Multistage Stochastic Integer Programs: An Introduction

Werner Römisch[1] and Rüdiger Schultz[2]

[1] Humboldt-University Berlin, Institute of Mathematics, Berlin
[2] Gerhard-Mercator University Duisburg, Department of Mathematics, Duisburg

Abstract. We consider linear multistage stochastic integer programs and study their functional and dynamic programming formulations as well as conditions for optimality and stability of solutions. Furthermore, we study the application of the Rockafellar-Wets dualization approach as well as the structure and algorithmic potential of corresponding dual problems. For discrete underlying probability distributions we discuss possible large scale mixed-integer linear programming formulations and three dual decomposition approaches, namely, *scenario*, *component* and *nodal decomposition*.

1 Introduction

Stochastic programming deals with the optimization of decision making under uncertainty over time. Typical objects of study are random optimization problems where outcomes of random data are unveiled over time, and the decisions to be optimized must not anticipate future outcomes (non-anticipativity). The latter provides a tight link to real-time optimization seen as the need for optimal "here-and-now" decision in an incomplete (or uncertain) data environment. Provided that probabilistic information on the uncertain data is available, operational models suitable for real-time optimization often may be formulated as multi-stage stochastic programs. Basic references for theory, algorithmics, and application of stochastic programming are the textbooks [7,24,34]. The edited volume [50] provides insight into recent research in the field.
Indispensability of integer requirements is a basic modeling experience in practical optimization. Like in other branches of mathematical optimization this has considerable consequences on structural properties and algorithm design in stochastic programming, too. The models best understood so far are (purely) linear stochastic programs. This is mainly due to the fact that the optimal value of a linear minimization problem is a convex function of the right-hand side and a concave function of the objective function vector. This enables application of the machinery of convex analysis in various contexts, such as duality, stability, and subgradient minimization. For an impression on these developments we refer to [12,14,36,49], with accent on theory, and to [5,42], with accent on computation.
With integer requirements, the above convexity/concavity observation is no

longer valid, and the mentioned functions become discontinuous. Thus, comparatively little is known on theory and algorithms for mixed-integer linear stochastic programs. A recent survey is provided in [26]. Impressions on developments in theory can be obtained from [2,43,46] and on algorithm design from [9,11,19,27,30,33,47], see also the Ph.D. thesises [8,32,48].
The present paper aims at a short introduction into some essential theoretical and algorithmic issues in multi-stage stochastic integer programming. Accent is placed on introducing approaches. Proofs are omitted, with references to the original sources instead. The main topics will be modeling, approximation, and algorithmics.

2 Multistage Stochastic Integer Programs

2.1 Modeling

We consider a finite horizon sequential decision process under uncertainty, in which a decision made at stage t is based only on information available at t $(1 \leq t \leq T)$. We assume that the information is given by a discrete time stochastic process $\{\xi_t\}_{t=1}^{T}$ defined on some probability space $(\Omega, \mathcal{F}, P)$ and with ξ_t taking values in $\mathbb{R}^{s_t}$. The information available at stage t consists of the random vector $\xi^t := (\xi_1, \ldots, \xi_t)$, and the stochastic decision x_t at stage t varying in $\mathbb{R}^{m_t}$ is assumed to depend only on ξ^t. The latter property is called *nonanticipativity* and is equivalent to the measurability of x_t with respect to the σ-algebra $\mathcal{F}_t \subseteq \mathcal{F}$ which is generated by ξ^t. Clearly, we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, T-1$ and, with no loss of generality, we may assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., ξ_1 and x_1 are deterministic, and that $\mathcal{F}_T = \mathcal{F}$.
More precisely, we consider a decision model where the objective is given by expected linear costs and the constraints consist of three groups: the measurability constraints on x_t, a linear constraint describing the relation between decisions at different stages, and constraints characterizing feasibility of the t-th stage decision x_t. The latter constraints consist of a linear inequality constraint and of the general constraint $x_t \in X_t$ where the (fixed) set X_t has the property that its convex hull $\mathrm{conv}(X_t)$ is polyhedral, allowing for mixed-integer decisions in all stages. Furthermore, the data ξ_t at stage t may enter all corresponding cost coefficients, matrices and right-hand sides. This leads to the following stochastic decision model:

$$\min\{E[\sum_{t=1}^{T} c_t\ (\xi_t)x_t] \ : \ x_t \text{ is measurable with respect to } \mathcal{F}_t, \tag{1}$$

$$x_t \in X_t\,,\ B_t(\xi_t)x_t \geq d_t(\xi_t)\,,\ P-a.s.\,,\ t = 1, \ldots, T, \tag{2}$$

$$\sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau \geq g_t(\xi_t)\,,\ P-a.s.\,,\ t = 2, \ldots, T\} \tag{3}$$

Throughout, the following is imposed: The sets X_t are nonempty and closed. The matrices $A_{t\tau}(\cdot)$, $B_t(\cdot)$ as well as the coefficients $c_t(\cdot)$ and the right-hand sides $d_t(\cdot)$, $g_t(\cdot)$ all depend affinely linearly on the corresponding component of ξ, for each $\tau = 1, \ldots, t$, $t = 1, \ldots, T$. In order to have the model (1)–(3) well defined, we need that the scalar products $c_t(\xi_t)x_t$ are integrable. The latter property is implied by the integrability of $\|\xi_t\|\|x_t\|$ and by the conditions $\xi_t \in L_{q_t}(\Omega, \mathcal{F}_t, P; \mathbb{R}^{s_t})$ and $x_t \in L_{r_t}(\Omega, \mathcal{F}_t, P; \mathbb{R}^{m_t})$ where $q_t, r_t \in [1, \infty]$ with $\frac{1}{q_t} + \frac{1}{r_t} = 1$. Since it is desirable to impose only weak conditions on the data process ξ and since we assume later on that the set X_t is bounded, we may restrict our attention to decisions $x_t \in L_\infty(\Omega, \mathcal{F}_t, P; \mathbb{R}^{m_t})$ and to the first order moment condition $\xi_t \in L_1(\Omega, \mathcal{F}_t, P; \mathbb{R}^{s_t})$ on the data at stage t for each $t = 1, \ldots, T$. Then the nonanticipativity constraint (1) may be expressed equivalently as

$$x_t \in L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{m_t}) \quad \text{and} \quad x_t = E[x_t|\mathcal{F}_t]\,,\ t = 1, \ldots, T, \tag{4}$$

by using the conditional expectation $E[\cdot|\mathcal{F}_t]$ with respect to the σ-algebra $\mathcal{F}_t$. Condition (4) describes a linear subspace $\mathcal{N}_{na}$ of the space $\times_{t=1}^T L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{m_t})$. This combination of functional and (P-a.s.) pointwise constraints in our model, i.e., the functional condition $x \in \mathcal{N}_{na}$ and the P-a.s. constraints (2) and (3), forms the theoretical and algorithmic challenge of multistage stochastic programs. A special role is played by the two-stage case (i.e., T=2) where $\mathcal{N}_{na}$ takes the specific form $\mathcal{N}_{na} = \mathbb{R}^{m_1} \times L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{m_2})$. An additional complication of the model (1)–(3) is caused by the mixed-integer constraints hidden in the condition $x_t \in X_t$, $t = 1, \ldots, T$.

2.2 Multistage Models, Dynamic Programming and Optimality

We adopt the setting of the previous section and assume that X_t is compact and $\xi_t \in L_1(\Omega, \mathcal{F}, P; \mathbb{R}^{s_t})$ for $t = 1, \ldots, T$. For each $\omega \in \Omega$ we define the subset $\mathcal{Y}(\omega)$ of $\mathcal{X} := \times_{t=1}^T \mathbb{R}^{m_t}$ by

$$\begin{aligned}\mathcal{Y}(\omega) := \{y \in \mathcal{X} : y_t \in X_t,\ B_t(\xi_t(\omega))y_t \geq d_t(\xi_t(\omega)),\ t = 1, \ldots, T,\\ \sum_{\tau=1}^{t} A_{t\tau}(\xi_t(\omega))y_\tau \geq g_t(\xi_t(\omega)),\ t = 2, \ldots, T\}\end{aligned} \tag{5}$$

and the extended real-valued function φ

$$\varphi(y_1, \ldots, y_T, \omega) := \begin{cases} \sum\limits_{t=1}^{T} c_t(\xi_t(\omega))y_t\ , & (y_1, \ldots, y_T) \in \mathcal{Y}(\omega), \\ +\infty & , \text{otherwise} \end{cases} \tag{6}$$

from $\mathcal{X} \times \Omega$ to $(-\infty, +\infty]$. With these notations, the model (1)–(3) is equivalent to the optimization problem

$$\min\{E[\varphi(x_1, \ldots, x_T, \omega)] : x_t \text{ is measurable w.r.t. } \mathcal{F}_t\,,\ t = 1, \ldots, T\}. \tag{7}$$

The real-valued function $(y, \omega) \mapsto \sum\limits_{t=1}^{T} c_t(\xi_t(\omega))y_t$ 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 [40]). Hence, the function φ is $\mathcal{B}(\mathcal{X}) \bigotimes \mathcal{F}$-measurable (cf. Example 14.32 in [40]). Furthermore, the following estimate is valid for each $y \in \times_{t=1}^{T} X_t$ and $\omega \in \Omega$:

$$|\varphi(y_1, \dots, y_T, \omega)| \leq \sum_{t=1}^{T} \|c_t(\xi_t(\omega))\| \sup_{y_t \in X_t} \|y_t\| \tag{8}$$

Hence, $E[\varphi(x_1, \dots, x_T, \omega)]$ 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 [16], we construct recursively two sequences of functions by putting $\psi_{T+1} := \varphi$ and

$$\varphi_t(y_1, \dots, y_t, \omega) := E^r[\psi_{t+1}(y_1, \dots, y_t, \cdot)|\mathcal{F}_t](\omega)\,, \tag{9}$$

$$\psi_t(y_1, \dots, y_{t-1}, \omega) := \inf_{y} \varphi_t(y_1, \dots, y_{t-1}, y, \omega)\,, \tag{10}$$

for $t = T, \dots, 1$, and for each $\omega \in \Omega$ and $y_\tau \in X_\tau$, $\tau = 1, \dots, T$. Here, $E^r[\cdot|\mathcal{F}_t]$ denotes the regular conditional expectation with respect to $\mathcal{F}_t$. We recall that the *regular conditional expectation* is a version of the conditional expectation (i.e., $E^r[\cdot|\mathcal{F}_t] = E[\cdot|\mathcal{F}_t]$, P-a.s.) having the property that the mapping $(z, \omega) \mapsto \Phi(z, \omega) := E^r[\Psi(z, \cdot)|\mathcal{F}_t](\omega)$ from $Z_t \times \Omega$ to $(\infty, +\infty]$ is $\mathcal{B}(\mathcal{Z}_t) \bigotimes \mathcal{F}_t$-measurable if Ψ is $\mathcal{B}(\mathcal{Z}_t) \bigotimes \mathcal{F}$-measurable. Here, Z_t denotes a closed subset of a Euclidean space. The regular conditional expectation exists if Ψ is $\mathcal{B}(\mathcal{Z}) \bigotimes \mathcal{F}$-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 [15]). Due to condition (8), relation (9) is well defined for $t = T$ and leads to a $\mathcal{B}(\mathcal{Z}) \bigotimes \mathcal{F}_T$-measurable function ϕ_T, where $Z := \times_{t=1}^{T} X_t$. It is shown in [16] that the relations (9) and (10) are well defined for all $t = T, \dots, 1$. Furthermore, the following optimality criterion and existence result for (7) or, equivalently, for (1)–(3) are valid.

Theorem 1. *Let the general assumptions be satisfied and assume that there exists a feasible solution of (1)–(3). Then $\{\bar{x}_t\}_{t=1}^{T}$ is a solution of (1)–(3) iff*

$$\varphi_t(\bar{x}^t(\omega), \omega) = \psi_t(\bar{x}^{t-1}(\omega), \omega)\,,\ P-a.s.\,,\ t = 1, \dots, T. \tag{11}$$

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

$$\min\{\varphi_1(x_1) = E[\psi_2(x_1, \omega)] \,:\, x_1 \in X_1\,,\ B_1(\xi_1)x_1 \geq d_1(\xi_1)\}, \tag{12}$$

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

The theorem is a special case of the more general results (Theorems 1 and 2) in [16]. Theorem 1 implies the existence of a solution to (1)–(3) and justifies the solution approach (11) which is usually called *dynamic programming approach.* Due to measurable selection arguments (cf. Chapter 14 in [40]), a feasible solution of (1)–(3) exists if the model (1)–(3) has *relatively complete recourse*, i.e., if $\mathcal{Y}(\omega) \neq \emptyset$ P-a.s.

2.3 Structure and Stability

We adopt the setting of the previous sections, denote by $\mathcal{P}(\Xi)$ the set of all Borel probability measures on some closed subset Ξ of $\mathbb{R}^s$ with $s = \sum_{t=1}^{T} s_t$, which is chosen such that it contains the support of ξ. By $\mu \in \mathcal{P}(\Xi)$ we denote the probability distribution of ξ. We consider the probability space $(\Xi, \mathcal{B}(\Xi), \mu)$ as the underlying probability space $(\Omega, \mathcal{F}, P)$ in Section 1, and define a function f from $X_1 \times \Xi$ to the extended real numbers $\overline{\mathbb{R}}$ by $f(x_1, \xi) := \psi_2(x_1, \xi)$, where ψ_2 is defined by (9) and (10). Then the first-stage optimization problem (12) can be rewritten in the following form:

$$\min\{\int_{\Xi} f(x_1,\xi)\mu(d\xi) \ : \ x_1 \in X_1 \ , \ B_1(\xi_1)x_1 \geq d_1(\xi_1)\} \tag{13}$$

The techniques exploited in [16] and used in the previous section imply that the integrand f is $\mathcal{B}(X_1) \bigotimes \mathcal{B}(\Xi)$-measurable. The recursions (9) and (10) together with the Fatou Lemma for (conditional) expectations as well as lower semicontinuity properties of infima in parametric minimization (e.g. Theorem 1.17 of [40]) imply lower semicontinuity of f with respect to x_1 and of the objective function $x_1 \mapsto \int_{\Xi} f(x_1, \xi)\mu(d\xi)$. If the multistage model (1)–(3) has relatively complete recourse, it holds that $|f(x_1, \xi)| \leq K(1+\max_{t=1,\ldots,T} \|\xi_t\|)$ for each feasible x_1, each $\xi \in \Xi$ and some constant $K > 0$. Hence, the integrand has a uniform and integrable upper bound, and the objective function is finite at all feasible $\bar{x}_1$. By Lebesgue's theorem, the objective function is continuous at some feasible $\bar{x}_1$ if $\mu(\{\xi \in \Xi \ : \ f(\cdot, \xi) \text{ is not continuous at } \bar{x}_1\}) = 0$. Such discontinuity sets of the integrand f have been studied in [43] for the two-stage situation with fixed recourse matrix A_{22} and recourse costs c_2.

When developing approximation schemes and algorithmic approaches for solving the model (1)–(3), the behaviour of its optimal value $\mathrm{val}(\mu)$ and of the set $\mathrm{Sol}(\mu)$ of first-stage solutions to (1)–(3) is important when perturbing or approximating the underlying distribution μ. We say that the model (1)–(3) is *stable* if $\mathrm{val}(\mu)$ and $\mathrm{Sol}(\mu)$ satisfy certain continuity properties with respect to some suitable convergence of probability measures. Here, we follow the presentation in [35] and consider the following distance

$$d_f(\mu,\nu) := \sup\{|\int_{\Xi} f(x_1,\xi)(\mu-\nu)(d\xi)| \ : \ x_1 \text{ is feasible}\} \tag{14}$$

of probability measures μ and ν belonging to the set $\mathcal{P}_1(\Xi) := \{\nu \in \mathcal{P}(\Xi) : \int_{\Xi} \|\xi\|\nu(d\xi) < \infty\}$. Then it holds for any perturbation ν of the original underlying probability distribution μ that

$$|\mathrm{val}(\mu) - \mathrm{val}(\nu)| \leq d_f(\mu,\nu) \tag{15}$$

$$\emptyset \neq \mathrm{Sol}(\nu) \subseteq \mathrm{Sol}(\mu) + \Psi(d_f(\mu,\nu))B_{m_1} \tag{16}$$

where B_{m_1} denotes the closed unit ball in $\mathbb{R}^{m_1}$ and Ψ is some monotonically increasing function on $\mathbb{R}_+$ with $\Psi(0) = 0$, which is related to the growth behaviour of the objective function near the set $\mathrm{Sol}(\mu)$ (see [35]). While (15) represents a Lipschitz type estimate for the optimal value at μ, the relation (16) says that the sets of first-stage solutions behave upper semicontinuously at μ with respect to d_f. In general, the distance d_f is rather involved and difficult to handle. Hence, it is of considerable interest to derive estimates of d_f in terms of simpler probability metrics and to expose relations to the classical concept of weak convergence of probability measures. For two-stage models with fixed recourse matrix and costs, such results are obtained in [44] and [35]. We also refer to relevant stability studies in [2,18,49]. Altogether, such stability results justify the approximation of the underlying distribution μ by simpler measures and provide techniques for designing approximation schemes. Next, we show that approximations by discrete measures having finitely many atoms or *scenarios* play a prominent role since they lead to specially structured large-scale mixed-integer linear programs.

2.4 Scenario Based Models

We assume throughout this section that Ω is finite, i.e., $\Omega = \{\omega_s\}_{s=1}^S$, $\mathcal{F}$ is the power set of Ω and $P(\{\omega_s\}) = p_s$, $s = 1, \ldots, S$. We denote by $\xi_t^s := \xi_t(\omega_s)$ the value of the data scenario s at stage t and by x_t^s the value of the decision scenario s at t for $s = 1, \ldots, S$, $t = 1, \ldots, T$. Since Ω is finite, there exists a finite subset $\mathcal{E}_t$ of the σ-algebra $\mathcal{F}_t$, for each $t = 1, \ldots, T$, such that $\mathcal{E}_t$ is a partition of Ω and that the smallest σ-algebra containing $\mathcal{E}_t$ is just $\mathcal{F}_t$. Then the conditional expectation w.r.t. $\mathcal{F}_t$ in the nonanticipativity condition (4) takes the form

$$\begin{aligned} E[x_t|\mathcal{F}_t] &= \sum_{C\in\mathcal{E}_t} \frac{1}{P(C)} \int_C x_t(\omega)P(d\omega)\chi_C \\ &= \sum_{C\in\mathcal{E}_t} (\sum_{\substack{s=1\\ \omega_s\in C}}^S p_s)^{-1} (\sum_{\substack{s=1\\ \omega_s\in C}}^S p_s x_t^s)\chi_C \end{aligned} \tag{17}$$

where χ_C denotes the characteristic function of the set $C \in \mathcal{E}_t$. Hence, the nonanticipativity condition (4) is equivalent to the following equality constraints

$$x_t^\sigma = \sum_{\substack{C\in\mathcal{E}_t\\ \omega_\sigma\in C}} (\sum_{\substack{s=1\\ \omega_s\in C}}^S p_s)^{-1} \sum_{\substack{s=1\\ \omega_s\in C}}^S p_s x_t^s\,,\ \sigma = 1, \ldots, S,\ t = 1, \ldots, T. \tag{18}$$

Clearly, for $t = 1$ we have $\mathcal{E}_1 = \{\Omega\}$ and, hence, condition (18) is equivalent to the equations $x_1^\sigma = \sum_{s=1}^S p_s x_1^s$, $\sigma = 1, \ldots, S$, i.e, to $x_1^1 = \ldots = x_1^S$.
Hence, the multistage stochastic program (1)–(3) takes the following form

which will be called its scenario formulation:

$$\min \{ \sum_{s=1}^{S}\sum_{t=1}^{T} p_s c_t(\xi_t^s) x_t^s \; : \; x \text{ satisfies the constraints (18)}, \tag{19}$$
$$x_t^s \in X_t \, , \; B_t(\xi_t^s) x_t^s \geq d_t(\xi_t^s) \, , \; s = 1,\dots,S, \; t = 1,\dots,T,$$
$$\sum_{\tau=1}^{t} A_{t\tau}(\xi_t^s) x_\tau^s \geq g_t(\xi_t^s) \, , \; s = 1,\dots,S, \, t = 2,\dots,T\}$$

Since $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, every element of $\mathcal{E}_{t+1}$ can be represented as the union of certain elements of $\mathcal{E}_t$. Furthermore, formula (17) shows that the number of elements in $\mathcal{E}_t$ coincides with the number of realizations of ξ and x at period t, respectively. Hence, representing the relations between the elements of $\mathcal{E}_t$ and $\mathcal{E}_{t+1}$ for $t = 1,\dots,T-1$, leads to a tree having the same structure as the sets of scenarios of ξ and x, respectively. Therefore, such a tree is called *scenario tree*. It is based on a finite set $\mathcal{N} \subseteq \mathbb{N}$ of nodes. Fig. 1 shows an example of a scenario tree where the t_k denote the branching points of the tree. The root node $n = 1$ stands for period $t = 1$. Every other node n has

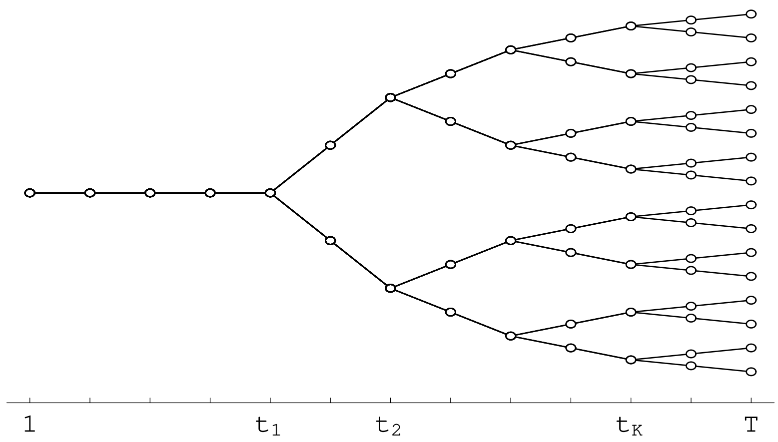

Fig. 1. Example of a (binary) scenario tree

a unique predecessor node n_- and a transition probability $\pi_{n/n_-} > 0$, which is the probability of n being the successor of n_-. The probability π_n of each node n is given recursively by $\pi_1 = 1 \, , \; \pi_n = \pi_{n/n_-} \pi_{n_-} \, , \; n > 1$. We denote by $\mathcal{N}_+(n)$ the set of successors to node n, by path(n) the path from the root to node n and by $t(n)$ its length, i.e., $t(n) := \text{card}(\text{path}(n))$. $\mathcal{N}_t$ denotes the set $\{n \in \mathcal{N} \, : \, t(n) = t\}$, and it holds $\sum_{n \in \mathcal{N}_t} \pi_n = 1$ for each period t. Nodes n with $\mathcal{N}_+(n) = \emptyset$ are called leaves; they constitute the terminal set $\mathcal{N}_T$. A scenario corresponds to a path from the root node to a leaf. Clearly, it holds that $\mathrm{card}(\mathcal{N}_T) = S$ and $\{\pi_n\}_{n\in\mathcal{N}_T} = \{p_s\}_{s=1}^S$. Conversely, given these scenario probabilities, the remaining node and transition probabilities are generated recursively by $\pi_n := \sum_{n_+\in\mathcal{N}_+(n)} \pi_{n_+}$, $\pi_{n_+/n} := \pi_{n_+}/\pi_n$ for $n_+ \in \mathcal{N}_+(n)$. We use the following notation for the sequence of predecessors of any node $n \in \mathcal{N}$: $n_0 := n$, $n_{-1} := n_-$ if $n > 1$, $n_{-(\kappa+1)} := (n_{-\kappa})_-$ if $t(\kappa) > 1$. Note that $t(n_{-\kappa}) = t(n) - \kappa$ for $\kappa = 1,\dots,t(n)-1$. Furthermore, we denote by $\{\xi^n\}_{n\in\mathcal{N}_t}$ the realizations of ξ_t and by $\{x^n\}_{n\in\mathcal{N}_t}$ the realizations of x_t. After these preparations the scenario tree formulation of the multistage stochastic program reads:

$$\begin{aligned} \min \{ \sum_{n\in\mathcal{N}} \pi_n c_{t(n)}(\xi^n)x^n \ : \ x^n \in X_{t(n)} \, , \, B_{t(n)}(\xi^n)x^n \geq d_{t(n)}(\xi^n), \quad (20) \\ \sum_{\kappa=0}^{t(n)-1} A_{t(n),t(n)-\kappa}(\xi^n)x^{n_{-\kappa}} \geq g_{t(n)}(\xi^n) \, , \, n \in \mathcal{N}\} \end{aligned}$$

Both formulations of the multistage stochastic program will be used for the description of decomposition approaches. We recall that the nonanticipativity condition appears explicitly in the scenario formulation (19), but disappears in the scenario tree formulation (20) because it is incorporated into the tree construction. Since it holds that $|\mathcal{N}| := \mathrm{card}(\mathcal{N}) << TS$, the dimensions of both model formulations are quite different. More precisely, the model (19) contains $(\sum_{t=1}^T m_t)S$ decision variables and $\sum_{t=1}^T (m_t + k_t + r_t)S$ linear constraints, whereas the model (20) contains $\sum_{n\in\mathcal{N}} m_{t(n)}$ decisions and $\sum_{n\in\mathcal{N}}(k_{t(n)} + r_{t(n)})$ linear constraints. Here, k_t and r_t denote the dimensions of $d_t(\cdot)$ and $g_t(\cdot)$, respectively, for $t = 1,\dots,T$.

2.5 Dualization and the Convex Case

We assume $\xi \in \times_{t=1}^T L_\infty(\Omega,\mathcal{F},P;\mathbb{R}^{s_t})$ and consider the multistage stochastic integer program of Section 1 as an abstract (infinite) optimization problem in the Banach space $\times_{t=1}^T L_\infty(\Omega,\mathcal{F},P;\mathbb{R}^{m_t})$, i.e., in the form

$$\min\{E\,[\sum_{t=1}^T c_t(\xi_t)x_t] \ : \ x \in \times_{t=1}^T L_\infty(\Omega,\mathcal{F},P;\mathbb{R}^{m_t}) \, , \, x \in \mathcal{N}_{na} \, , \quad (21)$$

$$x_t \in X_t \, , \, B_t(\xi_t)x_t \geq d_t(\xi_t) \, , \, P-a.s. \, , \, t = 1,\dots,T, \quad (22)$$

$$\sum_{\tau=1}^t A_{t\tau}(\xi_t)x_\tau \geq g_t(\xi_t) \, , \, P-a.s. \, , \, t = 2,\dots,T\} \, . \quad (23)$$

Let $F(\cdot)$ denote the objective function, i.e., $F(x) := E[\sum_{t=1}^T c_t(\xi_t)x_t]$.
Our aim is to introduce a Lagrangian associated with the essential groups of constraints of problem (21)–(23), namely, the (functional) nonanticipativity constraint $x \in \mathcal{N}_{na}$, the k_t coupling constraints $B_t(\xi_t)x_t \geq d_t(\xi_t)$ and r_t dynamic constraints (23). We make use of the concepts and results of

[38] and introduce the following sets $\Lambda_1 := \{\lambda_1 \in \times_{t=1}^T L_1(\Omega,\mathcal{F},P;\mathbb{R}^{m_t}) : E[\lambda_{1t}|\mathcal{F}_t] = 0\,,\, P-\text{a.s.}\,,\, t=1,\ldots,T\}$, $\Lambda_2 := \{\lambda_2 \in \times_{t=1}^T L_1(\Omega,\mathcal{F},P;\mathbb{R}^{k_t}) : \lambda_2 \geq 0\,,\, P-\text{a.s.}\}$ and $\Lambda_3 := \{\lambda_3 \in \times_{t=1}^T L_1(\Omega,\mathcal{F},P;\mathbb{R}^{r_t}) : \lambda_3 \geq 0\,,\, P-\text{a.s.}\}$ of Lagrange multipliers. The sets Λ_2 and Λ_3 are convex cones and Λ_1 is a linear space which is complementary to the nonanticipativity subspace $\mathcal{N}_{na}$ with respect to the dual pairing $\langle\cdot,\cdot\rangle$ of L_1 and L_∞, i.e., it holds $\langle\lambda_1, x\rangle := E[\sum_{t=1}^T \lambda_{1t}x_t] = 0$ for all $\lambda_1 \in \Lambda_1$ and $x \in \mathcal{N}_{na}$.
The Lagrangian is defined to be the function

$$\begin{aligned} L(x,\lambda) := E[\sum_{t=1}^{T} \{\, c_t(\xi_t)x_t - \lambda_{1t}x_t + \lambda_{2t}(d_t(\xi_t) - B_t(\xi_t)x_t)\} \qquad (24)\\ + \sum_{t=2}^{T} \lambda_{3t}(g_t(\xi_t) - \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau)] \end{aligned}$$

from $\times_{t=1}^T L_\infty(\Omega,\mathcal{F},P;\mathbb{R}^{m_t}) \times \Lambda$ to $\mathbb{R}$, where $\Lambda := \times_{i=1}^3 \Lambda_i$. The dual function D from Λ to $\mathbb{R}$ is defined by

$$\begin{aligned} D(\lambda) := \inf\{L(x,\lambda) : x \in \times_{t=1}^T L_\infty(\Omega,\mathcal{F},P;R^{m_t})\,, \qquad (25)\\ x_t \in X_t\,,\, P-\text{a.s.},\, t=1,\ldots,T\}, \end{aligned}$$

and the *dual problem* associated with (21)–(23) is

$$\max\{D(\lambda) \,:\, \lambda \in \Lambda\}. \qquad (26)$$

We assume again that the sets $X_t\,,\, t = 1,\ldots,T$, are compact. Then the Lagrangian L and the dual function D are well-defined, D is concave and the *weak duality* estimate

$$D(\lambda) \leq F(x) \quad \text{for all } \lambda \in \Lambda \text{ and all } x \text{ satisfying (21)–(23).} \qquad (27)$$

is valid. In the following, we say that the model (21)–(23) is *strictly feasible* if there exist $\tilde{x} \in \mathcal{N}_{na}$ and $\varepsilon > 0$ such that

$$\begin{aligned} \tilde{x}_t + \varepsilon B_{m_t} \subseteq \text{conv}(X_t)\,,\, B_t(\xi_t)\tilde{x}_t \geq d_t(\xi_t) + \varepsilon\,,\, P - a.s.\,,\, t = 1,\ldots,T,\\ \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)\tilde{x}_\tau \geq g_t(\xi_t) + \varepsilon\,,\, P - a.s.\,,\, t = 2,\ldots,T, \end{aligned}$$

where B_m denotes the closed unit ball in $\mathbb{R}^m$. Then we conclude from Theorem 1 and from Theorem 3 of [39] that the following holds.

Theorem 2. *Assume that the general assumptions are satisfied, that the sets $X_t\,,\, t = 1,\ldots,T$, are convex compact and that the model (21)–(23) has relatively complete recourse and is strictly feasible. Then there exist optimal solutions $\bar{\lambda}$ to (26) and $\bar{x}$ to (21)–(23), and it holds $D(\bar{\lambda}) = F(\bar{x})$.*

Since the sets X_t , $t = 1, \ldots, T$, fail to be convex, such a duality result is not available in our setting and, due to (27), we are faced with a *duality gap*

$$DG := F(\bar{x}) - \sup_{\lambda \in \Lambda} D(\lambda) \geq 0 \,. \tag{28}$$

This inequality is strict, in general. On the other hand, in case of a discrete underlying probability distribution, the theory of Lagrangian relaxation in mixed-integer linear programming (cf. e.g. Chapter II.3.6 of [31]) implies that the optimal value of (26) is greater than or equal to the optimal value of the linear programming relaxation to (19) or (20). In other words, the lower bound obtained by dualizing constraints is never worse the bound obtained by relaxing the integer requirements.

So far we have associated Lagrange multipliers with nonanticipativity, coupling as well as dynamic constraints. Of course, it is also of interest to consider restricted Lagrangians and restricted duals by associating multipliers with one or with two of these three groups of constraints, only. For such restricted dualization schemes, duality results for the convex case that are similar to Theorem 2 may be derived as well (see [37] for dualizing the nonanticipativity constraints and [38] for other inequality constraints). It is worth recalling that the duality gap increases when dualizing additional constraints (see Section 3.1 in [29]). Since small duality gaps are of algorithmic interest, we take a closer look at dualization schemes where either nonanticipativity or coupling or dynamic constraints are associated with Lagrange multipliers. We denote the corresponding dual functions from Λ_i to $\mathbb{R}$ by D_i for $i = 1, 2, 3$ and start with dualizing nonanticipativity constraints, i.e.,

$$\begin{aligned} D_1(\lambda_1) := \inf\{E[\sum_{t=1}^{T}(c_t(\xi_t)x_t - \lambda_{1t}x_t)] \,:\, & x \in \times_{t=1}^{T} L_\infty(\Omega, \mathcal{F}, P; R^{m_t}) \,, \\ & x_t \in X_t \,,\, B_t(\xi_t)x_t \geq d_t(\xi_t) \,,\, P-\text{a.s.}, \, t = 1, \ldots, T \,, \\ & \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau \geq g_t(\xi_t) \,,\, P-\text{a.s.}\,,\, t = 2, \ldots, T\} \\ = E[\inf\{\sum_{t=1}^{T}(c_t(\xi_t)x_t - \lambda_{1t}x_t) \,:\, & x_t \in X_t \,,\, B_t(\xi_t)x_t \geq d_t(\xi_t) \,, \\ & t = 1, \ldots, T \,,\, \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau \geq g_t(\xi_t) \,,\, t = 2, \ldots, T\}], \end{aligned}$$

where the infimum and expectation may be interchanged since the minimization problem only contains P-a.s. pointwise constraints (see e.g. Theorem 14.60 of [40]). Hence, the multistage stochastic program defining D_1 decomposes into pathwise minimization problems. This effect becomes more transparent if the underlying probability distribution of ξ is discrete, i.e., if $\Omega = \{\omega_1, \ldots, \omega_S\}$. Adopting the notation of Section 2.4, the dual function takes the form

$$\begin{aligned} D_1(\lambda_1) := \sum_{s=1}^{S} p_s \inf\{\sum_{t=1}^{T}[c_t(\xi_t^s)x_t^s - \lambda_{1t}^s x_t^s] \,:\, & x_t^s \in X_t\,, \\ & B_t(\xi_t^s)x_t^s \geq d_t(\xi_t^s)\,,\, t=1,\ldots,T, \\ & \sum_{\tau=1}^{t} A_{t\tau}(\xi_t^s)x_\tau^s \geq g_t(\xi_t^s)\,,\, t=2,\ldots,T\}, \end{aligned} \tag{29}$$

where $\lambda_1 \in \Lambda_1$ has the scenarios $\{\lambda_{1t}^s\}_{t=1}^T$ with probabilities p_s for $s=1,\ldots,S$ and Λ_1 is given by the linear subspace

$$\begin{aligned} \Lambda_1 &= \{\lambda_1 : E[\lambda_{1t}|\mathcal{F}_t] = \sum_{C\in\mathcal{E}_t} (\sum_{\substack{s=1\\ \omega_s\in C}}^{S} p_s)^{-1} \sum_{\substack{s=1\\ \omega_s\in C}}^{S} p_s\lambda_{1t}^s\chi_C = 0\,,\, t=1,\ldots,T\} \\ &= \{\lambda_1 : \sum_{\substack{s=1\\ \omega_s\in C}}^{S} p_s\lambda_{1t}^s = 0\,,\, C\in\mathcal{E}_t\,,\, t=1,\ldots,T\} \end{aligned} \tag{30}$$

of the Euclidean space of dimension $(\sum_{t=1}^T k_t + \sum_{t=2}^T r_t)S$. Since $\mathcal{E}_1 = \{\Omega\}$ and $\mathcal{E}_T = \{\{\omega_1\},\ldots,\{\omega_S\}\}$, the conditions for $t=1$ and $t=T$ in (30) are equivalent to $\sum_{s=1}^S p_s\lambda_{11}^s = 0$ and $\lambda_{1T}^s = 0\,,\, s=1,\ldots,S$, respectively. We note that the constraint $\lambda_1 \in \Lambda_1$ means that each $\lambda_1 \neq 0$ is *anticipative*, i.e., λ_{1t} is not $\mathcal{F}_t$-measurable for some t (see also the example in [22]). Sometimes, one might find it more convenient that the dual function is defined and maximized on the whole space, i.e, without regard to the subspace constraint $\lambda_1 \in \Lambda_1$. This can be done be replacing λ_{1t} in the right-hand side of (29) by $\lambda_{1t} - E[\lambda_{1t}|\mathcal{F}_t]$ for $t=1,\ldots,T$. Then the subspace constraint for the multiplier is automatically satisfied and the dual maximization problem is unconstrained.

Next we consider dualizations of certain inequality constraints by some multiplier, but leave the nonanticipativity constraint untouched. In contrast to the anticipativity of multipliers in the previous case, the multipliers may now be chosen nonanticipative, i.e., as elements of $\times_{t=1}^T L_1(\Omega,\mathcal{F}_t,P)$. This is due to the linear separability properties of (21)–(23) (Theorem 7 of [38]). In particular, when dualizing the coupling constraints, the restricted dual function

$$\begin{aligned} D_2(\lambda_2) := \inf\{E[\sum_{t=1}^{T}(c_t(\xi_t)x_t + \lambda_{2t}(d_t(\xi_t) - B_t(\xi_t)x_t))] \,:\, x \in \mathcal{N}_{na}\,, \\ x_t \in X_t\,,\, \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau \geq g_t(\xi_t)\,,\, P-a.s.\,,\, t=2,\ldots,T\}, \end{aligned} \tag{31}$$

has to be maximized on the convex cone $\Lambda_2 := \{\lambda_2 \in \times_{t=1}^T L_1(\Omega,\mathcal{F}_t,P;\mathbb{R}^{k_t}) :$ $\lambda_{2t} \geq 0\,,\, P-\text{a.s.},\, t=1,\ldots,T\}$. Dualizing the dynamic constraints leads to maximizing the restricted dual

$$D_3(\lambda_3) := \inf\{E[\sum_{t=1}^{T} c_t(\xi_t)x_t + \sum_{t=2}^{T} \lambda_{3t}(g_t(\xi_t) - \sum_{\tau=1}^{t} A_{t\tau}(\xi_t)x_\tau)] : \qquad (32)$$
$$x \in \mathcal{N}_{na}\,,\, x_t \in X_t\,,\, B_t(\xi_t)x_t \geq d_t(\xi_t)\,,\, P-a.s.\,,\, t = 1,\ldots,T\}.$$

subject to the convex cone $\Lambda_3 := \{\lambda_3 \in \times_{t=1}^{T} L_1(\Omega, \mathcal{F}_t, P; \mathbb{R}^{r_t}) : \lambda_{3t} \geq 0\,,\, P-\text{a.s.},\, t = 1,\ldots,T\}$. Clearly, both optimization problems on the right-hand sides of (31) and (32), respectively, are stochastic integer programs. While the program in (31) exhibits the typical multistage structure, the specific feature of the program in (32) is the lack of a dynamic constraint. In Sections 3.2 and 3.3 we gain further information on these programs in case of a discrete underlying probability distribution, i.e., when the data, decisions and multipliers form scenario trees.

3 Decomposition Methods

Due to the enormous size of scenario based models in multi-stage stochastic programming, decomposition is the method of choice when it comes to numerical solution. This is further enhanced by special structures met, both in the scenario formulation (19) and in the scenario tree formulation (20) of multi-stage stochastic programs. If integer requirements are missing in (21) - (23), powerful convexity and duality results (cf. Theorem 2) are the basis of efficient decomposition methods. These methods can be subdivided into primal and dual ones.

Primal decomposition methods employ the scenario tree formulation (20). Starting from the root node, primal proposals are passed down the tree where they are used to compute so called feasibility and optimality cuts that are passed upward to be included into convex optimization problems whose solutions lead to updated primal proposals that are again passed down the tree, and so on. This procedure (nested decomposition) is enhanced by regularization and cut deletion. Its mathematical backbone is convexity, in particular ideas from the area of bundle-trust and proximal point methods.

Dual decomposition circles around duality results such as Theorem 2. The approaches discussed in Section 2.5 then all benefit from a zero duality gap. Particular attention has been paid to dualizing nonanticipativity in the framework of augmented Lagrangians and related proximal point algorithms (progressive hedging, cf. [39]). The survey papers [5,42] provide further insights into both primal and dual decomposition of multi-stage stochastic linear programs.

With integer requirements in (21) - (23) the mentioned powerful convexity and duality results are lost. Approaches to decomposition, that have proven efficient for purely linear models, have to be rethought from their very beginnings.

The impact of integrality on *primal decomposition* is twofold: Feasibility and optimality cuts can no longer be obtained as linear functionals but as merely subadditive functionals instead. Primal proposals can no longer be obtained via convex programs but via merely lower semicontinuous (discontinuous) nonconvex programs instead. For algorithmic realization this leads to obstacles impossible to overcome with existing methods, [8,11]. Two-stage models have been tackled with limited success by solving the mentioned lower semicontinuous programs via enumeration [45] or branch-and-bound [1] and exploiting problem similarities in the second stage.
The impact of integrality on *dual decomposition* has already been mentioned in Section 2.5: Theorem 2 is no longer valid, and we face a non-zero duality gap (28). Although progressive hedging then is no longer formally justified, quite satisfactory results have been observed empirically for specific applications, [30,47].
In what follows, we will return to the dualization schemes introduced in Section 2.5 in case that the underlying probability distribution is discrete. We will discuss the solution of the corresponding dual maximization problems

$$\max\{D_i(\lambda_i) \,:\, \lambda_i \in \Lambda_i\} \quad (i = 1, 2, 3)$$

by subgradient type methods and examine the decoupling potential of the different dualizations. Under the conditions imposed in Section 2.5 the dual functions D_i are finite, concave and polyhedral. They have the form

$$D_i(\lambda_i) = \inf_x \{F(x) + \langle \lambda_i, G_i(x) \rangle\}\,, \tag{33}$$

where F is the objective function, G_i is some affine linear function from L_∞ to L_∞, and $\langle \cdot, \cdot \rangle$ denotes the dual pairing of L_1 and L_∞. Hence, $G_i(x_i(\lambda_i))$ is a subgradient of D_i if $x_i(\lambda_i)$ is a solution to the minimization problem (33) defining D_i. Furthermore, the solution sets of the dual problems are nonempty since their objectives are polyhedral and their suprema finite. Therefore, *subgradient bundle methods* may be used for solving the duals, [23,28,25]. Let us consider the *proximal bundle method* [17,23,25] in some more detail. Starting from an arbitrary point $\lambda_i^1 = \bar{\lambda}_i^1 \in \Lambda_i$, this method generates a sequence $\{\lambda_i^k\}_{k \in \mathbb{N}}$ in Λ_i converging to some solution of the dual problem, and trial points $\bar{\lambda}_i^k$ for evaluating the solutions $x_i^k = x_i(\bar{\lambda}_i^k)$ of (33), the subgradients $G_i(x_i^k)$ of D_i and its linearizations

$$D_i^k(\cdot) := D_i(\lambda_i^k) + \langle \cdot - \bar{\lambda}_i^k, G_i(x_i^k) \rangle \geq D_i(\cdot)\,.$$

Iteration k uses the polyhedral model $D_{ik}(\cdot) := \min_{l \in N^k} D_i^l(\cdot)$ with $k \in N^k \subset \{1, \ldots, k\}$ for finding the next trial point $\bar{\lambda}_i^{k+1}$ as a solution of the quadratic subproblem

$$\max\{D_{ik}(\lambda) - \frac{1}{2} u_k |\lambda - \lambda_i^k|^2 : \lambda \in \Lambda_i\}\,, \tag{34}$$

where the proximity weight $u_k > 0$ and the penalty term $|\cdot|^2 := \langle\cdot,\cdot\rangle$ should keep $\bar{\lambda}_i^{k+1}$ close to the prox-center λ_i^k. An ascent step to $\lambda_i^{k+1} = \bar{\lambda}_i^{k+1}$ occurs if $D_i(\bar{\lambda}_i^{k+1}) \geq D_i(\lambda_i^k) + \kappa\delta_k$, where $\kappa \in (0,1)$ is a fixed Armijo-like parameter and $\delta_k := D_{ik}(\bar{\lambda}_i^{k+1}) - D_i(\lambda_i^k) \geq 0$ is the predicted ascent (if $\delta_k = 0$ then λ_i^k is a solution and the method may stop). Otherwise, a null step $\lambda_i^{k+1} = \lambda_i^k$ improves the next model $D_{i,k+1}$ with the new linearization D_i^{k+1}. The choices of the weights u_k and of the index set N^{k+1} are dicussed in [17,25] (see also Section 3.4 of [19]). The quadratic subproblem (34) is essentially influenced by the dual pairing $\langle\cdot,\cdot\rangle$. The latter reads $\langle\lambda_i, y\rangle = \sum_{s=1}^S p_s \sum_{t=1}^T \lambda_{it}^s y_t^s$ and $\langle\lambda_i, y\rangle = \sum_{n\in\mathcal{N}} \pi_n \lambda_i^n y^n$ for the scenario and the node formulations, respectively.

3.1 Scenario Decomposition

Scenario decomposition rests on the dualization of nonanticipativity constraints if the probability distribution of ξ is discrete. This leads to the dual maximization problem

$$\max\{D_1(\lambda_1) \,:\, \lambda_1 \in \Lambda_1\} \tag{35}$$

where D_1 and Λ_1 are defined as in (29), (30) of Section 2.5. Since the computation of D_1 decomposes into solving pathwise minimization problems, function values and subgradients of D_1 are obtained by solving the single-scenario problems

$$\begin{aligned}\min\{\sum_{t=1}^T\{c_t(\xi_t^s)x_t^s - \lambda_{1t}^s x_t^s\} \;:\;\; & x_t^s \in X_t\,,\\ & B_t(\xi_t^s)x_t^s \geq d_t(\xi_t^s),\; t=1,\ldots,T,\\ & \sum_{\tau=1}^t A_{t\tau}(\xi_t^s)x_\tau^s \geq g_t(\xi_t^s),\; t=2,\ldots,T\}\end{aligned}$$

for all $s = 1,\ldots,S$.
Indeed, if $\bar{x}^s, s = 1,\ldots,S$, denote optimal solutions to these problems, then

$$D_1(\lambda_1) \;=\; \sum_{s=1}^S p_s(\sum_{t=1}^T\{c_t(\xi_t^s)\bar{x}_t^s - \lambda_{1t}^s\bar{x}_t^s\})\,,$$

and $G_1(\bar{x}) = \bar{x}$ is a subgradient of D_1 at λ_1, where $\bar{x}$ has the scenarios $\bar{x}^s$, $s = 1,\ldots,S$. Compared with the scenario formulation (19) of the multi-stage stochastic program (1)-(3), which is a mixed-integer linear program in dimension $S\cdot\sum_{t=1}^T m_t$, the above single-scenario problems are S mixed-integer linear programs each of dimension $\sum_{t=1}^T m_t$, only.
In view of (28), solving (35) provides a lower bound to the optimal value of the multi-stage stochastic integer program (19). If the single-scenario solutions

$\bar{x}_1^s, \ldots, \bar{x}_T^s$ for the optimal λ_1 in (35) fulfilled the nonanticipativity constraints then $\bar{x}$ would be optimal to (19). In general, however, one faces a non-zero duality gap (28). Therefore the lower bounding has to be accompanied by upper bounding procedures resting on the generation of "promising" feasible solutions. This can be accomplished by primal heuristics starting from the results of the dual optimization, i.e., from single-scenario solutions $\bar{x}_1^s, \ldots, \bar{x}_T^s$ corresponding to optimal or nearly optimal λ_1.

An algorithmic realization of scenario decomposition for the case $T = 2$, i.e., for two-stage stochastic integer programs, has been proposed in [8–10]. The nonanticipativity constraints then read $x_1^\sigma = \sum_{s=1}^{S} p_s x_1^s$, $\sigma = 1, \ldots, S$. In [8–10], the equivalent representation $x_1^1 = \ldots = x_1^S$ is employed, and the scenario formulation (19) is set up with (18) replaced by $x_1^1 = \ldots = x_1^S$. Then, the usual Lagrangian relaxation of mixed-integer linear programming is performed with respect to the constraints $x_1^1 = \ldots = x_1^S$. In particular, this leads to a non-probabilistic Lagrangian, in contrast to the probabilistic Lagrangian (24) introduced in Section 2.5. As a consequence, the Lagrangian dual of [8–10] is unconstrained and lives in dimension $(S-1) \cdot m_1$. In the setting of Section 2.5, cf. (30), we obtain a dual in dimension $S \cdot m_1$ constrained by $\sum_{s=1}^{S} p_s \lambda_{11}^s = 0$, i.e., essentially an unconstrained program in dimension $(S-1) \cdot m_1$ as well.

In [8–10], the scheme of lower and upper bounding outlined above is further enhanced by embedding into a branch-and-bound algorithm in the spirit of global optimization. As stated in (13), the stochastic program can be rewritten as a nonconvex global optimization problem. In the branching part of the algorithm, the feasible region of (13) is subdivided. On each member of the subdivision, the bounding part employs dualization of nonanticipativity for the lower and a primal heuristic for the upper bounds. For further details on scenario decomposition for two-stage stochastic integer programs we refer to [21].

Only little is known about algorithmic realizations of scenario decomposition for multi-stage stochastic integer programs with $T > 2$. First experiences on extending the approach of [8–10] will be reported in [4].

3.2 Component Decomposition

Dualization of component coupling constraints results in the dual maximization problem

$$\max\{D_2(\lambda_2) \,:\, \lambda_2 \in \Lambda_2\},$$

where D_2 and Λ_2 are defined in Section 2.5. We assume that the underlying probability distribution of the data process ξ is discrete and, hence, given in form of a scenario tree $\{\xi^n\}_{n \in \mathcal{N}}$, where $\mathcal{N}$ denotes the finite set of nodes. The notation of Section 2.4 is used, and we denote by $x = \{x^n\}_{n \in \mathcal{N}}$ the decision scenario tree and by $\lambda_2 = \{\lambda_2^n\}_{n \in \mathcal{N}}$ the multiplier scenario tree. Then the dual function (31) may be rewritten in the following form (see also (20)):

$$D_2(\lambda_2) := \inf\{\sum_{n\in\mathcal{N}} \pi_n\{c_{t(n)}(\xi^n)x^n + \lambda_2^n(d_{t(n)}(\xi^n) - B_{t(n)}(\xi^n)x^n)\} : \quad (36)$$
$$x^n \in X_{t(n)}, \sum_{\kappa=0}^{t(n)-1} A_{t(n),t(n)-\kappa}(\xi^n)x^{n-\kappa} \geq g_{t(n)}(\xi^n), n \in \mathcal{N}\}$$

where $\lambda_2 \in \Lambda_2 = \{\{\lambda_2^n\}_{n\in\mathcal{N}} : \lambda_2^n \geq 0\,, n \in \mathcal{N}\}$. In order to demonstrate the component decoupling potential hidden in D_2, we assume that X_t has the specific structure $X_t = \times_{i=1}^{m_t} X_{ti}$, where the X_{ti} are closed subsets of $\mathbb{R}$, that $m_t = m$, $k_t = k$ and $r_t = mr$ for $t = 1,\ldots,T$ and some $r \in \mathbb{N}$, and that the matrices $A_{t\tau}(\cdot)$ are block-diagonal with m blocks $a_{t\tau}^i(\cdot) \in \mathbb{R}^r$ for $i = 1,\ldots,m$. In particular, this condition means that the constraints in (36) are expressible as componentwise constraints. We denote by $c_t^i(\cdot)$ the i-th component of $c_t(\cdot)$, by $g_t^i(\cdot) \in \mathbb{R}^r$ the i-th component vector of $g_t(\cdot)$, and by $b_t^i(\cdot)$ the i-th column of the matrix $B_t(\cdot)$. With x_i^n denoting the i-th component of x^n, we obtain by exchanging summation w.r.t. n and i

$$\begin{aligned} D_2(\lambda_2) = \inf\{ &\sum_{n\in\mathcal{N}} \pi_n\{\sum_{i=1}^m [c_{t(n)}^i(\xi^n) - \lambda_2^n b_{t(n)}^i(\xi^n)]x_i^n + \lambda_2^n d_{t(n)}(\xi^n)\} : \\ &x_i^n \in X_{t(n)}^i\,, \sum_{\kappa=0}^{t(n)-1} a_{t(n),t(n)-\kappa}^i(\xi^n)x_i^{n-\kappa} \geq g_{t(n)}^i(\xi^n)\,, \\ &i = 1,\ldots,m\,,\, n \in \mathcal{N}\} \\ = &\sum_{i=1}^m D_{2i}(\lambda_2) + \sum_{n\in\mathcal{N}} \pi_n\lambda_2^n d_{t(n)}(\xi^n) \end{aligned}$$

where the functions D_{2i}, $i = 1,\ldots,m$, from Λ_2 to $\mathbb{R}$ are defined by

$$\begin{aligned} D_{2i}(\lambda_2) = \inf\{ &\sum_{n\in\mathcal{N}} \pi_n[c_{t(n)}^i(\xi^n) - \lambda_2^n b_{t(n)}^i(\xi^n)]x_i^n \,:\, x_i^n \in X_{t(n)}^i\,, \quad (37) \\ &\sum_{\kappa=0}^{t(n)-1} a_{t(n),t(n)-\kappa}^i(\xi^n)x_i^{n-\kappa} \geq g_{t(n)}^i(\xi^n)\,,\, n \in \mathcal{N}\}. \end{aligned}$$

By specifying (33) we obtain that $G_2(\bar{x}) = \{d_{t(n)}(\xi^n) - \sum_{i=1}^m b_{t(n)}^i(\xi^n)\bar{x}_i^n\}_{n\in\mathcal{N}}$ is a subgradient of D_2 at λ_2, where $\bar{x}_i = \{\bar{x}_i^n\}_{n\in\mathcal{N}}$ is a solution of (37). The dual function (36), which is defined by a multistage stochastic integer program of dimension $m|\mathcal{N}|$, decomposes into m functions each given by a multistage stochastic integer program of dimension $|\mathcal{N}|$. Since the dimension of the dual problem is $k|\mathcal{N}|$, the computational potential of this dualization approach takes effect in situations, where the number k of coupling constraints to be dualized is much smaller than the decision dimension m (i.e., $k << m$) and where the m subproblems (37) of dimension $|\mathcal{N}|$ can be solved

much faster than the original multistage model of dimension $m|\mathcal{N}|$. The latter could appear, for example, if complex mixed-integer models decompose into pure integer and pure linear programs.
Component decomposition has been applied successfully under the label *Lagrangian relaxation* of coupling constraints to solving hydro-thermal power management models under data uncertainty. Lagrangian relaxation has a long tradition for solving (deterministic) unit commitment problems of power systems operation. Recently, this technique has been extended to stochastic power management models, where the stochasticity enters the model, for example, via the electric load, streamflows to hydro units, and electricity prices. When letting the production decisions of individual power units play the role of components, the above dualization scheme leads to a decomposition into single (thermal or hydro) power unit models. Such approaches for determining lower bounds have been proposed and implemented in [3,13,19,33,41]. In [19,20,32] encouraging numerical results and computing times have been reported for both solving the dual and determining a nearly optimal primal solution by a Lagrangian based heuristic.

3.3 Nodal Decomposition

Finally, we return to the dualization of the dynamic constraints of (21) - (23) in case of a discrete underlying probability distribution and show that the dual function exhibits a nodewise decoupling structure. We let D_3 and Λ_3 be defined as in Section 2.5 and consider the corresponding dual problem

$$\max\{D_3(\lambda_3) \; : \; \lambda_3 \in \Lambda_3\}.$$

Let $\{\xi^n\}_{n\in\mathcal{N}}$ be the scenario tree representing the data process ξ, $\mathcal{N}$ the finite set of nodes, $\{\pi_n\}_{n\in\mathcal{N}}$ the node probabilities, and $\{x^n\}_{n\in\mathcal{N}}$ and $\{\lambda_3^n\}_{n\in\mathcal{N}}$ the corresponding scenario trees of the decision and of the multiplier process, respectively. Using the notation of Section 2.4, the dual function D_3 takes the following scenario tree representation

$$\begin{aligned} D_3(\lambda_3) = \inf\{c_1(\xi^1)x^1 + & \sum_{n\in\mathcal{N}\setminus\{1\}} \pi_n[c_{t(n)}(\xi^n)x^n + \lambda_3^n(g_{t(n)}(\xi^n) \\ & - \sum_{\kappa=0}^{t(n)-1} A_{t(n),t(n)-\kappa}(\xi^n)x^{n-\kappa})] \; : \\ & x^n \in X_{t(n)} \, , \, B_{t(n)}(\xi^n)x^n \geq d_{t(n)}(\xi^n) \, , \, n \in \mathcal{N}\} \, , \end{aligned} \tag{38}$$

where $\lambda_3 \in \Lambda_3 = \{\{\lambda_3^n\}_{n\in\mathcal{N}} \; : \; \lambda_3^n \geq 0 \, , \, n \in \mathcal{N}\}$. Since the minimization problem in (38) contains only node constraints for the decision tree, we rearrange its objective function with respect to the decision nodes and obtain

$$\begin{aligned} D_3(\lambda_3) = \inf\{ & \sum_{n\in\mathcal{N}} \pi_n (c_{t(n)}(\xi^n) - \sum_{\ell\in\mathrm{Tr}(n)} \pi_\ell \lambda_3^\ell A_{t(\ell),t(n)}(\xi^\ell))x^n \\ & + \sum_{n\in\mathcal{N}\setminus\{1\}} \pi_n \lambda_3^n g_{t(n)}(\xi^n) \ : \\ & x^n \in X_{t(n)}\, ,\, B_{t(n)}(\xi^n)x^n \geq d_{t(n)}(\xi^n)\, ,\, n\in\mathcal{N}\}\, , \end{aligned}$$

where $\mathrm{Tr}(1) := \mathcal{N}\setminus\{1\}$, and $\mathrm{Tr}(n)$ for $n > 1$ denotes the set of all nodes belonging to the subtree with root node n, i.e., $\mathrm{Tr}(n) := \cup_{n_T\in\mathcal{N}_T}\{\mathrm{path}(n_T) : n \in \mathrm{path}(n_T)\} \setminus \mathrm{path}(n_-)$. Now, we may interchange summation and minimization and arrive at the node decomposed formulation

$$D_3(\lambda_3) = \sum_{n\in\mathcal{N}} D_{3n}(\lambda_3) + \sum_{n\in\mathcal{N}\setminus\{1\}} \pi_n \lambda_3^n g_{t(n)}(\xi^n) \tag{39}$$

of D_3, where the functions D_{3n}, $n \in \mathcal{N}$, are defined on Λ_3 and given by

$$\begin{aligned} D_{3n}(\lambda_3) := \inf\{(\pi_n c_{t(n)}(\xi^n) - & \sum_{\ell\in\mathrm{Tr}(n)} \pi_\ell \lambda_3^\ell A_{t(\ell),t(n)}(\xi^\ell))x^n \ : \\ & x^n \in X_{t(n)}\, ,\, B_{t(n)}(\xi^n)x^n \geq d_{t(n)}(\xi^n)\}\, . \end{aligned} \tag{40}$$

Hence, the representation (39) of D_3 provides a decomposition of the original mixed-integer program of dimension $\sum_{n\in\mathcal{N}} m_{t(n)}$ into $|\mathcal{N}|$ subproblems (40) of dimension $m_{t(n)}$ for $n \in \mathcal{N}$. Formulas for computing subgradients of D_3 may be derived similarly to the previous section. Computational experience of such nodal decomposition schemes for determining lower bounds of multi-stage stochastic integer programs is not available yet.

Acknowledgement. We wish to thank Krzysztof C. Kiwiel (Systems Research Institute, Polish Academy of Sciences, Warsaw) for helpful discussions on the subject of this paper.

References

1. Ahmed, S., Tawarmalani, M., Sahinides, N.V.: A finite branch and bound algorithm for two-stage stochastic integer programs, preprint, University of Illinois, 2000
2. Artstein, Z., Wets, R.J-B: Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. SIAM J. Optim. **4** (1994) 537–550
3. Bacaud, L., Lemaréchal, C., Renaud, A., Sagastizábal, C.: Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners. Comput. Optim. Appl. (to appear)
4. Bachmann, H., Schultz, R.: Scenario grouping for multi-stage stochastic programs with integer requirements (in preparation)

5. Birge, J.R.: Stochastic programming computation and applications. INFORMS J. Comput. **9** (1997) 111–133
6. Birge, J.R., Dempster, M.A.H.: Stochastic programming approaches to stochastic scheduling, J. Glob. Optim. **9** (1996) 417–451
7. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York, 1997
8. Carøe, C.C.: Decomposition in stochastic integer programming. Ph.D. thesis, Institute of Mathematical Sciences, University of Copenhagen, 1998
9. Carøe, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. Oper. Res. Lett. **24** (1999) 37–45
10. Carøe, C.C., Schultz, R.: A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal system, Schwerpunktprogramm "Echtzeit-Optimierung großer Systeme" der Deutschen Forschungsgemeinschaft, Preprint 98-13, 1998, revised 1999
11. Carøe, C.C., Tind, J.: L-shaped decomposition of two-stage stochastic programs with integer recourse, Math. Progr. **83** (1998) 451-464
12. Dempster, M.A.H.: On stochastic programming II: Dynamic problems under risk. Stochastics **25** (1988) 15–42
13. Dentcheva, D., Römisch, W.: Optimal power generation under uncertainty via stochastic programming. In: Stochastic Programming Methods and Technical Applications (Marti, K., Kall, P. eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer-Verlag, Berlin 1998, 22–56
14. Dupačová, J.: Multistage stochastic programs: The state-of-the-art and selected bibliography. Kybernetika **31** (1995) 151–174
15. Dynkin, E.B., Evstigneev, I.V.: Regular conditional expectation of correspondences. Theory Probab. Appl. **21** (1976) 325–338
16. Evstigneev, I.: Measurable selection and dynamic programming. Math. Oper. Res **1** (1976) 267–272
17. Feltenmark, S., Kiwiel, K.C.: Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. SIAM J. Optim. **10** (2000) 697–721
18. Fiedler, O., Römisch, W.: Stability in multistage stochastic programming. Ann. Oper. Res. **56** (1995) 79–93
19. Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. Preprint 99-19, Institut für Mathematik, Humboldt-Universität Berlin, 1999 and IMA Volumes of Mathematics and its Applications, Springer-Verlag (to appear)
20. Gröwe-Kuska, N., Nowak, M.P., Römisch, W., Wegner, I.: Modeling of uncertainty for the real-time management of power systems. Submitted to this volume
21. Hemmecke, R., Schultz, R.: Decomposition methods for two-stage stochastic integer programs. Submitted to this volume
22. Higle, J.L., Sen, S.: Duality in multistage stochastic programs. In: Prague Stochastics '98 (Hušková, M., Lachout, P., Višek, J.Á., eds.), JČMF, Prague 1998, 233-236
23. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms II, Springer, Berlin, 1993
24. Kall, P., Wallace, S.W.: Stochastic Programming. Wiley, Chichester, 1994

25. Kiwiel, K.C.: Proximity control in bundle methods for convex nondifferentiable minimization. Math. Progr. **46** (1990) 105–122
26. Klein Haneveld, W.K., van der Vlerk, M.H.: Stochastic integer programming: General models and algorithms. Ann. Oper. Res. **85** (1999) 39–57
27. Laporte, G., Louveaux, F.V.: The integer L-shaped method for stochastic integer programs with complete recourse, Oper. Res. Lett. **13** (1993) 133-142
28. Lemaréchal, C.: Lagrangian decomposition and nonsmooth optimization: Bundle algorithm, prox iteration, augmented Lagrangian. In: Nonsmooth Optimization, Methods and Applications (Giannessi, F. ed.), Gordon & Breach, Philadelphia, 1992, 201–216
29. Lemaréchal, C., Renaud, A.: A geometric study of duality gaps, with applications. Math. Progr. (to appear)
30. Løkketangen, A., Woodruff, D.L.: Progressive hedging and tabu search applied to mixed integer (0,1) multi-stage stochastic programming, J. of Heuristics **2** (1996) 111-128
31. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization, Wiley, New York, 1988
32. Nowak, M.P.: Stochastic Lagrangian relaxation in power scheduling of a hydro-thermal system under uncertainty. Ph.D. thesis, Institute of Mathematics, Humboldt-University Berlin, 2000
33. Nowak, M.P., Römisch, W.: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. Ann. Oper. Res. (to appear)
34. Prékopa, A.: Stochastic Programming. Kluwer, Dordrecht, 1995
35. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics. Preprint 00-22, Institut für Mathematik, Humboldt-Universität Berlin, 2000
36. Rockafellar, R.T.: Duality and optimality in multistage stochastic programming. Ann. Oper. Res. **85** (1999) 1–19
37. Rockafellar, R.T., Wets, R.J-B: Nonanticipativity and L^1-martingales in stochastic optimization problems. Math. Progr. Study **6** (1976) 170–187
38. Rockafellar, R.T., Wets, R.J-B: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. SIAM J. Contr. Optim. **16** (1978) 16–36
39. Rockafellar, R.T., Wets, R.J-B: Scenarios and policy aggregation in optimization under uncertainty. Math. Oper. Res. **16** (1991) 119–147
40. Rockafellar, R.T., Wets, R.J-B: Variational Analysis. Springer-Verlag, Berlin, 1997
41. Römisch, W., Schultz, R.: Decomposition of a multi-stage stochastic program for power dispatch, Zeitschr. Angew. Math. Mech. 76 (1996) Suppl. 3, 29–32
42. Ruszczyński, A.: Decomposition methods in stochastic programming. Math. Progr. **79** (1997) 333–353
43. Schultz, R.: On structure and stability in stochastic programs with random technology matrix and complete integer recourse. Math. Progr. **70** (1995), 73–89
44. Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse. SIAM J. Optim. **6** (1996), 1138–1152
45. Schultz, R., Stougie, L., van der Vlerk, M. H.: Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions, Math. Progr. **83** (1998) 229-252

46. Stougie, L.: Design and Analysis of Algorithms for Stochastic Integer Programming, CWI Tract 37, Centrum voor Wiskunde en Informatica, Amsterdam, 1987
47. Takriti, S., Birge, J.R., Long, E.: A stochastic model for the unit commitment problem, IEEE Trans. Power Syst. **11** (1996), 1497–1508
48. van der Vlerk, M.H.: Stochastic programming with integer recourse, Ph.D. thesis, University of Groningen, 1995
49. Wets, R.J-B: Stochastic programming. in: Handbooks in Operations Research and Management Science, Vol. 1, Optimization (G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd eds.), North-Holland, Amsterdam 1989, 573–629
50. Wets, R.J-B, Ziemba, W.T.: Stochastic programming. State of the Art, 1998, Ann. Oper. Res. **85** (1999)