Mathematical Programming manuscript No.
(will be inserted by the editor)

J. Dupačová · N. Gröwe-Kuska · W. Römisch

Scenario Reduction in Stochastic Programming

An Approach using Probability Metrics

Received: date / Revised version: date

Abstract. Given a convex stochastic programming problem with a discrete initial probability distribution, the problem of optimal scenario reduction is stated as follows: Determine a scenario subset of prescribed cardinality and a probability measure based on this set that is the closest to the initial distribution in terms of a natural (or canonical) probability metric. Arguments from stability analysis indicate that Fortet-Mourier type probability metrics may serve as such canonical metrics. Efficient algorithms are developed that determine optimal reduced measures approximately. Numerical experience is reported for reductions of electrical load scenario trees for power management under uncertainty. For instance, it turns out that after 50% reduction of the scenario tree the optimal reduced tree still has about 90% relative accuracy.

Key words. Stochastic programming – quantitative stability – Fortet-Mourier metrics – scenario reduction – transportation problem – electrical load scenario tree

1. Introduction

Various important real-life decision problems can be formulated as convex stochastic programs which can be mostly written in the form

$$\min_{x\in X} \mathbb{E}_P f(\omega,x) = \int_\Omega f(\omega,x)P(d\omega). \tag{1}$$

Here, $X \subset \mathbb{R}^n$ is a given nonempty convex closed set, Ω a closed subset of $\mathbb{R}^s$ and $\mathcal{B}$ the Borel σ-field relative to Ω, the function f from $\Omega\times\mathbb{R}^n$ to the extended reals $\overline{\mathbb{R}}$ is measurable with respect to ω and lower semicontinuous and convex with respect to x, and P a fixed probability measure on $(\Omega,\mathcal{B})$, i.e., $P \in \mathcal{P}(\Omega)$, with $\mathbb{E}_P$ denoting expectation with respect to P. This formulation covers two- and multi-stage stochastic programs with recourse. In these cases, X is the set of feasible first-stage decisions and the function values $f(\omega,x)$ evaluate the best possible outcomes of decisions x in case that ω is observed.

J. Dupačová: Charles University Prague, Department of Probability and Mathematical Statistics, 18675 Prague 8, Czech Republic, e-mail: `dupacova@karlin.mff.cuni.cz`

N. Gröwe-Kuska: Humboldt-University Berlin, Institute of Mathematics, 10099 Berlin, Germany, e-mail: `nicole@mathematik.hu-berlin.de`

W. Römisch: Humboldt-University Berlin, Institute of Mathematics, 10099 Berlin, Germany, e-mail: `romisch@mathematik.hu-berlin.de`

Mathematics Subject Classification (1991): 90C15, 90C31

Apparently, (1) is a mathematical program. However, several important problems concerning its formulation and algorithmic approaches have to be resolved. The probability measure P does not need to be known precisely, the function f – the random objective – is given implicitly (e.g. as an (iterated) optimal value), and at the same time, (1) is often large scale. To solve (1), various problem specific procedures have been designed that approximate the model and exploit its special structure. Many approximation methods exploit a discrete probability measure having the property that the corresponding optimal value and the set of ε-approximate solutions are close to $v(P)$ and $S_\varepsilon(P)$, respectively. Here,

$$v(P) := \inf\{\mathbb{E}_P f(\omega, x) : x \in X\}$$
$$S_\varepsilon(P) := \{x \in X : \mathbb{E}_P f(\omega, x) \le v(P) + \varepsilon\}$$

for each $\varepsilon \ge 0$. Consistently, $S(P) := S_0(P)$ denotes the solution set of (1).
In this context, stability properties of the model (1) with respect to perturbations (approximations) of P become important (see e.g. the surveys [6,21] and references therein, and [10,19,22]). In Section 2 we present a quantitative stability result for optimal values and ε-approximate solution sets of (1) that is based on the general perturbation analysis in [1,18], and we demonstrate that *probability (semi-) metrics with ζ-structure* (cf. [15,24]), i.e.,

$$d_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} \left| \int_\Omega f(\omega) P(d\omega) - \int_\Omega f(\omega) Q(d\omega) \right| \tag{2}$$

with $\mathcal{F}$ being a class of measurable functions from Ω to $\mathbb{R}$ and P, Q belonging to $\mathcal{P}(\Omega)$, appear as natural and suitable distances of probability distributions for stability analysis. Furthermore, it is explained there that classes of (continuous) functions having the form

$$\mathcal{F}_c = \{f : \Omega \to \mathbb{R} : f(\omega) - f(\tilde\omega) \le c(\omega, \tilde\omega) \ \text{ for all } \ \omega, \tilde\omega \in \Omega\} \tag{3}$$

with a continuous symmetric function $c : \Omega \times \Omega \to \mathbb{R}_+$ having the property that $c(\omega, \tilde\omega) = 0$ iff $\omega = \tilde\omega$, are highly relevant in the context of convex stochastic programming models. The choice of c depends on the quantitative continuity properties of the integrand f with respect to ω and is discussed in Section 2. The following estimate is valid for the corresponding probability (semi-) metric ζ_c with ζ-structure

$$\zeta_c(P, Q) := d_{\mathcal{F}_c}(P, Q) \le \hat\mu_c(P, Q), \tag{4}$$

where $\hat\mu_c$ denotes the *Kantorovich functional*

$$\hat\mu_c(P, Q) := \inf\{ \int\limits_{\Omega\times\Omega} c(\omega, \tilde\omega)\eta(d(\omega, \tilde\omega)) : \eta \in \mathcal{P}(\Omega \times \Omega), \eta(B \times \Omega) = P(B), \tag{5}$$
$$\eta(\Omega \times B) = Q(B) \ \text{ for all } B \in \mathcal{B}\}.$$

Problem (5) is called *Monge-Kantorovich mass transportation problem* ([15,17]). In case that c has the particular form

$$c_h(\omega, \tilde\omega) = \|\omega - \tilde\omega\| \max\{1, h(\|\omega - \omega_0\|), h(\|\tilde\omega - \omega_0\|)\} \tag{6}$$

for all $\omega, \tilde{\omega} \in \Omega$, where $\|\cdot\|$ is some norm on $\mathbb{R}^n$, ω_0 some fixed element in $\mathbb{R}^s$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ continuous and nondecreasing, the corresponding metric with ζ-structure

$$\zeta_h(P, Q) := d_{\mathcal{F}_{c_h}}(P, Q)$$

is often called *Fortet-Mourier (type) metric* (cf. [15]).
An important instance is that the original probability measure P in (1) is itself discrete with finitely many atoms (or *scenarios*) or that a good discrete approximation of P is available. Its support may be very large so that due to computational complexity and running time limitations, this probability measure is further approximated by a probability measure Q carried by a (much) smaller number of atoms. In this case, the distance $\zeta_c(P, Q)$ and its upper bound, the Kantorovich functional $\hat{\mu}_c(P, Q)$, represent optimal values of certain linear programs. To be more precise, let P and Q be probability measures with scenarios $\{\omega_1, \ldots, \omega_N\}$ and $\{\tilde{\omega}_1, \ldots, \tilde{\omega}_M\}$, and probability weights $\{p_1, \ldots, p_N\}$ and $\{q_1, \ldots, q_M\}$, respectively. Thus, $P = \sum_{i=1}^{N} p_i \delta_{\omega_i}$ and $Q = \sum_{j=1}^{M} q_j \delta_{\tilde{\omega}_j}$, where $\delta_\omega \in \mathcal{P}(\Omega)$ denotes the Dirac measure placing unit mass at ω. Then the Kantorovich functional has the primal-dual representation

$$\begin{aligned}\hat{\mu}_c(P, Q) &= \min\{\sum_{i=1}^{N}\sum_{j=1}^{M} c(\omega_i, \tilde{\omega}_j)\eta_{ij} : \eta_{ij} \geq 0, \sum_{i=1}^{N} \eta_{ij} = q_j, \sum_{j=1}^{M} \eta_{ij} = p_i \forall i, j\} \\ &= \max\{\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{M} q_j v_j : u_i + v_j \leq c(\omega_i, \tilde{\omega}_j) \forall i, j\},\end{aligned}$$

i.e., $\hat{\mu}_c(P, Q)$ represents the optimal value of linear transportation problem. In particular, the functional $\hat{\mu}_c$ can be used to evaluate distances of specific probability measures obtained during a scenario-reduction process, i.e., in case that $\{\tilde{\omega}_1, \ldots, \tilde{\omega}_M\}$ is a subset of $\{\omega_1, \ldots, \omega_N\}$. Various reduction rules appear in the context of recent large-scale real-life applications. There are purely heuristic and ad hoc rules, e.g. [2,3], heuristic rules inspired by the contamination technique, cf. [7], and various sampling schemes. In [5] deletion is based on criteria for the expected value of perfect information (EVPI). A rule based on a uniform approximation of the random objective function f over the whole scenario space and independently on the decision vector is designed in [23]. A deletion rule based on maintaining first and second order moments is used in [4]. For more information on recent work for scenario generation and reduction we refer to [9].
In Section 3 we study a novel scenario reduction approach that is based on best approximations in terms of the functional $\hat{\mu}_c$. We show that the Kantorovich functional of a discrete original probability distribution P and the optimal reduced measure Q based on a given subset of scenarios of P as well as the optimal weights of Q can be computed explicitly, i.e., without solving a transportation problem. Furthermore, we derive two heuristic algorithms for determining the optimal subset of scenarios of P with given cardinality. Here, optimality always means closeness in terms of $\hat{\mu}_c$.
In Section 4 we report on numerical experience for the reduction of a scenario

tree that represents an approximation of the electrical load process in a power management model under uncertainty. It turns out that both algorithmic approaches for determining recursively the scenario subset to be deleted (*backward reduction*) and the set of remaining scenarios (*forward selection*), respectively, work reasonably well and efficient. The reduced subtrees obtained by forward selection are slightly better, but their computation requires higher CPU times. Somewhat surprisingly, a reduction of the scenario tree by 50% of the scenarios only implies a loss of about 10% relative accuracy. Furthermore, it is possible to determine a subtree containing less than 2% of the original number of scenarios that still carries about 50% relative accuracy.

2. On stability results and probability metrics

Given the original probability measure P of (1) and an approximation Q we derive quantitative estimates of the closeness of $v(Q)$ and $S_\varepsilon(Q)$ to $v(P)$ and $S_\varepsilon(P)$ in terms of a certain probability metric. This distance of probability measures is associated to the model (1) in a natural way. Namely, with the closed unit ball $\mathbb{B} := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ in $\mathbb{R}^n$, we consider the following set of probability measures and distances

$$\begin{aligned}\mathcal{P}_f := \{Q \in \mathcal{P}(\Omega) :\ -\infty < \int_\Omega \inf_{x \in X \cap \rho\mathbb{B}} f(\omega, x) Q(d\omega) \leq \\ \sup_{x \in X \cap \rho\mathbb{B}} \int_\Omega f(\omega, x) Q(d\omega) < \infty \ , \text{ for each } \rho > 0\}\end{aligned}$$

$$d_{f,\rho}(P, Q) := \sup_{x \in X \cap \rho\mathbb{B}} \left| \int_\Omega f(\omega, x) P(d\omega) - \int_\Omega f(\omega, x) Q(d\omega) \right|$$

for each $\rho > 0$ and $P, Q \in \mathcal{P}_f$. Note that, for any $Q \in \mathcal{P}_f$, the function $x \mapsto \mathbb{E}_Q f(\omega, x)$ is lower semicontinuous (by appealing to Fatou's lemma), proper (since $|\mathbb{E}_Q f(\omega, x)| < \infty$ for each $x \in X$) and convex on $\mathbb{R}^n$. Next we give a quantitative stability result for optimal values and (ε-approximate) solution sets.

Theorem 1. *Let $P \in \mathcal{P}_f$ and $S(P)$ be nonempty and bounded. Then there exist constants $\rho > 0$ and $\bar{\varepsilon} > 0$ such that*

$$|v(P) - v(Q)| \leq d_{f,\rho}(P, Q) \quad \textit{and} \quad \emptyset \neq S(Q) \subset S(P) + \Psi(d_{f,\rho}(P, Q))\mathbb{B}$$

whenever $Q \in \mathcal{P}_f$ with $d_{f,\rho}(P, Q) < \bar{\varepsilon}$, and that it holds for any $\varepsilon \in (0, \bar{\varepsilon})$

$$\mathrm{dl}_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{2\rho}{\varepsilon} d_{f,\rho+\varepsilon}(P, Q) \quad \textit{whenever} \quad Q \in \mathcal{P}_f, d_{f,\rho+\varepsilon}(P, Q) < \varepsilon.$$

Here $\Psi(\eta) := \eta + \psi^{-1}(2\eta), \eta \geq 0$, where $\psi(\tau) := \min\{\mathbb{E}_P f(\omega, x) - v(P) : d(x, S(P)) \geq \tau\}, \tau \geq 0$, is the conditioning function of model (1), $\mathrm{dl}_\infty(C, D) := \sup_{x \in \mathbb{R}^n} |d(x, C) - d(x, D)|$ is the Pompeiu-Hausdorff distance of nonempty closed sets $C, D \subseteq \mathbb{R}^n$ and $d(x, C) := \inf_{y \in C} \|x - y\|$ the distance of $x \in \mathbb{R}^n$ to $C \subseteq \mathbb{R}^n$.

Proof. Since the function $\mathbb{E}_P f(\omega,\cdot)$ is lower semicontinuous, proper and convex, we may apply Theorem 7.64 in [18]. Let $\bar{\rho} > 0$ be chosen such that $S(P) \subset \bar{\rho}\mathbb{B}$ and $v(P) \geq -\bar{\rho}$. Let $\rho > \bar{\rho}$ and $\bar{\varepsilon}$ be chosen such that $0 < \bar{\varepsilon} < \min\{\frac{1}{2}(\rho - \bar{\rho}), \frac{1}{2}\psi(\frac{1}{2}(\rho-\bar{\rho}))\}$. Then Theorem 7.64 in [18] says that

$$|v(P) - v(Q)| \leq \widehat{\mathrm{dl}}_\rho^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)) \quad \text{and}$$
$$\emptyset \neq S(Q) \subset S(P) + \Psi(\widehat{\mathrm{dl}}_\rho^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)))\mathbb{B}$$

holds for any $Q \in \mathcal{P}_f$ with $\widehat{\mathrm{dl}}_\rho^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)) < \bar{\varepsilon}$.
Here, $\widehat{\mathrm{dl}}_\rho^+$ denotes the auxiliary epi-distance (cf. Prop. 7.61 in [18])

$$\begin{aligned}\widehat{\mathrm{dl}}_\rho^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)) = \inf\{\eta \geq 0: \ & \text{for all } x \in \rho\mathbb{B} \text{ it holds} \\ & \inf_{y \in x+\eta\mathbb{B}} \mathbb{E}_Q f(\omega,y) \leq \max\{\mathbb{E}_P f(\omega,x), -\rho\} + \eta \text{ and} \\ & \inf_{y \in x+\eta\mathbb{B}} \mathbb{E}_P f(\omega,y) \leq \max\{\mathbb{E}_Q f(\omega,x), -\rho\} + \eta\}.\end{aligned}$$

Hence, the first part of the result is a consequence of the estimate

$$\widehat{\mathrm{dl}}_\rho^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)) \leq d_{f,\rho}(P,Q)$$

(cf. Example 7.62 in [18]). Noting that the function Ψ is increasing, completes the first part of the proof.
For the second part let $\varepsilon \in (0,\bar{\varepsilon})$ and $Q \in \mathcal{P}_f$ be such that $d_{f,\rho+\varepsilon}(P,Q) < \varepsilon$. Then

$$\emptyset \neq S(Q) \subset (\bar{\rho} + \Psi(\bar{\varepsilon}))\mathbb{B} \quad \text{and} \quad v(Q) \geq -(\bar{\rho}+\bar{\varepsilon}).$$

With $\rho > \hat{\rho} = \min\{\bar{\rho} + \Psi(\bar{\varepsilon}), \bar{\rho}+\bar{\varepsilon}\}$ and $\bar{\varepsilon} \leq \rho - \hat{\rho}$ it follows from Theorem 7.69 in [18] that

$$\widehat{\mathrm{dl}}_\rho(S_\varepsilon(P), S_\varepsilon(Q)) \leq \widehat{\mathrm{dl}}_{\rho+\varepsilon}^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot)),$$

where $\widehat{\mathrm{dl}}_\rho$ is the set distance $\widehat{\mathrm{dl}}_\rho(C,D) := \inf\{\eta \geq 0 : C \cap \rho\mathbb{B} \subset D + \eta\mathbb{B}, D\cap\rho\mathbb{B} \subset C + \eta\mathbb{B}\}$ for nonempty subsets C and D of $\mathbb{R}^n$. Using the same argument as above, we may estimate the auxiliary epi-distance $\widehat{\mathrm{dl}}_{\rho+\varepsilon}^+(\mathbb{E}_P f(\omega,\cdot), \mathbb{E}_Q f(\omega,\cdot))$ from above by $d_{f,\rho+\varepsilon}(P,Q)$. Moreover, since the functions $\mathbb{E}_P f(\omega,\cdot)$ and $\mathbb{E}_Q f(\omega,\cdot)$ are lower semicontinuous and convex, their level sets $S_{\bar{\varepsilon}}(P)$ and $S_{\bar{\varepsilon}}(Q)$ are also bounded. Hence, we may choose the constant ρ such that

$$\widehat{\mathrm{dl}}_\rho(S_\varepsilon(P), S_\varepsilon(Q)) = \mathrm{dl}_\infty(S_\varepsilon(P), S_\varepsilon(Q)).$$

This completes the proof. □

Theorem 1 is taken from the paper [20] which also contains more general results (e.g. allowing for unbounded solution sets $S(P)$). Its proof is included for convenience of the reader. The theorem illuminates the role of the distances $d_{f,\rho}(P,Q)$ for some $\rho > 0$, as *minimal information (m.i.) probability metrics* implying stability of optimal values and (approximate) solutions to (1). Here, m.i. means

that the distance $d_{f,\rho}(P,Q)$ processes the minimal information of problem (1) implying stability.
Clearly, the result remains valid when bounding $d_{f,\rho}(P,Q)$ from above by another distance $d(P,Q)$ and reducing $\mathcal{P}_f$ to a subset of $\mathcal{P}(\Omega)$ on which d is well defined. Such a distance d will be called a *canonical* or *ideal probability metric* associated with (1) if it has the form (2) with a class $\mathcal{F}$ of functions from Ω to $\overline{\mathbb{R}}$ that contains the integrands $f(\cdot,x)$ for each $x \in X \cap \rho\mathbb{B}$ and some relevant $\rho > 0$, as well as further functions carrying important analytical properties of $f(\cdot,x)$ without becoming too large. Typical analytical properties defining relevant classes $\mathcal{F}$ in the theory of probability metrics are: Hölder or Lipschitz continuity and m-th order differentiability together with Hölder or Lipschitz continuity of the m-th derivative (see [15]). Hence, the problem arises to explore analytical properties of integrands f in stochastic programming.
Typical integrands $f(\cdot,x)$, $x \in X$, in convex stochastic programming problems are nondifferentiable but locally Lipschitz continuous on Ω. More precisely, they often satisfy the following property: There exists a continuous symmetric function $c : \Omega \times \Omega \to \mathbb{R}_+$ having the property that $c(\omega,\tilde{\omega}) = 0$ holds iff $\omega = \tilde{\omega}$, and a nondecreasing function $g : \mathbb{R}_+ \to \mathbb{R}_+ \setminus \{0\}$ such that for each $x \in X$ and $\omega, \tilde{\omega} \in \Omega$,

$$|f(\omega,x) - f(\tilde{\omega},x)| \leq g(\|x\|)c(\omega,\tilde{\omega}). \tag{7}$$

From now we require that the function c is measurable and satisfies the properties
(C1) $c(\omega,\tilde{\omega}) = 0$ iff $\omega = \tilde{\omega}$;
(C2) $c(\omega,\tilde{\omega}) = c(\tilde{\omega},\omega)\ \forall \omega,\tilde{\omega} \in \Omega$ (symmetry);
(C3) $\sup\{c(\omega,\tilde{\omega}) : \omega,\tilde{\omega} \in B, \|\omega - \tilde{\omega}\| \leq \delta\}$ tends to 0 as $\delta \to 0$ for each bounded subset B of Ω ($\|\cdot\|$ denoting a norm on $\mathbb{R}^s$);
(C4) there exists a measurable function $\lambda : \Omega \to \mathbb{R}_+$ that is bounded on bounded sets and has the property $c(\omega,\tilde{\omega}) \leq \lambda(\omega) + \lambda(\tilde{\omega})\ \forall \omega,\tilde{\omega} \in \Omega$.

If c is a metric on Ω metrizing the norm topology, (C3) and (C4) are satisfied. If Ω is compact, (C4) is satisfied. (C3) is satisfied if c is continuous. An important example of a function c satisfying the conditions (C1)–(C4) is a function of the form (6). It clearly satisfies (C1)–(C3) and also (C4) by considering the function $\lambda_h(\omega) := 2\|\omega - \omega_0\| \max\{1, h(\|\omega - \omega_0\|)\}$. A typical choice for ω_0 is $\omega_0 = \mathbb{E}_P\omega$. If $c : \Omega \times \Omega \to \mathbb{R}_+$ is continuous and satisfies (C1)–(C4), the following duality result is valid for all probability measures $P, Q \in \mathcal{P}_c(\Omega) := \{Q \in \mathcal{P}(\Omega) : \int_\Omega \lambda(\omega)Q(d\omega) < \infty\}$ (see Sect. 5.3 in [15]).

$$\zeta_c(P,Q) = \overset{\circ}{\mu}_c (P,Q) = \inf\{ \int_{\Omega\times\Omega} c(\omega,\tilde{\omega})\eta(d(\omega,\tilde{\omega})) : \eta \text{ finite measure on } \Omega\times\Omega,$$
$$\eta(B\times\Omega) - \eta(\Omega\times B) = P(B) - Q(B) \ \text{ for each } B \in \mathcal{B}\}.$$

$\overset{\circ}{\mu}_c$ is called *Kantorovich-Rubinstein* or *Wasserstein functional* (cf. [15]). It holds $\overset{\circ}{\mu}_c (P,Q) \leq \hat{\mu}_c(P,Q)$ for all $P, Q \in \mathcal{P}_c(\Omega)$ and equality is valid if and only if c is a metric (Theorem 6.1.1 in [15]).
In the special case $h(r) = r^{p-1}$ for $r \in \mathbb{R}_+$ and $\omega_0 = 0$ in (6) we use the notation

c_p, $\zeta_p = \overset{\circ}{\mu}_p$, $\hat{\mu}_p$ and $\mathcal{P}_p(\Omega)$ for the corresponding function of the form (6), the polynomial Fortet-Mourier metric, Kantorovich functional and set of probability measures, respectively. In this case, both functionals $\overset{\circ}{\mu}_p$ and $\hat{\mu}_p$ generate the same topology on $\mathcal{P}_p(\Omega)$. The corresponding convergence is equivalent to weak convergence and convergence of p-th order moments. This fact and further relations between both functionals can be found in Sect. 6.2 of [15]. In particular, for each $1 \leq p < \infty$ the quantitative estimate

$$|\int_{\Omega} \|\omega\|^p (P - Q)(d\omega)| \leq p \overset{\circ}{\mu}_p (P, Q)$$

is valid. $\zeta_1 = \overset{\circ}{\mu}_1 = \hat{\mu}_1$ is also called *Kantorovich* or *L_1-Wasserstein metric*. Condition (7) motivates to consider the class (3) and the metric ζ_c on the set $\mathcal{P}_c(\Omega)$ as a canonical probability metric in convex stochastic programming, since it holds $(g(\rho))^{-1} f(\cdot, x) \in \mathcal{F}_c$ for each $x \in X \cap \rho \mathbb{B}$ and, hence,

$$d_{f,\rho}(P, Q) \leq g(\rho) \zeta_c(P, Q) \leq g(\rho) \hat{\mu}_c(P, Q).$$

Remark 1. (Choice of c for multistage stochastic programs)
It is shown in [16] that linear two-stage stochastic programs with fixed recourse enjoy quantitative stability properties with respect to the Fortet-Mourier metric ζ_2 (i.e., the corresponding integrand f satisfies condition (7) when setting $c = c_2$ where $c_2(\omega, \tilde{\omega}) := \|\omega - \tilde{\omega}\| \max\{1, \|\omega\|, \|\tilde{\omega}\|\} \forall \omega, \tilde{\omega} \in \Omega$). This result is extended in [20] to the case of linear multi-stage stochastic programming models with fixed recourse in all stages under additional assumptions on the underlying discrete-time stochastic process. The corresponding result asserts quantitative stability with respect to the metric ζ_K where K denotes the number of stages of the model. The result also says that such models are even quantitatively stable with respect to ζ_1 if only right-hand sides are random. Hence, polynomial Fortet-Mourier metrics serve as canonical distances for multistage models.

Let us now consider a stochastic program (1), and assume that the integrand f satisfies the condition (7) for some function c having the properties (C1)–(C4) and that the original probability measure belongs to $\mathcal{P}_c(\Omega)$. If the solution set $S(P)$ of (1) is nonempty and bounded, Theorem 1 applies and we may conclude Lipschitz stability properties of the optimal value v and the ε-approximate solution set S_ε at P with respect to the Fortet-Mourier metric $\zeta_c = \overset{\circ}{\mu}_c$.
This motivates to take one of the functionals $\overset{\circ}{\mu}_c$ and $\hat{\mu}_c$ as basis for approximating the original measure P. Let μ_c denote any of the functionals $\overset{\circ}{\mu}_c$ and $\hat{\mu}_c$. For instance, the principle of *optimal scenario generation* or *selection* for (1) may be formulated in the following way: Determine a discrete probability measure Q^* having a prescribed number M of scenarios in Ω such that

$$\mu_c(P, Q^*) = \min\{\mu_c(P, \sum_{j=1}^{M} q_j \delta_{\omega_j}) : \sum_{j=1}^{M} q_j = 1, q_j \geq 0, \omega_j \in \Omega \forall j\}. \tag{8}$$

Further constraints can be incorporated into (8), e.g., implying that the scenarios exhibit a certain prescribed tree structure.
Similarly, the principle of *optimal scenario reduction* of a given discrete approximation $Q = \sum_{j=1}^{M} q_j \delta_{\omega_j}$ to P may be written as: Determine an index set $J_* \subset \{1, \ldots, M\}$ of given cardinality $\#J_* = k$ and weights q_j^* for $j \notin J_*$ which are a solution of

$$\min\{\mu_c(P, \sum_{\substack{j=1 \\ j \notin J}}^{M} q_j \delta_{\omega_j}) : J \subset \{1, \ldots, M\}, \#J = k, \sum_{j \notin J} q_j = 1, q_j \geq 0\}. \tag{9}$$

We note that problem (8) represents a nondifferentiable nonconvex program that is large scale in many practical situations. Its algorithmic solution appears to be hopeless for general measures P, supports Ω, functions c and "large" numbers M of scenarios. An attempt for solving (8) is made in [14] in case of $c = c_1$ and $\Omega = \mathbb{R}^s$, where the scenarios $\{\omega_j : j = 1, \ldots, M\}$ are in addition tree-structured. The author of [14] develops a deterministic iteration scheme in case P is completely known and a stochastic approximation method based on empirical approximations of P. So far numerical experience is available for low dimensional test problems.
When assuming that P is discrete with finite support or replaced by a good discrete approximation, the situation becomes quite different. Then the functional $\mu_c(P, \sum_{\substack{j=1 \\ j \notin J}}^{M} q_j \delta_{\omega_j})$ is the optimal value of a (large scale) linear program with weights q_j and scenarios ω_j entering right-hand sides of linear constraints and the cost function, respectively. When looking at problem (9) and using $\mu_c := \hat{\mu}_c$, the situation becomes quite comfortable, as will be shown in the next section.

3. Scenario reduction

Assume that the original probability distribution P is discrete and carried by finitely many scenarios $\omega_i \in \Omega$ with weights $p_i > 0, i = 1, \ldots, N$, and $\sum_{i=1}^{N} p_i = 1$, i.e., $P = \sum_{i=1}^{N} p_i \delta_{\omega_i}$. Let $J \subset \{1, \ldots, N\}$ and consider the probability measure Q having scenarios ω_j with probabilities q_j, $j \in \{1, \ldots, N\} \setminus J$, i.e., compared to P the measure $Q = \sum_{j \notin J} q_j \delta_{\omega_j}$ is *reduced* by deleting all scenarios $\omega_j, j \in J$ and by assigning new probabilistic weights q_j to each scenario ω_j, $j \notin J$. The optimal reduction concept described in Section 2 (see (9)) advices to consider the functional

$$D(J; q) := \hat{\mu}_c(\sum_{i=1}^{N} p_i \delta_{\omega_i}, \sum_{j \notin J} q_j \delta_{\omega_j}), \tag{10}$$

where the function c is chosen such that the underlying stochastic program behaves stable with respect to the Fortet-Mourier metric ζ_c and, hence, with respect to the Kantorovich functional $\hat{\mu}_c$. We assume throughout this section that c satisfies (C1)–(C4). The reduction concept (9) says that the index set J is selected such that the distance $D(J;q)$ of the original and the reduced measure is optimal subject to all index sets with given cardinality. We distinguish two cases: *optimal* or *prescribed* weights q_j, $j \notin J$. Our first result provides an explicit representation of $D(J;q)$ in case of optimal weights q.

Theorem 2. *(optimal weights)*
Given $J \subset \{1,\dots,N\}$ we have

$$D_J = \min\{D(J;q) : q_j \geq 0, \sum_{j\notin J} q_j = 1\} = \sum_{i\in J} p_i \min_{j\notin J} c(\omega_i,\omega_j). \tag{11}$$

Moreover, the minimum is attained at $\bar{q}_j = p_j + \sum\limits_{i\in J_j} p_i$, for each $j \notin J$, where $J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min\limits_{j\notin J} c(\omega_i,\omega_j)$ for each $i \in J$ (optimal redistribution rule).

Proof. We set $c_{ij} := c(\omega_i,\omega_j)$ for $i,j \in I := \{1,\dots,N\}$ and make use of the primal as well as the dual representation of $D(J;q)$ for given J, i.e.,

$$\begin{aligned} D(J;q) &= \min\{\sum_{i,j} c_{ij}\eta_{ij} : \eta_{ij} \geq 0, \sum_{j\notin J}\eta_{ij} = p_i, i\in I, \sum_{i=1}^{N}\eta_{ij} = q_j, j\notin J\} \\ &= \max\{\sum_{i=1}^{N} p_i u_i + \sum_{j\notin J} q_j v_j : u_i + v_j \leq c_{ij}, i \in I, j\notin J\}. \end{aligned}$$

First, we show that the expression $\sum\limits_{j\in J} p_j \min\limits_{i\notin J} c_{ij}$ provides a lower bound of $D(J;q)$ for any feasible q. We set $u_i := \min\limits_{k\notin J} c_{ik}$ for each $i \in I$ and $v_j := 0$ for each $j \notin J$. We show that this choice of u_i and v_j is feasible for the dual representation of $D(J;q)$. Noting that $u_i = 0$ for any $i \notin J$ we obtain $u_i + v_j = u_i \leq c_{ij}$ for all $i \in J$ and $j \notin J$, and $u_i + v_j = 0 \leq c_{ij}$ for all $i,j \notin J$. Hence, it holds

$$\sum_{i=1}^{N} p_i u_i + \sum_{j\notin J} q_j v_j = \sum_{i\in J} p_i \min_{k\notin J} c_{ik} \leq D(J;q) \quad \text{for any feasible } q.$$

Next, we define elements $\bar{\eta}_{ij} := \begin{cases} p_i, \ i \in J_j, \\ p_i, \ i = j \notin J, \\ 0, \ \text{otherwise}, \end{cases}$ for each $i \in I$ and $j \notin J$,

and set $\bar{q}_j := \sum\limits_{i=1}^{N} \bar{\eta}_{ij} = p_j + \sum\limits_{i\in J_j} p_i$ for each $j \notin J$. We obtain that the $\bar{\eta}_{ij}$ are feasible for the primal representation of $D(J;\bar{q})$. Hence, it holds

$$D(J;\bar{q}) \leq \sum_{i,j} c_{ij}\bar{\eta}_{ij} = \sum_{i\in J} p_i \min_{j\notin J} c_{ij}.$$

We conclude that $\bar{q}$ is optimal and that $D(J;\bar{q}) = \sum\limits_{i\in J} p_i \min\limits_{j\notin J} c_{ij}$. □

The theorem provides an explicit formula for the optimal weights when the index set J of deleted scenarios is given. Its interpretation is that the new probability of a kept scenario is equal to the sum of its former probability and of all probabilities of deleted scenarios that are closest to it with respect to c.
When fixing the redistribution of the deleted weight $p_J = \sum_{i \in J} p_i$ of P by a rule of the form

$$q_j = p_j + \lambda_j p_J \quad \text{for each} \quad j \notin J, \tag{12}$$

where the redistribution weights $\lambda_j \geq 0, j \notin J$, with $\sum_{j \notin J} \lambda_j = 1$, are given, the following upper bound for $D(J;q)$ is valid.

Theorem 3. *(prescribed redistribution)*
When the index set $J \subset \{1, \ldots, N\}$ is given and q is redistributed by (12), we have

$$D(J;q) \leq \sum_{i \in J} p_i \sum_{j \notin J} \lambda_j c(\omega_i, \omega_j).$$

Moreover, equality holds if $\#J = 1$ and c satisfies the triangle inequality.

Proof. We use the primal representation of $D(J;q)$ (see the preceding proof) and set $\eta_{ij} := \begin{cases} p_i \lambda_j, & i \in J, j \notin J, \\ p_i, & i = j \notin J, \\ 0, & \text{otherwise.} \end{cases}$

Then we have $\sum\limits_{j \notin J} \eta_{ij} = p_i$ for $i = 1, \ldots, N$, and $\sum\limits_{i=1}^{N} \eta_{ij} = p_j + \lambda_j p_J = q_j$ for $j \notin J$. Hence, we obtain with $c_{ij} := c(\omega_i, \omega_j)$

$$D(J;q) \leq \sum_{i,j} c_{ij} \eta_{ij} = \sum_{i \in J} p_i \sum_{j \notin J} \lambda_j c_{ij}.$$

Finally, let $J = \{l\}$ and assume that c satisfies the triangle inequality. We set $u_i = -c_{il} = -v_i, i = 1, \ldots, N$, and note that $u_i + v_j = c_{jl} - c_{il} \leq c_{ij}$ holds for all $i, j \in \{1, \ldots, N\}, j \neq l$. Hence, we obtain from the dual representation of $D(J;q)$ that

$$D(J;q) \geq \sum_{i=1}^{N} (q_i - p_i) c_{il} = p_l \sum_{\substack{i=1 \\ i \neq l}}^{N} \lambda_i c_{il}.$$

□

Simple examples show that equality is lost in Theorem 3 in general if $\#J \geq 2$ and also if $\#J = 1$ and c does not satisfy the triangle inequality. We stress here that the latter property of c is not needed for the optimal redistribution in Theorem 2.
Next we discuss the optimal choice of an index set J for scenario reduction with optimal weights and fixed cardinality $\#J$, i.e., the solution of the problem

$$\min\{D_J = \sum_{i \in J} p_i \min_{j \notin J} c(\omega_i, \omega_j) : J \subset \{1, ..., N\}, \#J = k\} \tag{13}$$

for given $k \in \mathbb{N}, 1 \leq k < N$. First we consider the extremal cases of problem (13): deleting a single scenario and all but one scenarios.

Example 1. (single scenario deletion)
In case $\#J = 1$ the optimal deletion problem (13) takes the form

$$\min_{l \in \{1,...,N\}} p_l \min_{j \neq l} c(\omega_l, \omega_j). \tag{14}$$

If the minimum is attained at $l_* \in \{1,...,N\}$, i.e., the scenario ω_{l_*} is deleted, the optimal redistribution rule is $\bar{q}_l = p_l$ for each $l \notin \{l_*, j(l_*)\}$ and $\bar{q}_{j(l_*)} = p_{j(l_*)} + p_{l_*}$, where $j(l_*) \in \arg\min_{j \neq l_*} c(\omega_{l_*}, \omega_j)$. Of course, the optimal deletion of a single scenario may be repeated recursively until a prescribed number k of scenarios is deleted (as in Algorithm 1).

Example 2. (keeping only one scenario)
In case $\#J = N - 1$ the problem (13) has the form

$$\min_{u \in \{1,...,N\}} \sum_{i=1}^{N} p_i c(\omega_i, \omega_u). \tag{15}$$

If the minimum is attained at $u^* \in \{1,...,N\}$, the scenario ω_{u^*} is kept and the optimal redistribution rule provides $\bar{q}_{u^*} = p_{u^*} + \sum_{i \neq u^*} p_i = 1$.

Since (13) represents a special combinatorial optimization problem, efficient solution algorithms like for the cases $k = 1$ and $k = N - 1$ are hardly available in general. However, the next result provides lower and upper bounds for the optimal value of (13) that correspond to recursive extensions of the strategies (14) for $\#J = 1$ and (15) for $\#J = N - 1$. Moreover, as shown in [12], the lower bound is attained under additional assumptions.

Theorem 4. *(bounds)*

$$\sum_{i=1}^{k} p_{l_i} \min_{j \neq l_i} c(\omega_{l_i}, \omega_j) \leq \min\{D_J : J \subset \{1,...,N\}, \#J = k\} \leq \sum_{i \in J_u} p_i \min_{j \notin J_u} c(\omega_i, \omega_j)$$

where $J_u := \{1, \ldots, N\} \setminus \{u_1, \ldots, u_{N-k}\}$ *and the indices* l_i *and* u_j *are chosen recursively such that they are solutions of the minimization problems*

$$\min_{l \in \{1,...,N\} \setminus \{l_1,...,l_{i-1}\}} p_l \min_{j \neq l} c(\omega_l, \omega_j), i = 1, \ldots, k, \quad \textit{and} \tag{16}$$

$$\min_{u \notin \{u_1,...,u_{j-1}\}} \sum_{\substack{i=1 \\ i \notin \{u_1,...,u_{j-1},u\}}}^{N} p_i \min_{l \in \{u_1,...,u_{j-1},u\}} c(\omega_l, \omega_i), j = 1, \ldots, N - k, \tag{17}$$

respectively. Moreover, the index set $\{l_1, \ldots, l_k\}$ *is a solution of (13) if for each* $i = 1, \ldots, k$ *the set* $\arg\min_{j \neq l_i} c(\omega_{l_i}, \omega_j) \setminus \{l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_k\}$ *is nonempty.*

Proof. For any index set $J = \{j_1, \dots, j_k\} \subset \{1, \dots, N\}$ with $\#J = k$ we have from Theorem 2 that

$$D_J = \sum_{i=1}^{k} p_{j_i} \min_{j \notin \{j_1,\dots,j_k\}} c(\omega_{j_i}, \omega_j) \geq \sum_{i=1}^{k} p_{j_i} \min_{j \neq j_i} c(\omega_{j_i}, \omega_j)$$
$$\geq \sum_{i=1}^{k} p_{l_i} \min_{j \neq l_i} c(\omega_{l_i}, \omega_j)$$

where the last estimate is a consequence of the definition of the numbers $l_i, i = 1, \dots, k$. If $\arg\min_{j \neq l_i} c(\omega_{l_i}, \omega_j) \setminus \{l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_k\} \neq \emptyset$ holds for each $i = 1, \dots, k$, we obtain

$$\min_{j \neq l_i} c(\omega_{l_i}, \omega_j) = \min_{j \notin \{l_1,\dots,l_k\}} c(\omega_{l_i}, \omega_j)$$

for each $i = 1, \dots, k$. Hence, the above estimate may be continued to

$$D_J \geq \sum_{i=1}^{k} p_{l_i} \min_{j \notin \{l_1,\dots,l_k\}} c(\omega_{l_i}, \omega_j) = D_{\{l_1,\dots,l_k\}}.$$

Hence, the index set $\{l_1, \dots, l_k\}$ is a solution of (13). Finally, the obvious estimate

$$\min\{D_J : J \subset \{1, ..., N\}, \#J = k\} \leq D_{J_u} = \sum_{i \in J_u} p_i \min_{i \notin J_u} c(\omega_i, \omega_j)$$

completes the proof. □

Theorem 4 suggests two different *heuristic algorithms* for the optimal reduction of the N original scenarios to $N - k$ scenarios. The first one determines the k scenarios that have to be deleted by exploiting the lower bound technique (16). It will be called *backward reduction*, while the second one is called *forward selection*, because it selects the remaining $N - k$ scenarios by using the upper bound technique (17).

Algorithm 1. (backward reduction)
In the first step, an index k_1 with $1 \leq k_1 \leq k$ is determined such that $J_1 = \{l_1^{(1)}, \dots, l_{k_1}^{(1)}\}$ is a solution of (13) for $k = k_1$ by using the strategy described in Theorem 4. Next, the optimal redistribution rule of Theorem 2 is used and the probability measure P_1 is considered where the scenarios indexed by J_1 are deleted. If $k_1 < k$, then the measure P_1 is reduced in a second step by deleting all scenarios in some index set J_2 with $1 \leq k_2 = \#J_2 \leq k - k_1$ that is obtained in the same way using Theorem 4. This procedure is continued until in step r we have $\sum_{i=1}^{r} k_i = k$ and $J = \cup_{i=1}^{r} J_i$. Finally, the optimal redistribution rule is used again for the index set J.

A particular variant of Algorithm 1 consists in the case $k_i = 1$ for $i = 1, \dots, k$ (*backward reduction of single scenarios*). This variant (without the final optimal redistribution) is already announced in [8,11].

Algorithm 2. (forward selection)
The indices $u_j, j = 1, \dots, N-k$, are determined recursively as in (17). Finally, set $J := \{1, \dots, N\} \setminus \{u_1, \dots, u_{N-k}\}$ and redistribute for the index set J according to Theorem 2.

Both algorithms provide an approximate solution of problem (13) and hence lower and upper bounds for its optimal value. The performance of both algorithms is evaluated for a real-life test example in Section 4.
According to Theorem 3 the optimal choice of an index set J for scenario deletion with prescribed redistribution and fixed cardinality $\#J$ may be formulated as follows:

$$\min\{\sum_{i \in J} p_i \sum_{j \notin J} \lambda_j c(\omega_i, \omega_j) : J \subset \{1, ..., N\}, \#J = k\} \tag{18}$$

for given $k \in \mathbb{N}, 1 \leq k < N$ and weights $\lambda_j, j \notin J$. Let us consider the particular case $k = 1$.

Example 3. (deleting a single scenario with prescribed redistribution)
In case $\#J = 1$ problem (18) takes the form

$$\min_{i \in \{1, \dots, N\}} p_i \sum_{j \neq i} \lambda_j c(\omega_i, \omega_j).$$

When P is a uniform discrete distribution, i.e., $p_i = \frac{1}{N}$ for each i, it might be desirable that the reduced measure Q has uniform weights as well, i.e., $q_j = \frac{1}{N-1}$ for $j \notin J$. This corresponds to the choice $\lambda_j = \frac{1}{N-1}$ for $j \notin J$.

Although prescribing the weights for the reduced distribution might sometimes be useful, our theoretical results tend to preferring the optimal redistribution rule. For the latter rule no additional assumptions on c are needed, setwise reduction algorithms are available and Kantorovich functionals can be evaluated without solving transportation problems.

4. Numerical results for electrical load scenario trees

The optimization of electric power production in hydro-thermal generation systems for time horizons of one week or longer is inherently a stochastic decision problem. Indeed, forecasts on electrical load, flows to hydro reservoirs, and on fuel and electricity prices cannot be exact. For this reason, the weekly cost-optimal production of electric power in a hydro-thermal generation system is modeled in [11,13] as a multistage (mixed-integer) stochastic program. The optimal scheduling decisions for all power units minimize the expected production costs subject to the operational requirements of the generation system. The scheduling decisions for thermal units are: which units to commit in each period, and at what generating capacity. The decision variables for pumped storage hydro units are the generation and pumping levels for each period. Power contracts for delivery and purchase are regarded as special thermal units. The basic

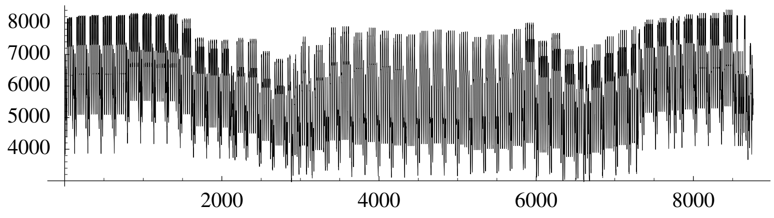

Fig. 1. Time plot of the load profile for one year

system requirement is to meet the electrical load and the spinning reserve capacity. Further operating constraints are capacity limits for thermal and hydro units, minimum up/down-time requirements for thermal units and operating ranges and dynamics of hydro storage units. The scheduling horizon of one week is discretized into hourly intervals. Accordingly, the stochastic data process is approximated by a discrete-time stochastic process. The numerical tests of the stochastic model are performed in [11,13] for a real-life hydro-thermal generation system consisting of 25 (coal-fired or gas-burning) thermal units and 7 pumped hydro units, and for stochastic electrical load (i.e., the remaining data were deterministic). The stochastic load process is approximated by a scenario tree. The resulting problem to be solved is a large-scale mixed-integer linear program with a special sparsity structure. The mixed-integer model is large even for a relatively small number of nodes in the tree. The single scenario model (i.e., 168 nodes) already contains 4200 binary and 6652 continous variables, 13441 constraints and 19657 nonzeros in the constraint matrix.
In [11], an initial load scenario tree was constructed according to the following steps:

1. Calibration of a time series model for the load, generation of a large number of load scenarios.
2. Construction of a (dense) initial load scenario tree using the sample means and standard deviations of the simulated load scenarios.

The time series model for the load d_t in period t was calibrated from a historical load profile of one year (cf. Figure 1). The time series model for the load process $\{d_t\}_{t\in\mathbb{Z}}$ is the SARIMA$(7,0,9)\times(0,1,0)_{168}$ model (cf. [11])

$$\begin{aligned} d_t = \hat{\phi}_1 d_{t-1} + \ldots + \hat{\phi}_7 d_{t-7} + d_{t-168} - \hat{\phi}_1 d_{t-169} - \ldots - \hat{\phi}_7 d_{t-175} \\ + Z_t + \hat{\theta}_1 Z_{t-1} + \ldots + \hat{\theta}_9 Z_{t-9}, \quad t \in \mathbb{Z}. \end{aligned} \tag{19}$$

The estimated model coefficients are

$$\begin{aligned} (\hat{\phi}_1,\ldots,\hat{\phi}_7) &= (2.79, -4.35, 5.16, -4.88, 3.67, -1.92, 0.50), \\ (\hat{\theta}_1,\ldots,\hat{\theta}_9) &= (-1.27, 1.53, -1.35, 0.88, -0.31, -0.06, 0.18, 0.11, 0.07) \end{aligned}$$

and Z_t, $t\in\mathbb{Z}$, are independent, normally distributed random variables with mean 0 and standard deviation 108.3.

For the generation system under consideration, the load forecast is reliable for the time span $t = 1, \ldots, 24$. A large number M of *simulated load scenarios* (sample paths) $\tilde{d}^\ell = (\tilde{d}^\ell_t)_{t=25}^{168}$, $\ell = 1, \ldots, M$, is generated from the SARIMA equation (19) using M i.i.d. realizations of Z_t, $t = 16, \ldots, 168$, and starting values d_t, $t = -150, \ldots, 24$. The *empirical means* $\bar{d}_t$ and *standard deviations* $\bar{\sigma}_t$ of the simulated load scenarios form the basis of the scenario tree building scheme. Since there exists a fixed daily time when already observable meteorological and load data provide the opportunity to re-adjust forecasts, $t_k = 24k$, $k = 1, \ldots, 6$ is a reasonable choice for the branching points. A balanced *ternary tree* with branching points t_k allows to distinguish the events with the verbal description "low load", "medium load" and "high load" in the time period $t = t_k + 1, \ldots, t_{k+1}$, $k = 1, \ldots, 6$. (For convenience of notation set $t_7 := 168$.) Thus, the tree consists of $N := 3^6 = 729$ scenarios $d^i = (d^i_t)_{t=1}^{168}$, $i = 1, \ldots, N$.
The scenarios of the initial load scenario tree and their probabilities can be assigned in various ways. The predicted load for the (first stage) planning period $t = 1, \ldots, 24$ yields the first 24 components for all scenarios. To each scenario i, $i = 1, \ldots, N$ we assign a vector $\omega^i = (\omega^i_k)_{k=2}^7$ with $\omega^i_k \in \{-1, 0, 1\}$ for $k = 2, \ldots, 7$. It provides a unique description of the path in the ternary tree that corresponds to scenario i. In particular, set $\omega^i_k := -1$ if the values of scenario i for $t = t_k + 1, \ldots, t_{k+1}$ are realizations of the event with the verbal description "low load" for this time span. Accordingly, we set $\omega^i_k := 0$ ($\omega^i_k := 1$) to describe the event "medium load" ("high load") for $t = t_k + 1, \ldots, t_{k+1}$. Then the value of scenario i at t is defined as

$$d^i_t := \bar{d}_t + \sum_{j=1}^{k-1} \omega^i_j \frac{\sqrt{3}\bar{\sigma}_{t_{j+1}}}{2^{(8-j)/2}} + \omega^i_k \frac{\sqrt{3}\bar{\sigma}_{t_{k+1}}}{2^{(8-k)/2}} \frac{t - t_k}{t_{k+1} - t_k} \tag{20}$$

for $t = t_k + 1, \ldots, t_{k+1}$, $k = 1, \ldots, 6$. We let all scenarios have equal probabilities $\frac{1}{N} = 3^{-6}$. (Alternative scenario probabilities might be computed from histograms of the simulated scenarios.) Figures 2 and 3 show the ternary load scenario tree $\{d^i_t\}_{t=1}^{168}$ and the corresponding mean shifted tree $\{d^i_t - \bar{d}_t\}_{t=1}^{168}$, respectively, where d^i_t is generated by the scheme (20), with $N = 3^6 = 729$ scenarios and branching points $t_k = 24k$, $k = 1, \ldots, 6$. This tree is used as (dense) initial scenario tree in order to test the performance of the reduction algorithms of Section 3. The test runs were performed on an HP 9000 (780/J280) Compute-Server with 180 MHz frequency and 768 MByte main memory under HP-UX 10.20. We compared the three algorithms *backward reduction of scenario sets* (Algorithm 1), *backward reduction of single scenarios* (variant of Algorithm 1) and *forward selection of scenarios* (Algorithm 2), where all of them were implemented in C. For computing the distances of the initial probability measure $P = \frac{1}{N}\sum_{i=1}^N \delta_{d^i}$ (i.e. given by the initial tree with identical scenario probabilities) and the reduced probability measure Q (reduced tree with optimal redistribution) we used the Kantorovich metric $\zeta_1 = \hat{\mu}_1$ and the explicit formula of Theorem 2. This choice of the metric is justified by Remark 1 when considering the stochastic power management model with fixed binary decisions and with electrical load appearing as stochastic right-hand side. Table 1 displays computing times (in

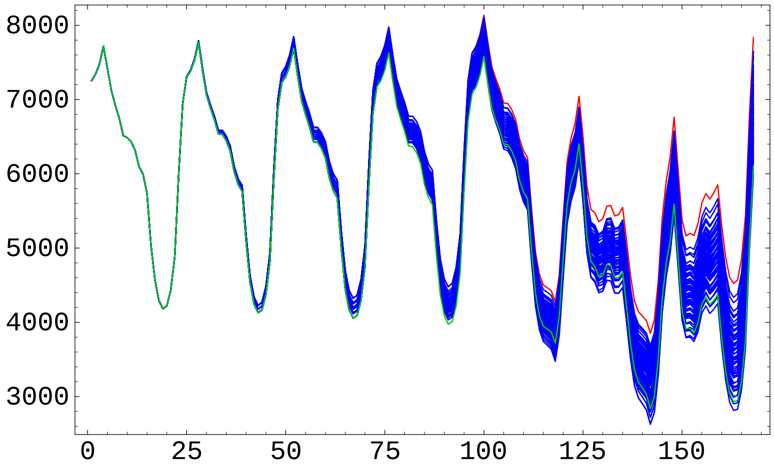

Fig. 2. Ternary load scenario tree for one week

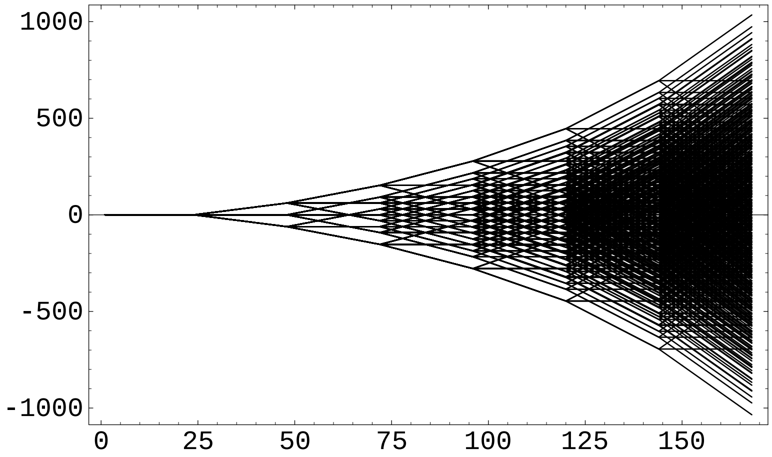

Fig. 3. Mean shifted ternary load scenario tree for one week

seconds) and (absolute and relative) ζ_1-distances of initial and reduced scenario trees for different numbers of scenarios. Here, the relative ζ_1-distance of P and Q is defined as the quotient

$$\frac{\zeta_1(P,Q)}{\zeta_1(P,\delta_{d^{i_*}})}$$

where $i_* \in \{1,\dots,N\}$ denotes the first index obtained by forward selection (see also Example 2) and $\delta_{d^{i_*}}$ is the corresponding Dirac measure. Hence, the distance $\zeta_1(P,\delta_{d^{i_*}})$ corresponds to the best possible ζ_1-distance of P to one of its scenarios endowed with unit mass. Thus, the relative ζ_1-distance reflects the quality (in percent) of the reduced scenario tree relative to the best possible

Number of scenarios	Backward reduction of scenario sets			Backward reduction of single scenarios			Forward selection of scenarios		
	time [s]	ζ_1-distance abs.	rel.[%]	time [s]	ζ_1-distance abs.	rel.[%]	time [s]	ζ_1-distance abs.	rel.[%]
600	8	66.63	3.37	8	66.63	3.37	8149	66.64	3.36
500	8	118.34	5.99	8	118.34	5.99	7637	118.32	5.99
400	8	176.24	8.92	8	176.25	8.92	6146	170.48	8.63
300	8	260.66	13.19	8	262.17	13.3	4280	235.76	11.93
200	8	348.77	17.65	8	357.19	18.08	2319	331.34	16.76
100	8	502.90	25.45	8	505.70	25.59	721	483.94	24.49
81	8	546.74	27.67	8	559.89	28.33	498	530.04	26.84
50	8	645.07	32.64	8	664.34	33.62	212	628.31	31.80
27	8	759.88	38.45	8	782.79	39.61	73	749.09	37.91
10	8	989.12	50.05	8	1019.73	51.60	19	951.02	48.13
9	8	1019.90	51.61	8	1049.02	53.09	17	970.23	49.10
8	8	1045.78	52.92	8	1071.23	54.21	15	1010.92	51.16
7	8	1073.14	54.31	8	1122.04	56.78	13	1051.64	53.22
6	8	1107.82	56.06	8	1147.86	58.09	12	1097.48	55.54
5	8	1153.44	58.37	8	1189.47	60.19	11	1143.42	57.86
4	8	1218.29	61.65	8	1290.15	65.29	10	1201.11	60.78
3	8	1303.74	65.98	8	1360.97	68.87	9	1259.25	63.73
2	8	1506.35	76.23	8	1666.22	84.20	9	1618.16	81.89
1	8	1976.07	100.0	8	2027.32	102.59	8	1976.07	100.0

Table 1. Results of load scenario tree reduction

deterministic approximation to P. The test runs showed that the "distances" $c_1(d^i, d^j) = \|d^i - d^j\|$ ($\|\cdot\|$ denoting the Euclidean norm) for each pair (d^i, d^j) of scenarios of the initial load tree are computed within 6s CPU time. These "distances" are needed in all algorithms.

Table 1 shows that all algorithms work reasonably well, and that backward reduction algorithms are (much) faster than forward selection. Furthermore, besides the very simple algorithmic structure, backward reduction of single scenarios is also competitive to backward reduction of scenario sets with respect to computing times and accuracy, although backward reductions of scenario sets always yield better results. The percentage of 102.59% appearing in the last row of Table 1 is due to the fact that the best possible scenario d^{i_*} has already been deleted in an earlier iteration of the recursive algorithm. For reduced trees with small scenario numbers ≤ 50 forward selection performs fast and produces better trees than backward reduction. Table 1 also reflects the surprising effect that a reduction of the scenario tree by 50% of the original scenarios only implies a loss of about 10% relative accuracy. Furthermore, it is possible to determine a subtree containing less than 2% of the original number of scenarios that still carries about 50% relative accuracy. Figures 4, 5 and 6 show (mean) shifted reduced load trees with 10 and 50 scenarios that are obtained by all three algorithms. The figures display the scenarios within the extremal paths of the initial scenario tree indicated by dashed lines and with grey levels proportional to scenario probabilities.

The different performance of the backward reduction and forward selection with respect to running times is due to the following algorithmic detail. The inner

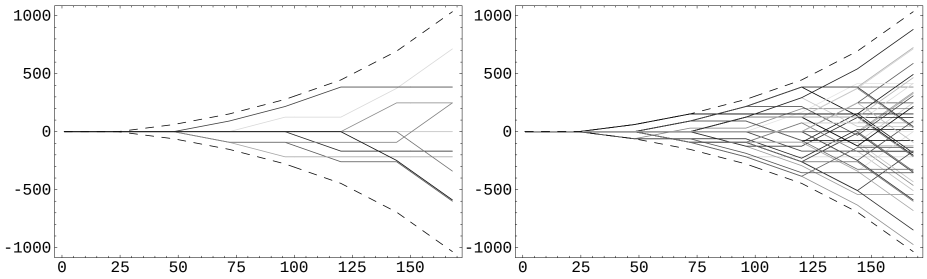

Fig. 4. Backward deletion of scenario sets: Reduced trees with 10 and 50 scenarios

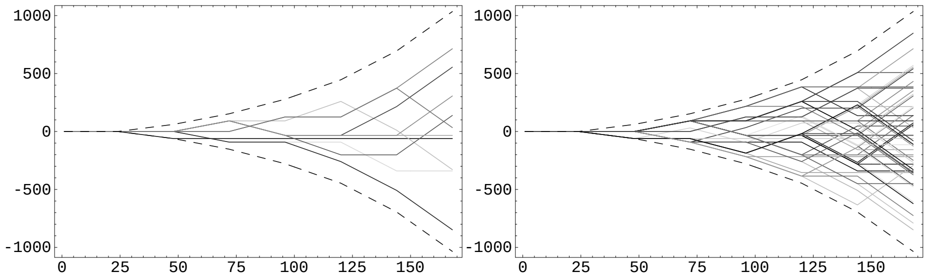

Fig. 5. Backward deletion of single scenarios: Reduced trees with 10 and 50 scenarios

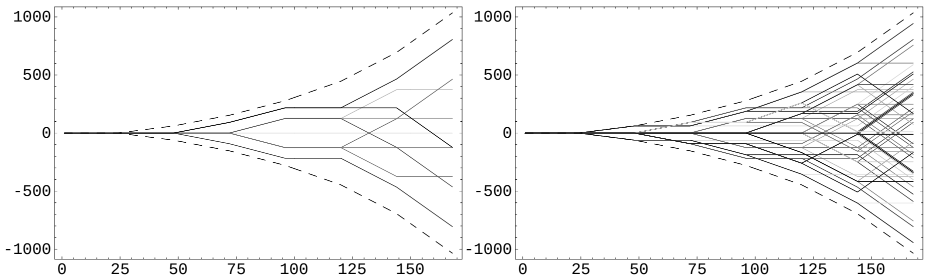

Fig. 6. Forward selection of scenarios: Reduced trees with 10 and 50 scenarios

minima and optimal indices in (14) and (16) are efficiently evaluated by the preprocessing step that sorts the "distances" $c(\cdot,\cdot)$ from each scenario to the remaining scenarios and stores the corresponding permutation vectors of scenario indices (requiring 1s CPU-time). After selecting the next scenario which enters the index set J the optimal indices of the inner minimization are updated by shifting pointers along the permutation vectors. Then the outer minimization for selecting the next scenario for J according to (14) and (16) can be easily performed. This explains the identical running times for both backward reduction algorithm for different reduced trees. The algorithm for fast evaluating the inner minima was adapted for the forward selection methods. However, the computing times indicate that the adaptation did not work as well since updating the pointers to the permutation vectors is more costly.

Acknowledgements. The first author was partly supported by the research project *Mathematical Methods in Stochastics* - MSM 113200008 and by the Grant Agency of the Czech Republic under grants No. 201/99/0264 and 402/99/1136, while the research of the other authors was supported by the Schwerpunktprogramm *Echtzeit-Optimierung grosser Systeme* of the Deutsche Forschungsgemeinschaft. We wish to thank these institutions as well as for the support from the collaboration agreement between the Charles University Prague and the Humboldt-University Berlin. We extend our gratitude to two anonymous referees for valuable suggestions, to Holger Heitsch (Humboldt-University Berlin) and Simon Siegrist (University of Zurich) for pointing out errors in an earlier version of this paper, to René Henrion (WIAS Berlin) for valuable discussions and to Isabel Wegner (Humboldt-University Berlin) for her assistance in providing the scenario trees and pictures of Section 4.

References

1. Attouch, H., Wets, R.J-B (1993): Quantitative stability of variational systems: III. ε-approximate solutions. Mathematical Programming **61**, 197–214
2. Beltratti, A., Consiglio, A., Zenios, S.A. (1999): Scenario modeling for the management of international bond portfolios. Annals of Operations Research **85**, 227–247
3. Bertocchi, M., Dupačová, J., Moriggia, V. (2000): Sensitivity of bond portfolio's behavior with respect to random movements in yield curve. Annals of Operations Research **99**, 267–286
4. Cariño, D.R., Myers, D.H., Ziemba, W.T. (1998): Concepts, technical issues, and uses of the Russel-Yasuda Kasai financial planning model. Operations Research **46**, 450–462
5. Dempster, M.A.H., Thompson, R.T. (1999): EVPI-based importance sampling solution procedure for multistage stochastic linear programs on parallel MIMD architectures. Annals of Operations Research **90**, 161–184
6. Dupačová, J. (1990): Stability and sensitivity analysis for stochastic programming. Annals of Operations Research **27**, 21–38
7. Dupačová, J. (1995): Scenario-based stochastic programs: Strategies for deleting scenarios. WP–95–014, IIASA Laxenburg
8. Dupačová, J. (1999): Portfolio optimization via stochastic programming: Methods of output analysis. Mathematical Methods of Operations Research **50**, 245–270
9. Dupačová, J., Consigli, G., Wallace, S.W. (2000): Scenarios for multistage stochastic programs. Annals of Operations Research **100**, 25–53
10. Dupačová, J., Römisch, W. (1998): Quantitative stability for scenario-based stochastic programs. In: *Prague Stochastics '98* (M. Hušková et al. eds.), JČMF Prague, 119–124
11. Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I. (2001): Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In: *Decision Making under Uncertainty: Energy and Power* (C. Greengard, A. Ruszczynski eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag New York, 39–70
12. Heitsch, H. (2001): Reduktion von Szenariobäumen in der stochastischen Optimierung. Masters thesis, Humboldt-University Berlin, Institute of Mathematics
13. Nowak, M.P., Römisch, W. (2000): Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. Annals of Operations Research **100**, 251–272
14. Pflug, G.Ch. (2001): Scenario tree generation for multiperiod financial optimization by optimal discretization. Mathematical Programming, Series B **89**, 251–271
15. Rachev, S.T. (1991): *Probability Metrics and the Stability of Stochastic Models.* Wiley
16. Rachev, S.T., Römisch, W. (2000): Quantitative stability in stochastic programming: The method of probability metrics. Preprint Nr. 00-22, Humboldt-University Berlin, Institute of Mathematics, 2000. Submitted to Mathematics of Operations Research
17. Rachev, S.T., Rüschendorf, L. (1998): *Mass Transportation Problems*, Vol. I and II. Springer-Verlag Berlin
18. Rockafellar, R.T., Wets, R.J-B (1997): *Variational Analysis.* Springer-Verlag Berlin

19. Römisch, W., Schultz, R. (1991): Stability analysis for stochastic programs. Annals of Operations Research **30**, 241–266
20. Römisch, W., Wets, R.J-B (2001): Stability of ε-approximate solutions to multistage stochastic programs. Manuscript, Humboldt-University Berlin, Institute of Mathematics
21. Schultz, R. (2000): Some aspects of stability in stochastic programming. Annals of Operations Research **100**, 55–84
22. Shapiro, A. (1994): Quantitative stability in stochastic programming. Mathematical Programming **67**, 99–108
23. Zenios, S.A., Shtilman, M.S. (1993): Constructing optimal samples from a binomial lattice. Journal of Information & Optimization Sciences **14**, 125–147
24. Zolotarev, V.M. (1983): Probability metrics. Theory of Probability and its Applications **28**, 278–302