Mathematical Programming (2022) 191:183–205
https://doi.org/10.1007/s10107-018-1337-6

FULL LENGTH PAPER

Series B

Problem-based optimal scenario generation and reduction in stochastic programming

R. Henrion[1] · **W. Römisch**[2]

Received: 17 February 2017 / Accepted: 25 September 2018 / Published online: 4 October 2018

Abstract
Scenarios are indispensable ingredients for the numerical solution of stochastic programs. Earlier approaches to optimal scenario generation and reduction are based on stability arguments involving distances of probability measures. In this paper we review those ideas and suggest to make use of stability estimates based only on problem specific data. For linear two-stage stochastic programs we show that the problem-based approach to optimal scenario generation can be reformulated as best approximation problem for the expected recourse function which in turn can be rewritten as a generalized semi-infinite program. We show that the latter is convex if either right-hand sides or costs are random and can be transformed into a semi-infinite program in a number of cases. We also consider problem-based optimal scenario reduction for two-stage models and optimal scenario generation for chance constrained programs. Finally, we discuss problem-based scenario generation for the classical newsvendor problem.

1 Introduction

Most numerical solution approaches in stochastic programming require the replacement of the underlying multivariate probability distribution by a discrete probability measure with a finite number of realizations or *scenarios*. The most used approach so far is Monte Carlo sampling (see, for example, [47, Chapter 6]). Another more classical approach for two-stage models uses discrete probability measures leading to lower and upper bounds for the expected recourse function. They are obtained by

Dedicated to the memory of Jitka Dupačová

✉ W. Römisch
romisch@math.hu-berlin.de

R. Henrion
henrion@wias-berlin.de

1 Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

2 Institute of Mathematics, Humboldt-University Berlin, Berlin, Germany

means of moment problems (see [26, Section 4.7.2]). More recently optimal quantization techniques (see [16,35]) and (randomized) Quasi-Monte Carlo methods (see [3,29,34]) are employed for solving two-stage stochastic programs. For a survey on scenario generation in stochastic programming see [44].

Jitka Dupačová was one of the pioneers for scenario generation and reduction. We recall her earlier paper [8] and the influential work [9,10].

Here, we study a problem-based approach to scenario generation and reduction for stochastic programming models without information constraints. A general form of such models is [26,47,49]

$$\min \left\{ \int_{\Xi} f_0(x,\xi) P(d\xi) : x \in X, \int_{\Xi} f_1(x,\xi) P(d\xi) \leq 0 \right\} \tag{1}$$

where X is a closed subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^s$, P is a Borel probability measure on Ξ abbreviated by $P \in \mathcal{P}(\Xi)$. The functions f_0 and f_1 from $\mathbb{R}^m \times \Xi$ to the extended reals $\overline{\mathbb{R}} = [-\infty, \infty]$ are normal integrands (in the sense of [42, Chapter 14]). For example, typical integrands f_0 in linear two-stage stochastic programming models are of the form [55], [47, Chapt. 2]

$$f_0(x,\xi) = \begin{cases} g(x) + \Phi(q(\xi), h(x,\xi)), & q(\xi) \in \mathcal{D} \\ +\infty, & \text{else} \end{cases} \quad \text{and} \quad f_1(x,\xi) \equiv 0, \tag{2}$$

where X and Ξ are convex polyhedral, $g(\cdot)$ is a linear function, Φ denotes the infimal function of the linear (second-stage) optimization problem

$$\Phi(q,t) := \inf\{\langle q, y\rangle : Wy = t, y \in Y\} \tag{3}$$

with a $(r, \bar{m})$ matrix W, a convex polyhedral cone $Y \subset \mathbb{R}^{\bar{m}}$ with $Y^\star$ denoting its polar cone, $q(\cdot)$ is an affine function, $h(\cdot,\xi)$ is affine for fixed ξ and $h(x,\cdot)$ is affine for fixed x, and $\mathcal{D} = \{u \in \mathbb{R}^{\bar{m}} : \{z \in \mathbb{R}^r : W^\top z - u \in Y^\star\} \neq \emptyset\}$ denotes the convex polyhedral dual feasibility set. Other examples of practical interest are infimal functions of linear-quadratic or second-order cone programming problems. Typical integrands f_1 appearing in chance constrained programming are of the form $f_1(x,\xi) = p - \mathbb{1}_{\mathcal{P}(x)}(\xi)$, where $\mathbb{1}_{\mathcal{P}(x)}$ is the characteristic function of the polyhedron $\mathcal{P}(x) = \{\xi \in \Xi : h(x,\xi) \leq 0\}$ depending on x.

Let $v(P)$ and $S(P)$ denote the infimum and solution set of (1). The notation indicates that their dependence on the underlying probability distribution is of particular interest. For general continuous multivariate probability distributions P such stochastic optimization models are not solvable in general. Even the computation of the involved integrals requires multivariate numerical integration methods. Many approaches for solving optimization models (1) numerically are based on discrete approximations of the probability measure P, i.e., on finding a discrete probability measure P_n in

$$\mathcal{P}_n(\Xi) := \left\{ \sum_{i=1}^{n} w_i \delta_{\xi^i} : \xi^i \in \Xi,\ i = 1, \ldots, n,\ (w_1, \ldots, w_n) \in \mathcal{S}_n \right\}$$

for some $n \in \mathbb{N}$, which approximates P in a suitable way. Here, $\mathcal{S}_n$ denotes the standard simplex $\mathcal{S}_n = \{w \in \mathbb{R}^n_+ : \sum_{i=1}^n w_i = 1\}$ and ξ^i, $i = 1, \ldots, n$, the scenarios. Of course, the notion 'suitable' should at least mean that the distance between the infima

$$|v(P) - v(P_n)| \tag{4}$$

becomes reasonably small. This is a consequence of stability results for stochastic programming problems which explore the behavior of infima and solution sets if the probability distribution is perturbed. To state a version of such results we introduce the following sets of functions and of probability distributions (both defined on Ξ)

$$\begin{aligned} \mathcal{F} &= \left\{ f_j(x, \cdot) : j = 0, 1,\ x \in X \right\}, \\ \mathcal{P}_{\mathcal{F}} &= \left\{ Q \in \mathcal{P}(\Xi) : -\infty < \int_{\Xi} \inf_{x \in X} f_j(x, \xi)\, Q(d\xi), \right. \\ &\qquad \left. \sup_{x \in X} \int_{\Xi} f_j(x, \xi)\, Q(d\xi) < +\infty,\ j = 0, 1 \right\} \end{aligned} \tag{5}$$

and the following (semi-) distance on $\mathcal{P}_{\mathcal{F}}$

$$d_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} \left| \int_{\Xi} f(\xi)(P - Q)(d\xi) \right| \quad (P, Q \in \mathcal{P}_{\mathcal{F}}). \tag{6}$$

The distance $d_{\mathcal{F}}$ is based on minimal information of the underlying optimization model (1). It is nonnegative, symmetric and satisfies the triangle inequality. At first sight the set $\mathcal{P}_{\mathcal{F}}$ seems to have a complicated structure. For typical applications, however, like for linear two-stage and chance constrained models, the set $\mathcal{P}_{\mathcal{F}}$ or appropriate subsets allow a simple characterization. For example as subsets of $\mathcal{P}(\Xi)$ satisfying certain moment conditions.

To state the following result we need a specific property for set-valued mappings. A set-valued map $M : \mathbb{R} \rightrightarrows \mathbb{R}^m$ with closed graph $\operatorname{gph} M = \{(y, x) : x \in M(y)\}$ has the *Aubin property* at $\bar{y} \in \mathbb{R}$ for $\bar{x} \in M(\bar{y})$ if there are neighborhoods V of $\bar{y}$ and W of $\bar{x}$, and a constant $\kappa \in \mathbb{R}_+$ such that

$$M(y') \cap W \subset M(y) + \kappa |y - y'| \mathbb{B} \quad \text{for all } y, y' \in V,$$

where $\mathbb{B}$ is the unit ball in $\mathbb{R}^m$ (see [42, Definition 9.36]). The next quantitative stability result for problem (1) is a consequence of [43, Theorems 5 and 9].

Proposition 1 *We consider the optimization model (1) with infimum $v(P)$ and solution set $S(P)$ for $P \in \mathcal{P}_{\mathcal{F}}$. Assume that X is compact and*

(i) *the function $x \to \int_{\Xi} f_0(x, \xi) P(d\xi)$ is Lipschitz continuous on X,*
(ii) *the set-valued mapping $y \rightrightarrows \left\{ x \in X : \int_{\Xi} f_1(x, \xi) P(d\xi) \le y \right\}$ has the Aubin property at $\bar{y} = 0$ for each $\bar{x} \in S(P)$.*

Then there exist constants $L > 0$ and $\delta > 0$ such that the estimates

$$|v(P) - v(Q)| \le L\, d_{\mathcal{F}}(P, Q) \tag{7}$$

$$\sup_{x \in S(Q)} d(x, S(P)) \le \Psi_P(L\, d_{\mathcal{F}}(P, Q)) \tag{8}$$

hold whenever $Q \in \mathcal{P}_{\mathcal{F}}$ and $d_{\mathcal{F}}(P, Q) < \delta$. The real-valued function Ψ_P is given by $\Psi_P(r) = r + \psi_P^{-1}(2r)$ for all $r \in \mathbb{R}_+$, where ψ_P is the growth function near the solution set $S(P)$ and $\psi_P(\tau)$ is defined for $\tau \ge 0$ as

$$\inf_{x \in X} \left\{ \int_{\Xi} f_0(x, \xi) P(d\xi) - v(P) : d(x, S(P)) \ge \tau,\ x \in X,\ \int_{\Xi} f_1(x, \xi) P(d\xi) \le 0 \right\}.$$

Note that in case $f_1 \equiv 0$ the estimates hold for $L = 1$ and any $\delta > 0$ and that Ψ_P is lower semicontinuous and increasing on $\mathbb{R}_+$ with $\Psi_P(0) = 0$.

The estimates (7) and (8) in Proposition 1 suggest to choose discrete approximations from $\mathcal{P}_n(\Xi)$ for solving (1) such that they solve the *best approximation problem*

$$\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \tag{9}$$

in order to bound (4) as tight as possible. Determining the scenarios of some solution to (9) may be called *optimal scenario generation*. This choice of discrete approximations was already suggested in [43, Section 4.2], but characterized there as a challenging task which is not solvable in most cases in reasonable time.

It is recommended in [37,43] to eventually enlarge the function class $\mathcal{F}$ such that $d_{\mathcal{F}}$ becomes a metric distance and has further nice properties. Following this idea, however, leads to coarse estimates of the original minimal information distance and, hence, may lead to unfavorable convergence rates of the sequence

$$\left(\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \right)_{n \in \mathbb{N}} \tag{10}$$

and to nonconvex nondifferentiable minimization problems (9) for determining the optimal scenarios.

In linear two-stage stochastic programming the class $\mathcal{F}$ contains piecewise linear-quadratic functions defined on Ξ if condition (A1) (see Sect. 2) is satisfied. If the linear two-stage model has even random recourse, $\mathcal{F}$ may contain more general piecewise polynomial functions (see [45]). Hence, a suitably enlarged class of functions may be chosen as the set

$$\mathcal{F}_r = \left\{ f : \Xi \mapsto \mathbb{R} : f(\xi) - f(\tilde{\xi}) \le \max\left\{1, \|\xi\|, \|\tilde{\xi}\|\right\}^{r-1} \|\xi - \tilde{\xi}\|,\ \forall \xi, \tilde{\xi} \in \Xi \right\} \tag{11}$$

of all locally Lipschitzian functions on Ξ with polynomially growing normalized local Lipschitz constants. Here, $\|\cdot\|$ denotes any norm on $\mathbb{R}^s$ and $r \geq 1$ characterizes the growth of the Lipschitz moduli. The corresponding distance

$$\zeta_r(P, Q) = d_{\mathcal{F}_r}(P, Q) \tag{12}$$

is defined on the set $\mathcal{P}_r(\Xi)$ of all probability measures on Ξ having rth order central moments and is called Fortet–Mourier metric of order r (see [36, Section 5.1]). The Fortet–Mourier metric has a dual representation as a transshipment problem (see [36, Section 5.3]). If Ξ is compact, ζ_r admits even a dual representation as transportation problem (see [38, Section 4.3]), namely, it holds

$$\zeta_r(P, Q) = \inf\left\{\int_{\Xi\times\Xi} c_r(\xi, \tilde{\xi})\eta(d\xi, d\tilde{\xi}) : \eta \circ \pi_1^{-1} = P, \eta \circ \pi_2^{-1} = Q\right\}, \tag{13}$$

where η is a probability measure on $\Xi \times \Xi$, π_1 and π_2 are the projections from $\Xi \times \Xi$ to the first and second component, respectively, c_r is a metric on $\mathbb{R}^s$ and $c_r(\xi, \tilde{\xi})$ is defined as

$$\inf\left\{\sum_{i=1}^{n-1} \max\{1, \|\xi_i\|, \|\xi_{i+1}\|\}^{r-1}\|\xi_i - \xi_{i+1}\| : n \in \mathbb{N}, \xi_i \in \Xi, \xi_1 = \xi, \xi_n = \tilde{\xi}\right\}$$

for all $\xi, \tilde{\xi} \in \Xi$. The representation (13) implies, in particular, that the best approximation problem (9) for $\mathcal{F} = \mathcal{F}_r$ is equivalent to

$$\min_{(\xi^1,\ldots,\xi^n)\in\Xi^n} \int_{\Xi} \min_{i=1,\ldots,n} c_r(\xi, \xi^i)P(d\xi), \tag{14}$$

where ξ^i, $i = 1, \ldots, n$, are the scenarios of $P_n \in \mathcal{P}_n(\Xi)$. This follows similarly as in [16, Lemma 4.2]. For $r = 1$ the probabilities w_i of ξ^i can be computed by $w_i = P(A_i)$, $i = 1, \ldots, n$, where the collection $\{A_i : i = 1, \ldots, n\}$ is a Voronoi partition of Ξ, i.e., A_i is Borel measurable and a subset of

$$\left\{\xi \in \Xi : \|\xi - \xi^i\| = \min_{j=1,\ldots,n} \|\xi - \xi^j\|\right\} \quad (i = 1, \ldots, n).$$

Note that the objective function in (14) is continuous and inf-compact on Ξ^n. Hence, the minimization problem (14) is solvable, but nonconvex for $n \geq 2$ even for $r = 1$. Furthermore, due to a classical result (see [6, Proposition 2.1]), the estimate

$$c\, n^{-\frac{1}{s}} \leq \zeta_1(P, P_n) \leq \zeta_r(P, P_n)$$

holds for each $P_n \in \mathcal{P}_n(\Xi)$, sufficiently large n and some constant $c > 0$ if P has a density on Ξ. Hence, the convergence rate (10) for $\mathcal{F} = \mathcal{F}_r$ is worse than the Monte Carlo rate $O(n^{-\frac{1}{2}})$ if the dimension s of Ξ is greater than two.

The approach to *optimal scenario reduction* for linear two-stage stochastic programs developed in [10] is based on Monge-Kantorovich functionals and applies to the Fortet–Mourier metric ζ_r [see (12)] due to the representation (13). Starting with a discrete probability measure P based on a large number N of scenarios, it selects a smaller number n of scenarios out of the original set of scenarios together with new probabilities such that the new discrete probability measure represents the best approximation to P with respect to ζ_r. More precisely, let P have the scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$. Using the dual representation (13) of ζ_r the best approximation problem

$$\min_{Q \in \mathcal{P}_n(\operatorname{supp} P)} \zeta_r(P, Q)$$

can be rewritten as combinatorial program

$$\min \left\{ \sum_{i,j=1}^{N} p_i x_{ij} c_r(\xi^i, \xi^j) \;\middle|\; \begin{array}{c} \sum_{i=1}^{N} x_{ij} = 1\,,\ j = 1, \ldots, N,\ \sum_{i=1}^{N} y_i \le n\,, \\ x_{ij} \le y_i\,,\ x_{ij} \in \{0,1\}\,,\ i, j = 1, \ldots, N\,, \\ y_i \in \{0,1\}\,,\ i = 1, \ldots, N \end{array} \right\}, \tag{15}$$

where the variable y_i decides whether scenario ξ^i remains and x_{ij} selects a scenario ξ^j that minimizes the distance $c_r(\cdot\,, \xi^i)$. We note that (15) is known as n-median problem (see [4]) which is NP-hard as shown in [27].

If J denotes a subset of $\{1, \ldots, N\}$ with cardinality $|J| = n$, the best approximation problem can be decomposed into finding the optimal index set J of remaining scenarios and into determining the optimal discrete probability measure given J. With P_J denoting any probability measure with support consisting of the scenarios ξ^j, $j \in J$, the best approximation problem has a solution $P_J^\star$ such that

$$\zeta_r(P, P_J^*) = \min_{P_J} \zeta_r(P, P_J) = \sum_{i \notin J} p_i \min_{j \in J} c_r(\xi^i, \xi^j) \tag{16}$$

$$\text{with } P_J^\star \text{ given by} \quad P_J^\star = \sum_{j \in J} \pi_j \xi^j\,, \quad \text{where} \quad \pi_j = p_j + \sum_{i \in I_j} p_i \quad (\forall j \in J) \tag{17}$$

and the index sets I_j, $j \in J$, are defined by $I_j := \{i \in \{1, \ldots, N\} \backslash J : j = j(i)\}$ with $j(i) \in \arg\min_{j \in J} c_r(\xi^i, \xi^j)$, $\forall i \notin J$. The formula (17) for the optimal weights is called redistribution rule in [10,19] where the results (16) and (17) are proved, too. For a survey of theory and algorithms for n-median problems we refer the interested reader to [4]. Presently local search heuristics [1] and a novel approximation algorithm [30] seem to be the most favorable algorithms with best approximation guarantees. Simple alternatives without approximation guarantees are forward and backward greedy heuristics developed and tested in [18, Algorithms 2.2 and 2.4], [19]. The scenario reduction approach described above has been extended to discrepancy distances in [20,21]. The latter distances are of the form

$$\alpha(P, Q) = \sup_{B \in \mathcal{B}} |P(B) - Q(B)| \quad (P, Q \in \mathcal{P}(\Xi)), \tag{18}$$

where $\mathcal{B}$ is a suitable class of Borel subsets of Ξ. Such distances are relevant for chance constrained stochastic programs if $\mathcal{B}$ contains the relevant sets (for example, the polyhedra $\mathcal{P}(x)$). We recall, however, that employing probability metrics like (12) and (18) means that decisions on reducing scenarios are based on *coarse estimates* of the minimal information distances (6) and, thus, do essentially not depend on the specific stochastic program. Possibly due to this observation several authors developed specific heuristic approaches to scenario generation and reduction for specific applications (see, for example, [13,32]). These developments served in turn as a motivation for the work reported in the present paper.

We will show in this paper that the optimal scenario generation problem (9) may have favorable solution properties if the set $\mathcal{F}$ remains as small as possible, i.e., as chosen in (5). In Sect. 2 we demonstrate this for linear two-stage stochastic programs. First we show that (9) can be formulated as generalized semi-infinite program (Theorem 1) which is convex in some cases (Theorem 2), enjoys stability (Theorem 3) and allows a transformation into a standard semi-infinite program in a number of cases. In Sect. 3 we revisit the problem of optimal scenario reduction for two-stage models and provide a new formulation based on the minimal information distance (6) as mixed-integer linear semi-infinite program. The latter decomposes into solving binary and linear semi-infinite programs recursively. Section 4 presents a mixed-integer linear semi-infinite program for optimal scenario generation in chance constrained programming. Finally we illustrate the approach to scenario generation for the classical newsvendor problem and finish with conclusions.

2 Optimal scenario generation for two-stage models by generalized semi-infinite programming

We consider a linear two-stage stochastic program (1) with the integrand (2), a probability distribution P on $\mathbb{R}^s$ and with Φ denoting the infimal value (3) of the second-stage program. Furthermore, we impose the following conditions in addition to the general assumptions made in Sect. 1:

(A0) X is a bounded polyhedron and Ξ is convex polyhedral.
(A1) $h(x, \xi) \in W(Y) = \{Wy : y \in Y\}$ and $q(\xi) \in \mathcal{D}$ hold for all $(x, \xi) \in X \times \Xi$,
(A2) P has a second order absolute moment.

Condition (A1) combines the usual conditions *relatively complete recourse* and *dual feasibility* and (A2) implies that all integrals are finite. Both conditions are standard for two-stage stochastic programs. In particular, (A0)–(A2) imply that the infima $v(P)$ and $v(P_n)$ are attained and the estimate

$$\begin{aligned} |v(P) - v(P_n)| &\leq \sup_{x \in X} \left| \int_{\Xi} f_0(x, \xi) P(d\xi) - \int_{\Xi} f_0(x, \xi) P_n(d\xi) \right| \\ &= \sup_{x \in X} \left| \int_{\Xi} \Phi(q(\xi), h(x, \xi)) P(d\xi) - \int_{\Xi} \Phi(q(\xi), h(x, \xi)) P_n(d\xi) \right| \end{aligned}$$

holds due to Proposition 1 for every $P_n \in \mathcal{P}_n(\Xi)$. Hence, the formulation of the optimal scenario generation problem for (1), (2) based on the minimal information distance (6) consists in solving the best uniform approximation problem

$$\min_{\substack{(\xi^1,\ldots,\xi^n)\in\Xi^n \\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \sup_{x\in X} \left| \int_\Xi \Phi(q(\xi), h(x,\xi)) P(d\xi) - \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x,\xi^i)) \right|. \quad (19)$$

It means that the convex expected recourse function $F_P : X \to \mathbb{R}$

$$F_P(x) := \int_\Xi \Phi(q(\xi), h(x,\xi)) P(d\xi) \quad (20)$$

has to be approximated uniformly on X by the best convex combination of n convex polyhedral functions appearing as integrand in F_P.

Note that the minimal class $\mathcal{F} = \{\Phi(q(\cdot), h(x,\cdot)) : x \in X\}$ of functions from Ξ to $\mathbb{R}$ enjoys specific properties. All functions are finite, continuous and piecewise linear-quadratic on Ξ. They are linear-quadratic on each convex polyhedral set

$$\Xi_j(x) = \{\xi \in \Xi : (q(\xi), h(x,\xi)) \in \mathcal{K}_j\} \quad (j = 1,\ldots,\ell),$$

where the convex polyhedral cones $\mathcal{K}_j$, $j = 1,\ldots,\ell$, represent a decomposition of the domain of Φ, which is itself a convex polyhedral cone in $\mathbb{R}^{\tilde{m}+r}$. The latter decomposition depends only on the matrix W [54]. In particular, the functions $\Phi(q(\cdot), h(x,\cdot))$ are locally Lipschitz continuous where the Lipschitz constants on the balls $\{\xi \in \Xi : \|\xi\| \le \rho\}$ grow linearly with ρ and can be chosen uniform with respect to $x \in X$ (see [43, Proposition 22]).

It is well-known that best uniform approximation problems may be reformulated as *semi-infinite programs* (SIP), i.e., as optimization problems with finitely many variables, but infinitely many constraints. We show next that (19) leads to a *generalized semi-infinite program* (GSIP), that is, to a SIP in which the index set of the constraints is infinite and depends on the decision. Theory and numerical methods for such models are studied in a number of publications. We mention here the monograph [50] and the tutorial [17].

Theorem 1 *Assume (A0)–(A2). Then (19) is equivalent to the GSIP*

$$\min_{\substack{t\ge 0 \\ (\xi^1,\ldots,\xi^n)\in\Xi^n \\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^n w_i \langle h(x,\xi^i), z_i\rangle \le t + F_P(x) \\ F_P(x) \le t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i\rangle \\ \forall (x,y,z) \in \mathcal{M}(\xi^1,\ldots,\xi^n) \end{array} \right\}, \quad (21)$$

where the set-valued mapping $\mathcal{M}$ from Ξ^n to $\mathbb{R}^{m+(\tilde{m}+r)n}$ is defined by

$$\mathcal{M}(\xi) = \{(x,y,z) \in X \times Y^n \times \mathbb{R}^{rn} : Wy_i = h(x,\xi^i), \; W^\top z_i - q(\xi^i) \in Y^*, \; i = 1,\ldots,n\} \quad (22)$$

for all $\xi = (\xi^1, \ldots, \xi^n) \in \Xi^n$ *and* $F_P : X \to \mathbb{R}$ *is given by* (20). *If the function* h *is affine, the feasible set of* (21) *is closed.*

Proof By the standard way of rewriting best uniform approximation problems one obtains first by introducing the auxiliary variable t that the semi-infinite program

$$\min_{\substack{t\geq 0\\ (\xi^1,\ldots,\xi^n)\in\Xi^n\\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x,\xi^i)) \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x,\xi^i)) \\ \forall x \in X \end{array} \right\} \tag{23}$$

is equivalent to (19). Next we exploit the duality relation

$$\Phi(q,t) = \inf\{\langle q, y\rangle : Wy = t,\, y \in Y\} = \sup\{\langle t, z\rangle : W^\top z - q \in Y^*\}$$

of the second-stage program for all pairs $(q,t) \in \mathcal{D} \times W(Y)$. Then the primal and dual program are both solvable. Due to (A1) the semi-infinite program (23) may be reformulated as

$$\min_{\substack{t\geq 0\\ (\xi^1,\ldots,\xi^n)\in\Xi^n\\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^n w_i \sup\{\langle h(x,\xi^i), z\rangle : W^\top z - q(\xi^i) \in Y^*\} \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \inf\{\langle q(\xi^i), y\rangle : Wy = h(x,\xi^i),\ y \in Y\} \\ \forall x \in X \end{array} \right\}. \tag{24}$$

Next we introduce $2n$ new variables $y_i \in Y$ with $Wy_i = h(x,\xi^i)$ and $z_i \in \mathbb{R}^r$ with $W^\top z_i - q(\xi^i) \in Y^*$, $i = 1, \ldots, n$, and consider the generalized linear semi-infinite program (21). Then any $t \geq 0$ and $(\xi^1, \ldots, \xi^n) \in \Xi^n$ solving problem (24) satisfies the constraints of (21).On the other hand, if $t \geq 0$ and $(\xi^1, \ldots, \xi^n) \in \Xi^n$ attain the minimum in (21), the two inequalities

$$\sum_{i=1}^n w_i \langle h(x,\xi^i), z_i\rangle \leq t + F_P(x) \quad \text{and} \quad F_P(x) \leq t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i\rangle$$

are satisfied for all $(x, y, z) \in \mathcal{M}(\xi^1, \ldots, \xi^n)$. Hence, the inequalities

$$\sum_{i=1}^n w_i \sup\{\langle h(x,\xi^i), z\rangle : W^\top z - q(\xi^i) \in Y^*\} \leq t + F_P(x)$$

$$F_P(x) \leq t + \sum_{i=1}^n w_i \inf\{\langle q(\xi^i), y\rangle : Wy = h(x,\xi^i),\ y \in Y\}$$

are satisfied for all $x \in X$. Hence, programs (24) and (21) are equivalent.

To show that the feasible set of (21) is closed, we know from [50, Corollary 3.1.21] (see also [17, Proposition 3.4]) that the lower semicontinuity of $\mathcal{M}$ on Ξ^n is a sufficient condition. Since the graph $\operatorname{gph}\mathcal{M}$ of $\mathcal{M}$ is of the form

$$\operatorname{gph}\mathcal{M} = \big\{(\xi^1,\dots,\xi^n,x,y,z) \in \Xi^n \times X \times Y^n \times \mathbb{R}^{rn} : Wy_i = h(x,\xi^i),\\ W^\top z_i - q(\xi^i) \in Y^*,\ i = 1,\dots,n\big\}$$

and h is affine, $\operatorname{gph}\mathcal{M}$ is convex polyhedral. Such set-valued mappings are even Hausdorff Lipschitz continuous on its domain (see, for example, [42, Example 9.35]) and, hence, on Ξ^n due to (A1). This completes the proof. □

In general the optimization model (21) is not convex even when the the weights w_i, $i = 1,\dots,n$, are fixed. However, we prove now that the model is convex if the function h is affine and either only right-hand sides or only costs are random.

Theorem 2 *Assume (A0)–(A2), let the function h be affine, the weights w_i, $i = 1,\dots,n$, be fixed and let either h or q be random. Then the feasible set of the GSIP (21) is closed and convex.*

Proof Let q be nonrandom. Then the feasible set M of (21) is of the form

$$M = \left\{(t,\xi^1,\dots,\xi^n) \in \mathbb{R}_+ \times \Xi^n \;\middle|\; \begin{array}{c} \sum_{i=1}^n w_i\langle h(x,\xi^i), z_i\rangle - t \le F_P(x) \\ F_P(x) \le t + \sum_{i=1}^n w_i\langle q, y_i\rangle \\ \forall (x,y,z) \in \mathcal{M}(\xi^1,\dots,\xi^n) \end{array}\right\}. \tag{25}$$

Let $\alpha \in [0,1]$ and $\xi_j = (\xi_j^1,\dots,\xi_j^n) \in \Xi^n$, $t_j \in \mathbb{R}_+$, be such that $(t_j,\xi_j) \in M$, $j = 1,2$. We have to show that $\alpha(t_1,\xi_1) + (1-\alpha)(t_2,\xi_2)$ belongs to M, too.

Let $x \in X$ and $z_i \in \{z \in \mathbb{R}^r : W^\top z - q \in Y^\star\}$ for $i = 1,\dots,n$ be chosen arbitrarily. Then we have

$$\begin{aligned} &\sum_{i=1}^n w_i\langle h(x,\alpha\xi_1^i + (1-\alpha)\xi_2^i), z_i\rangle - \alpha t_1 - (1-\alpha)t_2 \\ &\quad = \alpha\left(\sum_{i=1}^n w_i\langle h(x,\xi_1^i), z_i\rangle - t_1\right) + (1-\alpha)\left(\sum_{i=1}^n w_i\langle h(x,\xi_2^i), z_i\rangle - t_2\right) \\ &\quad \le \alpha F_P(x) + (1-\alpha)F_P(x) = F_P(x). \end{aligned}$$

Now, let $y_{ij} \in \{y \in Y : Wy = h(x,\xi_j^i)\}$ for $j = 1,2$, $i = 1,\dots,n$, be chosen arbitrarily in addition. We obtain $\alpha y_{i1} + (1-\alpha)y_{i2} \in \{y \in Y : Wy = h(x,\alpha\xi_1^i + (1-\alpha)\xi_2^i)\}$ and, hence,

$$\begin{aligned}
&\alpha t_1 + (1-\alpha)t_2 + \sum_{i=1}^{n} w_i \langle q, \alpha y_{i1} + (1-\alpha) y_{i2} \rangle \\
&= \alpha \Big(t_1 + \sum_{i=1}^{n} w_i \langle q, y_{i1} \rangle \Big) + (1-\alpha) \Big(t_2 + \sum_{i=1}^{n} w_i \langle q, y_{i2} \rangle \Big) \\
&\geq \alpha F_P(x) + (1-\alpha) F_P(x) = F_P(x).
\end{aligned}$$

This means $\alpha(t_1, \xi_1) + (1-\alpha)(t_2, \xi_2) \in M$ and M is convex. If q is random, but h nonrandom, the proof is similar. The closedness of M follows from Theorem 1. □

For fixed weights and given $n \in \mathbb{N}$ the GSIP (21) for determining the optimal scenarios $\xi^i, i = 1, \ldots, n$, is of dimension $n\,s+1$ and, thus, large scale in many cases. A difficulty of (21) is that the set $\mathcal{M}(\xi^1, \ldots, \xi^n)$ is unbounded even in general.

We note that $F_P(x)$ can only be calculated approximately even if the probability measure P is completely known. Hence, it becomes important that the optimization model (21) behaves stable when the function F_P is perturbed. The following result shows that even Lipschitz stability of the optimal values can be expected if the conditions of Theorem 2 are imposed.

Theorem 3 *Assume (A0) –(A2) and that the infimum $v(F_P)$ of (21) is positive. Let the function h be affine, let either h or q be random and the weights w_i, $i = 1, \ldots, n$, be fixed. Then there exist $\kappa > 0$ and $\delta > 0$ such that*

$$|v(F_P) - v(F)| \leq \kappa \sup_{x \in X} |F_P(x) - F(x)|, \tag{26}$$

for each continuous function F on X such that $\sup_{x \in X} |F_P(x) - F(x)| < \delta$. Here, $v(F)$ denotes the optimal value of (21) with F_P replaced by F.

Proof As in the proof of Theorem 2 we assume without loss of generality that q is nonrandom. We consider the set-valued mapping $(t, \xi^1, \ldots, \xi^n) \mapsto \Lambda(t, \xi^1, \ldots, \xi^n)$ from $\mathbb{R}_+ \times \Xi^n$ to the Banach space $C(X)$ of real-valued continuous functions on X with the standard norm $\|\cdot\|_\infty$, where

$$\Lambda(t, \xi^1, \ldots, \xi^n) = \left\{ f \in C(X) \;\middle|\; \begin{array}{c} \sum_{i=1}^{n} w_i \langle h(x, \xi^i), z_i \rangle - t - F_P(x) \leq f(x) \\ f(x) \leq t + \sum_{i=1}^{n} w_i \langle q, y_i \rangle - F_p(x) \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \ldots, \xi^n) \end{array} \right\}.$$

First, we show that the graph of the set-valued mapping Λ denoted by gph Λ is convex. Let $\alpha \in [0, 1]$ and $f_j \in C(X)$, $\xi_j = (\xi_j^1, \ldots, \xi_j^n) \in \Xi^n$, $t_j \in \mathbb{R}_+$, be such that $(t_j, \xi_j, f_j) \in$ gph Λ, $j = 1, 2$. Then we obtain as in the proof of Theorem 2

$$
\begin{aligned}
&\sum_{i=1}^{n} w_i \langle h(x, \alpha\xi_1^i + (1-\alpha)\xi_2^i), z_i\rangle - \alpha t_1 - (1-\alpha)t_2 - F_P(x) \\
&\quad = \alpha\Big(\sum_{i=1}^{n} w_i \langle h(x, \xi_1^i), z_i\rangle - t_1 - F_P(x)\Big) \\
&\qquad +(1-\alpha)\Big(\sum_{i=1}^{n} w_i \langle h(x, \xi_2^i), z_i\rangle - t_2 - F_P(x)\Big) \\
&\quad \le \alpha f_1(x) + (1-\alpha) f_2(x) \qquad\qquad \text{and} \\
&\qquad \alpha t_1 + (1-\alpha)t_2 + \sum_{i=1}^{n} w_i \langle q, \alpha y_{i1} + (1-\alpha) y_{i2}\rangle - F_P(x) \\
&\quad \ge \alpha f_1(x) + (1-\alpha) f_2(x),
\end{aligned}
$$

where $x \in X$, $z_i \in \{z \in \mathbb{R}^r : W^\top z - q \in Y^\star\}$ and $y_{ij} \in \{y \in Y : Wy = h(x, \xi_j^i)\}$ for $j = 1, 2, i = 1, \ldots, n$, are chosen arbitrary. This proves that gph Λ is convex. It is also closed as subset of $\mathbb{R}^{ns+1} \times C(X)$. Furthermore, we know that the null function $0 \in C(X)$ belongs to the range of Λ and that $\Lambda^{-1}(0)$ is just the feasible set of (21). Thus, there exists $(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n) \in \mathbb{R}_+ \times \Xi^n$ such that $0 \in \Lambda(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n)$. We know that $\bar{t} \ge v(F_P) > 0$ by assumption. Next we choose δ such that $0 < \delta < \bar{t}$ and conclude that the closed ball $\mathbb{B}(0, \delta)$ in $C(X)$ is contained in the range of Λ. The Robinson-Ursescu theorem (see [41, Theorem 2]) on continuity properties of set-valued mappings having closed convex graphs then implies that the inverse multifunction Λ^{-1} has the Aubin property at $f = 0$ for any point $(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n) \in \Lambda^{-1}(0)$ with $\bar{t} > 0$. This means that there exist neighborhoods U of 0 and W of $(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n)$, and a constant $\kappa \in \mathbb{R}_+$ such that

$$
\Lambda^{-1}(f) \cap W \subseteq \Lambda^{-1}(\tilde{f}) + \kappa \|f - \tilde{f}\|_\infty \mathbb{B} \tag{27}
$$

holds for all $f, \tilde{f} \in U$, where $\mathbb{B}$ is the unit ball in $\mathbb{R}^{ns+1}$.

Next we choose $f = F - F_P$ with $F \in C(X)$ and $\tilde{f} = 0$.

Let $\varepsilon > 0$ and $(\tilde{t}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n) \in \Lambda^{-1}(f) \cap W$ such that $v(f) \le \tilde{t} \le v(f) + \varepsilon$. Then there exists an element $(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n) \in \Lambda^{-1}(0)$ such that

$$
\|(\bar{t}, \bar{\xi}^1, \ldots, \bar{\xi}^n) - (\tilde{t}, \tilde{\xi}^1, \ldots, \tilde{\xi}^n)\| \le \kappa \|f\|_\infty
$$

holds for all $f \in U$ due to the Aubin property (27) of Λ^{-1} at 0. We note that $\Lambda^{-1}(F - F_P)$ is the constraint set of (21) with F_P replaced by F, respectively, and obtain that the estimates

$$
v(F_P) - v(F) \le \bar{t} - v(F) \le |\bar{t} - \tilde{t}| - \varepsilon \le \kappa \|F - F_P\|_\infty - \varepsilon\,.
$$

hold for all $F \in C(X)$ with $F - F_P \in U$. Since the latter estimate is valid for any $\varepsilon > 0$, we obtain $v(F_P) - v(F) \le \kappa \|F - F_P\|_\infty$ if $F - F_P \in U$. In the same way we can derive the estimate $v(F) - v(F_P) \le L\|F - F_P\|_\infty$ if $F - F_P \in U$. It remains to select $\delta > 0$ such that the open ball around 0 with radius δ in $C(X)$ is contained in V and to require $\|F - F_P\|_\infty < \delta$. Finally, we note that starting with the Aubin property

of Λ^{-1} at $0 \in C(X)$ the proof followed classical arguments of quantitative stability in optimization (see [28, Theorem 1]). □

Notice that the infimum $v(F_P)$ of (21) is always nonnegative and $v(F_P) = 0$ means that P has at most n scenarios. Hence, the assumption $v(F_P) > 0$ is natural and satisfied, for example, for any n if P is a continuous probability distribution. Theorem 3 applies, for example, if F_P is approximated by Monte Carlo or Quasi-Monte Carlo methods with a large sample size $N > n$. Let

$$F_P(x) \approx \frac{1}{N} \sum_{j=1}^{N} \Phi(q(\hat{\xi}^j), h(x, \hat{\xi}^j))$$

be such an approximate representation of $F_P(x)$ based on a sample $\hat{\xi}^j$, $j = 1, \ldots, N$. Inserting this approximation into (21) and exploiting again the duality relation then leads to the following approximate version of (21)

$$\min_{\substack{t \geq 0, (\xi^1, \ldots, \xi^n) \in \Xi^n \\ (w_1, \ldots, w_n) \in \mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^{n} w_i \langle h(x, \xi^i), z_i \rangle \leq t + \frac{1}{N} \sum_{j=1}^{N} \langle q(\hat{\xi}^j), \hat{y}_j \rangle \\ \frac{1}{N} \sum_{j=1}^{N} \langle h(x, \hat{\xi}^j), \hat{z}_j \rangle \leq t + \sum_{i=1}^{n} w_i \langle q(\xi^i), y_i \rangle \\ \forall (x, \hat{y}, \hat{z}) \in \mathcal{M}(\hat{\xi}^1, \ldots, \hat{\xi}^N) \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \ldots, \xi^n) \end{array} \right\}, \tag{28}$$

where the sample $\hat{\xi}^j$, $j = 1, \ldots, N$ is given. The latter problem may also be characterized as a *scenario clustering problem*: Given a large scenario set $\hat{\xi}^j$, $j = 1, \ldots, N$, we are looking for a smaller scenario set ξ^i, $i = 1, \ldots, n$, where each scenario ξ^j corresponds to a cluster $\hat{\xi}^i$, $i \in I_j$, of the original scenarios.

The specific structure of (21) and (28) as generalized semi-infinite programs is promising and allows for specific solution algorithms (see [17,50–52]).

In a number of cases it is even possible to reduce the GSIP (21) to a semi-infinite program by a transformation inspired by the recent paper [48]. To describe the idea, we consider the situation that only costs are random, the polyhedral cone Y is given by $Y = \mathbb{R}^{\bar{m}}_+$ and the transformation is defined by

$$t : \Xi \times \mathcal{U} \to \mathbb{R}^r, \quad t(\xi, u) = u + (W^+)^\top (q(\xi) - \bar{q}), \tag{29}$$

where $\mathcal{U} = \{u \in \mathbb{R}^r : W^\top u \leq \bar{q}\}$, $\bar{q} \in \mathbb{R}^{\bar{m}}$ and the $(\bar{m}, r)$ matrix W^+ denotes the Moore-Penrose inverse of W.

Proposition 2 *Assume (A0) and (A2), $h(x) \in W(\mathbb{R}^{\bar{m}}_+)$ for all $x \in X$ and that $\bar{q}$ and $q(\xi)$ belong to the range of $W^\top$ for all $\xi \in \Xi$. Then the generalized semi-infinite program (21) is equivalent to the semi-infinite program*

$$\min_{\substack{t\geq 0\\ (\xi^1,\ldots,\xi^n)\in\Xi^n\\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^{n} w_i \langle h(x), u_i + (W^+)^\top (q(\xi^i) - \bar{q})\rangle \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^{n} w_i \langle q(\xi^i), y_i\rangle \\ \forall (x, y_1, \ldots, y_n, u_1, \ldots, u_n) \in X \times \mathcal{Y}(x)^n \times \mathcal{U}^n \end{array} \right\}, \tag{30}$$

where $\mathcal{Y}(x) = \{y \in \mathbb{R}^{\tilde{m}}_+ : Wy = h(x)\}$ *for each* $x \in X$. *If the weights* w_i, $i = 1, \ldots, n$, *are fixed,* (30) *is a linear semi-infinite program.*

Proof The mapping t given by (29) has the property

$$W^\top t(\xi, u) = W^\top u + W^\top (W^\top)^+ (q(\xi) - \bar{q}) \leq \bar{q} + (q(\xi) - \bar{q}) = q(\xi)$$

for each pair $(\xi, u) \in \Xi \times \mathcal{U}$ as $q(\xi) - \bar{q} \in W^\top(\mathbb{R}^r)$ and $W^\top (W^\top)^+$ is just the orthogonal projection onto $W^\top(\mathbb{R}^r)$. Hence, it holds

$$t(\xi, \mathcal{U}) = \{z \in \mathbb{R}^r : W^\top z \leq q(\xi)\}.$$

The desired equivalence between (21) and (30) follows by setting $z_i = u_i + (W^+)^\top (q(\xi^i) - \bar{q})$, $i = 1, \ldots, n$, in (21). □

A similar result can be derived if only right-hand sides are random and the primal polyhedral constraint set of the linear second-stage problem is given in the form $\{y \in \mathbb{R}^{\tilde{m}} : Wy \leq h(x, \xi)\}$. Proposition 2 opens the possibility of using classical solution algorithms for semi-infinite programs, in particular, discretization and exchange methods (see the monographs [15,23] and the surveys [22,40]). In the "Appendix" we provide a short description of the discretization method due to [39].

Finally, we discuss the possible use of lower and upper bounds of $F_P(x)$ for scenario generation. There is a well-developed theory for deriving lower und upper bounds of expectation functionals of convex-concave integrands. While lower bounds are due to Jensen's classical result (e.g., see [7, Theorem 10.2.6]), upper bounds are known as Edmundson-Madansky bounds. They were further developed in the context of stochastic programming, for example, in [2,5,11,12,14,25]. Many upper bounds are derived via generalized moment problems appearing as duals of semi-infinite programs [12,25] (see also [26, Section 3.2.1]).

Let $l_P(x)$ and $u_P(x)$ denote lower and upper bounds of $F_P(x)$, respectively. Then the following optimization problem (derived from (21)) computes upper bounds of the infima to (19) or (21), respectively:

$$\min_{\substack{t\geq 0, (\xi_1,\ldots,\xi_n)\in\Xi^n\\ (w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^{n} w_i \langle h(x, \xi^i), z_i\rangle \leq t + l_P(x), \\ u_P(x) \leq t + \sum_{i=1}^{n} w_i \langle q(\xi^i), y_i\rangle, \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \ldots, \xi^n) \end{array} \right\}. \tag{31}$$

If $l_P(x)$ and $u_P(x)$ are exchanged, the optimization problem (31) provides lower bounds of the infima to (21). These observations may be of interest for the numerical solution of (21) if it is nonconvex.

3 Optimal scenario reduction for two-stage models

Next we discuss the scenario reduction approach for two-stage models based on the minimal information distance (5) and the best approximation problem (9).

As in Sect. 1 let ξ^i, $i = 1, \ldots, N$, be a large set of scenarios with probabilities p_i, $i = 1, \ldots, N$, that define a discrete probability measure P. For prescribed $n \in \mathbb{N}$, $n < N$, we intend to determine an index set $J \subset \{1, \ldots, N\}$ of cardinality $|J| = n$ and new weights $\bar{\pi}_j$, $j \in J$, such that the probability measure

$$P_J^* = \sum_{j \in J} \bar{\pi}_j \delta_{\xi^j}$$

solves the optimal scenario reduction problem

$$\min \left\{ \sup_{x \in X} \left| \sum_{j \in J} \pi_j \varphi_j(x) - \sum_{i=1}^{N} p_i \varphi_i(x) \right| : J \subset \{1, \ldots, N\}, |J| = n, \pi \in \mathcal{S}_n \right\}, \tag{32}$$

where the functions $\varphi_i(x) = \Phi(q(\xi^i), h(x, \xi^i))$, $i = 1, \ldots, N$, are convex polyhedral on X. Problem (32) represents a mixed-integer semi-infinite program. Compared with (15), (32) is based on Proposition 1 and, hence, on a (much) smaller upper bound for the difference of the optimal values. In addition, the solution of problem (32) depends on the data of the two-stage stochastic program.

Problem (32) decomposes into finding the optimal index set J of remaining scenarios and into determining the optimal weights π_j, $j \in J$, given J. The outer combinatorial optimization problem

$$\min \{D(J, P) : J \subset \{1, \ldots, N\}, |J| = n\}, \tag{33}$$

determines the index set J and can be reformulated as binary optimization problem similar to (15). Here, the objective function $D(J, P)$ denotes the infimum of the inner program

$$\min_{\pi \in \mathcal{S}_n} \sup_{x \in X} \left| \sum_{j \in J} \pi_j \varphi_j(x) - \sum_{i=1}^{N} p_i \varphi_i(x) \right|. \tag{34}$$

Any evaluation of the objective in (33) requires the solution of the best approximation problem (34). Hence, compared to the problem (16) in the introduction, the infimum of problem (34) cannot be computed explicitly.

For linear two-stage stochastic programs satisfying (A0)–(A2) the optimization model (34) contains finite functions and is equivalent to the reduced linear semi-infinite program

$$\min_{t\geq 0,\pi\in\mathcal{S}_n}\left\{t\;\middle|\;\begin{array}{c}\sum\limits_{j\in J}\pi_j\varphi_j(x)\leq t+\sum\limits_{i=1}^{N}p_i\varphi_i(x)\\ \sum\limits_{i=1}^{N}p_i\varphi_i(x)\leq t+\sum\limits_{j\in J}\pi_j\varphi_j(x)\\ \forall x\in X\end{array}\right\} \tag{35}$$

or to

$$\min_{t\geq 0,\pi\in\mathcal{S}_n}\left\{t\;\middle|\;\begin{array}{c}\sum\limits_{j\in J}\pi_j\langle h(x,\xi^j),z_j\rangle\leq t+\sum\limits_{i=1}^{N}p_i\langle q(\xi^i),y_i\rangle\\ \sum\limits_{i=1}^{N}p_i\langle h(x,\xi^i),z_i\rangle\leq t+\sum\limits_{j\in J}\pi_j\langle q(\xi^j),y_j\rangle\\ \forall(x,y,z)\in\mathcal{M}(\xi^1,\ldots,\xi^N)\end{array}\right\}, \tag{36}$$

where the set $\mathcal{M}(\xi^1,\ldots,\xi^N)$ is defined as in (22) with n replaced by N. Hence, the linear semi-infinite program (36) has a comparably low number $n+1$ of variables, but a (very) high-dimensional index set.

Problems (35) and (36) mean: For a given convex combination of many convex polyhedral functions $\varphi_i(\cdot)$ on X one is looking for the best convex combination of a given subset of convex polyhedral functions that approximates the former uniformly.

4 Scenario generation for chance constrained programs

We consider a chance constrained program

$$\min\{g(x) : x\in X,\ P(\mathcal{P}(x))\geq p\},$$

where $\mathcal{P}(x)=\{\xi\in\Xi : h(x,\xi)\leq 0\}$ is a polyhedron depending on x, g is a linear objective g, X and Ξ are polyhedral, h a function as described in Sect. 1 and $p\in(0,1)$ a given probability level. Then we have $f_0(x,\xi)=g(x)$ and $f_1(x,\xi)=p-\mathbb{1}_{\mathcal{P}(x)}(\xi)$, and the best approximation problem (9) is of the form

$$\min_{t\geq 0,\ P_n\in\mathcal{P}_n(\Xi)}\left\{t\;\middle|\;\begin{array}{c}P(\mathcal{P}(x))\leq t+P_n(\mathcal{P}(x))\\ P_n(\mathcal{P}(x))\leq t+P(\mathcal{P}(x))\\ \forall x\in X\end{array}\right\} \tag{37}$$

and, thus,

$$P_n(\mathcal{P}(x))=\sum_{i=1}^{n}w_i\,\mathbb{1}_{\mathcal{P}(x)}(\xi^i)=\sum_{i=1}^{n}w_i\,\mathbb{1}_{\mathbb{R}^r_-}(h(x,\xi^i))\quad(x\in X).$$

It is well-known that chance constrained optimization models with discrete probability distributions are nonconvex in general (see, for example, [26, Section 2.2.2]), but can be reformulated as mixed-integer programs. We follow here the presentation in [26, Section 2.2.2] and choose a constant $M>0$ such that

$$h(x, \xi^i) - M\, e \leq 0 \quad \forall x \in X, \tag{38}$$

holds for each $i = 1, \ldots, n$, where $e = (1, \ldots, 1)^\top \in \mathbb{R}^r$. Such constant M always exists as X is compact. This allows to introduce binary variables $z_i \in \{0, 1\}$ such that $z_i = 0$ if $h(x, \xi^i) \leq 0$ for all $x \in X$ and $z_i = 1$ otherwise, $i = 1, \ldots, n$.

Then it is possible to reformulate (37) as mixed-integer semi-infinite program

$$\min_{\substack{t \geq 0,\ (\xi^1, \ldots, \xi^n) \in \Xi^n \\ (w_1, \ldots, w_n) \in \mathcal{S}_n \\ (z_1, \ldots, z_n) \in \{0,1\}^n}} \left\{ t \;\middle|\; \begin{array}{c} P(\mathcal{P}(x)) \leq t + \sum_{i=1}^n w_i (1 - z_i) \\ \sum_{i=1}^n w_i (1 - z_i) \leq t + P(\mathcal{P}(x)) \\ h(x, \xi^i) - z_i M e \leq 0,\ i = 1, \ldots, n \\ \forall x \in X \end{array} \right\}. \tag{39}$$

If the weights w_i, $i = 1, \ldots, n$, are fixed, problem (39) is a mixed-integer linear semi-infinite optimization model.

Since mixed-integer linear programs containing 'big-M' type constraints are often difficult to solve, one is interested in strengthening the formulation of (39) by incorporating valid inequalities. A possible way consists in introducing precedence constraints based on partial orders $\preceq$ on the index set $\{1, \ldots, n\}$. Such a partial order $\preceq$ is called strongly consistent for (39) in [46] if for all $x \in X$

$$i \preceq j \wedge h(x, \xi^j) \leq 0 \Rightarrow h(x, \xi^i) \leq 0\,.$$

It follows as in [46] that the constraints

$$z_i \leq z_j \quad \text{for all } i, j \in \{1, \ldots, n\} \text{ such that } i \preceq j$$

are valid inequalities if $\preceq$ is a strongly consistent order for (39).

If the function h is of the special form $h(x, \xi) = \xi - T(\xi)x$ with a linear (s, m)-matrix function $T(\cdot)$, a strongly consistent order is $i \preceq j \Leftrightarrow \xi^i - T(\xi^i)x \leq \xi^j - T(\xi^j)x$, for all $x \in X$, where $\leq$ is the component-wise inequality between elements of $\mathbb{R}^s$. For the special function $h(x, \xi) = \xi - Tx$ and fixed weights $w_i, i = 1, \ldots, n$, problem (39) is a mixed-integer linear semi-infinite program of the form

$$\min_{\substack{t \geq 0,\ (\xi^1, \ldots, \xi^n) \in \Xi^n \\ (z_1, \ldots, z_n) \in \{0,1\}^n}} \left\{ t \;\middle|\; \begin{array}{c} P(\mathcal{P}(x)) \leq t + \sum_{i=1}^n w_i (1 - z_i) \\ \sum_{i=1}^n w_i (1 - z_i) \leq t + P(\mathcal{P}(x)) \\ \xi^i - Tx - z_i M e \leq 0,\ i = 1, \ldots, n \\ z_i \leq z_j \text{ if } \xi^i \leq \xi^j,\ i, j = 1, \ldots, n \\ \forall x \in X \end{array} \right\}. \tag{40}$$

The papers [31,53,56] are sources for deriving further valid inequalities.

5 Newsvendor with random demand: an illustration

We consider the classical newsvendor problem to illustrate the approach to scenario generation developed in Sect. 2. We recall that a newsvendor must place a daily order for a number of copies x of a newspaper. He has to pay c monetary units for each copy and sells a copy at r units, where $0 < c < r$. The daily demand ξ is a real random variable with (discrete) probability distribution $P \in \mathcal{P}(\mathbb{N})$, $\Xi = \mathbb{R}_+$, and the remaining copies $y(\xi) = \max\{0, x - \xi\}$ have to be removed. The newsvendor wishes that the order x maximizes his expected profit or, equivalently, minimizes his expected costs, i.e.,

$$\mathbb{E}[f_0(x,\xi)] = \int_{\mathbb{R}} f_0(x,\xi) dP(\xi) = (c-r)x + r\int_{\mathbb{R}} \max\{0, x-\xi\} P(d\xi) \quad (x \in \mathbb{R}).$$

The model may be reformulated as a linear two-stage stochastic program with the optimal value function $\Phi(t) = \max\{0, -t\}$. Starting from

$$\Phi(t) = \inf\{\langle q, y\rangle : Wy = t,\ y \geq 0\} = \sup\{\langle t, z\rangle : W^\top z \leq q\}$$

with $W = (-1, 1)$, $q = (0, r)^\top$, $Y = \mathbb{R}_+$ and $h(x,\xi) = \xi - x$, we obtain $D = \{z \in \mathbb{R} : W^\top z \leq q\} = [0, r]$ and for $x \in X = \mathbb{R}_+$

$$\mathbb{E}[f_0(x,\xi)] = (c-r)x + r\int_0^x (x-\xi) dG_P(\xi) = (c-r)x + r\int_0^x G_P(\xi) d\xi\,. \quad (41)$$

The latter is obtained using integration by parts, where G_P denotes the distribution function $G_P(x) = P(\{\xi \in \mathbb{R} : \xi \leq x\}) = \sum_{k \leq x} p_k$ of P and p_k is the probability of demand $k \in \mathbb{N}$. The unique solution is the $\frac{r-c}{r}$-quantile of P.

The corresponding optimal scenario generation problem is of the form

$$\min_{\substack{t \geq 0, (\xi^1, \ldots, \xi^n) \in \mathbb{R}^n \\ (w_1, \ldots, w_n) \in \mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{c} \sum_{i=1}^n w_i (\xi^i - x) z_i \leq t + F_P(x) \\ F_P(x) \leq t + r \sum_{i=1}^n w_i y_i \\ \forall (x, y, z) \in \mathbb{R}_+ \times \mathbb{R}_+^n \times \mathbb{R}^n : \\ y_i + x \geq \xi^i,\ 0 \leq z_i \leq r,\ i = 1, \ldots, n \end{array} \right\}, \quad (42)$$

where F_P is the convex expected recourse function

$$F_P(x) = r \sum_{k=1}^{\infty} p_k \max\{0, x - k\}\,. \quad (43)$$

We note that Theorems 2 and 3 apply to (42) if the weights w_i are fixed.

By incorporating F_P from (43), (42) is equivalent with the best approximation problem

$$\min_{\substack{(\xi^1,\ldots,\xi^n)\in\mathbb{R}^n\\(w_1,\ldots,w_n)\in\mathcal{S}_n}} \sup_{x\in\mathbb{R}_+} \left| \sum_{k=1}^{\infty} p_k \max\{0, x-k\} - \sum_{i=1}^{n} w_i \max\{0, x-\xi^i\} \right|. \tag{44}$$

We assume that the support of P is bounded, i.e., that the series representation in (43) is not infinite. Let $N \in \mathbb{N}$ be such that $p_k = 0$ for all $k > N$. Then F_P is piecewise linear convex on $\mathbb{R}_+$ with possible kinks at any $k \in \mathbb{N}$, $k \leq N$. The slope of F_P at k is $r \sum_{i=1}^{k} p_i$ and it holds $F_P(x) = r(x - \mathbb{E}[\xi])$ for $x \geq N$ where $\mathbb{E}[\xi]$ is the mean value of ξ, i.e., $\mathbb{E}[\xi] = \sum_{k=1}^{N} p_k k$.

Using the transformation idea from Proposition 2 (see [48]) we are able to transfer the generalized semi-infinite program (42) into a semi-infinite one. To this end we define the mapping

$$t : \Xi \times \mathcal{U} \to \mathbb{R}, \quad t(\xi, u) = u + \xi,$$

where $\mathcal{U} = \mathbb{R}_+ = \{u \in \mathbb{R}_+ : x + u \geq 0\}$ for each $x \in \mathbb{R}_+$. Then the transformation $y = t(\xi, u)$ leads to $t(\xi, \mathcal{U}) = \{y \in \mathbb{R}_+ : x + y \geq \xi\}$ for all $(x, \xi) \in \mathbb{R}_+^2$ and the optimization model (42) is of the form

$$\min_{\substack{t\geq 0,(\xi^1,\ldots,\xi^n)\in\mathbb{R}^n\\(w_1,\ldots,w_n)\in\mathcal{S}_n}} \left\{ t \;\middle|\; \begin{array}{l} \sum_{i=1}^{n} w_i(\xi^i - x) z_i \leq t + F_P(x) \\ F_P(x) \leq t + r \sum_{i=1}^{n} w_i(y_i + \xi^i) \\ \forall (x, y, z) \in \mathbb{R}_+ \times \mathbb{R}_+^n \times D^n \end{array} \right\}, \tag{45}$$

where u is replaced again by y. If the support of P is contained in $[0, M]$ for some $M \in \mathbb{N}$, we can also replace both $\mathbb{R}_+$ in (45) by $[0, M]$ and arrive at a compact index set of the semi-infinite program. Hence, a solution of (45) by a discretization method (see "Appendix") is possible.

6 Conclusions

The generation of scenarios is an important issue for solving applied stochastic programming models. Presently Monte Carlo sampling methods are the preferred approach (see [24]), but besides Quasi-Monte Carlo and sparse grid methods also best approximation methods are in use. The latter utilize metric distances of probability measures and suggest to determine discrete measures as best approximations to the underlying probability distribution (see [33,35]).

Existing scenario reduction methods [10,18] are based on the same theoretical background. However, we pointed out in Sect. 1 that stability results indicate that such probability metrics only lead to coarse estimates of distances of optimal values and solutions. Decisions on scenario generation and reduction based on such estimates

appear somewhat questionable and should at least be further examined. This is supported by slow convergence rates in terms of such probability metrics. But a stability result like Proposition 1 also suggests to make use of the minimal information distance $d_{\mathcal{F}}$ [see (6), (5)] as a basis for best approximation methods. This observation served as the guideline for the present paper. It turned out that at least for linear two-stage models the best approximation problem for scenario generation has favorable properties. It represents a best uniform approximation problem for the expected recourse function by a convex combination of polyhedral functions generated by scenarios. The latter can be rewritten as generalized semi-infinite optimization model and in many cases transformed into a standard semi-infinite program. If either only right-hand sides or only costs are random the optimization model is convex. In any case there exists a well-developed theory and a number of solution algorithms for such models (see [17,22,40,50–52]). Scenario reduction problems for linear two-stage models can be decomposed into solving a combinatorial optimization problem and a linear semi-infinite program, where the first determines the remaining scenarios and the second their new probabilities.

The characterization of scenario generation with respect to the distance $d_{\mathcal{F}}$ as best approximation problem for the expected recourse function provides a link to bounding schemes for the expected recourse (see [26, Section 3.2.1]). It reveals the close relationship of scenario generation, scenario reduction and bounding.

The aim of the present paper consisted in showing that employing minimal information distances for scenario generation and reduction leads to interesting optimization models. Their solution should result in improved decisions for scenario generation and reduction at least for two-stage models. In a next step we are planning to confirm this by numerical experiments.

Acknowledgements The first author gratefully acknowledges support by the FMJH Program Gaspard Monge in optimization and operations research including support to this program by EDF. Both authors wish to thank the referees and the Guest Editor for their valuable comments.

Appendix

We consider semi-infinite programs of the form

$$P[V] \qquad \min\{g_0(u) : u \in U,\ g_j(u, v) \le 0,\ j = 1, \ldots, p,\ \forall v \in V\},$$

where $U \subset \mathbb{R}^m$ is closed, $V \subset \mathbb{R}^k$ is compact and the functions $g_0 : U \to \mathbb{R}$, $g_j : U \times V \to \mathbb{R}$, $j = 1, \ldots, p$, are continuous. Let V_i, $i \in \mathbb{N}_0$, be an increasing sequence of finite subsets of V such that $\lim_{i\to\infty} \sup_{v\in V} \min_{v_i\in V_i} \|v - v_i\| = 0$.

Discretization algorithm:
Step 0: Set $i = 0$, $D_0 = V_0$.
Step 1: Find a solution u_i of $P[D_i]$.
Step 2: Find a solution v_i of $\max_{v\in V_{i+1}} \max_{j=1,\ldots,p} g_j(u_i, v)$.

Step 3: If $\gamma_i = \max_{j=1,\ldots,p} g_j(u_i, v_i) > 0$, then select a set D_{i+1} such that

$$D_i \cup \{v_i\} \subseteq D_{i+1} \subseteq V_{i+1}.$$

Step 4: If $\gamma_i \leq 0$ then stop.
Step 5: Set $i = i + 1$ and go to Step 1.

If the feasible set $F[V]$ of $P[V]$ is nonempty and the level set $\{u \in F[V_0] : g_0(u) \leq g_0(u_0)\}$ is bounded for some $u_0 \in F[V]$, the infima of $P[D_i]$ converge to the infimum of $P[V]$ and the sequence (u_i) has an accumulation point which solves $P[V]$. For a proof of this result we refer to [39, Theorem 2.1] and for further information and discussion to [40].

References

1. Arya, V., Garg, N., Khandekar, R., Meyerson, A., Munagala, K., Pandit, V.: Local search heuristics for k-median and facility location problems. SIAM J. Comput. **33**, 544–562 (2004)
2. Birge, J.R., Wets, R.J.-B.: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Math. Program. Study **27**, 54–102 (1986)
3. Dick, J., Kuo, F.Y., Sloan, I.H.: High-dimensional integration—the Quasi-Monte Carlo way. Acta Numerica **22**, 133–288 (2013)
4. Dohan, D., Karp, S., Matejek, B.: K-median algorithms: theory in practice. Working paper, Princeton, Computer Science (2015)
5. Dokov, S.P., Morton, D.P.: Second-order lower bounds on the expectation of a convex function. Math. Oper. Res. **30**, 662–677 (2005)
6. Dudley, R.M.: The speed of mean Glivenko–Cantelli convergence. Ann. Math. Stat. **40**, 40–50 (1969)
7. Dudley, R.M.: Real Analysis and Probability. Chapman & Hall, New York (1989)
8. Dupačová, J.: Scenario based stochastic programs: Strategies for deleting scenarios, IIASA Working Paper WP-95-014 (1995)
9. Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. Ann. Oper. Res. **100**, 25–53 (2000)
10. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: an approach using probability metrics. Math. Program. **95**, 493–511 (2003)
11. Edirisinghe, N.C.P.: New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. Oper. Res. **44**, 909–922 (1996)
12. Edirisinghe, N.C.P., Ziemba, W.T.: Bounding the expectation of a saddle function with application to stochastic programming. Math. Oper. Res. **19**, 314–340 (1994)
13. Feng, Y., Ryan, S.M.: Scenario construction and reduction applied to stochastic power generation planning. Comput. Oper. Res. **40**, 9–23 (2013)
14. Frauendorfer, K.: Solving SLP recourse problems with arbitrary multivariate distributions—the dependent case. Math. Oper. Res. **13**, 377–394 (1988)
15. Goberna, M.A., López, M.A.: Linear Semi-infinite Optimization. Wiley, Chichester (1998)
16. Graf, S., Luschgy, H.: Foundations of quantization for probability distributions. Lecture Notes in Mathematics, vol. 1730. Springer, Berlin (2000)
17. Guerra Vázquez, F., Rückmann, J.-J., Stein, O., Still, G.: Generalized semi-infinite programming: a tutorial. J. Comput. Appl. Math. **217**, 394–419 (2008)
18. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comput. Optim. Appl. **24**, 187–206 (2003)
19. Heitsch, H., Römisch, W.: A note on scenario reduction for two-stage stochastic programs. Oper. Res. Lett. **35**, 731–738 (2007)
20. Henrion, R., Küchler, C., Römisch, W.: Scenario reduction in stochastic programming with respect to discrepancy distances. Comput. Optim. Appl. **43**, 67–93 (2009)
21. Henrion, R., Küchler, C., Römisch, W.: Discrepancy distances and scenario reduction in two-stage stochastic integer programming. J. Ind. Manag. Optim. **4**, 363–384 (2008)

22. Hettich, R., Kortanek, K.O.: Semi-infinite programming: theory, methods, and applications. SIAM Rev. **35**, 380–429 (1993)
23. Hettich, R., Zencke, P.: Numerische Methoden der Approximation und Semi-Infiniten Optimierung. Teubner, Stuttgart (1982)
24. Homem-de-Mello, T., Bayraksan, G.: Monte Carlo sampling-based methods for stochastic optimization. Surv. Oper. Res. Manag. Sci. **19**, 56–85 (2014)
25. Kall, P.: Stochastic programming with recourse: upper bounds and moment problems. In: Guddat, J. (ed.) Advances in Mathematical Optimization, pp. 86–103. Akademie-Verlag, Berlin (1988)
26. Kall, P., Mayer, J.: Stochastic Linear Programming, 2nd edn. Springer, New York (2010)
27. Kariv, O., Hakimi, S.L.: An algorithmic approach to network location problems, II: the p-medians. SIAM J. Appl. Math. **37**, 539–560 (1979)
28. Klatte, D.: A note on quantitative stability results in nonlinear optimization. In: Lommatzsch, K. (ed.) Proceedings of the 19. Jahrestagung Mathematical Optimization. Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 90, pp. 77–86 (1987)
29. Leövey, H., Römisch, W.: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems. Math. Program. **151**, 315–345 (2015)
30. Li, S., Svensson, O.: Approximating k-median via pseudo-approximation. SIAM J. Comput. **45**, 530–547 (2016)
31. Luedtke, J., Ahmed, S., Nemhauser, G.L.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **122**, 247–272 (2010)
32. Morales, J.M., Pineda, S., Conejo, A.J., Carrion, M.: Scenario reduction for futures market trading in electricity markets. IEEE Trans. Power Syst. **24**, 878–888 (2009)
33. Pagès, G.: A space vector quantization method for numerical integration. J. Comput. Appl. Math. **89**, 1–38 (1997)
34. Pennanen, T., Koivu, M.: Epi-convergent discretizations of stochastic programs via integration quadratures. Numer. Math. **100**, 141–163 (2005)
35. Pflug, G.C., Pichler, A.: Approximations for probability distributions and stochastic optimization problems. In: Bertocchi, M.I., Consigli, G., Dempster, M.A.H. (eds.) Stochastic Optimization Methods in Finance and Energy, pp. 343–387. Springer, New York (2011)
36. Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, Hoboken (1991)
37. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. Math. Oper. Res. **27**, 792–818 (2002)
38. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems, vol. I. Springer, Berlin (1998)
39. Reemtsen, R.: Discretization methods for the solution of semi-infinite programming problems. J. Optim. Theory Appl. **71**, 85–103 (1991)
40. Reemtsen, R., Görner, S.: Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.-J. (eds.) Semi-Infinite Programming, pp. 195–275. Kluwer, Boston (1998)
41. Robinson, S.M.: Regularity and stability for convex multivalued functions. Math. Oper. Res. **1**, 130–143 (1976)
42. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
43. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.)Stochastic Programming. Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam, pp. 483–554 (2003)
44. Römisch, W.: Scenario generation. In: Cochran, J.J. (ed.) Wiley Encyclopedia of Operations Research and Management Science. Wiley, New York (2010)
45. Römisch, W., Wets, R.J.-B.: Stability of ε-approximate solutions to convex stochastic programs. SIAM J. Optim. **18**, 961–979 (2007)
46. Ruszczyński, A.: Probabilistic programming with discrete distributions and precedence constraints knapsack polyhedra. Math. Program. **93**, 195–215 (2002)
47. Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
48. Schwientek, J., Seidel, T., Küfer, K.-H.: A transformation-based discretization method for solving general semi-infinite optimization problems. Optim. Online Dig. (2017)
49. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming, MPS-SIAM Series on Optimization, Second Edition, Philadelphia (2014)
50. Stein, O.: Bi-level Strategies in Semi-infinite Programming. Kluwer, Boston (2003)

51. Stein, O., Still, G.: On generalized semi-infinite optimization and bilevel optimization. Eur. J. Oper. Res. **142**, 444–462 (2002)
52. Still, G.: Generalized semi-infinite programming: numerical aspects. Optimization **49**, 223–242 (2001)
53. Vielma, J.P., Ahmed, S., Nemhauser, G.L.: Mixed-integer linear programming formulations for probabilistic constraints. Oper. Res. Lett. **40**, 153–158 (2012)
54. Walkup, D., Wets, R.J.-B.: Lifting projections of convex polyhedra. Pac. J. Math. **28**, 465–475 (1969)
55. Wets, R.J.-B.: Stochastic programs with fixed recourse: the equivalent deterministic program. SIAM Rev. **16**, 309–339 (1974)
56. Zhao, M., Huang, K., Zeng, B.: A polyhedral study on chance constrained programs with random right-hand side. Math. Program. **166**, 19–64 (2017)