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

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

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:

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

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

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

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

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

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 function s 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

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

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

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

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*

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

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 $\nu \geq \nu_\varepsilon$,

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

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

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

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' in K, we have that for all ω

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

which implies that

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

Let us now take x' 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

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

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

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

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*

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

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 $\{\xi^l = (h^l, T^l), l = 1, \dots, \nu\}$ be a finite number of realizations of h and T, $x \in K_2$ and for $l = 1, \dots, \nu$,*

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

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 $l = 1, \dots, \nu$,

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

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 $\{\xi^l, l=1,\dots,\nu\}$ will yield a better approximation. For example, the ξ^l could be the conditional expectation of $\xi(\cdot)$ with respect to a partition $\mathcal{S}=\{S_l, l=1,\dots,\nu\}$ 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 $\nu=1,\dots,$ let $\mathcal{S}^\nu=\{S_l^\nu, l=1,\dots,L_\nu\}$ denote a partition of Ξ with*

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

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

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

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

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

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

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

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

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

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

then follows simply from the fact that $\mathcal{S}^{\nu+1}\supset\mathcal{S}^\nu$. Now observe that

$$\max_{1\le l\le L_\nu}[v^{\nu l}\xi+f(x,\xi^{\nu l})-v^{\nu l}\xi^{\nu l}]\ge g^\nu(x,\xi) \tag{3.10}$$

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

$$E_f^\nu(x)\ge E\{g^\nu(x,\xi(\omega))\}=\sum_{l=1}^{L_\nu}P[\xi(\omega)\in S_l^\nu]f(x,\xi^{\nu l})$$

which gives us

$$E_f(x)\ge\lim_{\nu\to\infty}E_f^\nu(x)\ge\lim_{\nu\to\infty}E\{g^\nu(x,\xi(\omega))\}=E_f(x);$$

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 $l=1,\dots,\nu$,*

$$y^l\in\operatorname{argmin}[q^ly\,|\,Wy=p-Tx,\ y\ge 0].$$

Then

$$\mathcal{Q}(x)\le E\{\min_{1\le l\le\nu}q(\omega)y^l\}. \tag{3.12}$$

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

$$Q(x,\xi)\le\min_{1\le l\le\nu}qy^l$$

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*

$$E_f(x)\le E_f^\nu(x)=\int f^\nu(x,\xi(\omega))P(\mathrm{d}\omega) \tag{3.14}$$

where

$$f^\nu(x,\xi)=\inf_{\lambda\in\mathbb{R}^\nu_+}\Big[\sum_{l=1}^\nu\lambda_l f(x,\xi^l)\,\Big|\,\xi=\sum_{l=1}^\nu\lambda_l\xi^l,\ 1=\sum_{l=1}^\nu\lambda_l\Big]. \tag{3.15}$$

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.

$$e^i = (0, \ldots, 0, 1, 0, \ldots, 0)^\mathrm{T}.$$

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

$$\begin{aligned} \rho(\hat{\xi}, \xi) = \inf[\, & cx + \hat{p} q \hat{y} + (1 - \hat{p}) q y_\xi] \\ \text{such that} \quad & Ax = b, \\ & \hat{T}x + W\hat{y} = \hat{h}, \\ & Tx + W y_\xi = h, \\ & x \geq 0, \ \hat{y} \geq 0, \ y_\xi \geq 0, \end{aligned} \tag{3.29}$$

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

$$\rho(\bar{\xi}, \xi) \leq cx^0 + \hat{p} q \bar{y} + (1 - \hat{p}) q y^0(\xi)$$

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

$$\xi^l = (h^l, T^l) = E\{(h(\omega), T(\omega)) | S^l\}$$

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

$$z^\nu \leq z^* = \inf\{cx + \mathcal{Q}(x) | Ax = b, x \geq 0\}.$$

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),

$$\psi_i(\chi_i, \xi_i) = \begin{cases} q_i^+(h_i - \chi_i) & \text{if } h_i \geq \chi_i, \\ q_i^-(\chi_i - h_i) & \text{if } h_i \leq \chi_i. \end{cases}$$

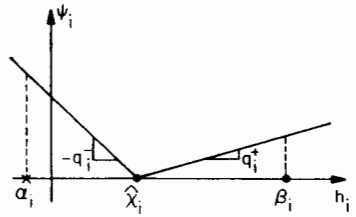

Fig. 4.8. The function $\psi_i(\chi_i, \cdot)$.

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 $l = 1, 2$,

$$h_i^l = E\{h_i(\omega) | S^l\} \quad \text{and} \quad p_{il} = P[h_i(\omega) \in S^l],$$

and $\bar{h}_i = E\{h_i(\omega)\}$. Note that

$$\Psi_i(\hat{\chi}_i) = q_i^- \int_{h_i(\omega) < \hat{\chi}_i} (\hat{\chi}_i - h_i(\omega)) P(\mathrm{d}\omega) + q_i^+ \int_{h_i(\omega) \geq \hat{\chi}_i} (h_i(\omega) - \hat{\chi}_i) P(\mathrm{d}\omega)$$

$$= \Psi_i^a(\hat{\chi}_i).$$

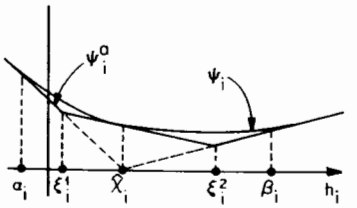

Fig. 4.9. The function Ψ^a.

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

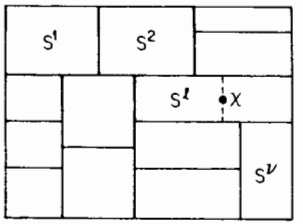

Fig. 4.11. Partition $\mathcal{S} = \{S^1, \ldots, S^\nu\}$ of Ξ.

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

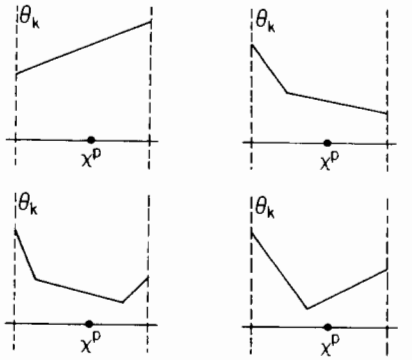

Fig. 4.12. The function θ_k on E_k.

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, $\mathcal{E}$) *with $\mathcal{E}$ 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 $A \in \mathcal{E}$. 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 $\mathcal{E}$ 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

$$G = \{\gamma = (\gamma_1, \ldots, \gamma_N) \mid \gamma_i = \alpha_i \text{ or } \beta_i,\ i = 1, \ldots, N\}.$$

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

$$\begin{aligned}
\mathcal{Q}(x) &= \int Q(x,\xi)\, \mathrm{d}F(\xi) \\
&\leq \max_{\{\hat{\xi} = (\hat{\xi}_2, \ldots, \hat{\xi}_N) \mid \hat{\xi} \in \hat{\Xi}\}} \int_{\alpha_1}^{\beta_1} Q(x, (\xi_1, \hat{\xi}_2, \ldots, \hat{\xi}_N))\, \mathrm{d}F(\xi_1, \hat{\xi})
\end{aligned}$$

$$= \max_{\{\hat{\xi}=(\hat{\xi}_2,\ldots,\hat{\xi}_N)\,|\,\hat{\xi}\in\hat{\Xi}\}} \left[\left(\frac{\beta_1 - \bar{\xi}_1(\hat{\xi})}{\beta_1 - \alpha_1} \right) Q(x, (\alpha_1, \hat{\xi})) \right.$$

$$\left. + \left(\frac{\bar{\xi}_1(\hat{\xi}) - \alpha_1}{\beta_1 - \alpha_1} \right) Q(x, (\beta_1, \hat{\xi})) \right] \tag{5.16}$$

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).

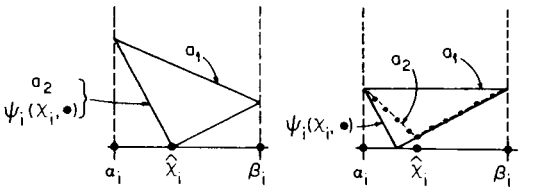

Fig. 5.19. Splitting $[\alpha_i, \beta_i]$ at $\hat{\chi}_i$.

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 $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{A})$, 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 $\mathcal{P}$. Since by assumption $P \in \mathcal{P}$, we have

$$\inf_{P' \in \mathcal{P}} \int f(x, \xi(\omega)) P'(\mathrm{d}\omega) \leq E_f(\chi) \leq \sup_{P' \in \mathcal{P}} \int f(x, \xi(\omega)) P'(\mathrm{d}\omega). \tag{6.2}$$

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 $\mathcal{P}$ is of utmost importance. On one hand we want $\mathcal{P}$ 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 $\mathscr{P}$ as a neighborhood of P such as

$$\mathscr{P} = \{\text{prob. meas. } Q \mid \sup_{A \in \mathscr{A}} |P(A) - Q(A)| \leq \varepsilon\}, \tag{6.3}$$

or even

$$\mathscr{P} = \left\{\text{prob. meas } Q \mid \sup_{\theta \in \Theta} \left| \int \theta(\omega) P(\mathrm{d}\omega) - \int \theta(\omega) Q(\mathrm{d}\omega) \right| \leq \varepsilon \right\}, \tag{6.4}$$

where Θ is a class of test functions. Or with F the distribution function of $\xi(\cdot)$ defined on $\mathbb{R}^N$

$$\mathscr{P} = \{\text{dist. funct. } G \mid \sup_{z} |F(z) - G(z)| \leq \varepsilon\}. \tag{6.5}$$

The class $\mathscr{P}$ 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; $\mathscr{P}$ 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 $\mathscr{P}$ be the set of probability measures Q on $(\Xi, \mathscr{B}^N)$ 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 $P \in \mathscr{P}$, 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 $\mathscr{P}$ is convex and compact (with respect to the weak* topology), and $\mathscr{P} = \mathrm{cl\,co}(\mathrm{ext}\,\mathscr{P})$. 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 $\mathscr{P}$. Moreover the extremal measures of $\mathscr{P}$ 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 $\mathscr{P}$, 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 $\mathscr{P}$ is easy to verify, the compactness for the space of measures on a compact metric space follows from Prohorov's Theorem, and $\mathscr{P} = \mathrm{cl\,co}(\mathrm{ext}\,\mathscr{P})$ 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

$$\sum_{l=1}^{\nu} p_l = 1,$$

$$\sum_{l=1}^{\nu} v_i(\xi^l) p_l \leq \alpha_i \quad \text{for } i = 1, \ldots, s,$$

$$\sum_{l=1}^{\nu} v_i(\xi^l)p_l = \alpha_i \quad \text{for } i = s+1, \ldots, M, \text{ and}$$

$$z = \sum_{l=1}^{\nu} v_0(\xi^l)p_l \text{ is maximized.} \tag{6.10}$$

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 $[1, v_1(\xi^l), \ldots, v_M(\xi^l)]^T$ that are linearly independent. And this holds for every possible objective $\int v_0(\xi)Q(\mathrm{d}\xi)$, 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:

$$\text{find } Q \text{ a measure on } (\Xi, \mathcal{B}^1)$$

$$\text{such that } Q \geq 0, \int_\alpha^\beta Q(\mathrm{d}\xi) = 1, \int_\alpha^\beta \xi Q(\mathrm{d}\xi) = \bar{\xi}$$

$$\text{and } \int_\alpha^\beta v_0(\xi)Q(\mathrm{d}\xi) \text{ is maximized.}$$

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,*

$$\mathcal{P} = \left\{ Q \,\middle|\, \int_\Xi \xi Q(\mathrm{d}\xi) = \bar{\xi} = \int_\Xi \xi P(\mathrm{d}\xi) \right\},$$

and $v_0: \Xi \to \mathbb{R}$ *is convex. Then there exists*

$$P_\nu \in \operatorname*{argmax}_{Q \in \mathcal{P}} \int_\Xi v_0(\xi) Q(\mathrm{d}\xi)$$

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

$$\int v_0(\xi) P_\nu(\mathrm{d}\xi) \leq \int v_0(\xi) P'_\nu(\mathrm{d}\xi)$$

but still

$$\int \xi P'_\nu(\mathrm{d}\xi) = \int \xi P_\nu(\mathrm{d}\xi) = \bar{\xi}.$$

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

$$\mathcal{P} = \left\{ Q \text{ prob. meas. on } \operatorname{ext}(\operatorname{co} \Xi) \,\middle|\, \int \xi Q(\mathrm{d}\xi) = \bar{\xi} \right\}. \qquad \square$$

6.16. Corollary [31; 17, Theorem 1]. *Suppose* K *is polytope that includes* Ξ, $\mathcal{P} = \{Q \mid \int_\Xi \xi Q(\mathrm{d}\xi) = \bar{\xi}\}$ *and* $v_0: K \to \mathbb{R}$ *is convex. Then*

$$\int v_0(\xi) P(\mathrm{d}\xi) \leq \max_{\lambda \geq 0} \left[\sum_{j=1}^{L} \lambda_j v_0(e^j) \,\middle|\, \sum_{j=1}^{L} \lambda_j e^j = \bar{\xi}, \sum_{j=1}^{L} \lambda_j = 1 \right] \tag{6.17}$$

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 $\mathcal{P}_K$. $\square$

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* $\operatorname{co}(e^1, \ldots, e^L) \supset \Xi$ *and* $h: \operatorname{co}(e^1, \ldots, e^L) \to \bar{\mathbb{R}}$ *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 $\pi_2^{\nu} \leq 0$, 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

$$P\{\omega \mid \xi^1(\omega) \preceq_C \xi^2(\omega)\} = 1;$$

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 $\preceq_{\mathbb{R}_+}$ when $\xi_i \leq \chi_i$ and $\psi_i(\chi_i, \cdot)$ is order preserving with respect to $\preceq_{\mathbb{R}_+}$ 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

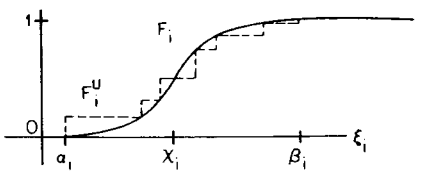

Fig. 7.10. Majorizing distribution function F_i^U.

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

find $\alpha = z_0 \leq z_1 \leq \cdots \leq z_L \leq z_{L+1} = \beta$

such that $\rho(z) = \sum_{l=1}^{L} [v(z_{l+1}) - v(z_l)][F(z_l) - F(z_{l+1})]$ is maximized. (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

find $p_j \geq 0, j = 1, \ldots, \nu$

such that $\sum_{l=1}^{j} p_l \leq F(z_j), j = 1, \ldots, \nu$, and $\sum_{j=1}^{\nu} v(z_j) p_j$ is maximized (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 $\mathcal{Q}^L(x) = \int Q(x, \xi) P_L(\mathrm{d}\xi)$. 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:

$$\text{find } \chi \in \mathbb{R}^{m_2} \text{ such that } \rho(\chi) + \Psi(\chi) \text{ is minimized.} \tag{8.7}$$

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

$$\begin{aligned} &\text{find } x \in \mathbb{R}^n_+,\ u^+ \in \mathbb{R}^{m_2}_+,\ u^- \in \mathbb{R}^{m_2}_+ \\ &\text{such that } Ax = b, \\ &\qquad Tx + u^+ - u^- = \chi^e \quad \text{and} \\ &\qquad cx + q^+u^+ + q^-u^- = z \quad \text{is minimized} \end{aligned} \tag{8.10}$$

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.

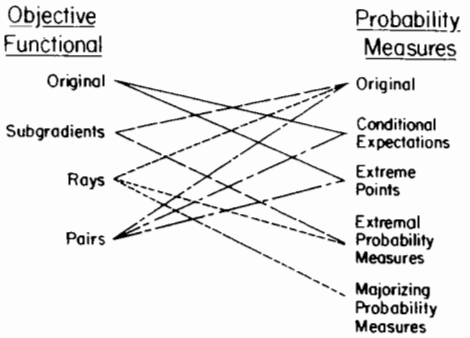

Fig. 9.1. Approximation pairs.

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: