Recent progress in two-stage mixed-integer stochastic programming with applications to power production planning

Werner Römisch and Stefan Vigerske

Abstract We present recent developments in two-stage mixed-integer stochastic programming with regard to application in power production planning. In particular, we review structural properties, stability issues, scenario reduction and decomposition algorithms for two-stage models. Furthermore, we describe an application to stochastic thermal unit commitment.

1 Introduction

Since its beginnings in the late 1980s mixed-integer stochastic programming has undergone a considerable development both in theory and computations. We refer to the excellent overviews in [LoS03, Sch03, Se05] and to the very comprehensive bibliography [vdV07].

The aim of this paper is to look at some of the more recent developments that bear further potential for applications to power systems modeling and optimization. First, we mention new results on structures and convex approximations [KSV06, vdV04], on estimating and approximating the underlying probability distribution [ER07, RV08], and as a consequence of the latter on scenario reduction in two-stage mixed-integer stochastic programs [HKR08, HKR09]. Another line of work deals with the consequences of replacing the (traditional) expectation functional in the objective by risk functionals on structural properties and algorithms (see [ER05, ST06] for example). Much work was directed to algorithmic issues and, in particular, to decomposition schemes [AEO03, ATS00, CS99, DR04, EG+07, LuS04, NS08b, SSV98, SeH05, SeS06], where much is due to the pioneering work of S. Sen and his co-workers.

Werner Römisch
Humboldt University, 10099 Berlin, Germany, e-mail: `romisch@math.hu-berlin.de`

Stefan Vigerske
Humboldt University, 10099 Berlin, Germany, e-mail: `vigerske@math.hu-berlin.de`

In the following, we review some of the recent work. We start with a review of structural properties, discuss stability issues, methods for scenario reduction, and decomposition algorithms. As an illustration, we finally discuss an application to the stochastic unit commitment problem in power production planning.

2 Models and structural properties

Stochastic programs with mixed-integer recourse arise as deterministic equivalents of linear programs containing a random parameter vector ξ (varying in Ξ) and being of the form

$$\min\{\langle c,x\rangle \mid x\in X,\, T(\xi)x\geq h(\xi)\},$$

where X is a closed subset of $\mathbb{R}^m$, $c\in\mathbb{R}^m$, the (technology) matrix $T(\xi)$ and the vector $h(\xi)$ may depend on ξ. Given a realization of ξ, a possible violation of $h(\xi)-T(\xi)x\leq 0$ is compensated by the recourse cost $\langle q_1(\xi),y_1(\xi)\rangle+\langle q_2(\xi),y_2(\xi)\rangle$, where the pair $(y_1(\xi),y_2(\xi))$ with integral y_2 satisfies the constraint $W_1y_1+W_2y_2\leq h(\xi)-T(\xi)x$. Here, the cost coefficients $q_1(\xi)$ and $q_2(\xi)$ may depend on ξ. The modelling idea consists in adding the expected recourse cost $\mathbb{E}(\langle q_1(\xi),y_2(\xi)\rangle+\langle q_2(\xi),y_2(\xi)\rangle)$ to the original cost $\langle c,x\rangle$ and in minimizing the total cost with respect to (y_1,y_2). This leads to the stochastic program with mixed-integer recourse

$$\min\left\{\int_\Xi f_0(x,\xi)dP(\xi)\,\middle|\, x\in X\right\}, \tag{1}$$

where the function f_0 is given by

$$f_0(\xi,x):=\langle c,x\rangle+\Phi(q(\xi),h(\xi)-T(\xi)x)\quad ((x,\xi)\in\mathbb{R}^m\times\Xi), \tag{2}$$

Φ is the infimum function of a mixed-integer linear program

$$\Phi(u,t):=\inf\{\langle u_1,y_1\rangle+\langle u_2,y_2\rangle \mid y_1\in\mathbb{R}^{m_1},y_2\in\mathbb{Z}^{m_2},W_1y_1+W_2y_2\leq t\} \tag{3}$$

for all pairs $(u,t)\in\mathbb{R}^{m_1+m_2}\times\mathbb{R}^r$, Ξ is a polyhedron in $\mathbb{R}^s$, W_1 and W_2 are (r,m_1)- and (r,m_2)-matrices, respectively, $q(\xi)\in\mathbb{R}^{m_1+m_2}$, $h(\xi)\in\mathbb{R}^r$, and the (r,m)-matrix $T(\xi)$ are affine functions of $\xi\in\mathbb{R}^s$, and P is a probability distribution on the set Ξ (shortly $P\in\mathscr{P}(\Xi)$). Since the decisions x and $y(\xi)$ are made before and after the realization of ξ, they are called first and second stage decisions, respectively.

The following conditions are imposed to have the model (1) well-defined:
(C1) The matrices W_1 and W_2 have only rational elements.
(C2) For each pair $(x,\xi)\in X\times\Xi$ it holds that $h(\xi)-T(\xi)x\in\mathscr{T}$, where

$$\mathscr{T}:=\{t\in\mathbb{R}^r \mid \exists y=(y_1,y_2)\in\mathbb{R}^{m_1}\times\mathbb{Z}^{m_2} \text{ such that } W_1y_1+W_2y_2\leq t\}.$$

(C3) For each $\xi\in\Xi$ the recourse cost $q(\xi)$ belongs to the dual feasible set

$$\mathscr{U} := \left\{ u = (u_1, u_2) \in \mathbb{R}^{m_1+m_2} \;\middle|\; \exists z \in \mathbb{R}^r_- \text{ such that } W_1^\top z = u_1,\, W_2^\top z = u_2 \right\}.$$

(C4) $P \in \mathscr{P}_2(\Xi)$, i.e., $P \in \mathscr{P}(\Xi)$ and $\int_\Xi \|\xi\|^2 P(d\xi) < +\infty$.

Condition (C2) means that a feasible second stage decision always exists (*relatively complete recourse*). Both (C2) and (C3) imply $\Phi(u,t)$ to be finite for all $(u,t) \in \mathscr{U} \times \mathscr{T}$. Clearly, it holds $(0,0) \in \mathscr{U} \times \mathscr{T}$ and $\Phi(0,t) = 0$ for every $t \in \mathscr{T}$. With the convex polyhedral cone

$$\mathscr{K} := \{t \in \mathbb{R}^r \mid \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } t \geq W_1 y_1\} = W_1(\mathbb{R}^{m_1}) + \mathbb{R}^r_+$$

one obtains the representation

$$\mathscr{T} = \bigcup_{z \in \mathbb{Z}^{m_2}} (W_2 z + \mathscr{K}). \tag{4}$$

The two extremal cases are (i) W_1 has rank r implying $\mathscr{K} = \mathbb{R}^r = \mathscr{T}$ (*complete recourse*) and (ii) $W_1 = 0$ (*pure integer recourse*) leading to $\mathscr{K} = \mathbb{R}^r_+$.

In general, the set $\mathscr{T}$ is connected (i.e., there exists a polygon connecting two arbitrary points of $\mathscr{T}$) and condition (C1) implies that $\mathscr{T}$ is closed. If, for each $t \in \mathscr{T}$, $Z(t)$ denotes the set

$$Z(t) := \{z \in \mathbb{Z}^{m_2} \mid \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } W_1 y_1 + W_2 z \leq t\},$$

the representation (4) implies that $\mathscr{T}$ may be decomposed into subsets of the form

$$\mathscr{T}(t_0) := \{t \in \mathscr{T} \mid Z(t) = Z(t_0)\} = \bigcap_{z \in Z(t_0)} (W_2 z + \mathscr{K}) \setminus \Big(\bigcup_{z \in \mathbb{Z}^{m_2} \setminus Z(t_0)} (W_2 z + \mathscr{K}) \Big) \tag{5}$$

for every $t_0 \in \mathscr{T}$. In general, the set $Z(t_0)$ is finite or countable, but condition (C1) implies that $Z(t_0)$ in the intersection in (5) may be replaced by a single element of $\mathscr{T}$ and $\mathbb{Z}^{m_2} \setminus Z(t_0)$ in the union by a finite subset of $\mathbb{Z}^{m_2}$, respectively (see [BG+82, Lemmas 5.6.1 and 5.6.2]). Hence, if (C1) is satisfied, there exist countably many elements $t_i \in \mathscr{T}$ and $z_{ij} \in \mathbb{Z}^{m_2}$ for j belonging to a finite subset N_i of $\mathbb{N}$, $i \in \mathbb{N}$, such that

$$\mathscr{T} = \bigcup_{i \in \mathbb{N}} \mathscr{T}(t_i) \quad \text{with} \quad \mathscr{T}(t_i) = (t_i + \mathscr{K}) \setminus \bigcup_{j \in N_i} (W_2 z_{ij} + \mathscr{K}). \tag{6}$$

The sets $\mathscr{T}(t_i)$, $i \in \mathbb{N}$, are nonempty and connected (even star-shaped cf. [BG+82, Theorem 5.6.3]), but nonconvex in general (see the illustration in Figure 1). If for some $i \in \mathbb{N}$ the set $\mathscr{T}(t_i)$ is nonconvex, it can be decomposed into a finite number of subsets of $\mathscr{T}(t_i)$ whose closures are convex polyhedra with facets parallel to suitable facets of $W_1(\mathbb{R}^{m_1})$ or of $\mathbb{R}^r_+$ (see Figure 1). By renumbering all such subsets (for every $i \in \mathbb{N}$) one obtains countably many subsets B_j, $j \in \mathbb{N}$, of $\mathscr{T}$ which form a partition of $\mathscr{T}$. Since the sets $Z(t)$ of feasible integer decisions do not change if t varies in some B_j, the function $\Phi(u, \cdot)$ is Lipschitz continuous (with modulus not depending on j) on B_j for every $j \in \mathbb{N}$ and every fixed $u \in \mathscr{U}$.

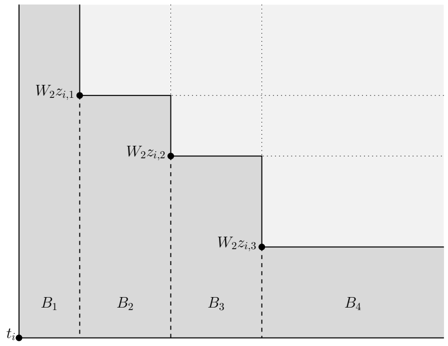


Fig. 1 Illustration of $\mathscr{T}(t_i)$ (see (6)) for $W_1 = 0$ and $r = 2$, i.e., $\mathscr{K} = \mathbb{R}^2_+$, with $N_i = \{1,2,3\}$ and its decomposition into the sets B_j, $j = 1,2,3,4$, whose closures are convex polyhedral (rectangular).

Now, let (C1)–(C3) be satisfied. Then the function Φ is lower semicontinuous and the function $(u,t) \mapsto \Phi(u,t)$ from $\mathscr{U} \times \mathscr{T}$ to $\mathbb{R}$ has the (convex) polyhedral continuity regions $\mathscr{U} \times B_j$, $j \in \mathbb{N}$. More precisely, the estimate

$$|\Phi(u,t) - \Phi(\tilde{u},\tilde{t})| \leq L(\max\{1,\|t\|,\|\tilde{t}\|\}\|u-\tilde{u}\| + \max\{1,\|u\|,\|\tilde{u}\|\}\|t-\tilde{t}\|) \quad (7)$$

holds for all pairs $(u,t),(\tilde{u},\tilde{t}) \in \mathscr{U} \times B_j$ and some constant $L > 0$. For proofs and further details the interested reader is referred to [BG+82, Chapter 5.6].

Next, we consider the integrand

$$f_0(x,\xi) = \langle c,x \rangle + \Phi(q(\xi),h(\xi) - T(\xi)x)$$

for all pairs $(x,\xi) \in X \times \Xi$ and study the continuity properties and growth behavior of $f_0(x,\cdot)$ on Ξ for fixed $x \in X$. The properties of Φ imply that, for every $x \in X$, there exists a partition $\{\Xi_{x,j}\}_{j\in\mathbb{N}}$ of Ξ given by

$$\Xi_{x,j} = \{\xi \in \Xi \mid h(\xi) - T(\xi)x \in B_j\} \qquad (j \in \mathbb{N}). \quad (8)$$

Furthermore, the function $f_0(x,\cdot)$ (on Ξ) satisfies the properties

$$|f_0(x,\xi) - f_0(x,\tilde{\xi})| \leq \hat{L}\max\{1,\|\xi\|,\|\tilde{\xi}\|\}\|\xi - \tilde{\xi}\| \quad (x \in X,\ \xi,\tilde{\xi} \in \Xi_{x,j}), \quad (9)$$

$$|f_0(x,\xi)| \leq C\max\{1,\|x\|\}\max\{1,\|\xi\|^2\} \quad (x \in X,\ \xi \in \Xi), \quad (10)$$

with some positive constants $\hat{L}$ and C. Due to (10), condition (C4) implies the existence of the integral in (1). We note that $f_0(x,\cdot)$ is globally Lipschitz continuous on $\Xi_{x,j}$ if the recourse cost $q(\xi)$ does not depend on ξ. It is even globally Lipschitz continuous on Ξ if only $q(\xi)$ depends on ξ. In both cases $|f_0(x,\cdot)|$ grows only linearly with $\|\xi\|$ and a finite first order moment of P, i.e., $P \in \mathscr{P}_1(\Xi)$ (instead of (C4)), implies the existence of the integral.

Since the objective function of (1) is lower semicontinuous if the conditions (C1)–(C4) are satisfied, solutions of (1) exist if X is closed and bounded. If the probability distribution P has a density, the objective function of (1) is continuous, but nonconvex in general. If the support of P is finite, the objective function is piecewise continuous with a finite number of polyhedral continuity regions. The latter is illustrated by Fig. 2, which shows the expected recourse function

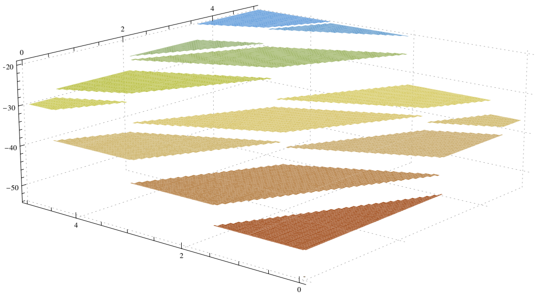
$$x \mapsto \int_{\Xi} \Phi(q, h(\xi) - Tx)\, dP(\xi) \qquad (x \in [0,5]^2)$$

with $r = s = 2$, $h(\xi) = \xi$, $m_1 = 0$, $W_1 = 0$, $m_2 = 4$, $q = (-16, -19, -23, -28)^\top$, the matrices

$$W_2 = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 6 & 1 & 3 & 2 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix},$$

and binary restrictions on the second stage variables as in [SSV98], but with a uniform probability distribution P having a smaller finite support than in [SSV98], namely, $\operatorname{supp}(P) = \{5, 10, 15\}^2$.

Fig. 2 Illustration of an expected recourse function with pure $0-1$ recourse, random right-hand side and discrete uniform probability distribution.



3 Stability

In this section, we review stability results for mixed-integer two-stage stochastic programs (1), i.e., results on the dependence of their solutions and optimal values on the underlying probability distribution P. Such results also provide information how an underlying probability distribution should be approximated such that approximate solutions and optimal values get close to the original ones.

In this context it is well known that the behavior of the (first stage) solution set

$$S(P) := \left\{ x \in X \;\middle|\; \int_{\Xi} f_0(x,\xi)P(d\xi) = v(P) \right\}$$

with respect to changes of P requires knowledge on the growth of the objective function

$$x \mapsto F_P(x) := \mathbb{E}(f_0(x,\xi)) = \int_{\Xi} f_0(x,\xi)P(d\xi)$$

near $S(P)$. Here, $v(P)$ denotes the infimum of the objective function or optimal value, i.e.,

$$v(P) := \inf\left\{ \int_{\Xi} f_0(x,\xi)P(d\xi) \;\middle|\; x \in X \right\}.$$

However, the growth behavior of F_P depends essentially on properties of the underlying probability distribution P. The situation is different for optimal values $v(P)$. Their behavior with respect to changes of P depends essentially on structural properties of the function f_0, which are well studied (cf. Section 2).

It is shown in [RV08] that the following distances of probability distributions are important for mixed-integer two-stage stochastic programs:

$$\zeta_{\ell,\mathscr{B}}(P,Q) := \sup\left\{ \left| \int_B f(\xi)P(d\xi) - \int_B f(\xi)Q(d\xi) \right| \;\middle|\; f \in \mathscr{F}_\ell(\Xi), B \in \mathscr{B} \right\}, \quad (11)$$

where $\ell \in \{1,2\}$ and $\mathscr{B}$ is a set of convex polyhedra, which contains the closures of $\Xi_{x,j}$, $j \in \mathbb{N}$, $x \in X$ (see (8)), and $\mathscr{F}_\ell(\Xi)$ contains all functions $f : \Xi \to \mathbb{R}$ such that

$$|f(\xi)| \le \max\{1, \|\xi\|^\ell\} \;\text{ and }\; |f(\xi) - f(\tilde{\xi})| \le \max\{1, \|\xi\|^{\ell-1}, \|\tilde{\xi}\|^{\ell-1}\}\|\xi - \tilde{\xi}\|$$

holds for all $\xi, \tilde{\xi} \in \Xi$. While the set $\mathscr{F}_\ell(\Xi)$ of functions has its origin in property (9) of the integrand f_0, but depends on the specific structure of the second stage program only with respect to $\ell \in \{1,2\}$, the class $\mathscr{B}$ of convex polyhedra strongly depends on that structure.

If the conditions (C1)–(C4) are satisfied and X is closed and bounded, there exists a constant $L > 0$ such that the estimate

$$|v(P) - v(Q)| \le L\varphi_P(\zeta_{\ell,\mathscr{B}}(P,Q)) \quad (12)$$

holds for every $Q \in \mathscr{P}_\ell(\Xi)$ with $\ell \in \{1,2\}$ and $\ell = 2$ if ξ enters $q(\xi)$ and, in addition, $h(\xi)$ or $T(\xi)$. Here, the function φ_P is defined by $\varphi_P(0) = 0$ and

$$\varphi_P(t) := \inf_{R\geq 1}\left\{R^{r+1}t + \int_{\{\xi\in\Xi\,|\,\|\xi\|>R\}} \|\xi\|^{\ell}P(d\xi)\right\} \qquad (t>0).$$

The function characterizes the tail behavior of P and is continuous at $t=0$. If P has a finite pth moment, i.e., if $\int_\Xi \|\xi\|^p P(d\xi) < +\infty$, for some $p>\ell$, the estimate

$$\varphi_P(t) \leq Ct^{\frac{p-\ell}{p+r-1}} \qquad (t\geq 0)$$

is valid for some constant $C>0$ and if Ξ is bounded, the estimate (12) simplifies to

$$|v(P)-v(Q)| \leq L\zeta_{\ell,\mathscr{B}}(P,Q).$$

If the set $\Xi\subset\mathbb{R}^s$ belongs to $\mathscr{B}$, we obtain from (11) by choosing $B:=\Xi$ and $f\equiv 1$, respectively,

$$\max\{\zeta_\ell(P,Q),\alpha_{\mathscr{B}}(P,Q)\} \leq \zeta_{\ell,\mathscr{B}}(P,Q) \tag{13}$$

for all $P,Q\in\mathscr{P}_\ell(\Xi)$. Here, ζ_ℓ and $\alpha_{\mathscr{B}}$ denote the ℓth order Fortet-Mourier metric (see [Ra91, Section 5.1]) and the polyhedral discrepancy

$$\zeta_\ell(P,Q) := \sup\left\{\left|\int_\Xi f(\xi)P(d\xi) - \int_\Xi f(\xi)Q(d\xi)\right| \,\middle|\, f\in\mathscr{F}_\ell(\Xi)\right\}, \tag{14}$$

$$\alpha_{\mathscr{B}}(P,Q) := \sup_{B\in\mathscr{B}}|P(B)-Q(B)|, \tag{15}$$

respectively. Hence, convergence of probability distributions with respect to $\zeta_{\ell,\mathscr{B}}$ implies their weak convergence, convergence of ℓth order absolute moments, and convergence with respect to the polyhedral discrepancy $\alpha_{\mathscr{B}}$. For bounded Ξ the technique in [Sch96, Proposition 3.1] can be employed to obtain

$$\zeta_{\ell,\mathscr{B}}(P,Q) \leq C_s\alpha_{\mathscr{B}}(P,Q)^{\frac{1}{s+1}} \qquad (P,Q\in\mathscr{P}(\Xi)) \tag{16}$$

for some constant $C_s>0$. In view of (13) and (16), the metric $\zeta_{\ell,\mathscr{B}}$ is stronger than $\alpha_{\mathscr{B}}$ in general, but in case of bounded Ξ both distances metrize the same convergence on $\mathscr{P}(\Xi)$.

For more specific models (1), improvements of the stability estimate (12) may be obtained by exploiting specific recourse structures, i.e., by using additional information on the shape of the sets B_j, $j\in\mathbb{N}$, and on the behavior of the function Φ on these sets. This may lead to stability estimates with respect to distances that are (much) weaker than $\zeta_{\ell,\mathscr{B}}$. For example, if $W_1=0$, Ξ is rectangular, T is fixed and some components of $h(\cdot)$ coincide with some of the components of ξ, the closures of $\Xi_{x,j}$, $x\in X$, $j\in\mathbb{N}$, are rectangular subsets of Ξ, i.e., belong to

$$\mathscr{B}_{\text{rect}} := \left\{I_1\times I_2\times\cdots\times I_s \mid \emptyset\neq I_j \text{ is a closed interval in } \mathbb{R},\ j=1,\ldots,s\right\} \tag{17}$$

and the stability estimate (12) is valid with respect to $\zeta_{\ell,\mathscr{B}_{\text{rect}}}$. As shown in [HKR09] convergence of a sequence of probability distributions with respect to $\zeta_{\ell,\mathscr{B}_{\text{rect}}}$ is equivalent to convergence with respect to both ζ_ℓ and $\alpha_{\mathscr{B}_{\text{rect}}}$. If, in addition to the previous assumptions, q is fixed and Ξ is bounded, the estimate (12) is valid with respect to the rectangular discrepancy $\alpha_{\mathscr{B}_{\text{rect}}}$ (see also [Sch96, Section 3]).

4 Scenario reduction

A well known approach for solving two-stage stochastic programs computationally consists in replacing the original probability distribution by a discrete distribution based on a finite number of scenarios. Let P be such a discrete distribution with scenarios ξ^i and probabilities p_i, $i = 1,\ldots,N$. The corresponding stochastic programming model is of the form

$$\min\left\{\langle c,x\rangle + \sum_{i=1}^{N} p_i(\langle q_1(\xi^i),y_1^i\rangle + \langle q_2(\xi^i),y_2^i\rangle)\;\middle|\; \begin{array}{l} W_1y_1^i + W_2y_2^i \le h(\xi^i) - T(\xi^i)x \\ y_1^i \in \mathbb{R}^{m_1}, y_2^i \in \mathbb{Z}^{m_2}, i = 1,\ldots,N, \\ x \in X \end{array}\right\}.$$

It may turn out that the computing times for solving the resulting mixed-integer linear programs are not acceptable. In such a case one might wish to reduce the number of scenarios entering the stochastic program. In [DGR03, HR07] a stability-based approach for scenario reduction in two-stage models without integrality requirements is developed. This approach suggests to look at stability results for optimal values and to use the corresponding distance of probability distributions for determining discrete distributions based on a smaller and prescribed number of scenarios as best approximations of P. According to the stability estimate (12) in Section 3, the distances $\zeta_{\ell,\mathscr{B}}$ or $\alpha_{\mathscr{B}}$ (if Ξ is bounded) appear as the right choice, where $\mathscr{B}$ is a set of convex polyhedra that depend on the structure of the stochastic program (1).

In [HKR08] the scenario reduction approach is elaborated for $\alpha_{\mathscr{B}}$ and a relevant set $\mathscr{B}$ of convex polyhedra. The numerical results show that the complexity of scenario reduction algorithms increases if $\mathscr{B}$ gets more involved. To avoid this effect, the distance $\zeta_{\ell,\mathscr{B}_{\text{rect}}}$ or, equivalently,

$$\mathbf{d}_\lambda(P,Q) := \lambda\,\alpha_{\mathscr{B}_{\text{rect}}}(P,Q) + (1-\lambda)\zeta_\ell(P,Q) \tag{18}$$

for some $\lambda \in (0,1)$ is considered in this section (and in [HKR09]).

Let Q_J denote a probability distribution whose support $\text{supp}(Q_J)$ contains the following subset of $\{\xi^1,\ldots,\xi^N\}$:

$$\text{supp}(Q_J) = \left\{\xi^i \,|\, i \in \{1,\ldots,N\}\setminus J\right\} \quad \text{and} \quad J \subset \{1,\ldots,N\}.$$

Let q_i ($i \notin J$) denote the probability of scenario ξ^i of Q_J. Now, the aim is to determine Q_J such that the distance $\mathbf{d}_\lambda(P,Q_J)$ is minimal, i.e., for arbitrary subsets J of $\{1,\ldots,N\}$, we are interested in

$$D_J := \min\left\{\mathbf{d}_\lambda(P,Q_J)\;\middle|\; q_i \ge 0,\, i \notin J,\, \sum_{i\notin J} q_i = 1\right\}. \tag{19}$$

In the following, we show that D_J can be computed as optimal value of a linear program. To this end, we assume without loss of generality that $J = \{n+1,\ldots,N\}$, i.e., $\mathrm{supp}(Q_J) = \{\xi^1,\ldots,\xi^n\}$ for some $1 \le n < N$. We consider the system of index sets

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

and obtain the following representation of the rectangular discrepancy

$$\alpha_{\mathscr{B}_{\mathrm{rect}}}(P,Q_J) = \sup_{B\in\mathscr{B}_{\mathrm{rect}}} |P(B)-Q_J(B)| = \max_{I\in\mathscr{I}_{\mathscr{B}_{\mathrm{rect}}}} \left| \sum_{i\in I} p_i - \sum_{j\in I\cap\{1,\ldots,n\}} q_j \right| \tag{20}$$

$$= \min\left\{ t_\alpha \,\middle|\, \begin{array}{l} -\sum_{j\in I\cap\{1,\ldots,n\}} q_j \le t_\alpha - \sum_{i\in I} p_i,\ I \in \mathscr{I}_{\mathscr{B}_{\mathrm{rect}}} \\ \sum_{j\in I\cap\{1,\ldots,n\}} q_j \le t_\alpha + \sum_{i\in I} p_i,\ I \in \mathscr{I}_{\mathscr{B}_{\mathrm{rect}}} \end{array} \right\} \tag{21}$$

Since the set $\mathscr{I}_{\mathscr{B}_{\mathrm{rect}}}$ may be too large to solve the linear program (21) numerically, we consider the system of reduced index sets

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

and the quantities

$$\gamma^{I^*} := \max\left\{ \sum_{i\in I} p_i \,\middle|\, I \in \mathscr{I}_{\mathscr{B}_{\mathrm{rect}}}, I\cap\{1,\ldots,n\} = I^* \right\}$$

$$\gamma_{I^*} := \min\left\{ \sum_{i\in I} p_i \,\middle|\, I \in \mathscr{I}_{\mathscr{B}_{\mathrm{rect}}}, I\cap\{1,\ldots,n\} = I^* \right\}$$

for every $I^* \in \mathscr{I}^*_{\mathscr{B}_{\mathrm{rect}}}$. Since any such index set I^* corresponds to some left-hand side of the inequalities in (21), γ^{I^*} and γ_{I^*} correspond to the smallest right-hand sides in (21). Hence, the rectangular discrepancy may be rewritten as

$$\alpha_{\mathscr{B}_{\mathrm{rect}}}(P,Q_J) = \min\left\{ t_\alpha \,\middle|\, \begin{array}{l} -\sum_{j\in I^*} q_j \le t_\alpha - \gamma^{I^*},\ I^* \in \mathscr{I}^*_{\mathscr{B}_{\mathrm{rect}}} \\ \sum_{j\in I^*} q_j \le t_\alpha + \gamma_{I^*},\ I^* \in \mathscr{I}^*_{\mathscr{B}_{\mathrm{rect}}} \end{array} \right\}. \tag{22}$$

Since the number of elements of $\mathscr{I}^*_{\mathscr{B}_{\mathrm{rect}}}$ is at most 2^n (compared to 2^N in $\mathscr{I}_{\mathscr{B}_{\mathrm{rect}}}$), passing from (21) to (22) indeed drastically reduces the maximum number of inequalities and may make the linear program (22) numerically tractable.

Due to duality arguments, the Fortet-Mourier distance $\zeta_{ell}(P,Q_J)$ (see (14)) allows the representation as linear program (cf. [HR07])

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

where $c_\ell(\xi,\tilde{\xi}) := \max\{1, \|\xi\|^{\ell-1}, \|\tilde{\xi}\|^{\ell-1}\}\|\xi-\tilde{\xi}\|$ for all $\xi,\tilde{\xi} \in \Xi = \{\xi^1,\ldots,\xi^N\}$ and $\hat{c}_\ell$ denotes the reduced costs

$$\hat{c}_\ell(\xi,\tilde{\xi}) := \inf\left\{ \sum_{k=1}^{K} c_\ell(\xi^{i_{k-1}},\xi^{i_k}) \;\middle|\; K\in\mathbb{N}, i_k\in\{1,\ldots,N\}, \xi^{i_0}=\xi, \xi^{i_K}=\tilde{\xi} \right\}.$$

Hence, extending the representation (22) of $\alpha_{\mathscr{B}_{\text{rect}}}$ we obtain the following linear program for determining D_J and the probabilities q_j, $j=1,\ldots,n$, of the discrete reduced distribution Q_J,

$$D_J = \min\left\{ \lambda\, t_\alpha + (1-\lambda)t_\zeta \;\middle|\; \begin{array}{l} t_\alpha, t_\zeta \geq 0,\, q_j \geq 0,\, \sum_{j=1}^n q_j = 1, \\ \eta_{ij} \geq 0, i=1,\ldots,N,\, j=1,\ldots,n, \\ t_\zeta \geq \sum_{i=1}^N \sum_{j=1}^n \hat{c}_\ell(\xi^i,\xi^j)\eta_{ij}, \\ \sum_{j=1}^n \eta_{ij} = p_i,\, i=1,\ldots,N, \\ \sum_{i=1}^N \eta_{ij} = q_j,\, j=1,\ldots,n, \\ -\sum_{j\in I^*} q_j \leq t_\alpha - \overline{\gamma}^{I^*},\, I^*\in\mathscr{I}^*_{\mathscr{B}_{\text{rect}}} \\ \quad \sum_{j\in I^*} q_j \leq t_\alpha + \underline{\gamma}_{I^*},\, I^*\in\mathscr{I}^*_{\mathscr{B}_{\text{rect}}} \end{array} \right\} \tag{23}$$

While the linear program (23) can be solved efficiently by available software, the determination of the index set $\mathscr{I}^*_{\mathscr{B}_{\text{rect}}}$ and the coefficients $\overline{\gamma}^{I^*}$, $\underline{\gamma}_{I^*}$ is more intricate.

It is shown in [HKR08, Section 3] that the parameters $\mathscr{I}^*_{\mathscr{B}_{\text{rect}}}$ and $\overline{\gamma}^{I^*}$, $\underline{\gamma}_{I^*}$ can be determined by studying the set $\mathscr{R}$ of supporting rectangles. A rectangle B in $\mathscr{B}_{\text{rect}}$ is called *supporting* if each of its facets contains an element of $\{\xi_1,\ldots,\xi_n\}$ in its relative interior (see also Fig. 3). Based on $\mathscr{R}$ the following representations are valid according to [HKR08, Prop. 1 and 2]:

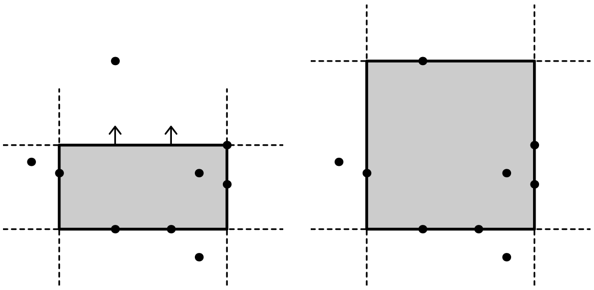


Fig. 3 Non-supporting rectangle (left) and supporting rectangle (right). The dots represent the remaining scenarios $\xi^1,\ldots,\xi^n$ for $s=2$.

$$\begin{aligned} \mathscr{I}^*_{\mathscr{B}_{\text{rect}}} &= \bigcup_{B\in\mathscr{R}} \left\{ I^*\subseteq\{1,\ldots,n\} \;\middle|\; \cup_{j\in I^*}\{\xi^j\} = \{\xi^1,\ldots,\xi^n\}\cap \operatorname{int} B \right\} \\ \overline{\gamma}^{I^*} &= \max\left\{ P(\operatorname{int} B) \;\middle|\; B\in\mathscr{R}, \cup_{j\in I^*}\{\xi^j\} = \{\xi^1,\ldots,\xi^n\}\cap\operatorname{int} B \right\} \\ \underline{\gamma}_{I^*} &= \sum_{i\in \underline{I}} p_i \text{ where } \underline{I} := \left\{ i\in\{1,\ldots,N\} \;\middle|\; \min_{j\in I^*}\xi^j_l \leq \xi^i_l \leq \max_{j\in I^*}\xi^j_l,\, l=1,\ldots,s \right\} \end{aligned}$$

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

An algorithm is developed in [HKR09] that constructs recursively l-dimensional supporting rectangles for $l = 1, \ldots, s$. Computational experiments show that its running time grows linearly with N, but depends on n and s via the expression $\binom{n+1}{2}^s$. Hence, while N may be large, only moderately sized values of n given s are realistic.

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

$$\min\{D_J \mid J \subset \{1, \ldots, N\}, \#J = n\} \tag{24}$$

which is known as n-median problem and as NP-hard. One possibility is to reformulate (24) as mixed-integer linear program and to solve it by standard software. Since, however, approximate solutions of (24) are sufficient, heuristic algorithms like *forward selection* are of interest, where u_k is determined in its kth step such that it solves the minimization problem

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

where $J^{[0]} = \{1, \ldots, N\}$, $J^{[k]} := J^{[k-1]} \setminus \{u_k\}$ $(k = 1, \ldots, n)$ and $J^{[n]} := \{1, \ldots, N\} \setminus \{u_1, \ldots, u_n\}$ serves as approximate solution to (24). Recalling that the complexity of

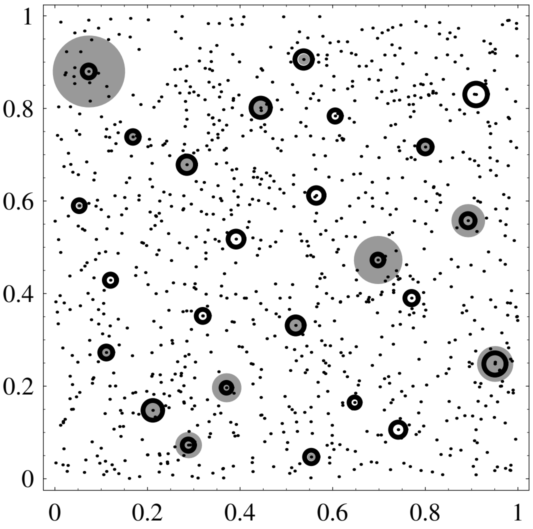


Fig. 4 $N = 1000$ samples ξ^i from the uniform distribution on $[0,1]^2$ and $n = 25$ points ξ^{u_k}, $k = 1, \ldots, n$, obtained via the first 25 elements z_k, $k = 1, \ldots, n$, of the *Halton sequence* (in the bases 2 and 3) (see [Ni92, p. 29]). The probabilities q_k of ξ^{u_k}, $k = 1, \ldots, n$, are computed for the distance $\mathbf{d}_\lambda$ with $\lambda = 1$ (gray balls) and $\lambda = 0.9$ (black circles) by solving (23). The diameters of the circles are proportional to the probabilities q_k, $k = 1, \ldots, 25$.

evaluating $D_{J^{[k-1]}\setminus\{u\}}$ for some $u \in J^{[k-1]}$ is proportional to $\binom{k+1}{2}^s$ shows that even the forward selection algorithm is expensive.

Hence, heuristics for solving (24) become important that require only a low number of D_J evaluations. For example, if P is a probability distribution on $[0,1]^s$ with independent marginal distributions P_j, $j = 1,\dots,s$, such a heuristic can be based on Quasi-Monte Carlo methods (cf. [Ni92]). The latter provide sequences of equidistributed points in $[0,1]^s$ that approximate the uniform distribution on the unit cube $[0,1]^s$. Now, let n Quasi-Monte Carlo points $z^k = (z_1^k,\dots,z_s^k) \in [0,1]^s$, $k = 1,\dots,n$, be given. Then we determine

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

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

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

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

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

and set again $J^{[k]} := J^{[k-1]} \setminus \{u_k\}$ for $k = 1,\dots,n$, where $J^{[0]} = \{1,\dots,N\}$. Figure 4 illustrates the results of such a Quasi-Monte Carlo based heuristic and Figure 5 shows the discrepancy $\alpha_{\mathscr{B}_{\text{rect}}}$ for different n and the running times of the Quasi-Monte Carlo based heuristic.

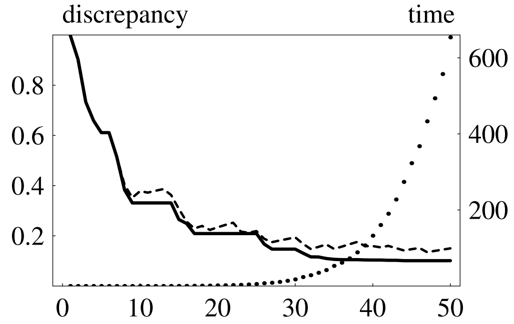


Fig. 5 Distance $\alpha_{\mathscr{B}_{\text{rect}}}$ between P (with $N = 1000$) and equidistributed QMC-points (dashed), QMC-points, whose probabilities are adjusted according to (23) (bold), and running times of the QMC-based heuristic (in seconds).

5 Decomposition algorithms

When the size of an optimization problem becomes intractable for standard solution approaches, a decomposition into small tractable subproblems by relaxing certain coupling constraints is often a possible resort. The task of the decomposition algorithm is then to coordinate the search in the subproblems in a way that their solutions can be combined into one that is feasible for the overall problem and has a "good" objective function value. Often, the algorithm also provides a certified lower bound on the optimal value which allows to evaluate the quality of a found solution. Since, on the one hand, mixed-integer stochastic programs easily reach a size that is intractable for standard solution approaches, but, on the other hand, are also very structured, many decomposition algorithms have been developed [LoS03, Sch03, Se05]. In the following, we discuss some of them in more detail.

Let us assume that the set of first stage feasible solutions X is given in the form

$$X = \{x \in \mathbb{Z}^{m_0} \times \mathbb{R}^{m-m_0} \mid Ax \leq b\},$$

where m_0 denotes the number of first stage variables with integrality restrictions, A is a (r_0, m)-matrix, and $b \in \mathbb{R}^{r_0}$. Further, we denote by

$$\bar{X} := \{x \in \mathbb{R}^m \mid Ax \leq b\}$$

the *linear relaxation* of X. We recall the *value function* (3),

$$\Phi(u,t) = \inf\{\langle u_1, y_1\rangle + \langle u_2, y_2\rangle \mid y_1 \in \mathbb{R}^{m_1}, y_2 \in \mathbb{Z}^{m_2}, W_1 y_1 + W_2 y_2 \leq t\},$$

and define the *expected recourse function* of model (1) by

$$\Psi(x) := \int_{\Xi} \Phi(q(\xi), h(\xi) - T(\xi)x) P(d\xi) \qquad (x \in \bar{X}).$$

For continuous ($m_2 = 0$) stochastic programs, the Benders decomposition is an established method [VSW69, Bi85]. It decomposes the decision on the first stage from the recourse decisions on the second stage by replacing the value function $\Phi(u,t)$ in (1) by an explicit approximation based on supporting hyperplanes. Unfortunately, Benders Decomposition relies heavily on the convexity of the value function $t \mapsto \Phi(u,t)$. Thus, in the view of Section 2, it cannot be directly applied to the case where discrete variables are present.

However, there are several approaches to overcome this difficulty. One of the first is the Integer L-shaped method [LaL93], which assumes that the first stage problem involves only binary variables. This property is exploited to derive linear inequalities that approximate the value function $\Phi(u,t)$ pointwise. While the algorithm makes only moderate assumptions on the second stage problem, its main drawback is the weak approximation of the value function due to lacking first order information about the value function. Thus, the algorithm might enumerate all feasible first stage solutions in order to find an optimal solution.

A cutting-plane algorithm is proposed in [CT97]. Here, the deterministic equivalent of (1) is solved by improving its linear relaxation with lift-and-project cuts. Decomposition arises here in two ways. First, the linear relaxation (including additional cuts) is solved by Benders Decomposition. Second, lift-and-project cuts are derived scenariowise. Further, in case of a fixed technology matrix $T(\cdot) \equiv T$, cut coefficients that have been computed for one scenario can also be reused to derive cuts for other scenarios. This algorithm can be seen as a predecessor of the dual decomposition approach presented in [SeH05]. While the cuts in [CT97] include variables from both stages, [SeH05] extends the Benders Decomposition approach to the mixed-integer case by sequentially convexifying the value function $\Phi(u,t)$. It is discussed in detail in Section 5.2.

In [KSV06] it is observed, that even though the value function $\Phi(u,t)$ might be nonconvex and difficult to handle, under some assumptions on the distribution of ξ, the expected recourse function $\Psi(x)$ can be convex. Starting with simple integer recourse models and then extending to more general classes of problems, techniques to compute tight convex approximations of the expected recourse function by perturbing the distribution of ξ are developed in a series of papers [KSV06, vdV04, vdV05]. We sketch this approach in more detail in Section 5.1.

In the case that the second stage problem is purely integer ($m_1 = 0$), the value function $\Phi(u,t)$ has the nice property to be constant on polyhedral subsets of $\mathscr{U} \times \mathscr{T}$. Therefore, in case of a finite distribution, also the expected recourse function $\Psi(x)$ is constant on polyhedral subsets of $\bar{X}$. This property allows to reduce the set X to a finite set of solution candidates that can be enumerated [SSV98]. Since the expected recourse function $\Psi(x)$ has to be evaluated for each candidate, many similar integer programs have to be solved. In [SSV98] a Gröbner basis for the second stage problem is computed once in advance (which is expensive) and then used for evaluation of $\Psi(x)$ for every candidate x (which is then cheap).

Another approach based on enumerating the sets where $\Psi(x)$ is constant is presented in [ATS00]. Instead of a complete enumeration, here a branch-and-bound algorithm is applied to the first stage problem to enumerate the regions of constant $\Psi(x)$ implicitely. Branching is thereby performed along lines of discontinuity of $\Psi(x)$, thereby reducing its discontinuity in generated subproblems.

While all approaches discussed so far explore the structure of the value or expected recourse function in some way, Lagrange decomposition is a class of algorithms where decomposition is achieved by relaxation of problem constraints. By moving certain coupling restrictions from the set of constraints into the objective function as penalty term, the problem decomposes into a set of subproblems, each of them often much easier to handle than the original problem. This relaxed problem then yields a lower bound onto the original optimal value, which is further improved by optimization of the penalty parameters. Since, in general, a solution of the relaxed problem violates the coupling constraints, heuristics and branch-and-bound approaches are applied to obtain good feasible solutions of the original problem. While there are several alternatives to choose a set of coupling constraints for relaxation, each one providing a lower bound of different quality [DR04], in general, scenario and geographical decomposition are the preferred strategies [CS99, NR00]. In scenario decomposition, nonanticipativity constraints are relaxed, so that the problem decomposes into one deterministic subproblem for each scenario. We discuss this approach in more detail in Section 5.3. In geographical decomposition, model-specific constraints are relaxed, which leads to one subproblem for each component of the model. Even though each subproblem then corresponds to a stochastic program itself, its structure often allows to develop specialized algorithms to solve them very efficiently. Similarly, the modelers knowledge can be explored to make solutions from the relaxed problem feasible for the original problem. Geographical decomposition is demonstrated for a unit commitment problem in Section 6.

5.1 Convexification of the expected recourse function

In a simple integer recourse model, the second stage variables are purely integer ($m_1 = 0$) and are partitioned into two sets $y^+, y^- \in \mathbb{Z}^s_+$ with $2s = m_2$. The cost-vector $q(\xi) \equiv (q^+, q^-)$ and technology matrix $T(\xi) \equiv T$ are fixed, $r = 2s$, $h(\xi) \equiv \xi$, and the value function takes the form

$$\Phi(q(\xi), h(\xi) - T(\xi)x) = \inf \left\{ \langle q^+, y^+ \rangle + \langle q^-, y^- \rangle \;\middle|\; \begin{array}{rcl} y^+ & \geq & \xi - Tx, \\ y^- & \geq & -(\xi - Tx) \\ y^+, y^- & \in & \mathbb{Z}^s_+ \end{array} \right\}.$$

The simple structure of the value function allows to write the expected recourse function in a separable form,

$$\Psi(x) = \sum_{i=1}^{s} q_i^+ \mathbb{E}[\lceil \xi_i - (Tx)_i \rceil^+] + q_i^- \mathbb{E}[\lfloor \xi_i - (Tx)_i \rfloor^-],$$

where $\lceil \alpha \rceil$ denotes the smallest integer that is at least α, $\lfloor \alpha \rfloor$ the largest integer that is at most α, $\alpha^+ = \max(0, \alpha)$, and $\alpha^- = \min(0, \alpha)$. Thus, it is sufficient to consider one-dimensional functions of the form $Q(z) := q^+ \mathbb{E}_\zeta[\lceil \zeta - z \rceil^+] + q^- \mathbb{E}_\zeta[\lfloor \zeta - z \rfloor^-]$ (with ζ a random variable).

In [KSV06], convex approximations of $Q(z)$ are derived from a piecewise linear function in the points $(z, Q(z))$, $z \in \alpha + \mathbb{Z}$, where $\alpha \in [0, 1)$ is a parameter. Further, if ζ has a continuous distribution, then the approximation of $Q(z)$ can be realized as expected recourse function of a continuous simple recourse model,

$$Q_\alpha(z) = q^+ \mathbb{E}_{\zeta_\alpha}[(\zeta_\alpha - z)^+] + q^- \mathbb{E}_{\zeta_\alpha}[(\zeta_\alpha - z)^-] + \frac{q^+ q^-}{q^+ + q^-},$$

where ζ_α is a discrete random variable with support in $\alpha + \mathbb{Z}$ [KSV06].

The results in [KSV06] are extended to derive convex approximations of the expected recourse function for models of the form (1), where $m_1 = 0$, $h(\xi) \equiv \xi$, and $q(\xi) \equiv q$ and $T(\xi) \equiv T$ are fixed [vdV04]. Further, the parameter α can be chosen such that the derived convex approximation underestimates the original expected recourse function. Since this convex underestimator is at least as good as an LP-based underestimator (obtained by relaxing the integrality condition on y) and even yield the convex hull of $\Psi(x)$ in the case that T is unimodular, it can be utilized to derive lower bounds in a Branch-and-Bound search for a solution of (1).

Another extension of the methodology from [KSV06] considers mixed-integer recourse models where $r = 1$ and the value function is semi-periodic, c.f. [vdV05].

5.2 Convexification of the value function

From now on we assume that the random vector ξ has only finitely many outcomes ξ^i with probability $p_i > 0$, $i = 1, \dots, N$. Thus, we can write the expected recourse function as

$$\Psi(x) = \sum_{i=1}^{N} p_i \Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x) \qquad (x \in \bar{X}).$$

As discussed in Section 2, the nonconvexity of the function $\Phi(u,t)$ forbids a representation by supporting hyperplanes as used in a Benders decomposition. However, while in the continuous case ($m_2 = 0$) the hyperplanes are derived from dual feasible solutions of the second stage problem, it is conceptually possible to carry over these ideas to the mixed-integer case by introducing (possibly nonlinear) dual price functions [TiW81]. Indeed, Chvátal and Gomory functions are sufficiently large classes of dual price functions that allow to approximate the value function $\Phi(u,t)$ [BJ82]. These dual functions can be obtained from a solution of (3) with a branch-and-bound or Gomory cutting plane algorithm [Wo81]. In [CT98] this approach is used to carry over the Benders decomposition algorithm for two-stage linear stochastic programs to the mixed-integer linear case by replacing the hyperplane approximation of the expected recourse function by an approximation based on dual price functions. While [CT98] does not discuss how the master problem with its (nonsmooth and nonconvex) dual price functions can be solved, the series of papers [SeH05, NS08b, SeS06, NS05, NS08a] show that a careful construction of dual price functions combined with a convexification step based on disjunctive programming allows to implement an efficient Benders decomposition for mixed-integer two-stage stochastic programs.

We consider the following *master problem* obtained from (1) by replacing the value functions $x \mapsto \Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x)$ by approximations $\Theta_i : \mathbb{R}^m \to \mathbb{R}$:

$$\min \left\{ \langle c, x \rangle + \sum_{i=1}^{N} p_i \Theta_i(x) \;\middle|\; x \in X \right\}, \tag{25}$$

where each function $\Theta_i(\cdot)$, $i = 1, \dots, N$, is given in the form

$$\Theta_i(x) := \max\{\min\{\eta_1(x), \dots, \eta_k(x)\} \mid (\eta_1(\cdot), \dots, \eta_k(\cdot)) \in C_i\}, \qquad x \in \bar{X},$$

and a tuple $\eta := (\eta_1(\cdot), \ldots, \eta_k(\cdot)) \in C_i$ consists of k (where k is allowed to vary with η) affine linear functions $\eta_j(\cdot) : \mathbb{R}^m \to \mathbb{R}$, $j = 1, \ldots, k$. The tuple η takes here the role of an optimality cut in Benders decomposition for the continuous case. That is, each $\eta \in C_i$ is constructed in a way such that for all $x \in X$

$$\Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x) \geq \eta_j(x) \text{ for at least one } j \in \{1, \ldots, k\}. \tag{26}$$

Hence, we have $\Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x) \geq \Theta_i(x)$ and the optimal value of problem (25) is a lower bound to the optimal value of (1). Before discussing the construction of the tuples η, we shortly discuss an algorithm to solve problem (25).

5.2.1 Solving the master problem

Note that problem (25) can be written as a disjunctive mixed-integer linear problem:

$$\min \left\{ \langle c, x \rangle + \sum_{i=1}^{N} p_i \theta_i \;\middle|\; \begin{array}{l} x \in X, \\ \theta_i \geq \eta_1(x) \vee \ldots \vee \theta_i \geq \eta_k(x),\ \eta \in C_i, i = 1, \ldots, N \end{array} \right\}. \tag{27}$$

Problem (27) can be solved by a branch-and-bound algorithm [NS08b]. To this end, assume that for each tuple $\eta \in C_i$ an affine linear function $\bar{\eta}(\cdot) : \mathbb{R}^m \to \mathbb{R}$ is known which underestimates each $\eta_j(\cdot)$, $j = 1, \ldots, k$, on X, i.e., $\eta_j(x) \geq \bar{\eta}(x)$ for $j = 1, \ldots, k$ and $x \in X$. $\bar{\eta}(\cdot)$ allows to derive a linear relaxation of problem (27):

$$\min \left\{ \langle c, x \rangle + \sum_{i=1}^{N} p_i \theta_i \;\middle|\; x \in \bar{X},\ \theta_i \geq \bar{\eta}(x),\ \eta \in C_i,\ i = 1, \ldots, N \right\}. \tag{28}$$

Let $(\hat{x}, \hat{\theta})$ be a solution of (28). If $\hat{x}$ is feasible for (27), then an optimal solution for (27) has been found. Otherwise, $\hat{x}$ either violates an integrality restriction on a variable x_j, $j = 1, \ldots, m_0$, or a disjunction $\theta_i \geq \min\{\eta_1(x), \ldots, \eta_k(x)\}$ for some tuple $\eta \in C_i$ (with $k > 1$) and some scenario i. In the former case, that is, $\hat{x}_j \notin \mathbb{Z}$, two subproblems of (28) are created with additional constraints $x_j \leq \lfloor \hat{x}_j \rfloor$ and $x_j \geq \lceil \hat{x}_j \rceil$, respectively. In the latter case, the tuple η is partitioned into two tuples $\eta' = (\eta_1(\cdot), \ldots, \eta_{k'}(\cdot))$ and $\eta'' = (\eta_{k'+1}(\cdot), \ldots, \eta_k(\cdot))$, $1 \leq k' < k$, corresponding linear underestimators $\bar{\eta}'(\cdot)$ and $\bar{\eta}''(\cdot)$ are computed (where $\bar{\eta}' = \eta_1$ if $k' = 1$ and $\bar{\eta}'' = \eta_k$ if $k' = k - 1$), and two subproblems where the tuple $\eta \in C_i$ is replaced by η' and η'', respectively, are constructed. Next, the same method is applied to each subproblem recursively. The first feasible solution for problem (27) is stored as "incumbent solution". In the following, new feasible solutions replace the incumbent solution if they have a better objective value. If a subproblem is infeasible or the value of its linear relaxation exceeds the current incumbent solution, then it can be discarded from the list of open subproblems. Since in each subproblem the number of feasible discrete values for a variable x_j or the length of a tuple $\eta \in C_i$ is reduced with respect to the ascending problem, the algorithm can generate only a finite number of subproblems and thus terminates with a solution of (27).

5.2.2 Convexification of disjunctive cuts

A linear function $\bar{\eta}(\cdot)$ in (28) that underestimates $\min\{\eta_1(\cdot),\dots,\eta_k(\cdot)\}$ can be constructed by means of disjunctive programming [Ba98, SeH05]: For a fixed scenario index i and a tuple $\eta \in C_i$, an inequality $\theta \geq \bar{\eta}(x)$ is valid for the feasible set of (27), if it is valid for $\bigcup_{j=1}^{k}\{(x,\theta)\in\mathbb{R}^{m+1} \mid x\in\bar{X}, \theta\geq\eta_j(x)\}$. That is, we require

$$\bar{\eta}(x) \leq \eta_j(x) \qquad \text{for all } x\in\bar{X}, \quad j=1,\dots,k. \tag{29}$$

We write $\bar{\eta}(x) = \bar{\eta}_0 + \langle\bar{\eta}_x, x\rangle$ and $\eta_j(x) = \eta_{j,0} + \langle\eta_{j,x}, x\rangle$ for some $\bar{\eta}_0, \eta_{j,0}\in\mathbb{R}$ and $\bar{\eta}_x, \eta_{j,x}\in\mathbb{R}^m$, $j=1,\dots,k$. Then (29) is equivalent to

$$\begin{aligned}\bar{\eta}_0 - \eta_{j,0} \leq &\min\{\langle\eta_{j,x}-\bar{\eta}_x, x\rangle \mid x\in\mathbb{R}^m, Ax\leq b\}\\ =&\max\{\langle\lambda_j, b\rangle \mid \lambda_j\in\mathbb{R}^{r_0}_{-},\ A^\top\lambda_j = \eta_{j,x}-\bar{\eta}_x\}.\end{aligned}$$

Therefore, choosing $\lambda_j\in\mathbb{R}^{r_0}_{-}$ and $\bar{\eta}_x\in\mathbb{R}^m$ such that $A^\top\lambda_j + \bar{\eta}_x = \eta_{j,x}$, and setting $\bar{\eta}_0 := \eta_{j,0} + \min\{\langle\lambda_j, b\rangle | j=1,\dots,k\}$ yields a function $\bar{\eta}(x)$ that satisfies (29).

[SeH05] note, that given an extreme point $\hat{x}$ of $\bar{X}$, the linear underestimator $\bar{\eta}(\cdot)$ can be chosen such that $\bar{\eta}(\hat{x}) = \min\{\eta_1(\hat{x}),\dots,\eta_k(\hat{x})\}$. Thus, if only extreme points of $\bar{X}$ are feasible for (1), then it is not necessary to branch on disjunctions η to solve (27). This is the case, e.g., if all first stage variables are restricted to be binary.

5.2.3 Approximation of $\Phi(u,t)$ by linear optimality cuts

The simplest way to construct a tuple η with property (26) is to derive a supporting hyperplane for the linear relaxation of $\Phi(u,t)$, which we denote by

$$\bar{\Phi}(u,t) := \min\{\langle u_1, y_1\rangle + \langle u_2, y_2\rangle \mid y_1\in\mathbb{R}^{m_1},\ y_2\in\mathbb{R}^{m_2},\ W_1y_1 + W_2y_2 \leq t\}. \tag{30}$$

It is well known, that $\bar{\Phi}(u,t)$ is piecewise linear and convex in t. Thus, if, for fixed $(\hat{u},\hat{t})\in\mathscr{U}\times\mathscr{T}$, $\hat{\pi}$ is a dual solution of (30), we obtain the inequality $\bar{\Phi}(\hat{u},t)\geq\bar{\Phi}(\hat{u},\hat{t}) + \langle\hat{\pi}, t-\hat{t}\rangle = \langle\hat{\pi}, t\rangle$ $(t\in\mathscr{T})$. Letting $\hat{u} = q(\xi^i)$ and $\hat{t} = h(\xi^i) - T(\xi^i)\hat{x}$ for a fixed scenario ξ^i and first stage decision $\hat{x}\in\bar{X}$, we obtain

$$\bar{\Phi}(q(\xi^i), h(\xi^i) - T(\xi^i)x) \geq \langle\hat{\pi}, h(\xi^i) - T(\xi^i)x\rangle =: \eta_1(x). \tag{31}$$

Since $\Phi(u,t)\geq\bar{\Phi}(u,t)$, (31) yields the *optimality cut* $\eta := (\eta_1(\cdot))$ (i.e., $k=1$). Due to the polyhedrality of $\bar{\Phi}(u,t)$, a finite number of such cuts for each scenario is sufficient to obtain an exact representation of $\bar{\Phi}(u,t)$ in the master problem (27).

5.2.4 Approximation of $\Phi(u,t)$ by lift-and-project

However, in order to capture the nonconvexity of the original value function $\Phi(u,t)$, *nonconvex optimality cuts* are necessary, i.e., tuples η of length $k>1$. For the case that the discrete variables in the second stage are all of binary type, the following method is proposed in [SeH05]: Let $\bar{x} \in X$ be a feasible solution of the master problem (25), let $(\bar{y}_1^i, \bar{y}_2^i)$ be a solution of the relaxed second stage problem (30) for $u = q(\xi^i)$ and $t = h(\xi^i) - T(\xi^i)\hat{x}$, $i = 1, \ldots, N$. If $\bar{y}_2^i \in \mathbb{Z}^{m_2}$ for all $i = 1, \ldots, N$, then a linear optimality cut (31) is derived, c.f. (31). Otherwise, let $j \in \{1, \ldots, m_2\}$ be an index such that $0 < \bar{y}_{2,j}^i < 1$. We now seek for inequalities $\langle \pi_1^i, y_1 \rangle + \langle \pi_2^i, y_2 \rangle \geq \pi_0^i(x)$, $i = 1, \ldots, N$, which are valid for (3) for all $x \in \bar{X}$, but cut off the solution $\bar{y}^i$ from (30) for at least one scenario i with fractional $\bar{y}_{2,j}^i$. That is, we search for inequalities that are valid for the disjunctive sets

$$\left\{ y \in \mathbb{R}^{m_1+m_2} \;\middle|\; \begin{aligned} W_1 y_1 + W_2 y_2 &\leq t, \\ y_{2,j} &\leq 0 \end{aligned} \right\} \cup \left\{ y \in \mathbb{R}^{m_1+m_2} \;\middle|\; \begin{aligned} W_1 y_1 + W_2 y_2 &\leq t, \\ -y_{2,j} &\leq -1 \end{aligned} \right\}, \tag{32}$$

where $t = h(\xi^i) - T(\xi^i)\bar{x}$, $i = 1, \ldots, N$. Observe, that points with fractional $y_{2,j}$ are not contained in (32). With an argumentation similar to the derivation of $\bar{\eta}(\cdot)$ before, it follows that, for fixed x, valid inequalities for (32) are described by the system

$$\begin{aligned}
W_1^\top \lambda_{1,1}^i &= \pi_1^i, & W_1^\top \lambda_{2,1}^i &= \pi_1^i, && \text{(33a)} \\
W_2^\top \lambda_{1,1}^i + e_j \lambda_{1,2}^i &= \pi_2^i, & W_2^\top \lambda_{2,1}^i - e_j \lambda_{2,2}^i &= \pi_2^i, && \text{(33b)} \\
\langle h(\xi^i) - T(\xi^i)x, \lambda_{1,1}^i \rangle &\geq \pi_0^i(x), & \langle h(\xi^i) - T(\xi^i)x, \lambda_{2,1}^i - \lambda_{2,2}^i \rangle &\geq \pi_0^i(x), && \text{(33c)} \\
\lambda_{1,1}^i \in \mathbb{R}_-^r, \lambda_{1,2}^i &\in \mathbb{R}_-, & \lambda_{2,1}^i \in \mathbb{R}_-^r, \lambda_{2,2}^i &\in \mathbb{R}_-, && \text{(33d)}
\end{aligned}$$

where $e_j \in \mathbb{R}^{m_2}$ is the j-th unit vector. Observe further, that the coefficients in (33a) and (33b) (i.e., W_1, W_2, e_j) are scenario independent. Thus, it is possible to use common cut coefficients $(\pi_1, \pi_2) \equiv (\pi_1^i, \pi_2^i)$ for all scenarios, thereby reducing the computational effort to the solution of a single linear program [SeH05]:

$$\max \left\{ \sum_{i=1}^N p_i(\pi_0^i(\bar{x}) - \langle \pi_1, \bar{y}_1^i \rangle - \langle \pi_2, \bar{y}_2^i \rangle) \;\middle|\; \begin{array}{l} \lambda_{1,1}, \lambda_{2,1} \in \mathbb{R}_-^r,\ \lambda_{1,2}, \lambda_{2,2} \in \mathbb{R}_-, \\ \pi_1 \in \mathbb{R}^{m_1},\ \pi_2 \in \mathbb{R}^{m_2},\ \pi_0^i(\bar{x}) \in \mathbb{R}, \\ W_1^\top \lambda_{1,1} = \pi_1, W_2^\top \lambda_{1,1} + e_j \lambda_{1,2} = \pi_2, \\ W_1^\top \lambda_{2,1} = \pi_1, W_2^\top \lambda_{2,1} - e_j \lambda_{2,2} = \pi_2, \\ \langle h(\xi^i) - T(\xi^i)\bar{x}, \lambda_{1,1} \rangle \geq \pi_0^i(\bar{x}), \\ \langle h(\xi^i) - T(\xi^i)\bar{x}, \lambda_{2,1} \rangle - \lambda_{2,2} \geq \pi_0^i(\bar{x}), \\ \|\pi_1\|_\infty \leq 1, \|\pi_2\|_\infty \leq 1, |\pi_0^i(\bar{x})| \leq 1, \\ i = 1, \ldots, N \end{array} \right\}$$

The objective function of this simple recourse problem maximizes the average violation of the computed cuts by $(\bar{y}_1^i, \bar{y}_2^i)$. The functions $\pi_0^i(\cdot)$, $i = 1, \ldots, N$, with $\langle \pi_1, y_1 \rangle + \langle \pi_2, y_2 \rangle \geq \pi_0^i(x)$ for all $x \in \bar{X}$ is derived from a solution of this LP as

$$\pi_0^i(x) := \min\{\langle h(\xi^i) - T(\xi^i)x, \lambda_{1,1} \rangle, \langle h(\xi^i) - T(\xi^i)x, \lambda_{2,1} \rangle - \lambda_{2,2}\}. \tag{34}$$

Adding these new cuts to (30) for $u = q(\xi^i)$ and $t = h(\xi^i) - T(\xi^i)\bar{x}$, $i = 1, \ldots, N$, yields the updated second stage linear relaxations

$$\min\left\{ \langle q_1(\xi^i), y_1\rangle + \langle q_2(\xi^i), y_2\rangle \;\middle|\; \begin{array}{l} W_1y_1 + W_2y_2 \leq h(\xi^i) - T(\xi^i)\bar{x} \\ -\langle \pi_1, y_1\rangle - \langle \pi_2, y_2\rangle \leq -\pi_0^i(\bar{x}) \\ y_1 \in \mathbb{R}^{m_1}, y_2 \in \mathbb{R}^{m_2} \end{array} \right\}. \quad (35)$$

A dual solution (μ, μ_0) of (35) can then be used to derive the inequality

$$\Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x) \geq \langle \mu, h(\xi^i) - T(\xi^i)x\rangle - \mu_0\pi_0^i(x).$$

However, the nonconvexity of the right-hand side $\pi_0^i(x)$ yields a nonconvex optimality cut $\eta := (\eta_1(\cdot), \eta_2(\cdot))$, where

$$\begin{aligned} \eta_1(x) &:= \langle \mu - \mu_0\lambda_{1,1}, h(\xi^i) - T(\xi^i)x\rangle, \\ \eta_2(x) &:= \langle \mu - \mu_0\lambda_{2,1}, h(\xi^i) - T(\xi^i)x\rangle + \mu_0\lambda_{2,2}. \end{aligned}$$

In a next iteration, when the second stage problems are revisited with a different first stage solution $\bar{x}$, the updated relaxation (35) takes the place of the original relaxation (30). Since the functions $\pi_0^i(\cdot)$ are known, the right-hand side of the added cut in (35) is updated when $\bar{x}$ changes.

5.2.5 Approximation of $\Phi(u,t)$ by branch-and-bound

For the general case where the discrete second stage variables can also be of integer type, the second stage problem (3) can be solved by a (partial) branch-and-bound algorithm and a (nonlinear) optimality cut η can be derived from the dual solutions of the linear programs in each leaf of the branch-and-bound tree [SeS06]: Let $\bar{x} \in X$ be again a feasible point to problem (25) and fix a scenario i. Assume that (3) with $\bar{u} = q(\xi^i)$ and $\bar{t} = h(\xi^i) - T(\xi^i)\bar{x}$ is (partially) solved by a branch-and-bound algorithm. Denote by $\mathscr{Q}$ the set of terminal nodes of the generated branch-and-bound tree. For any node $q \in \mathscr{Q}$ let $y_{2,l}^q$ and $y_{2,u}^q$ denote the vectors that define lower and upper bounds on the y_2 variables in the subproblem at node q. Then the LP relaxation of (3) for node $q \in \mathscr{Q}$ is given as

$$\min\left\{ \langle \bar{u}_1, y_1\rangle + \langle \bar{u}_2, y_2\rangle \;\middle|\; y \in \mathbb{R}^{m_1+m_2}, W_1y_1 + W_2y_2 \leq \bar{t}, \begin{array}{l} y_2 \leq y_{2,u}^q, \\ -y_2 \leq -y_{2,l}^q \end{array} \right\}. \quad (36)$$

We assume, that subproblems have been pruned if they are infeasible or their lower bound exceeds a known upper bound. Thus, all terminal nodes are associated with a feasible LP relaxation. The dual problem to (36) is

$$\max\left\{ \langle \mu, \bar{t}\rangle + \langle \pi_u, y_{2,u}^q\rangle - \langle \pi_l, y_{2,l}^q\rangle \;\middle|\; \begin{array}{ll} \mu \in \mathbb{R}^r_-, & W_1^\top\mu = \bar{u}_1 \\ \pi_l, \pi_u \in \mathbb{R}^{m_2}_-, & W_2^\top\mu + \pi_u - \pi_l = \bar{u}_2 \end{array} \right\}, \quad (37)$$

where we assume that a dual variable $\pi_{l,j}$, $\pi_{u,j}$ is fixed to 0 if the corresponding bound $y_{2,l,j}^q$, $y_{2,u,j}^q$ is $-\infty$ or $+\infty$, respectively, $j = 1, \ldots, m_2$. Based on a dual solution $(\mu^q, \pi_l^q, \pi_u^q)$ of (37), a supporting hyperplane of each nodes LP value function can be derived, c.f. (31). Since the branch-and-bound tree represents a partition of the feasible set of (3), it allows to state a disjunctive description of the function $t \mapsto \Phi(\bar{u},t)$ by combining the LP value function approximations in all nodes $q \in \mathcal{Q}$:

$$\Phi(\bar{u},t) \geq \langle \mu^q, t\rangle + \langle \pi_u^q, y_{2,u}^q\rangle - \langle \pi_l^q, y_{2,l}^q\rangle \quad \text{for at least one } q \in \mathcal{Q}. \tag{38}$$

This result directly translates into a nonlinear optimality cut $\eta := (\eta_1(\cdot),\ldots,\eta_q(\cdot))$ by letting $\eta_q(x) := \langle \mu^q, h(\xi^i) - T(\xi^i)x\rangle + \langle \pi_u^q, y_{2,u}^q\rangle - \langle \pi_l^q, y_{2,l}^q\rangle$.

5.2.6 Full Algorithm

We can now state a full algorithm for the solution of (1):

1. Solve the master problem (27) by branch-and-bound. If it is infeasible, then (1) is infeasible. Otherwise, let $(\bar{x},\bar{\theta})$ be a solution of (27).
2. Solve (3) for each scenario $i = 1,\ldots,N$. Let $\phi_i := \Phi(q(\xi^i), h(\xi^i) - T(\xi^i)\bar{x})$ be the optimal value of (3) for the first stage decision $\bar{x}$ in scenario i.
3. For scenarios i where $\phi_i > \bar{\theta}_i$, derive an optimality cut η of the value function $x \mapsto \Phi(q(\xi^i), h(\xi^i) - T(\xi^i)x)$ either via linearization of $\bar{\Phi}(u,t)$ (see (31)), via lift-and-project (Section 5.2.4), or from a (partial) branch-and-bound search (Section 5.2.5). Add η to C_i in (27).
4. If no new tuples η have been constructed, i.e., the master problem has not been updated, then finish: $\bar{x}$ is an optimal solution to (3). Otherwise, go back to 1.

Some remarks are in order:

- At the beginning, the sets C_i are empty, i.e., no information about the value function $\Phi(u,t)$ is available in (27). Thus, (27) should be solved either with the variables θ_i removed or bounded from below by a known lower bound on $\Phi(u,t)$.
- In the first iterations, when almost no information about $\Phi(u,t)$ is available, it is unnecessary to solve the master problem (27) and the second stage problems (3) to optimality. Instead, at first it is more efficient to ignore the integrality conditions and to construct a representation of the LP value function $\bar{\Phi}(u,t)$ by a usual Benders decomposition. Later, partial solves of (27) and the introduction of nonlinear optimality cuts η into (27) based on lift-and-project or partial branch-and-bound searches should be performed to capture the nonconvexity of $\Phi(u,t)$ in the master problem. Finally, to ensure convergence, first and second stage problems need to be solved to optimality, see also [SeH05, NS08b].

5.2.7 Extension to multistage problems

While the algorithms discussed so far allow an efficient extension of the Benders decomposition to two-stage mixed-integer stochastic programs, a further extension to the multistage case seems possible. While in the two-stage case we have a nonconvex value function only in the first stage, in the multistage setting we are faced with such a function in each node of the scenario tree other than the leaves. That is, the master problems in each node before the last stage are of the form (27). Approximation of the value function of such a master problem then requires to take the nonlinear optimality cuts which approximate the value functions of successor nodes into account. For that matter, we have seen how such a master problem can be solved by branch-and-bound (Section 5.2.1, [NS08b]) and how an optimality cut can be derived from a (partial) branch-and-bound search (Section 5.2.5, [SeS06]).

However, the efficiency of such an approach might suffer under the large number of disjunctions that are induced from optimality cuts on late stages into the master problems on early stages. That is, while in the two-stage case the disjunctions in (27) are caused only by integrality constraints on the second stage, in the multistage setting we have to deal with disjunctions that are induced by disjunctions on succeeding stages. Therefore, solving a fairly large mixed-integer multistage stochastic program to optimality with this approach seems questionable.

Nevertheless, an interesting application are multistage problem that can only be solved efficiently by a temporal decomposition, e.g., stochastic programs with recombining scenario trees [KV07]. For the latter, the recombining nature of the scenario tree leads to coinciding value functions, a property that can be explored by a nested Benders decomposition. Therefore, an extension to the mixed-integer case by application of the ideas discussed in this section seems promising.

5.3 Scenario Decomposition

Consider the following reformulation of (1) where the first stage variable x is replaced by one variable x^i for each scenario $i = 1,\dots,N$ and an explicit *nonanticipativity constraint* is added:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{N} p_i(\langle c,x^i\rangle + \langle q_1(\xi^i),y_1(\xi^i)\rangle + \langle q_2(\xi^i),y_2(\xi^i)\rangle) && \text{(39a)}\\
\text{such that} \quad & x^i \in X, \quad y_1^i \in \mathbb{R}^{m_1}, \quad y_2^i \in \mathbb{Z}^{m_2}, \qquad i = 1,\dots,N, && \text{(39b)}\\
& T(\xi^i)x^i + W_1 y_1^i + W_2 y_2^i \le h(\xi^i), \qquad i = 1,\dots,N, && \text{(39c)}\\
& x^1 = x^2 = \ldots = x^N. && \text{(39d)}
\end{aligned}
$$

Problem (39) decomposes into scenariowise subproblems by relaxing the coupling constraint (39d) [CS99]. The violation of the relaxed constraints is then added as a penalty into the objective function. That is, each subproblem has the form

$$
D_i(\lambda) := \min\left\{ L_i(x^i,y^i;\lambda) \;\middle|\; \begin{array}{c} x^i \in X,\, y_1^i \in \mathbb{R}^{m_1},\, y_2^i \in \mathbb{Z}^{m_2}, \\ T(\xi^i)x^i + W_1 y_1^i + W_2 y_2^i \le h(\xi^i), \end{array} \right\}, \tag{40}
$$

where $\lambda := (\lambda^1,\dots,\lambda^N) \in \mathbb{R}^{mN}$ is the *Lagrange multiplier* and

$$L_i(x^i,y^i;\lambda) := p_i(\langle c,x^i\rangle + \langle q_1(\xi^i),y_1(\xi^i)\rangle + \langle q_2(\xi^i),y_2(\xi^i)\rangle + \langle \lambda^i,x^i - x^1\rangle),$$

$i = 1,\dots,N$. For every choice of λ, a lower bound on (39) is obtained by computing

$$D(\lambda) := \sum_{i=1}^{N} D_i(\lambda). \tag{41}$$

That is, to compute (41), the deterministic problem (40) is solved for each scenario. To find the best possible lower bound, one now searches for an optimal solution to the *dual problem*

$$\max\{D(\lambda) \mid \lambda \in \mathbb{R}^{mN}\}. \tag{42}$$

The function $D(\lambda)$ is a piecewise linear concave function for which subgradients can be computed from a solution of (40). Thus, solution methods for the nonsmooth convex optimization problem (42) use a bundle of subgradients of $D(\lambda)$ to find promising values of λ [Ki90].

The primal solutions (x^i,y^i), $i = 1,\dots,N$, of (40), associated with a solution of (42), yield in general not a feasible solution to the original problem. To regain the relaxed nonanticipativity constraint, heuristics are employed that, e.g., select for x an average or a frequently occurring value among the x^i and then possibly resolve each second stage problem to ensure feasibility.

To find an optimal solution to (39), a branch-and-bound algorithm can be employed. Here, nonanticipativity constraints are insured by branching on the first stage variables. Since the additional bound constraints on x^i become part of the constraints in (40), the lower bound (42) improves by a branching operation.

An alternative to solving the dual problem (42) by a bundle method is proposed in [LuS04]: As shown in [CS99], the problem (42) is equivalent to the primal problem

$$\min \quad \sum_{i=1}^{N} p_i(\langle c,x^i\rangle + \langle q_1(\xi^i),y_1(\xi^i)\rangle + \langle q_2(\xi^i),y_2(\xi^i)\rangle) \tag{43a}$$

$$\text{such that} \quad (x^i,y^i) \in \operatorname{conv}\left\{(x,y_1,y_2) \;\middle|\; \begin{array}{c} x \in X,\ y_1 \in \mathbb{R}^{m_1},\ y_2 \in \mathbb{Z}^{m_2}, \\ T(\xi^i)x + W_1y_1 + W_2y_2 \le h(\xi^i) \end{array}\right\}, \tag{43b}$$

$$i = 1,\dots,N,$$

$$x^1 = x^2 = \dots = x^N. \tag{43c}$$

This problem is solved by a column generation approach, which constructs an inner approximation of the convex hull in (43b). Feasible solutions for the original problem are obtained by application of branch-and-bound.

For problems where all first stage variables are restricted to be binary, [AEO03] propose to relax both nonanticipativity and integrality constraints. Thereby, each scenario is associated with a branch-and-bound tree that enumerates the integer feasible solutions to the scenario's subproblem (i.e., the feasible set of (40)). Since each branch-and-bound fixes first stage variables to be either 0 or 1, a coordinated search across all n branch-and-bound trees allows to select feasible solutions from each subproblem that satisfy the nonanticipativity constraints. If also continuous variables are present in the first stage, [EG+07] propose to "cross over" to a Benders decomposition whenever the coordinated branch-and-bound search yields solutions which binary first stage variables satisfy the nonanticipativity constraints and second stage integer variables are fixed.

6 Application to stochastic thermal unit commitment

We consider an power generation system comprising thermal units and contracts for delivery and purchase, and describe a model for its cost-minimal operation under uncertainty in electrical load and in prices for fuel and electricity. Contracts for delivery and purchase of electricity are regarded as special thermal units. It is assumed that the time horizon is discretized into uniform (e.g., hourly) intervals. Let T and I denote the numbers of time periods and thermal units, respectively. For thermal unit i in period t, $u_{it} \in \{0,1\}$ is its *commitment* decision (1 if on, 0 if off), and x_{it} its *production*, with

$$u_{it} x_{it}^{\min} \leq x_{it} \leq x_{it}^{\max} u_{it} \qquad (i = 1,\ldots,I,\, t = 1,\ldots,T), \tag{44}$$

where $x_{it}^{\min}$ and $x_{it}^{\max}$ are the minimum and maximum capacities. Additionally, there are *minimum up/down-time requirements*: when unit i is switched on (off), it must remain on (off) for at least $\bar{\tau}_i$ ($\underline{\tau}_i$, resp.) periods, i.e.,

$$\begin{aligned} u_{i\tau} - u_{i,\tau-1} &\leq u_{it} \qquad (\tau = t - \bar{\tau}_i + 1,\ldots,t-1), && (45)\\ u_{i,\tau-1} - u_{i\tau} &\leq 1 - u_{it} \qquad (\tau = t - \underline{\tau}_i + 1,\ldots,t-1). && (46)\end{aligned}$$

for all $t = 1,\ldots,T$ and $i = 1,\ldots,I$. Let U_i denote the set of all pairs (x_i, u_i) satisfying the constraints (44), (45), and (46) for all $t = 1,\ldots,T$. The basic system requirement is to meet the electrical load d_t during all time periods $t = 1,\ldots,T$, i.e.,

$$\sum_{i=1}^{I} x_{it} \geq d_t \qquad (t = 1,\ldots,T). \tag{47}$$

The expected total system cost is given by the sum of expected startup and operating costs of all thermal units over the whole scheduling horizon, i.e.,

$$\mathbb{E}\left(\sum_{t=1}^{T}\sum_{i=1}^{I}(C_{it}(x_{it},u_{it}) + S_{it}(u_i))\right). \tag{48}$$

The *fuel cost* C_{it} for operating thermal unit i is assumed to be piecewise linear convex (concave for purchase contracts) during period t, i.e.,

$$C_{it}(x_{it},u_{it}) := \max_{l=1,\ldots,\bar{l}} \{ a_{ilt} x_{it} + b_{ilt} u_{it} \}$$

with cost coefficients a_{ilt}, b_{ilt}. The startup cost of unit i depends on its downtime; it may vary between a maximum cold-start value and a much smaller value when the unit is still relatively close to its operating temperature. This is modeled by the *startup cost*

$$S_{it}(u_i) := \max_{\tau=0,\ldots,\tau_i^c} c_{i\tau}\left(u_{it} - \sum_{\kappa=1}^{\tau} u_{i,t-\kappa}\right),$$

where $0 = c_{i0} < \ldots < c_{i\tau_i^c}$ are cost coefficients, τ_i^c is the cool-down time of unit i, $c_{i\tau_i^c}$ is its maximum cold-start cost, $u_i := (u_{it})_{t=1}^T$, and $u_{i\tau} \in \{0,1\}$ for $\tau = 1-\tau_i^c,\ldots,0$ are given initial values.

It is assumed that the stochastic input process $\xi = \{\xi\}_{t=1}^T$ is given by

$$\xi_t := (a_t, b_t, c_t, d_t) \quad (t = 1,\ldots,T).$$

or by some of its components. Furthermore, it is assumed that $\xi_1,\ldots,\xi_{t_1}$ (i.e., the input data for the first time period for which reliable forecasts are available), and, thus, the (first stage) decisions $\{(x_{it},u_{it}) \mid t = 1,\ldots,t_1,\, i = 1,\ldots,I\}$ are deterministic.

Minimizing the expected total cost (48) such that the operational constraints (44), (45), (46), and (47) are satisfied, represents a two-stage (linear) mixed-integer stochastic program with (random) second stage decision $\{(x_{it},u_{it}) \mid t = t_1+1,\ldots,T,\, i = 1,\ldots,I\}$.

In many cases it is possible to derive a model for the probability distribution P of ξ via time series analysis based on historical data (see, e.g., [ERW05, SYG06]). Sampling from P together with applying scenario reduction (see Section 4) then leads to a finite number of scenarios $\xi^j = (a_t^j, b_t^j, c_t^j, d_t^j)$ with probabilities p_j, $j = 1,\ldots,N$, for the stochastic process ξ and to the corresponding decision scenarios (x_i^j, u_i^j) (for unit i). The scenario-based unit commitment problem then reads

$$\min\left\{ \sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{I} p_j\big(C_{it}^j(x_{it}^j,u_{it}^j) + S_{it}^j(u_i^j)\big) \;\middle|\; \begin{array}{l} (x_i^j,u_i^j) \in U_i,\, i = 1,\ldots,I, \\ \sum_{i=1}^{I} p_{it}^j \geq d_t^j,\, t = 1,\ldots,T, \\ j = 1,\ldots,N \end{array} \right\} \tag{49}$$

where C_{it}^j and S_{it}^j denote the cost functions for scenario j.

Since the optimization problem (49) only contains NT (unit) coupling constraints while the number $2NTI$ of decision variables is typically (much) larger, *geographical decomposition* based on Lagrangian relaxation of the coupling constraints (47) seems to be promising. The Lagrangian function is of the form

$$\begin{aligned} L(x,u;\lambda) &= \sum_{j=1}^{N}\sum_{t=1}^{T} p_j\left(\sum_{i=1}^{I}\big(C_{it}^j(x_{it}^j,u_{it}^j) + S_{it}^j(u_i^j)\big) + \lambda_t^j\big(d_t^j - \sum_{i=1}^{I} x_{it}^j\big)\right) \\ &= \sum_{j=1}^{N}\sum_{t=1}^{T} p_j\left(\sum_{i=1}^{I}\big(C_{it}^j(x_{it}^j,u_{it}^j) + S_{it}^j(u_i^j) - \lambda_t^j x_{it}^j\big) + \lambda_t^j d_t^j\right), \end{aligned}$$

which leads to the dual function

$$D(\lambda) = \inf_{(x,u)} L(x,u;\lambda) = \sum_{j=1}^{N} p_j \left(\sum_{i=1}^{I} D_{ij}(\lambda^j) + \sum_{t=1}^{T} \lambda_t^j d_t^j \right)$$
$$D_{ij}(\lambda) = \inf_{u_i} \sum_{t=1}^{T} (\inf_{x_{it}} (C_{it}^j (x_{it}, u_{it}) - \lambda_t x_{it}) + S_{it}^j (u_i))$$

decomposing into unit subproblems for every scenario ξ^j and, hence, λ^j. While the inner minimization (with respect to x_{it}) can be solved explicitly, the outer minimization (with respect to u_i) can efficiently be done by dynamic programming. The dual concave (nondifferentiable) maximization problem

$$\max \left\{ D(\lambda) \mid \lambda \in \mathbb{R}_+^{NT} \right\} \tag{50}$$

can be solved by bundle subgradient methods (e.g., [Ki90]). If $(\bar{x}, \bar{u}, \bar{\lambda})$ is an (approximate) solution of (50), $D(\bar{\lambda})$ is a lower bound of the infimum of (49), but, in general, the (maximal) load constraints

$$\sum_{i=1}^{I} x_{it}^{\max} \bar{u}_{it}^j \geq d_t^j \qquad (t = 1, \ldots, T, j = 1, \ldots, N) \tag{51}$$

are violated for some scenarios j and some time intervals t, respectively. However, as shown in [Be82, Section 5.6.1], the relative duality gap gets small if the number I of units is large. In many practical situations this allows to apply simple Lagrangian heuristics (like [ZG88]) to modify $\bar{u}$ scenariowise such that (51) is satisfied for every pair (t, j). After having the commitment decision $\bar{u}$ fixed, a final scenariowise economic dispatch [vBL87] leads to good primal solutions $(\bar{x}, \bar{u})$.

The approach can be extended to multistage models by requiring in addition that the decisions (x_t, u_t) in (49) only depend on $(\xi_1, \ldots, \xi_t)$ (for $t > t_1$). We refer to the relevant work [CC+96, GK+02, GR05, NR00, PCW00, SYG06, TBL96, TKW00].

Furthermore, instead of the expected total system cost, a mean-risk objective of the form

$$\gamma \rho(Y_{t_1}, \ldots, Y_T) - (1-\gamma)\mathbb{E}(Y_T), \;\; Y_t := -\sum_{\tau=1}^{t} \sum_{i=1}^{I} (C_{it}(x_{it}, u_{it}) + S_{it}(u_i)) \;\; (t = t_1, \ldots, T)$$

may be considered, where $\gamma \in (0,1)$ and ρ is a multi-period risk functional (see [EHR09]). In this way, risk management is integrated into unit commitment. If the risk functional is polyhedral [ER05, EHR09], the scenario-based unit commitment model may be reformulated as mixed-integer linear program.

Extensions of the two-stage stochastic unit commitment model are discussed in [NR02] and [NSW05], respectively. In [NR02], a planning model is described whose (deterministic) first stage and (stochastic) second stage decisions are given on the whole time horizon $\{1, \ldots, T\}$. The first stage decisions are determined such that a transition from the first to the second stage and vice versa is always feasible and compatible. In [NSW05], day-ahead trading at a power exchange is incorporated into unit commitment.

Acknowledgements This work was supported by the DFG Research Center MATHEON *Mathematics for Key Technologies* in Berlin (`http://www.matheon.de`) and by the German Ministry of Education and Research (BMBF) under the grant 03SF0312E.

References

[BG+82] Bank, B., Guddat, J., Kummer, B., Klatte, D., Tammer, K.: Non-linear Parametric Optimization. Akademie-Verlag, Berlin (1982)

[AEO03] Alonso-Ayuso, A., Escudero, L.F., Teresa Ortuño, M.: BFC, a branch-and-fix coordination algorithmic framework for solving some types of stochastic pure and mixed 0-1 programs. Eur. J. Oper. Res. **151**, 503–519 (2003)

[ATS00] Ahmed, S., Tawarmalani, M., Sahinidis, N.V.: A finite branch and bound algorithm for two-stage stochastic integer programs. Math. Program. **100**, 355–377 (2004)

[Ba98] Balas, E.: Disjunctive programming: Properties of the convex hull of feasible points. Discrete Appl. Math. **89**, 3–44 (1998). originally MSRR#348, Carnegie Mellon University (1974)

[Be82] Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)

[Bi85] Birge, J.R.: Decomposition and Partitioning Methods for Multistage Stochastic Programming. Oper. Res. **33**(5), 989–1007 (1985)

[BJ82] Blair, C.E., Jeroslow, R.G.: The value function of an integer program. Math. Progr. **23**, 237–273 (1982)

[CS99] Carøe, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. Oper. Res. Lett. **24**(1-2), 37–45 (1999)

[CT97] Carøe, C.C., Tind, J.: A cutting-plane approach to mixed $0-1$ stochastic integer programs. Eur. J. Oper. Res. **101**(2), 306–316 (1997)

[CT98] Carøe, C.C., Tind, J.: L-shaped decomposition of two-stage stochastic programs with integer recourse. Math. Progr. **83**(3, Ser. A), 451–464 (1998)

[CC+96] Carpentier, P., Cohen, G., Culioli, J.-C., Renaud, A.: Stochastic optimization of unit commitment: A new decomposition framework. IEEE Trans. Power Systems **11**, 1067–1073 (1996)

[DR04] Dentcheva, D., Römisch, W.: Duality gaps in nonconvex stochastic optimization. Math. Progr. **101**(3, Ser. A), 515–535 (2004)

[DGR03] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenarios reduction in stochastic programming: An approach using probability metrics. Math. Progr. **95**, 493–511 (2003)

[EHR09] Eichhorn, A., Heitsch, H., Römisch, W.: Stochastic optimization of electricity portfolios: Scenario tree modeling and risk management. In: Kallrath, J., Pardalos, P. (eds.) Power Systems Handbook. Springer (to appear)

[ER05] Eichhorn, A., Römisch, W.: Polyhedral risk measures in stochastic programming. SIAM J. Optim. **16**, 69–95 (2005)

[ER07] Eichhorn, A., Römisch, W.: Stochastic integer programming: Limit theorems and confidence intervals. Math. Oper. Res. **32**, 118–135 (2007)

[ERW05] Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In: IEEE St. Petersburg PowerTech Proceedings (2005)

[EG+07] Escudero, L.F., Garín, A., Merino, M., Pérez, G.: A two-stage stochastic integer programming approach as a mixture of branch-and-fix coordination and Benders decomposition schemes. Ann. Oper. Res. **152**, 395–420 (2007)

[GK+02] Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In: Greengard, C., Ruszczyński, A. (eds.) Decision Making under Uncertainty: Energy and Power, pp. 39–70. Springer, New York (2002).

[GR05] Gröwe-Kuska, N., Römisch, W.: Stochastic unit commitment in hydro-thermal power production planning. In Wallace, S.W., Ziemba, W.T. (eds.) Applications of Stochastic Programming, pp. 633–653. MPS/SIAM Series on Optimization, SIAM, Philadelphia (2005)

[HR07] Heitsch, H., Römisch, W.: A note on scenario reduction for two-stage stochastic programs. Oper. Res. Lett. **35**, 731–738 (2007)

[HKR08] Henrion, R., Küchler, C., Römisch, W.: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming. J. Ind. Manag. Optim. **4**, 363–384 (2008)

[HKR09] Henrion, R., Küchler, C., Römisch, W.: A scenario reduction heuristic for two-stage stochastic integer programs. in preparation

[Ki90] Kiwiel, K.C.: Proximity control in bundle methods for convex nondifferentiable optimization. Math. Progr. **46**, 105–122 (1990)

[KSV06] Klein Haneveld, W.K., Stougie, L., van der Vlerk, M.H.: Simple integer recourse models: convexity and convex approximations. Math. Progr. **108**(2-3, Ser. B), 435–473 (2006)

[KV07] Küchler, C., Vigerske, S.: Decomposition of multistage stochastic programs with recombining scenario trees. Stoch. Progr. E-Print Series **9** (2007). `www.speps.org`

[LaL93] Laporte, G., Louveaux, F.V.: The integer *L*-shaped method for stochastic integer programs with complete recourse. Oper. Res. Lett. **13**(3), 133–142 (1993)

[LoS03] Louveaux, F.V., Schultz, R.: Stochastic Integer Programming. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, pp. 213–266. Handbooks in Operations Research and Management Science Vol. 10. Elsevier (2003)

[LuS04] Lulli, G., Sen, S.: A branch and price algorithm for multi-stage stochastic integer programs with applications to stochastic lot sizing problems. Manage. Sci. **50**, 786–796 (2004)

[Ni92] Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)

[NR00] Nowak, M.P., Römisch, W.: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. Ann. Oper. Res. **100**, 251–272 (2000)

[NSW05] Nowak, M.P., Schultz, R., Westphalen, M.: A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment. Optim. Eng. **6**, 163–176 (2005)

[NS05] Ntaimo, L., Sen, S.: The million-variable "march" for stochastic combinatorial optimization. J. Glob. Optim. **32**(3), 385–400 (2005)

[NS08a] Ntaimo, L., Sen, S.: A comparative study of decomposition algorithms for stochastic combinatorial optimization. Comput. Optim. Appl. **40**(3), 299–319 (2008)

[NS08b] Ntaimo, L., Sen, S.: A branch-and-cut algorithm for two-stage stochastic mixed-binary programs with continuous first-stage variables. Int. J. Comp. Sci. Eng. **3**, 232–241 (2008)

[NR02] Nürnberg, R., Römisch, W.: A two-stage planning model for power scheduling in a hydro-thermal system under uncertainty. Optim. Eng. **3**, 355–378 (2002)

[PCW00] Philpott, A.B., Craddock, M., Waterer, H.: Hydro-electric unit commitment subject to uncertain demand. Eur. J. Oper. Res. **125**, 410–424 (2000)

[Ra91] Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, Chichester (1991)

[RS01] Römisch, W., Schultz, R.: Multistage stochastic integer programs: An introduction. In: Grötschel, M., Krumke, S.O., J. Rambau, J. (eds.) Online Optimization of Large Scale Systems, pp. 581–600. Springer, Berlin (2001)

[RV08] Römisch, W., Vigerske, S.: Quantitative stability of fully random mixed-integer two-stage stochastic programs. Optim. Lett. **2**, 377–388 (2008)

[Sch96] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse. SIAM J. Optim. **6**, 1138–1152 (1996)

[Sch03] Schultz, R.: Stochastic programming with integer variables. Math. Progr. **97**, 285–309 (2003)

[SSV98] Schultz, R., Stougie, L., van der Vlerk, M.H.: Solving stochastic programs with integer recourse by enumeration: a framework using Gröbner basis reductions. Math. Progr. **83**(2, Ser. A), 229–252 (1998)

[ST06] Schultz, R., Tiedemann, S.: Conditional value-at-risk in stochastic programs with mixed-integer recourse. Math. Progr. **105**, 365–386 (2006)

[Se05] Sen, S.: Algorithms for stochastic mixed-integer programming models. In: Aardal, K., Nemhauser, G.L., Weismantel, R. (eds.) Handbook of Discrete Optimization, pp. 515–558. North-Holland Publishing Co. (2005)

[SYG06] Sen, S., Yu, L., Genc, T.: A stochastic programming approach to power portfolio optimization. Oper. Res. **54**, 55–72 (2006)

[SeH05] Sen, S., Higle, J.L.: The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: set convexification. Math. Progr. **104**(A), 1–20 (2005)

[SeS06] Sen, S., Sherali, H.D.: Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. Math. Progr. **106**(2, Ser. A), 203–223 (2006)

[TBL96] Takriti, S., Birge, J.R., Long, E.: A stochastic model for the unit commitment problem. IEEE Trans. Power Systems **11**, 1497–1508 (1996)

[TKW00] Takriti, S., Krasenbrink, B., Wu, L.S.-Y.: Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. Oper. Res. **48**, 268–280 (2000)

[TiW81] Tind, J., Wolsey, L.A.: An elementary survey of general duality theory in mathematical programming. Math. Progr. **21**, 241–261 (1981)

[vBL87] van den Bosch, P.P.J., Lootsma, F.A.: Scheduling of power generation via large-scale nonlinear optimization. J. Optim. Theory Appl. **55**, 313–326 (1987)

[vdV04] van der Vlerk, M.H.: Convex approximations for complete integer recourse models. Math. Progr. **99**(2, Ser. A), 297–310 (2004)

[vdV07] van der Vlerk, M.H.: Stochastic integer programming bibliography. `http://mally.eco.rug.nl/spbib.html` (1996-2007)

[vdV05] van der Vlerk, M.H.: Convex approximations for a class of mixed-integer recourse models. Ann. Oper. Res. (to appear) and Stoch. Progr. E-Print Series (2005)

[VSW69] Van Slyke, R.M., Wets, R.: L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming. SIAM J. Appl. Math. **17**(4), 638–663 (1969)

[Wo81] Wolsey, L.A.: Integer programming duality: Price functions and sensitivity analysis. Math. Progr. **20**, 173–195 (1981)

[ZG88] Zhuang, G., Galiana, F.D.: Towards a more rigorous and practical unit commitment by Lagrangian relaxation. IEEE Trans. Power Systems **3**, 763–773 (1988)