

Chapter 4

Stochastic Integer Programming

François V. Louveaux

*Département des Méthodes Quantitatives, Facultés Universitaires Notre Dame de la Paix,
Rempart de la Vierge 8, B-5000 Namur, Belgium, E-mail: flouveaux@fundp.ac.be*

Rüdiger Schultz

*Institut für Mathematik, Universität Duisburg-Essen, Lotharstraße 65, D-47048 Duisburg,
Germany, E-mail: schultz@math.uni-duisburg.de*

Abstract

When introducing integer variables into traditional linear stochastic programs structural properties and algorithmic approaches have to be rethought from the very beginning. Employing basics from parametric integer programming and probability theory we analyze the structure of stochastic integer programs. In the algorithmic part of the paper we review solution techniques from integer programming and discuss their impact on the specialized structures met in stochastic programming.

Key words: Stochastic integer programs, mixed-integer recourse, simple integer recourse, multi-stage models, decomposition schemes, cutting planes, Lagrangian relaxation, integer L -shaped algorithm, sampling methods.

1 Introduction

Like in other branches of mathematical optimization, integer variables are indispensable in many stochastic programming models. Integrality either may occur explicitly via indivisibles or Boolean decisions. Or it may occur in implicit fashion as a modelling tool, for instance when handling disjunctions and discontinuities, or when dealing with nonconvex piecewise linearity.

As an example let us reconsider the popular newsboy problem: a newsboy can purchase from a publisher a number of newspapers with purchase costs

c_i and selling costs q_i , $i = 1, \dots, s$. Each newspaper has a weight a_i , and a total weight b is available to the newsboy. The demand for newspapers, which is unknown to the newsboy at the time of purchase, is described by a random vector $h(\omega) \in \mathbb{R}^s$ on some probability space (Ω, \mathcal{F}, P) . Unsold newspapers cannot be returned to the publisher. The newsboy faces the problem to purchase from the publisher in such a way that the profit after selling is maximal. Maximizing the expected profit leads to the stochastic program

$$\min\{c^T x + \mathbb{E}_P[Q(x, h(\omega))]: 0 \leq a^T x \leq b\} \quad (1.1)$$

where

$$Q(x, h(\omega)) = \min\{-q^T y: 0 \leq y \leq h(\omega), y \leq x\}. \quad (1.2)$$

The newsboy problem is a specimen from a class of cost minimization (or profit maximization) problems where planning decisions must be taken before and operational decisions are taken after observation of a random demand. It is quite common to model these problems under the assumption that the commodities involved are infinitely divisible. As in (1.1), (1.2) this leads to decision variables in the real numbers. However, when dealing with indivisibles, as the newsboy obviously does, the precise modeling would require integer variables. The model (1.1), (1.2) then has to be supplemented by the conditions $x \in \mathbb{Z}_+^s$ and $y \in \mathbb{Z}_+^s$.

Now let us assume that, for renting a newsstand, for instance, the newsboy faces fixed charge costs f when selling the newspapers. Then, disregarding divisibility or indivisibility of goods, the discontinuity of the cost structure leads to an integer variable in (1.2):

$$Q(x, h(\omega)) = \min\{-q^T y + fu: 0 \leq y \leq h(\omega)u, y \leq x, u \in \{0, 1\}\}.$$

In the simple examples discussed above integrality, of course, does not pose a real challenge. This changes drastically if the above modeling paradigms occur in more complex situations. In principle, any integer or mixed-integer programming model arising in operations research may be affected by randomness and hence give rise to a stochastic programming extension.

In the present chapter we will see that the appearance of integer requirements has substantial structural and algorithmic consequences for stochastic programming models as discussed in Chapters 1 and 2, for instance.

2 Structural properties

2.1 Two-stage models with complete integer recourse

Introducing integer requirements into the two-stage stochastic linear program from Section 2 of Chapter 2 leads to the optimization problem

$$\min_x \{c^T x + \mathbb{E}_P[Q(x, \xi(\omega))]: x \in X\} \quad (2.3)$$

where $Q(x, \xi(\omega))$ is the optimal value of the second-stage problem

$$\min_{(y, y')} \{q^T y + q'^T y': Tx + Wy + W'y' = h(\omega), y \in \mathbb{Z}_+^{\bar{n}_2}, y' \in \mathbb{R}_+^{n'_2}\}. \quad (2.4)$$

We assume that all ingredients above have conformable dimensions, that W, W' are rational matrices, and that $X \subseteq \mathbb{R}^{n_1}$ is a nonempty closed polyhedron, possibly involving integrality constraints on components of the vector x . For ease of exposition, q, q', W, W', T are deterministic, such that the random variable $\xi(\omega) = h(\omega)$, living on some probability space (Ω, \mathcal{F}, P) , is the only stochastic ingredient. So far, most of the subsequent results on stochastic integer programs were obtained for that special case. Where appropriate, we will point to existing results under more general randomness.

As with the two-stage stochastic linear program, the second-stage value function $\Phi: \mathbb{R}^{m_2} \rightarrow \mathbb{R}$ with

$$\Phi(t) := \min_{(y, y')} \{q^T y + q'^T y': Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{n}_2}, y' \in \mathbb{R}_+^{n'_2}\} \quad (2.5)$$

is of prime importance for the structure of (2.3). Recall that, without integer requirements, Φ is convex on its domain of finiteness. To obtain a first impression on the impact of integrality let us start with two illustrative examples. They are both derived from

$$\Phi(t) = \min\{y^+ + y^-: y^+ - y^- = t, y^+ \in \mathbb{R}_+, y^- \in \mathbb{R}_+\}. \quad (2.6)$$

This is the classical simple-recourse situation. By duality one immediately obtains that $\Phi(t) = |t|$.

Now we add another variable v which we require to be integral:

$$\begin{aligned}\Phi(t) &= \min \left\{ \frac{1}{2} v + y^+ + y^- : v + y^+ - y^- = t, v \in \mathbb{Z}_+, y^+ \in \mathbb{R}_+, y^- \in \mathbb{R}_+ \right\} \\ &= \min \left\{ \frac{1}{2} v + |t - v| : v \in \mathbb{Z}_+ \right\}.\end{aligned}$$

Understanding $v \in \mathbb{Z}_+$ as indices, the function Φ can be seen as the pointwise minimum of staggered absolute values. Hence, Φ is no longer convex but still continuous, even Lipschitz continuous on \mathbb{R} .

The second example is basically derived by turning y^+ and y^- into integer variables:

$$\begin{aligned}\Phi(t) &= \min \{ v^+ + v^- : y + v^+ - v^- = t, y \in \mathbb{R}_+, v^+ \in \mathbb{Z}_+, v^- \in \mathbb{Z}_+ \} \\ &= \begin{cases} 0, & t \geq 0 \\ -t, & t < 0. \end{cases}\end{aligned}$$

Here $\lceil \cdot \rceil$ denotes the integer round-up operation. Obviously, Φ is no longer continuous, but still lower semicontinuous. Moreover, discontinuities occur in a set of Lebesgue measure zero only, and “jump heights” at discontinuities are globally bounded.

From these observations it becomes quite clear that, with stochastic integer programs, basic properties like convexity and duality, that are so important in the purely linear case, cannot be maintained for reasonable problem classes. Let us now study the mixed-integer value function in more detail.

Proposition 1. *Suppose that the recourse in (2.3)–(2.4) is complete, i.e.,*

$$W(\mathbb{Z}_+^{\bar{n}_2}) + W'(\mathbb{R}_+^{n'_2}) = \mathbb{R}^{m_2}, \quad (2.7)$$

and that the LP relaxation to the second-stage problem has a feasible dual, i.e.,

$$\{u \in \mathbb{R}^{m_2} : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset, \quad (2.8)$$

then the value function $\Phi(\cdot)$ is well-defined on \mathbb{R}^{m_2} .

Proof. Let $t \in \mathbb{R}^{m_2}$. By (2.7), the mixed-integer program defining $\Phi(t)$ is feasible. By the existence theorem of mixed-integer linear programming this problem then is solvable provided it is bounded. The latter, however, follows from the solvability of the LP relaxation to the mixed-integer program defining $\Phi(t)$ which is a consequence of the primal feasibility implied by (2.7) and the dual feasibility in (2.8). \square

Before proceeding further, let us have a quick look at the linear-programming counterpart to Φ that has been studied in more detail in Chapter 2:

$$\Phi_{\text{lin}}(t) := \min\{q'^T y' : W' y' = t, y' \in \mathbb{R}_+^{n'_2}\}. \quad (2.9)$$

If we assume that $W'(\mathbb{R}_+^{n'_2})$ is full-dimensional and that

$$\{u \in \mathbb{R}^{m_2} : W'^T u \leq q'\} \neq \emptyset,$$

then the latter set has vertices d_k , $k=1, \dots, K$, and it holds by linear programming duality that

$$\Phi_{\text{lin}}(t) = \max\{t^T u : W'^T u \leq q'\} = \max_{k=1, \dots, K} d_k^T t \quad \text{for all } t \in W'(\mathbb{R}_+^{n'_2}).$$

Hence, Φ_{lin} is convex and piecewise linear on its (conical) domain of definition.

Imposing the basic assumptions (2.7) and (2.8) we obtain

$$\begin{aligned} \Phi(t) &= \min\{q^T y + q'^T y' : Wy + W'y' = t, y \in \mathbb{Z}_+^{\bar{n}_2}, y' \in \mathbb{R}_+^{n'_2}\} \\ &= \min_y \{q^T y + \min_{y'} \{q'^T y' : W'y' = t - Wy, y' \in \mathbb{R}_+^{n'_2}\} : y \in \mathbb{Z}_+^{\bar{n}_2}\} \\ &= \min_y \{\Phi_y(t) : y \in \mathbb{Z}_+^{\bar{n}_2}\}, \end{aligned} \quad (2.10)$$

where

$$\Phi_y(t) = q^T y + \max_{k=1, \dots, K} d_k^T (t - Wy) \quad \text{for all } t \in Wy + W'(\mathbb{R}_+^{n'_2}).$$

Here, d_k , $k=1, \dots, K$ are the vertices of $\{u \in \mathbb{R}^{m_2} : W'^T u \leq q'\}$, and we have applied the argument about Φ_{lin} from the purely linear case. For $t \notin Wy + W'(\mathbb{R}_+^{n'_2})$ the problem $\min_{y'} \{q'^T y' : W'y' = t - Wy, y' \in \mathbb{R}_+^{n'_2}\}$ is infeasible, and we put $\Phi_y(t) = +\infty$. It is convenient to introduce the notation $Y(t) := \{y \in \mathbb{Z}_+^{\bar{n}_2} : \Phi_y(t) < +\infty\}$.

According to (2.10) the value function Φ is made up by the pointwise minimum of a family of convex, piecewise linear functions whose domains of definition are polyhedral cones arising as shifts of the cone $W'(\mathbb{R}_+^{n'_2})$. By our basic assumption $W(\mathbb{Z}_+^{\bar{n}_2}) + W'(\mathbb{R}_+^{n'_2}) = \mathbb{R}^{m_2}$, the cone $W'(\mathbb{R}_+^{n'_2})$ is full-dimensional.

Some first conclusions about the continuity of Φ may be drawn from the above observations:

- (1) Suppose that $t \in \mathbb{R}^{m_2}$ does not belong to any boundary of any of the sets $Wy + W'(\mathbb{R}_+^{n_2})$, $y \in \mathbb{Z}^{\bar{n}_2}$. Then the same is true for all points τ in some open ball B around t . Hence, $Y(\tau) = Y(t)$ for all $\tau \in B$. With an enumeration $(y_\nu)_{\nu \in \mathbb{N}}$ of $Y(t)$ we consider the functions $\Phi^\kappa(\tau) := \min\{\Phi_{y_\nu}(\tau) : \nu \leq \kappa\}$ for all $\tau \in B$. Then $\lim_{\kappa \rightarrow \infty} \Phi^\kappa(\tau) = \Phi(\tau)$ for all $\tau \in B$. Since, for any function Φ_y , its “slopes” are determined by the same, finitely many vectors d_k , $k = 1, \dots, K$, the functions Φ^κ , $\kappa \in \mathbb{N}$ are all Lipschitz continuous on B with a uniform Lipschitz constant. Thus, the family of functions Φ^κ , $\kappa \in \mathbb{N}$ is equicontinuous on B and has a pointwise limit there. Consequently, this pointwise limit Φ is continuous on B , in fact Lipschitz continuous with the mentioned uniform constant.
- (2) Any discontinuity point of Φ must be located at the boundary of some set $Wy + W'(\mathbb{R}_+^{n_2})$, $y \in \mathbb{Z}^{\bar{n}_2}$. Hence, the set of discontinuity points of Φ is contained in a countable union of hyperplanes. Since $W'(\mathbb{R}_+^{n_2})$ has only finitely many facets, this union of hyperplanes subdivides into finitely many classes, such that, in each class, the hyperplanes are parallel. By the rationality of the matrices W and W' , within each class, the pairwise distance of the hyperplanes is uniformly bounded below by some positive number.
- (3) Let $t_\nu \rightarrow t$ and $y \in \mathbb{Z}^{\bar{n}_2}$ such that $t_\nu \in Wy + W'(\mathbb{R}_+^{n_2})$ for all sufficiently large ν . Since the set $Wy + W'(\mathbb{R}_+^{n_2})$ is closed, this yields $t \in Wy + W'(\mathbb{R}_+^{n_2})$. Therefore, for sufficiently large ν , $Y(t_\nu) \subseteq Y(t)$. This paves the way for showing that $\liminf_{t_\nu \rightarrow t} \Phi(t_\nu) \geq \Phi(t)$, which is the lower semicontinuity of Φ at t .

The above analysis can be extended into the following result that dates back to a series of papers by Blair and Jeroslow, out of which we refer to [Blair and Jeroslow \(1977\)](#), and to the monographs [Bank et al. \(1982\)](#) and [Bank and Mandel \(1988\)](#).

Proposition 2. *Let W, W' be matrices with rational entries and assume that $W(\mathbb{Z}_+^{\bar{n}_2}) + W'(\mathbb{R}_+^{n_2}) = \mathbb{R}^{m_2}$ as well as $\{u \in \mathbb{R}^{m_2} : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$. Then it holds*

- (1) Φ is real-valued and lower semicontinuous on \mathbb{R}^{m_2} ,
- (2) there exists a countable partition $\mathbb{R}^{m_2} = \bigcup_{i=1}^\infty \mathcal{T}_i$ such that the restrictions of Φ to \mathcal{T}_i are piecewise linear and Lipschitz continuous with a uniform constant $L > 0$ not depending on i ,
- (3) each of the sets \mathcal{T}_i has a representation $\mathcal{T}_i = \{t_i + \mathcal{K}\} \setminus \bigcup_{j=1}^N \{t_{ij} + \mathcal{K}\}$ where \mathcal{K} denotes the polyhedral cone $W'(\mathbb{R}_+^{n_2})$ and t_i, t_{ij} are suitable points from \mathbb{R}^{m_2} , moreover, N does not depend on i ,
- (4) there exist positive constants β, γ such that $|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma$ whenever $t_1, t_2 \in \mathbb{R}^{m_2}$.

Let us now consider the expected value function

$$\phi(x) := \mathbb{E}_P[Q(x, \xi(\omega))] = \mathbb{E}_P[\Phi(\xi(\omega) - Tx)] = \mathbb{E}_\mu[\Phi(\xi - Tx)] \quad (2.11)$$

where μ denotes the image measure $P \circ \xi^{-1}$ on \mathbb{R}^{m_2} . Thanks to the lower semicontinuity in part (1) of [Proposition 2](#), the integrand $Q(x, \cdot)$ is measurable for all $x \in \mathbb{R}^{n_1}$. Moreover, lower semicontinuity and continuity are inherited according to the general principles displayed in [Proposition 14](#) of Chapter 1 and [Proposition 1](#) of Chapter 2. This is made precise in the following proposition.

Proposition 3. *Suppose that: (i) W, W' are rational, (ii) $W(\mathbb{Z}_+^{n_2}) + W'(\mathbb{R}_+^{n'_2}) = \mathbb{R}^{m_2}$, (iii) $\{u \in \mathbb{R}^{m_2} : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$, (iv) $\mathbb{E}_\mu[\|\xi\|] < \infty$. Then the expected value function ϕ is lower semicontinuous on \mathbb{R}^{n_1} . If, moreover, $\mu(E(x)) = 0$ where $E(x) := \{\xi \in \mathbb{R}^{m_2} : \Phi \text{ is discontinuous at } \xi - Tx\}$, then ϕ is continuous at x .*

Proof. Let $x \in \mathbb{R}^{n_1}$ and $x_k \rightarrow x$. Denote $r := \max_{k \in \mathbb{N}} \|x_k\|$. Assumptions (ii) and (iii) in particular imply that $\Phi(0) = 0$. Part (4) of [Proposition 2](#) then yields the estimate

$$\begin{aligned} \Phi(\xi - Tx_k) &\geq \Phi(0) - |\Phi(\xi - Tx_k) - \Phi(0)| \geq -\beta\|\xi - Tx_k\| - \gamma \\ &\geq -\beta\|\xi\| - \beta r\|T\| - \gamma \end{aligned}$$

such that, together with (iv), the function $-\beta\|\cdot\| - \beta r\|T\| - \gamma$ provides an integrable minorant for the functions $\Phi(\cdot - Tx_k)$, $k \in \mathbb{N}$. Fatou's Lemma and the lower semicontinuity of Φ now imply

$$\begin{aligned} \phi(x) &= \int_{\mathbb{R}^{m_2}} \Phi(\xi - Tx) \mu(d\xi) \leq \int_{\mathbb{R}^{m_2}} \liminf_{k \rightarrow \infty} \Phi(\xi - Tx_k) \mu(d\xi) \\ &\leq \liminf_{k \rightarrow \infty} \int_{\mathbb{R}^{m_2}} \Phi(\xi - Tx_k) \mu(d\xi) = \liminf_{k \rightarrow \infty} \phi(x_k) \end{aligned}$$

which proves the desired lower semicontinuity. For proving continuity we again employ part (4) of [Proposition 2](#) together with $\Phi(0) = 0$. We obtain the estimate

$$|\Phi(\xi - Tx_k)| = |\Phi(\xi - Tx_k) - \Phi(0)| \leq \beta\|\xi\| + \beta r\|T\| + \gamma$$

which, in view of (iv), provides us with an integrable majorant for the functions $|\Phi(\cdot - Tx_k)|$, $k \in \mathbb{N}$. By $\mu(E(x)) = 0$, we have

$$\lim_{k \rightarrow \infty} \Phi(\xi - Tx_k) = \Phi(\xi - Tx) \quad \text{for } \mu\text{-almost all } \xi \in \mathbb{R}^{m_2}.$$

Now Lebesgue's Dominated Convergence Theorem completes the proof:

$$\lim_{k \rightarrow \infty} \phi(x_k) = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^{m_2}} \Phi(\xi - Tx_k) \mu(d\xi) = \int_{\mathbb{R}^{m_2}} \Phi(\xi - Tx) \mu(d\xi) = \phi(x).$$

□

Recall that, by conclusion (1.2) in front of [Proposition 2](#), the set of discontinuity points of Φ , and hence $E(x)$, is contained in a countable union of hyperplanes, which is a set of Lebesgue measure zero. Assume that μ has a density and that (i)–(iv) are fulfilled. Then $\mu(E(x)) = 0$ for all $x \in \mathbb{R}^{n_1}$, and ϕ is continuous on \mathbb{R}^{n_1} . The next statement addresses Lipschitz continuity. To avoid further technicalities, we formulate the result for the case that μ has a density.

Proposition 4. *Suppose that: (i) q, q', W, W' all have rational entries, (ii) $W(\mathbb{Z}_+^{m_2}) + W'(\mathbb{R}_+^{m_2}) = \mathbb{R}^{m_2}$, (iii) $\{u \in \mathbb{R}^{m_2} : W^T u \leq q, W'^T u \leq q'\} \neq \emptyset$, (iv) $\mathbb{E}_\mu[\|\xi\|] < \infty$, (v) μ has a density and for any nonsingular linear transformation $B \in L(\mathbb{R}^{m_2}, \mathbb{R}^{m_2})$ all one-dimensional marginal distributions of $\mu \circ B$ have bounded densities which, outside some bounded interval, are monotonically decreasing with growing absolute value of the argument. Then ϕ is Lipschitz continuous on any bounded subset of \mathbb{R}^{n_1} .*

Proof. Let x', x'' belong to some bounded subset D of \mathbb{R}^{n_1} . Denote

$$S(x', x'') := \bigcup_{i \in \mathbb{N}} ((Tx' + \mathcal{T}_i) \cap (Tx'' + \mathcal{T}_i))$$

where \mathcal{T}_i are as in part (2) of [Proposition 2](#). Note that for all $\xi \in S(x', x'')$ we then have the estimate

$$|\Phi(\xi - Tx') - \Phi(\xi - Tx'')| \leq L \cdot \|Tx' - Tx''\|. \quad (2.12)$$

Now

$$\begin{aligned} |\phi(x') - \phi(x'')| &\leq \int_{S(x', x'')} |\Phi(\xi - Tx') - \Phi(\xi - Tx'')| \mu(d\xi) \\ &\quad + \int_{\mathbb{R}^{m_2} \setminus S(x', x'')} |\Phi(\xi - Tx') - \Phi(\xi - Tx'')| \mu(d\xi). \end{aligned}$$

Employing (2.12) and part (4) of [Proposition 2](#) gives the estimate

$$|\phi(x') - \phi(x'')| \leq (L + \beta) \cdot \|T\| \cdot \|x' - x''\| + \gamma \cdot \mu(\mathbb{R}^{m_2} \setminus S(x', x'')).$$

We complete the proof by deducing a Lipschitz estimate for the second term on the right. Let H_k , $k = 1, \dots, K$ be the hyperplanes containing the facets of the cone $W'(\mathbb{R}_+^{n_2})$ that arises in part (3) of [Proposition 2](#) and is full-dimensional by (ii). The set $\mathbb{R}^{m_2} \setminus S(x', x'')$ then is contained in a finite union of sets \mathcal{H}_k , $k = 1, \dots, K$ each of which is a countable union of “sandwiches” $\mathcal{H}_{k,i}$, $i \in \mathbb{N}$. Each “sandwich” $\mathcal{H}_{k,i}$ is the region in between and including the affine hyperplanes

$$t_i + Tx' + H_k \quad \text{and} \quad t_i + Tx'' + H_k$$

where t_i , $i \in \mathbb{N}$, are as in part (3) of [Proposition 2](#). By a nonsingular linear transformation B_k we map the hyperplane H_k to the hyperplane which is orthogonal to the first coordinate vector. Then $\mu(\mathcal{H}_k)$ can be estimated using the marginal density θ_k of the first component with respect to the image measure $\mu \circ B_k^{-1}$:

$$\mu(\mathcal{H}_k) \leq c_1 \cdot \sum_{i \in \mathbb{N}} \int_{\tau_{i,k}(x')}^{\tau_{i,k}(x'')} \theta_k(\tau) d\tau.$$

Here, $c_1 > 0$ is some constant, and $\tau_{i,k}(x')$, $\tau_{i,k}(x'')$ are the first components of $B_k(t_i + Tx')$ and $B_k(t_i + Tx'')$, respectively. Without loss of generality we assume that $\tau_{i,k}(x') < \tau_{i,k}(x'')$. Clearly, there exists a constant $c_2 > 0$ such that

$$\tau_{i,k}(x'') - \tau_{i,k}(x') \leq c_2 \cdot \|x' - x''\|. \quad (2.13)$$

Using the rationality of q , q' , W , W' one can show that the sequence $(\tau_{i,k}(x'))_{i \in \mathbb{N}}$ has no accumulation points. Since x' , x'' belong to the bounded set D , there exists an index $\bar{i} = \bar{i}(D)$, independent of x' , x'' , such that the intervals $[\tau_{i,k}(x'), \tau_{i,k}(x'')]$ (up to renumbering) meet the bounded interval from assumption (v) at most of $i \leq \bar{i}$.

According to assumption (v) we have an upper bound $\bar{\theta}_k$ for $\theta_k(\cdot)$. For $i > \bar{i}$, we denote by $\tilde{\tau}_{i,k}$ the left or right endpoint of $[\tau_{i,k}(x'), \tau_{i,k}(x'')]$ depending on whether θ_k is decreasing or increasing on that interval. This allows the estimate

$$\begin{aligned} \sum_{i \in \mathbb{N}} \int_{\tau_{i,k}(x')}^{\tau_{i,k}(x'')} \theta_k(\tau) d\tau &\leq \sum_{i \leq \bar{i}} \bar{\theta}_k \cdot (\tau_{i,k}(x'') - \tau_{i,k}(x')) \\ &\quad + \sum_{i > \bar{i}} \theta_k(\tilde{\tau}_{i,k}) \cdot (\tau_{i,k}(x'') - \tau_{i,k}(x')). \end{aligned}$$

In view of (2.13), there exists a constant $c_3 > 0$ such that the first sum on the right is estimated above by $c_3 \|x' - x''\|$. For the second sum we obtain the upper estimate $c_2 \sum_{i > \bar{i}} \theta_k(\tilde{\tau}_{i,k}) \cdot \|x' - x''\|$. It remains to show that $\sum_{i > \bar{i}} \theta_k(\tilde{\tau}_{i,k})$ is finite.

Let us do so for the sum over all $i > \bar{i}$ belonging to those $\tilde{\tau}_{i,k}$ around which θ_k is decreasing. For the remaining $i > \bar{i}$ a similar argument applies. Since the $\tilde{\tau}_{i,k}$ do not accumulate, it holds with some $\varepsilon > 0$

$$1 \geq \sum_i \int_{\tilde{\tau}_{i,k}-\varepsilon}^{\tilde{\tau}_{i,k}} \theta_k(\tau) d\tau \geq \sum_i \int_{\tilde{\tau}_{i,k}-\varepsilon}^{\tilde{\tau}_{i,k}} \theta_k(\tilde{\tau}_{i,k}) d\tau = \varepsilon \cdot \sum_i \theta_k(\tilde{\tau}_{i,k}),$$

providing the desired finiteness. Repeating the above arguments for all \mathcal{H}_k , $k = 1, \dots, K$, one confirms that there exists a constant $c > 0$ such that

$$\mu(\mathbb{R}^{m_2} \setminus S(x', x'')) \leq \sum_{k=1}^K \mu(\mathcal{H}_k) \leq c \cdot \|x' - x''\|,$$

and the proof is complete. \square

The following examples show that the boundedness and the monotonicity in assumption (v) of the above proposition are indispensable.

Example 5. Let $\Phi(t) = \min\{y : y \geq t, y \in \mathbb{Z}\} = \lceil t \rceil$ and μ be given by the (unbounded) density θ , for which $\theta(\tau) = \tau^{-1/2}$ if $0 < \tau \leq 1/4$ and $\theta(\tau) = 0$, otherwise. Then $\phi(x) = 1$ for $-3/4 \leq x \leq 0$ and $\phi(x) = 1 - 2\sqrt{x}$ if $0 \leq x \leq 1/4$. This function is not Lipschitz continuous on neighborhoods of $x_0 = 0$.

Example 6. Let $\Phi(t) = \min\{y : y \geq t, y \in \mathbb{Z}\} = \lceil t \rceil$ and μ be given by the density θ , for which $\theta(\tau) = 1/\nu$ if $\tau \in [\nu, \nu + \frac{1}{\nu^2} \cdot c]$, $c := (\sum_{k=1}^{\infty} 1/k^3)^{-1}$, $\nu \in \mathbb{N}$, and $\theta(\tau) = 0$, otherwise. This density violates the monotonicity assumption in

Proposition 4. We show that ϕ is not Lipschitz continuous on neighborhoods of $x_0 = 0$.

Assumptions (i)–(iii) of **Proposition 4** are clearly met. Moreover, θ is bounded and assumption (iv) is fulfilled, since

$$\int_{-\infty}^{+\infty} \tau \theta(\tau) \, d\tau \leq \sum_{v=1}^{\infty} (v+1) \cdot \frac{1}{v} \cdot \frac{1}{v^2} \cdot c = c \frac{\pi^2}{6} + 1.$$

For arbitrary $x \in \mathbb{R}$, $0 < x < 1$, it holds

$$\phi(0) - \phi(x) = \sum_{v=1}^{\infty} \int_v^{v+x} \theta(\tau) \, d\tau \geq \sum_{v=1}^{\bar{v}(x)} \frac{1}{v} \cdot x,$$

where $\bar{v}(x) := \lfloor \sqrt{\frac{c}{x}} \rfloor$. Consider $x_k := \frac{1}{k^2} \cdot c$, $k \in \mathbb{N}$. The above yields

$$\frac{1}{x_k} (\phi(0) - \phi(x_k)) \geq \sum_{v=1}^k \frac{1}{v}.$$

For $k \rightarrow \infty$, the left-hand side tends to infinity, proving that ϕ is not Lipschitz continuous on neighborhoods of $x_0 = 0$.

When studying the stability behaviour of the stochastic program (2.3) with respect to perturbations of the underlying probability measure, see the chapter “Stability of Stochastic Programming Problems” for an exposition of stability analysis, it is crucial to detect the continuity of the expected recourse function ϕ , jointly in the decision vector x and the probability measure μ . As a prerequisite, then a suitable convergence notion for probability measures is needed. Here, weak convergence of probability measures has proven both sufficiently general to cover relevant applications and sufficiently specific to enable substantial results. A sequence $(\mu_k)_{k \in \mathbb{N}}$ in the space $\mathcal{P}(\mathbb{R}^{m_2})$ of Borel probability measures on \mathbb{R}^{m_2} is said to converge weakly to $\mu \in \mathcal{P}(\mathbb{R}^{m_2})$, written $\mu_k \xrightarrow{w} \mu$, if for any bounded continuous function $g : \mathbb{R}^{m_2} \rightarrow \mathbb{R}$ we have

$$\int_{\mathbb{R}^{m_2}} g(\xi) \mu_k(d\xi) \rightarrow \int_{\mathbb{R}^{m_2}} g(\xi) \mu(d\xi) \quad \text{as } k \rightarrow \infty. \quad (2.14)$$

For establishing joint continuity of $\phi = \phi(x, \mu)$ a theorem on weak convergence of image measures attributed to Rubin will be very useful. This theorem says: Let g_k, g ($k \in \mathbb{N}$) be measurable functions from \mathbb{R}^{m_2} to \mathbb{R} and

denote $E := \{\xi \in \mathbb{R}^{m_2} : \exists \xi_k \rightarrow \xi \text{ such that } g_k(\xi_k) \not\rightarrow g(\xi)\}$. If $\mu_k \xrightarrow{w} \mu$ and $\mu(E) = 0$, then $\mu_k \circ g_k^{-1} \xrightarrow{w} \mu \circ g^{-1}$.

Proposition 7. Fix arbitrary $p > 1$ and $K > 0$, and denote $\Delta_{p,K}(\mathbb{R}^{m_2}) := \{\nu \in \mathcal{P}(\mathbb{R}^{m_2}) : \mathbb{E}_\nu[\|\xi\|^p] \leq K\}$. Let $\mu \in \Delta_{p,K}(\mathbb{R}^{m_2})$ be such that $\mu(E(x)) = 0$. Then the function $\phi : \mathbb{R}^{m_1} \times \Delta_{p,K}(\mathbb{R}^{m_2}) \rightarrow \mathbb{R}$ is continuous at (x, μ) .

Proof. Let $x_k \rightarrow x$ in \mathbb{R}^{m_1} and $\mu_k \xrightarrow{w} \mu$ in $\Delta_{p,K}(\mathbb{R}^{m_2})$. Introduce measurable functions $g_k, k \in \mathbb{N}$, and g by $g_k(\xi) := \Phi(\xi - Tx_k)$ and $g(\xi) := \Phi(\xi - Tx)$. For the corresponding exceptional set E a simple continuity argument provides $E(x)^c \subseteq E^c$ or, equivalently, $E \subseteq E(x)$. Hence, $\mu(E) = 0$, and Rubin's Theorem yields

$$\mu_k \circ g_k^{-1} \xrightarrow{w} \mu \circ g^{-1}. \quad (2.15)$$

Changing variables in the assertion

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^{m_2}} g_k(\xi) \mu_k(d\xi) = \int_{\mathbb{R}^{m_2}} g(\xi) \mu(d\xi)$$

yields the equivalent statement

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \tau \mu_k \circ g_k^{-1}(d\tau) = \int_{\mathbb{R}} \tau \mu \circ g^{-1}(d\tau).$$

For fixed $a \in \mathbb{R}_+$, consider the truncation $\kappa_a : \mathbb{R} \rightarrow \mathbb{R}$ with

$$\kappa_a(\tau) := \begin{cases} \tau, & |\tau| < a \\ 0, & |\tau| \geq a. \end{cases}$$

Now

$$\begin{aligned} & \left| \int_{\mathbb{R}} \tau \mu_k \circ g_k^{-1}(d\tau) - \int_{\mathbb{R}} \tau \mu \circ g^{-1}(d\tau) \right| \\ & \leq \left| \int_{\mathbb{R}} (\tau - \kappa_a(\tau)) \mu_k \circ g_k^{-1}(d\tau) \right| + \left| \int_{\mathbb{R}} \kappa_a(\tau) \mu_k \circ g_k^{-1}(d\tau) \right. \\ & \quad \left. - \int_{\mathbb{R}} \kappa_a(\tau) \mu \circ g^{-1}(d\tau) \right| + \left| \int_{\mathbb{R}} (\kappa_a(\tau) - \tau) \mu \circ g^{-1}(d\tau) \right|. \end{aligned} \quad (2.16)$$

The proof is completed by showing that, for a given $\varepsilon > 0$, each of the three expressions on the right becomes less than $\varepsilon/3$ provided that n and a are sufficiently large.

For the first expression we obtain

$$\begin{aligned} \left| \int_{\mathbb{R}} (\tau - \kappa_a(\tau)) \mu_k \circ g_k^{-1}(\mathrm{d}\tau) \right| &\leq \int_{\{\tau: |\tau| \geq a\}} |\tau| \mu_k \circ g_k^{-1}(\mathrm{d}\tau) \\ &= \int_{\{\xi: |g_k(\xi)| \geq a\}} |g_k(\xi)| \mu_k(\mathrm{d}\xi). \end{aligned} \quad (2.17)$$

Since $p > 1$,

$$\begin{aligned} \int_{\mathbb{R}^{m_2}} |g_k(\xi)|^p \mu_k(\mathrm{d}\xi) &\geq \int_{\{\xi: |g_k(\xi)| \geq a\}} |g_k(\xi)| \cdot |g_k(\xi)|^{p-1} \mu_k(\mathrm{d}\xi) \\ &\geq a^{p-1} \int_{\{\xi: |g_k(\xi)| \geq a\}} |g_k(\xi)| \mu_k(\mathrm{d}\xi). \end{aligned} \quad (2.18)$$

Therefore, the estimate in (2.17) can be continued by

$$\leq a^{1-p} \int_{\mathbb{R}^{m_2}} |g_k(\xi)|^p \mu_k(\mathrm{d}\xi). \quad (2.19)$$

Proposition 2, part (4), and $g_k(0) = 0$ imply

$$|g_k(\xi)|^p \leq (\beta \|\xi\| + \beta \|x_k\| \cdot \|T\| + \gamma)^p.$$

Since $(x_k)_{k \in \mathbb{N}}$ is bounded and all μ_k belong to $\Delta_{p,K}(\mathbb{R}^{m_2})$, there exists a positive constant c such that

$$\int_{\mathbb{R}^{m_2}} |g_k(\xi)|^p \mu_k(\mathrm{d}\xi) \leq c \quad \text{for all } k \in \mathbb{N}.$$

Hence, (2.19) can be estimated above by c/a^{p-1} which becomes less than $\varepsilon/3$ if a is sufficiently large.

We now turn to the second expression in (2.16). Since every probability measure on the real line has at most countably many atoms, we obtain that $\mu \circ g^{-1}(\{\tau : |\tau| = a\}) = 0$ for (Lebesgue-) almost all $a \in \mathbb{R}$. Therefore, κ_a is a measurable function whose set of discontinuity points D_{κ_a} has $\mu \circ g^{-1}$ -measure zero for almost all $a \in \mathbb{R}$. We apply Rubin's Theorem to the weakly convergent sequence $\mu_k \circ g_k^{-1} \xrightarrow{w} \mu \circ g^{-1}$, cf. (2.15), and the identical sequence of functions κ_a . The role of the exceptional set then is taken by D_{κ_a} , and Rubin's Theorem is working due to $\mu \circ g^{-1}(D_{\kappa_a}) = 0$. This yields the conclusion

$$\mu_k \circ g_k^{-1} \circ \kappa_a^{-1} \xrightarrow{w} \mu \circ g^{-1} \circ \kappa_a^{-1} \quad \text{for almost all } a \in \mathbb{R}. \quad (2.20)$$

Consider the bounded continuous function $\eta: \mathbb{R} \rightarrow \mathbb{R}$ given by

$$\eta(\tau') := \begin{cases} -a, & \tau' \leq -a \\ \tau', & -a \leq \tau' \leq a \\ a, & \tau' \geq a. \end{cases}$$

By the weak convergence in (2.20), we obtain for $n \rightarrow \infty$

$$\int_{\mathbb{R}} \eta(\tau') \mu_k \circ g_k^{-1} \circ \kappa_a^{-1}(d\tau') \rightarrow \int_{\mathbb{R}} \eta(\tau') \mu \circ g^{-1} \circ \kappa_a^{-1}(d\tau'). \quad (2.21)$$

Changing variables provides

$$\begin{aligned} \int_{\mathbb{R}} \eta(\tau') \mu_k \circ g_k^{-1} \circ \kappa_a^{-1}(d\tau') &= \int_{\kappa_a^{-1}(\mathbb{R})} \eta(\kappa_a(\tau)) \mu_k \circ g_k^{-1}(d\tau) \\ &= \int_{\mathbb{R}} \kappa_a(\tau) \mu_k \circ g_k^{-1}(d\tau). \end{aligned}$$

Analogously,

$$\int_{\mathbb{R}} \eta(\tau') \mu \circ g^{-1} \circ \kappa_a^{-1}(d\tau') = \int_{\mathbb{R}} \kappa_a(\tau) \mu \circ g^{-1}(d\tau).$$

The above identities together with (2.21) confirm that the second expression on the right-hand side of (2.16) becomes arbitrarily small for sufficiently large n and almost all sufficiently large a .

Let us finally turn to the third expression in (2.16). Analogously to (2.17), (2.18) and (2.19) we obtain

$$\left| \int_{\mathbb{R}} (\kappa_a(\tau) - \tau) \mu \circ g^{-1}(d\tau) \right| \leq a^{1-p} \int_{\mathbb{R}^{m_2}} |g(\xi)|^p \mu(d\xi).$$

The integral $\int_{\mathbb{R}^{m_2}} |g(\xi)|^p \mu(d\xi)$ is finite due to part (4) of Proposition 2 and $\mathbb{E}_\mu[\|\xi\|^p] \leq K$. Hence, the third expression in (2.16) becomes less than $\varepsilon/3$ if a is large enough. \square

Employing well-established arguments of parametric optimization Proposition 7, together with standard assumptions such as boundedness of the unperturbed solution set, leads to (qualitative) continuity of the (multi-) functions assigning to the underlying probability measure the optimal value and the set of optimal solutions to (2.3), respectively. Quantitative continuity of $\phi(\cdot, \cdot)$ and quantitative stability of (2.3) require the identification of suitable distances on the space $\mathcal{P}(\mathbb{R}^{m_2})$ of probability measures.

2.2 Simple integer recourse

Simple recourse models are two-stage stochastic programs where deviations of a first-stage bid Tx against the random outcome $h(\omega)$ have to be compensated at certain costs in the second-stage. In simple integer recourse this compensation must be integer. In (2.6) we had already seen a second-stage value function Φ corresponding to a simple recourse model in case $h(\omega)$ maps to \mathbb{R}^1 .

A two-stage stochastic program with simple integer recourse is given by

$$\min_x \{c^T x + \mathbb{E}_\mu[\Phi(\xi - Tx)]: x \in X\} \quad (2.22)$$

where

$$\Phi(t) := \min\{(q^+)^T y^+ + (q^-)^T y^- : y^+ \geq t, y^- \geq -t, y^+ \in \mathbb{Z}_+^s, y^- \in \mathbb{Z}_+^s\}. \quad (2.23)$$

It is easy to see that the optimization problem in (2.23) is feasible for any $t \in \mathbb{R}^s$, so the model has complete recourse. If, moreover, we assume that $q^+ \geq 0$ and $q^- \geq 0$, then the LP relaxation to the second-stage problem has a feasible dual, and $\Phi(t) \in \mathbb{R}$ for all $t \in \mathbb{R}^s$. Finally, the assumption $\mathbb{E}_\mu[\|\xi\|] < \infty$ will guarantee that the expectation in (2.22) is finite such that (2.22) becomes a well-defined optimization problem.

The crucial fact, allowing for a much richer analysis in simple integer recourse than in the general situation of (2.3)–(2.4), is the expressibility of the value function $\Phi(t)$ in closed form, namely

$$\Phi(t) = \sum_{i=1}^s (q_i^+ \lceil t_i \rceil^+ + q_i^- \lceil -t_i \rceil^+) \quad (2.24)$$

where $\lceil \cdot \rceil^+ := \max\{\lceil \cdot \rceil, 0\}$. For the expected value function this implies the following separable representation

$$\begin{aligned} \phi(x) &= \mathbb{E}_\mu[\Phi(\xi - Tx)] \\ &= \sum_{i=1}^s q_i^+ \mathbb{E}_{\mu_i}[\lceil \xi_i - (Tx)_i \rceil^+] + \sum_{i=1}^s q_i^- \mathbb{E}_{\mu_i}[\lceil (Tx)_i - \xi_i \rceil^+] \end{aligned}$$

where μ_i is the probability measure corresponding to the marginal distribution of the i -th component of ξ . Studying $\phi(x)$ thus is studying the functions

$$u_i(\chi_i) := \mathbb{E}_{\mu_i}[\lceil \xi_i - \chi_i \rceil^+] \quad \text{and} \quad v_i(\chi_i) := \mathbb{E}_{\mu_i}[\lceil \chi_i - \xi_i \rceil^+] \quad (2.25)$$

reflecting expected surplus and expected shortage, respectively. Both these functions in one variable are quite similar, such that we will restrict further considerations to the expected surplus function. The following proposition reveals a close relationship with the cumulative distribution function $F_i(\tau) := \mu_i(\{\xi_i : \xi_i \leq \tau\})$ of ξ_i . For notational convenience, we drop the index i from now on.

Proposition 8. *In the above setting it holds for all $\chi \in \mathbb{R}$*

$$u(\chi) = \sum_{k=0}^{\infty} (1 - F(\chi + k)). \quad (2.26)$$

Proof. We have

$$\begin{aligned}
 \sum_{k=0}^{\infty} (1 - F(\chi + k)) &= \sum_{k=0}^{\infty} \mu(\xi - \chi > k) \\
 &= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \mu(\lceil \xi - \chi \rceil^+ = j) \\
 &= \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \mu(\lceil \xi - \chi \rceil^+ = j) \\
 &= \sum_{j=1}^{\infty} j \cdot \mu(\lceil \xi - \chi \rceil^+ = j) \\
 &= \mathbb{E}_{\mu}[\lceil \xi - \chi \rceil^+] = u(\chi),
 \end{aligned}$$

and the proof is complete. \square

Continuity and smoothness properties of u now result from those of the distribution function F or, existence provided, from those of a suitable probability density function of ξ . In particular, it can be shown that u is Lipschitz continuous on \mathbb{R} if there exists a density of ξ with bounded variation. Moreover, u is differentiable on \mathbb{R} if there exists a continuous density of ξ .

The function u is convex if ξ has a probability density function that is piecewise constant on every interval $]\alpha + j, \alpha + j + 1[$, $j \in \mathbb{Z}$ for some $\alpha \in [0, 1[$. Hence, when relating with arbitrary probability measures piecewise constant densities depending on the shift parameter α , one obtains convex functions related with the in general nonconvex function u . Given $\mu \in \mathcal{P}(\mathbb{R})$, its cumulative distribution function F , and $\alpha \in [0, 1[$ we denote $\lfloor \tau \rfloor_{\alpha} := \lfloor \tau - \alpha \rfloor + \alpha$ and relate the following probability density function with μ

$$\theta_{\alpha}(\tau) := F(\lfloor \tau \rfloor_{\alpha} + 1) - F(\lfloor \tau \rfloor_{\alpha}), \quad \tau \in \mathbb{R}.$$

In the literature, the density θ_{α} and the corresponding measure μ_{α} as well as the resulting convex function related with u are called α -approximations. We will adopt this here although it is quite clear that the “distance” of μ and μ_{α} , and thus the approximation error in terms of the functions, in general cannot be made arbitrarily small.

Proposition 9. *For any $\alpha \in [0, 1[$ the α -approximation*

$$u_\alpha(\chi) := \mathbb{E}_{\mu_\alpha}[\lceil \xi - \chi \rceil^+]$$

of the expected surplus function u is a convex function on \mathbb{R} .

For the expected shortage function v an analogous statement is valid. Interestingly, the resulting convex α -approximation for

$$\tilde{\phi}(\chi) := q^+ u(\chi) + q^- v(\chi), \quad q^+ + q^- > 0,$$

then, up to an additive constant, arises as an expected value function of a simple recourse model with continuous variables where the underlying probability measure has been properly modified:

Proposition 10. *Let $\alpha \in [0, 1[$, $q^+ + q^- > 0$, and u_α, v_α denote the α -approximations for the expected surplus and shortage function u and v , respectively. Then it holds for all $\chi \in \mathbb{R}$:*

$$\begin{aligned} \tilde{\phi}_\alpha(\chi) &:= q^+ u_\alpha(\chi) + q^- v_\alpha(\chi) \\ &= q^+ \mathbb{E}_{v_\alpha}[\lceil \xi - \chi \rceil^+] + q^- \mathbb{E}_{v_\alpha}[\lceil \chi - \xi \rceil^+] + \frac{q^+ q^-}{q^+ + q^-} \end{aligned}$$

where v_α is a discrete probability measure such that for all $k \in \mathbb{Z}$

$$\begin{aligned} v_\alpha(\{\alpha + k\}) &:= \frac{q^+}{q^+ + q^-} (F(\alpha + k) - F(\alpha + k - 1)) \\ &\quad + \frac{q^-}{q^+ + q^-} (F(\alpha + k + 1) - F(\alpha + k)). \end{aligned}$$

The results outlined above are derived in detail in [van der Vlerk \(1995\)](#), see also the articles mentioned in the Bibliographical Notes below.

2.3 Multi-stage models

The models discussed so far assume a two-stage setting for the gain of information. Uncertainty is unveiled at once and decisions subdivide into those before and those after the unveiling. Often, a more complex view is appropriate at this place. Multistage stochastic programs address the situation

where uncertainty is unveiled stepwise with intermediate decisions that must not anticipate future information. We refer to Chapters 1 and 2 for basic statements about modeling principles and structure in the multi-stage situation. In what follows, we will adopt a fairly general modeling perspective, similar to Section 3.3 of Chapter 1. We will go beyond the setting of that section by adding integer requirements, and we will study some first implications of such a model extension.

Consider a finite horizon sequential decision process under uncertainty where the decision $x_t \in \mathbb{R}^{n_t}$ at stage $t \in \{1, \dots, T\}$ is based on information available up to time t only. Information is modeled as a discrete time stochastic process $\{\xi_t\}_{t=1}^T$ on some probability space (Ω, \mathcal{F}, P) with ξ_t taking values in \mathbb{R}^{m_t} . The random vector $\xi^t := (\xi_1, \dots, \xi_t)$ then reflects the information available up to time t . Nonanticipativity, i.e., the requirement that x_t must not depend on future information, is formalized by saying that x_t is measurable with respect to the σ -algebra $\mathcal{F}_t \subseteq \mathcal{F}$ which is generated by ξ^t , $t = 1, \dots, T$. Clearly, $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for all $t = 1, \dots, T-1$. As in the two-stage case, the first-stage decision x_1 usually is deterministic. Therefore, $\mathcal{F}_1 = \{\emptyset, \Omega\}$. Moreover, we assume that $\mathcal{F}_T = \mathcal{F}$.

The constraints of our multi-stage models are subdivided into three groups. The first group comprises conditions on x_t arising from the individual time stages:

$$\begin{aligned} x_t(\omega) \in X_t, \quad B_t(\xi_t(\omega))x_t(\omega) &\geq d_t(\xi_t(\omega)) \\ P\text{-almost surely, } t = 1, \dots, T. \end{aligned} \quad (2.27)$$

Here, $X_t \subseteq \mathbb{R}^{n_t}$ is a set whose convex hull is a polyhedron. In this way, integer requirements to components of x_t are allowed for. For simplicity we assume that X_t is compact. Note that, by the integrality in X_t , we go beyond the setting of all the multi-stage models analyzed in Chapters 1 and 2. As in the two-stage case, convexity will no longer be available. In particular, the arguments based on duality and conjugacy that led to Proposition 30 in Chapter 2, stating that the multi-stage stochastic program is convex, are not working anymore, and, in fact, the models become nonconvex.

The next group of constraints models linkage between different time stages:

$$\sum_{\tau=1}^t A_{t\tau}(\xi_t(\omega))x_\tau(\omega) \geq g_t(\xi_t(\omega)) \quad P\text{-almost surely, } t = 2, \dots, T. \quad (2.28)$$

Finally, there is the nonanticipativity of x_t , i.e.,

$$x_t \text{ is measurable with respect to } \mathcal{F}_t, \quad t = 1, \dots, T. \quad (2.29)$$

In addition to the constraints we have a linear objective function

$$\sum_{t=1}^T c_t(\xi_t(\omega))x_t(\omega).$$

The matrices $A_{t\tau}(\cdot)$, $B_t(\cdot)$ as well as the right-hand sides $d_t(\cdot)$, $g_t(\cdot)$ and the cost coefficients $c_t(\cdot)$ all have conformable dimensions and depend affinely linearly on the relevant components of ξ .

The decisions x_t are understood as members of the function spaces $L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{n_t})$, $t = 1, \dots, T$. The constraints (2.27), (2.28) then impose pointwise conditions on the x_t , whereas (2.29) imposes functional constraints, in fact, membership in a linear subspace of $\times_{t=1}^T L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{n_t})$.

The multistage extension of (2.3) is the minimization of expected minimal costs subject to nonanticipativity of decisions:

$$\min \left\{ \int_{\Omega} \min_{x(\omega)} \left\{ \sum_{t=1}^T c_t(\xi_t(\omega))x_t(\omega) : (2.27), (2.28) \right\} P(d\omega) : x \text{ fulfilling (2.29)} \right\} \quad (2.30)$$

The minimization in the integrand of (2.30) being separable with respect to $\omega \in \Omega$, it is possible to interchange integration and minimization, and the problem can be restated as

$$\min \left\{ \int_{\Omega} \sum_{t=1}^T c_t(\xi_t(\omega))x_t(\omega) P(d\omega) : x \text{ fulfilling (2.27), (2.28), (2.29)} \right\}. \quad (2.31)$$

Due to the mentioned interplay of pointwise and functional constraints it remains to check whether (2.31) is well defined, cf. Section 3.3 of Chapter 1 where this is addressed for a noninteger counterpart of (2.31).

Recall that X_t is compact and assume that $\xi_t \in L_1(\Omega, \mathcal{F}, P; \mathbb{R}^{m_t})$ for $t = 1, \dots, T$. For each $\omega \in \Omega$ we define the subset $\mathcal{Y}(\omega)$ of $\mathcal{X} := \times_{t=1}^T \mathbb{R}^{n_t}$

$$\mathcal{Y}(\omega) := \left\{ y \in \mathcal{X} : y_t \in X_t, B_t(\xi_t(\omega))y_t \geq d_t(\xi_t(\omega)), t = 1, \dots, T, \right. \\ \left. \sum_{\tau=1}^t A_{t\tau}(\xi_t(\omega))y_{\tau} \geq g_t(\xi_t(\omega)), t = 2, \dots, T \right\} \quad (2.32)$$

and the extended real-valued function φ

$$\varphi(y_1, \dots, y_T, \omega) := \begin{cases} \sum_{t=1}^T c_t(\xi_t(\omega))y_t, & (y_1, \dots, y_T) \in \mathcal{Y}(\omega), \\ +\infty, & \text{otherwise} \end{cases} \quad (2.33)$$

from $\mathcal{X} \times \Omega$ to $(-\infty, +\infty]$. With these notations, (2.31) is equivalent to

$$\min\{\mathbb{E}_P[\varphi(x_1, \dots, x_T, \omega)]: x_t \text{ measurable w.r.t. } \mathcal{F}_t, t = 1, \dots, T\}. \quad (2.34)$$

The real-valued function $(y, \omega) \mapsto \sum_{t=1}^T c_t(\xi_t(\omega))y_t$ is continuous in y for each $\omega \in \Omega$ and measurable in ω for each $y \in \mathcal{X}$, and the set-valued mapping \mathcal{Y} from Ω to \mathcal{X} is closed-valued and measurable (cf. Theorem 14.36 in [Rockafellar and Wets, 1997](#)). With $\mathcal{B}(\mathcal{X})$ denoting the σ -algebra of Borel sets in \mathcal{X} , the function φ is $\mathcal{B}(\mathcal{X}) \otimes \mathcal{F}$ -measurable (cf. Example 14.32 in [Rockafellar and Wets, 1997](#)). Furthermore, the following estimate is valid for each $y \in \times_{t=1}^T X_t$ and $\omega \in \Omega$:

$$|\varphi(y_1, \dots, y_T, \omega)| \leq \sum_{t=1}^T \|c_t(\xi_t(\omega))\| \sup_{y_t \in X_t} \|y_t\| \quad (2.35)$$

Hence, $\mathbb{E}_P[\varphi(x_1, \dots, x_T, \omega)]$ is finite for each decision $x = (x_1, \dots, x_T)$ such that $x(\omega) \in \mathcal{Y}(\omega)$ for P -almost all $\omega \in \Omega$.

As in [Evstigneev \(1976\)](#), we construct recursively two sequences of functions by putting $\psi_{T+1} := \varphi$ and

$$\varphi_t(y_1, \dots, y_t, \omega) := \mathbb{E}_P^r[\psi_{t+1}(y_1, \dots, y_t, \cdot) \mid \mathcal{F}_t](\omega), \quad (2.36)$$

$$\psi_t(y_1, \dots, y_{t-1}, \omega) := \inf_y \varphi_t(y_1, \dots, y_{t-1}, y, \omega), \quad (2.37)$$

for $t = T, \dots, 1$, and for each $\omega \in \Omega$ and $y_\tau \in X_\tau$, $\tau = 1, \dots, T$.

Here, $\mathbb{E}_P^r[\cdot \mid \mathcal{F}_t]$ denotes the regular conditional expectation with respect to \mathcal{F}_t . By definition, the regular conditional expectation is a version of the conditional expectation (i.e., $\mathbb{E}_P^r[\cdot \mid \mathcal{F}_t] = \mathbb{E}_P[\cdot \mid \mathcal{F}_t]$, P -a.s.) with the property that the mapping $(z, \omega) \mapsto \Phi(z, \omega) := \mathbb{E}_P^r[\Psi(z, \cdot) \mid \mathcal{F}_t](\omega)$ from $Z_t \times \Omega$ to $(-\infty, +\infty]$ is $\mathcal{B}(Z_t) \otimes \mathcal{F}_t$ -measurable if Ψ is $\mathcal{B}(Z_t) \otimes \mathcal{F}$ -measurable. Here, Z_t is allowed to be an arbitrary closed subset of a Euclidean space. The regular conditional expectation exists if Ψ is $\mathcal{B}(Z_t) \otimes \mathcal{F}$ -measurable and uniformly integrable, i.e., there exists a (real) random variable ζ with finite first moment such that $|\Psi(z, \omega)| \leq \zeta(\omega)$ for $z \in Z_t$ and $\omega \in \Omega$ (see [Dynkin and Evstigneev, 1976](#)).

Due to condition (2.35), relation (2.36) is well defined for $t = T$ and leads to a $\mathcal{B}(Z) \otimes \mathcal{F}_T$ -measurable function ϕ_T , where $Z := \times_{t=1}^T X_t$. It is shown in Evstigneev (1976) that the relations (2.36) and (2.37) are well defined for all $t = T, \dots, 1$. Furthermore, the following optimality criterion and existence result for (2.34) or, equivalently, for (2.31) is valid.

Proposition 11. *Adopt the above setting for (2.31) and assume that (2.31) has a feasible solution. Then $\{\bar{x}_t\}_{t=1}^T$ is an optimal solution of (2.31) iff*

$$\varphi_t(\bar{x}^t(\omega), \omega) = \psi_t(\bar{x}^{t-1}(\omega), \omega), \quad P - \text{a.s.}, \quad t = 1, \dots, T. \quad (2.38)$$

Moreover, there exists a solution \bar{x}_1 of the first-stage optimization problem

$$\min\{\varphi_1(x_1) = \mathbb{E}_P[\psi_2(x_1, \omega)]: x_1 \in X_1, B_1(\xi_1)x_1 \geq d_1(\xi_1)\}, \quad (2.39)$$

and given \mathcal{F}_τ -measurable functions \bar{x}_τ for $\tau = 1, \dots, t-1$, there exists an \mathcal{F}_τ -measurable function \bar{x}_t such that $\varphi_t(\bar{x}^t(\omega), \omega) = \psi_t(\bar{x}^{t-1}(\omega), \omega)$, P -a.s.

Relations (2.36) and (2.37) define the mixed-integer analogon to the nested formulation developed in Section 3.1 of Chapter 1 for the purely linear case, see also the general nested problem in continuous variables in Section 3.3 of Chapter 1. In particular, the optimal value $\psi_t(y_1, \dots, y_{t-1}, \omega)$ is the cost-to-go, cf. (3.3) in Chapter 2, and (2.38) states the fact that an optimal solution to (2.31) has to fulfil the dynamic programming equation, cf. (3.4) and Proposition 30 in Chapter 2. As in (3.5) of Chapter 1, in problem (2.39) all the subsequent stages are absorbed into the function $\varphi_1(x_1)$. Hence, (2.39) is a well-defined mixed-integer extension of (3.5) in Chapter 1, with the only difference that (3.5) has staircase whereas (2.39) triangular form.

In Proposition 30 of Chapter 2 convexity of the noninteger counterpart to $\varphi_1(\cdot)$ is shown. With integer requirements, this convexity already breaks down for two-stage problems and hence cannot be expected to hold. However, lower semicontinuity of $\varphi_1(\cdot)$ still can be established.

Proposition 12. *Adopt the above setting for (2.31) and assume that the matrices $A_{t\tau}(\cdot)$, $B_t(\cdot)$ as well as the cost coefficients $c_t(\cdot)$ all are deterministic. Then the objective function $\varphi_1(\cdot)$ of the first-stage optimization problem (2.39) is lower semicontinuous on its domain of definition.*

Proof. (Outline) The proof is done by induction over the time stages $t = T, T-1, \dots, 1$. According to the definitions in (2.33) and (2.37), at each stage an infimum of a parameter dependent function over a mixed-integer set constrained by linear inequalities is taken. Moreover, the latter set has parameters in the right-hand sides of the inequalities. At each stage, it can be

shown that the set-valued mapping assigning to the relevant right-hand side parameter the relevant mixed-integer constraint set is upper semicontinuous.

At stage T , the objective function of the above parametric optimization problem is linear and does not depend on a parameter. In fact, we have a mixed-integer linear program with right-hand side parameters, whose value function is lower semicontinuous according to the argument given in item (3) in front of [Proposition 2](#). A conditional expectation of this value function, cf. (2.36), which, by Fatou's Lemma for conditional expectations, is again lower semicontinuous, enters the objective at stage $T-1$, cf. (2.37).

Therefore, at stage $T-1$, as well as in all subsequent stages $t = T-2, \dots, 2$, we have a parametric program where, with respect to the relevant parameter, the objective function is lower semicontinuous and the constraint set mapping is upper semicontinuous. Basic results from parametric optimization, such as Theorem 1.17 of [Rockafellar and Wets \(1997\)](#) then imply that the value function of the optimization problem at stage $T-1$ is lower semicontinuous with respect to the relevant parameters. Fatou's Lemma for conditional expectations then inherits lower semicontinuity to the objective in stage $T-2$, and the above arguments can be repeated. \square

Structural properties for problem (2.31) beyond the above propositions are widely open and a field of current research.

3 Algorithms

3.1 Decomposition schemes

Consider a classical deterministic integer program

$$(\text{IP}) \quad \min\{c^T x : x \in X\}. \quad (3.40)$$

For simplicity, we consider the pure integer case, where $X \subseteq \mathbb{Z}^n$, although most of the definitions and properties presented here extend to the mixed integer case. In practice, the set X is described through a finite set of linear constraints, defining a polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ and through integrality restrictions.

A polyhedron $P \subseteq \mathbb{R}^n$ is a formulation for $X \subseteq \mathbb{Z}^n$ iff $X = P \cap \mathbb{Z}^n$. The ideal formulation would be to replace IP by the equivalent linear program

$$\min\{c^T x : x \in \text{conv}(X)\} \quad (3.41)$$

as all extreme points of the convex hull of X belong to X . Such a formulation naturally gives an integer solution to the linear programming relaxation of (3.41).

An inequality $\pi^T x \leq \pi_0$ is a valid inequality for $X \subseteq \mathbb{R}^n$ if $\pi^T x \leq \pi_0$ for all $x \in X$. The Chvatal–Gomory procedure to construct a valid inequality for $X = P \cap \mathbb{Z}^n$ is to consider a vector $u \in \mathbb{R}_+^m$ of nonnegative weights for the rows of A , take the linear combination of the rows

$$u^T A x \leq u^T b,$$

round down the coefficients in the l.h.s,

$$\sum_{j=1}^n \lfloor u^T a_j \rfloor x_j \leq u^T b$$

where a_j is the j th column of A , then round-down the r.h.s.

$$\sum_{j=1}^n \lfloor u^T a_j \rfloor x_j \leq \lfloor u^T b \rfloor. \quad (3.42)$$

The first transformation is justified by $u \geq 0$, the second by $\sum_{j=1}^n \lfloor u^T a_j \rfloor x_j \leq \sum_{j=1}^n u^T a_j x_j$, and the third as $\sum_{j=1}^n \lfloor u^T a_j \rfloor x_j$ is integer. Thus, for all $u \geq 0$, (3.42) is a valid inequality for X . More surprisingly, every valid inequality for X can be obtained by applying the Chvatal–Gomory procedure a finite number of times.

The separation problem SP associated with IP is the following : given $\bar{x} \in \mathbb{R}^n$, is $\bar{x} \in \text{conv}(X)$? If not, find a valid inequality $\pi^T x \leq \pi_0$ for X such that $\pi^T \bar{x} > \pi_0$. As the Chvatal–Gomory procedure is not a constructive one, a considerable amount of research has been devoted to characterize strong valid inequalities, which define facets of P , and related separation algorithms (SA). Among the popular SA, we may cite lifted cover inequalities for knapsack constraints, mixed integer rounding inequalities (a generalization of Gomory mixed integer cuts) and disjunctive inequalities. Separation algorithms are not always efficient, as sometimes the separation problem itself is NP-hard. Even when the separation algorithm is efficient, it may not be wise to generate all possible valid inequalities as the linear program would tend to become huge. Early attempt to use Gomory cuts appeared to be disappointing as successive cuts tend to be less and less efficient and the corresponding LP harder and harder to solve. Cutting planes are thus very often combined with a branch & bound scheme. Branching on a fractional value \bar{x}_j consists of defining two subregions $X \cap \{x: x_j \leq \lfloor \bar{x}_j \rfloor\}$ and $X \cap \{x: x_j \geq \lfloor \bar{x}_j \rfloor + 1\}$. These subregions can in turn be subdivided by later branchings. This generates a number of nodes N^ρ , $\rho = 1, \dots, R$, that form a partition of \mathbb{R}^n , i.e., $\mathbb{R}^n = \bigcup_{\rho=1, \dots, R} N^\rho$ and $N^\rho \cap N^\sigma = \emptyset$, $\rho \neq \sigma$.

Clearly,

$$\min\{c^T x: x \in X\} = \min_{\rho=1,\dots,R} \{\min\{c^T x: x \in X \cap N^\rho\}\}$$

The branch & bound procedure is finite as only finitely many nodes can be generated when X is compact. Nodes are fathomed for three reasons: when they have integer solution, when they have an optimal value, which is worse than an already known integer solution, and when they are infeasible. As already said, modern codes combine cut generation and branching in so called branch & cut procedures. Another basic technique in integer programming is Lagrangian relaxation. Its main idea is to relax “complicating” constraints to end up with manageable subproblems that are coordinated by a nonsmooth dual optimization. In general, integer requirements imply the lack of duality such that, typically, heuristics have to be employed for finding promising primal solutions on the basis of the results of the dual optimization. We now show how stochastic programming decomposition techniques could be designed using cutting planes and branch & bound in the second-stage as well as Lagrangian relaxation of nonanticipativity constraints.

Decomposition by cutting planes

Let the two stage stochastic integer problem be

$$(SIP) \quad \min\{c^T x + \mathbb{E}_\mu[\Phi(\xi - Tx)]: x \in X\} \quad (3.43)$$

with

$$\Phi(t) := \min\{q^T y: Wy = t, y \in Y\} \quad (3.44)$$

and $Y \subseteq \mathbb{Z}_+^{m_2}$. All notations are as in [Section 2](#). The deterministic equivalent program is

$$(DEP) \quad \min\{c^T x + Q(x): x \in X\} \quad (3.45)$$

where $Q(x) = \mathbb{E}_\mu[\Phi(\xi - Tx)]$. For all practical purposes, it can be transformed into

$$\min\{c^T x + \theta: x \in X, \theta \geq Q(x)\} \quad (3.46)$$

In view of the properties presented just above, the difficulty of solving (3.46) precisely lies in having a formulation to replace $\theta \geq Q(x)$. Any valid constraint in the \mathbb{R}^{m+1} space of (x, θ) is called an optimality cut. It takes the form of a constraint $\theta \geq f(x)$ that holds for every $x \in X$, and $\theta \geq Q(x)$.

Definition 13. A set of s optimality cuts $\{\theta \geq f_\ell(x), \ell = 1, \dots, s\}$ is said to be sufficient at $x \in X$ if

$$\theta \in \{\theta: \theta \geq f_\ell(x), \ell = 1 \dots s\} \Rightarrow \theta \geq Q(x).$$

In the classical L -shaped algorithm where the second-stage only involves continuous variables and ξ has a finite support, the linear programming duality theory implies that a sufficient set of optimality cuts can be found for all $x \in X$.

We now show how optimality cuts could be generated through a cutting plane solution of the second-stage program. For simplicity of presentation, we assume complete recourse for the second-stage so that feasibility cuts can be left aside. Take the case when the assumptions of [Proposition 4](#) are satisfied. We also assume $Y \subseteq \mathbb{Z}^{n_2}$. For each outcome $\xi^k \in \Xi$, the second-stage problem is

$$\Phi(\xi^k - Tx) = \min\{q^T y: Wy \geq \xi^k - Tx, y \in \mathbb{Z}^{n_2}\} \quad (3.47)$$

Let \mathbb{F} be the set of dual price functions. Although more general classes can be considered, take \mathbb{F} to be the class of functions $F: \mathbb{R}^{m_2} \rightarrow \mathbb{R}$ that are nondecreasing, subadditive and satisfy $F(0) = 0$. The dual problem of (3.47) is

$$\max_F \{F(\xi^k - Tx): F(w_j) = q_j, j = 1, \dots, n_2, F \in \mathbb{F}\} \quad (3.48)$$

where w_j is the j th column of W .

Proposition 14. Suppose \hat{F}^k , $k = 1, \dots, K$, are optimal dual price functions obtained by solving (3.48) with $x = x^v$ for each $\xi^k \in \Xi$. Then, an optimality cut at x^v is given by

$$\theta \geq \sum_{k=1}^K p^k \hat{F}^k(\xi^k - Tx) \quad (3.49)$$

Proof. For each $k = 1, \dots, K$, let \bar{y}^k and \hat{F}^k be optimal solutions of (3.47) and (3.48), respectively, corresponding to $x = x^v$. By integer duality, they satisfy $q^T \bar{y}^k = \hat{F}^k(\xi^k - Tx^v)$. For each feasible $F \in \mathbb{F}$ and each feasible $y \in \mathbb{Z}^{n_2}$, we have $F(\xi^k - Tx^v) \leq \hat{F}^k(\xi^k - Tx^v) = q^T \bar{y}^k \leq q^T y$. Now, for each $x \in X$ and corresponding optimal dual price \bar{F}^k , we have

$$Q(x) = \sum_{k=1}^K p^k \bar{F}^k(\xi^k - Tx) \geq \sum_{k=1}^K p^k \hat{F}^k(\xi^k - Tx),$$

with

$$Q(x^v) = \sum_{k=1}^K p^k \hat{F}^k(\xi^k - Tx^v). \quad \square$$

In a cutting plane procedure, the LP-relaxation of the integer program is considered. Valid inequalities are successively generated and added to the formulation, until the LP-relaxation optimal solution spontaneously meets the integrality restrictions. The valid inequalities can be written as

$$\sum_{j=1}^{m_2} F^{(\ell)}(w_j) y_j \geq F^{(\ell)}(q), \quad \ell = 1, \dots, s, \quad (3.50)$$

where $F^{(\ell)} \in \mathbb{F}$, $\ell = 1, \dots, s$. Let $(u_1, \dots, u_{m_2}, u_{m_2+1}, \dots, u_{m_2+s})$ be the dual variables associated to the optimum of the final LP-relaxation. We construct the function $F: \mathbb{R}^{m_2} \rightarrow \mathbb{R}$ as

$$F(t) := \sum_{i=1}^{m_2} u_i t_i + \sum_{i=1}^s u_{m_2+i} F^{(i)} t \quad (3.51)$$

By construction, $F \in \mathbb{F}$ and is an optimal solution to (3.48).

As an example, consider the case where valid inequalities are generated by the Gomory's fractional cutting plane algorithm.

Let $h = \xi^k - Tx^v$ in (3.47).

Let

$$y_{B_i} + \sum_{j \in NB} \bar{w}_{ij} y_j = \bar{h}_i \quad \text{for } i = 1, \dots, m_2 \quad (3.52)$$

be the optimal basis for second-stage problem (3.47) associated with a given k and given x^v . This optimal basis can be rewritten in such a way that $\bar{w}_{ij} \leq 0$, $j \in NB$, $\bar{h}_i \geq 0$, $i = 1, \dots, m_2$. NB is the set of nonbasic variables, while B_i is the variable basic in row i . Let y^* be the corresponding optimal solution. If y^* does not meet the integrality requirements, there is at least one row i with $\bar{h}_i \notin \mathbb{Z}^1$. Choosing such a row, the Chvatal–Gomory cut for row i is

$$y_{B_i} + \sum_{j \in NB} \lfloor \bar{w}_{ij} \rfloor y_j \leq \lfloor \bar{h}_i \rfloor$$

which, by eliminating y_{B_i} , can be rewritten as

$$\sum_{j \in \text{NB}} (\bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor) y_j \geq \bar{h}_i - \lfloor \bar{h}_i \rfloor$$

or

$$\sum_{j \in \text{NB}} f_{ij} y_j \geq f_i \quad (3.53)$$

where $f_{ij} = \bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor$, $f_i = \bar{h}_i - \lfloor \bar{h}_i \rfloor$. By construction, $0 \leq f_{ij} < 1$ and $0 < f_i < 1$. As $y_i^* = 0$, $i \in \text{NB}$ at the optimal LP solution, this inequality cuts off y^* .

Now let β be the row in the basis inverse corresponding to the row which generated the cut (3.53). Let $g = \beta - \lfloor \beta \rfloor$. Then, the Gomory cut (3.53) can be rewritten in terms of the original variables as the Chvatal–Gomory inequality

$$\sum_{j=1}^n \lfloor g w_j \rfloor y_j \leq \lfloor g h \rfloor,$$

namely

$$\sum_{j=1}^n \lfloor g w_j \rfloor y_j \leq \lfloor g(\xi^k - T x^\nu) \rfloor. \quad (3.54)$$

Thus, expressed in terms of the first-stage variable, the optimality cuts involve $\lfloor g(\xi^k - T x) \rfloor$ terms, i.e., rounding-down of combinations of x . Each one of these rounding down operations requires an additional constraint and an additional integer variable in the first-stage. The left-hand-side of (3.54) does not depend on x , so that similar cuts could be generated for various realizations of $\xi \in \Xi$. At the moment, it is not known how to use this property efficiently. The approach remains impracticable as the number of auxiliary variables and constraints will equal the total number of round down operations needed to generate the cuts.

Decomposition by branch & bound

We now indicate how optimality cuts can be obtained through a branch & bound solution procedure in the second-stage. Consider a given first-stage iterate point x^ν and a given realization ξ^k of ξ . Based on a full branching on the second-stage problem (3.47) for $x = x^\nu$, one obtains a partition of \mathbb{R}^{n_2} into R terminal nodes $Y^\rho = \{y: a^\rho \leq y \leq b^\rho\}$, $\rho = 1, \dots, R$. The optimal objective value of the second-stage program over Y^ρ is

$$Q^\rho(x^\nu, \xi^k) = \min\{q^T y: W y = \xi^k - T x^\nu, a^\rho \leq y \leq b^\rho\}$$

By linear programming duality, it is also

$$Q^\rho(x^\nu, \xi^k) = (\pi^\rho)^T(\xi^k - Tx^\nu) + (\underline{\pi}^\rho)^T a^\rho + (\bar{\pi}^\rho)^T b^\rho$$

where π^ρ , $\underline{\pi}^\rho$ and $\bar{\pi}^\rho$ are the optimal dual variables associated with the original constraints, lower and upper bounds on $y \in Y^\rho$, respectively.

To simplify notations, we represent this expression as

$$Q^\rho(x^\nu, \xi^k) = (\sigma_k^\rho)^T x^\nu + \tau_k^\rho$$

with $(\sigma_k^\rho)^T = -(\pi^\rho)^T \cdot T$ and $\tau_k^\rho = (\pi^\rho)^T \xi^k + (\underline{\pi}^\rho)^T a^\rho + (\bar{\pi}^\rho)^T b^\rho$.

Duality theory also implies that $Q^\rho(x, \xi^k) \geq (\sigma_k^\rho)^T x^\nu + \tau_k^\rho$.

Also, by construction of the branch & bound,

$$Q(x, \xi^k) = \min_{\rho=1, \dots, R} Q^\rho(x, \xi^k).$$

Thus,

$$\theta^k \geq p_k \min_{\rho=1, \dots, R} \{(\sigma_k^\rho)^T x + \tau_k^\rho\} \quad (3.55)$$

is a valid optimality cut for $Q(x, \xi^k)$. It can thus be embedded in a multicut representation $\theta = \sum_{k=1, \dots, K} \theta^k$. When SIP has not complete recourse, some of the terminal nodes may be infeasible, in which case their dual solutions contain unbounded rays with dual objective values going to $-\infty$, so that the minimum is restricted to the feasible terminal nodes.

The optimality cut (3.55) is a sufficient set of optimality cuts at x^ν for θ^k . Unfortunately, as it is well known, (3.55) is a nonlinear expression. R auxiliary binary variables and $R+1$ constraints are required to describe (3.55) in a polyhedral representation with mixed integer variables. This, plus solving a full branch & bound for each x^ν and each k , makes a decomposition by branch & bound of little practical value.

Scenario decomposition

In a general setting, scenario decomposition can be understood as a solution method for the multi-stage stochastic integer program (2.31) where, out of the constraints (2.27), (2.28), (2.29), the nonanticipativity condition (2.29) is subjected to Lagrangian relaxation. We will demonstrate scenario decomposition at the following two-stage model:

$$\min\{c^T x + \mathbb{E}_\mu[\Phi(\xi - Tx)]: x \in X\} \quad (3.56)$$

with

$$\Phi(t) := \min\{q^T y : Wy = t, y \in Y\}$$

where $X \subseteq \mathbb{R}^{n_1}$, $Y \subseteq \mathbb{R}^{n_2}$ are polyhedra, possibly involving integer requirements to components of x and y . We assume that all problem data have conformable dimensions, that W is rational, and that μ follows a discrete distribution with realizations (or scenarios) ξ_1, \dots, ξ_N and probabilities π_1, \dots, π_N . Then problem (3.56) can be written as the following mixed-integer linear program

$$\min_{x, y_j} \left\{ c^T x + \sum_{j=1}^N \pi_j q^T y_j : Tx + Wy_j = \xi_j, y_j \in Y, x \in X \right\}. \quad (3.57)$$

Due to the sheer size, general purpose mixed-integer linear programming solvers quickly fail at these problems. We reformulate (3.57) by introducing copies x_j , $j=1, \dots, N$, and adding the explicit nonanticipativity constraints $x_1 = \dots = x_N$, or an equivalent system. For notational convenience, the latter is written as $\sum_{j=1}^N H_j x_j = 0$ with proper (l, n_1) -matrices H_j , $j=1, \dots, N$. Problem (3.57) then becomes

$$\min \left\{ \sum_{j=1}^N \pi_j (c^T x_j + q^T y_j) : Tx_j + Wy_j = \xi_j, x_j \in X, y_j \in Y, \sum_{j=1}^N H_j x_j = 0 \right\}. \quad (3.58)$$

For $\lambda \in \mathbb{R}^l$ we consider the functions

$$L_j(x_j, y_j, \lambda) := \pi_j (c^T x_j + q^T y_j) + \pi_j \lambda^T H_j x_j, \quad j = 1, \dots, N, \quad (3.59)$$

and form the Lagrangian

$$L(x, y, \lambda) := \sum_{j=1}^N L_j(x_j, y_j, \lambda).$$

Later on, the Lagrangian will have to be minimized over a mixed-integer polyhedral set, which will be accomplished by mixed-integer linear programming solvers. To avoid nonlinearities, our Lagrangian is very much standard in that it is based on linear expressions and does not involve nonlinear augmentation terms.

Given that the objective in (3.58) is an expectation, it is natural to base the Lagrangian at a probabilistic inner product which, in (3.59), leads to the factors π_j in front of the terms $\lambda^T H_j x_j$. Conceptually, this follows the lines of dualization as developed in Rockafellar and Wets (1978). Furthermore, ill-conditioning in the Lagrangian dual when disregarding the probabilities in the second term of (3.59) is avoided this way, see Baccud et al. (2001) and Gröwe-Kuska et al. (2002) for respective observations.

The Lagrangian dual of (3.58) is the optimization problem

$$\max\{D(\lambda): \lambda \in \mathbb{R}^l\} \quad (3.60)$$

where

$$D(\lambda) = \min \left\{ \sum_{j=1}^N L_j(x_j, y_j, \lambda): Tx_j + Wy_j = \xi_j, x_j \in X, y_j \in Y \right\}. \quad (3.61)$$

The above minimization is separable, and we have

$$D(\lambda) = \sum_{j=1}^N D_j(\lambda) \quad (3.62)$$

where

$$D_j(\lambda) = \min\{L_j(x_j, y_j, \lambda): Tx_j + Wy_j = \xi_j, x_j \in X, y_j \in Y\}. \quad (3.63)$$

$D_j(\lambda)$ is the pointwise minimum of affine functions in λ , and hence piecewise affine and concave. Therefore, (3.60) is a nonsmooth concave maximization (or convex minimization) problem that can be solved by bundle methods from nondifferentiable optimization, see Hiriart-Urruty and Lemaréchal (1993) and Kiwiel (1990). At each iteration, these methods require the objective value and one subgradient of D . These are obtained by solving the optimization problem in (3.61) which, thanks to the separability in (3.62), reduces to solving N problems of single-scenario size. The latter are mixed-integer linear programs and very often within the reach of advanced general purpose solvers. Altogether, the optimal value φ_{LD} of (3.60) provides a lower bound to the optimal value φ of problem (3.57). Specifying a well-known result in Lagrangian relaxation of mixed-integer linear programs, see e.g., Nemhauser and Wolsey (1988), leads to the following proposition.

Proposition 15. *It holds $\varphi \geq \varphi_{\text{LD}}$. If for some multiplier $\lambda \in \mathbb{R}^l$ the optimal solutions (x_j, y_j) , $j=1, \dots, N$, to the optimization problem in (3.61) fulfil $\sum_{j=1}^N H_j x_j = 0$, then $\varphi = \varphi_{\text{LD}}$ and (x_j, y_j) , $j=1, \dots, N$, are optimal for (3.58).*

With φ_{LP} denoting the optimal value of the linear programming relaxation to (3.58) it holds $\varphi_{\text{LD}} \geq \varphi_{\text{LP}}$.

Equality of φ and φ_{LD} in Proposition 15 being a rare exception, “promising” feasible points for the original primal problem (3.58) are derived by heuristics using the results of the dual optimization. Since, in our situation, the relaxed constraints ($x_1 = \dots = x_N$) are particularly simple, ideas for such heuristics arise quite naturally. For example, examine the x_j -components, $j = 1, \dots, N$, of solutions to (3.63) for optimal or nearly optimal λ , and decide for the most frequent value arising or average and round if necessary. If the heuristic provides a feasible solution to (3.58), then the objective value of the latter yields an upper bound $\bar{\varphi}$ for φ .

The difference $\bar{\varphi} - \varphi_{\text{LD}}$ then indicates the quality of the feasible solution found. If desired, this quality certificate can be improved by embedding the procedure described so far into a branch & bound scheme for (3.56) seen as a nonconvex global optimization problem. Recall from (2.11) the notation $\phi(x) := \mathbb{E}_\mu[\Phi(\xi - Tx)]$. Let \mathbf{P} denote the list of current problems and $\varphi_{\text{LD}} = \varphi_{\text{LD}}(P)$ the Lagrangian lower bound for $P \in \mathbf{P}$. The scheme then consists of the following steps.

Scenario decomposition algorithm

Step 1 Initialization: Set $\bar{\varphi} = +\infty$ and let \mathbf{P} consist of problem (3.58).

Step 2 Termination: If $\mathbf{P} = \emptyset$ then the solution \hat{x} that yielded $\bar{\varphi} = c^T \hat{x} + \phi(\hat{x})$ is optimal.

Step 3 Node selection: Select and delete a problem P from \mathbf{P} and solve its Lagrangian dual. If the optimal value $\varphi_{\text{LD}}(P)$ hereof equals $+\infty$ (infeasibility of a subproblem) then go to step 2.

Step 4 Bounding: If $\varphi_{\text{LD}}(P) \geq \bar{\varphi}$ go to step 2 (this step can be carried out as soon as the value of the Lagrangian dual rises above $\bar{\varphi}$).

- (i) The scenario solutions x_j , $j = 1, \dots, N$, are identical: if $c^T x_j + \phi(x_j) < \bar{\varphi}$ then let $\bar{\varphi} = c^T x_j + \phi(x_j)$ and delete from \mathbf{P} all problems P' with $\varphi_{\text{LD}}(P') \geq \bar{\varphi}$. Go to step 2.
- (ii) The scenario solutions x_j , $j = 1, \dots, N$ differ: compute the average $\bar{x} = \sum_{j=1}^N \pi_j x_j$ and round it by some heuristic to obtain \bar{x}^R . If $c^T \bar{x}^R + \phi(\bar{x}^R) < \bar{\varphi}$ then let $\bar{\varphi} = c^T \bar{x}^R + \phi(\bar{x}^R)$ and delete from \mathbf{P} all problems P' with $\varphi_{\text{LD}}(P') \geq \bar{\varphi}$. Go to step 5.

Step 5 Branching: Select a component $x_{(k)}$ of x and add two new problems to \mathbf{P} obtained from P by adding the constraints $x_{(k)} \leq \lfloor \bar{x}_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor \bar{x}_{(k)} \rfloor + 1$, respectively (if $x_{(k)}$ is an integer component), or $x_{(k)} \leq \bar{x}_{(k)} - \varepsilon$ and $x_{(k)} \geq \bar{x}_{(k)} + \varepsilon$, respectively, where $\varepsilon > 0$ is a tolerance parameter to have disjoint subdomains.

This scheme is obviously finite if X is bounded and all x -components have to be integers. If x is mixed-integer some stopping criterion to avoid endless branching on the continuous components has to be employed.

As neither of the classical approaches for IP can provide a comprehensible decomposition for SIP, research has been devoted to a number of cases which receive a nice treatment. Subsequently, some of these cases will be addressed.

3.2 Simple integer recourse

A two-stage stochastic program with simple integer recourse was defined in [Section 2.2](#) as

$$\min_x \{c^T x + \mathbb{E}_\mu[\Phi(\xi - Tx)]: x \in X\}$$

where

$$\Phi(t) = \min\{(q^+)^T y^+ + (q^-)^T y^- : y^+ \geq t, y^- \geq -t, y^+ \in \mathbb{Z}_+^s, y^- \in \mathbb{Z}_+^s\}.$$

We again use the notation

$$u_i(\chi_i) := \mathbb{E}_{\mu_i}[\lceil \xi_i - \chi_i \rceil^+] \quad \text{and} \quad v_i(\chi_i) := \mathbb{E}_{\mu_i}[\lfloor \chi_i - \xi_i \rfloor^+]$$

to represent the expected surplus and the expected shortage, respectively. Letting

$$\Psi_i(\chi_i) = q_i^+ u_i(\chi_i) + q_i^- v_i(\chi_i) \tag{3.64}$$

the two-stage stochastic program with simple integer recourse can be transformed into

$$\min_x \left\{ c^T x + \sum_{i=1}^s \Psi_i(\chi_i) : \chi = Tx, x \in X \right\}. \tag{3.65}$$

To make the presentation simpler, we now consider the case with expected shortage only. It is defined as

$$\min_x \left\{ c^T x + \sum_{i=1}^s q_i^+ u_i(\chi_i) : \chi = Tx, x \in X \right\}. \tag{3.66}$$

Note that all results available for the expected shortage easily translate to the expected surplus, and therefore also to the functions $\Psi_i(\chi_i)$, $i = 1, \dots, s$. For notational convenience, we will drop the index i whenever it is not required.

Proposition 16. $u(\chi + 1) - u(\chi)$ is a nondecreasing function of χ .

Proof. From Proposition 8, we have

$$u(\chi) = \sum_{k=0}^{\infty} (1 - F(\chi + k)).$$

It follows that, for all $n \in \mathbb{Z}_+$, we have

$$u(\chi + n) = u(\chi) - \sum_{k=0}^{n-1} (1 - F(\chi + k)). \quad (3.67)$$

Taking $n = 1$, we get

$$u(\chi + 1) - u(\chi) = F(\chi) - 1.$$

The proposition holds as $F(\cdot)$ is a cumulative distribution function. \square

If we consider χ values which are integer apart, we may draw a piecewise linear function through successive points $(\chi \pm k, u(\chi \pm k))$, k integer. This piecewise linear function is convex by Proposition 16. It may sometimes be the convex hull of $u(\chi)$. A sufficient condition for that is that the support of ξ is a subset of \mathbb{Z} . But in general, it is not. Take the simple example where $\xi = 1/2$ or $3/2$ with probability $1/2$ each, and observe that $u(1/4) = 3/2 > \frac{1}{2}(u(0) + u(1/2))$, as $u(0) = 3/2$ and $u(1/2) = 1/2$. In any case, as this piecewise linear function is convex, we can derive valid inequalities in the $(\chi, u(\chi))$ space which are supporting half-lines of this function.

Proposition 17. Let $\eta \in \mathbb{Z}$. Define $\pi = u(\eta) - u(\eta + 1)$ and $e = (\eta + 1)u(\eta) - \eta u(\eta + 1)$. Then, for all $\chi \in \mathbb{Z}$,

$$u(\chi) \geq e - \pi\chi. \quad (3.68)$$

Moreover

$$u(\eta) = e - \pi \cdot \eta. \quad (3.69)$$

Proof. Consider the case where $\chi \geq \eta$. Then

$$u(\chi) - u(\eta) = \sum_{k=0}^{\chi-\eta-1} [u(\chi - k) - u(\chi - k - 1)].$$

By [Proposition 16](#), each term in the sum is bounded below by $u(\eta + 1) - u(\eta)$. Hence $u(\chi) - u(\eta) \geq (\chi - \eta)(u(\eta + 1) - u(\eta))$.

The inequality [\(3.68\)](#) follows. The case where $\chi \leq \eta$ is similar. Finally, [\(3.69\)](#) is obtained by straightforward computation of $e - \pi\eta$. \square

We now propose an exact algorithm for the case where χ is integer.

Algorithm SPSIR

Step 1 Initialization: $v := 0$, $r_i := 0$, $i = 1, \dots, s$.

Step 2 Current problem: Let $v := v + 1$. Solve the program

$$\min_x \left\{ c^T x + \sum_{i=1}^s q_i^+ \theta_i : \chi = Tx, x \in X, \pi_{\ell(i)} \chi_i + \theta_i \geq e_{\ell(i)}, \ell(i) = 1, \dots, r_i \right\}. \quad (3.70)$$

Let $(x^v, \theta_1^v, \dots, \theta_s^v)$ be an optimal solution to [\(3.70\)](#). If $r_i = 0$ for some i , θ_i^v is set to $-\infty$ and is not considered in the computation of x^v .

Step 3 Termination: Let $\chi^v = Tx^v$. If $\theta_i^v = u_i(\chi_i^v)$ for $i = 1, \dots, s$, then x^v is an optimal solution. Stop.

Step 4 Cut generation: For any i , $i = 1, \dots, s$, such that $\theta_i^v < u_i(\chi_i^v)$, compute

$$\pi_{r_i+1} = u_i(\chi_i^v) - u_i(\chi_i^v + 1)$$

and

$$e_{r_i+1} = (\chi_i^v + 1)u_i(\chi_i^v) - \chi_i^v u_i(\chi_i^v + 1).$$

Set $r_i := r_i + 1$. Go to step 2.

Proposition 18. Assume X is bounded. Also assume $\chi \in \mathbb{Z}^s$ for all $x \in X$. Then, the SPSIR algorithm finds an optimal solution to [\(3.66\)](#) in a finite number of steps, provided $u_i(\chi_i)$ can be obtained through a finite computation, $i = 1, \dots, s$.

Proof. X being bounded, there are only finitely many different values of χ_i , $i = 1, \dots, s$. For each χ_i , only one cut can be generated. Thus, r_i is finite. By [\(3.69\)](#), the same cut can only be generated once in step 4. \square

Example 19. Newsboy Problem Revisited. Assume now that the newsboy can purchase a number of publications (newspaper, magazines, ...). Each publication i , $i = 1, \dots, s$, has a purchase cost c_i , a selling cost s_i and a demand which is a random variable ξ_i with cumulative distribution function F_i . For simplicity, we assume unsold publications cannot be returned to the vendor.

Each publication has a weight (or volume) a_i . A total weight (or volume) b is available to the newsboy.

The problem is naturally an integer program as publications sell by the unit. It reads as follows

$$\min \left\{ \sum_{i=1}^s c_i x_i + \mathbb{E}_\mu \left[- \sum_{i=1}^s s_i y_i(\xi_i) \right] : \sum_{i=1}^s a_i x_i \leq b, \right. \\ \left. 0 \leq y_i(\xi_i) \leq \xi_i, y_i(\xi_i) \leq x_i, x \in \mathbb{Z}_+^s, y(\xi) \in \mathbb{Z}_+^s \right\}.$$

Letting $y_i^+(\xi_i) = \xi_i - y_i(\xi_i)$, one obtains an equivalent formulation

$$\min \left\{ \sum_{i=1}^s c_i x_i + \mathbb{E}_\mu \left[\sum_{i=1}^s s_i y_i^+(\xi_i) \right] - \mathbb{E}_\mu \left[\sum_{i=1}^s \xi_i \right] : \right. \\ \left. \sum_{i=1}^s a_i x_i \leq b, y_i^+(\xi_i) \geq \xi_i - x_i, x \in \mathbb{Z}_+^s, y^+(\xi) \in \mathbb{Z}_+^s \right\}.$$

Omitting the constant term and using the notation $u_i(x_i)$ previously defined for the expected surplus, the newsboy problem becomes

$$\min \left\{ \sum_{i=1}^s c_i x_i + \sum_{i=1}^s s_i u_i(x_i) : \sum_{i=1}^s a_i x_i \leq b, x \in \mathbb{Z}_+^s \right\}.$$

This problem obviously satisfies the assumptions in [Proposition 18](#): X is bounded through the upper limit on the total weight and $\chi = x$ is integer by definition.

It now remains to study a number of cases where the computation of $u(\chi)$ is finite.

(i) ξ has a finite range.

This case includes random variables with finite support, but also continuous random variables on a finite range. For instance, if ξ has a uniform density on $[0, a]$, then for $0 \leq x \leq a$,

$$u(x) = \sum_{k=0}^{\lceil a-x \rceil - 1} [1 - F(x+k)].$$

(ii) Closed form expressions can be found.

Let ξ have a negative exponential density with parameter $\lambda > 0$. Then, for $x \geq 0$,

$$\begin{aligned} u(x) &= \sum_{k=0}^{\infty} (1 - F(x+k)) = \sum_{k=0}^{\infty} e^{-\lambda(x+k)} \\ &= \frac{e^{-\lambda x}}{1 - e^{-\lambda}} \end{aligned}$$

(iii) The support of ξ is a subset of \mathbb{Z} .

By (3.67), we have

$$u(\chi + n) = u(\chi) - \sum_{k=0}^{n-1} (1 - F(\chi + k)).$$

Observe that $F(t) = F(\lfloor t \rfloor)$ for all $t \in \mathbb{R}$, as the support of $\xi \in \mathbb{Z}$. Thus, $u(t) = u(\lfloor t \rfloor)$ for all $t \in \mathbb{R}$. Consider $x \geq 0$. Apply (3.67) with $\chi=0$ and $n=\lfloor x \rfloor$. It follows that

$$u(x) = u(0) - n + \sum_{k=0}^{n-1} F(k).$$

Now, as the support of ξ is a subset of \mathbb{Z} , $u(0) = \mathbb{E}[\lceil \xi \rceil^+] = \mathbb{E}[\xi^+] = \mathbb{E}[\max(\xi, 0)]$. In particular, for such distributions with $\xi \geq 0$, $u(0) = \mathbb{E}[\xi]$. Such is the case for a Poisson distribution, for instance.

Example 20. Newsboy Problem Continued. Take the newsboy problem with $s=2$, $c^T=(1, 2)$, $s_1=3$, $s_2=7$, $a^T=(2, 3)$, $b=12$. Assume the demand for both publications follows a Poisson distribution with parameter 3. Assume a starting point of $x^T=(0, 0)$. The initial cuts are found as follows.

By definition, $u(0)=3$. Compute $u(1)=u(0)+F(0)-1=2.0498$. Then $\pi_1=0.9502$ and $e_1=3$.

The next iterate will be $x^T=(0, 4)$, with one extra cut generated, then $x^T=(3, 2)$ with two new cuts, then $x^T=(3, 2)$ again, which is optimal. Observe that the mean value optimum is $x^T=(1, 3)$.

Note, finally, that even for such continuous densities as the normal distribution, it is possible to compute $u(x)$ within a fixed tolerance in a finite number of steps (see Theorem 3.30 in Birge and Louveaux, 1997). This tolerance can be chosen equal to the machine tolerance so that the computation can be considered exact.

3.3 Binary first-stage variables

When the first-stage variables are binary, it is possible to obtain a sufficient set of optimality cuts at each $x \in X \cap \{0, 1\}^{n_1}$. A finite algorithm, called the integer L -shaped, has been designed. It can be made efficient for hard problems when lower bounding functionals are available.

Assumption 21. There exists a finite lower bound L satisfying

$$L \leq \min_x \{Q(x) : x \in X\}.$$

Assumption 22. For $x \in X$, $Q(x)$ is computable in a finite number of steps.

At a given stage of the algorithm, we consider the current problem (CP)

$$(\text{CP})_\rho \quad \min_{x, \theta} \{c^T x + \theta : x \in P \cap N^\rho, \theta \geq L, \theta \geq f_\ell(x), \ell = 1, \dots, s\}$$

where P is a formulation of X , N^ρ is node ρ of the first-stage branching scheme, $\mathbb{R}^{n_1} = \cup_{\rho=1, \dots, R} N^\rho$. L is a finite lower bound as in [Assumption 21](#) and $\theta \geq f_\ell(x)$, $\ell = 1, \dots, s$, are the optimality cuts. We now present a general scheme.

Integer L -shaped algorithm

Step 1 Initialization: Let $s=0$, $v=0$, $\rho=1$, $\bar{z} = \infty$, $\theta := L$. A list is created that contains one single node corresponding to the initial problem, i.e., $N^1 := \mathbb{R}^{n_1}$.

Step 2 Selection: Select one node ρ in the list, if none exists, stop.

Step 3 Solution: Set $v := v + 1$. Solve $(\text{CP})_\rho$. If it is infeasible, fathom node N^ρ and go to step 2. Otherwise, let (x^v, θ^v) be an optimal solution. If $c^T x^v + \theta^v \geq \bar{z}$, fathom node N^ρ and go to step 2.

Step 4 Branch & cut: Check for integrality restrictions. If some restriction is violated, apply a separation algorithm to find a valid inequality. If some is found, adapt P , then return to step 3. If not, create two new nodes following the usual branching. Append the nodes to the list, fathom node N^ρ and go to step 2.

Step 5 Second-stage value. Compute $Q(x^v)$ and $z^v = c^T x^v + Q(x^v)$. If $z^v < \bar{z}$, update $\bar{z} := z^v$.

Step 6 Optimality cuts: If $\theta^v \geq Q(x^v)$, fathom node N^ρ and go to step 2. Otherwise, find some optimality cut, set $s := s + 1$ and go to step 3.

The integer L -shaped method yields an optimal solution (when one exists) in a finite number of steps when a sufficient set of optimality cuts exists for each $x \in X$. As $X \subseteq \{0, 1\}^{n_1}$, step 6 can only be performed a finite number

of times. All other steps of the method are finite, as the branch & cut procedure is finite.

When needed, the integer L -shaped can be used in a multicut version. Then, θ is replaced by $\sum_{k=1,\dots,K} \theta^k$, and the restriction $\theta \geq Q(x)$ becomes $\theta^k \geq Q(x, \xi^k)$, $k = 1, \dots, K$. In this case, step 6 becomes

Step 6 *Optimality cuts in the multicut version:* If $(\theta^k)^v \geq Q(x^v, \xi^k)$, for all $k = 1, \dots, K$, fathom node N^ρ and go to step 2. Otherwise, for each k such that $(\theta^k)^v < Q(x^v, \xi^k)$, find some optimality cut, adapt s_k and go to step 3.

Based on the fact that the first-stage variables are binary, we easily obtain a sufficient set of optimality cuts.

Proposition 23. *Let $x_i = 1$, $i \in S$, and $x_i = 0$, $i \notin S$, be some first-stage feasible solution. Let $q_S = Q(x)$ be the corresponding recourse function value. Define the optimality cut as*

$$\theta \geq (q_S - L) \left(\sum_{i \in S} x_i - \sum_{i \notin S} x_i \right) - (q_S - L)(|S| - 1) + L. \quad (3.71)$$

Then, the optimality cut (3.71) is sufficient at x .

Proof. Let $\delta(x, S) = \sum_{i \in S} x_i - \sum_{i \notin S} x_i$. We have $\delta(S) \leq |S|$. $\delta(S) = |S|$ only if $x_i = 1$, $i \in S$, and $x_i = 0$, $i \notin S$. In that case, the right-hand side of (3.71) takes the value q_S and $\theta \geq q_S$ is valid as q_S is precisely $Q(x)$. In all other cases, $\delta(S) \leq |S| - 1$. Then the right hand side of (3.71) takes a value smaller than or equal to L and $\theta \geq L$ is valid by [Assumption 21](#). This single cut is sufficient at x since $Q(x) = q_S$. \square

An alternative is to consider the integer L -shaped as a particular case of the cutting plane decomposition of [Section 3.1.1](#), with

$$F(t) = \begin{cases} \phi(t) & \text{for } t \geq \xi - Tx^v, \\ L & \text{otherwise.} \end{cases}$$

To avoid sending too many optimality cuts, it is helpful to add a number of lower bounding functionals on $Q(x)$. One general possibility is to add cuts from the continuous L -shaped. Let

$$\Gamma(x, \xi) = \min\{q^T y : Wy = \xi - Tx, y \in \mathbb{R}_+^{n_2}\}$$

and $\Gamma(x) = \mathbb{E}_\mu[\Gamma(x, \xi)]$. Then,

$$\theta \geq \Gamma(x^\nu) + \partial\Gamma(x^\nu)^T(x - x^\nu) \quad (3.72)$$

is a valid lower bounding functional. We now present two different situations where the optimality cuts can be improved and the integer L -shaped becomes more efficient.

s-Neighbors

When more information is available on $Q(x)$, improvements on (3.71) can be obtained. Let

$$N(s, S) = \{x: \delta(x, S) = |S| - s, x \in X\}$$

be the s -neighbors of S . Assume we can find

$$\lambda(s, S) \leq \min\{Q(x): x \in N(s, S)\}, s = 0, \dots, |S|,$$

or at least a series of $\lambda(s, S)$, $s \leq t$. Observe that $\lambda(0, S) = q_S$.

Proposition 24. *Let $x_i = 1, i \in S, x_i = 0, i \notin S$, be a feasible solution to SIP, with $q_S = Q(x)$. Define $a = \max\{q_S - \lambda(1, S), (q_S - L)/2\}$. Then*

$$\theta \geq a\delta(x, S) + q_S - a|S| \quad (3.73)$$

is a sufficient optimality cut at x .

Proof. For $x \in N(s, S)$, the right-hand side of (3.73) is equal to $q_S - as$. We show that this value is a valid lower bound on $Q(x)$. This is obvious for $s = 0$. When $s = 1$, the r.h.s. is $q_S - a$. By definition of a , $q_S - a$ is bounded above by $q_S - (q_S - \lambda(1, S)) = \lambda(1, S)$, which is by definition a lower bound on one neighbors of S . When $s = 2$, $q - 2a \leq q_S - 2(q_S - L)/2 = L$. Finally, for $s \geq 3$, $q_S - as \leq q_S - 2a \leq L$. (3.73) is sufficient at x as $\theta \geq q_S$ for $s = 0$. \square

Geometrically, (3.73) defines a half-space in the (δ, θ) space, above a line passing through the two points $(|S|, q_S)$ and $(|S| - 1, \lambda(1, S))$ when $a = q_S - \lambda(1, S)$, or the two points $(|S|, q_S)$ and $(|S| - 2, L)$ where $a = (q_S - L)/2$.

Proposition 25. *Let $x_i = 1, i \in S, x_i = 0, i \notin S$, be a feasible solution, with $q_S = Q(x)$. Let $1 \leq t \leq |S|$ be some integer. Then (3.73) holds with*

$$a = \max_{s \leq t} \{\max(q_S - \lambda(s, S)/s; (q_S - L)/(t + 1)\}. \quad (3.74)$$

Proof. If $x \in N(s, S)$, the right-hand-side of (3.73) is $q_S - as$. By (3.74), for all $s \leq t$, $q_S - as \leq q_S - (q_S - \lambda(s, S)) = \lambda(s, S)$ which is a lower bound on $Q(x)$ by definition. For $s > t$, $q_S - as \leq q_S - a(t+1) \leq q_S - (q_S - L) = L$ which is also valid. \square

Proposition 26. Assume $q_S > \lambda(1, S)$. If $\lambda(s-1, S) - \lambda(s, S)$ is nonincreasing in s for $1 \leq s \leq \lfloor (q_S - L)/(q_S - \lambda(1, S)) \rfloor$, then (3.73) holds with $a = q_S - \lambda(1, S)$.

Proof. It suffices to show that in (3.74), the maximum in the right-hand side is obtained when $s = 1$. Let $t = \lfloor (q_S - L)/(q_S - \lambda(1, S)) \rfloor$. For $s \leq t$, we have $q_S - \lambda(s, S) = \sum_{i=1}^s (\lambda(i-1, S) - \lambda(i, S))$. By assumption, each term of the sum is smaller than the first term of the sum, so the total is less than s times $q_S - \lambda(1, S)$. By definition of t , we have $t+1 \geq (q_S - L)/(q_S - \lambda(1, S))$ or $q_S - \lambda(1, S) \geq (q_S - L)/(t+1)$. \square

Example 27. Let $Q(x) = \sum_{j=1}^{m_2} Q_j(x)$ with

$$Q_j(x, \xi) = \min \left\{ r_j y_j : d_j y_j \geq d_j - \sum_{i \in T(j)} \xi_{ij} x_i, y_j \in \{0, 1\} \right\}$$

and

$$Q_j(x) = \mathbb{E}_\mu[Q_j(x, \xi)].$$

Assume $x_i \in \{0, 1\}$, $i \in T$, with $T = \cup_{j=1, \dots, m_2} T(j)$ and $T(j) \cap T(k) = \emptyset$, $j \neq k$. This can be seen as a number of investments x_i , $i \in T$, which are made in a first stage. They have a random yield ξ_{ij} in product j . Any deficiency in attaining the target d_j for product j results in a penalty r_j . The second-stage value is then simply

$$Q(x) = \sum_{j=1}^{m_2} r_j P \left(\sum_{i \in S(j)} \xi_{ij} < d_j \right) \quad (3.75)$$

where

$$S(j) = \{i \in T(j) : x_i = 1\}, \quad j = 1, \dots, m_2, \text{ at the current solution } x.$$

To apply Proposition 26, take $S = \cup_{j=1, \dots, m_2} S(j)$. By definition, $q_S = Q(x)$ and is easily computed from (3.75) when the sum of the random variables ξ_{ij} , $i \in S(j)$, has a known distribution, $j = 1, \dots, m_2$. Such is case for standard distributions as the Poisson or normal distribution. There are two ways 1-neighbors can be obtained. First, for one j , $1 \leq j \leq m_2$, one x_i , $i \in S(j)$, goes

from one to zero (and all other x_i 's are unchanged). In that case, $Q(x)$ is increased and any value smaller than or equal to q_S is a valid lower bound. Second, we may have one single x_i , $i \notin S$, going from zero to one, and again all other x_i 's unchanged. As the $T(j)$, $j = 1, \dots, m_2$, form a partition, one single term is modified in (3.75). We easily obtain a lower bound by simply assuming this term vanishes. Thus

$$q_S - \lambda(1, S) \leq \max_{j=1, \dots, m_2} \left\{ r_j P \left(\sum_{i \in S(j)} \xi_{ij} < d_j \right) \right\}$$

Without loss of generality, order the j 's in decreasing order of $r_j P(\sum_{i \in S(j)} \xi_{ij} < d_j)$. Repeating the argument above, we get

$$q_S - \lambda(s, S) \leq \sum_{j=1}^s r_j P \left(\sum_{i \in S(j)} \xi_{ij} < d_j \right)$$

and **Proposition 26** applies.

Vehicle routing problems

The integer L -shaped method can be used to solve hard problems for specific applications. We now illustrate this on an example of a routing problem. The capacitated vehicle routing problem, is defined on an undirected graph $G = (V, E)$ where $V = \{v_1, \dots, v_n\}$ is a vertex set and $E = \{(v_i, v_j) : v_i, v_j \in V, i < j\}$ is an edge set. Vertex v_1 is a depot at which are based m identical vehicles of capacity D , while the remaining vertices are customers. A symmetric travel cost matrix $C = (c_{ij})$ is defined on E . With each customer v_i is associated a nonnegative demand to be collected or delivered, but not both. Without loss of generality, we consider the first case here. In the classical VRP, each v_i has a known demand d_i . The problem then consists of designing m vehicle routes: (i) each starting and ending at the depot, (ii) such that every customer is visited only once by one vehicle, (iii) the total demand of any route does not exceed D , and (iv) the total routing cost is minimized. The VRP is known to be NP-hard.

In the stochastic case, each customer v_i has a stochastic demand ξ_i . The consequence of having stochastic demands is that a planned vehicle route may fail at a given customer location whenever the accumulated demand exceeds D . In such a case, a failure is said to occur and a recourse action must be implemented. The stochastic VRP can easily be represented as

$$\min\{cx + Q(x) : x \in X\}$$

where X defines the usual restrictions on the routes (degree constraints at the nodes, no subtour, and expected demand of any route does not exceed the vehicle capacity). $x = (x_{ij})$ is defined as the arc variables, with $x_{ij} = 1$ if (v_i, v_j) belongs to a route and $x_{ij} = 0$ otherwise, and $Q(x)$ is the expected cost of recourse actions in case of failure. This expected cost is separable in the routes and must be computed for each of its two orientations

$$Q(x) = \sum_{k=1}^m \min\{Q^{k,1}, Q^{k,2}\}$$

where $Q^{k,\delta}$ denotes the expected cost of recourse corresponding to route k and orientation $\delta = 1$ or 2 . For a given route k defined by $V_k = (v_{i_1} = v_1, v_{i_2}, \dots, v_{i_t+1} = v_1)$, one orientation corresponds to following the route with customers visited in the natural order. The other orientation corresponds to visiting them in backward order. Observe that orientation does matter in a stochastic setting, while it does not in a deterministic one. Assume the recourse action simply consists of the following steps: return to the depot, unload the vehicle then resume the route at the customer where failure occurs. This recourse action is called a return trip to the depot. Then, the expected recourse cost for the first orientation is

$$Q^{k,1} = 2 \sum_{j=2}^t \sum_{\ell=1}^{\infty} P\left(\sum_{s=2}^{j-1} \xi_{i_s} \leq \ell D \leq \sum_{s=2}^j \xi_{i_s}\right) c_{1i_j}. \quad (3.76)$$

It can be rewritten as

$$Q^{k,1} = 2 \sum_{j=2}^t \sum_{\ell=1}^{\infty} [F^{j-1}(\ell D) - F^j(\ell D)] c_{1i_j} \quad (3.77)$$

where $F^j(\ell D) = P(\sum_{s=2}^j \xi_{i_s} \leq \ell D)$. If $\xi_i \leq D$ a.s. for all i , then the upper limit in the second summation in (3.76) and (3.77) can be brought down to $j-1$.

- (i) Lower bounding functionals on $Q(x)$ can be obtained at fractional first-stage solutions. They are based on the concept of “partial routes”. A partial route h is specified by two ordered vertex sets $S_h = \{v_1, \dots, v_{s_h}\}$ and $T_h = \{v_1, \dots, v_{t_h}\}$ satisfying $S_h \cap T_h = \{v_1\}$, and a third set U_h satisfying $S_h \cap U_h = \{v_{s_h}\}$ and $T_h \cap U_h = \{v_{t_h}\}$.

For simplicity, we write $(v_i, v_j) \in S_h$ or T_h if v_i and v_j are consecutive in S_h or T_h . The partial route h induced by these sets is made up of the two chains (v_1, \dots, v_{s_h}) , (v_1, \dots, v_{t_h}) and of the unstructured set U_h . Let $R_h = S_h \cup T_h \cup U_h$. Define

$$W_h(x) = \sum_{(v_i, v_j) \in R_h} x_{ij} - |R_h| + 1.$$

Let P be a lower bound on any solution containing the partial routes and, as usual, L a lower bound on $Q(x)$.

Proposition 28. *The constraint*

$$\theta \geq L + (P - L) \left(\sum_{h=1}^r W_h(x) - r + 1 \right) \quad (3.78)$$

is a valid inequality for SVRP.

Proof. By construction $W_h(x) \leq 1$. Hence, $\sum_{h=1}^r W_h(x) - r + 1 \leq 1$. It is only 1 when $W_h(x) = 1$ for all h . In this case, (3.78) becomes $\theta \geq P$. Otherwise (3.78) is redundant. \square

A greedy heuristic to find out partial routes proves to be an efficient separation algorithm to detect violated inequalities (3.78). We construct $P = \sum_{h=1}^{r+1} P_h$ as follows. For $h \leq r$, we create an artificial customer v_a with demand $\xi_a = \sum_{v_i \in U_h \setminus \{v_{s_h}, v_{t_h}\}} \xi_i$ and $c_{1a} = \min_{v_i \in U_h \setminus \{v_{s_h}, v_{t_h}\}} \{c_{1i}\}$. Then construct route k equal to $\{v_1, \dots, v_{s_h}, v_a, v_{t_h}, \dots, v_1\}$ and compute $P_h = \min\{Q^{k,1}, Q^{k,2}\}$ as before for this artificial route. P_{r+1} is a lower bound on the expected recourse restricted to the customer set $V \setminus \bigcup_{h=1}^r R_h$ and $m-r$ vehicles. It is similar to the computation of L for $Q(x)$, that is now described in the next paragraph.

- (ii) Lower bound on $Q(x)$. Relabel all customers in nondecreasing order of their distance to the depot. Denote by X_k the random demand on route k , $k = 1, \dots, m$, and let $F_k(\cdot)$ be its distribution function. Let X_T be the total random demand and $F_T(\cdot)$ its distribution function.

Proposition 29. *Let $\pi(F_k, D)$ be a lower bound on the probability of having at least one failure on a route whose demand is defined by F_k . A valid lower bound on $Q(x)$ is given by*

$$\begin{aligned} L = & \inf_{(F_1, \dots, F_m)} \left\{ 2 \sum_{k=1}^m c_{1,k+1} \pi(F_k, D) : \right. \\ & \int \cdots \int_{x_1 + \dots + x_m \leq x} dF_1(x_1) \cdots dF_m(x_m) = F_T(x) \quad \text{for all } x, \\ & \left. F_k \in \mathbb{F}_k, k = 1, \dots, m \right\} \end{aligned} \quad (3.79)$$

where \mathbb{F}_k is a family of distribution functions to be specified. In particular, it must be such that $\mathbb{E}_\mu[X_k] \leq D$.

Proof. From (3.77), the expected cost $Q^{k,\delta}$ of route k with orientation δ is obtained by computing the cost of having the ℓ th failure at the j th customer, then summing up over all ℓ and j . Each of these terms contribute to $Q^{k,\delta}$ by a nonnegative amount. A valid lower bound is obtained by only considering the first failure, $\ell=1$. By definition of $\pi(F_k, D)$, we obtain $Q(x) \geq 2 \sum_{k=1}^m \gamma_k \cdot \pi(F_k, D)$ for any lower bound γ_k on the distance of a customer in route k to the depot. We may then replace $\gamma_1, \dots, \gamma_m$ by the m least distances to the depot, to obtain the objective function in (3.79). \square

Apply this proposition to the case where \mathbb{F}_k is the set of normal distributions. Let $x_k = \mathbb{E}[X_k]$ and $y_k = \text{Var}[X_k]$. Let also $\mu_T = \mathbb{E}[X_T]$ and $\sigma_T^2 = \text{Var}[X_T]$. Let $x = (x_1, \dots, x_m)$, $y = (y_1, \dots, y_m)$.

Then (3.79) reduces to

$$L = \min_{x, y} \left\{ 2 \sum_{k=1}^m c_{1,k+1} \left[1 - G\left(\frac{D - x_k}{\sqrt{y_k}}\right) \right] : \sum_{k=1}^m x_k = \mu_T, \sum_{k=1}^m y_k = \sigma_T^2, y \geq 0, x \geq 0 \right\}. \quad (3.80)$$

where $G(t) = P(Z \leq t)$, $Z \sim N(0, 1)$.

As the objective function is neither convex nor concave, we obtain a more workable problem by replacing y_k by some lower bound y_0 . We know that for each route $x_k \geq x_0 := \max\{\mu_T - (m-1)D; \min_{i=2, \dots, n} \mathbb{E}[\xi_i]\}$. Then, $y_0 = \min_Z \{\sum_{i=2}^n \text{Var}[\xi_i] z_i : \sum_{i=2}^n \mathbb{E}[\xi_i] z_i \geq x_0, z_i \in \{0, 1\}\}$ is a lower bound on y_k . We may thus replace (3.80) by

$$L = \min_x \left\{ 2 \sum_{k=1}^m c_{1,k+1} \left[1 - G\left(\frac{D - x_k}{\sqrt{y_0}}\right) \right] : \sum_{k=1}^m x_k = \mu_T, x \geq 0 \right\}. \quad (3.81)$$

The objective function in (3.81) is convex in x_k for $x_k \leq D$, so that (3.81) can be solved by applying the Karush–Kuhn–Tucker conditions. Let γ be the multiplier on $\sum_{k=1}^m x_k = \mu_T$. Define $b = -\gamma \sqrt{\pi y_0 / 2}$. Then

$$x_k = \begin{cases} D & \text{if } c_{1,k+1} \leq b \\ D - \sqrt{2y_0 \ln(c_{1,k+1}/b)} & \text{if } c_{1,k+1} \geq b. \end{cases}$$

As x_k , $k = 1, \dots, m$, is a nondecreasing function of b , they can be determined recursively up to the moment where $\sum_{k=1}^m x_k = \mu_T$ within some tolerance. Similar results apply for other cases, such as the Poisson distribution.

The combination of lower bounding functionals on partial routes and the calculation of a lower bound L based on (3.81) for the normal case prove to be very efficient.

3.4 Second-stage integer variables

We now consider the case where $Y \subseteq \mathbb{Z}_+^{n_2}$. Assume again T and W are fixed. Suppose also that the second-stage program for a given ξ can be represented as

$$\Psi(\chi, \xi(\omega)) = \min\{q^T y : Wy \geq \xi - \chi, y \in Y\}$$

where $Y \subseteq \mathbb{Z}_+^{n_2}$. Assume also that Ξ has a finite support.

Proposition 30. *Let W be integer. Assume $\{u \in \mathbb{R}^{m_2} : W^T u \leq q\} \neq \emptyset$. For every $\xi \in \Xi$ and $j = 1, \dots, m_2$, $\Psi(\chi_j, \xi)$ is lower semicontinuous and nondecreasing in χ_j . Moreover, for any $h \in \mathbb{Z}$, $\Psi(\chi_j, \xi)$ is constant over $\chi_j \in (h - \xi_j - 1, h - \xi_j]$, $\xi \in \Xi$, $j = 1, \dots, m_2$.*

Proof. The first part of the proposition comes from Proposition 2. Any solution $y \in \arg \min\{q^T y : Wy \geq \xi - \chi', y \in Y\}$ belongs to $\{y : Wy \geq \xi - \chi, y \in Y\}$ when $\chi' \leq \chi$. So $\Psi(\chi, \xi(w)) \leq \Psi(\chi', \xi(w))$. Define $h = \lceil \xi_j - \chi'_j \rceil$. With integral W , the j th constraint $(Wy)_j \geq \xi_j - \chi'_j$ implies $(Wy)_j \geq h_j$. Hence $\Psi(\chi_j, \xi(\omega))$ is constant over $(h - \xi_j - 1, h - \xi_j]$. \square

It is thus possible to partition the χ space in an orthogonal complex \mathcal{C} , where each cell $C \in \mathcal{C}$ is of the form $\prod_{j=1, \dots, m_2} (a_j, b_j]$. On each of these cells, $\Psi(\chi)$ is constant. The closure of a nonempty cell forms a full-dimensional hyper-rectangle. The nonempty cells form a partition of \mathbb{R}^{m_2} . A branch & bound algorithm can be constructed, where the branching consists of considering other cells of this orthogonal complex than those already considered.

Define, as usual,

$$\Psi(\chi) = \mathbb{E}_\mu[\Psi(\chi, \xi(\omega))]$$

Define the current problem associated to a set S as

$$\text{CP}(S) \quad z = \inf_{x, \chi, \theta} \{cx + \theta : x \in X, \chi \in S, Tx = \chi, \theta \geq f_\ell(x), \ell = 1, \dots, s\}$$

where, as before, $\theta \geq f_\ell(x)$ are optimality cuts or lower bounding functionals. Typically, (3.72) can be used here.

Algorithm for second-stage integer variables

Step 1 Initialization: Set $v=0$, $\bar{z} = \infty$. Choose s and find s valid inequalities $\theta \geq f_\ell(x)$, $\ell = 1, \dots, s$. A list is created that contains the single node $N^1 := \chi$. Set $\rho = 1$.

Step 2 Selection: Select a node N^ρ in the list; if none exists, stop.

Step 3 Current solution: Set $v := v + 1$. Solve $\text{CP}(N^\rho)$. If N^ρ is infeasible, fathom node N^ρ , go to step 2. Otherwise, let (x^v, χ^v, θ^v) be an optimal solution. If $cx^v + \theta^v \geq \bar{z}$, fathom the current node and go to step 2.

Step 4 Cell: Find the cell $C = \prod_{j=1, \dots, m_2} (a_j, b_j]$ s.t. $\chi^v \in C$. Let $b = (b_1, \dots, b_{m_2})$. Compute $\Psi(b, \xi)$ for all $\xi \in \Xi$ and let $\Psi^v = \mathbb{E}_\mu[\Psi(b, \xi)]$.

Step 5 Solution value: Solve

$$v = \inf_{x, \chi} \{c^T x : x \in X, \chi \in C, Tx = \chi\}.$$

Let $\bar{x}, \bar{\chi}$ be an optimal solution. Compute $z^v = c\bar{x} + \Psi^v$. If $\bar{z} < z^v$, update $\bar{z} := z^v$.

Step 6 Branch: Select a component j , $1 \leq j \leq m_2$. Create two new nodes $N^\rho \cap \{\chi : \chi_j \leq a_j\}$ and $N^\rho \cap \{\chi : \chi_j > b_j\}$. Fathom node N^ρ and go to step 2.

The algorithm is finite as the number of cells of the complex is finite since X is compact. Additional valid inequalities can be found along the way in step 3. In practice, open sets and inf would be replaced by closed sets and min, through a well-chosen tolerance on open sets. In the following case, the definition of the cells is easier. When the discrete random variable is such that all realizations of component ξ_j , $j = 1, \dots, m_2$, are of the form $h_j + k$, $k \in \mathbb{Z}$ for some h_j . Then, all cells can be defined as $\prod_{j=1, \dots, m_2} (a_j, a_{j+1}]$.

Sampling

In many cases, the second-stage value function may be hard or impossible to compute, even for a fixed value of the first-stage solution. It is natural then to design methods based on some form of sampling. Such methods are currently used in continuous stochastic programming and are described elsewhere in this handbook. We provide here an introduction to some specific aspects of stochastic discrete programs.

Stochastic branch & bound

At a given stage of the algorithm, one considers a current problem

$$(\text{CP})_\rho \quad z^*(N^\rho) = \min\{c^T x + \mathbb{E}_\mu[\Psi(x, \xi)] : x \in P \cap N^\rho\} \quad (3.82)$$

where P is a formulation of X , N^ρ , is a node of the first-stage branching scheme, $\mathbb{R}^n = \cup_{\rho=1, \dots, R} N^\rho$ and

$$\Psi(x, \xi) = \min\{q^T(\xi)y : W(\xi)y \leq h(\xi) - T(\xi)x, y \in Y\}. \quad (3.83)$$

As $z^*(N^\rho)$ is too difficult to compute, the stochastic branch & bound does try to find a lower bound function $L(N^\rho)$ and an upper bound function $U(N^\rho)$ which satisfy the following two conditions

- (i) for every $N^\rho \neq \emptyset$, $L(N^\rho) \leq z^*(N^\rho) \leq U(N^\rho)$
- (ii) for every singleton $x \in X$, $L(\{x\}) = z^*(\{x\}) = U(\{x\})$. As usual, $z^* = \min_{\rho=1, \dots, R} z^*(N^\rho)$. Also, $z^* \leq \bar{z} = \min_{\rho=1, \dots, R} U(N^\rho)$. Thus, a node N^ρ can only be fathomed when $L(N^\rho) \geq \bar{z}$. Alternatively, a node $N_\ell = \emptyset$ can also be fathomed.

There are various ways to obtain lower and upper bound functions in stochastic discrete programs. One such way is to use the classical wait-and-see (WS) and expected result of using the EV solution (EEV) values. Let $\bar{x}(\xi) \in \arg \min\{c^T x + \Psi(x, \xi) : x \in P \cap N^\rho\}$ be the WS solution for a given $\xi \in \Xi$ and node N^ρ . Similarly, let $x^* \in \arg \min\{c^T x + \mathbb{E}_\mu[\Psi(x, \xi)] : x \in P \cap N^\rho\}$ be a solution of the current problem $(CP)_\rho$. For any $\xi \in \Xi$, $c^T \bar{x}(\xi) + \Psi(\bar{x}(\xi), \xi) \leq c^T x^* + \Psi(x^*, \xi)$. The lower bound function is obtained by taking the expectation of the left-hand-side of this inequality:

$$L(N^\rho) := \mathbb{E}_\mu[\min\{c^T x + \Psi(x, \xi) : x \in P \cap N^\rho\}].$$

By definition, the expectation of the right-hand of the same inequality is precisely $z^*(N^\rho)$, which proves the condition $L(N^\rho) \leq z^*(N^\rho)$ holds. Moreover, for every singleton x , we have $L(\{x\}) = z^*(\{x\}) = c^T x + \mathbb{E}_\mu[\Psi(x, \xi)]$.

Similarly, let $\bar{x}_\rho \in \arg \min\{c^T x + \Psi(x, \mathbb{E}_\mu[\xi]) : x \in P \cap N^\rho\}$ be the solution of the EV problem on node N^ρ . We define

$$U(N^\rho) := \mathbb{E}_\mu[c^T \bar{x}_\rho + \Psi(\bar{x}_\rho, \xi)]$$

which is the EEV value for node N^ρ . As \bar{x}_ρ is a feasible solution to $(CP)_\rho$, it follows that $z^*(N^\rho) \leq U(N^\rho)$ holds. Clearly, for every singleton, $U(\{x\}) = z^*(\{x\})$. We now illustrate why finding $U(N^\rho)$ and, especially, $L(N^\rho)$ is still a hard problem.

Example 31. Project Financing. Assume we can invest in n projects. Project j can be started at a cost c_j . In the second stage, projects which have been started can be continued or not. If project j is continued, it provides a revenue $-q_j(\xi)$. Project j requires t_{ij} units of resource i in the first stage and an additional $w_{ij}(\xi)$ units to be continued. A total of $h_i(\xi)$ units of resource i will be available over the two stages. The two-stage stochastic program reads as $\min\{c^T x + \mathbb{E}_\mu[\Psi(x, \xi)] : x \in P\}$ where P is a formulation of some first-stage constraints and

$$\Psi(x, \xi) = \min\{q^T(\xi)y : W(\xi)y \leq h(\xi) - Tx, y \leq x, y \in \{0, 1\}^n\}. \quad (3.84)$$

The current problem is as (3.82). To obtain an upper bound on N^ρ , we solve the expected value problem

$$\begin{aligned} \min \{ & c^T x + \mathbb{E}[q^T(\xi)]y : Tx + \mathbb{E}[W(\xi)]y \leq \mathbb{E}[h(\xi)], \\ & y \leq x, y \in \{0, 1\}^n, x \in P \cap N^\rho \}. \end{aligned} \quad (3.85)$$

This is a multiknapsack problem. For simplicity, assume $P \subseteq \{0, 1\}^n$. Now, N^ρ will include additional restrictions on x . If $x \in N^\rho$ implies $x_j = 0$ for some j , then x_j and y_j can be removed from the computation in (3.84). If $x \in N^\rho$ implies $x_j = 1$ for some j , then x_j is also removed from the computation in (3.84). Finally, if $x \in N^\rho$ implies $x_j \in \{0, 1\}$, then it follows that (3.84) can be simplified as $y_j = x_j$ in any deterministic solution. Thus, finding \bar{x}_ρ is relatively easy. Yet, finding $U(N^\rho)$ still requires computing $\mathbb{E}_\mu[\Psi(\bar{x}_\rho, \xi)]$. This involves solving a multiknapsack problem for each $\xi \in \Xi$. Unless the support of Ξ is discrete with low cardinality, $\mathbb{E}_\mu[\Psi(\bar{x}_\rho, \xi)]$ cannot be computed exactly, but only estimated through sampling. Similarly, computing $L(N^\rho)$ amounts to solving a deterministic multiknapsack for each $\xi \in \Xi$. The same simplifications as above are available, yet $L(N^\rho)$ cannot be computed exactly but only estimated through sampling.

A stochastic lower bound function can be obtained by Monte Carlo simulation. For i.i.d. observations ξ^k , $k = 1, \dots, S$, one obtains

$$\kappa(N^\rho) = \frac{1}{S} \sum_{k=1}^S \min \{ c^T x + \Psi(x, \xi^k) : x \in P \cap N^\rho \}.$$

For the same sample, one obtains a stochastic upper bound

$$\eta(N^\rho) = \frac{1}{S} \sum_{k=1}^S (c^T \bar{x}_\rho + \Psi(\bar{x}_\rho, \xi^k))$$

where, as before, \bar{x}_ρ is the solution of the EV problem on node N^ρ .

When the lower and upper bounds are random variables, fathoming of the nodes becomes more problematic. Deletion of a node on the basis of the statistical estimates may lead to the loss of the optimal solution. On the other hand, the stochastic lower and upper bounds $\kappa(N^\rho)$ and $\eta(N^\rho)$ can be updated each time a new sample is drawn, in a manner similar to the one which updates cuts in a stochastic decomposition method. As $\lim_{S \rightarrow \infty} \kappa(N^\rho) \rightarrow L(N^\rho)$, a.s. and $\lim_{S \rightarrow \infty} \eta(N^\rho) = U(N^\rho)$ a.s., the errors $|\kappa(N^\rho) - L(N^\rho)|$ and $|\eta(N^\rho) - U(N^\rho)|$ can be bounded in probability. Then, deletion can only occur after a sufficiently larger number of iterations where the estimations of the bounds are improved. Convergence of the stochastic

branch & bound relies on two arguments. As just mentioned, repeated sampling lets the stochastic bounds tend to the lower and upper bounds $L(N^\rho)$ and $U(N^\rho)$. On the other hand, repeated partitioning lets the nodes tend to singletons for which, by construction, $L(\{x\}) = z^*(\{x\}) = U(\{x\})$. A detailed description can be found in [Norkin et al. \(1998\)](#).

Sample average approximation

Consider a sample ξ^k , $k = 1, \dots, S$, of sample scenarios generated from ω according to the probability distribution P . The Sample Average Approximation problem is the following

$$(SAA) \quad z_S = \min \left\{ c^T x + \frac{1}{S} \sum_{k=1}^S \Psi(x, \xi^k) : x \in P \right\}$$

where P is a formulation of X . It is a stochastic program with discrete distribution. It can be solved by the classical L -shaped method if the second-stage is continuous, combined with a branch & cut scheme to recover integrality in the first-stage. The SAA method proceeds by solving problem (SAA) repeatedly. Assume M independent samples, each of size S , are generated and, for each, the corresponding (SAA) problem is solved. We obtain optimal values $z_S^1, z_S^2, \dots, z_S^M$ and associated candidate solutions $\hat{x}^1, \hat{x}^2, \dots, \hat{x}^M$. It is then natural to pick-up one of these as the (estimated) optimal solution. This would require finding the one which minimizes $c^T x + \mathbb{E}_\mu[\Psi(x, \xi)]$. As this calculation is again impossible to carry out, this value is estimated by

$$\hat{z}_N(x) = c^T x + \frac{1}{N} \sum_{n=1}^N \Psi(x, \xi^n)$$

where $\{\xi^1, \dots, \xi^N\}$ is a sample of size N . Typically, N is chosen to be quite large and must be independent of the M samples generated above.

The estimated optimal solution is

$$\hat{x}^* \in \arg \min \{ \hat{z}_N(x) : x \in \{\hat{x}^1, \hat{x}^2, \dots, \hat{x}^M\} \}.$$

Now, for any x , $\hat{z}_N(x)$ is an unbiased estimator of $c^T x + \mathbb{E}_\mu[\Psi(x, \xi)]$, and therefore, for any feasible x , we have $\mathbb{E}_\mu[\hat{z}_N(x)] \geq z^*$. On the other hand, we may also consider the average of the M optimal values of the SAA problems

$$\bar{z}_S = \frac{1}{M} \sum_{m=1}^M z_S^m.$$

As $\mathbb{E}[\bar{z}_S] = WS$, it follows that $\mathbb{E}[\bar{z}_S] \leq z^*$. Thus, the quality of the solution \hat{x}^* can be evaluated by computing the optimality gap estimate $\hat{z}_N(\hat{x}^*) - \bar{z}_S$. As this quantity is random, it is helpful to estimate its variance. As the samples used to compute those quantities are independent, the variance of the gap is simply the sum of the variances of the two terms. These are

$$\hat{\sigma}_{\bar{z}_S}^2 = \frac{1}{(M-1)M} \sum_{m=1}^M (z_S^m - \bar{z}_S)^2$$

and

$$\hat{\sigma}_{\hat{z}_N(\hat{x}^*)}^2 = \frac{1}{(N-1)N} \sum_{n=1}^N (c^T \hat{x}^* + \Psi(\hat{x}, \xi^n) - \hat{z}_N(\hat{x}^*))^2.$$

The above procedure for statistical evaluation of a candidate solution was suggested in [Mak et al. \(1999\)](#). Convergence properties of the SAA method were studied in [Kleywegt et al. \(2001\)](#) and [Ahmed and Shapiro \(2002\)](#). An example of application to routing problems can be found in [Verweij et al. \(2003\)](#).

3.5 Bibliographical notes

The systematic investigation of stochastic integer programs started during the 1990s, only. Survey and introductory articles were published by [Klein Haneveld and van der Vlerk \(1999\)](#) and [Römis and Schultz \(2001\)](#). An annotated bibliography was compiled by [Stougie and van der Vlerk \(1997\)](#). The textbook [Birge and Louveaux \(1997\)](#) has a chapter on stochastic integer programs. Integer programming basics are laid out in the textbooks [Nemhauser and Wolsey \(1988\)](#) and [Wolsey \(1998\)](#), for instance.

The first continuity result for the expected recourse function of a stochastic linear program with integer recourse is due to [Stougie \(1985\)](#). The sufficient conditions for lower semicontinuity, continuity, and Lipschitz continuity of [Propositions 3 and 4](#), with extensions to random technology matrix T in (2.4), were derived in [Schultz \(1993, 1995\)](#). Joint qualitative (semi-) continuity of the expected recourse as a function of the first-stage decision and the integrating probability measure together with conclusions towards stability were addressed in [Artstein and Wets \(1994\)](#), [Schultz \(1992, 1995\)](#). Quantitative joint continuity of the expected recourse function and quantitative stability of the problem (2.3) were studied in [Rachev and Römis \(2002\)](#) and in [Schultz \(1996\)](#).

For the first time, stochastic programs with simple integer recourse were investigated in [Louveaux and van der Vlerk \(1993\)](#). The analysis was pushed ahead in [Klein Haneveld and van der Vlerk \(1994\)](#), [Klein Haneveld et al. \(1996, 1997\)](#), see also the dissertation [van der Vlerk \(1995\)](#).

So far, very little is known about the structure of multi-stage stochastic integer programs. An introduction into this problem class was given in Römisch and Schultz (2001). Our exposition in Section 2.3 essentially follows parts of that paper. Power problems leading to multi-stage stochastic integer programs and specialized methods for their solution were discussed in Dentcheva and Römisch (1998), Nowak (2000), Nowak and Römisch (2000) and Takriti et al. (1996).

More details on the definitions and properties presented in Section 3.1 can be found in Wolsey (1998).

The first use of decomposition methods in stochastic programs having discrete decisions in the second-stage is the integer L -shaped method of Laporte and Louveaux (1993) for the case of first-stage binary variables. This method has been applied to solve a variety of location and routing problems. The example considered in Section 3.3.2 is taken from Laporte et al. (2002). The generalization of the integer L -shaped to mixed integer first-stage was done by Carøe (1998) in his doctoral dissertation and presented in Carøe and Tind (1997, 1998).

The first attempt to design a method based on first-stage integer variables branching is due to Ahmed et al. (2000), see the exposition in Section 3.4. In Schultz et al. (1998) a solution procedure based on enumeration and bounding in the first-stage while handling the second-stage by algebraic methods exploiting problem similarities has been proposed. Another application of algebraic methods to stochastic integer programming has been carried out in Hemmecke and Schultz (2003). The paper deals with test set decomposition. It identifies building blocks that enable generation of improving vectors for the stochastic program and that can be computed by a completion procedure from computational algebra.

The scenario decomposition method displayed in Section 3.1.3 stems from Carøe and Schultz (1999). In Løkketangen and Woodruff (1996) a scenario decomposition method employing Augmented Lagrangians for the dualization and tabu search for the resulting quadratic mixed-integer subproblems has been developed.

Recently, Alonso-Ayuso et al. (2003) have proposed a branch-and-fix coordination approach to solving multi-stage stochastic integer programs. Their idea is to formulate the stochastic program in a splitting variable representation, to perform scenario-wise an LP-based branch & bound, and to establish the nonanticipativity in the course of the branching. The latter is achieved by a coordinated fixing of variables that puts variables for different scenarios at identical values provided a nonanticipativity condition requires the identity.

References

- Ahmed, S., A. Shapiro (2002). The sample average approximation method for stochastic programs with integer recourse, Preprint, Georgia Institute of Technology, Atlanta.

- Ahmed, S., M. Tawarmalani, N.V. Sahinidis (2000). A finite branch and bound algorithm for two-stage stochastic integer programs. Stochastic Programming E-Print Series 16—2000, <http://dochost.rz.hu-berlin.de/speps/>.
- Alonso-Ayuso, A., L.F. Escudero, A. Garín, M.T. Ortuño, G. Pérez (2003). An approach for strategic supply chain planning under uncertainty based on stochastic 0-1 programming. *Journal of Global Optimization* 26, 97–124.
- Artstein, Z., R.J.-B. Wets (1994). Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM Journal on Optimization* 4, 537–550.
- Bacaud, L., C. Lemaréchal, A. Renaud, C. Sagastizábal (2001). Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners. *Computational Optimization and Applications* 20, 227–244.
- Bank, B., J. Guddat, D. Klatte, B. Kummer, K. Tammer (1982). *Non-linear Parametric Optimization*, Akademie-Verlag, Berlin.
- Bank, B., R. Mandel (1988). *Parametric Integer Optimization*, Akademie-Verlag, Berlin.
- Birge, J.R., F.V. Louveaux (1997). *Introduction to Stochastic Programming*, Springer-Verlag, New York.
- Blair, C.E., R.G. Jeroslow (1977). The value function of a mixed integer program: I. *Discrete Mathematics* 19, 121–138.
- Carøe, C.C. (1998). Decomposition in stochastic integer programming, Ph.D. Thesis, Institute of Mathematical Sciences, Department of Operations Research, University of Copenhagen, Denmark.
- Carøe, C.C., R. Schultz (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters* 24, 37–45.
- Carøe, C.C., J. Tind (1997). A cutting plane approach to mixed 0–1 stochastic integer programs. *European Journal of Operations Research* 101, 306–316.
- Carøe, C.C., J. Tind (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming* 83, 139–152.
- Dentcheva, D., W. Römisch (1998). Optimal power generation under uncertainty via stochastic programming, in: K. Marti, P. Kall (eds.), *Stochastic Programming Methods and Technical Applications. Lecture Notes in Economics and Mathematical Systems*, Vol. 458, Springer-Verlag, Berlin, pp. 22–56.
- Dynkin, E.B., I.V. Evstigneev (1976). Regular conditional expectation of correspondences. *Theory of Probability and Applications* 21, 325–338.
- Evstigneev, I. (1976). Measurable selection and dynamic programming. *Mathematics of Operations Research* 1, 267–272.
- Gröwe-Kuska, N., K.C. Kiwiel, M.P. Nowak, W. Römisch, I. Wegner (2002). Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation, in: C. Greengard, A. Ruszczyński (eds.), *Decision Making under Uncertainty: Energy and Power IMA Volumes in Mathematics and its Applications*, Vol. 128, Springer-Verlag, New York.
- Hemmecke, R., R. Schultz (2003). Decomposition of test sets in stochastic integer programming. *Mathematical Programming* 94, 323–341.
- Hiriart-Urruty, J.B., C. Lemaréchal (1993). *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin.
- Kiwiel, K.C. (1990). Proximity control in bundle methods for convex nondifferentiable optimization. *Mathematical Programming* 46, 105–122.
- Klein Haneveld, W.K., M.H. van der Vlerk (1994). On the expected value function of a simple integer recourse problem with random technology matrix. *Journal of Computational and Applied Mathematics* 56, 45–53.
- Klein Haneveld, W.K., L. Stougie, M.H. van der Vlerk (1996). An algorithm for the construction of convex hulls in simple integer recourse programming. *Annals of Operations Research* 64, 67–81.
- Klein Haneveld, W.K., L. Stougie, M.H. van der Vlerk (1997). Convex approximations for simple integer recourse models by perturbing the underlying distribution. Research Report 97A19, SOM, University of Groningen.
- Klein Haneveld, W.K., M.H. van der Vlerk (1999). Stochastic integer programming: General models and algorithms. *Annals of Operations Research* 85, 39–57.

- Kleywegt, A.J., A. Shapiro, T. Homem-de-Mello (2001). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12, 479–502.
- Laporte, G., F.V. Louveaux (1993). The integer L -shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* 13, 133–142.
- Laporte, G., F.V. Louveaux, L. Van Hamme (2002). An integer L -shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research* 50, 415–423.
- Løkketangen, A., D.L. Woodruff (1996). Progressive hedging and tabu search applied to mixed integer (0,1) multi-stage stochastic programming. *Journal of Heuristics* 2, 111–128.
- Louveaux, F.V., M.H. van der Vlerk (1993). Stochastic programming with simple integer recourse. *Mathematical Programming* 61, 301–325.
- Mak, W.K., D.P. Morton, R.K. Wood (1999). Monte-Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* 24, 47–56.
- Nemhauser, G.L., L.A. Wolsey (1988). *Integer and Combinatorial Optimization*, Wiley, New York.
- Norkin, V.I., Y.M. Ermoliev, A. Ruszczyński (1998). On optimal allocation of indivisibles under uncertainty. *Operations Research* 46, 381–395.
- Nowak, M.P. (2000). Stochastic Lagrangian relaxation in power scheduling of a hydro-thermal system under uncertainty. Ph.D. Thesis, Institute of Mathematics, Humboldt-University Berlin.
- Nowak, M.P., W. Römis (2000). Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Annals of Operations Research* 100, 251–272.
- Rachev, S.T., W. Römis (2002). Quantitative stability in stochastic programming: The method of probability metrics. *Mathematics of Operations Research* 27, 792–818.
- Rockafellar, R.T., R.J.-B. Wets (1978). The optimal recourse problem in discrete time: L^1 -multipliers for inequality constraints. *SIAM Journal on Control and Optimization* 16, 16–36.
- Rockafellar, R.T., R.J.-B. Wets (1997). *Variational Analysis*, Springer-Verlag, Berlin.
- Römis, W., R. Schultz (2001). Multistage stochastic integer programs: an introduction, in: M. Grötschel, S.O. Krumke, J. Rambau (eds.), *Online Optimization of Large Scale Systems*, Springer-Verlag, Berlin, pp. 581–600.
- Schultz, R. (1992). Continuity and stability in two-stage stochastic integer programming, in: Marti, K. (ed.), *Stochastic Optimization, Numerical Methods and Technical Applications. Lecture Notes in Economics and Mathematical Systems*, Vol. 379, Springer, Berlin.
- Schultz, R. (1993). Continuity properties of expectation functions in stochastic integer programming. *Mathematics of Operations Research* 18, 578–589.
- Schultz, R. (1995). On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Mathematical Programming* 70, 73–89.
- Schultz, R. (1996). Rates of convergence in stochastic programs with complete integer recourse. *SIAM Journal on Optimization* 6, 1138–1152.
- Schultz, R., L. Stougie, M.H. van der Vlerk (1998). Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Mathematical Programming* 83, 229–252.
- Stougie, L. (1985). Design and analysis of algorithms for stochastic integer programming. Ph.D. Thesis, Center for Mathematics and Computer Science, Amsterdam.
- Stougie, L., M.H. van der Vlerk (1997). Stochastic integer programming: *Annotated Bibliographies in Combinatorial Optimization*, in: M. Dell’Amico, F. Maffioli, S. Martello (eds.), Wiley, Chichester, pp. 127–141.
- Takriti, S., J.R. Birge, E. Long (1996). A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems* 11, 1497–1508.
- van der Vlerk, M.H. (1995). Stochastic programming with integer recourse. Ph.D. Thesis, University of Groningen.
- Verweij, B., S. Ahmed, A.J. Kleywegt, G.L. Nemhauser, A. Shapiro (2003). The sample average approximation method applied to stochastic routing problems: a computational study. *Computational Optimization and Applications* 24, 289–333.
- Wolsey, L.A. (1998). *Integer Programming*, Wiley, New York.