Scenario Reduction Techniques in Stochastic Programming

Werner Römisch

Humboldt-University Berlin, Department of Mathematics, Rudower Chaussee 25, 12489 Berlin, Germany

Abstract. Stochastic programming problems appear as mathematical models for optimization problems under stochastic uncertainty. Most computational approaches for solving such models are based on approximating the underlying probability distribution by a probability measure with finite support. Since the computational complexity for solving stochastic programs gets worse when increasing the number of atoms (or scenarios), it is sometimes necessary to reduce their number. Techniques for scenario reduction often require fast heuristics for solving combinatorial subproblems. Available techniques are reviewed and open problems are discussed.

1 Introduction

Many stochastic programming problems may be reformulated in the form

$$\min\Big\{\mathbb{E}(f_0(x,\xi)) = \int_{\mathbb{R}^s} f_0(x,\xi)P(d\xi) : x \in X\Big\}, \tag{1}$$

where X denotes a closed subset of $\mathbb{R}^m$, the function f_0 maps from $\mathbb{R}^m \times \mathbb{R}^s$ to the extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, $\mathbb{E}$ denotes expectation with respect to P and P is a probability distribution on $\mathbb{R}^s$.

For example, models of the form (1) appear as follows in economic applications. Let $\xi = \{\xi_t\}_{t=1}^T$ denote a discrete-time stochastic process of d-dimensional random vectors ξ_t at each $t \in \{1, \ldots, T\}$ and assume that decisions x_t have to be made such that the total costs appearing in an economic process are minimal. Such an optimization model may often be formulated as

$$\min\Big\{\mathbb{E}\Big(\sum_{t=1}^{T} f_t(x_t,\xi_t)\Big) : x_t \in X_t,\ \sum_{\tau=0}^{t-1} A_{t\tau}(\xi_t)x_{t-\tau} = h_t(\xi_t),\ t = 1,\ldots,T\Big\}. \tag{2}$$

Typically, the sets X_t are polyhedral, but they may also contain integrality conditions. In addition, the decision vector $(x_1, \ldots, x_T)$ has to satisfy a dynamic constraint (i.e. x_t depends on the former decisions) and certain balancing conditions. The matrices $A_{t\tau}(\xi_t)$, $\tau = 0, \ldots, t-1$, (e.g. containing technical parameters) and the right-hand sides $h_t(\xi_t)$ (e.g. demands) are (partially) random. The functions f_t describe the costs at time t and may also be (partially) random

O. Watanabe and T. Zeugmann (Eds.): SAGA 2009, LNCS 5792, pp. 1–14, 2009.

(e.g. due to uncertain market prices). The objective in (2) consists in minimizing the expected total costs.

The time $t = 1$ represents the present, hence, we assume that ξ_1 is deterministic and require that the decision x_1 is deterministic, too. The latter condition is modeled by the constraint $x_1 = \mathbb{E}(x_1)$. Then we may reformulate the stochastic program (2) as optimization model

$$\min\Big\{f_1(x_1.\xi_1) + \mathbb{E}\Big(\Phi(x_1,\hat{\xi})\Big) : x_1 = E(x_1),\, x_1 \in X_1,\, A_{10}x_1 = h_1(\xi_1)\Big\} \qquad (3)$$

for the decision x_1 at time $t = 1$. With $\hat{\xi}$ denoting $\hat{\xi} := (\xi_2, \ldots, \xi_T)$ the uncertain future input process the function Φ is defined by

$$\Phi(x_1,\hat{\xi}) := \inf\Big\{\mathbb{E}\Big(\sum_{t=2}^{T} f_t(x_t,\xi_t)\Big) : x_t \in X_t, \sum_{\tau=0}^{t-1} A_{t\tau}x_{t-\tau} = h_t(\xi_t),\, t = 2,\ldots,T\Big\}. \qquad (4)$$

A solution of (3) minimizes the first period costs and also the minimal expected future costs. It is called *first-stage solution* while a (stochastic) solution $(x_2, \ldots, x_T)$ of (4) is a *second-stage solution*. Consequently, the model (3) is named *two-stage stochastic program*. We note that any first-stage solution depends on the probability distribution of the stochastic process ξ.

An often more realistic condition is to require that the decision x_t at time t depends only on the available information $(\xi_1, \ldots, \xi_t)$ and that the information flow evolves over time. This requirement is modeled by the constraints

$$x_t = \mathbb{E}(x_t|\xi_1, \xi_2, \ldots, \xi_t) \quad (t = 1, \ldots, T) \qquad (5)$$

which has to be incorporated into the constraint set of (2) and (4), respectively. The expression $\mathbb{E}(\cdot\,|\xi_1, \xi_2, \ldots, \xi_t)$ on the right-hand side of (5) is the conditional expectation with respect to the random vector $(\xi_1, \xi_2, \ldots, \xi_t)$ (and is assumed to be well defined). The constraints (5) are often called *nonanticipativity constraints* and (2) including (5) is a *multi-stage stochastic program*. We note that the constraint for $t = 1$ in (5) coincides with the former condition $x_1 = \mathbb{E}(x_1)$ in (2).

If we set $x = x_1$ and

$$X := \{x \in X_1 : x = \mathbb{E}(x),\, A_{10}x = h_1(\xi_1)\}$$
$$f_0(x,\xi) := \begin{cases} f_1(x,\xi_1) + \Phi(x,\hat{\xi}) \;, & \text{if } x \in X \text{ and } \Phi(x,\hat{\xi}) \text{ is finite} \\ +\infty \qquad\qquad\;\; , & \text{else} \end{cases}$$

the optimization model (2) is of the form (1). Suitable assumptions on ξ together with duality arguments often imply $\Phi(x,\hat{\xi}) > -\infty$. In that case, finiteness of $\Phi(x,\hat{\xi})$ is guaranteed if a feasible decision $(x_2, \ldots, x_T)$ exists for given x_1.

Most of the approaches for solving (1) numerically are based on replacing the probability distribution P by a discrete distribution with finite support $\{\xi^1, \ldots, \xi^N\}$, where the atom or *scenario* ξ^i appears with probability $p_i > 0$, $i = 1, \ldots, N$, and it holds $\sum_{i=1}^{N} p_i = 1$. The approximate stochastic program associated to (1) then is of the form

$$\min\Big\{\sum_{i=1}^{N} p_i f_0(x,\xi^i) : x \in X\Big\}. \tag{6}$$

In Section 2 we discuss how the discrete distribution should approximate the original probability distribution P. It turns out that properties of the integrand $f_0(x,\cdot)$ as function of ξ together with properties of P characterize the kind of approximation needed in stochastic programming. When looking at (3), (4) we recall that evaluations of f_0 at a pair (x,ξ) may be expensive. This leads us to one of the main *numerical challenges* in stochastic programming: A good approximation of P requires a large N, but solving (6) in a reasonable running time prefers a small(er) N.

Hence, when solving applied stochastic programming models, one might start with a large(r) N, but running time requirements might force a (much) smaller number n of scenarios. Should $\xi^{n+1},\dots,\xi^N$ just be thrown away ? In this paper, we argue that one should take advantage of the information contained in $\{\xi^1,\dots,\xi^N\}$ and look for a better approximation of P based on n scenarios out of $\{\xi^1,\dots,\xi^N\}$ by *scenario reduction*. The concept of scenario reduction and recent results in this area are reviewed in Section 3.

In this paper, we restrict our attention to two-stage stochastic programs. We take a look at the underlying theory, review available algorithms for linear and mixed-integer two-stage stochastic programs, and point out which heuristic algorithms are important in this context and where we see possibilities for improvements.

Multi-stage models and the tree structure of scenarios (due to the constraint (5)) require a (slightly) different theoretical basis and will not be considered in this paper, although scenario reduction techniques may be very helpful in this respect (see [8, 9]). Instead we refer to the literature, e.g., [2, 3, 13, 14, 15, 16, 18, 19, 25], to our recent work [8, 9] and to excellent sources of material on theory and applications of stochastic programs in [24, 26].

2 Stability and Distances of Probability Distributions

Solving stochastic programs computationally requires to replace the original probability distribution P by a probability measure Q having a finite number of scenarios and associated probabilities. Of course, Q has to be determined in such a way that infima and solutions of (1) do not change much if P is replaced by Q. In the following, we use the notations

$$v(P) := \inf\Big\{\int_{\mathbb{R}^s} f_0(x,\xi)P(d\xi) : x \in X\Big\}$$

$$S(P) := \Big\{x \in X : \int_{\mathbb{R}^s} f_0(x,\xi)P(d\xi) = v(P)\Big\}$$

for the infimum and solution set, respectively, of (1) with respect to P. If the set X is compact, the following estimates are valid

$$|v(P) - v(Q)| \leq \sup_{x \in X} \Big| \int_{\mathbb{R}^s} f_0(x, \xi)(P - Q)(d\xi) \Big| \tag{7}$$

$$\sup_{y \in S(Q)} d(y, S(P)) \leq \psi_P^{-1}\Big(\sup_{x \in X} \Big| \int_{\mathbb{R}^s} f_0(x, \xi)(P - Q)(d\xi) \Big| \Big). \tag{8}$$

In (7) $d(y, S(P))$ denotes the distance of $y \in S(Q)$ to $S(P)$. The mapping ψ_P^{-1} denotes the inverse of the growth function ψ_P of the objective in a neighborhood of the solution set $S(P)$, i.e.,

$$\psi_P(t) := \inf \Big\{ \int_{\mathbb{R}^s} f_0(x, \xi) P(d\xi) - v(P) : d(x, S(P)) \geq t,\, x \in X \Big\}.$$

The growth function ψ_P is monotonically increasing on $[0, +\infty)$ and it holds $\psi_P(0) = 0$. The estimates (7) and (8) elucidate that the distance

$$d(P, Q) := \sup_{x \in X} \Big| \int_{\mathbb{R}^s} f_0(x, \xi) P(d\xi) - \int_{\mathbb{R}^s} f_0(x, \xi) Q(d\xi) \Big| \tag{9}$$

of the two probability distributions P and Q becomes important. But, this distance is difficult to evaluate numerically since the function f_0 is often very involved (e.g. an infimum function of an optimization model as in the example in Section 1).

However, in several important model instances of functions f_0 it is possible to derive estimates of $d(P, Q)$ by other distances of probability distributions that are easier to evaluate. We mention the following possibilities:

(1) The first one is the classical Koksma-Hlawka inequality (see [17, Section 2.2] and [5]) if the functions $f_0(x, \cdot\,)$ are of bounded variation in the sense of Hardy and Krause (uniformly with respect to $x \in X$) on the support $[0, 1]^s$ of both probability distributions P and Q. The relevant distance of probability measures is the so-called $*$-discrepancy and is defined in (14).

(2) A second idea is to determine a set $\mathcal{F}$ of functions from a closed subset Ξ of $\mathbb{R}^s$ containing the support of both probability distributions P and Q to $\mathbb{R}$ such that $Cf_0(x, \cdot\,) \in \mathcal{F}$ for some constant $C > 0$ not depending on x. Then it holds

$$d(P, Q) \leq \frac{1}{C} d_{\mathcal{F}}(P, Q)$$

$$d_{\mathcal{F}}(P, Q) := \sup_{f \in \mathcal{F}} \Big| \int_{\Xi} f(\xi) P(d\xi) - \int_{\Xi} f(\xi) Q(d\xi) \Big|.$$

Of course, the set $\mathcal{F}$ should be selected such that the distance $d_{\mathcal{F}}(P, Q)$ is easier to evaluate than $d(P, Q)$.

For two-stage stochastic programs that are (i) *linear* or (ii) *mixed-integer linear* the following sets $\mathcal{F}$ of functions are relevant (cf. [21, 23, 22]):

(i) $\mathcal{F}_r(\Xi) = \{f : \Xi \to \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq c_r(\xi, \tilde{\xi}),\, \forall \xi, \tilde{\xi} \in \Xi\}$,

(ii) $\mathcal{F}_{r,\mathcal{B}}(\Xi) = \{f\mathbb{1}_B : f \in \mathcal{F}_r(\Xi),\ |f(\xi)| \le \max\{1, |\xi|^r\},\ \forall \xi \in \Xi,\ B \in \mathcal{B}\}$,

where $r \in \mathbb{N}$, $\mathcal{B}$ denotes a set of polyhedral subsets of Ξ, $|\cdot|$ is a norm on $\mathbb{R}^s$ and $\mathbb{1}_B$ is the characteristic function of the set B, i.e., $\mathbb{1}_B(\xi) = 1$ for $\xi \in B$ and $\mathbb{1}_B(\xi) = 0$ for $\xi \in \Xi \setminus B$. Furthermore, the function c_r is defined by

$$c_r(\xi, \tilde{\xi}) := \max\{1, |\xi|^{r-1}, |\tilde{\xi}|^{r-1}\}|\xi - \tilde{\xi}| \quad (\xi, \tilde{\xi} \in \Xi).$$

We note that the choice of r and of $\mathcal{B}$ depends on the structure of the stochastic program.

While the second stage infima are locally Lipschitz continuous as functions of ξ in case of linear two-stage models and belong to $\mathcal{F}_r(\Xi)$ for some $r \in \mathbb{N}$, they are discontinuous on boundaries of certain polyhedral sets for mixed-integer linear models. Hence, the set $\mathcal{F}_{r,\mathcal{B}}(\Xi)$ is relevant in the latter case.

An important special case of the class $\mathcal{B}$ is the set $\mathcal{B}_{\text{rect}}$ of all rectangular sets in $\mathbb{R}^s$, i.e.

$$\mathcal{B}_{\text{rect}} = \{I_1 \times \cdots \times I_s : \emptyset \neq I_j \text{ is a closed interval in } \mathbb{R},\ j = 1, \ldots, s\}, \tag{10}$$

which is relevant for the stability of pure integer second stage models.

In case (i) so-called *Fortet-Mourier metrics* of order r

$$\zeta_r(P, Q) = \sup_{f \in \mathcal{F}_r(\Xi)} \left| \int_\Xi f(\xi) P(d\xi) - \int_\Xi f(\xi) Q(d\xi) \right| \tag{11}$$

appear as special instances of $d_{\mathcal{F}}$, and in case (ii) the metrics

$$\zeta_{r,\mathcal{B}}(P, Q) = \sup_{f \in \mathcal{F}_r(\Xi), B \in \mathcal{B}} \left| \int_B f(\xi) P(d\xi) - \int_B f(\xi) Q(d\xi) \right|. \tag{12}$$

Since a sequence of probability measures on $\mathbb{R}^s$ converges with respect to $\zeta_{r,\mathcal{B}}$ if and only if it converges with respect to ζ_r and with respect to the $\mathcal{B}$-discrepancy

$$\alpha_{\mathcal{B}}(P, Q) := \sup_{B \in \mathcal{B}} |P(B) - Q(B)|,$$

respectively (cf. [12]), one may consider the 'mixed' distance

$$d_\lambda(P, Q) = \lambda\, \alpha_{\mathcal{B}}(P, Q) + (1 - \lambda)\, \zeta_r(P, Q) \tag{13}$$

for some $\lambda \in (0, 1]$ instead of the more complicated metric $\zeta_{r,\mathcal{B}}$. Two specific discrepancies are of interest in this paper. The first one is the rectangular discrepancy $\alpha_{\mathcal{B}_{\text{rect}}}$ which is needed in Section 3.2. The second one is the $*$-discrepancy

$$\alpha^*(P, Q) = \sup_{x \in [0,1]^s} |P([0, x)) - Q([0, x))|, \tag{14}$$

where P and Q are probability measures on $[0, 1]^s$ and $[0, x) = \times_{i=1}^s [0, x_i)$. As is known from [11] discrepancies are (much) more difficult to handle in scenario reduction compared to Fortet-Mourier metrics (see also Section 3).

If both P and Q are discrete probability distributions (with finite support), the Fortet-Mourier distance $\zeta_r(P,Q)$ may be computed as optimal value of certain linear programs (see also [7]). We show in Section 3.2 that the distance $d_\lambda(P,Q)$ may be computed as well by linear programming (see also [10, 12]).

If the set Ξ is compact the Fortet-Mourier metric ζ_r admits a dual representation as Monge-Kantorovich transportation functional. In particular, it holds for all probability measures on Ξ with finite rth moment (see [20, Section 4.1]):

$$\zeta_r(P,Q) := \inf\Big\{ \int_{\Xi\times\Xi} \hat{c}_r(\xi,\tilde{\xi})\eta(d\xi, d\tilde{\xi}) : \eta \in M(P,Q)\Big\}, \tag{15}$$

where $M(P,Q)$ denotes the set of all probability measures η on $\Xi\times\Xi$ whose marginal distributions on Ξ are just P and Q, respectively. Furthermore, the function $\hat{c}_r$ is the *reduced cost function* associated with the cost c_r and is defined by

$$\hat{c}_r(\xi,\tilde{\xi}) := \inf\Big\{ \sum_{j=1}^{n+1} c_r(z_{j-1}, z_j) : z_0 = \xi,\ z_{n+1} = \tilde{\xi},\ z_j \in \Xi,\ n \in \mathbb{N}\Big\}. \tag{16}$$

The function $\hat{c}_r$ is a metric on Ξ and it holds $\hat{c}_r \le c_r$ (cf. [20, Chapter 4.1.3]).

3 Optimal Scenario Reduction

Let P be a probability distribution on $\mathbb{R}^s$ with finite support consisting of N scenarios ξ^i and their probabilities p_i, $i \in I := \{1,\dots,N\}$. The basic idea of optimal scenario reduction consists in determining a probability distribution Q_n which is the best approximation of P with respect to a given distance d of probability measures and whose support consists of a subset of $\{\xi^1,\dots,\xi^N\}$ with $n < N$ elements. This means

$$d(P,Q_n) = \inf\{d(P,Q) : Q(\mathbb{R}^s) = 1, \mathrm{supp}(Q) \subset \mathrm{supp}(P), |\mathrm{supp}(Q)| = n\}. \tag{17}$$

An equivalent formulation of (17) may be obtained as follows: Let Q_J denote a probability measure on $\mathbb{R}^s$ with $\mathrm{supp}(Q_J) = \{\xi^i : i \in \{1,\dots,N\} \setminus J\}$ for some index set $J \subset \{1,\dots,N\}$ and let q_i, $i \in \{1,\dots,N\} \setminus J$, be the probability of scenario indexed by i. Then the minimization problem

$$\min\Big\{ d(P,Q_J) : J \subset I, |J| = N-n, q_i \ge 0, i \in I\setminus J, \sum_{i\in I\setminus J} q_i = 1\Big\} \tag{18}$$

determines some index set J^* and $q_i^* \in [0,1]$ such that the probability measure with scenarios ξ^i and probabilities q_i^* for $i \in \{1,\dots,N\}\setminus J^*$ solves (17). The second formulation (18) of the optimal scenario reduction problem has the advantage to provide a decomposition into an *inner* and an *outer* minimization problem, namely,

$$\min_J\Big\{ \inf_q \Big\{ d(P,Q_J) : q_i \ge 0, i \in I\setminus J, \sum_{i\in I\setminus J} q_i = 1\Big\} : J \subset I, |J| = N-n\Big\}. \tag{19}$$

As suggested in [4] the distance d has to be selected such that the stochastic program (1) behaves stable with respect to d (cf. Section 2).

3.1 Two-Stage Stochastic Linear Programs

Here, the distance d is just the Fortet-Mourier metric ζ_r for some $r \in \mathbb{N}$ which depends on the structure of the underlying stochastic programming model. For a discussion of the choice of r it is referred to [21, 23].

Since the distance $\zeta_r(P, Q_J)$ has a dual representation as mass transportation problem for any index set J (see (15)), the inner minimization problem in (19) can be solved explicitly. Namely, it holds

$$D_J := \min\Big\{\zeta_r(P, Q_J) : q_i \geq 0,\, i \notin J, \sum_{i \notin J} q_i = 1\Big\} = \sum_{j \in J} p_j \min_{i \notin J} \hat{c}_r(\xi^i, \xi^j), \quad (20)$$

where the reduced costs $\hat{c}_r$ have the simplified representation

$$\hat{c}_r(\xi^i, \xi^j) := \min\Big\{\sum_{k=1}^{n-1} c_r(\xi^{l_k}, \xi^{l_{k+1}}) : n \in \mathbb{N}, l_k \in I, l_1 = i, l_n = j\Big\} \quad (21)$$

as shortest path problem compared with the general form (16). The optimal probabilities q_i^*, $i \notin J$, are given by the *redistribution rule*

$$q_i^* = p_i + \sum_{\substack{j \in J \\ i(j) = i}} p_j \quad \text{and} \quad i(j) \in \arg\min_{i \notin J} \hat{c}_r(\xi^i, \xi^j), \quad i \notin J, \quad (22)$$

The redistribution rule consists in adding the probability of a deleted scenario $j \in J$ to the probability of a scenario that is nearest to ξ^j with respect to the distance $\hat{c}_r$ on $\mathbb{R}^s$.

The outer minimization problem is of the form

$$\min\Big\{D_J = \sum_{j \in J} p_j \min_{i \in I \setminus J} \hat{c}_r(\xi^i, \xi^j) : J \subset I, |J| = N - n\Big\} \quad (23)$$

and represents a combinatorial optimization problem of n-median type. It is $\mathcal{NP}$-hard and suggests to use heuristic or approximation algorithms. We also refer to [1] where a branch-and-cut-and-algorithm is developed and tested on large scale instances.

In our earlier work [6, 7] we proposed two simple heuristic algorithms: the *forward (selection)* and the *backward (reduction) heuristic*. To give a short description of the two algorithms, let

$$c_{ij} := \hat{c}_r(\xi^i, \xi^j) \quad (i, j = 1, \ldots, N).$$

The basic idea of the forward algorithm originates from the simple structure of (23) for the special case $n = 1$. It is of the form

$$\min_{u \in \{1, \ldots, N\}} \sum_{\substack{j=1 \\ j \neq u}}^{N} p_j c_{uj}.$$

If the minimum is attained at u^*, the index set $J = \{1, \ldots, N\} \setminus \{u^*\}$ solves (23). The scenario ξ^{u^*} is taken as the first element of $\mathrm{supp}(Q)$. Then the separable

structure of D_J is exploited to determine the second element of $\mathrm{supp}(Q)$ while the first element is fixed. The process is continued until n elements of $\mathrm{supp}(Q)$ are selected.

Forward algorithm for scenario reduction

Step 0: $J^{[0]} := \{1, \dots, N\}$.

Step k: $u_k \in \arg\min_{u \in J^{[k-1]}} \sum_{j \in J^{[k-1]} \setminus \{u\}} p_j \min_{i \notin J^{[k-1]} \setminus \{u\}} c_{ij}$,

$J^{[k]} := J^{[k-1]} \setminus \{u_k\}$.

Step n+1: Redistribution with $J := J^{[n]}$ via (22).

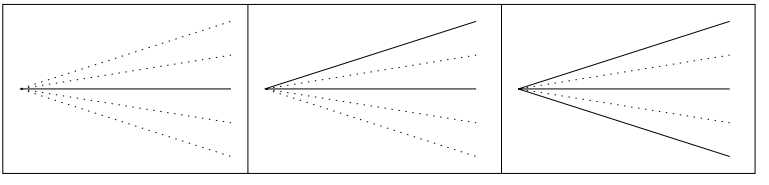

Fig. 1. Illustration of selecting the first, second and third scenario out of $N = 5$

The idea of the backward algorithm is based on the second special case of (23) for $n = N - 1$. It is of the form

$$\min_{l \in \{1,\dots,N\}} p_l \min_{i \neq l} c_{il}.$$

If the minimum is attained at l^*, the index set $J = \{l^*\}$ solves (23) in case $n = N - 1$. After fixing the remaining index set $I \setminus \{l^*\}$ a second scenario is reduced etc. The process is continued until $N - n$ scenarios are reduced.

Backward algorithm of scenario reduction

Step 0: $J^{[0]} := \emptyset$.

Step k: $l_k \in \arg\min_{l \notin J^{[k-1]}} \sum_{j \in J^{[k-1]} \cup \{l\}} p_j \min_{i \notin J^{[k-1]} \cup \{l\}} c_{ij}$,

$J^{[k]} := J^{[k-1]} \cup \{l_k\}$.

Step N-n+1: Redistribution with $J := J^{[N-n]}$ via (22).

It is shown in [6] that both heuristics are polynomial time algorithms and that the forward heuristics is recommendable at least if $n \leq \frac{N}{4}$.

3.2 Two-Stage Stochastic Mixed-Integer Programs

The relevant probability metric is now $\zeta_{r,\mathcal{B}}$, where $r \in \mathbb{N}$ and $\mathcal{B}$ is a set of polyhedral subsets of a (polyhedral) set Ξ that contains the support of P. It was

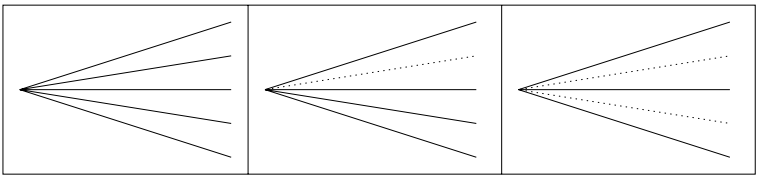

Fig. 2. Illustration of the original scenario set with $N = 5$ and of reducing the first and the second scenario

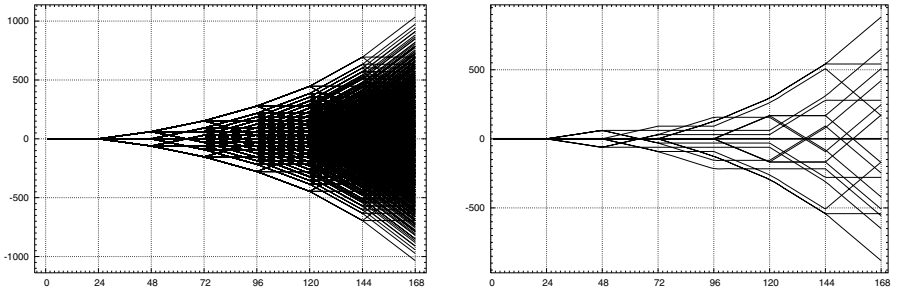

Fig. 3. Left: $N = 729$ equally distributed scenarios in form of a regular ternary tree. Right: Reduced scenario set with $n = 20$ and line width proportional to scenario probabilities obtained by the forward heuristic with $r = 2$.

mentioned in Section 2 that both r and $\mathcal{B}$ depend on the structure of the particular stochastic program. Since the complexity of scenario reduction algorithms increase if the sets in $\mathcal{B}$ get more involved, we follow here [12] and consider only the set $\mathcal{B}_{\text{rect}}$ (see (10)) and the distance

$$d_\lambda(P, Q) := \lambda \alpha_{\mathcal{B}_{\text{rect}}}(P, Q) + (1 - \lambda)\zeta_r(P, Q) \tag{24}$$

for some $\lambda \in (0, 1]$.

As in Section 3.1 we are interested in computing

$$D_J := \min \Big\{ d_\lambda(P, Q_J) : q_i \geq 0,\ i \notin J,\ \sum_{i \notin J} q_i = 1 \Big\}. \tag{25}$$

In the following, we show that D_J can be computed as optimal value of a linear program (but a closed formula for D_J as in Section 3.1 is not available in general). To this end, we assume without loss of generality that $J = \{n + 1, \ldots, N\}$, i.e., $\text{supp}(Q_J) = \{\xi^1, \ldots, \xi^n\}$ for some $1 \leq n < N$, consider the system of index sets

$$\mathcal{I}_{\mathcal{B}_{\text{rect}}} := \{I(B) := \{i \in \{1, \ldots, N\} : \xi^i \in B\} : B \in \mathcal{B}_{\text{rect}}\}$$

and obtain the following representation of the rectangular discrepancy

$$\alpha_{\mathcal{B}_{\rm rect}}(P,Q_J) = \sup_{B\in\mathcal{B}_{\rm rect}} |P(B)-Q_J(B)| = \max_{I\in\mathcal{I}_{\mathcal{B}_{\rm rect}}} \left| \sum_{i\in I} p_i - \sum_{j\in I\cap\{1,\ldots,n\}} q_j \right| \quad (26)$$

$$= \min\left\{ t_\alpha \,\middle|\, \begin{array}{l} -\sum_{j\in I\cap\{1,\ldots,n\}} q_j \le t_\alpha - \sum_{i\in I} p_i,\ I\in\mathcal{I}_{\mathcal{B}_{\rm rect}} \\ \sum_{j\in I\cap\{1,\ldots,n\}} q_j \le t_\alpha + \sum_{i\in I} p_i,\ I\in\mathcal{I}_{\mathcal{B}_{\rm rect}} \end{array} \right\} \quad (27)$$

Since the set $\mathcal{I}_{\mathcal{B}_{\rm rect}}$ may be too large to solve the linear program (27) numerically, we consider the system of reduced index sets

$$\mathcal{I}^*_{\mathcal{B}_{\rm rect}} := \{I(B)\cap\{1,\ldots,n\} : B\in\mathcal{B}_{\rm rect}\}$$

and the quantities

$$\gamma^{I^*} := \max\Big\{ \sum_{i\in I} p_i : I\in\mathcal{I}_{\mathcal{B}_{\rm rect}}, I\cap\{1,\ldots,n\} = I^* \Big\}$$

$$\gamma_{I^*} := \min\Big\{ \sum_{i\in I} p_i : I\in\mathcal{I}_{\mathcal{B}_{\rm rect}}, I\cap\{1,\ldots,n\} = I^* \Big\}$$

for every $I^*\in\mathcal{I}^*_{\mathcal{B}_{\rm rect}}$. Since any such index set I^* corresponds to some left-hand side of the inequalities in (27), γ^{I^*} and γ_{I^*} correspond to the smallest right-hand sides in (27). Hence, the rectangular discrepancy may be rewritten as

$$\alpha_{\mathcal{B}_{\rm rect}}(P,Q_J) = \min\left\{ t_\alpha \,\middle|\, \begin{array}{l} -\sum_{j\in I^*} q_j \le t_\alpha - \gamma^{I^*},\ I^*\in\mathcal{I}^*_{\mathcal{B}_{\rm rect}} \\ \sum_{j\in I^*} q_j \le t_\alpha + \gamma_{I^*},\ I^*\in\mathcal{I}^*_{\mathcal{B}_{\rm rect}} \end{array} \right\}. \quad (28)$$

Since the number of elements of $\mathcal{I}^*_{\mathcal{B}_{\rm rect}}$ is at most 2^n (compared to 2^N in $\mathcal{I}_{\mathcal{B}_{\rm rect}}$), passing from (27) to (28) indeed drastically reduces the maximum number of inequalities and may make the linear program (28) numerically tractable.

Due to (15) the Fortet-Mourier distance $\zeta_r(P,Q_J)$ allows the representation as linear program

$$\zeta_r(P,Q_J) = \inf\left\{ \sum_{i=1}^N\sum_{j=1}^n \eta_{ij}\hat{c}_r(\xi^i,\xi^j) \,\middle|\, \eta_{ij}\ge 0,\ \begin{array}{l} \sum_{i=1}^N \eta_{i,j} = q_j,\ j=1,\ldots,n \\ \sum_{j=1}^n \eta_{i,j} = p_i,\ i=1,\ldots,N \end{array} \right\}$$

where the reduced cost function $\hat{c}_r$ is given by (21).

Hence, extending the representation (28) of $\alpha_{\mathcal{B}_{\rm rect}}$ we obtain the following linear program for determining D_J and the probabilities q_j, $j=1,\ldots,n$, of the discrete reduced distribution Q_J,

$$D_J = \min\left\{ \lambda\, t_\alpha + (1-\lambda) t_\zeta \,\middle|\, \begin{array}{l} t_\alpha, t_\zeta \ge 0,\ q_j\ge 0,\ \sum_{j=1}^n q_j = 1, \\ \eta_{ij}\ge 0, i=1,\ldots,N,\ j=1,\ldots,n, \\ t_\zeta \ge \sum_{i=1}^N\sum_{j=1}^n \hat{c}_r(\xi^i,\xi^j)\eta_{ij}, \\ \sum_{j=1}^n \eta_{ij} = p_i,\ i=1,\ldots,N, \\ \sum_{i=1}^N \eta_{ij} = q_j,\ j=1,\ldots,n, \\ -\sum_{j\in I^*} q_j \le t_\alpha - \gamma^{I^*},\ I^*\in\mathcal{I}^*_{\mathcal{B}_{\rm rect}} \\ \sum_{j\in I^*} q_j \le t_\alpha + \gamma_{I^*},\ I^*\in\mathcal{I}^*_{\mathcal{B}_{\rm rect}} \end{array} \right\} \quad (29)$$

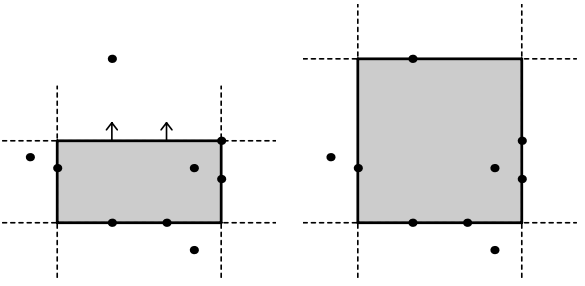

Fig. 4. Non-supporting rectangle (left) and supporting rectangle (right). The dots represent the remaining scenarios $\xi^1, \ldots, \xi^n$ for $s = 2$.

While the linear program (29) can be solved efficiently by available software, determining the index set $\mathcal{I}^*_{\mathcal{B}_{\text{rect}}}$ and the coefficients γ^{I^*}, γ_{I^*} is more intricate.

It is shown in [10, Section 3] that the parameters $\mathcal{I}^*_{\mathcal{B}_{\text{rect}}}$ and γ^{I^*}, γ_{I^*} can be determined by studying the set $\mathcal{R}$ of supporting rectangles. A rectangle B in $\mathcal{B}_{\text{rect}}$ is called *supporting* if each of its facets contains an element of $\{\xi_1, \ldots, \xi_n\}$ in its relative interior (see also Fig. 4). Based on $\mathcal{R}$ the following representations are valid according to [10, Prop. 1 and 2]:

$$\mathcal{I}^*_{\mathcal{B}_{\text{rect}}} = \bigcup_{B \in \mathcal{R}} \left\{ I^* \subseteq \{1, \ldots, n\} : \cup_{j \in I^*} \{\xi^j\} = \{\xi^1, \ldots, \xi^n\} \cap \operatorname{int} B \right\}$$

$$\gamma^{I^*} = \max \left\{ P(\operatorname{int} B) : B \in \mathcal{R} \cup_{j \in I^*} \{\xi^j\} = \{\xi^1, \ldots, \xi^n\} \cap \operatorname{int} B \right\}$$

$$\gamma_{I^*} = \sum_{i \in \underline{I}} p_i \quad \text{where } \underline{I} := \left\{ i \in \{1, \ldots, N\} : \min_{j \in I^*} \xi^j_l \leq \xi^i_l \leq \max_{j \in I^*} \xi^j_l,\ l = 1, \ldots, s \right\}$$

for every $I^* \in \mathcal{I}^*_{\mathcal{B}_{\text{rect}}}$. Here, $\operatorname{int} B$ denotes the interior of the set B.

An algorithm is developed in [12] that constructs recursively l-dimensional supporting rectangles for $l = 1, \ldots, s$. Computational experiments show that its running time grows linearly with N, but depends on n and s via the expression $\binom{n+1}{2}^s$. Hence, while N may be large, only moderately sized values of n given s seem to be realistic.

Since an algorithm for computing D_J is now available, we finally look at determining a scenario index set $J \subset I = \{1, \ldots, N\}$ with cardinality $|J| = N - n$ such that D_J is minimal, i.e., at solving the combinatorial optimization problem

$$\min\{D_J : J \subset I, |J| = N - n\} \tag{30}$$

which is known as n-median problem and as $\mathcal{NP}$-hard. One possibility is to reformulate (30) as mixed-integer linear program and to solve it by standard software. Since, however, approximate solutions of (30) are sufficient, heuristic algorithms like *forward selection* are of interest, where u_k is determined in its kth step such that it solves the minimization problem

$$\min\left\{ D_{J^{[k-1]}\setminus\{u\}} \;\middle|\; u \in J^{[k-1]} \right\},$$

where $J^{[0]} = \{1,\dots,N\}$, $J^{[k]} := J^{[k-1]} \setminus \{u_k\}$ $(k = 1,\dots,n)$ and $J^{[n]} := \{1,\dots,N\} \setminus \{u_1,\dots,u_n\}$ serves as approximate solution to (30). Recalling that the complexity of evaluating $D_{J^{[k-1]}\setminus\{u\}}$ for some $u \in J^{[k-1]}$ is proportional to

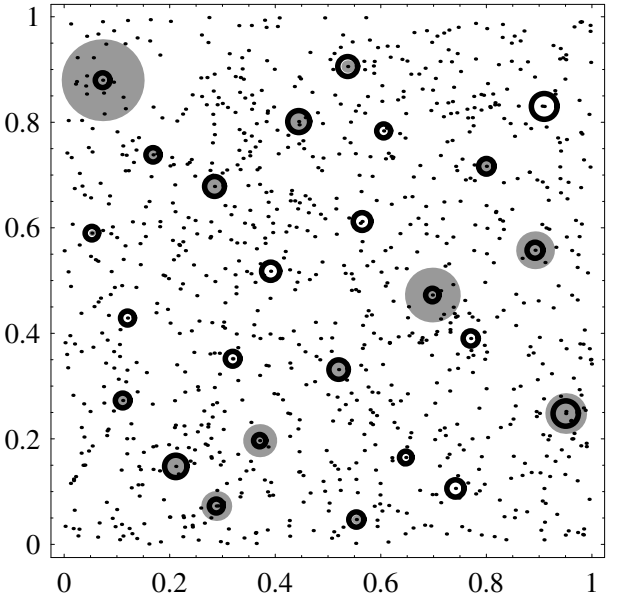

Fig. 5. $N = 1\,000$ samples ξ^i from the uniform distribution on $[0,1]^2$ and $n = 25$ points ξ^{u_k}, $k = 1,\dots,n$, obtained via the first 25 elements z_k, $k = 1,\dots,n$, of the *Halton sequence* (in the bases 2 and 3) (see [17, p. 29]). The probabilities q_k of ξ^{u_k}, $k = 1,\dots,n$, are computed for the distance d_λ with $\lambda = 1$ (gray balls) and $\lambda = 0.9$ (black circles) by solving (29). The diameters of the circles are proportional to the probabilities q_k, $k = 1,\dots,25$.

$\binom{k+1}{2}^s$ shows that even the forward selection algorithm is expensive.

Hence, heuristics for solving (30) become important that require only a low number of D_J evaluations. For example, if P is a probability distribution on $[0,1]^s$ with independent marginal distributions P_j, $j = 1,\dots,s$, such a heuristic can be based on Quasi-Monte Carlo methods (cf. [17]). The latter provide sequences of equidistributed points in $[0,1]^s$ that approximate the uniform distribution on the unit cube $[0,1]^s$. Now, let n Quasi-Monte Carlo points $z^k = (z_1^k,\dots,z_s^k) \in [0,1]^s$, $k = 1,\dots,n$, be given. Then we determine

$$y^k := \left(F_1^{-1}(z_1^k),\dots,F_s^{-1}(z_s^k)\right) \qquad (k = 1,\dots,n),$$

where F_j is the (one-dimensional) distribution function of P_j, i.e.,

$$F_j(z) = P_j((-\infty, z]) = \sum_{i=1,\, \xi_j^i \le z}^{N} p_i \qquad (z \in \mathbb{R})$$

and $F_j^{-1}(t) := \inf\{z \in \mathbb{R} : F_j(z) \ge t\}$ $(t \in [0,1])$ its inverse $(j = 1, \ldots, s)$. Finally, we determine u_k as solution of

$$\min_{u \in J^{[k-1]}} |\xi^u - y^k|$$

and set again $J^{[k]} := J^{[k-1]} \setminus \{u_k\}$ for $k = 1, \ldots, n$, where $J^{[0]} = \{1, \ldots, N\}$. Figure 5 illustrates the results of such a Quasi-Monte Carlo based heuristic and Figure 6 shows the discrepancy $\alpha_{\mathcal{B}_{\text{rect}}}$ for different n and the running times of the Quasi-Monte Carlo based heuristic.

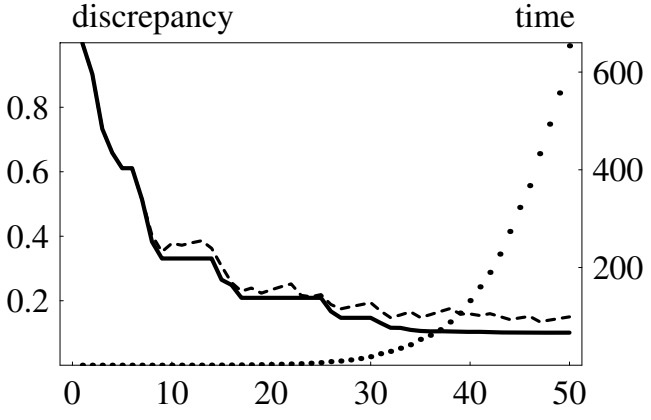

Fig. 6. Distance $\alpha_{\mathcal{B}_{\text{rect}}}$ between P (with $N = 1000$) and equidistributed QMC-points (dashed), QMC-points, whose probabilities are adjusted according to (29) (bold), and running times of the QMC-based heuristic (in seconds)

Acknowledgements. The work of the author is supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

References

[1] Avella, P., Sassano, A., Vasil'ev, I.: Computational study of large scale p-median problems. Math. Program. 109, 89–114 (2007)
[2] Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming. Math. Oper. Res. 30, 615–631 (2005)
[3] Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. Ann. Oper. Res. 100, 25–53 (2000)
[4] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics. Math. Program. 95, 493–511 (2003)

[5] Götz, M.: Discrepancy and the error in integration. Monatsh. Math. 136, 99–121 (2002)
[6] Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comp. Optim. Appl. 24, 187–206 (2003)
[7] Heitsch, H., Römisch, W.: A note on scenario reduction for two-stage stochastic programs. Oper. Res. Let. 35, 731–736 (2007)
[8] Heitsch, H., Römisch, W.: Scenario tree modeling for multistage stochastic programs. Math. Program. 118, 371–406 (2009)
[9] Heitsch, H., Römisch, W.: Scenario tree reduction for multistage stochastic programs. Comp. Manag. Science 6, 117–133 (2009)
[10] Henrion, R., Küchler, C., Römisch, W.: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming. J. Ind. Manag. Optim. 4, 363–384 (2008)
[11] Henrion, R., Küchler, C., Römisch, W.: Scenario reduction in stochastic programming with respect to discrepancy distances. Comp. Optim. Appl. 43, 67–93 (2009)
[12] Henrion, R., Küchler, C., Römisch, W.: A scenario reduction heuristic for two-stage stochastic mixed-integer programming (in preparation)
[13] Hochreiter, R., Pflug, G.C.: Financial scenario generation for stochastic multistage decision processes as facility location problems. Ann. Oper. Res. 152, 257–272 (2007)
[14] Høyland, K., Kaut, M., Wallace, S.W.: A heuristic for moment-matching scenario generation. Comp. Optim. Appl. 24, 169–185 (2003)
[15] Kuhn, D.: Generalized bounds for convex multistage stochastic programs. LNEMS, vol. 548. Springer, Berlin (2005)
[16] Kuhn, D.: Aggregation and discretization in multistage stochastic programming. Math. Program. 113, 61–94 (2008)
[17] Niederreiter, H.: Random number generation and Quasi-Monte Carlo methods. CBMS-NSF Regional Conference Series in Applied Mathematics Vol. 63. SIAM, Philadelphia (1992)
[18] Pagès, G., Pham, H., Printems, J.: Optimal quantization methods and applications to numerical problems in finance. In: Rachev, S.T. (ed.) Handbook of Computational and Numerical Methods in Finance, pp. 253–297. Birkhäuser, Boston (2004)
[19] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures. Math. Program. 116, 461–479 (2009)
[20] Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems, vol. I, II. Springer, Berlin (1998)
[21] Römisch, W.: Stability of Stochastic Programming Problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10, pp. 483–554. Elsevier, Amsterdam (2003)
[22] Römisch, W., Vigerske, S.: Quantitative stability of fully random mixed-integer two-stage stochastic programs. Optim. Lett. 2, 377–388 (2008)
[23] Römisch, W., Wets, R.J.-B.: Stability of ε-approximate solutions to convex stochastic programs. SIAM J. Optim. 18, 961–979 (2007)
[24] Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
[25] Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems. Math. Meth. Oper. Res. 58, 57–68 (2003)
[26] Wallace, S.W., Ziemba, W.T. (eds.): Applications of Stochastic Programming. MPS-SIAM Series in Optimization (2005)