Dantzig-Wolfe Decomposition for Monotone Two-Stage Stochastic Mixed-Integer Programs Applied to a Power Distribution System Resilience Problem

Laurens Elderhorst
Department of Operations, Faculty of Economics and Business, University of Groningen
Nettelbosje 2, 9747AE Groningen, The Netherlands
l.elderhorst@rug.nl

Ward Romeijnders
Department of Operations, Faculty of Economics and Business, University of Groningen
Nettelbosje 2, 9747AE Groningen, The Netherlands
w.romeijnders@rug.nl

Miguel A. Lejeune
Department of Decision Sciences, School of Business, The George Washington University
2201 G Street NW, Washington, District of Columbia 20052, USA
mlejeune@gwu.edu

Payman Dehghanian
Department of Electrical and Computer Engineering, The George Washington University
800 22nd St. NW, Suite 6630, Washington, District of Columbia 20052, USA
payman@gwu.edu

Abstract

We develop a novel Dantzig-Wolfe (DW) decomposition algorithm for monotone two-stage stochastic mixed-integer programs (SMIPs). The key novelty in this algorithm is to relax the non-anticipativity constraints (NACs) in line with the monotonicity in the problem. We prove (i) that the relaxed restricted master problem (RMP) faster identifies dominated columns that cannot improve the RMP objective value, (ii) that our monotone DW algorithm does not increase the optimality gap in special cases, and (iii) that for SMIPs with pure-binary first-stage decisions, our DW algorithm can be embedded within a simple branch-and-bound procedure. Numerically, our DW algorithm may find solutions up to 92% faster than the traditional DW algorithm. We apply our new algorithm to a power distribution system resilience problem in which nodes in the electricity network may be equipped with electricity apparatus, i.e., electricity adapters, such that mobile emergency power sources (e.g., generators or batteries) can be deployed following a disaster, to temporarily provide power while the electricity grid is being repaired. We find that installing such adapters may reduce expected power loss by up to 40%, and by up to 19% compared to installing them according to the highest loads policy, which is a popular heuristic.

Key words: stochastic programming, Dantzig-Wolfe decomposition, monotonicity, distribution system resilience, extreme-weather events, mobile power sources.

September 4, 2026

1 Introduction

Many decision-making problems in sectors, such as healthcare (Saadouli et al., 2015), energy (Hafiz et al., 2019), manufacturing (Li et al., 2009), and logistics (Stewart Jr and Golden, 1983), involve complex decisions that must be made while key information is unknown. To model such problems realistically, stochastic programming (see, e.g., Birge and Louveaux (1997), and Shapiro et al. (2021)) and two-stage stochastic mixed-integer linear programs (SMIPs) offer a powerful paradigm that explicitly takes uncertainty into account.

In this paper, we focus on SMIPs that are *monotone*, that is, where the second-stage value functions are non-increasing, non-decreasing, or constant with respect to some first-stage variables. Traditional methods such as integer L-shaped cuts (Laporte and Louveaux, 1993), dual decomposition (Carøe and Schultz, 1999), Lagrangian cuts (Zou et al., 2019), and scaled-cuts (van der Laan and Romeijnders, 2024) do not exploit monotonicity in monotone SMIPs, making them inefficient for such problems. As a result, specialized algorithms have been developed, such as monotone integer L-shaped cuts for stochastic vehicle routing (Parada et al., 2024), monotonic cuts for discrete monotonic optimization (Tuy et al., 2006), and the MIDAS algorithm for stochastic dynamic programs with monotonic Bellman functions (Philpott et al., 2020). In line with this literature, we develop a novel Dantzig-Wolfe (DW) decomposition algorithm for monotone two-stage SMIPs.

The development of this algorithm is motivated by our case study on power distribution system resilience (DSR), but it can be applied to any two-stage monotone SMIP, such as capacity expansion problems, see, e.g., Park and Baldick (2015), and Basciftci et al. (2024), in which future operational costs are non-increasing in the installed capacity, or hydro-thermal scheduling, see, e.g., de Queiroz (2016), in which future thermal power supply is non-decreasing in the current water reservoir levels.

In our DSR problem, we need to prepare an electricity distribution grid for future extreme-weather events, including storms, tornados, and floods, see, e.g., Shi et al. (2022). Due to climate change, such events occur more frequently (Myhre et al., 2019), damaging distribution lines and resulting in power loss for households, hospitals, and other consumers. To reduce power loss in disconnected regions, mobile power sources (MPSs) can travel to these regions to temporarily provide power, but only if the required electricity apparatus, i.e., electricity adapters, are installed there. The goal in our DSR problem is to install a limited number of adapters at selected nodes in the distribution system to minimize the expected future power loss as a result of extreme-weather events, see Figure 1 for a schematic illustration of the DSR problem.

We consider a two-stage SMIP formulation for the DSR problem. Various heuristics for the DSR problem have been developed, see, e.g., (Li et al., 2022) and (Sun et al., 2025). However, to the best of our knowledge, we are the first to solve it to optimality.

The key insight in developing an efficient DW decomposition algorithm for the DSR problem is to exploit the monotonicity in the problem. Indeed, if the number of installed adapters increases, then the power loss does not increase. This applies to all future scenarios, and thus also to the expected power loss. For scenarios, however, in which node i remains connected to the main substation, the power loss is not only non-increasing in the first-stage decision x_i,

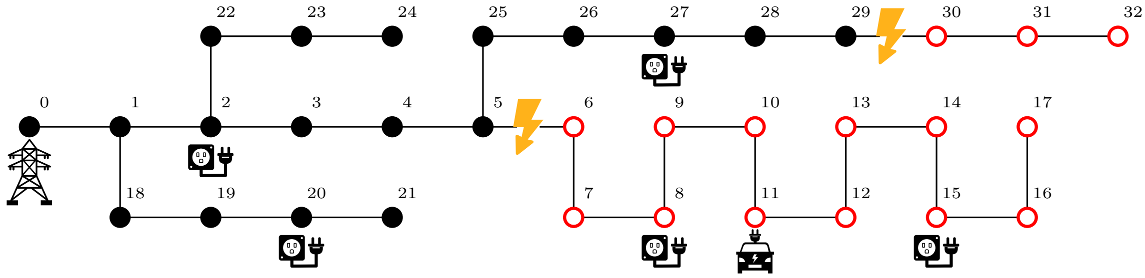


Figure 1: Illustration of how MPSs (electric car symbol) can supply power via electricity adapters (power outlet symbol). A main substation (power pole symbol) which can supply power to connected consumers, is located at node 0. Black solid nodes represent consumers that remain connected to the substation and red hollow nodes represent disconnected consumers. Yellow lightning symbols indicate damaged lines.

indicating whether to install an adapter at node i or not, but in fact constant in x_i. Our DW decomposition algorithm exploits this type of monotonicity as well. In general, there exist many forms of monotonicity, where the type of monotonicity, e.g., non-decreasing, non-increasing, or constant, may differ per component, and even per scenario, as it does in the DSR problem. To the best of our knowledge, we are the first to exploit such complex cases of monotonicity in SMIP algorithms.

The key idea in our algorithm is to relax some of the non-anticipativity constraints (NACs) in line with the type of monotonicity in the problem. We show that this leads to compelling computational benefits when these relaxations are applied in a DW decomposition algorithm. In particular, the restricted master problem (RMP) with relaxed NACs has a significantly larger feasible region than its non-relaxed counterpart, so that the DW master problem is attained in less iterations. Moreover, the relaxed RMP is able to identify faster than the traditional RMP that certain columns of the DW master problem (MP) cannot improve the objective value in the RMP. We show that this generally causes the termination criterion of DW – that no remaining columns in the MP can improve the RMP solution – to trigger sooner. Relaxing NACs is related to relaxing constraints in Lagrangian relaxations (Beltran et al., 2006). The main difference is that they enforce *additional* relaxed constraints to strengthen the Lagrangian relaxation, while we relax NACs to weaken the Lagrangian relaxation.

We note that relaxed NACs for the special case where the second-stage value functions are non-increasing in all scenarios have been observed previously in the literature (Singh et al., 2009). To the best of our knowledge, we are the first to study this relaxation in full generality—allowing the type of monotonicity to differ across scenarios and individual components, and permitting the complete removal of NACs when some second-stage value functions are constant with respect to some first-stage variables.

We apply our novel DW decomposition algorithm to the DSR problem. In fact, we implement our DW algorithm asynchronously parallel to solve large-scale instances of the DSR problem. Our numerical results indicate that the optimal SMIP solution may decrease power loss by up to 19 percent compared to the highest loads policy (HLP), which is a popular existing heuristic that installs adapters at nodes with the highest demand. Additionally, it reduces expected power loss by up to 40% compared to installing no electricity adapters. We observe that it is optimal to spread adapter installations across the electricity grid to allow for greater flexibility in operational planning. Moreover, the majority of adapters should be installed in downstream areas of the grid to minimize the expected power loss, since these regions are most likely to require backup power from MPSs.

The main contributions of this paper are summarized as follows:

- We consider two-stage monotone SMIPs with general forms of component-wise monotonicity, and show how to relax NACs in such models. We prove that within a DW decomposition algorithm, relaxing these NACs leads to significant computational advantages. This *monotone* DW algorithm is applicable to all *monotone* two-stage SMIPs, such as SMIPs related to the capacity expansion problem, and the hydrothermal planning problem.

- We formulate the DSR problem as a two-stage monotone SMIP, and solve it to optimality with our DW decomposition algorithm. The optimal solution of this SMIP results in 19% less expected power loss than the solution obtained from the HLP. Moreover, we observe that it is optimal to spread adapter installations across the electricity grid, and that adapters should be installed in regions that are commonly disconnected from the substation.

The remainder of the paper is organized as follows. In Section 2, we introduce two-stage SMIPs, and we propose a DW algorithm for *monotone* two-stage SMIPs, in which NACs are relaxed in line with the monotonicity in the SMIP. Additionally, we explain the benefits of relaxing NACs within a DW decomposition algorithm. In Section 3, we provide the formulation of the DSR problem, and we present our numerical results in Section 4. Finally, we conclude in Section 5. Moreover, we note that throughout this paper, proofs of propositions and theorems are moved to the appendix.

2 DW Decomposition for Monotone Two-Stage SMIPs

Let $N := \{1, \dots, n_1^{\mathrm{I}} + n_1^{\mathrm{C}}\}$ denote the index set of the first-stage decision variables $x = (x_i)_{i \in N}$, where n_1^{I} and n_1^{C} are the numbers of integer and continuous first-stage variables, respectively. The *compact formulation* of a two-stage SMIP is

$$z^\star_{\mathrm{CF}} = \min_{x \in \mathcal{X}} \Big\{ c^\top x + \sum_{s \in S} p_s \, v_s(x) \Big\} \tag{1}$$

where $\mathcal{X} = \{x \in \mathbb{Z}_+^{n_1^{\mathrm{I}}} \times \mathbb{R}_+^{n_1^{\mathrm{C}}} : Ax \geq b\}$ is the set of feasible first-stage decisions, and the second-stage value function v_s of scenario $s \in S$ is defined by

$$v_s(x) = \min_{y_s \in \mathcal{Y}_s(x)} \quad q_s^\top y_s \tag{2}$$

with $\mathcal{Y}_s(x) = \{y_s \in \mathbb{Z}_+^{n_2^{\mathrm{I}}} \times \mathbb{R}_+^{n_2^{\mathrm{C}}} : W_s y_s \geq h_s - T_s x\}$. The numbers of integer and continuous second-stage variables are n_2^{I} and n_2^{C}, respectively. We impose the following standard assumptions.

Assumption 1. *We assume that $|S| < +\infty$, and moreover that*

1. *for every $s \in S$ and all $x \in \mathcal{X}$, $-\infty < v_s(x) < \infty$, and*
2. *$\mathcal{X}$ is bounded and $\mathcal{Y}_s(x)$ is bounded for all $x \in \mathcal{X}$ and $s \in S$.*

Assumption 1(1) is the *relatively complete recourse* and *sufficiently expensive recourse* condition: for every feasible first-stage decision, there exists an optimal second-stage solution with a finite cost in every scenario. Assumption 1(2) states that all feasible regions are bounded; since these regions may be arbitrarily large, this poses no practical restriction. Finally, $|S| < +\infty$ holds if scenarios are drawn from historical data or generated by sample average approximation (Kleywegt et al., 2002).

An equivalent reformulation of SMIP (1), which is known as the *split-variable formulation*, see, e.g., Romeijnders et al. (2026), introduces a scenario-specific copy x_s of the first-stage variables for each $s \in S$ and couples them to a *consensus variable* $\hat{x}$ via *non-anticipativity constraints* (NACs):

$$z^\star_{\mathrm{SV}} = \min_{\hat{x},\, x_s \in \mathcal{X}} \Big\{ c^\top \hat{x} + \sum_{s \in S} p_s\, v_s(x_s) : \, x_s = \hat{x} \ \ \forall s \in S \Big\}. \tag{3}$$

The NACs enforce that the scenario copies x_s have the same first-stage decision $\hat{x}$, so that $z^\star_{\mathrm{SV}} = z^\star_{\mathrm{CF}}$.

2.1 Monotone Two-Stage SMIPs and Relaxed NACs

A central observation of our work is that, for a broad and practically important class of SMIPs, the equality NACs in (3) can be *relaxed*—replaced by weaker inequality constraints or removed completely—without altering the optimal objective value. The enabling condition is *monotonicity* of the recourse functions v_s.

Informally, v_s is *non-increasing* in x_i if investing more in the i-th first-stage decision can only (weakly) reduce second-stage costs. This structure arises naturally whenever, for example, first-stage variables represent purchases of resources, such as vehicles or capacity, that *expand* the set of feasible second-stage recourse actions: a larger first-stage allocation can never restrict the set of feasible decisions in the second stage, so greater investment 'here-and-now' cannot increase 'future' recourse costs.

It turns out – as we will later show in Proposition 1 – that for all $s \in S$ and all $i \in N$, we may relax the NAC constraint $x_{s,i} = \hat{x}_i$ according to the type of monotonicity that v_s exhibits in x_i as follows.

1. If v_s is *non-increasing* in x_i, then replace replace $x_{s,i} = \hat{x}_i$ by $x_{s,i} \leq \hat{x}_i$.
2. If v_s is *non-decreasing* in x_i, then replace replace $x_{s,i} = \hat{x}_i$ by $x_{s,i} \geq \hat{x}_i$.
3. If v_s is *constant* in x_i, then drop the NAC for component i entirely.
4. In all other cases, retain the standard equality $x_{s,i} = \hat{x}_i$.

In order to present the relaxed NACs compactly while allowing the type of monotonicty of v_s to vary for all scenarios $s \in S$ and every component x_i of x, we introduce $\bowtie_s$ in Definition 1 as a vector of relationship operators.

Definition 1. *Let Δ_s denote the collection of all vectors of component-wise relationship operators $\bowtie_s = (\bowtie_{s,i})_{i \in N}$, where each $\bowtie_{s,i} \in \{\leq, \geq, =, \square\}$. We write $x^1 \bowtie_s x^2$ to mean $x^1_i \bowtie_{s,i} x^2_i$ for all $i \in N$, where $\square$ is a symbol indicating that this constraint is always satisfied, i.e., the constraint $a \square b$ is satisfied for all values of $a, b \in \mathbb{R}$.*

Using the definition of $\bowtie_s$ (Definition 1), we can now compactly express a relaxation of $x_s = \hat{x}$ in (3) as $x_s \bowtie_s \hat{x}$. However, as explained above, an NAC corresponding to $x_{s,i}$ may only be relaxed if v_s exhibits the corresponding type of monotonicity in x_i. To this end, we define Θ_s as the collection of vectors of relationship operators that are in line with the type of monotonicity that v_s exhibits.

Definition 2. *Given the two-stage SMIP* (1)*, define for each $s \in S$:*

$$\Theta_s := \left\{ \bowtie_s \in \Delta_s : \begin{array}{l} v_s(x^1) \geq v_s(x^2) \text{ for all} \\ x^1, x^2 \in \mathcal{X} \text{ satisfying } x^1 \bowtie_s x^2 \end{array} \right\}.$$

Observe that the definition of Θ_s aligns with the four types of monotonicity as follows. If $\bowtie_s$ equals $(\leq)_{i \in N}$, then for all $x^1, x^2 \in \mathcal{X}$ with $x^1 \leq x^2$ we have that $v_s(x^1) \geq v_s(x^2)$, i.e., v_s is *non-increasing* in all components of x; symmetrically, $\bowtie_s$ equals $(\geq)_{i \in N}$ corresponds to v_s being *non-decreasing* in all components of x. If $\bowtie_s$ equals $(\square)_{i \in N}$, then $v_s(x^1) = v_s(x^2)$ for all $x^1, x^2 \in \mathcal{X}$. Finally, $\bowtie_s$ equal to $(=)_{i \in N}$ imposes no monotonicity assumption on v_s. Therefore, it can be used when v_s does not exhibit any form of monotonicity, leaving the standard NACs in place. For ease of exposition, these examples use identical operators for all components in x. However, observe that for all $s \in S$, the definition of $\bowtie_s$ also allows the type of monotonicity to differ between all components $i \in N$.

We propose for every $\bowtie_s \in \Theta_s$ the following split-variable formulation with relaxed NACs as

$$\begin{aligned} z^\star_{\text{SV-R}} = \min_{\hat{x}, x_s \in \mathcal{X}} \quad & c^\top \hat{x} + \sum_{s \in S} p_s v_s(x_s) \\ \text{s.t.} \quad & x_s \bowtie_s \hat{x} \qquad\qquad s \in S. \end{aligned} \tag{4}$$

Note that, for every $s \in S$, we have $\Theta_s \neq \emptyset$, since the condition in Definition 2 is satisfied when $\bowtie_s$ equals $(=)_{i \in N}$. In fact, the formulation (4) is then identical to the split-variable formulation with traditional NACs in (3). However, when v_s monotone in some components $i \in N$, the set Θ_s is typically larger, and thus, the *monotone* split-variable formulation (4) is a strict relaxation of the traditional split-variable formulation (3) for such $\bowtie_s \in \Theta_s$. In Proposition 1 below, we prove that indeed these two formulations yield the same optimal solution.

Proposition 1 ($z^\star_{\text{SV-R}} = z^\star_{\text{SV}}$)**.** *The optimal objective values of SMIP with traditional NACs in* (3) *and of the SMIP with relaxed NACs in* (4) *are equal if $\bowtie_s \in \Theta_s$ for all $s \in S$.*

The set of vectors $\bowtie_s \in \Theta_s$ generally contains multiple elements for all $s \in S$ if the second-stage value functions v_s are monotone. In such cases, any $\bowtie_s \in \Theta_s$ can thus be used to obtain the optimal solution $\hat{x}^\star$ in (3). However, for the decomposition algorithm introduced in Section 2.2, it is typically better to relax the operator types in $\bowtie_s$ as much as possible for all $s \in S$ to reduce computation time. For example, operator $\square$ is less restrictive than operators $\leq$ and $\geq$, and operator $=$ is more restrictive than $\leq$ and $\geq$. Proposition 2 provides a simple sufficient condition based on the signs of the elements in the columns $T_{s,i}$ of the technology matrix T_s to determine relaxed relationship operators $\bowtie_s \in \Theta_s$.

Proposition 2. *If for all $s \in S$ and all $i \in N$ we have that $\bowtie_{s,i}$ equals '$\square$' if $T_{s,i} = \mathbf{0}$, '$\geq$' if $T_{s,i} \leq \mathbf{0}$ and $T_{s,i} \neq \mathbf{0}$, '$\leq$' if $T_{s,i} \geq \mathbf{0}$ and $T_{s,i} \neq \mathbf{0}$, and '$=$', otherwise. Then, $\bowtie_s \in \Theta_s$ for all $s \in S$.*

2.2 Monotone Dantzig-Wolfe Decomposition

In this section, we introduce the monotone Dantzig-Wolfe decomposition algorithm. Before doing so, we first rewrite the problem (4) as

$$\begin{aligned} z^\star_{\text{SV-R}} = \min_{\substack{\hat{x} \in \mathcal{X} \\ (x_s, y_s) \in \mathcal{F}_s}} \quad & c^\top \hat{x} + \sum_{s \in S} p_s q_s^\top y_s \\ \text{s.t.} \quad & x_s \bowtie_s \hat{x} \qquad\qquad s \in S, \end{aligned} \tag{5}$$

where $\mathcal{F}_s$ denotes the set $\{(x_s, y_s) \in (\mathbb{Z}_+^{n_1^{\text{I}}} \times \mathbb{R}_+^{n_1^{\text{C}}}) \times (\mathbb{Z}^{n_2^{\text{I}}} \times \mathbb{R}^{n_2^{\text{C}}}) : x_s \in \mathcal{X}, y_s \in \mathcal{Y}_s(x_s)\}$. We solve this large-scale mixed-integer program to optimality with the Dantzig-Wolfe decomposition algorithm combined with branch-and-bound. At the root node of the branch-and-bound tree, we use the following relaxation of program (5), obtained by replacing $\mathcal{F}_s$ by $\text{conv}(\mathcal{F}_s)$ and $\mathcal{X}$ by $\text{conv}(\mathcal{X})$:

$$\begin{aligned} z^\star_{\text{DW-R}} = \min_{\substack{\hat{x} \in \text{conv}(\mathcal{X}) \\ (x_s, y_s) \in \text{conv}(\mathcal{F}_s)}} \quad & c^\top \hat{x} + \sum_{s \in S} p_s q_s^\top y_s \\ \text{s.t.} \quad & x_s \bowtie_s \hat{x} \qquad s \in S. \end{aligned} \tag{6}$$

This linear relaxation is often called the Dantzig-Wolfe reformulation, see, e.g., (Wolsey and Nemhauser, 1999). The main difference, between the relaxed DW formulation in (6) and the traditional one, where $\bowtie_s$ equals $(=)_{i \in N}$ for all $s \in S$, is that we may relax some NACs based on the monotonicity of v_s for all $s \in S$. As a result of this relaxation, we must add the constraint $\hat{x} \in \text{conv}(\mathcal{X})$, whereas this constraint may be omitted in the traditional DW formulation, since it is implied by the fact that $x_s \in \text{conv}(\mathcal{X})$ and that the traditional NACs state that $\hat{x} = x_s$ for all $s \in S$.

Following the standard DW approach, we can express, using Minikowski's Theorem, any point $(x_s, y_s) \in \text{conv}(\mathcal{F}_s)$ as a convex combination of the extreme points of $\text{conv}(\mathcal{F}_s)$. For all $s \in S$, let K_s denote the index set corresponding to the extreme points (x_s^k, y_s^k) of $\text{conv}(\mathcal{F}_s)$, and let $\lambda_s^k \geq 0$ represent the weight of extreme point (x_s^k, y_s^k). Additionally, we require that $\sum_{k \in K_s} \lambda_s^k = 1, s \in S$, which implies that $(x_s, y_s) = \sum_{k \in K_s} (x_s^k, y_s^k) \lambda_s^k$. The resulting DW Master Problem (MP) is thus:

$$
\begin{aligned}
z^\star_{\text{DW-R}} = \min_{\substack{\hat{x}\in\operatorname{conv}(\mathcal{X})\\ \lambda\geq 0}} \quad & c^\top \hat{x} + \sum_{s\in S}\sum_{k\in K_s} p_s q_s^\top y_s^k \lambda_s^k \\
\text{s.t.} \quad & \sum_{k\in K_s} x_s^k \lambda_s^k \bowtie_s \hat{x} \quad s\in S \\
& \sum_{k\in K_s} \lambda_s^k = 1 \qquad s \in S,
\end{aligned}
\tag{7}
$$

where $\lambda = (\lambda_s^k)_{s\in S, k\in K_s}$ denotes the vector of decision variables that represent the weight of extreme points (x_s^k, y_s^k) for all $s \in S$ and all $k \in K_s$.

Remark 1. *As an alternative to formulation* (7)*, the first-stage consensus decision variable* $\hat{x}$ *may also be expressed as a convex combination of the extreme points of* $\operatorname{conv}(\mathcal{X})$*.*

We denote $z^\star_{\text{DW}}$ as an optimal solution of MP (7) where all NACs use equality relationship operators. Since directly solving MP (7) is generally computationally intractable because the original formulation admits an exponentially large number of extreme points, making an explicit enumeration or representation is impractical. Therefore, we solve the restricted master problem (RMP), which starts with only a small subset $\bar{K}_s \subseteq K_s$ of the variables corresponding to extreme points. For convenience of notation, we define $\bar{K} = (\bar{K}_s)_{s\in S}$ and the RMP of (7) for a given index set $\bar{K}$ as

$$
\begin{aligned}
z^\star_{\text{RMP-R}}(\bar{K}) = \min_{\substack{\hat{x}\in\operatorname{conv}(\mathcal{X})\\ \lambda\geq 0}} \quad & c^\top \hat{x} + \sum_{s\in S}\sum_{k\in \bar{K}_s} p_s q_s^\top y_s^k \lambda_s^k \\
\text{s.t.} \quad & \sum_{k\in \bar{K}_s} x_s^k \lambda_s^k \bowtie_s \hat{x} \quad s\in S \\
& \sum_{k\in \bar{K}_s} \lambda_s^k = 1 \qquad s \in S,
\end{aligned}
\tag{8}
$$

where π_s and α_s denote the dual variables of the first and second set of constraints, respectively. The RMP is solved to obtain optimal dual variables $\bar{\pi}_s$, $\bar{\alpha}_s$ for all $s \in S$. Afterwards, the pricing problem (9) below is solved to obtain columns of the MP that could improve the RMP's solution:

$$
r_s^\star(\bar{\pi}_s, \bar{\alpha}_s) = \min_{(x_s,y_s)\in\mathcal{F}_s} \Big\{ p_s q_s^\top y_s - \bar{\pi}_s^\top x_s - \bar{\alpha}_s \Big\}. \tag{9}
$$

If $r_s^\star(\bar{\pi}_s, \bar{\alpha}_s) \geq 0$ for all $s \in S$, it follows that $z^\star_{\text{RMP-R}}(\bar{K}) = z^\star_{\text{DW-R}}$ (Desrosiers and Lübbecke (2005)).

2.3 The effects of relaxing NACs in Dantzig-Wolfe Decomposition

To be able to compare the RMPs of the standard and relaxed DW formulation, we first reformulate the RMP (8) where we make the dependence on $\bar{K}_s$ and $\bowtie_s$ explicit, that is, we define $z^\star_{\text{RMP-R}}$ as

$$
z^\star_{\text{RMP-R}}(\bar{K}) = \min_{\hat{x}\in\operatorname{conv}(\mathcal{X})} \quad c^\top \hat{x} + \bar{Q}\Big(\hat{x}, (\bar{K}_s, \bowtie_s)_{s\in S}\Big), \tag{10}
$$

where $\bar{Q}\Big(x, (\bar{K}_s, \bowtie_s)_{s\in S}\Big) = \sum_{s\in S} p_s \bar{v}_s(x, \bar{K}_s, \bowtie_s)$, and $\bar{v}_s(x, \bar{K}_s, \bowtie_s)$ denotes the approximation of the second-stage cost $v_s(x)$ in scenario s, for a given subset of columns $\bar{K}_s \subseteq K_s$ and NAC operators $\bowtie_s \in \Theta_s$, i.e., $\bar{v}_s(x, \bar{K}_s, \bowtie_s)$ is defined as

$$\bar{v}_s(x, \bar{K}_s, \bowtie_s) = \min_{(\lambda_s^k)_{k\in\bar{K}_s}\geq 0} \quad \sum_{k\in\bar{K}_s} \lambda_s^k q_s^\top y_s^k \tag{11a}$$

$$\text{s.t.} \quad \sum_{k\in\bar{K}_s} \lambda_s^k = 1 \tag{11b}$$

$$\sum_{k\in\bar{K}_s} \lambda_s^k x_s^k \bowtie_s x. \tag{11c}$$

The traditional RMP without relaxed NACs corresponds to $\bar{v}_s(x, \bar{K}_s, (=)_{i\in N})$, that is, $\bowtie_{s,i}$ equals '=' for all $s \in S$ and all $i \in N$. Moreover, $z^\star_{\text{RMP}}$ denotes the optimal objective value of the traditional RMP (10). We compare the traditional RMP without relaxed NACs with the relaxed RMP by investigating the difference between $\bar{v}_s(x, \bar{K}_s, (=)_{i\in N})$ and $\bar{v}_s(x, \bar{K}_s, \bowtie_s)$.

Relaxing the NACs in the DW decomposition affects the MP and RMP in four fundamental ways. First, the effective domain of the approximated second-stage value functions (11) enlarges, reducing the plateau effect (Vanderbeck and Savelsbergh, 2006), which is the effect that optimal objective values of the RMP (8) remain constant for several iterations of a DW algorithm, see, e.g., Desrosiers and Lübbecke (2005). Second, more columns of the MP (7) become dominated in the RMP (8), so the pricing problem (9) produces fewer columns. Third, the error in the second-stage cost approximation may increase, potentially widening the optimality gap. However, we also identify a broad class of instances where no gap arises. Fourth, we show that for SMIPs with pure-binary first-stage variables, the approximated second-stage cost $\bar{v}_s(x, K_s, \bowtie_s)$ is equal to the true second-stage cost $v_s(x)$ at all binary first-stage solutions $x \in \mathcal{X} \subseteq \mathbb{B}^N$. Consequently, we show that branching solely on consensus variables $\hat{x}$ in (10) suffices when first-stage variables are binary. However, in general, spatial branching or branching on second-stage variables might be required to find optimal integer-feasible solutions (Kim and Dandurand, 2022). In the next subsections, we discuss all four effects in turn.

2.3.1 Larger effective domains

We begin by showing that relaxing NACs enlarges the effective domain in the RMP (10). To do so, we define the *effective domains* of $\bar{v}_s(., \bar{K}_s, \bowtie_s)$ and $\bar{Q}\Big(., (\bar{K}_s, \bowtie_s)_{s\in S}\Big)$ in Equations (11) and (10) as $\mathcal{X}_s(\bar{K}_s, \bowtie_s) := \big\{x \in \mathbb{R}^N_+ : \bar{v}_s(x, \bar{K}_s, \bowtie_s) < \infty\big\}$, and

$$\mathcal{X}\Big((\bar{K}_s, \bowtie_s)_{s\in S}\Big) := \bigcap_{s\in S} \mathcal{X}_s(\bar{K}_s, \bowtie_s). \tag{12}$$

Then, we prove in in Proposition 3 that these effective domains increase when the NACs are relaxed.

Proposition 3. *If $\bowtie_s \in \Theta_s$ for all $s \in S$, then $\mathcal{X}_s(\bar{K}_s, (=)_{i\in N}) \subseteq \mathcal{X}_s(\bar{K}_s, \bowtie_s)$ for all $s \in S$, and consequently $\mathcal{X}\big((\bar{K}_s, (=)_{i\in N})_{s\in S}\big) \subseteq \mathcal{X}\big((\bar{K}_s, \bowtie_s)_{s\in S}\big)$.*

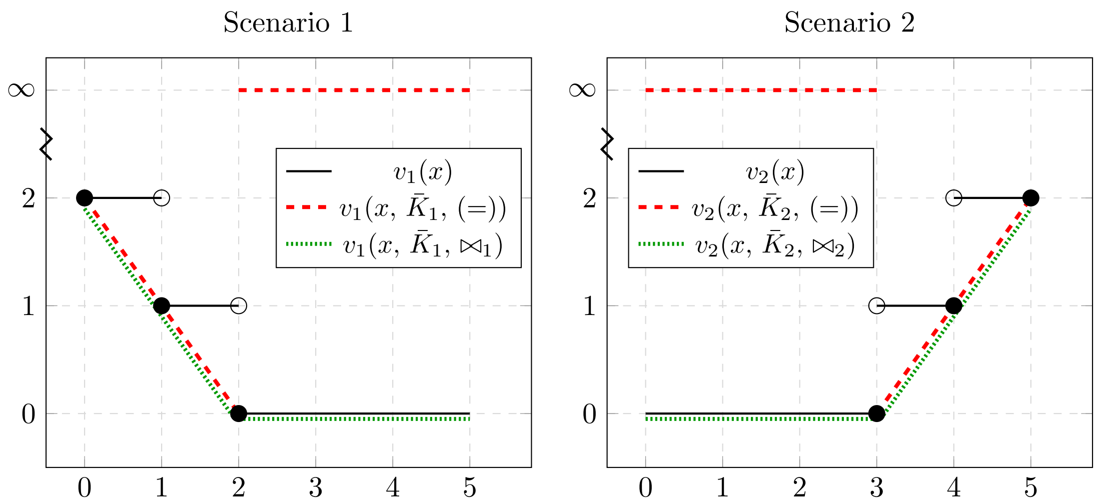


Figure 2: Second-stage value functions $v_s(.)$ (black, solid), RMP approximation without relaxed NACs $\bar{v}_s(., \bar{K}_s, (=)_{i\in N})$ (red, striped), and with relaxed NACs $\bar{v}_s(., \bar{K}_s, \bowtie_s)$ (green, dotted).

As a result of these increased effective domains, the RMP (10) is thus able to find better solutions when the NACs are relaxed. This effect is amplified when the number of scenarios $|S|$ is large. To understand this, observe that effective domain of $\bar{Q}\Big(., (\bar{K}_s, \bowtie_s)_{s\in S}\Big)$ in (12) involves an intersection over all scenarios $s \in S$. Therefore, this effective domain shrinks as $|S|$ grows, which can significantly limit the number of feasible solutions of the RMP if there are only limited number of columns in the RMP. Relaxing the NACs partially counteracts this effect by increasing each per-scenario domain $\mathcal{X}_s(\bar{K}_s, \bowtie_s)$. In Example 1, we provide an example of a SMIP where relaxing the NACs indeed increases the effective domains.

Example 1. *Consider the SMIP* $z^\star = \min_{x\in\mathcal{X}} \frac{1}{2} v_1(x) + \frac{1}{2} v_2(x)$*, where* $\mathcal{X} := \{x \in \mathbb{R}_+ : x \leq 5\}$*,* $v_s(x) := \min\{y_s : y_s \in \mathcal{Y}_s(x)\}$*,* $\mathcal{Y}_1(x) := \{y \in \mathbb{Z}_+ : y \leq 2,\ y \geq 2 - x\}$*, and* $\mathcal{Y}_2(x) := \{y \in \mathbb{Z}_+ : y \leq 2,\ y \geq x - 3\}$*.*

Observe that v_1 *is monotonically non-increasing, since the feasible region* $\mathcal{Y}_1(x)$ *is strictly increasing with respect to* x*. Similarly,* v_2 *is monotonically non-decreasing. Thus, according to Proposition 1, we may choose* $\bowtie_1$ *and* $\bowtie_2$ *equal to '*$\leq$*' and '*$\geq$*', respectively. Moreover, observe that two of the extreme points of* $\mathcal{F}_1$ *are* $(x_1^{k_1}, y_1^{k_1}) = (0, 2)$ *and* $(x_1^{k_2}, y_1^{k_2}) = (2, 0)$*, and two of those of* $\mathcal{F}_2$ *are* $(x_2^{k_3}, y_2^{k_3}) = (3, 0)$ *and* $(x_2^{k_4}, y_2^{k_4}) = (5, 2)$*. Let* $\bar{K}_1 = \{k_1, k_2\}$*, and* $\bar{K}_2 = \{k_3, k_4\}$*. Figure 2 visualizes the true value functions* $v_s(.)$*, the RMP approximations without relaxed NACs* $\bar{v}_s(., \bar{K}_s, (=)_{i\in N})$*, and the RMP approximations with relaxed NACs* $\bar{v}_s(., \bar{K}_s, \bowtie_s)$*.*

The resulting effective domains are:

$$
\begin{aligned}
&\mathcal{X}_1(\bar{K}_1, (=)) = [0, 2] \qquad &&\mathcal{X}_1(\bar{K}_1, (\leq)) = [0, \infty) \\
&\mathcal{X}_2(\bar{K}_2, (=)) = [3, 5] \qquad &&\mathcal{X}_2(\bar{K}_2, (\geq)) = (-\infty, 5] \\
&\mathcal{X}\big((\bar{K}_s, (=))_{s\in S}\big) = \emptyset \qquad &&\mathcal{X}(\bar{K}, \bowtie) = [0, 5]
\end{aligned}
$$

where $(\bar{K}, \bowtie) = \big((\bar{K}_1, (\leq)), (\bar{K}_2, (\geq))\big)$*. Observe that the effective domains with relaxed NACs (in the right column) are strictly larger than those corresponding to non-relaxed NACs (in the left column). Moreover, observe in particular that the effective domain of* $\bar{Q}\Big(., (\bar{K}_s, \bowtie_s)_{s\in S}\Big)$ *(in* *the third row) is empty if the NACs are not relaxed (in the left column), while this is not the case when the NACs are relaxed (in the right column).*

The enlarged effective domain as a result of relaxing NACs generally reduces the *plateau effect* (Vanderbeck and Savelsbergh, 2006), i.e., the effect in which objective values of RMPS remain constant over consecutive iterations. To understand this, recall that in the DW algorithm, see, e.g., Desrosiers and Lübbecke (2005), continuously a new column with index k' of the MP (7) which is not yet in the RMP (8), i.e., $k' \in K_s \setminus \bar{K}_s$, is included in the set of columns of the RMP, i.e. $\bar{K}_s \leftarrow \bar{K}_s \cup k'$. This column with index k' can only improve the solution of the RMP if there exists a feasible solution of the RMP where its corresponding weight $\lambda_s^{k'} > 0$. Since the weights λ_s^k are used to represent first-stage decisions in the effective domain of $\bar{v}_s(., \bar{K}_s, \bowtie_s)$, enlarging this effective domain makes it more likely that a new solution of the RMP with $\lambda_s^{k'} > 0$ exists, which in turn may result in a new optimal objective value. We demonstrate in numerical experiments for our case study, see, e.g., Figure 4 in Section 4, that relaxing NACs indeed strongly reduced the plateau effect in early iterations of the algorithm.

2.3.2 More dominated columns

Dominated columns, see, e.g., Achterberg et al. (2020), are columns in an LP for which there exists a convex combination of other columns in the LP which have a better objective value and a less negative effect on the constraints in the LP. Therefore, (re-)assigning the weight of the dominated column to the convex combination of columns that dominate it will result in a lower objective value while still satisfying the constraints of the LP. Following the definition of (weakly) dominated columns (Achterberg et al., 2020), the set of dominated columns $\tilde{K}_s(\bar{K}_s, \bowtie_s)$ in the MP, by columns $\bar{K}_s$ in the RMP under NACs with $\bowtie_s$ is defined as

$$\tilde{K}_s(\bar{K}_s, \bowtie_s) = \left\{ \tilde{k} \in K_s : \text{ s.t. } 13(\text{a})\text{-}13(\text{d}) \right\} \tag{13}$$

where equations 13(a)-13(d) denote $\exists (\mu_s^k)_{k \in \bar{K}_s} \geq 0$, $\sum_{k \in \bar{K}_s} \mu_s^k = 1$, $\sum_{k \in \bar{K}_s} \mu_s^k x_s^k \bowtie_s x_s^{\tilde{k}}$, and $\sum_{k \in \bar{K}_s} \mu_s^k q_s^\top y_s^k \leq q_s^\top y_s^{\tilde{k}}$, respectively. Intuitively, any weight $\lambda_s^{\tilde{k}}$ that the RMP assigns to a dominated column with index $\tilde{k}$ can instead be assigned to the dominating combination $(\mu_s^k)_{k \in \bar{K}_s}$, which leads to a lower objective value while still satisfying the constraints in the RMP (8).

Proposition 4 shows that, for all scenarios $s \in S$ and for a fixed set of columns $\bar{K}_s$ in s, the RMP with relaxed NACs generally has more dominated columns than the RMP without relaxed NACs.

Proposition 4. *Let $\bowtie_s \in \Theta_s$ for all $s \in S$ be given. Then, $\tilde{K}_s(\bar{K}_s, (=)_{i \in N}) \subseteq \tilde{K}_s(\bar{K}_s, \bowtie_s)$ holds for all $s \in S$.*

This effect is particularly pronounced when $\mathcal{X} \subseteq \mathbb{B}^N$. Under non-relaxed NACs, every x_s^k is a vertex of $\text{conv}(\mathcal{X})$ and can only be dominated by convex combinations of columns sharing the same x-value. This is the case because otherwise constraint $\sum_{k \in \bar{K}_s} \mu_s^k x_s^k \bowtie_s x_s^{\tilde{k}}$ would not be satisfied. Or intuitively, an extreme point of $\text{conv}(\mathcal{X})$ can only be obtained by taking a convex combination of points that coincide with it, see, e.g., Zou et al. (2019). Under relaxed NACs, columns with *different* x-values may also dominate it, substantially enlarging the dominated set.

The practical relevance of dominated columns is that they have non-negative reduced costs at any optimal RMP dual solution, as we show next in Proposition 5 below.

Proposition 5. *Let $\tilde{k} \in \tilde{K}_s(\bar{K}_s, \bowtie_s)$ be given, and let $(\pi_s, \alpha_s)_{s\in S}$ be an optimal dual solution of the RMP* (8)*. Then, the reduced cost of the column corresponding to $\tilde{k}$, $r_s^{\tilde{k}}(\pi_s, \alpha_s) := p_s\, q_s^\top y_s^{\tilde{k}} - \pi_s^\top x_s^{\tilde{k}} - \alpha_s$, is non-negative.*

Thus, by enlarging the set of dominated columns, the RMP with relaxed NACs classifies more columns as having non-negative reduced costs. Consequently, the RMP with relaxed NACs tends to reach its termination criterion — where no columns with negative reduced costs remain - more rapidly than the traditional RMP without relaxed NACs.

2.3.3 Weaker bounds on the second-stage value functions

Relaxing NACs also has a downside, namely, the second-stage cost approximations $\bar{v}_s(x, K_s, \bowtie_s)$ provide a weaker lower bound on the true second-stage cost v_s than its non-relaxed counterpart $\bar{v}_s(x, K_s, (=)_{i\in N})$. In this section, we characterize when this gap may exist, and when it vanishes. Since the minimization with relaxed NACs in (11) is a relaxation of the one with NAC operators $(=)_{i\in N}$, we immediately have

$\bar{v}_s(x, \bar{K}_s, \bowtie_s) \le \bar{v}_s(x, \bar{K}_s, (=)_{i\in N})$ and $\bar{v}_s(x, K_s, \bowtie_s) \le \bar{v}_s(x, K_s, (=)_{i\in N})$. Recall that the traditional MP without relaxed NACs and with all columns recovers the convex envelope: $\bar{v}_s(x, K_s, (=)_{i\in N}) = \mathrm{co}_{\mathcal{X}}(v_s)(x)$ (Romeijnders et al., 2026). Hence the MP without relaxed NACs already underestimates $v_s(x)$, and relaxing NACs may amplify this underestimation. Proposition 6 below provides a useful characterization of the second-stage value function approximation $\bar{v}_s(x, \bar{K}_s, \bowtie_s)$ of v_s.

Proposition 6. *For all $x \in \mathbb{R}_+^N$ and $\bar{K}_s \subseteq K_s$,*

$$\bar{v}_s(x, \bar{K}_s, \bowtie_s) = \min_{x_s} \big\{ \bar{v}_s(x_s, \bar{K}_s, (=)_{i\in N}) \ : \ x_s \bowtie_s x \big\}. \tag{14}$$

In particular, when $\bar{K}_s = K_s$,

$$\bar{v}_s(x, K_s, \bowtie_s) = \min_{x_s} \big\{ \mathrm{co}_{\mathcal{X}}(v_s)(x_s) \ : \ x_s \bowtie_s x \big\}. \tag{15}$$

Proposition 6 reveals the mechanism behind the relaxed-NAC approximation: it selects the point x_s in the region $\{x_s \in \mathrm{conv}(\mathcal{X}) : x_s \bowtie_s x\}$ that minimizes the convex envelope $\mathrm{co}_{\mathcal{X}}(v_s)$. According to Proposition 7 below, if $\mathrm{co}_{\mathcal{X}}(v_s)$ is monotone in the directions prescribed by $\bowtie_s$, then the minimizer is $x_s = x$ itself, and no gap arises.

Proposition 7. *Let $\bowtie_s \in \Theta_s$ be given. If $\mathrm{co}_{\mathcal{X}}(v_s)(x^1) \ge \mathrm{co}_{\mathcal{X}}(v_s)(x^2)$ for all $x^1, x^2 \in \mathrm{conv}(\mathcal{X})$ with $x^1 \bowtie_s x^2$, then $\bar{v}_s(x, K_s, \bowtie_s) = \mathrm{co}_{\mathcal{X}}(v_s)(x)$ for all $x \in \mathrm{conv}(\mathcal{X})$.*

Proposition 7 highlights that the gap between $\bar{v}_s(x, K_s, \bowtie_s)$ and $\mathrm{co}_{\mathcal{X}}(v_s)(x)$ occurs because $\bowtie_s \in \Theta_s$ merely guarantees monotonicity of v_s on $\mathcal{X}$, but the convex envelope $\mathrm{co}_{\mathcal{X}}(v_s)$ need not inherit this monotonicity on $\mathrm{conv}(\mathcal{X})$ in general. However, the next theorem identifies a broad class of instances where the optimality gap is null.

Theorem 1. *Let $\bowtie_s \in \Theta_s$ be given, and $\mathcal{X} = \{x \in \mathbb{Z}_+^{n_1^I} \times \mathbb{R}_+^{n_1^C} : Ax \geq b\}$ with $N = 1, \dots, |n_1^I| + |n_1^C|$. If $\operatorname{conv}(\mathcal{X})$ is a hyper-rectangle, i.e., $\operatorname{conv}(\mathcal{X}) = \{x \in \mathbb{R}_+^N : l_i \leq x_i \leq u_i\}$, then*

$$\bar{v}_s(x, K_s, \bowtie_s) = \mathrm{co}_{\mathcal{X}}(v_s)(x) \textit{ for all } x \in \operatorname{conv}(\mathcal{X}).$$

Theorem 1 covers all instances where the first-stage feasible region consists of individually bounded variables, including the common case $\mathcal{X} = \mathbb{B}^N$. In such instances, the relaxed-NAC MP approximates v_s equally as good as the traditional MP, so the algorithmic benefits (larger effective domain, more dominated columns) come at no cost in approximation quality. Outside the hyper-rectangle setting, a gap may exist.

Proposition 8. *There exist SMIP, and corresponding $\hat{x} \in \operatorname{conv}(\mathcal{X})$ and $\bowtie_s \in \Theta_s$, such that*

$$\bar{v}_s(\hat{x}, K_s, \bowtie_s) < \bar{v}_s(\hat{x}, K_s, (=)_{i \in N}).$$

2.3.4 Branch-and-bound procedure

In Section 2.3.3, we show that relaxing NACs can increase the optimality gap of the MP (8). In turn, this may also interfere with the B&B procedure used to obtain optimal solutions that satisfy integrality restrictions. In general, DW algorithms for two-stage SMIPs (with or without monotonicity) may require a branch-and-bound procedure that uses spatial-branching, or branching on second-stage variables (Kim and Dandurand, 2022).

However, we observe that in the spacial case where the first-stage variables are pure binary, branching solely on first-stage consensus variables $\hat{x}$ in MP (8) with relaxed NACs is sufficient to obtain optimal solutions that satisfy integrality restrictions. To prove this, we introduce a minimization problem similar to the MP (7), where additional integrality restrictions are enforced on the first-stage consensus variables. Recall that the MP (7) is equivalent to the RMP (10) that includes all columns, i.e., $\bar{K}_s = K_s$ for all $s \in S$. This minimization problem is defined as:

$$z^\star_{\text{DW-R}} = \min_{\hat{x} \in \mathcal{X}} \Big\{ c^\top \hat{x} + \sum_{s \in S} p_s \bar{v}_s(x, K_s, \bowtie_s) \Big\}. \tag{16}$$

Observe that a branch-and-bound procedure that branches solely on first-stage consensus variables $\hat{x}$ in the MP (10) (with $\bar{K}_s = K_s$ for all $s \in S$) yields an optimal solution of the minimization problem in (16) above. It turn out that this solution is also an optimal solution of the original SMIP (1) if the first-stage variables are pure-binary. Note that a sufficient condition is that the *monotone* DW second-stage value approximation $\bar{v}_s(x, K_s, \bowtie_s)$ equals the true second-stage value function $v_s(x)$, at all integer feasible first-stage solutions $x \in \mathcal{X}$. Indeed, we prove this in Theorem 2 below.

Theorem 2. *If $\mathcal{X} \subseteq \mathbb{B}^N$, and $\bowtie_s \in \Theta_s$ for all $s \in S$, then $\bar{v}_s(x, K_s, \bowtie_s)$ in* (11) *equals $v_s(x)$ in* (2) *for all $x \in \mathcal{X}$ and $s \in S$.*

The key benefit of branching solely on consensus first-stage variables $\hat{x}$, is that $\mathcal{F}_s$ remains unaffected by branch-and-bound restrictions for all $s \in S$. As a result, all columns corresponding to extreme points of $\operatorname{conv}(\mathcal{F}_s)$ are valid in all nodes of the branch-and-bound scheme, and thus need never be removed. This may be computationally beneficial, because the effective domains in the RMPs will be larger if no columns are removed, yielding the benefits of larger effective domains as explained in a previous section. Another key benefit that follows from Theorem 2 is that it becomes easy to verify feasibility of RMP solutions. That is, if the consensus variable $\hat{x}$ in RMP (6) satisfies binary integrality restrictions – even when for some scenario $s \in S$, copy of first-stage variable, x_s, does not – then this RMP solution is feasible.

3 Case Study: Power Distribution System Resilience

In this section, we present a case study on the power distribution system resilience (DSR) problem. We present a novel two-stage stochastic programming formulation for the DSR problem and solve it exactly using our new *monotone* DW algorithm. Consequently, we provide novel insights into optimal decisions in the DSR problem, and we compare these solutions with heuristic ones.

3.1 Problem Description

Consider an electricity distribution grid, modeled as a connected graph $G = (N, B)$, with N representing the set of nodes and B the set of distribution branches in the grid. Electricity may flow over branches in B between customers at nodes N. Additionally, a substation provides power to this grid and is located at node $n^1 \in N$. Moreover, the consumers at node n have a positive real and reactive power demand, represented by D_n^+ and D_n^- per time period, respectively. We assume that under normal operating conditions, the substation has sufficient capacity to supply the entire grid's power demand.

However, extreme weather events, such as storms, tornadoes, and floods, can severely disrupt the system by damaging distribution lines and the surrounding road infrastructure. The DSR problem addresses both the strategic preparation before and the operational response after an extreme weather event. In particular, before any extreme weather event occurs, we may equip a limited number of nodes, $X^{\max}$, with an electricity adapter, so that mobile power sources (MPSs) may temporarily provide backup power via such adapters after a disaster has happened. In fact, after such an event has occurred, we jointly employ repair crews $r \in R$ and MPSs $m \in M$ over a discrete time horizon with $|T|$ periods to minimize total power loss. The DSR problem is to determine which nodes in the network to equip with an electricity adapter in order to minimize total expected future power loss.

3.2 DSR SMIP Formulation

Before an extreme weather event occurs, electricity adapters may be installed at candidate nodes $N^c \subseteq N$. Let $x_i \in \{0, 1\}$ be equal to 1 if an electricity adapter is installed at candidate node n, and 0, otherwise. We propose the following SMIP to solve the DSR problem:

$$z^\star_{\mathrm{DSR}} = \min_{x \in \mathcal{X}} \Big\{ \sum_{s \in S} p_s v_s(x) : \sum_{n \in N^c} x_n \leq X^{\max} \Big\}, \tag{17}$$

where $\mathcal{X} := \{0,1\}^{|N^c|}$ and the constraint in (17) states that at most $X^{\max}$ electricity adapters may be installed at candidate nodes $N^c \subseteq N$, p_s denotes the probability that scenario s occurs, $v_s(x)$ denotes the total power loss in scenario s, and $x = (x_n)_{n \in N^c}$ denotes the first-stage decision vector. We define for every $s \in S$ the second-stage value function $v_s(x)$ as

$$v_s(x) = \min_{y_s \in \mathcal{Y}_s} \sum_{n \in N} \sum_{t \in T} \pi^+_{n,t} \tag{18}$$

where the objective is to minimize the aggregated power loss $\sum_{n \in N} \sum_{t \in T} \pi^+_{n,t}$ over all nodes $n \in N$, and all time periods $t \in T$. Moreover, $\mathcal{Y}_s = \mathcal{Y}_s^{\text{MPS}} \cap \mathcal{Y}_s^{\text{RC}} \cap \mathcal{Y}_s(x)^{\text{pow}} \cap \mathcal{Y}_s^{\text{flow}} \cap \mathcal{Y}_s^{\text{volt}}$, where the sets $\mathcal{Y}_s^{\text{MPS}}$, $\mathcal{Y}_s^{\text{RC}}$, $\mathcal{Y}_s(x)^{\text{pow}}$, $\mathcal{Y}_s^{\text{flow}}$, and $\mathcal{Y}_s^{\text{volt}}$ contain constraints that are related to different aspects of the DSR problem. The vector $y_s = (l^s, \rho^s, a^s, c^s, p^s, f^s, \pi^s, o^s, \nu^s)$ denotes the aggregated vector of all second-stage decision variables. In the next subsections, we elaborate on the second-stage variables and constraints. The second-stage problem in which operational decisions are made is based on the MIP proposed by Anokhin et al. (2021).

3.2.1 MPS routing

Consider a road transportation network which is also modeled as a graph (N_2, G_2) where N_2 denotes the set of locations in the transportation network, and G_2 denotes the roads between the locations in the transportation network. We assume that every node in the DS is located at exactly one location in the transportation network. Moreover, we assume electrical branches (power lines) can be accessed by RCs via the nodes in the transportation network that corresponds to the DS node closest to the substation (at node 0), that is also adjacent to the branch. Each damage scenario $s \in S$ specifies a subset of damaged branches B_s^d in the electrical network and roads G_s^d in the transportation network, and we assume that these damaged roads cannot be repaired or traversed throughout the entire restoration period. This is in line with existing literature, see, e.g., Anokhin et al. (2021), Su et al. (2023), and Su et al. (2024). For modeling purposes, it is not explicitly required to know where MPSs and RCs are located in the transportation network, since this is already implied by their locations in the DS. Therefore, we only consider the transportation network implicitly, by pre-computing the travel distances $T_{n_1,n_2,s}$ between all nodes $n_1, n_2 \in N$ in the DS using the travel distances in the transportation network under road damages G_s^d in scenario s.

The binary variable $l^s_{n,m,t}$ equals 1 if MPS m is located at node n at time t in scenario s, and 0, otherwise. The depot of MPSs is located at node n_1^{MPS} in the DS. The MPSs are pre-positioned at this node in the first time period, $t^1 = 0$. Afterwards, they may travel over the transportation network to provide temporary power at other nodes in the DS. For modeling purposes, however, it is sufficient to only keep track, for all time periods $t \in T$, of whether an MPS m is located at the starting location n_1^{MPS}, or at disconnected candidate nodes $N_s^c \subseteq N$ in scenario s. The set of nodes between which the MPSs travel, N_s^d, is thus $N_s^d := \{n_1^{\text{MPS}}\} \cup N_s^c$. Moreover, we write $y_s \in \mathcal{Y}_s^{\text{MPS}}$ if $l^s \in \mathcal{L}_s := \{0,1\}^{|N_s^d| \times |M| \times |T|}$ satisfies

$$\sum_{n \in N_s^d} l_{n,m,t}^s \leq 1, \qquad \forall m \in M,\ t \in T \tag{19a}$$

$$l_{n_1^{\text{MPS}},m,t^1}^s = 1, \qquad \forall m \in M \tag{19b}$$

$$l_{n_1,m,t_1}^s + l_{n_2,m,t_2}^s \leq 1, \qquad \forall m \in M, (n_1, n_2, t_1, t_2) \in \mathcal{T}_s, \tag{19c}$$

where Equation (19a) states that in every time period t, each MPS m cannot be located at more than a single location $n \in N_s^d$ in scenario s. Equation (19b) states that all MPSs start at the depot location corresponding to node n_1^{MPS}, at time $t^1 = 1$ in scenario s. Finally, Equation (19c) contains the travel restrictions for scenario s.

In general, it is not possible to travel from node n_1 at time t_1 to node n_2 at time t_2 in scenario s if the required travel time $T_{n_1,n_2,s}$ exceeds $|t_2 - (t_1 + 1)|$. For all $n_1, n_2 \in N$ and all $t_1, t_2 \in T$, we have that $(n_1, n_2, t_1, t_2) \in \mathcal{T}_s$ if it is not possible to travel from node n_1 in period t_1 to node n_2 in period t_2 in scenario s.

3.2.2 RC routing

Power cannot flow over damaged branches $B_s^d \subseteq B$. However, RCs can repair damaged branches B_s^d to make them operational again. To do so, they can travel across the transportation network to the damaged branches where they spend T^{RC} time periods to repair the damage. For modeling purposes, we may assume, without loss of generality, that RCs travel to and repair damaged branches as quickly as possible, meaning that only the order in which RCs visit damaged branches needs to be modeled. We use the travel times using damaged roads G_s^d in scenario s, and the repair time T^{RC} to pre-compute when damaged branches B_s^d become operational if a certain repair route is assigned to an RC. To do so, we introduce route indices $z \in Z_s$, where Z_s denotes all feasible orders in which an RC can visit and repair damaged branches in scenario s. The second-stage decision variable $\rho_{z,r}^s$ equals 1 if RC r is assigned to repair route z in scenario s, and 0, otherwise. The second-stage decision variable $a_{b,t}^s$ equals 1 if branch b is operational at time t in scenario s, and the parameter $R_{r,b,t,s}$ equals 1 if repair route z is such that branch b is operational at time t in scenario s if an RC is assigned to route z in scenario s. Moreover, we write $y_s \in \mathcal{Y}_s^{\text{RC}}$ if $\rho^s \in \{0,1\}^{|Z_s| \times |R|}$, and $a^s \in \{0,1\}^{|B| \times |T|}$, satisfying

$$a_{b,t}^s = 1 \qquad b \in B \setminus B_s^d,\ t \in T \tag{20a}$$

$$a_{b,t}^s = \sum_{z \in Z_s} \sum_{r \in R} R_{z,b,t,s} \rho_{z,r}^s \qquad b \in B_s^d,\ t \in T \tag{20b}$$

$$\sum_{z \in Z_s} \rho_{z,r}^s \leq 1 \qquad r \in R \tag{20c}$$

$$\rho_{z_1,r_1}^s \leq \sum_{z \geq z_1} \rho_{z,r_2}^s \qquad z_1 \in Z_s,\ r_1, r_2 \in R,\ r_1 \leq r_2 \tag{20d}$$

where Equation (20a) states that non-damaged branches are always operational. Equation (20b) states that any damaged branch b is operational at time t in scenario s if some RC assigned to repair route z has already repaired the damaged branch ($R_{z,b,t,s} = 1$). Equation (20c) states that every crew can be assigned to at most one route in scenario s. Finally, Equation (20d) is a symmetry breaking constraint that states that repair crews with a higher index must always perform a repair route with a higher index. This constraint is not required but strengthens the LP relaxation. As opposed to Anokhin et al. (2021), we use variables that represent an entire restoration route over time, instead of variables for each time period indicating what is being repaired by RCs. We do this since explicit enumeration of such routes is tractable when the number of damaged branches is small.

3.2.3 Power output constraints

The following constraints dictate how much power can be injected into the grid at every node. Moreover, they model that an MPS can only provide power at node n if it is located there and node n is equipped with an electricity adapter. Let $p^{s,+}_{n,t}, p^{s,-}_{n,t} \in \mathbb{R}_+$ denote the total real and reactive power output of MPSs at node n at time t in scenario s, respectively. Additionally, let P^+_m and P^-_m denote the maximum real and reactive output capacities of MPS m, respectively. Moreover, second-stage decision variable $c^s_{n,m,t}$ equals 1 if MPS m is connected to the electricity adapters at node n at time t in scenario s, and 0, otherwise. Moreover, we write $y_s \in \mathcal{Y}^{\text{pow}}_s$ if $c^s \in \{0,1\}^{|N^c_s| \times |M| \times |T|}$, $l^s \in \mathcal{L}_s := \{0,1\}^{|N^d_s| \times |M| \times |T|}$, $p^{s,+}, p^{s,-} \in \mathbb{R}^{|N^c_s| \times |T|}$ satisfying

$$
\begin{aligned}
& c^s_{n,m,t} \leq l^s_{n,m,t} && n \in N^c_s, m \in M, t \in T && \text{(21a)}\\
& c^s_{n,m,t} \leq x_n && n \in N^c_s, m \in M, t \in T && \text{(21b)}\\
& p^{s,+}_{n,t} \leq \sum_{m \in M} c^s_{n,m,t} \cdot P^+_m && n \in N^c_s, t \in T && \text{(21c)}\\
& p^{s,-}_{n,t} \leq \sum_{m \in M} c^s_{n,m,t} \cdot P^-_m && n \in N^c_s, t \in T, && \text{(21d)}
\end{aligned}
$$

where Equation (21a) states that an MPS can only connect to a node n if it is located there. Moreover, Equation (21b) states that an MPS m can only connect to the electricity adapter at node n if an electricity adapter is installed at node n. Furthermore, Equation (21c) and (21d) state that the total real and reactive power supplied by MPSs at time t, node n, in scenario s is bounded by the sum of of the maximum capacities of connected MPSs.

Remark 2. *Observe the following monotonicity in the second-stage value functions v_s for all $s \in S$. First, observe that v_s is non-increasing with respect to all first-stage variables since the size of the feasible region of the second-stage variables increases if x increases, see, Equation* (21b)*. Second, v_s is constant with respect to x_n for $n \in N^c \setminus N^c_s$ since no second-stage constraint contains x_n. Therefore, the feasible region of the minimization problem in v_s is unaffected by the value of x_n. Recall that it is sufficient to solely consider MPSs connecting to disconnected nodes from the substation, N^c_s, because we assume that the substation is able to provide power to all other nodes.*

3.2.4 Power balance constraints

The following second-stage variables and constraints dictate how power may flow through the DS. The second-stage decision variables $f^{s,+}_{b,t}, f^{s,-}_{b,t} \in \mathbb{R}$ denote the real and reactive power flow in branch b at time t in scenario s, respectively. Additionally, $\pi^{s,+}_{n,t}, \pi^{s,-}_{n,t} \in \mathbb{R}_+$ denote the real and reactive power loss at node n at time t in scenario s. The auxilliary variables $\phi^{s,+}_{n,t}$ and $\phi^{s,-}_{n,t}$ represent the amount of real and reactive power inflow and outflow of node n at time t in scenario s, respectively. We omit them in the implementation but consider them in equations below to write some equations more compactly. Moreover, $o^{s,+}_t, o^{s,-}_t \in \mathbb{R}_+$ denote the real and reactive power output of the substation at time t in scenario s. Moreover, we write $y_s \in \mathcal{Y}^{\text{flow}}_s$ if $f^{s,+}, f^{s,-} \in \{0,1\}^{|B|\times|T|}$, $\pi^{s,+}, \pi^{s,-} \in \{0,1\}^{|N|\times|T|}$, $o^{s,+}, o^{s,-} \in \{0,1\}^{|T|}$, $p^{s,+}, p^{s,-} \in \mathbb{R}^{|N^c_s|\times|T|}$, $a^s \in \{0,1\}^{|B|\times|T|}$, satisfying

$$\phi^{s,+}_{n,t} = \sum_{b\in B^{\text{OUT}}_n} f^{s,+}_{b,t} + (D^+_n - \pi^{s,+}_{n,t}) \qquad n \in N,\ t \in T \tag{22a}$$

$$\phi^{s,-}_{n,t} = \sum_{b\in B^{\text{OUT}}_n} f^{s,-}_{b,t} + (D^-_n - \pi^{s,-}_{n,t}) \qquad n \in N,\ t \in T \tag{22b}$$

$$\phi^{s,+}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,+}_{b,t} + o^{s,+}_t \qquad n \in \{n^1\},\ t \in T \tag{22c}$$

$$\phi^{s,+}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,+}_{b,t} + p^{s,+}_{n,t} \qquad n \in N^c_s \setminus \{n^1\},\ t \in T \tag{22d}$$

$$\phi^{s,+}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,+}_{b,t} \qquad n \in N \setminus \{N^c_s \cup \{n^1\}\},\ t \in T \tag{22e}$$

$$\phi^{s,-}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,-}_{b,t} + o^{s,-}_t \qquad n \in \{n^1\},\ t \in T \tag{22f}$$

$$\phi^{s,-}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,-}_{b,t} + p^{s,-}_{n,t} \qquad n \in N^c_s \setminus \{n^1\},\ t \in T \tag{22g}$$

$$\phi^{s,-}_{n,t} = \sum_{b\in B^{\text{IN}}_n} f^{s,-}_{b,t} \qquad n \in N \setminus \{N^c_s \cup \{n^1\}\},\ t \in T \tag{22h}$$

$$0 \le \pi^{s,+}_{n,t} \le D^+_n \qquad n \in N,\ t \in T \tag{22i}$$

$$(D^-_n - \pi^{s,-}_{n,t}) \le (D^+_n - \pi^{s,+}_{n,t})\frac{D^-_n}{D^+_n} \qquad n \in N \setminus \{n^1\},\ t \in T \tag{22j}$$

$$-a^s_{b,t}F^+_b \le f^{s,+}_{b,t} \le a^s_{b,t}F^+_b \qquad b \in B,\ t \in T \tag{22k}$$

$$-a^s_{b,t}F^-_b \le f^{s,-}_{b,t} \le a^s_{b,t}F^-_b \qquad b \in B,\ t \in T \tag{22l}$$

$$o^{s,+}_t \le O^+ \qquad t \in T \tag{22m}$$

$$o^{s,-}_t \le O^- \qquad t \in T \tag{22n}$$

where $B^{\text{IN}}_n, B^{\text{OUT}}_n \subset B$ denote the branches that flow into and out of node n, respectively. Moreover, n^1 denotes the node in the DS where the substation is located. Equations (22a) and (22b) state that $\phi^{s,+}_{n,t}$ and $\phi^{s,-}_{n,t}$ represent the amount of real and reactive power outflow of node n at time t in scenario s, respectively. Equations (22c)-(22e) enforce Kirchhoff's first law which stipulates that the total incoming power flow must be equal to the total outgoing power flow for real power, since the right-hand sides of these equations define the total power inflow at all nodes. That is, first, at all nodes power may flow from incoming branches B^{IN}_n via $f^{s,+}_{b,t}$. Second, at nodes in N^c_s, MPSs may supply power via $p^{s,+}_{n,t}$. Third, at node n^1, where the substation is located, the substation supplies power via $o^{s,+}_l$. Equations (22f)-(22h) state the equivalent of Equations (22c)-(22e) for reactive power flow. The parameters D^+_n and D^-_n denote the demand for real and reactive power at node n, respectively. Equation (22i) ensures that the power loss in all nodes is bounded by their demands. The relationship between the restored real and restored reactive load is defined by Equation (22j). The real and reactive power flows in online branches are limited by their real and reactive power capacities F_b^+ and F_b^-, respectively. This is enforced in Equation (22k) and Equation (22l). Equation (22m) and Equation (22n) enforce that the real and reactive power generation at the substation node are bounded by their maximum capacities, O^+ and O^-, respectively.

3.2.5 Power flow constraints

These constraints dictate how the squared voltage magnitude affects the power flow. The minimum and maximum squared voltage at node n are denoted by $V_n^{\min}$ and $V_n^{\max}$, respectively. Moreover, the resistance and reactance of branch b are denoted by R_b^+ and R_b^-, respectively. The second-stage decision variable $v_{n,t}^s$ denotes the squared voltage magnitude at node n at time t in scenario s. Moreover, we write $y_s \in \mathcal{Y}_s^{\text{volt}}$ if $\nu^s \in \mathcal{V}_s := \{0,1\}^{|N|\times|T|}$, $f^{s,+}, f^{s,-} \in \{0,1\}^{|B|\times|T|}$, and $a^s \in \{0,1\}^{|B|\times|T|}$, satisfying

$$\nu_{o(b),t}^s - \nu_{d(b),t}^s \leq (1 - a_{b,t}^s)U + 2\big(R_b^+ f_{b,t}^{s,+} + R_b^- f_{b,t}^{s,-}\big) \qquad b \in B, t \in T, s \in S \tag{23a}$$

$$\nu_{o(b),t}^s - \nu_{d(b),t}^s \geq (a_{b,t}^s - 1)U + 2\big(R_b^+ f_{b,t}^{s,+} + R_b^- f_{b,t}^{s,-}\big) \qquad b \in B, t \in T, s \in S \tag{23b}$$

$$V_n^{\min} \leq \nu_{n,t}^s \leq V_n^{\max} \qquad n \in N, t \in T, s \in S \tag{23c}$$

where Equation (23a) and (23b) state that if branch b is damaged (not connected) at time t in scenario s, then the difference in squared voltage between the two nodes at the start of branch b, $o(b)$, and at the end of branch b, $d(b)$, is between $-U$ and U. That is, the squared voltage is bounded by the negative of the maximum squared voltage and the positive of the maximum squared voltage, because $2(R_b^+ f_{b,t}^{s,+} + R_b^- f_{b,t}^{s,-}) = 0$ when $a_{b,t}^s = 0$ due to equation (22k) and (22l). Additionally, if branch b is not damaged (i.e., remains connected and online) at time t in scenario s, then $a_{b,t}^s = 1$ and the inequalities (23a) and (23b) collapse into an equality, where the voltages at the two ends become physically coupled, simply representing the linearized voltage-drop equation across the branch. The bounds $V_n^{\max}$ and $V_n^{\min}$ of the squared voltage magnitude at node n at time t in scenario s are enforced in Equation (23c).

3.3 Experimental Design

We evaluate the proposed DSR model on the well-known IEEE 33-bus distribution system, see, e.g., Anokhin et al. (2021) or Figure 1 in Section 1, paired with a road transportation network designed to match this system (Li et al., 2022). The distribution system has a radial structure, meaning that a single branch failure disconnects all downstream consumers from the substation. We identify 13 operationally significant nodes as candidates $N^c \subset N$ for electricity adapter installation. These are nodes: 2, 4, 7, 9, 12, 14, 17, 19, 22, 24, 27, 29, and 32. The planning horizon consists of $T = 36$ discrete time periods of 10 minutes each, representing a 6-hour restoration window. This window ensures that on average over all extreme-weather scenarios, more than 99% of the power demand is restored at the end of the 6 hour restoration window.

Extreme weather scenarios $s \in S$ are generated by sampling three damaged distribution branches and three damaged road segments uniformly at random. This is a realistic assumption when, e.g., the wind speeds of a hurricane are similar across the entire network (Javanbakht and Mohagheghi, 2014), which may be reasonable when the nodes in the network are geographically close together. To ensure that RCs and MPSs can reach all nodes, we enforce a reachability constraint: all inter-node travel times must be finite. Scenarios in which road damage renders some node unreachable are resampled. In the base case, we assume the following for the parameters. First, a budget of $X^{\max} = 5$ for electricity adapters. Second, three MPS units, each providing 400 normalized units of real power and 300 normalized units of reactive power. Third, one repair crew with a fixed repair duration of 60 minutes (six time periods) per branch. Fourth, the starting locations of MPSs and RCs, n_1^{MPS} and n_1^{RC}, coincide with the node n^1 where the substation is located. Fifth, we consider $|S| = 100$ scenarios in total. The complete dataset is available at the Github referenced in Section 6.

We solve the DSR problem using our proposed monotone DW decomposition and compare three variants based on different NAC types: $\text{DW}_=$, $\text{DW}_\leq$, and $\text{DW}_{\bowtie^\star}$. These correspond to the operators types used in (6), where $\text{DW}_=$ and $\text{DW}_\leq$ correspond to using $\bowtie_s$ equal to $(=)_{i\in N}$ and $(\leq)_{i\in N}$ for all $s \in S$, and $\text{DW}_{\bowtie^\star}$ corresponds to using most relaxed NACs as determined by the sufficiency condition in Proposition 2. All three variants are embedded in branch-and-bound frameworks, see, e.g., Kim and Dandurand (2022). We implement the DW algorithm asynchronously parallel, see, e.g., Basso and Ceselli (2022). To ensure initial feasibility of the RMP (8), we include columns corresponding to $x^0 = \mathbf{0}$ (no adapters installed), such that the RMP always admits a feasible solution. Moreover, in $\text{DW}_=$ we also include a new feasible solution whenever branching results in an infeasible node. For $\text{DW}_\leq$, and $\text{DW}_{\bowtie^\star}$ this is never required by design of x^0.

First, we benchmark the DW algorithms against the large-scale deterministic equivalent formulation (LSDE) which solves the SMIP (1) directly. Second, we benchmark the stochastic optimization policy (SOP) obtained by solving the full DSR formulation of Section 3.2 against the Highest Loads Policy (HLP), a heuristic that allocates electricity adapters to the nodes with the highest individual power demands, see, e.g., Li et al. (2022).

All experiments were performed on the high-performance computing cluster Hábrók of the University of Groningen, using 20 cores of an AMD EPYC 7763 CPU with a clock speed of 2.45–3.5 GHz. In the DW algorithms, one core is assigned to the RMP (8) and 19 cores to the pricing problems (9). All mathematical programs are solved using Gurobi 11.0.0.

4 Numerical Results

We provide numerical results on the computational efficiency of our *monotone* DW algorithm in Section 4.1. Furthermore, we provide practical insights into the DSR problem in Section 4.2.

Algorithm	$\lvert N^c\rvert = 8$					$\lvert N^c\rvert = 10$					$\lvert N^c\rvert = 13$					Average
	$\lvert S\rvert$					$\lvert S\rvert$					$\lvert S\rvert$					
	20	40	60	80	100	20	40	60	80	100	20	40	60	80	100	
LSDE	313	1,622	> 13,867	> 14,400	> 14,400	1,812	> 14,400	> 14,400	> 14,400	> 14,400	> 14,400	> 14,400	> 14,400	> 14,400	> 14,400	> 11,734
$\mathrm{DW}_=$	464	1,953	2,721	4,177	9,197	1,413	5,435	10,163	> 14,400	> 14,400	2,985	> 13,296	> 14,400	> 14,400	> 14,400	> 8,254
$\mathrm{DW}_\leq$	160	317	**323**	327	755	**424**	**1,103**	1,855	4,465	7,578	**1,125**	3,782	4,948	4,685	7,137	2,599
$\mathrm{DW}_{\bowtie^\star}$	**159**	**231**	352	**326**	**750**	465	1,318	**1,776**	**3,693**	**6,783**	1,346	**3,400**	**4,550**	**4,541**	**6,348**	**2,403**

Table 1: Average computation time (in seconds) of the LSDE, $\mathrm{DW}_=$, $\mathrm{DW}_\leq$, and $\mathrm{DW}_{\bowtie^\star}$ over 5 repeated simulations for each instance. If the computation time exceeded 14,400 seconds (4 hours) in any simulation of an instance with a method, '>' indicates that the average computation time over all selected simulations of this instance and method exceeds the specified time. The final column contains the average computation times over all instances for each method.

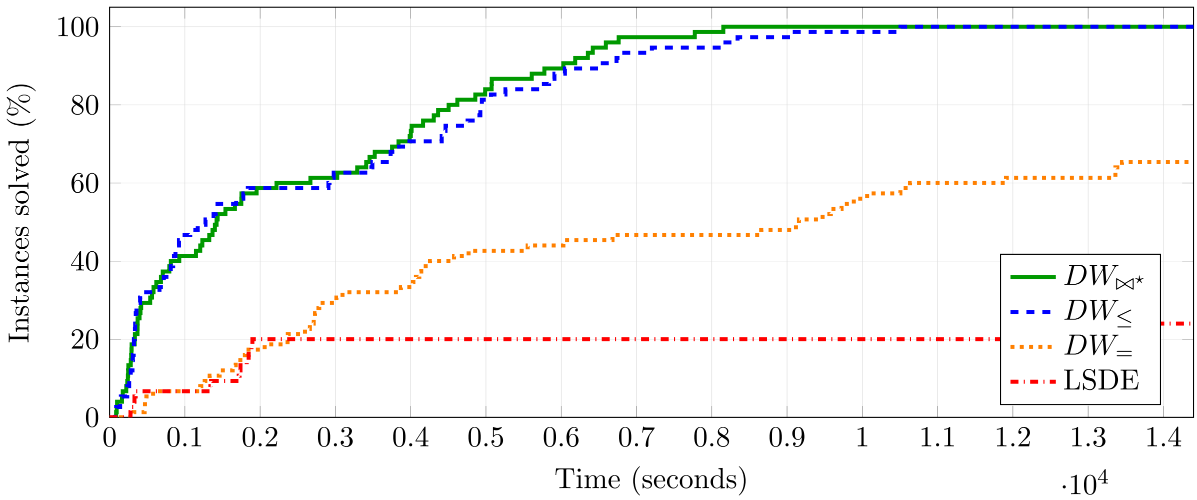


Figure 3: Percentage of DSR instances of Table 1 solved over time (in seconds) of $DW_{\bowtie^\star}$ (green, solid), $DW_\leq$ (blue, dashed), $DW_=$ (orange, dotted), and LSDE (red, dash-dotted). Each instance is repeated five times for each method.

4.1 Algorithmic Efficiency of the Monotone DW Algorithm

In this section, we compare the computational performance of $\mathrm{DW}_=$, $\mathrm{DW}_\leq$, and $\mathrm{DW}_{\bowtie^\star}$ as defined in Section 3.3. Table 1 reports computation times for the base case DSR instance, varying the number of scenarios $|S|$ and the number of candidate nodes $|N^c|$.

The following observations stand out. First, the LSDE is unable to solve many instances within 4 hours, in particular the instances comprising more than 40 scenarios. Second, $\mathrm{DW}_=$ is substantially slower than the other two DW algorithms across all instances. Its computation time exceeds 4 hours in all 5 repetitions of five instances, whereas $\mathrm{DW}_\leq$ and $\mathrm{DW}_{\bowtie^\star}$ manage to solve all instances within 7,578 and 6,783 seconds on average, respectively. Most strikingly, $\mathrm{DW}_{\bowtie^\star}$ solves the instance with $|S| = 100$ and $|N^c| = 8$, 92% faster on average than $\mathrm{DW}_=$, and the instance with $|S| = 80$ and $|N^c| = 8$, 98% faster on average than the LSDE. Third, the computation time of $\mathrm{DW}_\leq$ and $\mathrm{DW}_{\bowtie^\star}$ are within the same order of magnitude for all considered instances. However, algorithm $\mathrm{DW}_{\bowtie^\star}$ outperforms $\mathrm{DW}_\leq$ by 7.5%, on average over all instances. We also visualize the performance of these methods in the performance plot in Figure 3 below.

Figure 3 shows the percentage of instances of Table 1 solved within a certain amount of

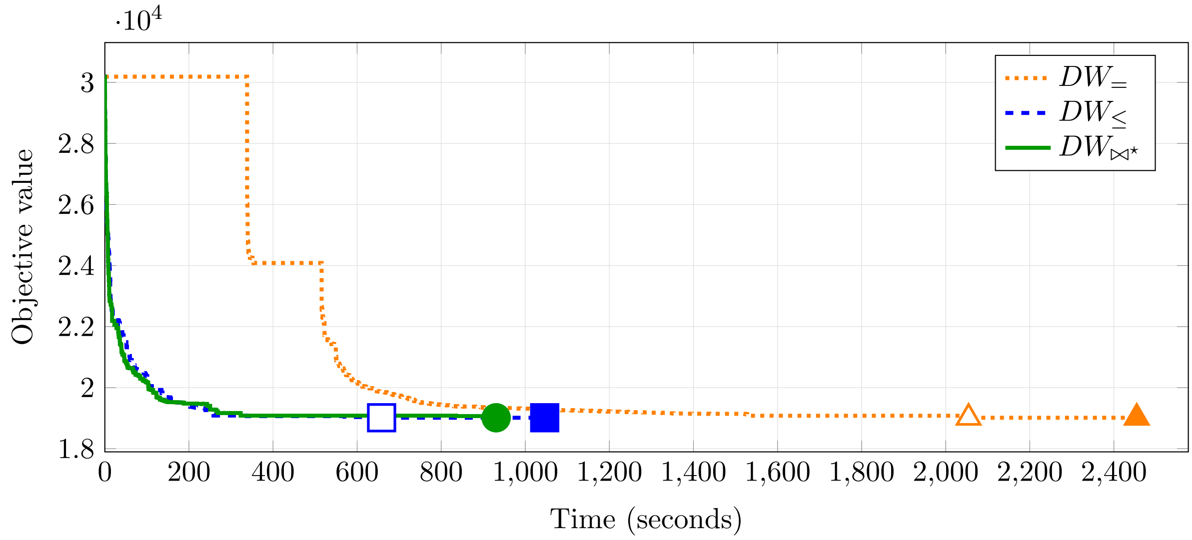


Figure 4: Optimal objective value over time of the RMP (8) for $\mathrm{DW}_=$ (orange, dotted, triangle), $\mathrm{DW}_\leq$ (blue, dashed, square), and $\mathrm{DW}_{\bowtie^\star}$ (green, solid, circle) on the instance with $|S| = 20$ and $|N^c| = 13$. Hollow and solid markers denote when the optimal solutions are found for the first time, and proven to be optimal, respectively.

time. Each instance is solved five times to reduce the variance of the results. Observe again, that both $DW_\leq$, and $DW_{\bowtie^\star}$ are able to solve significantly more instances than the LSDE and the traditional DW formulation with equality NACs $DW_=$ within the same amount of time.

To highlight the differences between the three DW algorithms, we present the objective values of the RMP of these algorithms over time for one DSR instance, in Figure 4.

Two observations follow. First, the objective value of $\mathrm{DW}_=$ remains constant for the first 350 seconds, and again later from 350-500 seconds. As discussed in Section 2.3, this is a consequence of the highly restrictive effective domain in RMP (10). Recall that the effective domain in RMP (10) is given by $\bigcap_{s\in S} \mathrm{conv}(\{x_s^1, ..., x_s^{|K_s|}\}$ when all NACs are equality constraints, i.e., $\bowtie_s$ equals $(=)_{i\in N}$ for all $s \in S$. While the $\mathrm{DW}_=$ iteratively includes many new extreme points (first-stage decision variables x_s^k and their corresponding second-stage cost $v_s(x_s^k)$) in the RMP (8), this intersection $\bigcap_{s\in S} \mathrm{conv}(\{x_s^1, ..., x_s^{|K_s|}\})$ remains equal to its initial solution $x^0 = (0, ..., 0)$ in early iterations of the algorithm. Therefore, the optimal solution also remains equal. The algorithms $\mathrm{DW}_\leq$ and $\mathrm{DW}_{\bowtie^\star}$ do not exhibit this behavior as a consequence of their relaxed NACs. In fact, since they also start with initial solution $x^0 = (0, ..., 0)$, the effective domains in RMP in (10) with relaxed NACs $\bowtie_s = (\leq)_{i\in N}$ or $\bowtie_s = \bowtie_s^\star$ for all $s \in S$, are already at their maximum size when the initial solution $x^0 = (0, ..., 0)$ is included in RMP (8). That is, their effective domains are both equal to the $\mathrm{conv}(\mathcal{X})$. This is the case because for any $\hat{x} \in \mathrm{conv}(\mathcal{X})$, the solution $(\hat{x}, x_1, ..., x_S) = (\hat{x}, x^0, ..., x^0)$ of RMP (8) is a feasible solution when the NACs state that $x_s \leq \hat{x}$ or $x_s \square \hat{x}$, thus this solution has a finite objective value, and consequently, $\hat{x}$ is in the effective domain in RMP (10). As a result, any columns that are included to the RMP in (8) with relaxed NACs, even in early iterations, can always be used to generate a new solution of RMP (8).

Second, the objective values of $\mathrm{DW}_\leq$ and $\mathrm{DW}_{\bowtie^\star}$ are broadly similar over time. However,

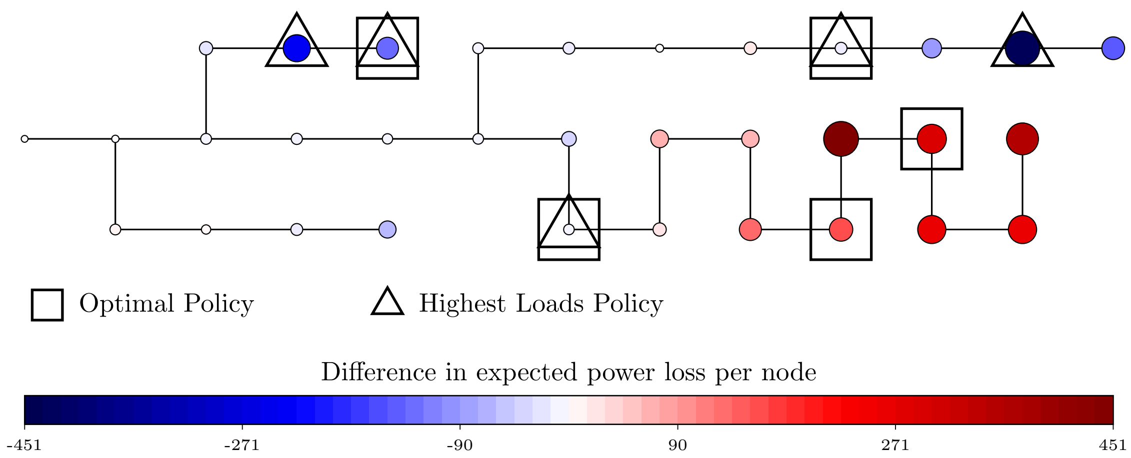


Figure 5: Expected power loss difference per node between the stochastic optimization policy (SOP) and the Highest Loads Policy (HLP) for the base case instance. Red (blue) nodes have lower (higher) expected power loss under the SOP. Node size reflects the magnitude of the difference in expected power loss per node between the SOP and HLP. Squares and triangles mark nodes where electricity adapters are installed under the SOP and HLP, respectively.

while $\mathrm{DW}_{\bowtie^\star}$ finds the optimal solution later than $\mathrm{DW}_{\leq}$, it is able to prove optimality of this solution faster than $\mathrm{DW}_{\leq}$. In fact, $\mathrm{DW}_{\bowtie^\star}$ was able to prove optimality immediately after finding the optimal solution, which is why the hollow green circle marker is not observed. This can be explained by the fact that the MP corresponding to $\mathrm{DW}_{\bowtie^\star}$ has more dominated columns than the MP corresponding to $\mathrm{DW}_{\leq}$, which may trigger the termination criterion sooner, as explained in Section 2.3.

4.2 Managerial Insights for the DSR Problem

4.2.1 Structural Limitations of Demand-Based Adapter Placement

We compare the optimal adapter placements and expected power loss of the stochastic optimization policy (SOP), and the highest-loads policy (HLP) for the base case instance with 13 candidate nodes for electricity adapters. Figure 5 visualizes both solutions on the IEEE 33-bus test system and highlights the per-node difference in power loss. The comparison reveals three structural advantages of optimization-based adapter placement, such as the SOP, over demand-greedy heuristics like the HLP. As a result, the SOP has 4% less expected power loss overall.

The first advantage is as follows. The HLP evaluates nodes in isolation, ignoring that an adapter serves all nodes in its disconnected region, not just the node at which it is installed. In radial networks, a single upstream failure can disconnect an entire subtree whose aggregate demand may be large, even if no individual node ranks among the highest loads. The SOP accounts for this by placing adapters in high-aggregate-demand regions that the HLP leaves uncovered. In Figure 5, this effect is most visible in the subtree with red nodes, where the SOP achieves substantially lower expected power loss by installing more electricity adapters in that region.

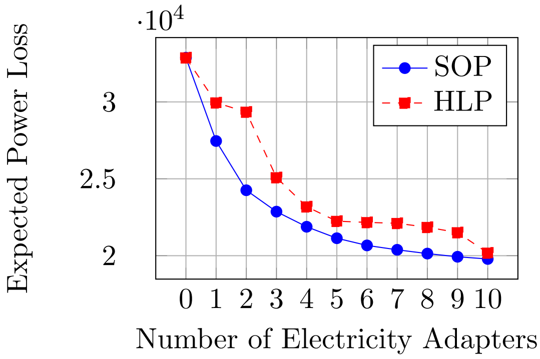


Figure 6: Expected power loss vs. number of electricity adapters for the SOP(blue, solid) and the HLP(red, dashed).

Second, even when two candidate nodes can serve the same disconnected region, they may differ in how quickly an MPS can reach them. The SOP exploits this by preferring nodes with shorter travel times from the MPS depot. More generally, optimization-based placement jointly considers network topology and travel distances, whereas demand-based heuristics ignore accessibility entirely.

Third, the SOP distributes adapters across topologically distinct parts of the network, favouring downstream regions that are most vulnerable to disconnection. This spatial spread reduces the risk that a damage scenario leaves entire regions without adapter access. It also complements the natural repair strategy: repair crews typically restore upstream branches first, allowing the substation to re-energize large portions of the grid, while MPSs — via downstream adapters — provide temporary power for the isolated consumers in downstream areas. In contrast, demand-based heuristics tend to cluster adapters in high-load areas, leaving other regions without backup power access.

4.2.2 Diminishing Returns of Adapter Investment

Figure 6 shows the expected power loss as a function of the number of electricity adapters for the base-case DSR instance, under both the SOP and the HLP. First of all, we observe that installing adapters may reduce expected power loss by up to 40% compared to installing no electricity adapters.

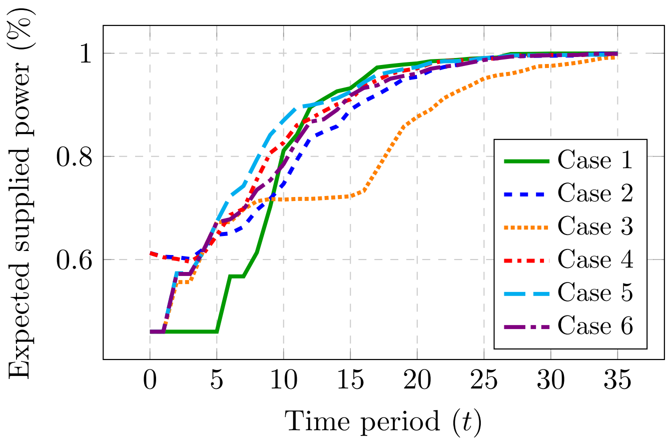
Second, the expected power loss of the SOP exhibits clear diminishing marginal returns with respect to the number of electricity adapters that may be installed. The first few adapters yield the largest reductions in expected power loss, as they grant MPS access to high-aggregate-demand regions that are frequently disconnected after a disaster. Once these high-value regions are covered, additional adapters primarily serve as alternative access points that reduce MPS travel time, offering incrementally smaller improvements. A moderate adapter budget therefore already captures most of the attainable resilience benefit.

The HLP does not exhibit the same monotone diminishing-returns pattern in expected power loss. Because the HLP assigns adapters to nodes with the highest loads irrespective of their location in the DS, it may place several adapters in the same region of the DS, yielding only marginal reductions in expected power loss. Subsequently, adapters may be installed in more isolated regions of the DS, and thus produce larger improvements in expected power loss. This mechanism explains why sharp drops follow after shallow decreases in expected power loss, in Figure 6 between 2 and 3 adapters, and again between 9 and 10 adapters.

Finally, observe that the SOP clearly outperforms the HLP in all 11 instances, irrespective of the number of electricity adapters that may be installed. On average over these 11 instances, the SOP reduces the expected power loss of the HLP by 7.0%. Additionally, there is a large variance in the difference in expected power loss between the SOP and HLP, which means that using the HLP has a risk of extremely poor performance. In particular, in the instance with 2 charging stations, the SOP reduces the expected power loss of the HLP by 17.3%.

4.2.3 Joint Placement of MPSs and RCs Depots Determines Expected Power Loss

We examine how depot locations of MPSs and RCs affect expected power loss by considering six configurations, in which depots are located at the substation (node 0), centrally (node 5), or downstream (node 32). Cases 1–3 co-locate both depots at the same node, while Cases 4–6 separate them.



Case	MPS Depot	RC Depot
1	Substation	Substation
2	Central	Central
3	Downstream	Downstream
4	Central	Substation
5	Downstream	Substation
6	Downstream	Central

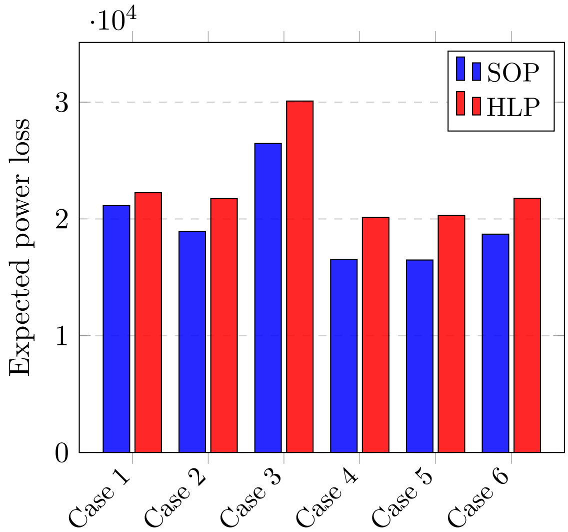
Figure 7: The percentage of expected power demand supplied over time.

Figure 7 shows the expected service (power) recovery over time for optimal policies of these six instances, suggesting the following two findings. First, the RC depot location seems to determine long-run restoration quality: Cases 1, 4, and 5 all have the RC depot at the substation node and consistently achieve the highest recovery from $t = 10$ onward. When the RC depot is placed downstream (Case 3), RCs tend to travel to upstream branches, such that they can enable the substation to re-energize the network beyond those damaged branches. This causes a slower restoration process, and thus a lower percentage of expected power demand met from period $t = 10$ onward.

Second, the MPS depot location mainly determines early-period supply: a central or downstream depot (Cases 2-6) allows MPSs to begin serving disconnected loads within two periods, whereas a substation depot (Case 1) introduces a delay of 6 periods until the MPS supply backup power. Cases 4 and 5 achieve the best overall outcome by combining both advantages: the MPS depot is close to outage areas, allowing MPSs to begin serving loads from $t = 1$ and $t = 2$, respectively, while the RC depot is located upstream at the substation, allowing RCs to restore upstream branches efficiently, which yields the highest cumulative power recovery across the horizon since the substation at node 0 can supply power to these newly connected nodes. The mixed configurations in Cases 4–6 all outperform Case 1 and 3, which involve an MPS too far from the outage area or an RC too far from the upstream areas, respectively.

Overall, these results show that MPS and RC depot placement have complementary and significant effects on expected power loss: the MPS depot should be positioned near probable outage regions to minimize supply delays, while the RC depot should be positioned upstream to enable fast network restoration.

In Figure 8 below, we present the expected power loss under the SOP and the HLP for the six cases in consideration. Observe that the SOP outperforms the HLP significantly in all six instances. In particular, the percentage difference in expected power loss between the SOP and HLP is large when the MPS depot is located at a central or downstream node (Cases 2-6). This can be explained by the fact that nodes in these regions are disconnected from the substation. As a result, the travel times of MPSs to such regions is short, and thus, MPSs can supply backup power in more periods before all damaged branches are repaired by RCs. Therefore, the benefit of installing electricity adapters optimally, is larger. In the six cases, the SOP reduces the expected power loss between 5% and 19% compared to the HLP with an average of 13.5%.



Case	MPS Depot	RC Depot
1	Substation	Substation
2	Central	Central
3	Downstream	Downstream
4	Central	Substation
5	Downstream	Substation
6	Downstream	Central

Figure 8: Expected power loss under SOP and HLP.

5 Conclusion

We propose a Dantzig–Wolfe decomposition algorithm that exploits monotone structure in second-stage value functions by relaxing selected non-anticipativity constraints. We prove that the relaxed formulation preserves optimality while substantially accelerating convergence, reducing computation times by up to 92% and 98% compared to the traditional DW algorithm and the LSDE, respectively. A key practical advantage is that the relaxed algorithm avoids the stalling behaviour of standard DW, where the restrictive domain of the approximated value functions can prevent any improvement over the initial solution for extended time. We also introduce a simple procedure to determine the types of monotonicity that the second-stage value functions exhibit. This may help researchers to identify whether the monotone DW algorithm is applicable for given SMIPs.

Furthermore, we propose a novel two-stage stochastic programming formulation for the DSR problem, which we solve to optimality using our newly developed monotone DW algorithm combined with a standard branch-and-bound procedure. We implement our DW algorithm asynchronously parallel using 20 cores. Applying the proposed algorithm to the DSR problem yields the following novel insights into optimal policies for the DSR problem. Electricity adapters should be distributed across topologically distinct parts of the grid, with a preference for downstream locations. This placement complements the natural repair strategy in which repair crews restore upstream branches to re-energize the grid from the substation, while mobile power sources supply downstream regions that remain disconnected during repairs. Compared to installing no adapters, the optimal placement reduces expected power loss by up to 40%. Compared to the demand-based Highest Loads Policy, the optimal policy reduces power loss by up to 19%. This highlights the importance of accounting for aggregate regional demand, stochastic damage patterns, and travel times of MPSs, in making investment decision on where electricity adapters must be installed.

Several directions for future research emerge. On the methodological side, relaxed NACs could be explored in other settings, such as risk-averse optimization, multi-stage SMIPs, and distributionally robust optimization. On the application side, the DSR model could be extended to other (larger-scale) distribution systems and transportation networks. Finally, this research could be extended with alternative technologies such as mobile energy storage systems, switchable distribution lines, or investments in more robust underground electricity branches.

6 Code and Data Disclosure

The code and data to support the numerical experiments in this paper can be found at github.com/LaurensElderhorst/monotone_dw_and_distribution_system_resilience_paper.

7 Acknowledgements

The research of Laurens Elderhorst and Ward Romeijnders has been supported by the Netherlands Organisation for Scientific Research [Grant VI.Vidi.211.066]. Moreover, we thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high performance computing cluster.

References

Achterberg, T., Bixby, R. E., Gu, Z., Rothberg, E., and Weninger, D. (2020). Presolve reductions in mixed integer programming. *INFORMS Journal on Computing*, 32(2):473–506.

Anokhin, D., Dehghanian, P., Lejeune, M. A., and Su, J. (2021). Mobility-as-a-service for resilience delivery in power distribution systems. *Production and Operations Management*, 30(8):2492–2521.

Basciftci, B., Ahmed, S., and Gebraeel, N. (2024). Adaptive two-stage stochastic programming with an analysis on capacity expansion planning problem. *Manufacturing & Service Operations Management*, 26(6):2121–2141.

Basso, S. and Ceselli, A. (2022). Distributed asynchronous column generation. *Computers & Operations Research*, 146:105894.

Beltran, C., Tadonki, C., and Vial, J. P. (2006). Solving the p-median problem with a semi-lagrangian relaxation. *Computational Optimization and Applications*, 35(2):239–260.

Birge, J. R. and Louveaux, F. (1997). *Introduction to stochastic programming.* Springer.

Carøe, C. C. and Schultz, R. (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1-2):37–45.

de Queiroz, A. R. (2016). Stochastic hydro-thermal scheduling optimization: An overview. *Renewable and Sustainable Energy Reviews*, 62:382–395.

Desrosiers, J. and Lübbecke, M. E. (2005). A primer in column generation. In *Column generation*, pages 1–32. Springer.

Hafiz, F., de Queiroz, A. R., Fajri, P., and Husain, I. (2019). Energy management and optimal storage sizing for a shared community: A multi-stage stochastic programming approach. *Applied energy*, 236:42–54.

Javanbakht, P. and Mohagheghi, S. (2014). A risk-averse security-constrained optimal power flow for a power grid subject to hurricanes. *Electric Power Systems Research*, 116:408–418.

Kim, K. and Dandurand, B. (2022). Scalable branching on dual decomposition of stochastic mixed-integer programming problems. *Mathematical Programming Computation*, 14(1):1–41.

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on optimization*, 12(2):479–502.

Laporte, G. and Louveaux, F. V. (1993). The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142.

Li, C., Liu, F., Cao, H., and Wang, Q. (2009). A stochastic dynamic programming based model for uncertain production planning of re-manufacturing system. *International Journal of Production Research*, 47(13):3657–3668.

Li, Z., Tang, W., Lian, X., Chen, X., Zhang, W., and Qian, T. (2022). A resilience-oriented two-stage recovery method for power distribution system considering transportation network. *International Journal of Electrical Power & Energy Systems*, 135:107497.

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., et al. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific reports*, 9(1):16063.

Parada, L., Legault, R., Côté, J.-F., and Gendreau, M. (2024). A disaggregated integer L-shaped method for stochastic vehicle routing problems with monotonic recourse. *European Journal of Operational Research*, 318(2):520–533.

Park, H. and Baldick, R. (2015). Stochastic generation capacity expansion planning reducing greenhouse gas emissions. *IEEE Transactions on Power Systems*, 30(2):1026–1034.

Philpott, A. B., Wahid, F., and Bonnans, J. F. (2020). Midas: A mixed integer dynamic approximation scheme. *Mathematical Programming*, 181(1):19–50.

Romeijnders, W., Zhang, Y., and Sen, S. (2026). Decomposition algorithms for computational stochastic mixed-integer programming: A survey. *European Journal of Operational Research*, 335(2):335–353.

Saadouli, H., Jerbi, B., Dammak, A., Masmoudi, L., and Bouaziz, A. (2015). A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers & Industrial Engineering*, 80:72–79.

Shapiro, A., Dentcheva, D., and Ruszczynski, A. (2021). *Lectures on stochastic programming: modeling and theory*. SIAM.

Shi, Q., Liu, W., Zeng, B., Hui, H., and Li, F. (2022). Enhancing distribution system resilience against extreme weather events: Concept review, algorithm summary, and future vision. *International Journal of Electrical Power & Energy Systems*, 138:107860.

Singh, K. J., Philpott, A. B., and Wood, R. K. (2009). Dantzig-Wolfe decomposition for solving multistage stochastic capacity-planning problems. *Operations Research*, 57(5):1271–1286.

Stewart Jr, W. R. and Golden, B. L. (1983). Stochastic vehicle routing: A comprehensive approach. *European Journal of Operational Research*, 14(4):371–385.

Su, J., Anokhin, D., Dehghanian, P., and Lejeune, M. A. (2023). On the use of mobile power sources in distribution networks under endogenous uncertainty. *IEEE transactions on control of network systems*, 10(4):1937–1949.

Su, J., Zhang, R., Dehghanian, P., Kapourchali, M. H., Choi, S., and Ding, Z. (2024). Renewable-dominated mobility-as-a-service framework for resilience delivery in hydrogen-accommodated microgrids. *International Journal of Electrical Power & Energy Systems*, 159:110047.

Sun, W., Wang, Y., Hao, Y., Alharthi, Y. Z., and Wang, Y. (2025). A resilience-oriented optimization framework for smart grid operation and recovery before, during, and after natural disasters. *Applied Energy*, 398:126414.

Tuy, H., Minoux, M., and Hoai-Phuong, N. (2006). Discrete monotonic optimization with application to a discrete location problem. *SIAM Journal on Optimization*, 17(1):78–97.

van der Laan, N. and Romeijnders, W. (2024). A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models. *Operations Research*, 72(5):2190–2214.

Vanderbeck, F. and Savelsbergh, M. W. (2006). A generic view of dantzig–wolfe decomposition in mixed integer programming. *Operations Research Letters*, 34(3):296–306.

Wolsey, L. A. and Nemhauser, G. L. (1999). *Integer and combinatorial optimization*. John Wiley & Sons.

Zou, J., Ahmed, S., and Sun, X. A. (2019). Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502.

8 Appendix: Proofs

Proof of Proposition 1. Since the SMIP with relaxed NACs in (4) is a relaxation of the SMIP with traditional NACs in (3), it follows immediately that $z^\star_{\text{SV-R}} \le z^\star_{\text{SV}}$. To prove that $z^\star_{\text{SV-R}} \ge z^\star_{\text{SV}}$ as well, let $(\hat{x}^\star, x^\star_1, \ldots, x^\star_{|S|})$ denote an optimal solution of the SMIP with relaxed NACs in (4) with optimal objective value $z^\star_{\text{SV-R}} = c^\top \hat{x}^\star + \sum_{s\in S} p_s v_s(x^\star_s)$. We will show that $(\hat{x}^\star, \hat{x}^\star, \ldots, \hat{x}^\star)$ is a feasible solution of the SMIP with traditional NACs in (3) with an objective value of at most $c^\top \hat{x}^\star + \sum_{s\in S} p_s v_s(x^\star_s)$.

Observe that, first, $\hat{x} \in \mathcal{X}$ since $(\hat{x}^\star, x^\star_1, \ldots, x^\star_{|S|})$ denotes an optimal solution of the SMIP with relaxed NACs in (4). Second, $x_s = \hat{x}$ holds for all $s \in S$ in solution $(\hat{x}^\star, \hat{x}^\star, ..., \hat{x}^\star)$. Thus, $x_s \in \mathcal{X}$ for all $s \in S$. It follows that $(\hat{x}^\star, \hat{x}^\star, \ldots, \hat{x}^\star)$ is indeed a feasible solution of the SMIP with traditional NACs in (3).

Moreover, since $\bowtie_s \in \Theta_s$ for all $s \in S$, it follows that $v_s(x^1_s) \ge v_s(x^2_s)$ for all $x^1_s, x^2_s \in \mathcal{X}$ with $x^1_s \bowtie_s x^2_s$. Additionally, since we also have that the optimal solution $(\hat{x}^\star, x^\star_1, \ldots, x^\star_{|S|})$ of the SMIP with relaxed NACs in (4) satisfies $x^\star_s \bowtie_s \hat{x}^\star$ for all $s \in S$, it follows that $v_s(x^\star_s) \ge v_s(\hat{x}^\star)$ for all $s \in S$. Thus, the objective of the feasible solution $(\hat{x}^\star, \hat{x}^\star, \ldots, \hat{x}^\star)$ of the SMIP with traditional NACs in (3) is $c^\top \hat{x}^\star + \sum p_s v_s(\hat{x}^\star) \le c^\top \hat{x}^\star + \sum_{s\in S} p_s v_s(x^\star_s) = z^\star_{\text{SV-R}}$. We have now shown that $z^\star_{\text{SV-R}} \le z^\star_{\text{SV}}$ and $z^\star_{\text{SV-R}} \ge z^\star_{\text{SV}}$, which implies that $z^\star_{\text{SV-R}} = z^\star_{\text{SV}}$. □

Proof of Proposition 2. Observe that for all $s \in S$, we have that $\bowtie_{s,i} \in \{\square, \ge, \le, =\}$, which implies that $\bowtie_s \in \Delta_s$ (see Definition 1). To show that $\bowtie_s \in \Theta_s$ (see Definition 2) for all $s \in S$, it remains to be shown that $v_s(x^1_s) \ge v_s(x^2_s)$ for all $x^1_s, x^2_s \in \mathcal{X}$ with $x^1_s \bowtie_s x^2_s$.

To prove this, we will show that for any $x^1_s, x^2_s \in \mathcal{X}$ with $x^1_s \bowtie_s x^2_s$, it holds that $T_s x^1_s \le T_s x^2_s$, and thus the feasible region in the second-stage minimization problem corresponding to $v_s(x^1_s)$ in (2) is smaller than that corresponding to $v_s(x^2_s)$ which implies that $v_s(x^1_s) \ge v_s(x^2_s)$.

Let us now show that $T_s x^1_s \le T_s x^2_s$ indeed holds. Observe that a sufficient condition is that $T_{s,i} x^1_{s,i} \le T_{s,i} x^2_{s,i}$ for all $s \in S$ and all $i \in N$. For an arbitrary $s \in S$ and $i \in N$, we will show that this holds for all $\bowtie_{s,i} \in \{\square, \ge, \le, =\}$, by considering all four cases explicitly.

First, if $\bowtie_{s,i}$ equals $\square$, then $T_{s,i} = \mathbf{0}$, and it follows that $T_{s,i} x^1_{s,i} = \mathbf{0} \le \mathbf{0} = T_{s,i} x^2_{s,i}$. Second, if $\bowtie_{s,i}$ equals $\ge$, then $T_{s,i} \le \mathbf{0}$ and $T_{s,i} \ne \mathbf{0}$, and it follows that $x^1_{s,i} \bowtie_{s,i} x^2_{s,i}$ denotes $x^1_{s,i} \ge x^2_{s,i}$, which implies $T_{s,i} x^1_{s,i} \le T_{s,i} x^2_{s,i}$. Third, if $\bowtie_{s,i}$ equals $\le$, then $T_{s,i} \ge \mathbf{0}$ and $T_{s,i} \ne \mathbf{0}$, and it follows that $x^1_{s,i} \bowtie_{s,i} x^2_{s,i}$ denotes $x^1_{s,i} \le x^2_{s,i}$, which implies $T_{s,i} x^1_{s,i} \le T_{s,i} x^2_{s,i}$. Fourth, if $\bowtie_{s,i}$ equals $=$, then $x^1_{s,i} \bowtie_{s,i} x^2_{s,i}$ denotes $x^1_{s,i} = x^2_{s,i}$, which implies $T_{s,i} x^1_{s,i} \le T_{s,i} x^2_{s,i}$. The above cases cover all possibilities, and the statement holds in each. Since $s \in S$ and $i \in N$ were chosen arbitrarily, it follows that $T_{s,i} x^1_{s,i} \le T_{s,i} x^2_{s,i}$ for all $s \in S, i \in N$. This completes the proof. □

Proof of Proposition 3. Observe that by definition, $\mathcal{X}_s(\bar{K}_s, (=)_{i\in N})$ and $\mathcal{X}_s(\bar{K}_s, \bowtie_s)$ are the effective domains of functions $\bar{v}_s(., \bar{K}_s, (=)_{i\in N})$ and $\bar{v}_s(., \bar{K}_s, \bowtie_s)$, respectively. That is, the set of first-stage values where the optimal objective value of the minimization problem corresponding to $\bar{v}_s(., \bar{K}_s, (=)_{i\in N})$ and $\bar{v}_s(., \bar{K}_s, \bowtie_s)$ are finite, respectively. Since the minimization problem corresponding to $\bar{v}_s(., \bar{K}_s, \bowtie_s)$ is a relaxation of that corresponding to $\bar{v}_s(., \bar{K}_s, (=)_{i\in N})$ (See Equation (11)), finiteness of the latter implies finiteness of the former. □

Proof of Proposition 4. Observe that $\tilde{K}_s(\bar{K}_s, (=)_{i\in N})$ and $\tilde{K}_s(\bar{K}_s, \bowtie_s)$ only have one difference

(See Definition 13), which is that $\sum_{k\in\bar{K}_s}\mu_s^k x_s^k = x_s^{\tilde{k}}$ holds for all columns in the former, while $\sum_{k\in\bar{K}_s}\mu_s^k x_s^k \bowtie_s x_s^{\tilde{k}}$ holds for all columns in the latter. Since $\sum_{k\in\bar{K}_s}\mu_s^k x_s^k \bowtie_s x_s^{\tilde{k}}$ is a relaxation of $\sum_{k\in\bar{K}_s}\mu_s^k x_s^k = x_s^{\tilde{k}}$, it follows that all columns in $\tilde{K}_s(\bar{K}_s,(=)_{i\in N})$ also satisfy all constraints in $\tilde{K}_s(\bar{K}_s,\bowtie_s)$, and are thus also columns in $\tilde{K}_s(\bar{K}_s,\bowtie_s)$. □

Proof of Proposition 5. By LP optimality, the reduced cost of every column $k\in\bar{K}_s$ is non-negative:

$$r_s^k(\pi_s,\alpha_s) = p_s\, q_s^\top y_s^k - \pi_s^\top x_s^k - \alpha_s \geq 0.$$

Since $\tilde{k}\in\tilde{K}_s(\bar{K}_s,\bowtie_s)$, there exist $(\mu_s^k)_{k\in\bar{K}_s}\geq 0$ with $\sum_k \mu_s^k x_s^k \bowtie_s x_s^{\tilde{k}}$, $\sum_k \mu_s^k q_s^\top y_s^k \leq q_s^\top y_s^{\tilde{k}}$, and $\sum_k \mu_s^k = 1$. Dual feasibility of the RMP implies that for each component i, the sign of $\pi_{s,i}$ is consistent with the direction $\bowtie_{s,i}$ (non-negative for $\geq$, non-positive for $\leq$, unrestricted for $=$). Combined with the dominance condition, this gives $\pi_{s,i}(\sum_k \mu_s^k x_{s,i}^k - x_{s,i}^{\tilde{k}})\geq 0$ for each $i\in N$, and hence $\pi_s^\top\sum_k \mu_s^k x_s^k \geq \pi_s^\top x_s^{\tilde{k}}$. Therefore,

$$\begin{aligned} r_s^{\tilde{k}} &\geq p_s \sum_{k\in\bar{K}_s}\mu_s^k q_s^\top y_s^k - \pi_s^\top \sum_{k\in\bar{K}_s}\mu_s^k x_s^k - \alpha_s \\ &= \sum_{k\in\bar{K}_s}\mu_s^k\, r_s^k \geq 0. \end{aligned} \tag{24}$$

□

Proof of Proposition 6. Introduce an auxiliary variable x_s and write $\bar{v}_s(x,\bar{K}_s,\bowtie_s)$ in (11) as

$$\begin{aligned} \min_{x_s}\ \min_{(\lambda_s^k)_{k\in\bar{K}_s}}\quad & \sum_{k\in\bar{K}_s}\lambda_s^k q_s^\top y_s^k \\ \text{s.t.}\quad & \sum_{k\in\bar{K}_s}\lambda_s^k x_s^k = x_s \\ & x_s \bowtie_s x, \\ & \sum_{k\in\bar{K}_s}\lambda_s^k = 1, \\ & \lambda_s^k \geq 0. \end{aligned}$$

The inner minimization over λ at fixed x_s equals $\bar{v}_s(x_s,\bar{K}_s,(=)_{i\in N})$ in (11), yielding the first claim. The second follows from $\bar{v}_s(\cdot,K_s,(=)_{i\in N}) = \text{co}_{\mathcal{X}}(v_s)$, see, e.g., (Romeijnders et al., 2026). □

Proof of Proposition 7. The inequality $\bar{v}_s(x,K_s,\bowtie_s)\leq \text{co}_{\mathcal{X}}(v_s)(x)$ follows from the fact that $\bar{v}_s(.,K_s,(=)_{i\in N}) = \text{co}_{\mathcal{X}}(v_s)$, see, e.g., (Romeijnders et al., 2026), and the minimization in the problem corresponding to $\bar{v}_s(x,K_s,\bowtie_s)$ in (11) relaxes that corresponding to $\bar{v}_s(.,K_s,(=)_{i\in N})$ in (11). For the reverse, it is sufficient to show that $\bar{v}_s(x,K_s,\bowtie_s)\geq \text{co}_{\mathcal{X}}(v_s)(x)$ for an arbitrary $x\in\text{conv}(\mathcal{X})$. By Proposition 6, we have that

$$\bar{v}_s(x,K_s,\bowtie_s) = \min_{x_s}\big\{\text{co}_{\mathcal{X}}(v_s)(x_s)\ :\ x_s\bowtie_s x\big\}. \tag{25}$$

Consider an arbitrary $x \in \text{conv}(\mathcal{X})$. By assumption, we have that all $x_s \in \text{conv}(\mathcal{X})$ with $x_s \bowtie_s x$ satisfy $\text{co}_{\mathcal{X}}(v_s)(x_s) \geq \text{co}_{\mathcal{X}}(v_s)(x)$. Moreover, $\text{co}_{\mathcal{X}}(v_s)(x_s) = \infty$ for $x_s \notin \text{conv}(\mathcal{X})$. Thus, the minimum in (25) is attained at $x_s = x$. Equality NACs also hold in this solution, which implies that this is also a feasible solution of the problem corresponding to $\bar{v}_s(x, K_s, (=)_{i\in N})$ in (11). $\square$

Proof of Theorem 1. By Proposition 7, it suffices to show that $\text{co}_{\mathcal{X}}(v_s)(x^1) \geq \text{co}_{\mathcal{X}}(v_s)(x^2)$ for all $x^1, x^2 \in \text{conv}(\mathcal{X})$ with $x^1 \bowtie_s x^2$. Note that since $\text{co}_{\mathcal{X}}(v_s) = \bar{v}_s(., K_s, (=)_{i\in N})$ (Romeijnders et al., 2026), this is equivalent to showing that $\bar{v}_s(x^1, K_s, (=)_{i\in N}) \geq \bar{v}_s(x^2, K_s, (=)_{i\in N})$ for all $x^1, x^2 \in \text{conv}(\mathcal{X})$ with $x^1 \bowtie_s x^2$. To do so, fix arbitrary $s \in S$ and $x_s^1, x_s^2 \in \text{conv}(\mathcal{X})$ with $x_s^1 \bowtie_s x_s^2$, and let $\lambda_s(0)$ be an optimal solution of the minimization corresponding to $\bar{v}_s(x_s^1, K_s, (=)_{i\in N})$ in (11).

We construct a sequence $\lambda_s(0), \lambda_s(1), \ldots, \lambda_s(|N|)$ in which $\lambda_s(i)$ is obtained from $\lambda_s(i-1)$ by shifting weights of $\lambda_s(i-1)$ such that $\sum_{k\in K_s} \lambda_s^k(i)\, x_{s,i}^k$ equals $x_{s,i}^2$ instead of $x_{s,i}^1$, while not affecting the other components $j \in N \setminus \{i\}$. This ensures that $\lambda_s(i)$ satisfies the NAC of component i corresponding to the problem in $\bar{v}_s(x_s^2, K_s, (=)_{i\in N})$ instead of the one in $\bar{v}_s(x_s^1, K_s, (=)_{i\in N})$. We will show by induction that $\lambda_s(|N|)$ then satisfies all NACs in $\bar{v}_s(x_s^2, K_s, (=)_{i\in N})$, besides all other constraints in $\bar{v}_s(x_s^2, K_s, (=)_{i\in N})$. Moreover, its objective is at most $\bar{v}_s(x_s^1, K_s, (=)_{i\in N})$, proving the claim. Formally, each $\lambda_s(i)$ satisfies:

(a) $\sum_{k\in K_s} \lambda_s^k(i) = 1$, and $\lambda_s^k(i) \geq 0$ for all $k \in K_s$.

(b) $\sum_{k\in K_s} \lambda_s^k(i)\, x_{s,j}^k = \begin{cases} x_{s,j}^2, & j \in N, j \leq i, \\ x_{s,j}^1, & j \in N, j > i. \end{cases}$

(c) $\sum_{k\in K_s} \lambda_s^k(i)\, v_s(x_s^k) \leq \text{co}_{\mathcal{X}}(v_s)(x^1)$.

Observe that $\lambda_s(0)$ satisfies (a)–(c) since it is an optimal solution of the problem corresponding to $\bar{v}_s(x_s^1, K_s, (=)_{i\in N})$ in (11). As the inductive step, we will show that if $\lambda_s(\hat{\imath})$ satisfies (a)–(c) for arbitrary $\hat{\imath} \in 0, ..., |N-1|$, then $\lambda_s(\hat{\imath}+1)$ will satisfy (a)–(c). For each $k \in K_s$, define $\hat{k}(k, \hat{\imath}+1)$ as the index of a column with $x_{s,j}^{\hat{k}(k,\hat{\imath}+1)} = x_{s,j}^k$ for all $j \neq \hat{\imath}+1$, and

$$x_{s,\hat{\imath}+1}^{\hat{k}(k,\hat{\imath}+1)} = \begin{cases} u_{\hat{\imath}+1} & \text{if } x_{s,\hat{\imath}+1}^1 < x_{s,\hat{\imath}+1}^2, \\ l_{\hat{\imath}+1} & \text{if } x_{s,\hat{\imath}+1}^1 > x_{s,\hat{\imath}+1}^2, \\ x_{s,\hat{\imath}+1}^k & \text{if } x_{s,\hat{\imath}+1}^1 = x_{s,\hat{\imath}+1}^2. \end{cases}$$

Observe that $\hat{k}(k, \hat{\imath}+1)$ is not necessarily in K_s since it might not correspond to an extreme point of $\mathcal{F}_s$. However, we may include $\hat{k}(k, \hat{\imath}+1)$ in K_s without loss of generality because of the following. We have that $\text{conv}(\mathcal{X})$ is a hyper-rectangle, so replacing coordinate $\hat{\imath}+1$ of $x_s^k \in \mathcal{X}$ by $l_{\hat{\imath}+1}$ or $u_{\hat{\imath}+1}$ preserves membership in $\mathcal{X}$: values in dimension other than $\hat{\imath}+1$ are unchanged, $l_{\hat{\imath}+1} \leq l_{\hat{\imath}+1}, u_{\hat{\imath}+1} \leq u_{\hat{\imath}+1}$, and integrality holds because $l_{\hat{\imath}+1}, u_{\hat{\imath}+1}$ are extreme-point values. Thus $x_s^{\hat{k}(k,\hat{\imath}+1)} \in \mathcal{X}$.

We establish two key properties. First, we claim there exists $f \in [0, 1]$ satisfying

$$
\begin{aligned}
&(1-f)\sum_{k\in K_s}\lambda_s^k(\hat{\imath})\,x^k_{s,\hat{\imath}+1}\\
&\quad+f\sum_{k\in K_s}\lambda_s^k(\hat{\imath})\,x^{\hat{k}(k,\hat{\imath}+1)}_{s,\hat{\imath}+1}=x^2_{s,\hat{\imath}+1}.
\end{aligned}
\tag{26}
$$

Indeed, by (b), the first sum equals $x^1_{s,\hat{\imath}+1}$. If $x^1_{s,\hat{\imath}+1}<x^2_{s,\hat{\imath}+1}$, the second sum equals $u_{\hat{\imath}+1}\geq x^2_{s,\hat{\imath}+1}$, so $x^2_{s,\hat{\imath}+1}$ lies between the two sums. If $x^1_{s,\hat{\imath}+1}>x^2_{s,\hat{\imath}+1}$, the second sum equals $l_{\hat{\imath}+1}\leq x^2_{s,\hat{\imath}+1}$, again placing $x^2_{s,\hat{\imath}+1}$ between them. If $x^1_{s,\hat{\imath}+1}=x^2_{s,\hat{\imath}+1}$, both sums equal $x^1_{s,\hat{\imath}+1}$ and any f works.

Second, we claim $x^k_s \bowtie_s x_s^{\hat{k}(k,\hat{\imath}+1)}$ for all $k\in K_s$. For $j\neq\hat{\imath}+1$ this is immediate since $x^k_{s,j}=x^{\hat{k}(k,\hat{\imath}+1)}_{s,j}$. For $j=\hat{\imath}+1$: if $x^1_{s,\hat{\imath}+1}<x^2_{s,\hat{\imath}+1}$ then $\bowtie_{s,\hat{\imath}+1}\in\{\leq,\square\}$ (since $x^1_s\bowtie_s x^2_s$) and $x^k_{s,\hat{\imath}+1}\leq u_{\hat{\imath}+1}=x^{\hat{k}(k,\hat{\imath}+1)}_{s,\hat{\imath}+1}$; if $x^1_{s,\hat{\imath}+1}>x^2_{s,\hat{\imath}+1}$ then $\bowtie_{s,\hat{\imath}+1}\in\{\geq,\square\}$ and $x^k_{s,\hat{\imath}+1}\geq l_{\hat{\imath}+1}=x^{\hat{k}(k,\hat{\imath}+1)}_{s,\hat{\imath}+1}$; if equal, then $\bowtie_{s,\hat{\imath}+1}\in\{=,\geq,\leq,\square\}$ and $x^k_{s,\hat{\imath}+1}=x^{\hat{k}(k,\hat{\imath}+1)}_{s,\hat{\imath}+1}$. Thus, in all cases $x^k_s\bowtie_s x^{\hat{k}(k,\hat{\imath}+1)}_s$ holds. Moreover, since $\bowtie_s\in\Theta_s$, and $x^k_s, x^{\hat{k}(k,\hat{\imath}+1)}_s\in\mathcal{X}$ for all $k\in K_s$, it follows that

$$
v_s(x^k_s)\geq v_s(x^{\hat{k}(k,\hat{\imath}+1)}_s)\quad\text{for all } k\in K_s. \tag{27}
$$

Now define for all $l\in K_s$

$$
\lambda^l_s(\hat{\imath}+1)=(1-f)\lambda^l_s(\hat{\imath})+f\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,\mathbf{1}_{l=\hat{k}(k,\hat{\imath}+1)}.
$$

Verification of (a): Each k maps to exactly one $\hat{k}(k,\hat{\imath}+1)$, so summing $\lambda^l_s(\hat{\imath}+1)$ over l gives $(1-f)+f=1$. Non-negativity follows from $f\in[0,1]$ and $\lambda_s(\hat{\imath})\geq 0$.

Verification of (b): Substituting the definition of $\lambda_s(\hat{\imath}+1)$ and using that each k maps to exactly one $\hat{k}(k,\hat{\imath}+1)$ gives, for any $j\in N$,

$$
\begin{aligned}
\sum_{l\in K_s}\lambda^l_s(\hat{\imath}+1)\,x^l_{s,j}&=(1-f)\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,x^k_{s,j}\\
&\quad+f\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,x^{\hat{k}(k,\hat{\imath}+1)}_{s,j}.
\end{aligned}
\tag{28}
$$

For $j\neq\hat{\imath}+1$ this reduces to $\sum_k\lambda^k_s(\hat{\imath})\,x^k_{s,j}$ since $x^{\hat{k}(k,\hat{\imath}+1)}_{s,j}=x^k_{s,j}$, satisfying (b) by induction. For $j=\hat{\imath}+1$ the expression equals $x^2_{s,\hat{\imath}+1}$ by (26).

Verification of (c): Substituting the definition of $\lambda_s(\hat{\imath}+1)$ and using that each k maps to exactly one $\hat{k}(k,\hat{\imath}+1)$, combined with (27), gives

$$
\begin{aligned}
\sum_{l\in K_s}\lambda^l_s(\hat{\imath}+1)\,v_s(x^l_s)&=(1-f)\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,v_s(x^k_s)\\
&\quad+f\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,v_s(x^{\hat{k}(k,\hat{\imath}+1)}_s)\\
&\leq\sum_{k\in K_s}\lambda^k_s(\hat{\imath})\,v_s(x^k_s),
\end{aligned}
\tag{29}
$$

satisfying (c) by induction. It follows that $\lambda_s(|N|)$ is a feasible solution of $\bar{v}_s(x^2_s,K_s,(=)_{i\in N})$ in (10) since it satisfies (a) and (b). Moreover, since it also satisfies (c), it also follows that

$$
\bar{v}_s(x^1_s,K_s,(=)_{i\in N})\geq\bar{v}_s(x^2_s,K_s,(=)_{i\in N}),
$$

which completes the proof. $\square$

Proof of Proposition 8. Consider the single-scenario SMIP

$$z^\star = \min_{\substack{x_1,x_2,x_3\in\{0,1\}\\ x_1+x_2+x_3\le 2}} -x_1 + v_1(x_1,x_2,x_3), \tag{30}$$

where $v_1(x) = \min\{y_1 \in \{0,1\} : y_1 \ge 1 - x_2,\ y_1 \ge 1 - x_3\}$, and $N = \{1, 2, 3\}$. It follows from Proposition 2 that $(\le)_{i\in N} \in \Theta_1$. Observe that for the problem (30), we have that the optimal objective value of DW without relaxed NACs $z^\star_{\text{DW}}$ in (6) can alternatively be formulated as

$$z^\star_{\text{DW}} = \min_{x,y_1} \Big\{ x_1 + y_1 : (x_1, x_2, x_3, y_1) \in \text{conv}(\mathcal{F}_1) \Big\}, \tag{31}$$

where an extreme-point of $\text{conv}(\mathcal{F}_1)$ must be optimal. Hence, such a solution is an integer feasible solution of the original problem (30). Observe that all optimal solutions satisfy $x_1 = y_1$, since $y_1 = 0$ is feasible if $x_1 = 0$ but infeasible if $x_1 = 1$. Thus, $z^\star_{\text{DW}} = 0$.
On the other hand, the MP of (30) with relaxed NACs $x_i \le \hat{x}_i$ for $i \in N$, admits the feasible solution

$$\hat{x} = \big(1,\ \tfrac{1}{2},\ \tfrac{1}{2}\big), \qquad x = \big(\tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2}\big), \qquad y_1 = \tfrac{1}{2},$$

where we take $(x_1, x_2, x_3, y_1) \in \text{conv}(\mathcal{F}_1)$ as a convex combination between the extreme points (1, 0, 0, 1) and (0, 1, 1, 0) both with a weight of $\frac{1}{2}$. The corresponding objective value is $-\frac{1}{2}$. Hence, $z^\star_{\text{DW-R}} \le -\frac{1}{2} < 0 = z^\star_{\text{DW}}$. Moreover, at $\hat{x} = (1, \frac{1}{2}, \frac{1}{2})$, we have $\bar{v}_1(\hat{x},\ K_1,\ (\le)_{i\in N}) \le -\frac{1}{2}$ and $\bar{v}_1(\hat{x},\ K_1,\ (=)_{i\in N}) \ge 0$, which means that relaxing NACs in the MP (7) can strictly weaken the lower bound on $v_s(x)$ and $\sum_{s\in S} p_s v_s(x)$ obtained via the MP (7). $\square$

Proof of Theorem 2. Consider any arbitrary $x \in \mathcal{X}$ and $s \in S$, and define $K_s(x) := \{k \in K_s : x_s^k \bowtie_s x\}$. The inequality $\bar{v}_s(x, K_s, \bowtie_s) \le v_s(x)$ holds because the minimization corresponding to $\bar{v}_s(x, K_s, \bowtie_s)$ relaxes that corresponding to $v_s(x)$. It remains to show that $\bar{v}_s(x, K_s, \bowtie_s) \ge v_s(x)$.

First, we show that every feasible solution $(\lambda_s^k)_{k\in K_s}$ for the minimization in $\bar{v}_s(x, K_s, \bowtie_s)$ is such that $\lambda_s^k = 0$ for all $k \notin K_s(x)$. Second, we show that this implies that $\bar{v}_s(x, K_s, \bowtie_s) \ge v_s(x)$. Consider an arbitrary feasible solution $(\lambda_s^k)_{k\in K_s}$ of the problem in $\bar{v}_s(x, K_s, \bowtie_s)$. Suppose $\lambda_s^{\bar{k}} > 0$ for some $\bar{k} \notin K_s(x)$.

Then there exists $i \in N$ with $x_{s,i}^{\bar{k}} \bowtie_{s,i} x_i$ violated. Since both values are binary, $x_{s,i}^{\bar{k}} = 1 - x_i$. If $x_i = 0$, then $\bowtie_{s,i} \in \{=, \le\}$ since $x_{s,i}^{\bar{k}} \bowtie_{s,i} x_i$ would otherwise not be violated. Observe that

$$\textstyle\sum_{k\in K_s} \lambda_s^k x_{s,i}^k \ \ge\ \lambda_s^{\bar{k}} \cdot 1 \ >\ 0 = x_i,$$

violating the NAC $\sum_{k\in K_s} \lambda_s^k x_{s,i}^k \bowtie_{s,i} x_i$. If $x_i = 1$, then $\bowtie_{s,i} \in \{=, \ge\}$ and

$$\textstyle\sum_{k\in K_s} \lambda_s^k x_{s,i}^k \ =\ \sum_{k\neq\bar{k}} \lambda_s^k x_{s,i}^k \ \le\ 1 - \lambda_s^{\bar{k}} \ <\ 1 = x_i,$$

again violating this NAC. In both cases $(\lambda_s^k)_{k\in K_s}$ is infeasible, which is a contradiction. It follows that $\lambda_s^k = 0$ for all $k \notin K_s(x)$. Thus,

$$\bar{v}_s(x, K_s, \bowtie_s) = \min_{(\lambda_s^k)_{k \in K_s(x)}} \sum_{k \in K_s(x)} \lambda_s^k q_s^\top y_s^k \tag{32a}$$

$$\text{s.t.} \sum_{k \in K_s(x)} \lambda_s^k = 1 \tag{32b}$$

$$\sum_{k \in K_s(x)} \lambda_s^k x_s^k \bowtie_s x \tag{32c}$$

$$\lambda_s^k \geq 0 \quad k \in K_s(x). \tag{32d}$$

Observe that the NACs in (32) are always satisfied because $x_s^k \bowtie_s x$ holds for all $k \in K_s(x)$, so it also holds for a convex combination. Thus, the NACs can be dropped from the formulation. The remaining problem minimizes $\sum_k \lambda_s^k q_s^\top y_s^k$ over the simplex, hence equals $\min_{k \in K_s(x)} q_s^\top y_s^k$. Finally,

$$\min_{k \in K_s(x)} q_s^\top y_s^k \;\geq\; \min_{k \in K_s(x)} v_s(x_s^k) \;\geq\; v_s(x),$$

where the first inequality follows from the definition of v_s, and the second from the fact that $\bowtie_s \in \Theta_s$ and $x_s^k \bowtie_s x$ for all $k \in K_s(x)$. This completes the proof. □