*)$$

where

$$Y^*(u^*) = \left\{ y: y \in Y, \bar{\varphi}(u^*) = q^0(y) + \sum_{i=1}^m u_i^* q^i(y) \right\}.$$

3. *There exists a finite set $Y_t = \{y^1, \ldots, y^t\}$, $t \leq m+1$, and a solution H^* of problem* (1)-(3) *such that* $\operatorname{supp} H^* = Y_t$, *i.e., the probability measure H^* can be expressed as a collection of $t \leq m+1$ pairs $\{(y^1, \bar{p}_1), \ldots, (y^t, \bar{p}_t)\}$.*

The probabilities $\bar{p}_1, \ldots, \bar{p}_t$ are solutions of the following linear programming problem:

$$\min_p \sum_{i=1}^t p_i q^0(y^i),$$

$$\sum_{i=1}^{t} p_i q^j(y^i) \leq 0, \quad j = \overline{1, m},$$

$$\sum_{i=1}^{t} p_i = 1, \quad p_i \geq 0, \quad i = \overline{1, t}.$$

Combining Theorem 1 and Lemma 1 we obtain the following result:

Theorem 2. *Suppose that* $\Psi^0(H^*) \leq \Psi^0(H)$ *for all* $H \in G$ *and that the conditions of Lemma* 1 *are satisfied. We make the additional assumptions*:

(a) *Set* $Y \subset \mathbb{R}^n$ *is compact.*

(b) *Functions* $f^i(y, H^*)$, $i = \overline{0, m}$ *are continuous on* Y.

Then

1. *We have*

$$\int_Y f^0(y, H^*)\, dH^*(y) = \max_{u \in U^+} \varphi(u)$$

where

$$\varphi(u) = \min_{y \in Y} \left(f^0(y, H^*) + \sum_{i \in I^0} u_i f^i(y, H^*) \right) - \sum_{i \in I^0} u_i c_i,$$

$$c_i \equiv \int_Y f^i(y, H^*)\, dH^*(y).$$

2. *There exists a* u^*, $\varphi(u^*) = \max_{u \in U^+} \varphi(u)$, $u^* \in U^+$ *such that*

$$\operatorname{supp} H^* \subseteq Y^*(u^*)$$

where

$$Y^*(u^*) = \left\{ y : y \in Y, \varphi(u^*) = f^0(y, H^*) + \sum_{i \in I^0} u_i^* f^i(y) - \sum_{i \in I^0} u_i^* c_i \right\}.$$

The situation that arises when the constraints are linear deserves special attention. Suppose that the feasible set G is defined in the following way:

$$G = \left\{ H : \int_Y q^i(y)\, dH(y) = 0 \text{ for } i = \overline{1, m_1}, \int_Y q^i(y)\, dH(y) \leq 0 \text{ for } i = \overline{m_1, m} \right\} \tag{10}$$

and we are trying to solve the problem $\min_{H \in G} \Psi^0(H)$.

Theorem 2a. *Suppose that* $\Psi^0(H^*) \leq \Psi^0(H)$ *for all* $H \in G$ *and the following assumptions are satisfied*:

(a) *Set* Y *is compact and functions* $f^0(y, H^*)$, $q^i(y)$, $i = \overline{1, m}$ *are continuous on* Y.

(b) $0 \in \operatorname{int} \operatorname{co} Z$ *where*

$$Z = \{z: z = (q^1(y), \ldots, q^m(y)), y \in Y\}.$$

Then

1. *We have*

$$\int_Y f^0(y, H^*)\, dH^*(y) = \max_{u \in \bar{U}^+} \varphi(u)$$

where

$$\varphi(u) = \min_{y \in Y} \left(f^0(y, H^*) + \sum_{i=1}^m u_i q^i(y) \right),$$

$$\bar{U}^+ = \{u: u \in \mathbb{R}^m, u_i \geq 0 \text{ for } i = \overline{m_1, m}\}.$$

2. *There exists a* u^*, $\varphi(u^*) = \max_{u \in \bar{U}^+} \varphi(u)$, $u^* \in \bar{U}^+$ *such that*

$$\operatorname{supp} H^* \subseteq Y^*(u^*)$$

where

$$Y^*(u^*) = \left\{ y: y \in Y, \varphi(u^*) = f^0(y, H^*) + \sum_{i=1}^m u_i^* q^i(y) \right\}.$$

This last theorem is considered in more detail in [14].

3. Linearization method

It is now possible to construct a method which finds points satisfying the necessary conditions of Lemma 1 or, in the convex case, global minima. This method is of the linearization type [13].

Algorithm 1

1. Begin with an initial distribution H^1.
2. Suppose we have an approximate solution H^s before starting iteration number s. Then at the sth iteration we do the following:

(i) Find a distribution $\bar{H}^s$ such that

$$\int_Y f^0(y, H^s)\, d\bar{H}^s(y) \leq z_s + \varepsilon_s$$

where

$$z_s = \inf_H \int_Y f^0(y, H^s)\, dH(y) \tag{11}$$

subject to constraints

$$\Psi^i(H^s)+\int_Y f^i(y, H^s)\,\mathrm{d}(H-H^s)\leq 0, \qquad i=\overline{1, m}, \tag{12}$$

and $\varepsilon_s>0$ is the accuracy with which problem (11)-(12) is to be solved. It is not necessary to know the value of ε_s, only that $\varepsilon_s\to 0$ as $s\to 0$.

(ii) Choose a stepsize ρ_s: $0<\rho_s\leq 1$ and calculate a new approximation to the optimal solution:

$$H^{s+1}=(1-\rho_s)H^s+\rho_s\bar{H}^s. \tag{13}$$

Then go to step (i).

Remark. The stepsize can be chosen according to a number of different rules:

$$\text{(a)} \qquad \rho_s\to 0, \ \sum_{s=0}^{\infty}\rho_s=\infty. \tag{14}$$

In the case where the constraints are linear (see (2)) and the initial point is a member of G, the following two line-search techniques could be used:

$$\text{(b)} \qquad \rho_s=\arg\min_{\alpha\geq 0}\Psi^0(H^s+\alpha(\bar{H}^s-H^s)). \tag{15}$$

(c) Take a sequence α_s, where

$$\alpha_s\to 0, \quad \sum_{s=0}^{\infty}\alpha_s=\infty$$

and

$$\rho_s=\min\{\arg\min_{\alpha\geq 0}\Psi^0(H^s+\alpha(\bar{H}^s-H^s)), \alpha_s\}. \tag{16}$$

If the constraints are nonlinear analogous methods could be used.

We shall now prove the convergence of the algorithm given above.

Theorem 3. *Suppose that the following statements are true:*

1. *Functions* $\Psi^i(H)$, $i=\overline{0, m}$ *satisfy* (8) *where*

$$|\tau^i(\alpha, H_1, H_2)|\leq\tilde{\tau}(\alpha)<\infty, \quad 0\leq\alpha\leq 1, \qquad \tilde{\tau}(\alpha)/\alpha\to 0 \quad \text{as } \alpha\to 0,$$

and are convex.

2. *There exists a distribution* $\tilde{H}$ *such that* $\Psi^i(\tilde{H})<-\sigma<0$ *for* $1\leq i\leq m$.
3. $|f^i(y, H)|\leq K<\infty$ *for* $y\in Y$, $H\in G(Y)$.
4. $\varepsilon_s\to 0$.
5. $\rho_s\geq 0$, $\rho_s\to 0$, $\sum_{s=0}^{\infty}\rho_s=\infty$.

Then $\lim_{s\to\infty}\Psi^0(H^s)=\inf_{H\in G}\Psi^0(H)$ *and* $\overline{\lim}_{s\to\infty}\Psi^i(H^s)\leq 0$ *for* $1\leq i\leq m$.

Proof. 1. Let $\bar{H}$ be an arbitrary distribution such that $\Psi^i(\bar{H})\leq 0$ for all $1\leq i\leq m$.

If we now consider the convex function $F^i(\alpha) = \Psi^i(H^s + \alpha(\bar{H} - H^s))$ we have

$$F^i(\alpha_1) - F^i(\alpha_2) \leq F^i_\alpha(\alpha_1)(\alpha_1 - \alpha_2)$$

$$= (\alpha_1 - \alpha_2) \int_Y f^i(y, H^s + \alpha_1(\bar{H} - H^s))\, \mathrm{d}(\bar{H} - H^s)$$

which, after setting $\alpha_1 = 0$, $\alpha_2 = 1$, gives:

$$\Psi^i(H^s) + \int_Y f^i(y, H^s)\, \mathrm{d}(\bar{H} - H^s) \leq \Psi^i(\bar{H}).$$

Therefore all $H \in G$ satisfy constraints (12), problem (11)-(12) always has a solution and the algorithm is defined correctly.

2. We shall now prove that $\overline{\lim}_{s\to\infty} \Psi^i(H^s) \leq 0$ for all $1 \leq i \leq m$. From (8) we have the estimate

$$\Psi^i(H^{s+1}) = \Psi^i(H^s + \rho_s(\bar{H}^s - H^s))$$

$$\leq \Psi^i(H^s) + \rho_s \int_Y f^i(y, H^s)\, \mathrm{d}(\bar{H}^s - H^s) + \tilde{\tau}(\rho_s)$$

$$\leq (1 - \rho_s)\Psi^i(H^s) + \tilde{\tau}(\rho_s).$$

Since $\tilde{\tau}(\rho_s)/\rho_s \to 0$ we have for any $\gamma > 0$ an $\bar{s}$ such that

$$\Psi^i(H^{s+1}) < (1 - \rho_s/2)\Psi^i(H^s) \quad \text{if } s > \bar{s} \text{ and } \Psi^i(H^s) \geq \gamma,$$

$$\Psi^i(H^{s+1}) < \gamma(1 + \rho_s/2) \quad \text{if } s > \bar{s} \text{ and } \Psi^i(H^s) < \gamma.$$

This and the fact that $\prod_s^\infty (1 - \rho_s/2) = 0$ since $\sum_{i=0}^\infty \rho_i = \infty$ finally show that $\max\{0, \Psi^i(H^s)\} \to 0$.

3. Consider

$$\gamma_s = \int_Y f^0(y, H^s)\, \mathrm{d}H^s(y) - z_s, \qquad \tilde{\gamma}_s = \Psi^0(H^s) - \inf_{H\in G} \Psi^0(H),$$

where z_s is defined in (11). We shall now derive some useful relations between γ_s and $\tilde{\gamma}_s$. Take arbitrary $\varepsilon > 0$ and $H_\varepsilon \in G$ such that $\Psi^0(H_\varepsilon) - \inf_{H\in G} \Psi^0(H) < \varepsilon$ and consider the convex function $F^0(\alpha) = \Psi^0(H^s + \alpha(H_\varepsilon - H^s))$.

Just as in the first part of the proof we obtain

$$-\gamma_s \leq \int_Y f^0(y, H^s)\, \mathrm{d}(H_\varepsilon - H^s) \leq \Psi^0(H_\varepsilon) - \Psi^0(H^s) \leq -\tilde{\gamma}_s + \varepsilon$$

which gives $\gamma_s \geq \tilde{\gamma}_s$ because ε is an arbitrary positive number. Thus, for any sequence s_k such that $\max\{0, \gamma_{s_k}\} \to 0$ we have

$$\max\{0, \Psi^0(H^{s_k}) - \inf_{H\in G} \Psi^0(H)\} \to 0$$

and $\gamma_s > \delta$ whenever $\tilde{\gamma}_s > \delta$.

4. Let us now prove that $\tilde{\gamma}_s \to 0$, which will complete the proof of the theorem. We begin with the following estimate:

$$\Psi^0(H^{s+1}) = \Psi^0(H^s + \rho_s(\bar{H}^s - H^s))$$

$$\leq \Psi^0(H^s) + \rho_s \int_Y f^0(y, H^s)\, \mathrm{d}(\bar{H}^s - H^s) + \tilde{\tau}(\rho_s)$$

$$\leq \Psi^0(H^s) - \rho_s(\gamma_s - \varepsilon_s - \bar{\tau}(\rho_s)) \tag{17}$$

where $\bar{\tau}(\rho_s) \to 0$ as $\rho_s \to 0$; $\varepsilon_s \to 0$, due to assumptions 1 and 4 of the theorem. Suppose now that there exists an $\bar{s}$ such that for $s > \bar{s}$

$$\gamma_s > \varepsilon_s + \alpha + \bar{\tau}(\rho_s)$$

where $\alpha > 0$. Now from (17) we get:

$$\Psi^0(H^{s+1}) \leq \Psi^0(H^s) - \alpha\rho_s. \tag{18}$$

Summing (18) from $s > \bar{s}$ to k we obtain:

$$\Psi^0(H^k) \leq \Psi^0(H^s) - \alpha \sum_{i=s}^{k-1} \rho_i.$$

Observe now that $|\Psi^0(H)| < K_1$ for some $K_1 < \infty$ and all $H \in G(Y)$ from assumptions 1 and 3 of the theorem. This contradicts the previous inequality because $\sum_{i=1}^{\infty} \rho_i = \infty$. Therefore a subsequence n_k exists such that

$$\max\{0, \gamma_{n_k} - \varepsilon_{n_k} - \bar{\tau}(\rho_{n_k})\} \to 0.$$

From part 3 of the proof we have $\max\{0, \Psi^0(H^{n_k}) - \inf_{H \in G} \Psi^0(H)\} \to 0$. Now suppose that there is a subsequence m_k such that $\tilde{\gamma}_{m_k} > 2\alpha > 0$ for some $\alpha > 0$. We may assume without loss of generality that

$$n_k < m_k < n_{k+1} < m_{k+1} < \cdots.$$

Let us take a sequence l_k such that

$$n_k \leq l_k \leq m_k, \qquad \tilde{\gamma}_i \geq \alpha \quad \text{for } l_k \leq i \leq m_k, \qquad \tilde{\gamma}_{l_k - 1} < \alpha.$$

We deduce from (17) that $\max\{0, \Psi^0(H^{s+1}) - \Psi^0(H^s)\} \to 0$ as $s \to \infty$ and therefore $\tilde{\gamma}_{l_k}$ exists, which satisfies previous inequalities and $\tilde{\gamma}_{l_k} < 2\alpha$ for k sufficiently large. From part 3 of the proof this will give $\gamma_i \geq \alpha$ for $l_k \leq i \leq m_k$, which together with (17) gives

$$\Psi^0(H^{i+1}) \leq \Psi^0(H^i) - \rho_i(\alpha - \varepsilon_i - \bar{\tau}(\rho_i)) \leq \Psi^0(H^i)$$

for $l_k \leq i < m_k$ and sufficiently large k. Combining this with $\tilde{\gamma}_{l_k} < 2\alpha$ gives $\tilde{\gamma}_{m_k} < 2\alpha$, which contradicts the initial assumption. Therefore $\max\{0, \tilde{\gamma}_s\} \to 0$.

Let us now take $\Delta_s = \max\{0, \max_{1 \leq i \leq m} \Psi^i(H^s)\}$ and $\tilde{H}^s = H^s + (\Delta_s/\sigma)(\tilde{H} - H^s)$. The convexity of functions Ψ^i, $i = \overline{1, m}$ and the fact that $\Psi^i(\tilde{H}) < -\sigma$ imply that

$\Psi^i(\tilde{H}^s) < 0$, $1 \leq i \leq m$. Thus $\tilde{H}^s$ satisfies all constraints (12), so that

$$\tilde{\gamma}_s \geq -|\Psi^0(\tilde{H}^s) - \Psi^0(H^s)| \geq -\frac{2K\Delta_s}{\sigma} - \tilde{\tau}\left(\frac{\Delta_s}{\sigma}\right)$$

and $\tilde{\gamma}_s \to 0$ since $\Delta_s \to 0$. This completes the proof.

Remark. We did not come to any conclusion about the convergence of distributions H^s because we did not introduce any topology on the set $G(Y)$. Let us now do this using weak* convergence topology, which by definition is the weakest topology in the space of all probability measures such that the map

$$H \to \int g(y)\, \mathrm{d}H(y)$$

is continuous wherever $g(y)$ is bounded and continuous. Note that in this topology the set of all probability measures with compact support Y is compact. If we assume that $\Psi^i(H)$ is continuous in this topology then we can deduce the existence of an H^* such that

$$\Psi^0(H^*) = \inf_{H \in G} \Psi^0(H).$$

Under the assumptions of the theorem all of the limit distributions of the sequence H^s minimize $\Psi^0(H)$.

The choice of the stepsize using (15), (16) when the constraints are linear is discussed in more detail in [14], which also gives some results for nonconvex functions $\Psi^0(H)$. In order to obtain a practical method from the general framework described in this section, we have to specify ways of performing step 2(i). This is the purpose of the next section.

4. Solving the linear subproblem using cutting-plane techniques

We shall now consider a method for solving linear subproblem (9) which reduces step 2(i) of algorithm 1 to the solution of one finite-dimensional linear programming problem. This method is based on generalized linear programming [3, 23], cutting-plane algorithms [19], and has much in common with the method proposed in [10] for solution of linear problem (1)–(3). The method is based on the duality relations for problem (1)–(3), which were studied in [10].

Let us assume that the assumptions of Theorems 1 and 3 are satisfied. Then, under the assumptions of Theorem 1, the optimal value z_s of problem (11)–(12) equals the optimal value of the finite-dimensional minimax problem:

$$z_s = \max_{u \in U^+} \varphi^s(u)$$

where

$$\varphi^s(u) = \min_{y \in Y} \left(f^0(y, H^s) + \sum_{i=1}^{m} u_i f^i(y, H^s) \right) - \sum_{i=1}^{m} u_i b_i^s,$$

$$b_i^s = \int_Y f^i(y, H^s)\, dH^s - \Psi^i(H^s).$$

Suppose that distribution H^s is fixed. Then it is possible to solve the problem

$$\max_{u \in U^+} \varphi^s(u) \tag{19}$$

using the following method.

Algorithm 2

1. First select $m+1$ points $y^1, y^2, \ldots, y^{m+1}$ and set $\nu = m+1$. These points are used to approximate function $\varphi^s(u)$ by the function

$$\varphi^s(u, 0) = \min_{1 \leq j \leq \nu} \left(f^0(y^j, H^s) + \sum_{i=1}^{m} u_i f^i(y^j, H^s) \right) - \sum_{i=1}^{m} u_i b_i^s.$$

The initial approximation u^0 to the solution of problem (19) maximizes the function $\varphi^s(u, 0)$:

$$u^0 = \arg \max_{u \in U^+} \varphi^s(u, 0)$$

so that we have to solve a linear programming problem. The points $y^1, \ldots, y^{m+1}$ are selected in such a way that the solution of this problem exists and bounded.

2. Suppose that before beginning iteration number k we have ν points $y^1, y^2, \ldots, y^\nu$ and the current estimate of the minimum u^{k-1}. Then iteration number k involves the following stages:

(i) Take $\nu = \nu + 1$

(ii) Find

$$y^\nu = \arg \min_{y \in Y} \left(f^0(y, H^s) + \sum_{i=1}^{m} u_i^{k-1} f^i(y, H^s) \right) \tag{20}$$

(iii) Calculate the next approximation to the optimal solution u^k:

$$u^k = \arg \max_{u \in U^+} \varphi^s(u, k) \tag{21}$$

where $\varphi^s(u, k)$ is the current approximation of function $\varphi^s(u)$:

$$\varphi^s(u, k) = \min_{1 \leq j \leq \nu} \left(f^0(y^j, H^s) + \sum_{i=1}^{m} u_i f^i(y^j, H^s) \right) - \sum_{i=1}^{m} u_i b_i^s.$$

It should be realized that this is only a general framework for solution—much has already been done to avoid increasing the number of points y^i stored and to implement approximate solutions of problem (20) (for details see [7, 10]). The advantage of this method is that it becomes possible to obtain approximate solutions of the initial problem (9) during the solution of problem (19). These approximations

are discrete distributions containing no more than $m+1$ points with positive probabilities:

$$(y^1, \bar{p}_1), (y^2, \bar{p}_2), \ldots, (y^{m+1}, \bar{p}_{m+1})$$

where the $\bar{p}_i$ are nonzero solutions of the following linear programming problem:

$$\min_{p} \sum_{i=1}^{\nu} p_i f^0(y^i, H^s),$$

$$\sum_{i=1}^{\nu} p_i f^j(y^i, H^s) \leqslant b_j^s, \qquad j = \overline{1, m},$$

$$\sum_{i=1}^{\nu} p_i = 1, \quad p_i \geqslant 0, \quad i = \overline{i, \nu},$$

and the y^i are the corresponding points. Note that the above problem is actually dual to the linear program equivalent to (21), and therefore both problems can be solved simultaneously.

Some other ways of solving minimax problems of the type (19) were considered in [4, 12].

What we actually need when implementing algorithm 1 is not a precise solution of problem (11)–(12) at each step, but rather to track its changing extreme value. The approximate solutions of (11)–(12) may be very rough for the first few iterations, gradually increasing in accuracy. It appears that algorithm 2 can be used to follow the extreme value by tracking the changing optimal solution of dual problem (19). It is only necessary to make one iteration of algorithm 2 for each iteration of algorithm 1.

In the remainder of this section we shall simplify the notation, writing

$$f^0(y, H^s) = f_s^0(y), \qquad f^i(y, H^s) - b_s^i = f_s^i(y).$$

We now want an algorithm which allows us to follow the optimal solution of problem (19) as the current distribution H^s changes.

Algorithm 2a

1. First select $m+1$ points $y^1, y^2, \ldots, y^{m+1}$ and set $\nu = m+1$. The inital approximation u^0 to the solution of problem (19) maximizes the function $\varphi^0(u, 0)$:

$$u^0 = \arg\max_{u \in U^+} \varphi^0(u, 0), \qquad \varphi^0(u, 0) = \min_{1 \leqslant j \leqslant \nu} \left(f_0^0(y^j) + \sum_{i=1}^{m} u_i f_0^i(y^j) \right).$$

2. Suppose that before beginning iteration number s we have ν points y^1, $y^2, \ldots, y^\nu$ and the current estimate of the minimum u^{s-1}. Then iteration number s involves the following stages:

(i) Take $\nu = \nu + 1$.

(ii) Find

$$y^\nu = \arg\min_{y\in Y}\left(f_s^0(y)+\sum_{i=1}^m u_i^{s-1}f_s^i(y)\right).$$

(iii) Calculate the next approximation to the optimal solution u^s:

$$u^s = \arg\max_{u\in U^+}\varphi^s(u,s), \qquad \varphi^s(u,s)=\min_{1\le j\le \nu}\left(f_s^0(y^j)+\sum_{i=1}^m u_i f_s^i(y^j)\right)$$

The following theorem proves the convergence of this method.

Theorem 4. *Assume that*

(a) *Set Y is compact.*

(b) *Functions $f_s^i(y)$, $i=\overline{0,m}$ are uniformly on s continuous for $y\in Y$ and*

$$\max_{y\in Y}|f_{s+1}^i(y)-f_s^i(y)|\to 0$$

as $s\to\infty$. In addition $|f_s^i(y)|<K<\infty$ for $0\le i\le m$, $y\in Y$.

(c) *The sequence u^s is bounded.*

Then $\max_{u\in U^+}\varphi^s(u)-\varphi^s(u^s)\to 0$, *where*

$$\varphi^s(u)=\min_{y\in Y}\left(f_s^0(y)+\sum_{i=1}^m u_i f_s^i(y)\right).$$

Proof. Suppose, arguing by contradiction, that the theorem is not true and that there exists an $\alpha>0$ and a sequence s_k such that

$$\max_{u\in U^+}\varphi^{s_k}(u)-\varphi^{s_k}(u^{s_k})>\alpha.$$

From the boundedness of sequence u^s and assumptions (a) and (b) we may assume without loss of generality that

$$u^{s_k}\to u^*, \qquad \|u^{s_k}-u^{s_{k+1}}\|\to 0, \qquad \varphi^{s_k}(u^*)\to\varphi^*, \qquad y^{\nu_k}\to y^*, \qquad f_{s_k}^i(y^*)\to f_*^i$$

where $\nu_k=s_k+1+m+1$.

We shall now estimate the difference

$$\max_{u\in U^+}\varphi^{s_{k+1}}(u)-\varphi^{s_{k+1}}(u^{s_{k+1}}).$$

From assumption (b) and definition of the function $\varphi^s(u)$ we have:

$$|\varphi^s(u_1)-\varphi^s(u_2)|\le\|u_1-u_2\|\max_{y\in Y}|f_s^i(y)|\le K\|u_1-u_2\|$$

for all $u_1, u_2\in U^+$. This gives

$$\varphi^{s_{k+1}}(u^{s_{k+1}})\ge\varphi^{s_{k+1}}(u^{s_k})-K\|u^{s_k}-u^{s_{k+1}}\|. \tag{22}$$

We also know that $\varphi^{s_k}(u^*)\to\varphi^*$, which together with assumption (b) implies that

$$\varphi^{s_{k+1}}(u^{s_k})\to\varphi^* \quad\text{and}\quad \varphi^{s_k}(u^{s_k})\to\varphi^*.$$

Hence

$$\varphi^{s_{k+1}}(u^{s_k}) \geq \varphi^{s_k}(u^{s_k}) - o(1). \tag{23}$$

We have assumed that

$$\max_{y \in Y} |f^i_{s+1}(y) - f^i_s(y)| \to 0$$

as $s \to \infty$ and this gives

$$\varphi^{s_k}(u^{s_k}) \geq \varphi^{s_k+1}(u^{s_k}) - o(1). \tag{24}$$

According to the step (ii) of the algorithm we have

$$\varphi^{s_k+1}(u^{s_k}) = f^0_{s_k+1}(y^{\nu_k}) + \sum_{i=1}^{m} u^{s_k}_i f^i_{s_k+1}(y^{\nu_k})$$

which gives from assumption (b):

$$\varphi^{s_k+1}(u^{s_k}) \geq f^0_{s_k}(y^{\nu_k}) + \sum_{i=1}^{m} u^{s_k}_i f^i_{s_k}(y^{\nu_k}) - o(1).$$

Taking into account uniform with respect to s continuity of $f^i_s(y)$ and our previous assumption $y^{\nu_k} \to y^*$, $f^i_{s_k}(y^*) \to f^i_*$ we obtain:

$$\varphi^{s_k+1}(u^{s_k}) \geq f^0_{s_k+1}(y^{\nu_k}) + \sum_{i=1}^{m} u^{s_k}_i f^i_{s_k+1}(y^{\nu_k}) - o(1) \tag{25}$$

where $\nu_k = s_k + 1 + m + 1$. From estimate (25) we obtain:

$$\begin{aligned}
&f^0_{s_{k+1}}(y^{\nu_k}) + \sum_{i=1}^{m} u^{s_k}_i f^i_{s_{k+1}}(y^{\nu_k}) \\
&\geq \min_{1 \leq j \leq \nu_{k+1}-1} \left[f^0_{s_{k+1}}(y^j) + \sum_{i=1}^{m} u^{s_k}_i f^i_{s_{k+1}}(y^j) \right] \\
&\geq \min_{1 \leq j \leq \nu_{k+1}-1} \left[f^0_{s_{k+1}}(y^j) + \sum_{i=1}^{m} u^{s_{k+1}}_i f^i_{s_{k+1}}(y^j) \right] - \|u^{s_k} - u^{s_{k+1}}\| \max_{y \in Y,\, 1 \leq i \leq m} |f^i_{s_{k+1}}(y)| \\
&\geq \max_{u \in U^+} \min_{1 \leq j \leq \nu_{k+1}-1} \left[f^0_{s_{k+1}}(y^j) + \sum_{i=1}^{m} u_i f^i_{s_{k+1}}(y^j) \right] - K \|u^{s_k} - u^{s_{k+1}}\| \\
&\geq \max_{u \in U^+} \varphi^{s_{k+1}}(u) - K \|u^{s_k} - u^{s_{k+1}}\|.
\end{aligned} \tag{26}$$

Combining (22)–(26) gives:

$$\max_{u \in U^+} \varphi^{s_{k+1}}(u) - \varphi^{s_{k+1}}(u^{s_{k+1}}) \leq 2K \|u^{s_k} - u^{s_{k+1}}\| + o(1) \leq o(1).$$

This contradicts the initial assumption

$$\max_{u \in U^+} \varphi^s(u) - \varphi^s(u^s) > \alpha > 0$$

and thus completes the proof.

There are various ways to ensure the boundedness of sequence u^s which depend on the nature of the functions f_s^i. Let us now turn again to our initial problem, assuming that the constraints are linear (see (2)). In this case $f_i^s(y) \equiv q^i(y)$, $i = \overline{1, m}$. Suppose also that we have a distribution $\tilde{H}$ such that

$$\int_Y q^i(y)\, d\tilde{H}(y) < 0, \qquad i = \overline{1, m}. \tag{27}$$

In this case it is possible to select the initial $m+1$ points in a way that will guarantee the desired boundedness. Let us consider the following problem:

$$\min_H \mu, \tag{28}$$

$$\int_Y q^j(y)\, dH(y) \leq \mu, \quad j = \overline{1, m} \tag{29}$$

$$\int_Y dH(y) = 1. \tag{30}$$

According to Theorem 1 there exist $m+1$ points $\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^{m+1}$ such that one of the solutions of (28)-(30) is a measure with support consisting of points $\bar{y}^i$ with probabilities which are solutions of the following linear programming problem:

$$\min_{p,\mu} \mu, \tag{31}$$

$$\sum_{i=1}^{m+1} p_i q^j(\bar{y}^i) \leq \mu, \quad j = \overline{1, m}, \tag{32}$$

$$\sum_{i=1}^{m+1} p_i = 1, \quad p_i \geq 0, \quad i = \overline{1, m+1}. \tag{33}$$

According to (27) the optimal solution μ^* of this problem is negative. The problem dual to (31)-(33) has the form:

$$\max_{u \in U^+} u_{m+1}, \tag{34}$$

$$\sum_{j=1}^{m} u_j q^j(\bar{y}^j) - u_{m+1} \geq 0, \quad i = \overline{1, m+1}, \tag{35}$$

$$\sum_{j=1}^{m} u_j = 1, \tag{36}$$

with solution $u_{m+1}^* < 0$. This implies that there exists a $\gamma > 0$ such that for any $u \in U^+$, $\|u\| = 1$ there exists an $i \in \{1, \ldots, m+1\}$ for which

$$\sum_{j=1}^{m} u_j q^j(\bar{y}^i) < -\gamma.$$

We shall prove that, under the assumptions of Theorem 4, sequence u^s will be bounded if $\bar{y}^i$, $i = \overline{1, m}$ are taken as initial points. Take any point $\bar{u} \in U^+$ and estimate

the value of $\varphi^s(u, s)$ at this point. Select $i \leq m+1$ such that

$$\sum_{j=1}^{m} \bar{u}_j q^j(\bar{y}^i) < -\gamma \|\bar{u}\|,$$

which will always be possible. Then

$$\varphi^s(\bar{u}, s) \leq f_s^0(\bar{y}^i) - \gamma \|\bar{u}\|. \tag{37}$$

According to assumption (b) of the theorem there exists a constant K such that $|f_s^0(y)| \leq K$. Combining this with (37) yields

$$\varphi^s(\bar{u}, s) \leq K - \gamma \|\bar{u}\|$$

and

$$\varphi^s(0, s) \geq -K.$$

These two inequalities lead to

$$\varphi^s(0, s) - \varphi^s(\bar{u}, s) \geq \gamma \|\bar{u}\| - 2K,$$

which implies that the norm of any point u^* which maximizes $\varphi^s(u, s)$ is bounded, i.e.,

$$\|u^*\| \leq \frac{2K}{\gamma}$$

where constants K and γ do not depend on s. This proves that sequence u^s is bounded, because u^s maximizes $\varphi^s(u, s)$.

It is not necessary to solve problem (28)-(30) precisely—in fact, any distribution satisfying (27) with support consisting of $m+1$ points would do. If the constraints are nonlinear it is not possible to guarantee the boundedness of the sequence u^s by choosing initial points. In the case when $\Psi^i(\tilde{H}) < 0$ for all $1 \leq i \leq m$ and for some $\tilde{H}$, and functions $\Psi^i(\tilde{H}) < 0$ are convex for all $1 \leq i \leq m$, one possible strategy is to update the set of $m+1$ initial points periodically so that (27) is always satisfied for some distribution with the initial points as support.

5. Version of algorithm 1 based on generalized linear programming techniques

We shall now describe an algorithm based on the results obtained in Sections 3 and 4. It is assumed that the conditions of Theorem 3 are met.

Algorithm 1a

1. We begin by choosing an initial distribution H^1 such that the problem

$$\max_{u \in U^+} \min_{y \in Y} \left[f^0(y, H^1) + \sum_{i=1}^{m} u_i f^i(y, H^1) \right]$$

has a solution. This may be done by applying algorithm 2 to the problem (28)-(30)

until a distribution is obtained which satisfies (27). Under the assumptions of Theorem 3 this will occur in a finite number of steps. If the constraints are linear, such a choice will also guarantee the boundedness of the sequence u^s. The initial step of algorithm 1a therefore involves the following stages:

(i) Take $\nu_0 = m+1$, where ν_0 is the number of points in distribution H^1.

(ii) Obtain the initial distribution H^1, where

$$H^1 = \{(y^1, p_1^1), (y^2, p_2^1), \ldots, (y^{m+1}, p_{m+1}^1)\},$$

by applying algorithm 2 to problem (28)-(30). This algorithm will produce a sequence of distributions $\bar{H}^k$ which, after substitution in problem (28)-(30), yields the corresponding values of μ_k. Take as H^1 the first $\bar{H}^k$ with $\mu_k < 0$.

(iii) Take $u^1 \in U^+$ as an initial point for solution of the dual problem.

2. Suppose that before beginning iteration number s we have the current approximation H^s to the optimal solution, which consists of ν_s points:

$$H^s = \{(y^1, p_1^s), \ldots, (y^{\nu_s}, p_{\nu_s}^s)\}$$

and point u^s. Iteration number s then involves the following operations, where steps (i)-(iii) correspond to steps (1)-(iii) of algorithm 2a and step (i) of algorithm 1, and step (iv) corresponds to step (ii) of algorithm 1:

(i) Take $\nu_{s+1} = \nu_s + 1$.

(ii) Find a new point $y^{\nu_{s+1}}$, where

$$y^{\nu_{s+1}} = \arg\min_{y \in Y} \left[f^0(y, H^s) + \sum_{i=1}^{m} u_i^s f^i(y, H^s) \right]. \tag{38}$$

(iii) Solve the following linear programming problem:

$$\min_{p} \sum_{i=1}^{\nu_{s+1}} p_i f^0(y^i, H^s), \tag{39}$$

$$\sum_{i=1}^{\nu_{s+1}} p_i f^j(y^i, H^s) \leq 0, \quad j = \overline{1, m}, \tag{40}$$

$$\sum_{i=1}^{\nu_{s+1}} p_i = 1, \quad p_i \geq 0, \quad i = \overline{1, \nu_{s+1}} \tag{41}$$

together with its dual:

$$\max_{u} u_{m+1} \tag{42}$$

$$f^0(y^j, H^s) + \sum_{i=1}^{m} u_i f^i(y^j, H^s) - u_{m+1} \geq 0, \quad j = \overline{1, \nu_{s+1}}, \tag{43}$$

$$u_i \geq 0, \quad i = \overline{1, m}.$$

This will give us the next approximation to the solution of the dual problem, u^{s+1},

and also a vector $\bar{p}^{s+1}$:

$$\bar{p}^{s+1} = (\bar{p}_1^{s+1}, \ldots, \bar{p}_{\nu_{s+1}}^{s+1})$$

which will have no more than $m+1$ nonzero elements, say $(\bar{p}_{k_1}^{s+1}, \ldots, \bar{p}_{k_{m+1}}^{s+1})$.

(iv) Take the family of distributions

$$H^{s+1}(\alpha) = \{(y^1, p_1^{s+1}(\alpha)), \ldots, (y^{\nu_{s+1}}, p_{\nu_{s+1}}^{s+1}(\alpha))\}$$

where

$$p_i^{s+1}(\alpha) = \begin{cases} p_i^s(1-\alpha), & \text{if } i \neq k_j \text{ for } j = \overline{1, m+1}, \\ p_{k_j}^s(1-\alpha) + \alpha \bar{p}_{k_j}^{s+1}, & \text{otherwise.} \end{cases}$$

Then take

$$H^{s+1} = H^{s+1}(\rho_s) \quad \text{where } \sum_{s=0}^{\infty} \rho_s = \infty, \ \rho_s \to 0$$

and go to step 2(i).

Remark 1. If the constraints are linear step (iv) can be performed in the following way [14]:

Set

$$\alpha_s = \arg \min_{0 \leq \alpha \leq 1} \Psi^0(H^{s+1}(\alpha)).$$

Take $\rho_s = \min\{\alpha_s, \beta_s\}$, where

$$\beta_s \to 0, \qquad \sum_{i=1}^{\infty} \beta_i = \infty.$$

Then take

$$H^{s+1} = H^{s+1}(\rho_s)$$

and go to step 2(i).

Remark 2. It is not necessary to solve nonlinear programming problem (38) with great accuracy. All we need is a point $y^{\nu_{s+1}}$ such that

$$\lim_{s \to \infty} \left\{ \left[f^0(y^{\nu_{s+1}}, H^s) + \sum_{i=1}^{m} u_i^s f^i(y^{\nu_{s+1}}, H^s) \right] - \min_{y \in Y} \left[f^0(y, H^s) + \sum_{i=1}^{m} u_i^s f^i(y, H^s) \right] \right\} = 0.$$

It is also possible to avoid increases in the dimension of linear programming problem (39)-(41) by considering only points y^i which satisfy some additional inequality (see [10]).

Remark 3. Algorithm 1a adds one additional point to the current approximation H^s of the optimal solution at each iteration, which may not be convenient if we have restrictions on the amount of memory available for storing the distribution. In this case measures should be taken to avoid this expansion, perhaps at the expense of accuracy. Some possible ways of achieving this are discussed below.

1. Suppose we want to find approximation, which consists of N points, to the optimal solution of (5)–(7). (It is assumed that some additional memory is available for storing N further points.) We then proceed as follows:

(i) Run algorithm 1a until the current distribution H^s contains $2N$ points. Arrange these points in order of decreasing probabilities:

$$H^s = \{(y^{l_1}, p_{l_1}), \ldots, (y^{l_{2N}}, p_{l_{2N}})\}.$$

(ii) Start algorithm 1a again from the distribution

$$H^1 = \{(y^1, p_1^1), \ldots, (y^N, p_N^1)\},$$

$$y^i = y^{l_i}, \quad p_i^1 = \frac{p_{l_i}}{1-\vartheta}, \quad i = \overline{1, N}, \quad \vartheta = \sum_{j=N+1}^{2N} p_{l_j}.$$

(iii) Continue this process until the new N-point distribution no longer has a better value of $\Psi^0(H)$ than the previous one.

2. Suppose that we want to find an approximation to the optimal solution using at most N points.

(i) Run algorithm 1a until the current distribution contains N points. Let $\bar{p} = \max_{1 \leq i \leq N} p_i^s$. Divide the set $\{1, \ldots, N\}$ into two subsets:

$$I_1 = \{i: p_i^s \geq \chi \bar{p}\}, \qquad I_2 = \{i: p_i^s < \chi \bar{p}\}$$

where $\chi > 0$ should be chosen before the run.

(ii) Start algorithm 1a again from distribution H^1:

$$H^1 = \left\{(y^i, p_i^1), i \in I_1, p_i^1 = \frac{p_i^s}{1 - \sum_{j \in I_2} p_j^s}\right\}.$$

(iii) Continue this process for as long as the value of $\Psi^0(H)$ in consecutive N-point distributions improves and set I_2 is not empty.

3. Another possibility is to use approximation techniques to fit discrete distributions by continuous ones. For example, splines can be used when the dimensions are small or when the distributions H^s have independent components. This approach needs further study.

The convergence of algorithm 1a follows directly from Theorems 3 and 4. The next result may be derived from these theorems.

Theorem 5. *Let the following conditions be satisfied*:

1. *Y is a compact set.*

2. *Functions* $\Psi^i(H)$, $i=\overline{0, m}$ *satisfy* (8) *where*

$$\tau^i(\alpha, H_1, H_2) \leq \tilde{\tau}(\alpha) < \infty, \quad 0 \leq \alpha \leq 1, \qquad \tilde{\tau}(\alpha)/\alpha \to 0 \quad \text{as } \alpha \to 0.$$

3. *Functions* $f^i(y, H)$ *are uniformly continuous with respect to* $y \in Y$ *over* $H \in G(Y)$; $|f^i(y, H)| \leq K < \infty$ *for* $y \in Y$, $H \in G(Y)$.

4. *There exists a distribution* $\tilde{H}$ *such that* $\Psi^i(\tilde{H}) < -\sigma < 0$ *for* $1 \leq i \leq m$.

5. *Sequence* u^s *is bounded.*

Then

$$\lim_{s\to\infty} [\Psi^0(H^s) - \min_{H\in G} \Psi^0(H)] = 0.$$

If the constraints are linear then assumption 5 *is satisfied automatically for algorithm* 1a.

It is interesting to compare this algorithm with the methods for solving stochastic problems with recourse recently proposed by Nazareth and Wets [22]. Although applied to quite different problems, they both use generalized linear programming techniques and on each iteration require solution of one linear programming problem and one nonlinear optimization problem.

6. Some generalizations

Another way of dealing with problem (5)-(7) is to apply the dual approach directly. When the functions $\Psi^i(H)$, $i=\overline{0, m}$ are convex and the constraints satisfy the Slater condition, (5)-(7) is equivalent (under certain assumptions) to the problem

$$\max_{v_i \geq 0} \min_{H\in G} \left[\Psi^0(H) + \sum_{i=1}^{m} v_i \Psi^i(H) \right] \tag{44}$$

where linear constraints of type (2) may be included in the set G. If the functions $\Psi^i(H)$ are directionally differentiable (see (8)) then the following algorithm can be used.

1. Take initial points v^1 and H^1 where H^1 is selected as in algorithm 1a.
2. At step s compute

$$\bar{v}^s = v^s + \beta_s \Psi^i(H^s)$$

and take

$$v_i^{s+1} = \max\{0, \bar{v}_i^s\}.$$

Then perform one step of algorithm 1a.

Problem (44) is a particular case of the minimax problem

$$\min_{v\in V} \max_{H\in G} \Psi(v, H),$$

where $V \subset \mathbb{R}^n$, G is defined by (7) or (10) and function $\Psi(v, H)$ has directional derivatives of type (8) with respect to H. In this case algorithm 1a can be modified to include some kind of gradient descent with respect to variables v at the same time as the inner problem is being solved. Problems of this kind occur in optimal experiment design when the underlying regression problem depends nonlinearly on the parameters to be estimated. More general minimax problems arise when finding the solution of a game problem with mixed strategies:

$$\min_{P \in Q} \max_{H \in G} \Psi(P, H)$$

where sets Q and G are both defined analogously to G in (7) and the function Ψ possesses some regularity properties in addition to directional differentiability.

In this case the analogue of algorithm 1a will include two cutting plane processes.

Acknowledgements

The author is grateful to Yu. Ermoliev, R. Wets and L. Nazareth for valuable comments.

References

[1] J. Birge and R. Wets, "Designing approximation schemes for stochastic optimization problems, in particular for stochastic problems with recourse", Working paper WP-83-114, International Institute for Applied Systems Analysis (Laxenburg, Austria, 1983).

[2] T. Cipra, "Class of unimodal distributions and its transformations", *Časopis pro pěstovani Matematiky* 103 (1978) 17-26.

[3] G.B. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, NJ, 1963).

[4] V. Demyanov and L. Vasilyev, *Nondifferentiable optimization* (Nauke, Moscow, 1981).

[5] J. Dupačova, "Minimax stochastic programs with nonconvex nonseparable penalty functions", in: *A. Prekopa, ed., Colloquia Mathematica Societatis Janos Bolyai 12, Progress in Operations Research Eger* 1974 (North-Holland, Amsterdam, 1976) pp. 303-316.

[6] J. Dupačova, "Minimax approach to stochastic linear programming and the moment problem", *Zeitschrift fur Angewandte Mathematik und Mechanik* 58 (1978) T466-T467.

[7] B.C. Eaves and W.I. Zangwill, "Generalized cutting plane algorithms", *SIAM Journal on Control* 9 (1971) 529-542.

[8] Yu. Ermoliev, "Method for stochastic programming in randomized strategies", *Kibernetika* 1 (1970) 3-9 (in Russian).

[9] Yu. Ermoliev and C. Nedeva, "Stochastic optimization problems with partially known distribution functions", Collaborative paper CP-82-60, International Institute for Applied Systems Analysis (Laxenburg, Austria, 1982).

[10] Yu. Ermoliev, A. Gaivoronski and C. Nedeva, "Stochastic optimization problems with incomplete information on distribution functions", Working paper WP-83-113, International Institute for Applied Systems Analysis (Laxenburg, Austria, 1983).

[11] V. Fedorov, *Theory of optimal experiments* (Academic Press, New York, 1972).

[12] V. Fedorov, *Numerical methods of maximin* (Nauka, Moscow, 1979).

[13] M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3 (1956) 95-110.

[14] A. Gaivoronski, "Optimization of functionals which depend on distribution functions: 1. Nonlinear functional and linear constraints", Working paper WP-83-114, International Institute for Applied Systems Analysis (Laxenburg, Austria, 1983).
[15] A. Golodnikov, "Optimization problems with distribution functions", Ph.D. Thesis, V. Glushkov Institute for Cybernetics (Kiev, 1979, in Russian).
[16] A. Golodnikov, A. Gaivoronski and Le Ding Fung, "A method for solving limit extremum problems applied to the minimization of functionals", *Kibernetika* 1 (1980) 99-103.
[17] P.J. Huber, *Robust Statistics* (John Wiley & Sons, New York, 1981).
[18] A. Karr, "Extreme points of certain sets of probability measures, with applications", *Mathematics of Operations Research* 8 (1983) 74-85.
[19] J.E. Kelley, "The cutting-plane method for solving convex programs", *Journal of Society for Industrial and Applied Mathematics* 8 (1960) 703-712.
[20] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems", *Annals of mathematical statistics* 30 (1959) 271-294.
[21] M. Krein and A. Nudelman, *The Markov moment problem and extremal problems*, Translation Mathem. Monographs 50 (American Mathematical Society, Providence, RI, 1977).
[22] L. Nazareth and R.J.-B. Wets, "Algorithms for stochastic programs: the case of nonstochastic tenders", Working paper WP-83-5 (Revised version), International Institute for Applied Systems Analysis (Laxenburg, Austria, 1983).
[23] R. Wets, *Grundlagen Konvexer Optimierung*, Lecture Notes in Economics and Mathematical Systems 137 (Springer-Verlag, Berlin, 1976).
[24] P. Whittle, "Some general points in the theory of optimal experimental design", *Journal of the Royal Statistical Society* (*B*) 35 (1973) 123-130.
[25] J. Žačkova, "On minimax solutions of stochastic linear programming problems", *Časopis pro pěstovani matematiky* 91 (1966) 423-430.