Duality gaps in nonconvex stochastic optimization *

Darinka Dentcheva and Werner Römisch

Stevens Institute of Technology
Mathematical Sciences
Hoboken, NJ 07030, U.S.A.
ddentche@stevens-tech.edu

Humboldt University Berlin
Institute of Mathematics
D-10099 Berlin, Germany
romisch@mathematik.hu-berlin.de

Abstract

We consider multistage stochastic optimization models. Logical or integrality constraints, frequently present in optimization models, limit the application of powerful convex analysis tools. Different Lagrangian relaxation schemes and the resulting decomposition approaches provide estimates of the optimal value. We formulate convex optimization models equivalent to the dual problems of the Lagrangian relaxations. Our main results compare the resulting duality gap for these decomposition schemes. Attention is paid also to programs that model large systems with loosely coupled components.

Key words: Stochastic programming, Lagrangian relaxation, nonconvex, mixed-integer, duality gap, decomposition
2000 MSC: 90C15, 90C11

1 Introduction

Stochastic dynamic programs arise as optimization models of systems driven by some discrete-time stochastic process $\{\xi_t : t = 1, 2, \ldots\}$. Let ξ_t be defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^{s_t}$. More precisely, let $\xi_t \in L_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{s_t})$. Let us assume that our modeling time horizon includes T time periods, and we make sequential decisions $x_t \in \mathbb{R}^{q_t}$ at every time interval (stage) $t = 1, 2, \ldots, T$ based on the information available at that time. We shall denote the information available at time period t by $\zeta_t := (\xi_1, \xi_2, \ldots, \xi_t)$. The condition that x_t may depend only on ζ_t is known as *nonanticipativity* condition. This property is equivalent to the measurability of x_t with respect to the σ-algebra $\mathcal{F}_t \subseteq \mathcal{F}$ that is generated by ζ_t. Clearly, the set $\{\mathcal{F}_t\}$ forms a filtration, i.e., $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, and we assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. Furthermore, there are constraints associated with each time period expressed in the following general form:

$$x_t \in X_t(\zeta_t), \quad t = 1, 2, \ldots, T.$$

*Supported by the Priority Programme *Online Optimization of Large Scale Systems* of the Deutsche Forschungsgemeinschaft

We shall assume that the sets X_t are compact $\mathbb{P}$-almost surely. Furthermore, we assume that X_t are nonconvex sets, which may be due to different reasons. A typical example are mixed-integer optimization problems. Some integrality requirements may be incorporated into the definition of the set X_t, e.g., $(x_t)_j \in \mathbb{Z}, j \in J_t$ for some index set $J_t : |J_t| \leq q_t$, where $\mathbb{Z}$ is the set of all integers. Other constraints in the description of the set X_t may involve nonconvex functions as well.
The dynamics of the system is described by the following inequality:

$$\sum_{\tau=1}^{t} A_{t,\tau}(\zeta_t) x_\tau \geq c_t(\zeta_t).$$

We shall consider the following stochastic optimization problem:

$$\min_x \; \mathbb{E}\Big[\sum_{t=1}^{T} f_t(\zeta_t, x_t)\Big] \tag{1}$$

$$\text{subject to} \; \sum_{\tau=1}^{t} A_{t,\tau}(\zeta_t) x_\tau \geq c_t(\zeta_t) \quad t = 1, \dots T, \quad \mathbb{P}\text{-a.s.}, \tag{2}$$

$$x_t \in X_t(\zeta_t) \quad t = 1, \dots T, \quad \mathbb{P}\text{-a.s.}, \tag{3}$$

$$x_t = \mathbb{E}\Big[x_t | \mathcal{F}_t\Big] \quad t = 1, \dots, T, \quad \mathbb{P}\text{-a.s.}. \tag{4}$$

Throughout the paper we shall assume that all necessary integrability requirements are satisfied, so that the problem is well defined. Moreover, all functions f_t, $t = 1, 2, \dots T$ will be assumed lower semi-continuous with respect to the second argument.

Let us summarize that the optimization model (1)-(4) consists in minimizing the expected cost subject to three groups of constraints. The first group (2) expresses the dynamics of the system. The second group are constraints describing feasibility requirements (3) for the decisions x_t at stage t, which may include integrality requirements for some of the decisions. The nonanticipativity condition (4) is a linear constraint on the decision process. All constraints are imposed $\mathbb{P}$-almost surely.

In order to solve such a stochastic optimization problem, the stochastic process is approximated by a set of scenarios. They form the basis of a deterministic optimization problem which replaces the stochastic problem. We shall refer to the scenario-based deterministic problem as a multistage stochastic programming problem. The approximation typically leads to a model of very large dimensions. The large size, the combination of different type of constraints, and the nonconvexity (e.g., integrality requirements) turn the multistage problem into a theoretical and numerical challenge. In the recent years different decomposition approaches to these type of problems were suggested (see [6, 26] for an overview). Let us mention the primal nested Benders decomposition and the regularized decomposition methods (see [4, 23, 24]). Dual approaches associate Lagrange multipliers with some group of the constraints and make use of the solution of some "dual" problem. Most of the approaches as progressive hedging ([19]), and the augmented Lagrangian decomposition suggested in ([25, 18]), relax the nonanticipativity constraints (4). Lagrange multipliers properties of some

inequality constraints like those that may be incorporated in (3) are investigated in [20] but no relaxation or decomposition is further pursued there. Nodal decomposition is a technique that associates Lagrange multipliers with dynamics constraints (2) (see [22]). Decomposition methods for stochastic programming models with integrality constraints are suggested in [7, 8, 21]. Our analysis will focus on comparing the duality gaps for the established decomposition techniques. The main result, presented in Section 4 will show that the scenario decomposition provides a better bound (smaller duality gap) than the nodal decomposition. The issue how to estimate the duality gap for mixed-integer optimization problems, and how to compare different relaxation approaches is investigated in [1, 3, 13, 12, 16, 14]. The general results and techniques in these papers are not applicable to estimate and compare the stochastic programming decomposition approaches which we consider. The specific structure of the stochastic models requires different approach, however, our study is inspired by those works. In the context of stochastic programming some related research is presented in [5, 27, 28]. In the papers [5, 28] the authors try to establish quantitative estimates for the duality gap arising when scenario decomposition approach is applied to solving mixed-integer multistage stochastic programs. The example presented in the paper [27] contradicts some of the results obtained in [5].
Another decomposition approach has been suggested for the unit commitment problem under uncertain load (cf.[10, 17]). For related work we refer to [2, 9, 29]. The decomposition is based on a certain type of structure of the model, which is frequently present in models of large systems. Many complex systems consist of components, which require their own independent model and have costs associated with their operation. The operation of the components is further coordinated by constraints linking their models. Therefore, the model of the whole system has to large extend a separable structure with respect to the system's components. We shall place a special attention to the decomposition technique, which takes advantage of this property. We call this approach geographical decomposition. The precise description of the model and its analysis are presented in Section 5. The geographical decomposition is compared to the nodal and scenario decomposition in Section 6. For an optimization problem we introduce a measure of sensitivity to relaxation of constraints and use it to characterize the relative effectiveness of the decomposition approaches.

2 Formulations of multistage stochastic programming problems based on scenarios

Let us consider the multistage problem based on S scenarios. We may think of it as a special case of the stochastic program in which the set Ω is finite, i.e., $\Omega = \{\omega_1, \ldots, \omega_S\}$. For simplicity we shall identify each scenario ω_s with its index s, and from now on, we shall assume that $\Omega = \{1, 2, \ldots, S\}$. In this case $\mathcal{F}$ is the power set of Ω and $\mathbb{P}(\{s\}) = p_s$, $s = 1, 2, \ldots, S$ with $\sum_{s=1}^{S} p_s = 1$. We denote the value of the process ξ for the scenario s at stage t by $\xi_{s,t}$. Correspondingly, $x_{s,t}$ will denote the value of the decision for the scenario s at stage t, where $s = 1, 2, \ldots, S$ and $t = 1, 2, \ldots, T$. There exists a finite partition $\mathcal{E}_t$ of Ω for each t, $t = 1, 2, \ldots, T$, such that $\mathcal{E}_t$ generates the σ-algebra $\mathcal{F}_t$. Moreover, since $\mathcal{F}_t$ form a filtration, each element C of $\mathcal{E}_t$ is further partitioned into sets C_j, $j = 1, \dots, S_C$, which are members of the partition $\mathcal{E}_{t+1}$. Clearly, the partition $\mathcal{E}_T$ consists of the members $\{s\}$, $s = 1, 2, \dots, S$. Using the partitioning we can express the conditional expectation with respect to $\mathcal{F}_t$ in the nonanticipativity constraint as the following sum:

$$\mathbb{E}[x_t | \mathcal{F}_t] = \sum_{C \in \mathcal{E}_t} \frac{1}{\mathbb{P}(C)} \int_C x_t(\omega)\mathbb{P}(d\omega)\chi_C = \sum_{C \in \mathcal{E}_t} (\sum_{\omega_s \in C} p_s)^{-1} (\sum_{\omega_s \in C} p_s x_{s,t})\chi_C.$$

Here χ_C denotes the characteristic function of the set $C \in \mathcal{E}_t$. We define for each scenario σ the set $C_{\sigma,t}$ to be the unique element of $\mathcal{E}_t$ that contains σ. Then the nonanticipativity constraint corresponds the following system of equations

$$x_{\sigma,t} = (\sum_{s \in C_{\sigma,t}} p_s)^{-1} \sum_{s \in C_{\sigma,t}} p_s x_{s,t} \quad \sigma = 1, 2, \dots, S, \; t = 1, 2, \dots, T.$$

For $t = 1$ we have $\mathcal{E}_1 = \{\Omega\}$, and the latter condition reads

$$x_{\sigma,1} = \sum_{s=1}^{S} p_s x_{s,1} \quad \sigma = 1, 2, \dots, S,$$

i.e., the nonanticipativity condition is equivalent to the equations $x_{1,1} = x_{2,1} = \cdots = x_{S,1}$.

The nonanticipativity constraint can be formulated in another way which yields a simple and sparse structure (see [18]). Here is how it goes. Let the last common stage for the scenarios ζ and θ be denoted by

$$t^{\max}(\zeta, \theta) = \max\{t : \zeta_\tau = \theta_\tau, \; \tau = 1, \dots, t\}.$$

We can always order the scenarios $1, 2, \dots s$ so that $t^{\max}(\zeta, \zeta + 1) = \max\{t^{\max}(\zeta, j) : j > \zeta\}$. Next for every scenario s and every stage t, we define the sibling of s as follows:

$$\alpha(s,t) = \begin{cases} s+1 & \text{if } t^{\max}(s, s+1) \geq t \\ \min\{j : \, t^{\max}(s,j) \geq t\} & \text{otherwise} \end{cases}$$

Observe that for every t, the mapping $\alpha(\cdot, t)$ defines a permutation of Ω, which maps bundles of undistinguishable scenarios onto themselves. Using this mapping we can formulate the nonanticipativity constraint as

$$x_{s,t} = x_{\alpha(s,t),t} \text{ for all } t = 1, \dots T, \text{ and } s \in \Omega.$$

We can formulate the stochastic dynamic problem as a multistage stochastic programming problem in the following manner:

$$\min_x \quad \mathbb{E}\Big[\sum_{t=1}^{T} \sum_{s=1}^{S} f_{s,t}(\zeta_{s,t}, x_{s,t}) \Big] \tag{5}$$

subject to

$$\sum_{\tau=1}^{t} A_{t,\tau}(\zeta_{s,t}) x_{s,\tau} \geq c_{s,t}(\zeta_{s,t}) \quad t = 1, \dots, T, \tag{6}$$

$$x_{s,t} \in X_{s,t} \quad t = 1, \dots, T, \tag{7}$$

$$x_{s,t} = x_{\alpha(s,t),t} \quad t = 1, \dots, T, \; s = 1, 2, \dots, S. \tag{8}$$

Generally, the number of elements in the partition $\mathcal{E}_t$ corresponds to the number of realizations or sample paths of ξ at time period t and to the number of realization of x at time period t. Graphical representation of the relations between the elements of $\mathcal{E}_t$ and the elements of $\mathcal{E}_{t+1}$ leads to a tree. The nodes stand for the elements of the partitions $\mathcal{E}_t$, e.g., the root node stands for the value of ξ at the time period $t = 1$, and a branching node corresponds to a member of $\mathcal{E}_t$ that is partitioned and represented as union of sets members of $\mathcal{E}_{t+1}$. The tree expresses the relations between the scenarios. The structure of the decisions can be represented by the same tree as well because any $x_{s,t}$ and $x_{l,t}$ are undistinguishable for $s \in C$, $l \in C$, and $C \in \mathcal{E}_t$ due to the nonanticipativity constraint. Therefore, we can associate both the realizations of the process $\xi_{s,t}$ and the decisions $x_{s,t}$ with the nodes of the tree. Let us enumerate the nodes of the tree setting $n = 1$ for the root node. Every other node n has a unique predecessor node $a(n)$. Every node n has a set of successors $\mathcal{S}(n)$. The set of successors is empty for the terminal nodes (leaves). Each member C of any partition $\mathcal{E}_t$, $t = 1, 2, \ldots, T$, is associated in an unique way with a node n. Let us observe that this is a unique correspondence $\beta : (s,t) \to n$ assigning a node n to a scenario s at a certain stage t. Note that $\zeta_{s,t}$ represents a path from the root to a certain node $n = \beta(s,t)$. We shall denote this path by ζ_n and its length by $t(n)$. Note also that $t(n)$ is the number of the stage at which the node occurs. A scenario is a path from the root to a leaf of the tree. Let us observe that $\mathcal{S}(n) = \emptyset$ for all nodes $n = \beta(s,T)$. These are exactly the terminal nodes representing a member $\{s\}$ of $\mathcal{E}_T$. Thus, the correspondence $\gamma : n \to s$ is well defined by setting $\gamma(n) = s$ if $n = \beta(s,T)$.
Using the probabilities of the scenarios, we can define probabilities associated with the nodes of the tree according to the following recursive procedure:

$$\pi_n = \begin{cases} p_{\gamma(n)} & \text{for all } n \text{ such that } \mathcal{S}(n) = \emptyset \\ \pi_n = \sum_{m \in \mathcal{S}(n)} \pi_m & \text{for all other nodes } n. \end{cases}$$

Let N_t for $t = 1, 2, ..., T$ denote the set of nodes at stage t, i.e., $N_t = \{n : t(n) = t\}$. It holds $\sum\limits_{n \in N_t} \pi_n = 1$ for each period $t = 1, 2, \ldots, T$. Clearly, $|N_T| = S$. Let us denote the number of all nodes by N. Using the mapping β, we can introduce a shorthand notation as follows. For all (s,t) let $n = \beta(s,t)$, then

$$\begin{aligned} f_n(x_n) &:= f_t(\zeta_{s,t}, x_{s,t}) \\ X_n &:= X_{s,t} \\ A_{n,\beta(s,\tau)} &:= A_{t,\tau}(\zeta_{s,t}) \\ c_n &:= c_t(\zeta_{s,t}). \end{aligned}$$

At this point we can formulate the multistage problem based on the scenario tree in the following way:

$$\min_x \quad \sum_{n=1}^{N} \pi_n f_n(x_n) \tag{9}$$

subject to

$$\sum_{\tau \in \zeta_n} A_{n,\tau} x_\tau \geq c_n, \quad n = 1, 2, \ldots, N, \tag{10}$$

$$x_n \in X_n, \quad n = 1, 2, \ldots, N. \tag{11}$$

This formulation of the multistage problem is commonly called a primal formulation. Let us observe that the formulation with respect to scenarios can be obtained by splitting the decision variables in (9)-(11) for each scenario and formulating the constraints for all split variables and introducing the explicit nonanticipativity constraint (8). The latter constraint is omitted in the model (9)-(11) because it is reflected in the tree structure of the decisions.

3 Lagrangian relaxation approaches and their dual equivalent convex problems

We can distinguish three relaxation ideas which lead to decomposition of the multistage optimization model:

- Lagrange multipliers are associated with the nonanticipativity constraints in the extended formulation (5)-(8). The problem decomposes into $\mathcal{S}$ subproblems, each one in $(\sum_{t=1}^{T} n_t)$-dimensional decision space expressing the optimal operation of the system under the realization of each scenario. This approach is frequently called *scenario decomposition* or *scenario disaggregation*.

- Lagrange multipliers are associated with the dynamic constraints at each node of the scenario tree. Let us emphasize that this formulation involves generally a much smaller decision space ($N << ST$). The problem decomposes into N subproblems each one in an n_t-dimensional decision space (for $t = 1, 2, \ldots, T$). The subproblems model the optimal operation of the system under the conditions determined by the node of the tree. This approach is called *nodal decomposition* in [22].

- Lagrangian relaxation by decoupling system components when the multistage program has loosely coupled structure. This approach will be called *geographical decomposition* in Section 5.

For more information about the scenario and nodal decomposition the reader is referred to [22, 19]. In the next subsections we shall describe the Lagrange function, the dual functional, and review some features of the first two relaxation approaches. The third approach requires a more precise description, so we defer its discussion to the next section.

Let us first recall some basic notation and theory. The *conjugate function* $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as follows:

$$f^*(y) = \sup \left\{ \langle y, x \rangle - f(x) \, : \, x \in \mathbb{R}^n \right\}.$$

Here $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Convexity is not needed for the conjugacy operation to make sense. We just need to assume that f is not identical to infinity and there exists an affine minorant of f. Note that this assumption implies that $f(x) > -\infty$ for all $x \in \mathbb{R}^n$.

The biconjugate function f^{**} of f is defined by

$$f^{**}(x) = (f^*)^*(x) = \sup\left\{\langle y, x\rangle - f^*(y) :\ y \in \mathbb{R}^n\right\}.$$

It is known that this operation provides the close-convexification of f.

Lemma 3.1 ([15], Ch.X, Theorem 1.3.5, p.45) *Assume that f is not identical to infinity and there exists an affine minorant of f. Then*

$$\mathrm{epi} f^{**} = \overline{\mathrm{co}}(\mathrm{epi} f).$$

Here epi$f$ refers to the epigraph of f and $\overline{\mathrm{co}}$ denotes the operation of taking the convex hull and closure. Thus, the biconjugate function f^{**} is the closed convex hull of the function f, i.e., the largest lower semicontinuous convex function below f.

Let us observe the following fact.

Lemma 3.2 *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ has an affine minorant. If f is a sum of a linear and nonlinear functions, i.e., $f(x) = a^T x + g(x)$ for some $a \in \mathbb{R}^n$, then $f^{**}(x) = a^T x + g^{**}(x)$.*

Proof: We have $f^*(y) = \sup\{y^T x - a^T x - g(x) : x \in \mathbb{R}^n\} = g^*(y - a)$ by definition of the conjugate function. Furthermore,

$$\begin{aligned} f^{**}(z) &= \sup\{z^T y - f^*(y) : y \in \mathbb{R}^n\} = \sup\{z^T (y-a) - g^*(y-a) + a^T z : y \in \mathbb{R}^n\} \\ &= \sup\{z^T (y-a) - g^*(y-a) : y \in \mathbb{R}^n\} + a^T z = g^{**}(z) + a^T z. \end{aligned}$$

This proves the statement. □

The indicator of a nonempty set A will be denoted by

$$\delta_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise} \end{cases}$$

Clearly, it holds that $[\delta_A]^{**} = \delta_{A^{**}}$, where $A^{**} = \overline{\mathrm{co}}\, A$.
Given the optimization problem

$$\begin{aligned} \min f(x) \text{ subject to} & \\ x \in X & \\ a_i x \le b_i, &\quad i = 1, \ldots, l, \end{aligned}$$

where f is a finite lower semicontinuous function, we denote its dual function by $D(y)$, i.e.,

$$D(y) = \min\left\{f(x) - \sum_{i=1}^{l} y_i (b_i - a_i x) : x \in X\right\}, \quad y \in \mathbb{R}^l_+.$$

Further, we can collect the nonlinearities and nonconvexities of f and X into f using the indicator function of X. We define

$$f_X = f + \delta_X.$$

We recall a result due to the pioneering work [11], also formulated in [16], which will be applied here.

Theorem 3.3 ([11, 16]) *Assume that X is a compact set. The function D is also the dual function associated with the following problem:*

$$\inf[f_X]^{**}(x) \quad \text{subject to } a_i x \le b_i, \quad i = 1, \dots, l. \tag{12}$$

Moreover, assume that there is a feasible point $\bar{x}$ lying in the relative interior of $dom[f_X]^{**}$. Then D attains its maximum, which is equal to the infimal value of (12).

3.1 Scenario decomposition

For simplicity, we assume that the dimension of the decision vector q_t is the same for all stages, i.e., $q_t = r$. All observations of this paper can be carried out without this assumption, which merely simplifies notation.

We transform the scenario formulation of the multistage problem in the following way:

$$\min_x \quad \sum_{t=1}^{T} \sum_{s=1}^{S} p_s v_t^s \tag{13}$$

subject to

$$f_{s,t}(\zeta_{s,t}, x_{s,t}) \le v_t^s, \quad x_{s,t} \in X_{s,t}, \tag{14}$$

$$\sum_{\tau=1}^{t} A_{t,\tau}(\zeta_{s,t}) x_{s,\tau} \ge c_{s,t}(\zeta_{s,t}), \tag{15}$$

$$x_{s,t} = x_{\alpha(s,t),t}, \quad s = 1, 2, \dots, S, \quad t = 1, 2, \dots, T. \tag{16}$$

We associate a Lagrange multiplier $\mu \in \mathbb{R}^{(TS-N)r}$ with the nonanticipativity constraint (16). Both formulations of the nonanticipativity constraint can be used for our analysis. The only issue relevant to our considerations is that the nonanticipativity condition is expressed by linear equality constraints. The Lagrange function reads:

$$L^{na}(x, v, \mu) = \sum_{s=1}^{S} p_s \sum_{t=1}^{T} \left[v_t^s + \mu_{s,t}(x_{s,t} - x_{\alpha(s,t),t}) \right]$$

The dual functional is defined as follows:

$$D^{na}(\mu) = \inf \left\{ L^{na}(x, v, \mu) : (x, v) \in \mathop{\times}_{s=1}^{S} \mathcal{X}_s^{na} \right\}$$

where the set of feasible solutions $\mathcal{X}^{na} = \times_{s=1}^{S} \mathcal{X}_s^{na}$ is defined by setting $\mathcal{X}_s^{na}$ for each scenario s, $s = 1, 2, \dots, S$, to be the set

$$\mathcal{X}_s^{na} = \Big\{ y^s \in \mathbb{R}^{T(r+1)} : f_{s,t}(x_{s,t}) \le v_t^s, \ x_{s,t} \in X_{s,t}, \\ \sum_{\tau=1}^{t} A_{t,\tau}(s) x_{s,\tau} \ge c_{s,t}, \ t = 1, \dots T \Big\},$$

where $y^s = (x_{s,1}, \dots, x_{s,T}, v_1^s, \dots, v_T^s)$. The dual problem is:

$$\sup \left\{ D^{na}(\mu) : \mu \in \mathbb{R}^{(ST-N)r} \right\}. \tag{17}$$

The dual functional decomposes into S subproblems associated with each scenario. Each subproblem optimizes the operation of the system when the stochastic process follows a particular scenario. We introduce the functions

$$F^s(y^s) := \sum_{t=1}^{T} p_s v_t^s,$$

which is the part of the objective function associated with the scenario s ($s = 1, \dots, S$). Let us consider the following convex optimization problem:

$$\inf \sum_{s=1}^{S} \left[F^s_{\mathcal{X}^{na}_s} \right]^{**}(y^s) \quad \text{subject to (16).} \tag{18}$$

Proposition 3.4 *The functional D^{na} is also the dual functional of problem (18). Moreover, assume that the problem (13)-(16) has a feasible solution lying in the relative interior of the set $[\mathcal{X}^{na}]^{**}$. Then D^{na} attains its maximum, which is equal to the infimal value of (18).*

Proof: According to Theorem 3.3 the dual functional D^{na} is also a dual functional to the following problem:

$$\min \left[\Big(\sum_{s=1}^{S} F^s \Big)_{\mathcal{X}^{na}} \right]^{**}(y) \quad \text{subject to} \quad (16).$$

Due to the linearity of F_s (Lemma 3.2) we obtain:

$$\left[\Big(\sum_{s=1}^{S} F^s \Big)_{\mathcal{X}^{na}} \right]^{**}(y) = [\sum_{s=1}^{S} F^s + \delta_{\mathcal{X}^{na}}]^{**}(y) = \sum_{s=1}^{S} F^s(y^s) + [\delta_{\mathcal{X}^{na}}]^{**}(y)$$

Further, due to the separability of the set $\mathcal{X}^{na}$ it holds true that

$$[\delta_{\mathcal{X}^{na}}]^{**}(y) = \delta_{[\mathcal{X}^{na}]^{**}}(y) = \sum_{s=1}^{S} \delta_{[\mathcal{X}^{na}_s]^{**}}(y^s) = \sum_{s=1}^{S} \left[\delta_{\mathcal{X}^{na}_s} \right]^{**}(y^s)$$

Combining the two equations, we obtain:

$$\begin{aligned} \left[\Big(\sum_{s=1}^{S} F^s \Big)_{\mathcal{X}^{na}} \right]^{**}(y) &= \sum_{s=1}^{S} F^s(y^s) + \sum_{s=1}^{S} [\delta_{\mathcal{X}^{na}_s}]^{**}(y^s) = \sum_{s=1}^{S} \left[(F^s + \delta_{\mathcal{X}^{na}_s}) \right]^{**}(y^s) \\ &= \sum_{s=1}^{S} \left[F^s_{\mathcal{X}^{na}_s} \right]^{**}(y^s) \end{aligned}$$

Let us observe also that $\operatorname{dom}\left[\Big(\sum_{s=1}^{S} F^s \Big)_{\mathcal{X}^{na}} \right]^{**} = \left[\mathcal{X}^{na} \right]^{**}$. Thus the constraint qualification of Theorem 3.3 is satisfied and we may conclude that the dual function $D^{na}(\mu)$ attains its maximum, which is equal to the infimum value of the problem (18). $\square$

The proposition gives us a primal convex optimization problem, namely (18), which is equivalent to the dual problem in the Lagrange relaxation of the nonanticipativity constraints. The Fenchel duality reveals the essence of this approach. This relaxation is equivalent to the "convexification" of the objective function and feasible set separately for each scenario.

3.2 Nodal decomposition

The next relaxation is associated with the primal formulation of the multistage problem. We shall simplify the dynamics of the system, assuming that there is a dependence only between two successive stages. This assumption does not limit the application of the results obtained here. It simplifies the notation and provides more clarity in the presentation of the main ideas. We introduce again an equivalent formulation converting the objective function to a linear objective.

$$\min_{x} \quad \sum_{n=1}^{N} \pi_n v_n \tag{19}$$

$$\text{subject to} \quad A_{n,n} x_n + A_{n,a(n)} x_{a(n)} \geq c_n, \tag{20}$$

$$f_n(x_n) \leq v_n, \quad x_n \in X_n, \quad n = 1, 2, \ldots, N. \tag{21}$$

The nodal decomposition associates Lagrange multipliers $\nu \in \mathbb{R}^{Nm}$ with the dynamic constraints (20), where we have assumed for simplicity that m is the fixed dimension of c_n for each $n \in N$. The set of feasible solutions $\mathcal{X}^d$ becomes

$$\mathcal{X}^d = \mathop{\times}_{n=1}^{N} \mathcal{X}_n^d = \mathop{\times}_{n=1}^{N} \big\{ (x_n, v_n) \in \mathbb{R}^{r+1} \,:\, f_n(x_n) \leq v_n,\, x_n \in X_n \big\}.$$

The Lagrange function and the dual functional are defined by

$$L^d(x, v, \nu) = \sum_{n=1}^{N} \pi_n \big[v_n + \nu_n (c_n - A_{n,n} x_n - A_{n,a(n)} x_{a(n)}) \big] \quad \text{and}$$

$$D^d(\nu) = \inf\{ L^d(x, v, \nu) : (x, v) \in \mathcal{X}^d \},$$

respectively, and the dual problem is

$$\sup\{ D^d(\nu) : \nu \in \mathbb{R}_+^{Nm} \}. \tag{22}$$

The dual functional D^d decomposes into N subproblems associated with each node of the scenario tree. Each subproblem optimizes the operation of the system for the data identified with the node. For each node $n = 1, 2, \ldots, N$ we introduce the function

$$\tilde{F}^n(y_n) := \pi_n v_n, \quad \text{where} \quad y_n = (x_n, v_n).$$

Let us consider the following convex optimization problem:

$$\inf \sum_{n=1}^{N} \left[\tilde{F}^n_{\mathcal{X}_n^d} \right]^{**} (y_n) \quad \text{subject to} \quad (20). \tag{23}$$

Proposition 3.5 *The functional D^d is also the dual functional of problem (23). Moreover, assume that there is a feasible solution of the problem (19)-(21) lying in the relative interior of the set $[\mathcal{X}^d]^{**}$. Then D^d attains its maximum, which is equal to the infimal value of (23).*

Proof: We follow the same line of arguments as in the proof of Proposition 3.4. According to Theorem 3.3 the dual functional D^d is also a dual functional to the following problem:

$$\min \Big[\Big(\sum_{n=1}^{N} \tilde{F}^n\Big)_{\mathcal{X}^d}\Big]^{**}(y) \quad \text{subject to} \quad (20).$$

Due to the linearity of the objective function (Lemma 3.2) and the separability of the feasible set $\mathcal{X}^d$, the following chain of equations holds true:

$$\begin{aligned}\Big[\Big(\sum_{n=1}^{N} \tilde{F}^n\Big)_{\mathcal{X}^d}\Big]^{**}(y) &= \Big[\sum_{n=1}^{N} \tilde{F}^n + \delta_{\mathcal{X}^d}\Big]^{**}(y) = \sum_{n=1}^{N} \tilde{F}^n(y_n) + [\delta_{\mathcal{X}^d}]^{**}(y) \\ &= \sum_{n=1}^{N} \tilde{F}^n(y_n) + \sum_{n=1}^{N} [\delta_{\mathcal{X}^d_n}]^{**}(y_n) = \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}^d_n}\Big]^{**}(y)\end{aligned}$$

Observe, that the constraint qualification of Theorem 3.3 is satisfied as well and we conclude that the dual function $D^d(\nu)$ attains its maximum, which is equal to the infimal value of the problem (23). □

Due to Proposition 3.5, problem (23) is a primal convex program which is equivalent to the dual problem in the Lagrange relaxation of the dynamic constraints. This relaxation is equivalent to the "convexification" of the objective function and of the feasible set separately for each node of the scenario tree.

4 Scenario versus nodal decomposition

Now, we are ready to compare the duality gap of the introduced Lagrangian relaxations for multistage stochastic programs.

Theorem 4.1 *Assume that the convex hull of the feasible set of the problem (19)-(21) has nonempty relative interior, then the scenario decomposition provides a better bound for the optimal value of the multistage problem than the nodal decomposition, i.e., the following inequality holds true:*

$$\sup_{\nu} D^d(\nu) \le \sup_{\mu} D^{na}(\mu).$$

Proof: Let us introduce the following notation, which will simplify the presentation. The following set is associated with the dynamics constraints:

$$G = \big\{y = (x, v) \in \mathbb{R}^{N(r+1)} : A_{n,n}x_n + A_{n,a(n)}x_{a(n)} \ge c_n, \quad n = 1, 2, \dots, N\big\}$$

For each scenario s, $s = 1, 2, \dots, S$, we set $x^s = (x_{s,1}, \dots, x_{s,T})$ and v^s accordingly, and define the sets

$$G^s = \big\{(x^s, v^s) \in \mathbb{R}^{T(r+1)} : A^s_{t,t}x^s_t + A_{t,t-1}x^s_{t-1} \ge c^s_t, \quad t = 1, 2, \dots, T\big\}$$

and

$$Y_t^s = \big\{(x_{s,t}, v_{s,t}) \in \mathbb{R}^{r+1} : x_{s,t} \in X_{s,t}, f_{s,t}(x_{s,t}) \le v_t^s\big\}.$$

Furthermore, let $Y^s = \times_{t=1}^T Y_t^s$. Let us observe that for each scenario s the relation

$$\mathcal{X}_s^{na} = G^s \cap Y^s$$

holds. Furthermore, if the problem (19)-(21) has a feasible solution in the relative interior of the set $[\mathcal{X}^d \cap G]^{**}$, then this solution is contained in the set G and in the relative interior of the larger set $[\mathcal{X}^d]^{**}$. Consequently, the constraint qualification of Proposition 3.5 is satisfied. Moreover, we can split the variables for each scenario and obtain a feasible solution of the problem (13)-(16). This solution will be contained in the relative interior of $[\mathcal{X}^{na}]^{**}$ by construction. Therefore, the assumptions of Proposition 3.4 are satisfied as well. By relaxing the nonanticipativity constraints one obtains a lower bound $\hat{D}^{na}$ of the objective function

$$\hat{D}^{na} = \max_{\mu} D^{na}(\mu) \tag{24}$$

According to Proposition 3.4 there exists a solution $\bar{y} = (\bar{x}, \bar{v}) \in \mathbb{R}^{TS(r+1)}$ of the convex equivalent problem (18) such that:

$$\hat{D}^{na} = \inf\{\sum_{s=1}^S \Big[F^s_{\mathcal{X}_s^{na}}\Big]^{**}(y^s): \quad \text{subject to (16)}\} = \sum_{s=1}^S \Big[F^s_{\mathcal{X}_s^{na}}\Big]^{**}(\bar{y}_s) \tag{25}$$

and $\bar{y}$ satisfies the nonanticipativity constraint. Using the linearity of F_s and separability of the set $\mathcal{X}^{na}$, we obtain

$$\begin{aligned}
\hat{D}^{na} &= \sum_{s=1}^S \Big[F^s_{\mathcal{X}_s^{na}}\Big]^{**}(\bar{y}_s) = \sum_{s=1}^S \Big[F^s + \delta_{G^s\cap Y^s}\Big]^{**}(\bar{y}_s) \\
&= \sum_{s=1}^S \Big[F^s(\bar{y}_s) + \delta_{[G^s\cap Y^s]^{**}}\Big](\bar{y}_s).
\end{aligned} \tag{26}$$

Let us observe that $[G^s \cap Y^s]^{**} \subseteq (G^s)^{**} \cap (Y^s)^{**} = G^s \cap (Y^s)^{**}$. Consequently,

$$\delta_{[G^s\cap Y^s]^{**}}(\bar{y}_s) \ge \delta_{G^s\cap (Y^s)^{**}}(\bar{y}_s) = \delta_{G^s}(\bar{y}_s) + \delta_{(Y^s)^{**}}(\bar{y}_s).$$

We can continue the chain of transformations (26) in the following way:

$$\begin{aligned}
\sum_{s=1}^S \Big[F^s(\bar{y}_s) + \delta_{[G^s\cap Y^s]^{**}}(\bar{y}_s)\Big] &\ge \sum_{s=1}^S \Big[F^s(\bar{y}_s) + \delta_{G^s}(\bar{y}_s) + \delta_{(Y^s)^{**}}(\bar{y}_s)\Big] \\
&= \sum_{s=1}^S F^s(\bar{y}_s) + \delta_{(\times_{s=1}^S G^s)}(\bar{y}) + \delta_{\times_{s=1}^S (Y^s)^{**}}(\bar{y})
\end{aligned} \tag{27}$$

Let us define $\tilde{y} = (\tilde{x}, \tilde{v}) \in \mathbb{R}^{N(r+1)}$ to be the projection of $\bar{y}$ corresponding to the node formulation of the multistage problem. That is, we set: $\tilde{y}_n := \bar{y}_{s,t}$ for $n = \beta(s,t)$, and $\pi_n = \sum_{(s,t):\beta(s,t)=n} p_s$. This definition is non ambiguous because $\bar{y}$ satisfies the nonanticipativity constraint. Therefore,

$$\sum_{s=1}^{S} F^s(\bar{y}_s) = \sum_{n=1}^{N} \tilde{F}_n(\tilde{y}_n).$$

We can represent the set $\times_s(Y^s)$ as follows: $\times_s(Y^s) = \times_s \times_t Y^s_t$. By the definition of the models (19)-(21) and (13)-(16) the sets $Y^s_t = Y^{\alpha(s,t)}_t$ for all $s = 1, 2, \ldots, S$, $t = 1, 2, \ldots, T$. Therefore, we can write $Y^s_t = \mathcal{X}^d_n$ for the node $n = \beta(s,t)$. Since $\bar{y}_s \in [\mathcal{X}^{na}_s]^{**}$ we obtain that

$$\tilde{y}_n \in [\mathcal{X}_n]^{**} \text{ and } \delta_G(\tilde{y}) = 0. \tag{28}$$

Thus, we obtain that

$$\begin{aligned}
\sum_{s=1}^{S} F^s(\bar{y}_s) + \delta_{\times_{s,t}(Y^s_t)^{**}}(\bar{y}) + \delta_{\times_s G^s}(\bar{y}) &= \sum_{n=1}^{N} \tilde{F}_n(\tilde{y}_n) + \delta_{\times_n \mathcal{X}^{**}_n}(\tilde{y}) + \delta_G(\tilde{y}) \\
= \sum_{n=1}^{N} \left[\tilde{F}_n + \delta_{\mathcal{X}_n}\right]^{**}(\tilde{y}_n) + \delta_G(\tilde{y}) &= \sum_{n=1}^{N} \left[\tilde{F}_{n\mathcal{X}_n}\right]^{**}(\tilde{y}_n) + \delta_G(\tilde{y}) \\
&\geq \min\left\{\sum_{n=1}^{N} \left[\tilde{F}_{n\mathcal{X}_n}\right]^{**}(y_n) + \delta_G(y)\right\} \quad (29)
\end{aligned}$$

According to Proposition 3.5 the right hand side of the latter inequality is equal to the optimal value $\max_\nu D^d(\nu)$ of the dual problem associated with the relaxation of the dynamic constraints. From the chain of inequalities (26), (27), and (29), we obtain:

$$\max_\mu D^{na}(\mu) = \sum_{s=1}^{S} [F^s_{\mathcal{X}^{na}_s}]^{**}(\bar{y}_s) \geq \sum_{n=1}^{N} \left[\tilde{F}_{n\mathcal{X}_n}\right]^{**}(\tilde{y}_n) + \delta_G(\tilde{y}) \geq \max_\nu D^d(\nu)$$

which is the desired inequality. □

In general, the estimate in Theorem 4.1 is strict. To see this consider a multistage problem with several scenarios which happen to coincide. In that case, the relaxation of the nonanticipativity constraints does not induce a duality gap, i.e., the supremum of D^{na} coincides with the infimum of the primal problem. However, in the same situation, relaxing the dynamic constraints induces a duality gap if the dynamics of the system impose restrictions to the decisions of consecutive states.

5 Geographical decomposition

In this section, we turn to complex systems with loosely coupled components. We shall assume that system components require their own models which are coordinated by several linking constraints. Furthermore, we assume that the objective function is separable with respect to components. Therefore, the model of the entire system has to large extent separable structure with respect to the system's components. We would like to associate geographical locations with the components and refer further to them as locations. Let us assume that the modeling system comprises I locations and $x^i \in \mathbb{R}^{n_i}$ is the portion of the decisions associated with location i, $i = 1, 2, \ldots, I$, and $\sum_{i=1}^{I} n_i = r$.

We shall deal with the following multistage stochastic optimization problem:

$$\min_{x} \quad \mathbb{E}\Big[\sum_{t=1}^{T}\sum_{i=1}^{I} f_t^i(\zeta_t, x_t^i)\Big] \tag{30}$$

$$\text{subject to} \quad \sum_{\tau=1}^{t} A_{t,\tau}^i(\zeta_t)x_\tau^i \geq c_t^i(\zeta_t), \quad i = 1, 2, \ldots, I, t = 1, 2, \ldots, T, \tag{31}$$

$$\sum_{i=1}^{I} B_t^i(\zeta_t)x_t^i \geq d_t(\zeta_t), \quad t = 1, 2, \ldots, T, \tag{32}$$

$$x_t^i \in X_t^i(\zeta_t), \quad i = 1, 2, \ldots, I, t = 1, 2, \ldots, T, \tag{33}$$

$$x_t^i = \mathbb{E}\big[x_t^i|\mathcal{F}_t\big], \quad t = 1, 2, \ldots, T, i = 1, 2, \ldots, I. \tag{34}$$

Here B_t^i are $K \times n_i$-dimensional matrices and d_t are K-dimensional vectors. All constraints are imposed $\mathbb{P}$-a.s. Thus, there are K constraints (32) for each sequential decision that are coupling the models of the locations. The assumption that the modeled system consists of *loosely coupled locations* means that $K << I$. The model written in a primal formulation with the transformations we have adopted in the previous section reads:

$$\min_{(x,v)} \quad \sum_{n=1}^{N}\sum_{i=1}^{I} \pi_n v_n^i \tag{35}$$

$$\text{subject to} \quad A_{n,n}^i(x_n^i) + A_{n,a(n)}^i x_{a(n)}^i \geq c_n^i, \; n = 1, 2, \ldots, N, i = 1, 2, \ldots, I, \tag{36}$$

$$x_n^i \in X_n^i, \quad f_n^i(x_n^i) \leq v_n^i, \quad n = 1, 2, \ldots, N, i = 1, 2, \ldots, I, \tag{37}$$

$$\sum_{i=1}^{I} B_n^i x_n^i \geq d_n, \quad n = 1, 2, \ldots, N. \tag{38}$$

As the third decomposition approach we consider the decoupling of locations. Let us associate Lagrange multipliers $\lambda \in \mathbb{R}^{NK}$ with the coupling constraints (38). Then we obtain the following Lagrange function and dual functional:

$$L^c(x, v, \lambda) = \sum_{n=1}^{N} \Big[\sum_{i=1}^{I} \pi_n v_n^i + \lambda_n(d_n - \sum_{i=1}^{I} B_n^i x_n^i)\Big] \quad \text{and}$$

$$D^c(\lambda) = \inf\{L^c(x, v, \lambda) : y \in \mathcal{X}^c\},$$

respectively. Here the set of feasible solutions $\mathcal{X}^c = \times_{i=1}^{I} \mathcal{X}_i^c$ decomposes into I components, defined in the following way:

$$\mathcal{X}_i^c = \big\{y^i = (x^i, v^i) \in \mathbb{R}^{N(r+1)} \; : \; x_n^i \in X_n^i, \, f_n^i(x_n^i) \leq v_n^i, n = 1, 2, \ldots, N, \\ A_{n,n}^i(x_n^i) + A_{n,a(n)}^i x_{a(n)}^i \geq c_n^i, n = 1, 2, \ldots, N\big\}$$

The dual problem reads

$$\sup\left\{D^c(\lambda) : \lambda \in \mathbb{R}_+^{KN}\right\}. \tag{39}$$

The dual functional decomposes into I subproblems associated with each location. For each $i = 1, \ldots, I$ we introduce the functions

$$\hat{F}^i(y^i) := \sum_{n=1}^{N} \pi_n v_n^i.$$

We can now formulate a convex optimization problem that will be the dual equivalent to the Lagrangian relaxation of the coupling constraints. Let us consider the following problem:

$$\inf \sum_{i=1}^{I} \left[\hat{F}^i_{\mathcal{X}_i^c}\right]^{**}(y^i) \quad \text{subject to (38).} \tag{40}$$

Proposition 5.1 *The functional D^c is also the dual functional of problem (40). Moreover, assume that there is a feasible solution of the problem (35)-(38) lying in the relative interior of $[\mathcal{X}^c]^{**}$. Then D^c attains its maximum, which is equal to the infimal value of (40).*

Proof: The proof follows the same lines of arguments as those of the Propositions 3.4 and 3.5. □

The proposition shows that the Lagrange relaxation of the coupling constraints is equivalent to the convexification of the objective function and of the feasible set separately for each geographical location.

6 Geographical decomposition versus scenario and nodal decomposition

We shall derive necessary and sufficient conditions for comparing the geographical decomposition with the two other approaches. For this purpose we need a measure of stability of a problem with respect to Lagrange relaxations when its feasible set (and possibly also its objective function) is nonconvex. Given a problem $\inf f_{\mathcal{A}\cap\mathcal{B}}$, we can evaluate the change of the optimal value of the Lagrangian dual problems when the constraints defining the set $\mathcal{B}$ are relaxed, i.e., $\inf[(f_{\mathcal{A}})^{**} + \delta_{\mathcal{B}^{**}}]$, compared to the optimal value of the dual problem with no relaxation $\inf[f_{\mathcal{A}\cap\mathcal{B}}]^{**}$. It provides us also with some measure of effectiveness of the particular relaxation.

Definition 6.1 *The measure of sensitivity of the problem* $\inf f_{\mathcal{A}\cap\mathcal{B}}$ *with respect to relaxation of the constraint set $\mathcal{B}$ is given by*

$$\rho(f, \mathcal{A}; \mathcal{B}) = \inf[f_{\mathcal{A}\cap\mathcal{B}}]^{**} - \inf[(f_{\mathcal{A}})^{**} + \delta_{\mathcal{B}^{**}}].$$

Obviously, problems where f is a convex function and $\mathcal{A}$ and $\mathcal{B}$ are closed convex sets satisfying a constraint qualification, are insensitive to relaxations by virtue of the strong duality theorem. One can easily see that in this case also $\rho(f, \mathcal{A}; \mathcal{B}) = 0$.

In [14] the notion of "convexity with respect to a set" is introduced for a similar purpose. There, a set $\mathcal{A}$ is called *$\mathcal{B}$-convex* if $(\mathcal{A} \cap \mathcal{B})^{**} = \mathcal{A}^{**} \cap \mathcal{B}^{**}$. Note that if the set $\mathcal{A}$ is $\mathcal{B}$-convex, then also the set $\mathcal{B}$ is $\mathcal{A}$-convex. Let us observe also, that if the set $\mathcal{A}$ is $\mathcal{B}$-convex, then $\rho(f, \mathcal{A}; \mathcal{B}) = 0$ for all linear functions f.

In order to compare the relaxations, we shall need a dual formulation of the model, where the variables are split for all scenarios, and the nonanticipativity is formulated as a system of equality constraints. We introduce the set M of points in $\mathbb{R}^{TS(r+I)}$ which satisfy the coupling constraints, and the set $\mathcal{N}$ of points in $\mathbb{R}^{TS(r+I)}$ which satisfy the nonanticipativity constraints. We define

$$M_s = \Big\{ y = (x, v) \in \mathbb{R}^{TS(r+I)} : \sum_{i=1}^{I} B^i_{s,t} x^i_{s,t} \geq d^s_t, \quad t = 1, 2, \ldots, T \Big\},$$

$$\mathcal{N}^i = \Big\{ y = (x, v) \in \mathbb{R}^{TS(r+I)} : y^i_{s,t} = y^i_{\alpha(s,t),t} \quad t = 1, 2, \ldots, T, s = 1, 2, \ldots, S \Big\}.$$

Then, we can represent the sets $M = \cap_{s=1}^{S} M_s$ and $\mathcal{N} = \cap_{i=1}^{I} \mathcal{N}^i$. For each scenario $s = 1, 2, \ldots, S$ and each location $i = 1, 2, \ldots, I$ we define the cylindrical sets:

$$\begin{aligned} \hat{\Gamma}^i_s = \{ y = (x, v) \in \mathbb{R}^{TS(r+I)} : & \, x^i_{s,t} \in X^i_{s,t}, f^i_{s,t}(x^i_{s,t}) \leq v^i_{s,t} \\ & A^{s,i}_{t,t}(x^i_{s,t}) + A^{s,i}_{t,t-1} x^i_{s,t-1} \geq c^i_{s,t}, t = 1, 2, \ldots, T \} \end{aligned}$$

Furthermore, for each $i = 1, 2, \ldots, I$ and $s = 1, 2, \ldots, S$ we set

$$\Gamma^i = \bigcap_{s=1}^{S} \hat{\Gamma}^i_s, \quad , \Gamma_s = \bigcap_{i=1}^{I} \hat{\Gamma}^i_s \quad \text{and} \quad \Gamma = \bigcap_{i=1}^{I} \Gamma^i.$$

We denote the objective function of the multistage problem by $F : \mathbb{R}^{TS(r+I)} \to \mathbb{R}$, i.e.,

$$F(y) = \sum_{s=1}^{S} \sum_{i=1}^{I} \sum_{t=1}^{T} p_s v^i_{s,t}$$

Theorem 6.2 *Assume that the convex hull of the feasible set for the problem (35)-(38) has nonempty relative interior. The geographical decomposition provides a better bound for the optimal value than the scenario decomposition, i.e.,*

$$\sup_{\mu} D^{na}(\mu) \leq \sup_{\lambda} D^c(\lambda)$$

if and only if the following inequality holds true:

$$\rho(F, \Gamma \cap M, \mathcal{N}) - \rho(F, \Gamma \cap \mathcal{N}, M) \geq 0.$$

Proof: According to Proposition 5.1 by relaxing the coupling constraints one obtains a lower bound $\hat{D}^c$ of the objective function such that:

$$\hat{D}^c = \sup_{\lambda} D^c(\lambda) = \inf \Big\{ \sum_{i=1}^{I} \left[\hat{F}^i \chi_{\mathcal{X}^c_i} \right]^{**} : \text{ subject to } \quad (38) \Big\}. \tag{41}$$

According to Proposition 3.4 by relaxing the nonanticipativity constraints one obtains another lower bound $\hat{D}^{na}$ of the objective function such that:

$$\hat{D}^{na} = \sup_{\mu} D^{na}(\mu) = \inf\Big\{ \sum_{s=1}^{S} \Big[F^{s}_{\mathcal{X}^{na}_{s}} \Big]^{**} \ :\ \text{subject to (16)} \Big\}. \tag{42}$$

We shall transform the latter problem by using the definition of the sensitivity measure. Let V denote the optimal value of the multistage problem before relaxation. Then

$$\begin{aligned}
\hat{D}^{na} &= \inf\Big\{ \sum_{s=1}^{S} \Big[F^{s}_{\mathcal{X}^{na}_{s}} \Big]^{**} + \delta_{\mathcal{N}} \Big\} \\
&= V - \rho(F, \Gamma \cap M, \mathcal{N}) \\
&= \inf\Big\{ \sum_{s=1}^{S} \Big[\sum_{i=1}^{I}\sum_{t=1}^{T} p_s v^i_{s,t} + \delta_{(\Gamma_s)^{**}} \Big] + \delta_{\mathcal{N}\cap M} \Big\} \\
&\quad + \rho(F, \Gamma, \mathcal{N}\cap M) - \rho(F, \Gamma \cap M, \mathcal{N}) \\
&= \inf\Big\{ \sum_{s=1}^{S}\sum_{i=1}^{I}\sum_{t=1}^{T} p_s v^i_{s,t} + \sum_{s=1}^{S} \delta_{(\Gamma_s)^{**}} + \delta_M + \delta_{\mathcal{N}} \Big\} \\
&\quad + \rho(F, \Gamma, \mathcal{N}\cap M) - \rho(F, \Gamma \cap M, \mathcal{N})
\end{aligned}$$

Let us observe that

$$\sum_{s=1}^{S} \delta_{(\Gamma_s)^{**}} = \sum_{s=1}^{S}\sum_{i=1}^{I} \delta_{(\hat{\Gamma}^i_s)^{**}} = \sum_{i=1}^{I} \delta_{(\Gamma^i)^{**}} \tag{43}$$

For this purpose, we note that the sum on the left hand side $\sum\limits_{s=1}^{S} \delta_{(\Gamma_s)^{**}}(y) = 0$ for the point $y \in \mathbb{R}^{ST(r+I)}$ if and only if $y \in \Gamma^{**}_s$ for all s. Furthermore, Γ_s are cylindrical sets of the form $\mathbb{R}^{T(r+I)(s-1)} \times C_s \times \mathbb{R}^{T(r+I)(S-s)}$ for some set C_s. Consequently, $\Gamma^{**}_s = \mathbb{R}^{T(r+I)(s-1)} \times C^{**}_s \times \mathbb{R}^{T(r+I)(S-s)}$. This implies that

$$\Gamma^{**} = \Big(\bigcap_{s=1}^{S} \Gamma_s \Big)^{**} = \Big(\mathop{\times}_{s=1}^{S} C_s \Big)^{**} = \mathop{\times}_{s=1}^{S} C^{**}_s = \bigcap_{s=1}^{S} \Gamma^{**}_s$$

using the separability of the sets C_s. Otherwise the sum $\sum\limits_{s=1}^{S} \delta_{(\Gamma_s)^{**}}(y) = \infty$. By the same arguments

$$\Gamma^{**} = \sum_{s=1}^{S}\sum_{i=1}^{I} \delta_{(\hat{\Gamma}^i_s)^{**}} \quad \text{and} \quad \Gamma^{**} = \sum_{i=1}^{I} \delta_{(\Gamma^i)^{**}}$$

which proves the equality (43). Therefore, we can continue the transformation of the dual equivalent problem as follows:

$$\begin{aligned}\hat{D}^{na} &= \inf\Big\{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v_{s,t}^i + \sum_{i=1}^{I}\delta_{(\Gamma^i)^{**}} + \delta_M + \delta_{\mathcal{N}}\Big\} \\ &+ \rho(F,\Gamma,\mathcal{N}\cap M) - \rho(F,\Gamma\cap M,\mathcal{N}) \\ &= \inf\Big\{\sum_{i=1}^{I}\Big[\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v_{s,t}^i + \delta_{(\Gamma^i)^{**}} + \delta_{\mathcal{N}^i}\Big] + \delta_M\Big\} \\ &+ \rho(F,\Gamma,\mathcal{N}\cap M) - \rho(F,\Gamma\cap M,\mathcal{N})\end{aligned} \tag{44}$$

On the other hand for each point $y = (x, v) \in \mathbb{R}^{ST(r+I)} : y \in (\Gamma^i \cap \mathcal{N}_i)^{**}$, we can associate a point $\hat{y} = (\hat{x}, \hat{v}) \in \mathbb{R}^{N(r+1)}$ such that

$$\sum_{i=1}^{I}\Big[\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v_{s,t}^i + \delta_{(\Gamma^i\cap\mathcal{N}_i)^{**}}\Big](y) = \sum_{i=1}^{I}\Big[\sum_{n=1}^{N}\pi_n v_n^i + \delta_{\mathcal{X}_i^{c**}}(\hat{y}^i)\Big] = \sum_{i=1}^{I}\hat{F}^i{}_{\mathcal{X}_i^c}(\hat{y}^i).$$

with $n = \beta(s,t)$ because nonanticipativity is satisfied. Using again the measure of sensitivity, we obtain

$$\begin{aligned}\hat{D}^{c} &= \inf\Big\{\sum_{i=1}^{I}\Big[\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v_{s,t}^i + \delta_{(\Gamma^i\cap\mathcal{N}_i)^{**}}\Big] + \delta_M\Big\} \\ &= V - \rho(F,\Gamma\cap\mathcal{N},M) \\ &= \inf\Big\{\sum_{i=1}^{I}\Big[\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v_{s,t}^i + \delta_{(\Gamma^i)^{**}} + \delta_{\mathcal{N}^i}\Big] + \delta_M\Big\} \\ &+ \rho(F,\Gamma,\mathcal{N}\cap M) - \rho(F,\Gamma\cap\mathcal{N},M)\end{aligned} \tag{45}$$

Putting the two equalities (44) and (45) together yields

$$\hat{D}^{na} = \hat{D}^{c} + \rho(F,\Gamma\cap\mathcal{N},M) - \rho(F,\Gamma\cap M,\mathcal{N})$$

This completes the proof. □

The next corollary is an immediate consequence of the theorem and the definition of "set-convexity".

Corollary 6.3 *Assume that the convex hull of the feasible set for the problem (35)-(38) has nonempty relative interior.*

(1) If the set $\Gamma\cap\mathcal{N}$ is M-convex, then the geographical decomposition provides a better bound for the optimal value than the scenario decomposition, i.e.,

$$\sup_{\lambda} D^c(\lambda) \geq \sup_{\mu} D^{na}(\mu)$$

(2) *If the set $\Gamma \cap M$ is $\mathcal{N}$-convex, then the scenario decomposition provides a better bound for the optimal value than the geographical decomposition, i.e.,*

$$\sup_{\lambda} D^{c}(\lambda) \leq \sup_{\mu} D^{na}(\mu)$$

Now, we turn to the relations between the geographical and the nodal decomposition. We shall use the relations and properties of the following sets:

$$C_n = \Big\{ y = (x, v) \in \mathbb{R}^{N(r+I)} : \sum_{i=1}^{I} B_n^i x_n^i \geq d_n \Big\},$$

$$G^i = \Big\{ y = (x, v) \in \mathbb{R}^{N(r+I)} : A_{n,n}^i(x_n^i) + A_{n,a(n)}^i x_{a(n)}^i \geq c_n^i \quad n = 1, 2, \ldots, N \Big\}$$

Then the sets $C = \cap_{n=1}^{N} C_n$ and $G = \cap_{i=1}^{I} G^i$. For each node $n = 1, 2, \ldots, N$ and each location $i = 1, 2, \ldots, I$, we define the sets:

$$\hat{Y}_s^i = \Big\{ y = (x, v) \in \mathbb{R}^{N(r+1)} : x_n^i \in X_n^i,\ f_n^i(x_n^i) \leq v_n^i \Big\}$$

Furthermore, for each $i = 1, 2, \ldots, I$ and $n = 1, 2, \ldots, N$ we set

$$Y^i = \bigcap_{n=1}^{N} \hat{Y}_n^i, \quad , Y_n = \bigcap_{i=1}^{I} \hat{Y}_n^i, \quad \text{and} \quad Y = \bigcap_{i=1}^{I} Y^i$$

Again, we denote the objective function of the multistage problem by $F : \mathbb{R}^{N(r+I)} \to \mathbb{R}$:

$$F(y) = \sum_{i=1}^{I} \sum_{n=1}^{N} \pi_n v_n^i$$

Theorem 6.4 *Assume that the convex hull of the feasible set for the problem (35)-(38) has nonempty relative interior. The geographical decomposition provides a better bound for the optimal value than the nodal decomposition, i.e.,*

$$\sup_{\nu} D^{d}(\nu) \leq \sup_{\lambda} D^{c}(\lambda)$$

if and only if the following inequality holds true:

$$\rho(F, Y \cap C, G) - \rho(F, Y \cap G, C) \geq 0.$$

Proof: According to Proposition 5.1 by relaxing the coupling constraints one obtains a lower bound of the objective function $\hat{D}^c$ such that:

$$\hat{D}^c = \sup_{\lambda} D^c(\lambda) = \inf \Big\{ \sum_{i=1}^{I} \Big[\hat{F}^i{}_{\mathcal{X}_i^c} \Big]^{**} (y^i) : \quad (38) \Big\}.$$

According to Proposition 3.5 the dynamic relaxation provides a lower bound of the objective function D^d such that:

$$\hat{D}^d = \sup_{\nu} D^d(\nu) = \inf \Big\{ \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}_n^d} \Big]^{**} (y_n) : \quad \text{subject to (6)} \Big\}.$$

We shall transform the latter problem by using the definition of the sensitivity measure. Observe that $\mathcal{X}_n^d = Y_n \cap C_n$ and $\mathcal{X}^d = Y \cap C$. Let V denote the optimal value of the multistage problem before relaxation. Then

$$\begin{aligned}
\hat{D}^d &= \inf\Big\{ \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}_n^d}\Big]^{**} + \delta_G \Big\} \\
&= \inf\Big\{ \sum_{n=1}^{N} \Big[\sum_{i=1}^{I} \pi_n v_n^i + \delta_{Y_n \cap C_n}\Big]^{**} + \delta_G \Big\} \\
&= \inf\Big\{ \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i + \sum_{n=1}^{N} \delta_{(Y_n \cap C_n)^{**}} + \delta_G \Big\} \\
&= V - \rho(F, Y \cap C, G) \\
&= \inf\Big\{ \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i + \sum_{n=1}^{N} \delta_{(Y_n)^{**}} + \delta_C + \delta_G \Big\} \\
&\quad + \rho(F, Y, G \cap C) - \rho(F, Y \cap C, G)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\hat{D}^c &= \inf\Big\{ \sum_{i=1}^{I} \Big[\hat{F}^i{}_{\mathcal{X}_i^c}\Big]^{**} + \delta_C \Big\} \\
&= \inf\Big\{ \sum_{i=1}^{I} \Big[\sum_{n=1}^{N} \pi_n v_n^i + \delta_{Y^i \cap G^i}\Big]^{**} + \delta_C \Big\} \\
&= \inf\Big\{ \sum_{i=1}^{I} \sum_{n=1}^{N} \pi_n v_n^i + \sum_{i=1}^{I} \delta_{(Y^i \cap G^i)^{**}} + \delta_C \Big\} \\
&= V - \rho(F, Y \cap G, C) \\
&= \inf\Big\{ \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i + \sum_{i=1}^{I} \delta_{(Y^i)^{**}} + \delta_C + \delta_G \Big\} \\
&\quad + \rho(F, Y, G \cap C) - \rho(F, Y \cap G, C)
\end{aligned}$$

By the same arguments as in the previous proof, we observe that

$$\sum_{n=1}^{N} \delta_{(Y_n)^{**}} = \sum_{n=1}^{N} \sum_{i=1}^{I} \delta_{(\hat{Y}_n^i)^{**}} = \sum_{i=1}^{I} \delta_{(Y^i)^{**}} \tag{46}$$

Therefore, we can write the equality:

$$
\begin{aligned}
\hat{D}^d &= \inf \Big\{ \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i + \sum_{n=1}^{N} \delta_{(Y_n)^{**}} + \delta_C + \delta_G \Big\} \\
&\quad + \rho(F, Y, G \cap C) - \rho(F, Y \cap C, G) \\
&= \inf \Big\{ \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i + \sum_{i=1}^{I} \delta_{(Y^i)^{**}} + \delta_C + \delta_G \Big\} \\
&\quad + \rho(F, Y, G \cap C) - \rho(F, Y \cap C, G) \\
&= \hat{D}^c + \rho(F, Y, G \cap C) - \rho(F, Y \cap C, G) - \rho(F, Y, G \cap C) + \rho(F, Y \cap G, C) \\
&= \hat{D}^c - \rho(F, Y \cap C, G) + \rho(F, Y \cap G, C)
\end{aligned}
$$

This proves the assertion. □

Analogously to the comparison of scenario and geographical decomposition we obtain the following corollary.

Corollary 6.5 *Assume that the convex hull of the feasible set for the problem (35)-(38) has nonempty relative interior.*

(1) If the set $Y \cap G$ is C-convex, then the geographical decomposition provides a better bound for the optimal value than the nodal decomposition, i.e.,

$$\sup_{\lambda} D^c(\lambda) \geq \sup_{\nu} D^d(\nu)$$

(2) If the set $Y \cap C$ is G-convex, then the nodal decomposition provides a better bound for the optimal value than the geographical decomposition, i.e.,

$$\sup_{\nu} D^d(\nu) \geq \sup_{\lambda} D^c(\lambda)$$

Some sufficient conditions for set-convexity are provided in [14].

References

[1] J.-P. Aubin, I. Ekeland: Estimates of the duality gap in nonconvex optimization, *Mathematics of Operations Research* 1 (1976), 225–245.

[2] L. Bacaud, C. Lemaréchal, A. Renaud, C. Sagastizábal: Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners, *Computational Optimization and Applications* 20 (2001), 227–244.

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

[4] J.R. Birge: Decomposition and partitioning methods for multi-stage stochastic linear programs, *Operations Resarch* 33 (1985), 989–1007.

[5] J.R. Birge and M.A.H. Dempster: Stochastic programming approaches to stochastic scheduling, *Journal of Global Optimization* 9 (1996), 417–451.

[6] J.R. Birge and F. Louveaux: Introduction to Stochastic Programming, Springer Series in Operation Research, Springer Verlag, 1997.

[7] C. Carøe and R. Schultz: Dual decomposition in stochastic integer programming, *Operations Research Letters* 24 (1999), 37–45.

[8] C. Carøe and J. Tind: L-shaped decomposition of two-stage stochastic programs with integer recourse, *Mathematical Programming* 83 (1998), 451–464.

[9] P. Carpentier, G. Cohen, J.-C. Culioli and A. Renaud: Stochastic optimization of unit commitment: a new decomposition framework, *IEEE Transactions on Power Systems* 11 (1996), 1067–1073.

[10] D. Dentcheva and W. Römisch: Optimal power generation under uncertainty via stochastic programming, in: *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer-Verlag, Berlin 1998, 22–56.

[11] J.E. Falk: Lagrange multipliers and nonconvex programs, *SIAM Journal on Control* 7 (1969), 534–545.

[12] S. Feltenmark and K.C. Kiwiel: Dual applications of proximal bundle methods, incluing lagrangian relaxation of nonconvex problems, *SIAM Journal on Optimization* 10 (2000), 697–721.

[13] A.M. Geoffrion: Lagrangian relaxation and its uses in integer programming, *Mathematical Programming Study* 2 (1974), 82–114.

[14] M. Guignard and S. Kim: Lagrangean decomposition: a model yielding stronger Lagrangean bounds, *Mathematical Programming* 39 (1987), 215–228.

[15] J.-B. Hiriart-Urruty and C. Lemaréchal: *Convex Analysis and Minimization Algorithms I and II*, Springer-Verlag, Berlin 1993.

[16] C. Lemaréchal and A. Renaud: A geometric study of duality gaps, with applications, *Mathematical Programming* 90 (2001), 399–427.

[17] M.P. Nowak and W. Römisch: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-termal system under uncertainty. *Annals of Operations Research* 100 (2000), 251–272.

[18] J.M. Mulvey and A. Ruszczyński: A new scenario decomposition method for large-scale stochastic optimization, *Operations Research* 43 (1995), 477–490.

[19] R.T. Rockafellar and R.J-B Wets: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints, *SIAM Journal on Control and Optimization* 16 (1978), 16–36.

[20] R.T. Rockafellar and R.J-B Wets: Scenario and policy aggregation in optimization under uncertainty, *Mathematics of Operations Research* 16 (1991), 119–147.

[21] W. Römisch and R. Schultz: Multistage stochastic integer programs: An introduction, in: *Online Optimization of Large Scale Systems* (M. Grötschel, S.O. Krumke, J. Rambau eds.), Springer-Verlag, Berlin 2001, 579–598.

[22] C.H. Rosa and A. Ruszczyński: On augmented Lagrangian decomposition methods for multistage stochastic programs, *Annals of Operations Research* 64 (1996), 289–309.

[23] A. Ruszczyński: A regularized decomposition method for minimizing a sum of polyhedral functions, *Mathematical Programming* 35 (1986), 309-333.

[24] A. Ruszczyński: Parallel decomposition of multistage stochastic programming problems, *Mathematical Programming* 58 (1993), 201-228.

[25] A. Ruszczyński: On convergence of an augmented Lagrangian decomposition method for sparse convex optimization, *Mathematics of Operations Research* 20 (1995) 634–656.

[26] A. Ruszczyński: Decomposition methods in stochastic programming, *Mathematical Programming* 79 (1997), 333–353.

[27] S. Sen, J.L. Higle, and J.R. Birge: Duality gaps in stochastic integer programming, *Journal of Global Optimization* 18 (2000), 189–194.

[28] S. Takriti and J.R. Birge: Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs, *Operations Research* 48 (2000), 91–98.

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