Mathematical Programming manuscript No.
(will be inserted by the editor)

Darinka Dentcheva · Werner Römisch

Duality gaps in nonconvex stochastic optimization

Received: date / Revised version: date

Abstract. We consider multistage stochastic optimization models containing nonconvex constraints, e.g., due to logical or integrality requirements. We study three variants of Lagrangian relaxations and of the corresponding decomposition schemes, namely, *scenario*, *nodal* and *geographical* decomposition. Based on convex equivalents for the Lagrangian duals, we compare the duality gaps for these decomposition schemes. The first main result states that scenario decomposition provides a smaller or equal duality gap than nodal decomposition. The second group of results concerns large stochastic optimization models with loosely coupled components. The results provide conditions implying relations between the duality gaps of geographical decomposition and the duality gaps for scenario and nodal decomposition, respectively.

Key words. Stochastic programming - Nonconvex optimization - Lagrangian relaxation - Duality gap - Decomposition - Integer programming

1. Introduction

Stochastic dynamic programs arise as optimization models of systems driven by some discrete-time stochastic process $\{\xi_t : t = 1, 2, \ldots\}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some finite-dimensional Euclidean space. Our modeling time horizon includes T time periods, and we make sequential decisions $x_t \in \mathbb{R}^r$ at every time interval $t = 1, 2, \ldots, T$ on the basis of 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 $\mathcal{F}_t \subseteq \mathcal{F}$, which 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 $\mathcal{F}_T = \mathcal{F}$. Nonanticipativity can be briefly expressed by the equality constraints

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

The dynamics of the system is described by the inequalities

$$A_t^t(\zeta_t)x_t + A_{t-1}^t(\zeta_t)x_{t-1} \geq c_t(\zeta_t), \ t = 1, \ldots, T, \ \mathbb{P} - \text{a.s.}, \tag{2}$$

Darinka Dentcheva: Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, U.S.A. e-mail: ddentche@stevens-tech.edu

Werner Römisch: Humboldt University Berlin, Institute of Mathematics, D-10099 Berlin, Germany e-mail: romisch@mathematik.hu-berlin.de

Mathematics Subject Classification (1991): 90C15

where $A_\tau^t(\zeta_t)$, $\tau = t-1, t$, and $c_t(\zeta_t)$ are matrices and vectors, respectively, of appropriate dimension and depending measurably on ζ_t, $t = 1, \ldots, T$.
The remaining constraints associated with each time period are expressed in the form

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

where it is assumed that the sets $X_t(\zeta_t)$, $t = 1, \ldots, T$, are measurable w.r.t. $\omega \in \Omega$ and compact $\mathbb{P}$-a.s. We *do not assume* that these sets are convex. Typical examples are mixed-integer stochastic programs, where some integrality requirements are incorporated into the definition of the sets $X_t(\zeta_t)$. The presence of nonconvex functions in the description of these sets is possible as well.

In this paper, we consider the following stochastic optimization model:

$$\text{Minimize} \quad \mathbb{E}\Big[\sum_{t=1}^{T} f_t(\zeta_t, x_t)\Big] \quad \text{subject to the constraints } (1), (2) \text{ and } (3). \tag{4}$$

Here, we assume that all functions $f_t(\zeta_t(\cdot), \cdot)$, $t = 1, \ldots, T$, are finite normal integrands on $\Omega \times \mathbb{R}^r$. Then the problem (1) – (4) is well defined.

In order to solve such a model, the stochastic process $\{\xi_t\}_{t=1}^T$ is approximated by a process having a finite number of scenarios. In this way a deterministic optimization problem is generated, which replaces the stochastic problem. We shall refer to the scenario-based deterministic problem as a multistage problem. The approximation typically leads to a model of very large dimensions. The large size, the combination of different types of constraints, and the nonconvexity (e.g., integrality requirements) turn the multistage problem into a theoretical and numerical challenge. In recent years various decomposition methods for solving multistage problems were suggested (see [6, 28] for an overview). These include the primal nested Benders decomposition and the regularized decomposition methods (see [4, 26]). Dual decomposition techniques associate Lagrange multipliers with some group of the constraints and make use of the solution of some "dual" problem. Most of the dual approaches such as progressive hedging [22], and the augmented Lagrangian decomposition suggested in [27, 19], relax the nonanticipativity constraints (1). Some dual problems are investigated in [21], where Lagrange multipliers are associated with some inequality constraints describing the sets $X_t(\zeta_t)$. Nodal decomposition is a technique that associates Lagrange multipliers with dynamics constraints (2) (see [25]). Another decomposition approach was suggested for the unit commitment problem in power generation under uncertain load in [10, 18] (see also [2, 9, 32] for related work). The decomposition exploits the specific structure of many large systems that each component has a separate model. The joint operation is then coordinated by coupling constraints. Therefore, to a large extent the model of the whole system has a separable structure. We shall pay special attention to this decomposition approach, which we call geographical decomposition. For further advances in duality in stochastic programming problems we refer the interested reader to [20]. Decomposition methods for stochastic programming models with integrality constraints were suggested in a few papers. Primal methods based on Benders and test sets for the two-stage situation are presented in [8] and [15]. Dual schemes for the multistage case are discussed in [7] and [24].

Estimates of duality gaps for mixed-integer programs, and comparisons of some relaxation techniques are found in [1,3,13,12,14,17]. The results and techniques of these papers require separation of constraints into two or more groups, while the decomposition methods in stochastic programming lead to subproblems where certain sets of constraints are copied, and others are separated. Moreover, the process of duplicating constraints within each decomposition scheme is complicated. The specific structure of the multistage models requires a special approach. Some relevant research on duality gaps in stochastic programming is contained in [5,30,31]. The authors of [5,31] try to establish quantitative estimates for the duality gap arising when the scenario decomposition approach is applied to solving mixed-integer multistage stochastic programs. The relative duality gap per scenario term is estimated in [31]. In [5] the authors state sufficient conditions that lead to the vanishing of the duality gap in the scenario decomposition if the number of scenarios tends to infinity. However, an example with non-vanishing duality gap while increasing the number of scenarios is constructed in [30].

Our analysis will focus on comparing the duality gaps for the three established dual decomposition techniques in stochastic programming: scenario, nodal and geographical decomposition. We derive convex programs whose duals are equivalent to the dual problems associated with three different Lagrangian relaxations. This analysis for the scenario and nodal decomposition is presented in Section 3. The main result presented in Section 4 shows that the scenario decomposition provides a better lower bound (smaller or equal duality gap) than the nodal decomposition. The precise description and the analysis of the geographical decomposition are presented in Section 5. The geographical decomposition is compared to the nodal and scenario decomposition in Section 6. For a general optimization problem we introduce a measure of sensitivity to relaxation of constraints. This sensitivity measure is used to characterize the relative effectiveness of the decomposition approaches. Finally, sufficient conditions implying how the duality gap for geographical decomposition relates to the corresponding gaps for scenario and nodal decomposition are provided. The latter conditions are discussed for stochastic integer programs.

2. Formulations of multistage problems

We consider the multistage problem based on an approximation of the stochastic process ξ by finitely many scenarios. We may think of it as a special case of the stochastic program in which the set Ω is finite, i.e., $\Omega = \{1, 2, \ldots, S\}$ and each scenario is represented by its index s. We denote the value of the process ξ for the scenario s at t by $\xi_{s,t}$. Correspondingly, $x_{s,t}$ will denote the value of the decision for the scenario s at t, where $s = 1, 2, \ldots, S$ and $t = 1, 2, \ldots, T$. Setting $\mathbb{P}(\{s\}) = p_s,\ s = 1, 2, \ldots, S$, with $\sum_{s=1}^{S} p_s = 1$, the nonanticipativity constraint corresponds to the following system of equations

$$x_{\sigma,t} = x_{s,t} \quad \text{whenever} \quad \zeta_{\sigma,t} = \zeta_{s,t}.$$

For $t = 1$ the latter condition reads $\quad x_{1,1} = x_{2,1} = \cdots = x_{S,1}$.
The nonanticipativity constraints require the decisions to belong to a certain subspace.

The particular algebraic representation of this subspace is inessential to our study, and, therefore, we adopt the following general form:

$$\sum_{s=1}^{S} W_{s,t} x_{s,t} = 0\,, \;\; t = 1, 2, \ldots, T, \tag{5}$$

where the $W_{s,t}$ denote matrices of appropriate dimensions. For particular forms of $W_{s,t}$ see [22] and [19].
To simplify notation we set $f_{s,t} := f(\zeta_{s,t}, \cdot)$, $A^t_{s,\tau} := A^t_\tau(\zeta_{s,t})$, $\tau = t-1, t$, and $X_{s,t} := X_t(\zeta_{s,t})$ for $s = 1, 2, \ldots, S$ and $t = 1, 2, \ldots, T$.
Then the multistage stochastic program is of the form

$$\min_{(x,v)} \; \sum_{t=1}^{T} \sum_{s=1}^{S} p_s v_{s,t} \tag{6}$$

subject to

$$A^t_{s,t} x_{s,t} + A^t_{s,t-1} x_{s,t-1} \geq c_{s,t}\,, \;\; t = 2, \ldots, T,\; s = 1, \ldots, S, \tag{7}$$

$$x_{s,t} \in X_{s,t},\; f_{s,t}(x_{s,t}) \leq v_{s,t}\,, \;\; t = 1, \ldots, T,\; s = 1, \ldots, S, \tag{8}$$

$$\sum_{s=1}^{S} W_{s,t} x_{s,t} = 0\,, \;\; t = 1, 2, \ldots, T, \tag{9}$$

where additional variables $v = (v_{s,t})$ were introduced to arrive at a model with linear objective and with all potential nonconvexities contained in (8).

A graphical representation of the relations between the atoms in the filtration leads to a tree. We can associate nodes of the tree with the realizations of the process as well as with the decisions. We enumerate the nodes setting $n = 1$ for the root node. Any other node n has a unique predecessor node $a(n)$. Every node n has a set of successors $\mathcal{S}(n)$, which is empty for the terminal nodes (leaves). We denote the number of all nodes by N. There is a unique correspondence $\beta : (s,t) \to n$ assigning a node n to a scenario s at a certain time t. Furthermore, if n is a terminal node, then 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 may associate probabilities with the nodes of the tree according to the following recursive procedure:

$$\pi_n = \begin{cases} p_{\gamma(n)} & \text{for all } n \text{ such that } \mathcal{S}(n) = \emptyset \\ \sum\limits_{m \in \mathcal{S}(n)} \pi_m & \text{for all other nodes } n. \end{cases} \tag{10}$$

By setting $x_n = x_{s,t}$, $v_n = v_{s,t}$, $X_n = X_{s,t}$, $f_n = f_{s,t}$, $c_n = c_{s,t}$, $A_{n,n} = A^t_{s,t}$ and $A_{n,a(n)} = A^t_{s,t-1}$, respectively, if $n = \beta(s,t)$, and by using the node probabilities (10), we arrive at the following scenario-tree-based formulation of the multistage problem:

$$\min_{(x,v)} \; \sum_{n=1}^{N} \pi_n v_n \tag{11}$$

subject to

$$A_{n,n} x_n + A_{n,a(n)} x_{a(n)} \geq c_n\,, \;\; n = 2, \ldots, N, \tag{12}$$

$$x_n \in X_n,\; f_n(x_n) \leq v_n\,, \;\; n = 1, 2, \ldots, N. \tag{13}$$

The latter form of the multistage problem is commonly called a primal formulation. By splitting the decision variables in (11)-(13), i.e., by setting $x_{s,t} := x_n$ for each node n at time t and each $s \in \{1, \dots, S\}$ satisfying $\beta(s,t) = n$, and by reformulating the constraints for all splitted variables and introducing the explicit nonanticipativity constraint (9), we arrive again at the scenario-based formulation (6)–(9). The constraint (9) is omitted in the model (11)-(13) 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 that lead to the decomposition of the multistage optimization model:

- Lagrange multipliers are associated with the nonanticipativity constraints (9). This relaxation is frequently called *scenario decomposition* or *scenario disaggregation*.
- Lagrange multipliers are associated with the dynamic constraints (12). This techniques is called *nodal decomposition* in [25].
- Lagrangian relaxation by decoupling system components when the multistage problem has a loosely coupled structure. This approach will be called *geographical decomposition*.

For more information about the scenario and nodal decomposition the reader is referred to [25, 22]. In the next subsections we shall review some features of the first two relaxation approaches. The third approach requires a more precise description, and we defer its discussion to Section 5.

Recall that 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:

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

Here $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We assume that f is not identical to $+\infty$ and that there exists an affine minorant of f. The latter condition implies that $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. The biconjugate function f^{**} of f is defined by

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

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

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

(cf. [16, Chapt. X, Theorem 1.3.5], [23, Theorem 11.1]). Here $\mathrm{epi} f$ refers to the epigraph of f and $\overline{\mathrm{co}}$ denotes the operation of taking the convex hull and closure.

The following fact can be proved easily and will be used repeatedly.

Lemma 1. *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ has an affine minorant. If f is a sum of a linear and of an extended real-valued nonlinear function g, i.e., $f(x) = a^T x + g(x)$ for some $a \in \mathbb{R}^n$, then $f^{**}(x) = a^T x + g^{**}(x)$.*

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 $A^{**} = \overline{\text{co}}\, A$.

Consider the optimization problem

$$\min\ f(x) \quad \text{subject to} \quad x \in X,\ a_i x \le b_i,\ i = 1, \ldots, l, \tag{14}$$

where f is a finite lower semicontinuous function on $\mathbb{R}^n$, X a subset of $\mathbb{R}^n$ and $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ for $i = 1, \ldots, l$. When relaxing the linear constraints, the dual function is denoted by D, i.e.,

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

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

$$f_X = f + \delta_X .$$

Next we recall a result on convex programs that are dually equivalent to (14). It is due to the pioneering work in [11] and also formulated in a more general version in [17].

Theorem 1. *Assume that X is compact. The function D is also the dual function associated with the following problem:*

$$\min [f_X]^{**}(x) \quad \textit{subject to} \quad a_i x \le b_i,\ i = 1, \ldots, l. \tag{15}$$

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 (15).*

3.1. Scenario decomposition

We consider the scenario-based formulation (6)–(9) of the multistage problem and associate a Lagrange multiplier $\mu \in \mathbb{R}^{(TS-N)r}$ with the nonanticipativity constraint (9). Then the Lagrange function reads

$$L^{na}(x, v, \mu) = \sum_{s=1}^{S} \sum_{t=1}^{T} \big[p_s v_{s,t} + \mu_t W_{s,t} x_{s,t} \big].$$

The dual function is given by

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

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 $s = 1, 2, \ldots, S$, to be the set

$$\mathcal{X}_s^{na} = \Big\{ y_s \in \mathbb{R}^{T(r+1)} : x_{s,t} \in X_{s,t},\ f_{s,t}(x_{s,t}) \le v_{s,t}\ ,\ t = 1, \ldots T,$$
$$A_{s,t}^t x_{s,t} + A_{s,t-1}^t x_{s,t-1} \ge c_{s,t}\ ,\ t = 2, \ldots T \Big\},$$

where $y_s = (x_{s,1}, \ldots, x_{s,T}, v_{s,1}, \ldots, v_{s,T})$. The dual problem is

$$\max \left\{ D^{na}(\mu) : \mu \in \mathbb{R}^{(ST-N)r} \right\}. \tag{16}$$

The dual function decomposes into S scenario subproblems. Each subproblem optimizes the operation of the system when the stochastic process follows a particular scenario. For each scenario s we introduce its objective

$$F^s(y_s) := \sum_{t=1}^{T} p_s v_{s,t}\ .$$

Next we show that the convex program

$$\min \sum_{s=1}^{S} \left[F^s_{\mathcal{X}_s^{na}} \right]^{**}(y_s) \quad \text{subject to} \quad (9) \tag{17}$$

is equivalent to the Lagrangian dual (16).

Proposition 1. *The function D^{na} is also the dual function of problem (17). Moreover, assume that the problem (6)-(9) has a feasible solution lying in the relative interior of the set $[\mathcal{X}^{na}]^{**}$. Then D^{na} attains its supremum, which is equal to the infimal value of (17).*

Proof: According to Theorem 1 the function D^{na} is also a dual function to the problem

$$\min \left[\Big(\sum_{s=1}^{S} F^s \Big)_{\mathcal{X}^{na}} \right]^{**}(y) \quad \text{subject to} \quad (9).$$

Due to the linearity of the functions F^s we obtain from Lemma 1 that

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

Further, due to the separability of the set $\mathcal{X}^{na}$, the biconjugate of its indicator function is of the form

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

Combining the two equations, we obtain

$$\Big[\Big(\sum_{s=1}^{S} F^s\Big)_{\mathcal{X}^{na}}\Big]^{**}(y) = \sum_{s=1}^{S} F^s(y_s) + \sum_{s=1}^{S} [\delta_{\mathcal{X}_s^{na}}]^{**}(y_s) = \sum_{s=1}^{S} \Big[(F^s + \delta_{\mathcal{X}_s^{na}})\Big]^{**}(y_s)$$
$$= \sum_{s=1}^{S} \Big[F^s_{\mathcal{X}_s^{na}}\Big]^{**}(y_s)\,.$$

Furthermore, it holds that $\operatorname{dom}\Big[\Big(\sum_{s=1}^{S} F^s\Big)_{\mathcal{X}^{na}}\Big]^{**} = \Big[\mathcal{X}^{na}\Big]^{**}$. Thus, the constraint qualification of Theorem 1 is satisfied and we may conclude that the dual function D^{na} attains its supremum, which is equal to the infimum value of (17). □

This result demonstrates that the relaxation of the nonanticipativity constraints 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 (11)–(13) of the multistage problem. The nodal decomposition associates Lagrange multipliers $\nu \in \mathbb{R}^{(N-1)m}$ with the dynamic constraints (12), where m is the dimension of c_n for each n. The set of feasible solutions $\mathcal{X}^d$ becomes

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

The Lagrange function and the dual function are given by

$$L^d(x, v, \nu) = v_1 + \sum_{n=2}^{N} \pi_n \big[v_n + \nu_n(c_n - A_{n,n}x_n - A_{n,a(n)}x_{a(n)})\big] \qquad \text{and}$$
$$D^d(\nu) = \inf\{L^d(x, v, \nu) : (x, v) \in \mathcal{X}^d\},$$

respectively, and the dual problem is

$$\sup\{D^d(\nu) : \nu \in \mathbb{R}_+^{Nm}\}. \tag{18}$$

The dual problem decomposes across nodes, i.e., into $N-1$ subproblems of dimension $r+1$. Each subproblem models the optimal operation of the system under the conditions determined by the relevant node. It is worth noting that the dimension of $\mathcal{X}^d$ is much smaller than that of $\mathcal{X}^{na}$ as $N << ST$.
We introduce the notation $\tilde{F}^n(y_n) = \pi_n v_n$, where $y_n = (x_n, v_n)$, and consider the convex optimization problem

$$\min \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}_n^d}\Big]^{**}(y_n) \quad \text{subject to} \quad (12) \tag{19}$$

and show that it is equivalent to the dual problem in the Lagrange relaxation of the dynamic constraints.

Proposition 2. *The function D^d is also the dual function of problem (19). Moreover, assume that there is a feasible solution of the problem (11) – (13) 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 (19).*

Proof: We follow the same line of arguments as in the proof of Proposition 1. According to Theorem 1 the dual function D^d is also a dual function to the problem

$$\min \Big[\Big(\sum_{n=1}^{N} \tilde{F}^n\Big)_{\mathcal{X}^d}\Big]^{**}(y) \quad \text{subject to} \quad (12).$$

Arguing analogously to the proof of Proposition 1, we obtain

$$\Big[\Big(\sum_{n=1}^{N} \tilde{F}^n\Big)_{\mathcal{X}^d}\Big]^{**}(y) = \sum_{n=1}^{N} \tilde{F}^n(y_n) + \sum_{n=1}^{N} [\delta_{\mathcal{X}_n^d}]^{**}(y_n) = \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}_n^d}\Big]^{**}(y)\,.$$

Since the constraint qualification of Theorem 1 is satisfied, the second assertion follows, too. □

The proposition demonstrates that 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 2. *Assume that the convex hull of the feasible set of the problem (11) – (13) 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:

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

For each scenario s set $x_s := (x_{s,t})$ and $v_s = (v_{s,t})$. We consider the sets

$$G_s = \big\{y_s = (x_s, v_s) \in \mathbb{R}^{T(r+1)} : A^t_{s,t}x_{s,t} + A^t_{s,t-1}x_{s,t-1} \ge c_{s,t},\ t = 2, \ldots, T\big\}$$

and

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

Furthermore, let $Y_s = \times_{t=1}^{T} Y_{s,t}$. Clearly, it holds for each scenario s that

$$\mathcal{X}_s^{na} = G_s \cap Y_s \,.$$

Furthermore, if the problem (11)-(13) 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 2 is satisfied. Moreover, we can split the variables as described shortly after (13) and obtain a feasible solution of the problem (6) – (9). This solution will be contained in the relative interior of $[\mathcal{X}^{na}]^{**}$ by construction. Therefore, the assumptions of Proposition 1 are satisfied as well. By relaxing the nonanticipativity constraints one obtains a lower bound $\hat{D}^{na}$ of the objective function, where

$$\hat{D}^{na} = \max_{\mu} D^{na}(\mu) \,. \tag{20}$$

According to Proposition 1 there exists a solution $\bar{y} = (\bar{x}, \bar{v}) \in \mathbb{R}^{TS(r+1)}$ of the convex equivalent problem (17) such that

$$\hat{D}^{na} = \inf\{\sum_{s=1}^{S} \left[F^s_{\mathcal{X}_s^{na}}\right]^{**}(y_s) : \ \text{ subject to } \ (9)\} = \sum_{s=1}^{S} \left[F^s_{\mathcal{X}_s^{na}}\right]^{**}(\bar{y}_s) \tag{21}$$

and $\bar{y}$ satisfies the nonanticipativity constraint. Using the linearity of F_s and the separability of the set $\mathcal{X}^{na}$, we continue (21) and obtain

$$\hat{D}^{na} = \sum_{s=1}^{S} \left[F^s + \delta_{G_s \cap Y_s}\right]^{**}(\bar{y}_s) = \sum_{s=1}^{S} \left[F^s(\bar{y}_s) + \delta_{[G_s \cap Y_s]^{**}}\right](\bar{y}_s). \tag{22}$$

Observe that $[G_s \cap Y_s]^{**} \subseteq (G_s)^{**} \cap (Y_s)^{**} = G_s \cap (Y_s)^{**}$ and, thus,

$$\delta_{[G_s \cap Y_s]^{**}}(\bar{y}_s) \geq \delta_{G_s \cap (Y_s)^{**}}(\bar{y}_s) = \delta_{G_s}(\bar{y}_s) + \delta_{(Y_s)^{**}}(\bar{y}_s)$$

holds. Hence, we may continue the chain (22) of transformations as follows

$$\begin{aligned} \sum_{s=1}^{S} \left[F^s(\bar{y}_s) + \delta_{[G_s \cap Y_s]^{**}}(\bar{y}_s)\right] &\geq \sum_{s=1}^{S} \left[F^s(\bar{y}_s) + \delta_{G_s}(\bar{y}_s) + \delta_{(Y_s)^{**}}(\bar{y}_s)\right] \\ &= \sum_{s=1}^{S} F^s(\bar{y}_s) + \delta_{(\underset{s=1}{\overset{S}{\times}} G_s)}(\bar{y}) + \delta_{\underset{s=1}{\overset{S}{\times}} (Y_s)^{**}}(\bar{y}). \end{aligned} \tag{23}$$

Let us define an element $\tilde{y} = (\tilde{x}, \tilde{v}) \in \mathbb{R}^{N(r+1)}$ by setting $\tilde{y}_n := \bar{y}_{s,t}$ if $n = \beta(s,t)$. This definition is non ambiguous because $\bar{y}$ satisfies the nonanticipativity constraint. We also set $\pi_n = \sum_{s:\beta(s,t)=n} p_s$ and obtain

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

The set $\times_{s=1}^{S} Y_s$ can be represented in the form $\times_{s=1}^{S} Y_s = \times_{s=1}^{S} \times_{t=1}^{T} Y_{s,t} =: \times_{s,t} Y_{s,t}$. Furthermore, due to the definition of the models (11)-(13) and (6)-(9), the sets $Y_{s,t}$

satisfy the condition $Y_{s,t} = Y_{\sigma,t}$ whenever $\zeta_{s,t} = \zeta_{\sigma,t}$ for all $s, \sigma = 1, 2, \ldots, S$ and $t = 1, 2, \ldots, T$. Therefore, we may replace $Y_{s,t}$ by $\mathcal{X}_n^d$ for the node $n = \beta(s,t)$. Since $\bar{y}_s \in [\mathcal{X}_s^{na}]^{**}$ holds, we have

$$\tilde{y}_n \in [\mathcal{X}_n]^{**} \quad \text{and} \quad \delta_G(\tilde{y}) = 0. \tag{24}$$

Thus, we obtain that

$$\begin{aligned}\sum_{s=1}^{S} F^s(\bar{y}_s) + \delta_{\times_{s,t}(Y_{s,t})^{**}}(\bar{y}) + \delta_{\times_{s=1}^{S} G_s}(\bar{y}) &= \sum_{n=1}^{N} \tilde{F}_n(\tilde{y}_n) + \delta_{\times_{n=1}^{N} \mathcal{X}_n^{**}}(\tilde{y}) + \delta_G(\tilde{y}) \\ = \sum_{n=1}^{N} \left[\tilde{F}_n + \delta_{\mathcal{X}_n}\right]^{**}(\tilde{y}_n) + \delta_G(\tilde{y}) &= \sum_{n=1}^{N} \left[\tilde{F}_n \mathcal{X}_n\right]^{**}(\tilde{y}_n) + \delta_G(\tilde{y}) \\ &\geq \min\left\{ \sum_{n=1}^{N} \left[\tilde{F}_n \mathcal{X}_n\right]^{**}(y_n) + \delta_G(y) \right\}. \end{aligned} \tag{25}$$

According to Proposition 2 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 (22), (23), and (25), we obtain

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

which is the desired inequality. □

In general, the estimate in Theorem 2 is strict, as we will see in the next example.

Example 1. We consider a two-stage model with two scenarios (three nodes in the scenario tree) in a primal formulation:

$$\begin{aligned} \min \quad & -0.2x_1 - \frac{1}{2}x_2 - x_3 \\ \text{subject to} \quad & \\ & x_1 + x_2 \leq 1.5 \qquad \text{(dynamics in the first scenario)} \\ & x_1 + x_3 \leq 1.2 \qquad \text{(dynamics in the second scenario)} \\ & x_n \in [0,2],\ n = 1,2,3; \quad x_2, x_3 \text{ integer} \qquad \text{(local constraints)} \end{aligned}$$

The feasible set of the problem is: $\Big\{x \in \mathbb{R}^3 : x_2 \in \{0,1\},\, x_3 \in \{0,1\},\, x_1 \in [0, \min\{1.5 - x_2, 1.2 - x_3\}]\Big\}$. The optimal solution of this problem is obtained for $x_1 = 0.2$, $x_2 = x_3 = 1$ and the optimal value is -1.54. The relaxation of the nonanticipativity yields the same optimal value and, therefore, no duality gap. The same optimal solution can be obtained by solving the convex problem equivalent to the dual of the scenario decomposition according to Proposition 1. The feasible set of the latter problem is $\Big\{x \in \mathbb{R}^3 : x_2 \in [0,1],\, x_3 \in [0,1],\, x_1 \in [0, \min\{1.5 - x_2, 1.2 - x_3\}]\Big\}$. In contrast, the nodal decomposition results in a smaller optimal value -1.95, which

creates a duality gap of 0.41. The convex problem equivalent to the dual of the nodal decomposition according to Proposition 2 has a larger feasible set, namely $\{x \in \mathbb{R}^3 : x_2 \in [0,2],\ x_3 \in [0,2],\ x_1 \in [0, \min\{1.5 - x_2, 1.2 - x_3\}]\}$, which creates the gap in this example.

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. We would like to associate geographical locations with the components and refer to them as locations in the following. Let us assume that the modeling system comprises I locations and $x_n^i \in \mathbb{R}^{r_i}$ is the decision at node n associated with location i, $i = 1, 2, \ldots, I$, and $\sum_{i=1}^{I} r_i = r$.

We shall deal with the following multistage stochastic optimization problem, written in a primal formulation with the transformations we have adopted in the previous section

$$\min_{(x,v)} \sum_{n=1}^{N} \sum_{i=1}^{I} \pi_n v_n^i \tag{26}$$

subject to (27)

$$A_{n,n}^i(x_n^i) + A_{n,a(n)}^i x_{a(n)}^i \geq c_n^i,\ \ n = 2, \ldots, N,\ i = 1, \ldots, I, \tag{28}$$

$$x_n^i \in X_n^i, \quad f_n^i(x_n^i) \leq v_n^i,\ \ n = 1, \ldots, N,\ i = 1, \ldots, I, \tag{29}$$

$$\sum_{i=1}^{I} B_n^i x_n^i \geq d_n,\ \ n = 1, \ldots, N. \tag{30}$$

Here B_n^i are $K \times r_i$-dimensional matrices and d_n are K-dimensional vectors. Thus, there are K constraints (30) for each sequential decision that are coupling the models of the locations. The assumption that the modeled system consists of *loosely coupled l ocations* means that $K << I$.

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 (30). Then we obtain the following Lagrangian and dual function

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

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

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

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

The dual problem reads

$$\sup \left\{ D^c(\lambda) : \lambda \in \mathbb{R}_+^{KN} \right\}. \tag{31}$$

The dual function decomposes into I subproblems representing the operation models of each location. For each $i = 1, \ldots, I$ we introduce the functions

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

The following convex optimization problem will turn out to be the dual equivalent to the Lagrangian relaxation of the coupling constraints.

$$\min \sum_{i=1}^{I} \left[\hat{F}^i_{\mathcal{X}_i^c} \right]^{**} (y^i) \quad \text{subject to} \quad (30). \tag{32}$$

Proposition 3. *The function D^c is also the dual function of problem (32). Moreover, assume that there is a feasible solution of the problem (26)-(30) lying in the relative interior of $[\mathcal{X}^c]^{**}$. Then D^c attains its maximum, which is equal to the infimal value of (32).*

Proof: The proof follows the same lines of arguments as the proofs of the Proposition 1 and Proposition 2. □

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 real-valued function on $\mathbb{R}^n$ and subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathbb{R}^n$, we consider the problem $\min f_{\mathcal{A}\cap\mathcal{B}}$. 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, i.e., $\inf [f_{\mathcal{A}\cap\mathcal{B}}]^{**}$. It provides us also with some measure of effectiveness of the particular relaxation.

Definition 1. *The measure of sensitivity of the problem $\min f_{\mathcal{A}\cap\mathcal{B}}$ with respect to the relaxation of the constraint set $\mathcal{B}$ is given by*

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

where it is assumed that at least one of the infima is finite.

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. This property is a kind of constraint qualification as we will see in the following example.

Example 2. First, let us consider convex subsets $\mathcal{A}$ and $\mathcal{B}$ of some Euclidean space. Then the $\mathcal{B}$-convexity of $\mathcal{A}$ translates into the condition

$$\mathrm{cl}(\mathcal{A} \cap \mathcal{B}) = \mathrm{cl}\mathcal{A} \cap \mathrm{cl}\mathcal{B}\,,$$

which is valid if $\mathrm{ri}\,\mathcal{A} \cap \mathrm{ri}\,\mathcal{B} \neq \emptyset$ [16, Proposition 2.1.10], where cl and ri denote closure and relative interior, respectively.
Secondly, let $\mathcal{A}$ be the set of all r-dimensional integral vectors, i.e, $\mathcal{A} := \mathbb{Z}^r$, and $\mathcal{B}$ be a polyhedral subset of $\mathbb{R}^r$. Then $\mathcal{B}$-convexity of $\mathcal{A}$ reads

$$(\mathbb{Z}^r \cap \mathcal{B})^{**} = \mathcal{B} \tag{33}$$

and means that the integer hull of $\mathcal{B}$, i.e., the closed convex hull of the set of all integral vectors contained in $\mathcal{B}$, coincides with $\mathcal{B}$. The integer hull of $\mathcal{B}$ is polyhedral if $\mathcal{B}$ is a rational polyhedron [29, Theorem 16.1]. If $\mathcal{B}$ is rational and condition (33) is satisfied, $\mathcal{B}$ is said to be an integral polyhedron. The polyhedron $\mathcal{B}$ is integral iff each face of $\mathcal{B}$ contains an integral vector. Important concepts for detecting the integrality of a polyhedron are the *total unimodularity* of matrices and the *total dual integrality* of rational linear inequalities (see [29, Part IV]). Here, we briefly recall the first concept. A matrix B is called totally unimodular if each subdeterminant of B is 0, $+1$ or -1. In particular, each entry in a totally unimodular matrix is 0, $+1$ or -1. A polyhedron $\mathcal{B} = \{x : Bx \leq b\}$ is integral if B is totally unimodular and b is integral (cf. [29, Theorem 19.1]). A complete characterization of total unimodularity is provided by the Hoffman–Kruskal theorem: If B is an integral matrix, then B is totally unimodular if and only if for each integral vector b the polyhedron $\mathcal{B} = \{x : x \geq 0,\ Bx \leq b\}$ is integral (cf. [29, Corollary 19.2a]). In particular, network matrices are totally unimodular [29, Section 19.3].

Lemma 2. *For any linear function f the sensitivity measure $\rho(f, \mathcal{A}; \mathcal{B})$ is nonnegative if it is finite, and it vanishes if the set $\mathcal{A}$ is $\mathcal{B}$-convex.*

Proof: Using Lemma 1 and the inclusion $(\mathcal{A} \cap \mathcal{B})^{**} \subseteq \mathcal{A}^{**} \cap \mathcal{B}^{**}$ we have

$$\begin{aligned} \inf [f_{\mathcal{A}\cap\mathcal{B}}]^{**} &= \inf [f + \delta_{\mathcal{A}\cap\mathcal{B}}]^{**} = \inf [f + (\delta_{\mathcal{A}\cap\mathcal{B}})^{**}] \\ &\geq \inf [f + (\delta_{\mathcal{A}})^{**} + (\delta_{\mathcal{B}})^{**}] = \inf [(f_{\mathcal{A}})^{**} + \delta_{\mathcal{B}^{**}}]. \end{aligned}$$

If $\mathcal{A}$ is $\mathcal{B}$-convex, the only inequality in this chain becomes an equality. □

In order to compare the relaxations, we shall need a dual formulation of the model (26) – (30), where the variables are split for all scenarios as described after (13), and the

nonanticipativity is formulated as a system of equality constraints. We introduce the set M of points in $\mathbb{R}^{TS(r+I)}$ that satisfy the coupling constraints, and the set $\mathcal{N}$ of points in $\mathbb{R}^{TS(r+I)}$ satisfying the nonanticipativity constraints. We define

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

$$\mathcal{N}^i = \Big\{ y = (x, v) \in \mathbb{R}^{TS(r+I)} : \sum_{s=1}^{S} \hat{W}^i_{s,t} y^i_{s,t} = 0, \quad t = 1, \ldots, T \Big\},$$

$$M = \bigcap_{s=1}^{S} M_s \quad \text{and} \quad \mathcal{N} = \bigcap_{i=1}^{I} \mathcal{N}^i.$$

Here the matrices $\hat{W}^i_{s,t}$ give the corresponding algebraic formulation of the nonanticipativity subspace for the vectors $y^i_{s,t}$. For each scenario $s = 1, \ldots, S$ and each location $i = 1, \ldots, I$ we define the cylindrical sets

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

Furthermore, for each $i = 1, \ldots, I$ and $s = 1, \ldots, S$ 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 3. *Assume that the convex hull of the feasible set for the problem (26)-(30) 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 the Propositions 3 and 1 by relaxing the coupling and nonanticipativity constraints, one obtains a lower bound $\hat{D}^c$ and $\hat{D}^{na}$, respectively, of the objective function such that

$$\hat{D}^c = \sup_{\lambda} D^c(\lambda) = \inf \Big\{ \sum_{i=1}^{I} \Big[\hat{F}^i_{\mathcal{X}^c_i} \Big]^{**} : \text{ subject to } (30) \Big\},$$

$$\hat{D}^{na} = \sup_{\mu} D^{na}(\mu) = \inf \Big\{ \sum_{s=1}^{S} \Big[F^s_{\mathcal{X}^{na}_s} \Big]^{**} : \text{ subject to } (9) \Big\}.$$

The latter problem is transformed by using the definition of the sensitivity measure. Let V denote the optimal value of the multistage problem before relaxation. Then

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

We claim that

$$\sum_{s=1}^{S}\delta_{(\Gamma_s)^{**}} = \sum_{s=1}^{S}\sum_{i=1}^{I}\delta_{(\hat{\Gamma}^i_s)^{**}} = \sum_{i=1}^{I}\delta_{(\Gamma^i)^{**}}\,. \tag{34}$$

Indeed, the sum $\sum_{s=1}^{S}\delta_{(\Gamma_s)^{**}}(y)$ on the left hand side vanishes for the point $y \in \mathbb{R}^{ST(r+I)}$ if and only if $y \in \Gamma_s^{**}$ for all s. Furthermore, the 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, we have $\Gamma_s^{**} = \mathbb{R}^{T(r+I)(s-1)}\times C_s^{**}\times\mathbb{R}^{T(r+I)(S-s)}$. Using the separability of the sets C_s this implies that

$$\Gamma^{**} = \Big(\bigcap_{s=1}^{S}\Gamma_s\Big)^{**} = \Big(\bigtimes_{s=1}^{S}C_s\Big)^{**} = \bigtimes_{s=1}^{S}C_s^{**} = \bigcap_{s=1}^{S}\Gamma_s^{**}\,.$$

We obtain by the same arguments

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

In case $\sum_{s=1}^{S}\delta_{(\Gamma_s)^{**}}(y) = \infty$, then $y \notin \Gamma_s^{**}$ for some s which implies the the equality (34). Therefore, we can continue the transformation of the dual equivalent problem as follows:

$$\begin{aligned}
\hat{D}^{na} &= \inf\Big\{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v^i_{s,t} + \sum_{i=1}^{I}\delta_{(\Gamma^i)^{**}} + \delta_M + \delta_{\mathcal{N}}\Big\} \\
&\quad +\rho(F,\Gamma,\mathcal{N}\cap M) - \rho(F,\Gamma\cap M,\mathcal{N}) \\
&= \inf\Big\{\sum_{i=1}^{I}\Big[\sum_{s=1}^{S}\sum_{t=1}^{T} p_s v^i_{s,t} + \delta_{(\Gamma^i)^{**}} + \delta_{\mathcal{N}^i}\Big] + \delta_M\Big\} \\
&\quad +\rho(F,\Gamma,\mathcal{N}\cap M) - \rho(F,\Gamma\cap M,\mathcal{N})\,.
\end{aligned} \tag{35}$$

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

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

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

$$\begin{aligned} \hat{D}^c &= \inf \Big\{ \sum_{i=1}^{I} \Big[\sum_{s=1}^{S} \sum_{t=1}^{T} p_s v^i_{s,t} + \delta_{(\Gamma^i \cap \mathcal{N}_i)^{**}} \Big] + \delta_M \Big\} \qquad (36) \\ &= V - \rho(F, \Gamma \cap \mathcal{N}, M) \\ &= \inf \Big\{ \sum_{i=1}^{I} \Big[\sum_{s=1}^{S} \sum_{t=1}^{T} p_s v^i_{s,t} + \delta_{(\Gamma^i)^{**}} + \delta_{\mathcal{N}^i} \Big] + \delta_M \Big\} \\ &\quad + \rho(F, \Gamma, \mathcal{N} \cap M) - \rho(F, \Gamma \cap \mathcal{N}, M)\,. \end{aligned}$$

Putting the two equalities (35) and (36) 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 Theorem 3 and Lemma 2.

Corollary 1. *Assume that the convex hull of the feasible set for the problem (26)-(30) 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:

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

$$G^i = \Big\{ y = (x, v) \in \mathbb{R}^{N(r+I)} : A^i_{n,n}(x^i_n) + A^i_{n,a(n)} x^i_{a(n)} \geq c^i_n\,,\ n = 2, \ldots, N \Big\},$$

$$C = \bigcap_{n=1}^{N} C_n \quad \text{and} \quad G = \bigcap_{i=1}^{I} G^i\,.$$

For each node $n = 1, \ldots, N$ and each location $i = 1, \ldots, I$, we define the sets:

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

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

Abusing notation, we denote the objective function of the multistage problem in the primal form 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 4. *Assume that the convex hull of the feasible set for the problem (26)-(30) has nonempty relative interior. The geographical decomposition provides a better bound for the optimal value than the nodal decomposition, i.e.,*

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

if and only if the following inequality holds true:

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

Proof: According to Propositions 3 and 2 by relaxing the coupling and dynamic constraints one obtains a lower bound of the objective functions $\hat{D}^c$ and D^d, respectively, such that

$$\hat{D}^c = \sup_{\lambda} D^c(\lambda) = \Big\{ \inf \sum_{i=1}^{I} \Big[\hat{F}^i{}_{\mathcal{X}_i^c} \Big]^{**}(y^i) : \quad \text{subject to} \quad (30) \Big\},$$

$$\hat{D}^d = \sup_{\nu} D^d(\nu) = \Big\{ \inf \sum_{n=1}^{N} \Big[\tilde{F}^n_{\mathcal{X}_n^d} \Big]^{**}(y_n) : \quad \text{subject to} \quad (28) \Big\}.$$

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 we have

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

On the other hand,

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

By using the same arguments as in the previous proof, we obtain that

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

Therefore, we arrive at the following chain of equalities:

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

This proves the assertion. □

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

Corollary 2. *Assume that the convex hull of the feasible set for the problem (26)-(30) 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.,

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

In general, the calculation or estimation of the measure ρ is a difficult task. A general method for evaluating the measure is beyond the scope of this paper. However, the sufficient conditions in terms of set-convexity provided in both corollaries are verifiable for some stochastic integer optimization problems.

Example 3. We consider a multistage stochastic integer program with a linear objective function and loosely coupled locations. Since there is no need to introduce the variable v, the set Γ in Theorem 3 is of the form

$$\Gamma = X \cap G = \bigcap_{t=1}^{I} \bigcap_{s=1}^{S} \left[(\bigcap_{t=1}^{T} X^i_{s,t}) \cap G^i_s \right],$$

where $G^i_s := \left\{ x \in \mathbb{R}^{TSr} \,:\, A^{t,i}_{s,t} x^i_{s,t} + A^{t,i}_{s,t-1} x^i_{s,t-1} \geq c^i_{s,t}\,,\ t = 2, \ldots, T \right\}$. Let M denote the polyhedron given by the coupling constraints and $\mathcal{N}$ be the nonanticipativity subspace of $\mathbb{R}^{TSr}$. We assume that $G \cap M \cap \mathcal{N}$ is a rational and bounded polyhedral subset of $\mathbb{R}^{TSr}$. Then we may assume that X is a bounded subset of the set $\mathbb{Z}^{TSr}$ of all integral vectors of dimension TSr having the property

$$\mathbb{Z}^{TSr} \cap G \cap M \cap \mathcal{N} = X \cap G \cap M \cap \mathcal{N}.$$

If the polyhedron $G \cap M \cap \mathcal{N}$ is integral (cf. Example 2), we have

$$\left(\mathbb{Z}^{TSr} \cap G \cap M \cap \mathcal{N}\right)^{**} = G \cap M \cap \mathcal{N}.$$

In this case, both set-convexity conditions of Corollary 1 are satisfied. Hence, if the relative interior of $G \cap M \cap \mathcal{N}$ is nonempty, the duality gaps of scenario and geographical decomposition are identical and both are smaller than for nodal decomposition (Theorem 2).

Acknowledgement: This work was supported by the Priority Programme *Online Optimization of Large Scale Systems* of the Deutsche Forschungsgemeinschaft. The authors wish to thank Andrzej Ruszczyński (Rutgers University) for helpful discussions.

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, New York 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, 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, including 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. Hemmecke, R,; Schultz, R.: Decomposition of test sets in stochastic integer programming, *Mathematical Programming* 94 (2003), 323–341.
16. J.-B. Hiriart-Urruty and C. Lemaréchal: *Convex Analysis and Minimization Algorithms I and II*, Springer, Berlin 1993.
17. C. Lemaréchal and A. Renaud: A geometric study of duality gaps, with applications, *Mathematical Programming* 90 (2001), 399–427.
18. 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.
19. J.M. Mulvey and A. Ruszczyński: A new scenario decomposition method for large-scale stochastic optimization, *Operations Research* 43 (1995), 477–490.
20. R.T. Rockafellar: Duality and optimality in multistage stochastic programming, *Annals of Operations Research* 85 (1999), 1–19.
21. 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.
22. R.T. Rockafellar and R.J-B Wets: Scenario and policy aggregation in optimization under uncertainty, *Mathematics of Operations Research* 16 (1991), 119–147.
23. R.T. Rockafellar and R.J-B Wets: *Variational Analysis*, Springer, Berlin 1998.
24. 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, Berlin 2001, 579–598.
25. C.H. Rosa and A. Ruszczyński: On augmented Lagrangian decomposition methods for multistage stochastic programs, *Annals of Operations Research* 64 (1996), 289–309.
26. A. Ruszczyński: A regularized decomposition method for minimizing a sum of polyhedral functions, *Mathematical Programming* 35 (1986), 309-333.
27. A. Ruszczyński: On convergence of an augmented Lagrangian decomposition method for sparse convex optimization, *Mathematics of Operations Research* 20 (1995) 634–656.
28. A. Ruszczyński: Decomposition methods in stochastic programming, *Mathematical Programming* 79 (1997), 333–353.
29. A. Schrijver: *Theory of Linear and Integer Programming*, Wiley, Chichester 1986.
30. S. Sen, J.L. Higle, and J.R. Birge: Duality gaps in stochastic integer programming, *Journal of Global Optimization* 18 (2000), 189–194.
31. S. Takriti and J.R. Birge: Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs, *Operations Research* 48 (2000), 91–98.
32. 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.