

Solving Chance Constrained Programs via a Penalty based Difference of Convex Approach

Zhiping Li¹

Nan Jiang²

Rujun Jiang^{3, †}

March 13, 2026

Abstract

We develop two penalty based difference of convex (DC) algorithms for solving chance constrained programs. First, leveraging a rank-based DC decomposition of the chance constraint, we propose a proximal penalty based DC algorithm in the primal space that does not require a feasible initialization. Second, to improve numerical stability in the general nonlinear settings, we derive an equivalent lifted formulation with complementary constraints and show that, after minimizing primal variables, the penalized lifted problem admits a tractable DC structure in the dual space over a simple polyhedron. We then develop a penalty based DC algorithm in the lifted space with a finite termination guarantee. We establish exact penalty and stationarity guarantees under mild constraint qualifications and identify the relationship of the local minimizers between the two formulations. Numerical experiments demonstrate the efficiency and effectiveness of our proposed methods compared with state-of-the-art benchmarks.

1 Introduction

As a powerful tool for addressing uncertainty in decision-making, the chance constrained program (CCP) has been extensively studied in the literature. It has been widely applied across fields such as finance, energy systems, and supply chain management [2, 32]. In general, a chance constrained program can be written as follows:

$$\min_{\mathbf{x} \in \mathcal{X}} \left\{ f(\mathbf{x}) : \Pr\{g(\mathbf{x}, \tilde{\xi}) \leq 0\} \geq 1 - \alpha \right\}, \quad (1)$$

where $\tilde{\xi}$ is a random vector supported on $\Xi \subseteq \mathbb{R}^{d_\xi}$ and $\alpha \in (0, 1)$ is a prescribed violation risk level. The feasible region \mathcal{X} is a deterministic set contained in an open set $\mathcal{U} \subseteq \mathbb{R}^d$, the functions $f : \mathcal{U} \rightarrow \mathbb{R}$, $g : \mathcal{U} \times \Xi \rightarrow \mathbb{R}$ are real-valued functions. We use $\tilde{\xi}$ to denote the random vector and ξ as its realization. The goal of Problem (1) is to minimize the objective function $f(\mathbf{x})$ subject to the requirement that the constraint $g(\mathbf{x}, \tilde{\xi}) \leq 0$ be satisfied with probability at least $(1 - \alpha)$.

Problem (1) is generally difficult to solve, especially when the distribution information of the random variable $\tilde{\xi}$ is unknown [1]. First, verifying whether a candidate solution satisfies the chance constraint can be computationally challenging. Moreover, the feasible region of a CCP is generally

¹IEDA, HKUST and School of Economics, Fudan University. Email: lizhiping13624@gmail.com

²IEDA, HKUST. Email: nanjiang@ust.hk

³School of Data Science, Fudan University. Email: rjjiang@fudan.edu.cn

[†]Corresponding author.

nonconvex even when \mathcal{X} is convex, implying that finding an optimal solution with a provable guarantee can be elusive.

Given the challenges mentioned above, there are two major strategies to address the CCP. The first strategy uses tractable conservative approximations, where the CCP is formulated as a convex optimization problem that can be efficiently solved and yields a feasible solution to the original CCP [43, 3, 15]. The second approach is the Sample Average Approximation (SAA), also sometimes known as the scenario approach, and has been extensively studied in [2, 37, 40]. It is a popular method to approximately solve Problem (1), especially when the distribution of $\tilde{\xi}$ is unknown, and we can only access a sample of S i.i.d. realizations $\{\xi^s\}_{s=1}^S$.

In this paper, we consider the SAA approximation of Problem (1) as follows:

$$v^* = \min_{\mathbf{x} \in \mathcal{X}} \left\{ f(\mathbf{x}) : \frac{1}{S} \sum_{s=1}^S \mathbb{I}\{g(\mathbf{x}, \xi^s) \leq 0\} \geq 1 - \alpha \right\}, \quad (2)$$

where the indicator function $\mathbb{I}\{g(\mathbf{x}, \xi^s) \leq 0\} = 1$ if $g(\mathbf{x}, \xi^s) \leq 0$ and 0 otherwise. Moreover, we focus on the convex setting and make the following assumptions.

Assumption 1. *The feasible region $\mathcal{X} \subseteq \mathbb{R}^d$ is compact, convex, and has a nonempty relative interior.*

Assumption 2. *The function $f : \mathcal{U} \rightarrow \mathbb{R}$ is continuously differentiable and convex. The function $g : \mathcal{U} \times \Xi \rightarrow \mathbb{R}$ takes the form $g(\mathbf{x}, \xi) = \max_{i \in [I]} h_i(\mathbf{x}, \xi)$, where $h_i : \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ is continuously differentiable and convex in \mathbf{x} for every $\xi \in \Xi$ and $i \in [I]$. Moreover, $\mathcal{U} \subseteq \mathbb{R}^d$ is an open set that contains \mathcal{X} .*

Assumptions 1 and 2 are common in the existing literature [50, 30, 53, 4] and cover several common settings, e.g., the objective function $f(\mathbf{x})$ is linear, and the feasible region \mathcal{X} is a polyhedron [38]. Note that Assumptions 1 and 2 imply that $\nabla f(\mathbf{x})$ is bounded on \mathcal{X} and thus f is Lipschitz continuous on \mathcal{X} . That is, there exists $L_f > 0$ such that

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq L_f \|\mathbf{x} - \mathbf{x}'\|, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}.$$

Problem (2) is known as a single chance constrained program if $I = 1$ and a joint chance constrained program otherwise. Other assumptions will be introduced if needed. Given a sample of S realizations $\{\xi^s\}_{s=1}^S$, we use $g_s(\mathbf{x})$ to denote $g(\mathbf{x}, \xi^s)$ and use $h_{i,s}(\mathbf{x})$ to denote $h_i(\mathbf{x}, \xi^s)$ for all $s \in [S]$ and $i \in [I]$ in the rest of the paper.

1.1 Relevant Literature

Chance Constrained Programs. Chance constrained programming has a long history as a framework for incorporating uncertainty directly into decision models. Since its early development [12, 11], it has become a central tool in applications where decisions must remain feasible with high probability. In energy systems, for instance, chance constrained programs provide a mechanism for coping with operational randomness such as renewable generation variability, fluctuating demand profiles, and unexpected network contingencies [16, 42]. Such models enable system operators to enforce reliability targets while acknowledging the stochastic nature of real-world power systems. Similar ideas appear in financial decision-making, where probabilistic constraints are used

to balance portfolio performance against the risk generated by volatile market conditions [19]. By controlling the likelihood of unacceptable losses, these models support investment strategies that remain robust in uncertain environments. In logistics and supply chain management [20, 22], chance constraints play a comparable role: they allow planners to ensure service, routing, or inventory requirements are satisfied with prescribed confidence levels despite uncertain demand or transportation disruptions. We refer interested readers to [2, 32] for a comprehensive review.

A central difficulty in working with chance constrained programs (1) is that the feasible region is typically nonconvex, which limits the applicability of standard convex optimization tools. To address this issue, a commonly used strategy is the sample average approximation (SAA), in which the probabilistic constraint is replaced with a finite set of scenarios. This yields a deterministic mixed-integer formulation that can be solved by modern mixed-integer optimization solvers. While this provides a viable computational framework, its practical applicability is generally limited to instances of moderate scale [37, 44]. Another line of research focuses on constructing convex inner approximations of the original chance constraint [40]. These methods replace the probabilistic requirement with a tractable approximation that can be handled using standard convex optimization techniques. A commonly used example is the conditional value-at-risk (CVaR) reformulation, which provides a convex upper bound on the probability of violation. While this approximation preserves feasibility with respect to the chance constraint, it can be conservative and generally does not recover the optimal objective value of the original problem. More recent developments aim to mitigate the conservatism of single-shot convex approximations by employing iterative refinement procedures [30, 31]. The ALSO-X method is a representative example of this direction. When the deterministic feasible region \mathcal{X} is convex, it has been shown that ALSO-X can yield solutions that improve upon those obtained from the CVaR approximation, due to its iterative updates.

Building on ALSO-X, an enhanced variant, ALSO-X+, has been proposed to improve solution quality further. In numerical experiments, ALSO-X+ has been observed to achieve better objective values than ALSO-X and CVaR approximation; see, for example, [54, 51, 46]. At the same time, these works also document that the computational time required by ALSO-X+ is substantially higher than that of ALSO-X and the CVaR approximation, especially on larger instances. A further limitation is that the theoretical properties of the solutions produced by ALSO-X+ are not fully understood. In particular, current analyses in [30] do not establish convergence to a stationary point of the underlying chance constrained problem. Hence, the quality of the limiting solutions lacks a rigorous optimality guarantee.

Difference of Convex (DC) Program. DC programs are optimization problems that aim to minimize a DC function subject to DC constraints. DC programs play an important role in nonconvex optimization and have been extensively studied for decades [28, 47]. One of the most important methods for solving DC programs is the DC algorithm (DCA) [27, 36], which solves a sequence of convex subproblems by replacing the concave part with its first-order approximation at the current iterate. Recently, [50] proposed a DCA to solve the SAA of chance constrained programs, with subproblems that are easy to implement and are amenable to the off-the-shelf solvers. Another popular method for solving the DC program is exact penalty [35, 23, 29], which penalizes the DC constraint in the objective function and solves a sequence of DC programs using DCA with increasing penalty parameters. Compared with the DCA, the exact penalty method does not require an initial feasible solution. It allows constraints to be violated in the initial

updates, which can sometimes lead to a lower objective value [35].

1.2 Summary of Contributions

In this paper, we study the penalty based DC algorithms for solving chance constrained programs. Our contributions can be summarized as follows:

- (i) Building on the rank-based DC decomposition of the SAA chance constraint proposed in [50], we first propose a penalty based DC method in the primal space. The method does not require an initial feasible solution and can alleviate over-conservatism caused by infeasible subproblems. This algorithm improves both solution quality and running time in the polyhedral setting, where both \mathcal{X} and g are polyhedral, but exhibits pronounced oscillations in the nonlinear setting.
- (ii) To mitigate the aforementioned oscillations that may arise in the primal space under nonlinear settings, we derive an equivalent lifted formulation with complementary constraints, inspired by the Toland duality framework for DC programs [48]. This leads to a tractable DC structure over a simple polyhedron and yields a proximal DC algorithm in the lifted space. The subproblem admits finite termination, is amenable to effective warm starts, and exhibits substantially improved numerical stability.
- (iii) We establish an exact penalty for both the primal and lifted formulations under mild constraint qualification assumptions, including global exactness, local exactness for the primal formulation, and equivalence between strongly stationary points of the lifted formulation and stationary points of the penalized counterpart. We further analyze the relationship between lifted and primal local minimizers, identify the possibility of spurious lifted local minimizers, and provide a sufficient condition under which such spurious solutions are ruled out.
- (iv) Numerical experiments demonstrate that the proposed methods substantially reduce over-conservatism. Moreover, to the best of our knowledge, the algorithm in the lifted space remains computationally efficient and outperforms state-of-the-art baselines in both solution quality and runtime.

Organization. The remainder of the paper is organized as follows. In Section 2, we develop the algorithmic framework: we first review the DC decomposition of the chance constraint and the exact penalty framework in the primal space. Then we derive an equivalent lifted formulation with complementary constraints and the algorithm in the lifted space. Section 3 establishes the exact penalty and stationarity guarantees for both formulations, including global and local results. Section 4 investigates the relationship between local minimizers of the lifted and primal formulations, characterizes when spurious lifted local minimizers may arise, and provides conditions ensuring that lifted local optimality implies primal local optimality. Section 5 reports numerical experiments on real and synthetic instances and compares against state-of-the-art baselines. In Section 6, we conclude the paper and provide some future directions.

1.3 Notations and Preliminaries

We first introduce the notation used throughout the paper. For a positive integer n , we write $[n] := \{1, 2, \dots, n\}$. We use \mathbb{R}_+^n to denote the nonnegative orthant. For a vector $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\|$

denotes the Euclidean norm and $\langle \cdot, \cdot \rangle$ denotes the standard inner product. We use $\|\mathbf{x}\|_1$ to denote the ℓ_1 -norm and $\|\mathbf{x}\|_0$ to denote the ℓ_0 pseudo norm, i.e., the number of nonzero elements. For a scalar $t \in \mathbb{R}$, define $[t]_+ := \max\{t, 0\}$. For $\bar{\mathbf{x}} \in \mathbb{R}^d$ and $\varepsilon > 0$, $B(\bar{\mathbf{x}}, \varepsilon) := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \bar{\mathbf{x}}\| < \varepsilon\}$ denotes the open neighborhood with radius ε . Given a set $\mathcal{S} \subseteq \mathbb{R}^d$, the indicator function is

$$\delta_{\mathcal{S}}(\mathbf{x}) := \begin{cases} 0, & \mathbf{x} \in \mathcal{S}, \\ +\infty, & \mathbf{x} \notin \mathcal{S}, \end{cases} \text{ and the Euclidean projection operator is } \Pi_{\mathcal{S}}(\mathbf{x}) \in \arg \min_{\mathbf{y} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|.$$

Note that the projection is unique if \mathcal{S} is convex. The distance from a point to a set is $\text{dist}(\mathbf{x}, \mathcal{S}) := \inf\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \mathcal{S}\}$, and $\text{diam}(\mathcal{S}) := \sup\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x}, \mathbf{y} \in \mathcal{S}\}$. For a closed convex set \mathcal{X} , $\mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}})$ and $\mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$ denote the tangent cone and normal cone at $\bar{\mathbf{x}} \in \mathcal{X}$.

For an extended-real-valued function $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$, the directional derivative at $\bar{\mathbf{x}}$ along \mathbf{d} is

$$f'(\bar{\mathbf{x}}; \mathbf{d}) := \lim_{t \rightarrow 0} \frac{f(\bar{\mathbf{x}} + t\mathbf{d}) - f(\bar{\mathbf{x}})}{t}.$$

If f is locally Lipschitz, its Clarke directional derivative is

$$f^\circ(\bar{\mathbf{x}}; \mathbf{d}) := \limsup_{t \downarrow 0} \frac{f(\bar{\mathbf{x}} + t\mathbf{d}) - f(\bar{\mathbf{x}})}{t},$$

and its Clarke subdifferential is

$$\partial f(\bar{\mathbf{x}}) := \left\{ \mathbf{v} \in \mathbb{R}^d : \langle \mathbf{v}, \mathbf{d} \rangle \leq f^\circ(\bar{\mathbf{x}}; \mathbf{d}) \quad \forall \mathbf{d} \in \mathbb{R}^d \right\}.$$

When f is convex, the Clarke subdifferential coincides with the usual convex subdifferential, and $f^\circ(\bar{\mathbf{x}}; \mathbf{d}) = f'(\bar{\mathbf{x}}; \mathbf{d})$. For a closed proper convex function f , its Fenchel conjugate is

$$f^*(\mathbf{p}) := \sup_{\mathbf{x} \in \mathbb{R}^d} \{\langle \mathbf{p}, \mathbf{x} \rangle - f(\mathbf{x})\}.$$

For a DC program in the form $\min_{\mathbf{x}} \{\phi(\mathbf{x}) - \psi(\mathbf{x})\}$ with ϕ, ψ closed proper convex, its Toland dual [48] is $\min_{\mathbf{p}} \{\psi^*(\mathbf{p}) - \phi^*(\mathbf{p})\}$, which is also a DC program. A DC program is called a polyhedral DC program if either ϕ or ψ is polyhedral, i.e., the epigraph is a polyhedron. The polyhedral DC program has some interesting properties on local optimality and finite convergence of the DC algorithm. The reader may refer to [47] for a detailed review.

For a vector $\mathbf{y} \in \mathbb{R}^S$, let $y_{(1)} \leq \dots \leq y_{(S)}$ denote the nondecreasing rearrangement of (y_1, \dots, y_S) and $|y|_{(1)} \leq \dots \leq |y|_{(S)}$ denote the nondecreasing rearrangement of $(|y_1|, \dots, |y_S|)$. The sum of the r largest components is defined as $\text{T}_r(\mathbf{y}) = \sum_{i=S-r+1}^S y_{(i)}$. The Ky-Fan norm is defined as $\|\mathbf{y}\|_{(r)} := \sum_{i=S-r+1}^S |y|_{(i)}$.

Next, we list some classical results on exact penalization of constrained optimization problems. To make the paper self-contained, we provide the proofs in Appendix A. The reader may also refer to [10] for a detailed review. Consider the single constrained problem

$$\min_{\mathbf{x} \in \mathbb{R}^d} \{f(\mathbf{x}) : g(\mathbf{x}) \leq 0, \mathbf{x} \in \mathcal{X}\}, \quad (3)$$

where $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ are locally Lipschitz and $\mathcal{X} \subseteq \mathbb{R}^d$ is convex and compact. Denote the feasible set by

$$\mathcal{F} := \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) \leq 0\}.$$

Let $g^\circ(\bar{\mathbf{x}}; \mathbf{d})$ be the Clarke directional derivative of g at $\bar{\mathbf{x}}$ along \mathbf{d} . We say that the generalized Mangasarian–Fromovitz constraint qualification (GMFCQ) holds at a boundary feasible point $\bar{\mathbf{x}} \in \mathcal{F}$ with $g(\bar{\mathbf{x}}) = 0$ if there exists a direction $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}})$ such that

$$g^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0. \quad (4)$$

Let $\mathcal{G}(\mathbf{x}) := g(\mathbf{x}) + \mathbb{R}_+ = \{t + g(\mathbf{x}) : t \geq 0\}$ for $\mathbf{x} \in \mathcal{X}$ and $\mathcal{G}^{-1}(y) = \{\mathbf{x} \in \mathcal{X} : y \in \mathcal{G}(\mathbf{x})\}$ for $y \in \mathbb{R}$. Under the GMFCQ, the feasible region \mathcal{F} can be shown to be metrically regular [10, Definition 2.2] around $\bar{\mathbf{x}}$. That is, there exists $\kappa, \epsilon > 0$ such that

$$\text{dist}(\mathbf{x}, \mathcal{G}^{-1}(y)) \leq \kappa \text{dist}(y, \mathcal{G}(\mathbf{x})), \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, \epsilon), y \in B(0, \epsilon).$$

We state this in the following proposition.

Proposition 1.1. *Suppose GMFCQ (4) holds at $\bar{\mathbf{x}}$. Define $\mathcal{G}(\mathbf{x}) := g(\mathbf{x}) + \mathbb{R}_+$ for $\mathbf{x} \in \mathcal{X}$. Then \mathcal{F} is metrically regular at $\bar{\mathbf{x}}$. Consequently, there exist $\kappa > 0$ and $\epsilon > 0$ such that the local linear error bound*

$$\text{dist}(\mathbf{x}, \mathcal{F}) \leq \kappa [g(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, \epsilon) \quad (5)$$

holds.

The error bound (5) together with the Lipschitz continuity of f implies local exact penalization. Similar proofs can be found in [33, 13, 39].

Theorem 1.2 (Local exact penalty). *Suppose f is Lipschitz continuous on the compact and convex set \mathcal{X} with constant L_f , and that the local error bound (5) holds at some feasible point $\bar{\mathbf{x}} \in \mathcal{F}$. For $\sigma > 0$, consider the ℓ_1 -penalty problem*

$$\min_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \sigma [g(\mathbf{x})]_+\}. \quad (6)$$

Then the following statements are equivalent:

- (i) For any $\sigma > 0$, if $\bar{\mathbf{x}}$ is a local minimizer of the penalty problem (6) and satisfies $g(\bar{\mathbf{x}}) \leq 0$, then $\bar{\mathbf{x}}$ is a local minimizer of the constrained problem (3).
- (ii) If $\bar{\mathbf{x}}$ is a local minimizer of the constrained problem (3), then there exists $\bar{\sigma} > 0$ such that, for every $\sigma > \bar{\sigma}$, $\bar{\mathbf{x}}$ is a local minimizer of the penalty problem (6).

When the constrained system admits a global linear error bound, i.e.,

$$\text{dist}(\mathbf{x}, \mathcal{F}) \leq \kappa [g(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X}, \quad (7)$$

the same argument as in Theorem 1.2 yields a global exact penalization result. Since the proof is analogous, we only state the theorem and omit its proof.

Theorem 1.3 (Global exact penalty). *Suppose that f is Lipschitz continuous on the compact and convex set \mathcal{X} with constant L_f , and that the global error bound (7) holds. Then the following statements are equivalent:*

- (i) For any $\sigma > 0$, if $\bar{\mathbf{x}}$ is a global minimizer of the penalty problem (6) and satisfies $g(\bar{\mathbf{x}}) \leq 0$, then $\bar{\mathbf{x}}$ is a global minimizer of the constrained problem (3).
- (ii) If $\bar{\mathbf{x}}$ is a global minimizer of the constrained problem (3), then there exists $\bar{\sigma} > 0$ such that for every $\sigma > \bar{\sigma}$ the point $\bar{\mathbf{x}}$ is a global minimizer of the penalty problem (6).

2 Algorithm Frameworks

In this section, we present the overall algorithmic foundations of our penalty based approach for solving chance constrained program (2). We first review the DC reformulation of the SAA chance constraint in the primal space and discuss why the exact penalty in the primal space can be fragile in practice, especially in the nonlinear settings. Motivated by these limitations and inspired by the general Toland duality of DC programs [48], we propose an alternative exact penalty approach and then derive an equivalent lifted formulation, which is a mathematical program with complementarity constraints (MPCC), and show that the resulting value-function representation yields a tractable DC structure. This leads to a polyhedral DC program with a finite convergence guarantee and is amenable to effective warm starts.

2.1 Algorithm in the Primal Space

To obtain a tractable reformulation of Problem (2), [50] recently proposed a difference of convex based reformulation of Problem (2) along with a difference of convex based algorithm (DCA). The algorithm is easy to implement, with subproblems that can be solved directly by state-of-the-art solvers, and is especially efficient in the polyhedral setting. We begin this section by briefly reviewing its key elements. Let $m := \lfloor \alpha S \rfloor$ and let $g_{(1)}(\mathbf{x}) \leq \dots \leq g_{(S)}(\mathbf{x})$ denote the nondecreasing rearrangement of the vector $(g_1(\mathbf{x}), \dots, g_S(\mathbf{x}))$. Define

$$G_1(\mathbf{x}) := \sum_{s=S-m}^S g_{(s)}(\mathbf{x}), \quad G_2(\mathbf{x}) := \sum_{s=S-m+1}^S g_{(s)}(\mathbf{x}). \quad (8)$$

The following equivalence is standard, and we include it for completeness.

Proposition 2.1. [50, lemma 1] *For any $\mathbf{x} \in \mathcal{X}$, the sample based chance constraint*

$$\sum_{s=1}^S \mathbb{I}\{g_s(\mathbf{x}) \leq 0\} \geq S - m \quad (9)$$

holds if and only if $G_1(\mathbf{x}) - G_2(\mathbf{x}) \leq 0$.

By Proposition 2.1, Problem (2) is equivalent to

$$v^* = \min_{\mathbf{x}} \left\{ f(\mathbf{x}) : \mathbf{x} \in \hat{\mathcal{X}} \right\}, \quad (10)$$

where the feasible region $\hat{\mathcal{X}}$ of (10) is defined as follows:

$$\hat{\mathcal{X}} := \{ \mathbf{x} \in \mathcal{X} : \phi(\mathbf{x}) := G_1(\mathbf{x}) - G_2(\mathbf{x}) \leq 0 \}.$$

We next summarize several properties of the functions G_1 and G_2 established in [50].

Proposition 2.2. [50, lemma 1, 2] Under Assumption 1 and Assumption 2, both $G_1(\mathbf{x})$ and $G_2(\mathbf{x})$ defined in (8) are convex and continuous on \mathcal{X} . Moreover, let $\mathcal{M}_{G_2}(\mathbf{x})$ denote the active index set of $G_2(\mathbf{x})$ at \mathbf{x} by

$$\mathcal{M}_{G_2}(\mathbf{x}) := \left\{ (s_1, \dots, s_m) \subseteq [S] : \sum_{t=1}^m g_{s_t}(\mathbf{x}) = G_2(\mathbf{x}) \right\}.$$

Then for any $\mathbf{x} \in \mathcal{X}$, the subdifferential of G_2 admits the explicit representation

$$\partial G_2(\mathbf{x}) = \text{conv} \left\{ \bigcup_{(s_1, \dots, s_m) \in \mathcal{M}_{G_2}(\mathbf{x})} \sum_{t=1}^m \partial g_{s_t}(\mathbf{x}) \right\},$$

where for all $s \in [S]$

$$\partial g_s(\mathbf{x}) = \text{conv} \left\{ \nabla h_{s,i}(\mathbf{x}) : i \in \mathcal{M}_s(\mathbf{x}) \right\},$$

and $\mathcal{M}_s(\mathbf{x})$ denotes the active index set of g_s at \mathbf{x} by

$$\mathcal{M}_s(\mathbf{x}) := \{i \in [I] : h_{s,i}(\mathbf{x}) = g_s(\mathbf{x})\}.$$

We make the following remarks on Proposition 2.2.

- (i) The sample based chance constraint (9) is equivalent to an empirical value-at-risk (VaR) constraint. The requirement that at least $(S - m)$ scenarios satisfy $g_s(\mathbf{x}) \leq 0$ holds if and only if

$$g_{(S-m)}(\mathbf{x}) \leq 0,$$

i.e., the empirical $(1 - \alpha)$ -quantile of $\{g_s(\mathbf{x})\}_{s=1}^S$ is nonpositive.

- (ii) Similarly, both G_1 and G_2 can be viewed as the empirical Conditional Value-at-Risk (CVaR) [43] up to a scaling factor, and each admits an equivalent minimization formulation based on strong duality of linear programs, which will be useful in the following section:

$$G_1(\mathbf{x}) = \min_{\alpha \in \mathbb{R}} \left\{ (m+1)\alpha + \sum_{s=1}^S [g_s(\mathbf{x}) - \alpha]_+ \right\}, \quad G_2(\mathbf{x}) = \min_{\beta \in \mathbb{R}} \left\{ m\beta + \sum_{s=1}^S [g_s(\mathbf{x}) - \beta]_+ \right\}.$$

- (iii) Under Assumption 1–2, G_1 and G_2 are convex and continuous on the compact set \mathcal{X} . Therefore, G_1 , G_2 and their difference $\phi(\mathbf{x})$ are Lipschitz continuous on \mathcal{X} and their Clarke subdifferentials are well-defined [17, section 2.1].

The authors in [50] proposed a difference of convex based algorithm (DCA) for solving Problem (10). At the k th iteration, let $\mathbf{x}^k \in \hat{\mathcal{X}}$. We pick a subgradient $\mathbf{n}^k \in \partial G_2(\mathbf{x}^k)$, linearize the concave part in the constraint, and solve the following subproblem as the update:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) : G_1(\mathbf{x}) - G_2(\mathbf{x}^k) - \langle \mathbf{n}^k, \mathbf{x} \rangle \leq 0\}. \quad (11)$$

We note that the above DCA scheme has several intrinsic limitations despite its strong numerical performance:

- (i) DCA can be viewed as a conservative approximation of Problem (2). The affine majorization of the concave part can sometimes be overly conservative, so the subproblem (11) may become infeasible. A similar problem may also arise in other conservative approximation techniques, e.g., the CVaR approximation [43].
- (ii) The DCA requires an initial feasible point (i.e., $\mathbf{x}^0 \in \hat{\mathcal{X}}$) to linearize the constraint, and finding such a point can be computationally demanding.
- (iii) Problem (10) is a polyhedral DC program when $h_{i,s}$ are affine for all $i \in [I]$ and $s \in [S]$. While DCA can be computationally efficient in the polyhedral setting due to the finite convergence guarantee, it often performs poorly with nonlinear constraints due to pronounced oscillations and slow progress (see [50] and the numerical experiments in Section 5 for details).

To address the first two limitations, we propose an exact penalty based approach for the DC program (10), which has been extensively studied in the literature [35, 33] and does not require an initial feasible solution $\mathbf{x}^0 \in \hat{\mathcal{X}}$. Let $\sigma > 0$ be the penalty parameter, and we consider the following penalized problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \sigma[\phi(\mathbf{x})]_+\}. \quad (12)$$

It is well-known ([28, proposition 2.1]) that the ℓ_1 exact penalty of a DC function still admits an explicit DC decomposition. Therefore, Problem (12) admits a DC decomposition as follows:

$$\min_{\mathbf{x} \in X} f(\mathbf{x}) + \sigma \max\{G_1(\mathbf{x}), G_2(\mathbf{x})\} - \sigma G_2(\mathbf{x}). \quad (13)$$

For a fixed $\sigma > 0$, we may also adopt a proximal DC algorithm to solve Problem (13). Similar to the DCA, for the k th iteration, we can pick a subgradient $\mathbf{n}^k \in \partial G_2(\mathbf{x}^k)$ by Proposition 2.2. We then need to solve the following subproblem:

$$\min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sigma \max\{G_1(\mathbf{x}), G_2(\mathbf{x})\} - \sigma \langle \mathbf{n}^k, \mathbf{x} \rangle \right\}. \quad (14)$$

We remark that the subproblem (14) is feasible under the mild assumption that \mathcal{X} is nonempty. To stabilize the iterates, we can also add a proximal term with $\rho > 0$. By introducing an auxiliary epigraph variable $t \in \mathbb{R}$ and leveraging the dual representation of CVaR, we can obtain an equivalent formulation as follows:

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}, t, \eta_1, \eta_2, \mathbf{u} \in \mathbb{R}_+^S, \mathbf{v} \in \mathbb{R}_+^S} \quad & f(\mathbf{x}) + \sigma t - \sigma \langle \mathbf{n}^k, \mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}^k\|^2 \\ \text{s.t.} \quad & t \geq (m+1)\eta_1 + \sum_{s=1}^S u_s, \quad u_s \geq g_s(\mathbf{x}) - \eta_1, \quad \forall s \in [S], \\ & t \geq m\eta_2 + \sum_{s=1}^S v_s, \quad v_s \geq g_s(\mathbf{x}) - \eta_2, \quad \forall s \in [S]. \end{aligned} \quad (15)$$

We now present the full penalty framework for Problem (13) in Algorithm 1, which is a double-loop algorithm. The inner loop solves the DC program (13) via update (14) to a critical point, and the outer loop updates the parameter σ by enlarging it with a constant factor $\beta > 0$. Following [41, Chapter 17], we initialize the algorithm with a small penalty σ^0 and then increase σ progressively, so that the violation of the chance constraint (equivalently, $[\phi(\mathbf{x})]_+$) is gradually driven to zero.

Similar to the standard analysis of DC program [28], the sequence generated by Algorithm 1 subsequently converges to a critical point of Problem (13). Let $H(\mathbf{x}) = \max\{G_1(\mathbf{x}), G_2(\mathbf{x})\}$. We say $\mathbf{x}^* \in \mathcal{X}$ is a critical point of Problem (13) if there exists $\mathbf{n}^* \in \partial G_2(\mathbf{x}^*)$, $\mathbf{h}^* \in \partial H(\mathbf{x}^*)$ such that the following condition holds:

$$\mathbf{0} \in \nabla f(\mathbf{x}^*) + \sigma \mathbf{h}^* - \sigma \mathbf{n}^* + \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*).$$

Algorithm 1 Penalty Based Proximal DC Algorithm in the Primal Space

Require: Initial $\sigma^0 > 0$, $\beta > 1$.

- 1: **for** $t = 0, 1, 2, \dots$ **do**
 - 2: **for** $k = 0, 1, 2, \dots$ **do**
 - 3: Find $\mathbf{g}^{k,t} \in \partial G_2(\mathbf{x}^{k,t})$ via Proposition 2.2
 - 4: Solve the subproblem (15) with $\sigma = \sigma^t$
 - 5: **end for**
 - 6: Update $\sigma^{t+1} \leftarrow \beta \sigma^t$
 - 7: **end for**
-

We conclude this subsection with some remarks on Algorithm 1.

- (i) Algorithm 1 admits effective warm starts. Specifically, the outer penalty loop can reuse the output solution of the previous penalty level as the initialization for the next penalty level, which often accelerates convergence in practice. In addition, when σ is small, and constraint enforcement is mild, it is usually unnecessary to solve the proximal DC subproblem with fixed- σ to high precision. Instead, one may adopt a loose inner stopping tolerance at early penalty stages and gradually tighten it as σ increases, thereby reducing the overall runtime.
- (ii) In our experiments, we set ρ to a small constant ($10^{-4} - 10^{-3}$) to improve numerical stability and mitigate oscillations, which are observed in the numerical experiments when $\rho = 0$.
- (iii) The primal-space framework extends directly to the case where the objective itself admits a DC decomposition, namely $f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x})$ with f_1, f_2 being convex. In this case, the penalized objective can be written as

$$(f_1(\mathbf{x}) + \sigma H(\mathbf{x})) - (f_2(\mathbf{x}) + \sigma G_2(\mathbf{x})),$$

which is still a DC function on \mathcal{X} . A similar DCA algorithm can also be applied by solving a sequence of subproblems where the concave part $(f_2(\mathbf{x}) + \sigma G_2(\mathbf{x}))$ is linearized using first-order approximation at the current iterate.

Similar to the DCA algorithm in [50], Algorithm 1 is particularly efficient when applied to a polyhedral DC program, which has the property of finite convergence under the mild assumption [47, theorem 6]. A sufficient condition for such property is that \mathcal{X} is a polyhedron and $f, h_{i,s}$ are affine for all i, s . In the numerical experiments (see Section 5), we observe that Algorithm 1 yields a better solution with a lower objective value than the DCA algorithm in the polyhedral setting. The running time of both algorithms is comparable in the setting with a linear objective function. At the same time, Algorithm 1 exhibits substantial speedup over the DCA algorithm when the objective function is quadratic.

In contrast, when the constraints are nonlinear, Algorithm 1 may exhibit pronounced oscillations, similar to the DCA algorithm, and require substantially more iterations and runtime to solve a single subproblem, degrading its practical performance. Moreover, Algorithm 1 fails to return a feasible solution within a reasonable runtime even though we use a larger initial penalty σ^0 and a more aggressive growth factor β . This phenomenon motivates us to propose an alternative exact penalty approach with better numerical stability in the general setting.

2.2 Equivalent Formulation in the Lifted Space

Motivated by the limitations discussed above, we next consider equivalent formulations of Problem (2), which will serve as the basis for our subsequent algorithmic development. Observe that the chance constraint (9) admits an equivalent reformulation [55] as follows:

$$\sum_{s=1}^S \mathbb{I}\{g_s(\mathbf{x}) \leq 0\} \geq S - m \iff \sum_{s=1}^S \mathbb{I}\{[g_s(\mathbf{x})]_+ > 0\} \leq m.$$

Introducing an auxiliary variable $\mathbf{y} \in \mathbb{R}_+^S$, we obtain an equivalent reformulation of Problem (2) with a cardinality constraint:

$$v^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0}} \{f(\mathbf{x}) : \|\mathbf{y}\|_0 \leq m, g_s(\mathbf{x}) \leq y_s, \forall s \in [S]\},$$

where the cardinality constraint $\|\mathbf{y}\|_0 \leq m$ enforces that the number of violated scenarios cannot exceed m .

It is well-known in the literature that the ℓ_0 -norm admits a DC decomposition [18, 23], that is

$$\|\mathbf{y}\|_0 \leq m \iff \|\mathbf{y}\|_1 - \|\mathbf{y}\|_{(m)} = 0,$$

where $\|\mathbf{y}\|_{(m)}$ denotes the Ky-Fan norm of \mathbf{y} . In [23], the authors considered a penalty approach based on the DC decomposition of the cardinality constraint. In our setting, since $\mathbf{y} \geq \mathbf{0}$, the penalized problem can be simplified as follows:

$$v^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0}} \left\{ f(\mathbf{x}) + \sigma(\mathbf{1}^\top \mathbf{y} - T_m(\mathbf{y})) : g_s(\mathbf{x}) \leq y_s, \forall s \in [S] \right\}, \quad (16)$$

where $T_m(\mathbf{y}) = \max_{\mathbf{u}} \{\langle \mathbf{u}, \mathbf{y} \rangle : \mathbf{0} \leq \mathbf{u} \leq \mathbf{1}, \mathbf{1}^\top \mathbf{u} \leq m\}$ denotes the sum of the m -largest elements of \mathbf{y} . To obtain a simplified reformulation, define the convex polyhedron

$$\mathcal{C} := \left\{ \mathbf{z} \in \mathbb{R}^S : 0 \leq z_s \leq 1 \forall s \in [S], \mathbf{1}^\top \mathbf{z} \geq S - m \right\}.$$

We can now obtain an equivalent reformulation of Problem (16) using the support function $\pi_{\mathcal{C}}$ of the polyhedron \mathcal{C} as follows:

$$v^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0}} \{f(\mathbf{x}) - \pi_{\mathcal{C}}(-\sigma \mathbf{y}) : g_s(\mathbf{x}) \leq y_s, \forall s \in [S]\}, \quad (17)$$

where $\pi_{\mathcal{C}}(\mathbf{y}) = \max_{\mathbf{z}} \{\mathbf{z}^\top \mathbf{y} : \mathbf{z} \in \mathcal{C}\}$.

A key observation is that Problem (17) is a polyhedral DC program since the nonsmooth concave part in the objective function $-\pi_{\mathcal{C}}(-\sigma \mathbf{y})$ is polyhedral. Moreover, its Fenchel conjugate takes a

simple form: the indicator function of the same polyhedron composed with a linear mapping. Therefore, a natural idea is to lift this problem into the dual space using the Toland duality [48, 28, 5] for DC programs, which can transfer the nonsmoothness in the objective function to simple polyhedral constraints. Compared with the primal problem (17), the dual problem takes the form of minimizing a concave function on a polyhedron, for which the standard DC algorithm guarantees finite convergence. This simple structure can potentially improve numerical stability regardless of the nonlinearity of the constraints and objective. We formally state this in the following proposition.

Proposition 2.3. *Suppose that Assumptions 1–2 hold. Consider Problem (17) in the unconstrained form*

$$\min_{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^S} \left\{ \zeta(\mathbf{x}, \mathbf{y}) - \eta(\mathbf{x}, \mathbf{y}) \right\},$$

where

$$\zeta(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + \delta_{\mathcal{X}}(\mathbf{x}) + \delta_{\mathbb{R}_+^S}(\mathbf{y}) + \delta_{\{(\mathbf{x}, \mathbf{y}) : g_s(\mathbf{x}) \leq y_s, \forall s \in [S]\}}(\mathbf{x}, \mathbf{y}), \quad \eta(\mathbf{x}, \mathbf{y}) := \pi_{\mathcal{C}}(-\sigma \mathbf{y}).$$

Let $(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^d \times \mathbb{R}^S$ denote the dual variables associated with (\mathbf{x}, \mathbf{y}) . Then the Fenchel conjugates of η and ζ satisfy:

(i)

$$\eta^*(\mathbf{p}, \mathbf{q}) = \delta_{\{\mathbf{0}\}}(\mathbf{p}) + \inf_{\mathbf{z} \in \mathbb{R}^S} \left\{ \delta_{\mathcal{C}}(\mathbf{z}) : -\sigma \mathbf{z} = \mathbf{q} \right\}.$$

(ii)

$$\zeta^*(\mathbf{p}, \mathbf{q}) = \sup_{\substack{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0} \\ g_s(\mathbf{x}) \leq y_s, \forall s \in [S]}} \left\{ \langle \mathbf{p}, \mathbf{x} \rangle + \langle \mathbf{q}, \mathbf{y} \rangle - f(\mathbf{x}) \right\}.$$

Consequently, by parameterizing the dual variable of the support function directly by $\mathbf{z} \in \mathcal{C}$, the Toland dual of Problem (17) can be written as

$$\min_{\mathbf{z} \in \mathcal{C}} \left(-\zeta^*(\mathbf{0}, -\sigma \mathbf{z}) \right) = \min_{\mathbf{z} \in \mathcal{C}} \inf_{\substack{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0} \\ g_s(\mathbf{x}) \leq y_s, \forall s \in [S]}} \left\{ f(\mathbf{x}) + \sigma \mathbf{z}^\top \mathbf{y} \right\}. \quad (18)$$

Proof. We derive the two Fenchel conjugates separately. The Toland dual problem (18) can then be derived via straightforward calculation.

(i): Computing η^* . Since η does not depend on \mathbf{x} , its conjugate separates as

$$\eta^*(\mathbf{p}, \mathbf{q}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \langle \mathbf{p}, \mathbf{x} \rangle + \sup_{\mathbf{y} \in \mathbb{R}^S} \left\{ \langle \mathbf{q}, \mathbf{y} \rangle - \pi_{\mathcal{C}}(-\sigma \mathbf{y}) \right\}.$$

The first term equals $\delta_{\{\mathbf{0}\}}(\mathbf{p})$. For the second term, write $h(\mathbf{u}) = \pi_{\mathcal{C}}(\mathbf{u})$, so that

$$\eta(\mathbf{y}) = (h \circ A)(\mathbf{y}), \quad A := -\sigma I.$$

Since h is the support function of \mathcal{C} , we have

$$h^*(\mathbf{z}) = \delta_{\mathcal{C}}(\mathbf{z}).$$

By the conjugacy rule for composition with a linear mapping,

$$(h \circ A)^*(\mathbf{q}) = \inf_{\mathbf{z} \in \mathbb{R}^S} \left\{ h^*(\mathbf{z}) : A^\top \mathbf{z} = \mathbf{q} \right\}.$$

Because $A^\top = -\sigma I$, it follows that

$$\sup_{\mathbf{y} \in \mathbb{R}^S} \left\{ \langle \mathbf{q}, \mathbf{y} \rangle - \pi_{\mathcal{C}}(-\sigma \mathbf{y}) \right\} = \inf_{\mathbf{z} \in \mathbb{R}^S} \left\{ \delta_{\mathcal{C}}(\mathbf{z}) : -\sigma \mathbf{z} = \mathbf{q} \right\}.$$

Therefore

$$\eta^*(\mathbf{p}, \mathbf{q}) = \delta_{\{\mathbf{0}\}}(\mathbf{p}) + \inf_{\mathbf{z} \in \mathbb{R}^S} \left\{ \delta_{\mathcal{C}}(\mathbf{z}) : -\sigma \mathbf{z} = \mathbf{q} \right\},$$

which proves part (i).

(ii): Computing ζ^* . By definition,

$$\zeta^*(\mathbf{p}, \mathbf{q}) = \sup_{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^S} \left\{ \langle \mathbf{p}, \mathbf{x} \rangle + \langle \mathbf{q}, \mathbf{y} \rangle - \zeta(\mathbf{x}, \mathbf{y}) \right\}.$$

Substituting the definition of ζ yields

$$\zeta^*(\mathbf{p}, \mathbf{q}) = \sup_{\substack{\mathbf{x} \in \mathcal{X}, \mathbf{y} \geq \mathbf{0} \\ g_s(\mathbf{x}) \leq y_s, \forall s \in [S]}} \left\{ \langle \mathbf{p}, \mathbf{x} \rangle + \langle \mathbf{q}, \mathbf{y} \rangle - f(\mathbf{x}) \right\},$$

which proves part (ii). □

We provide an alternative approach to deriving the same formulation (18) based on integer programming techniques, as discussed in the literature. Starting from (2), we introduce a binary vector $\mathbf{z} \in \{0, 1\}^S$ to represent the indicator functions. To obtain a lifted (higher-dimensional) reformulation, we can introduce another nonnegative vector $\mathbf{y} \in \mathbb{R}_+^S$ to indicate constraint violation instead of using the big-M formulation. A similar approach has been proposed in [30]. To make the paper self-contained, we state the result formally in the following proposition without proof.

Proposition 2.4. [30, proposition 1] *Problem (2) can be viewed as the following equivalent formulation:*

$$v^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_+^S, \mathbf{z} \in [0, 1]^S} \left\{ f(\mathbf{x}) : \sum_{s=1}^S z_s \geq S - m, V(\mathbf{y}, \mathbf{z}) = 0, g_s(\mathbf{x}) \leq y_s, \forall s \in [S] \right\}, \quad (19)$$

where $V(\mathbf{y}, \mathbf{z}) = \sum_{s=1}^S y_s z_s$.

We remark that the variable \mathbf{z} in formulation (19) can be either continuous or discrete. From this perspective, Problem (19) can be viewed as a special case of a mathematical program with complementarity constraint (MPCC) [34, 45]. In [30], the authors proposed an algorithm named ALSO-X+ to solve Problem (19) using alternating minimization with an iterative procedure. The algorithm can outperform the CVaR approximation [43] under Assumptions 1 and 2, and exhibits applicability in various settings. However, this algorithm lacks a stationarity guarantee and is inefficient due to the iterative procedure.

Motivated by these limitations, we instead study the following penalized formulation of Problem (19), which is equivalent to the Toland dual problem (18):

$$v_\sigma^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_+^S, \mathbf{z} \in [0,1]^S} \left\{ f(\mathbf{x}) + \sigma V(\mathbf{y}, \mathbf{z}) : \sum_{s=1}^S z_s \geq S - m, g_s(\mathbf{x}) \leq y_s, \forall s \in [S] \right\}, \quad (20)$$

where $\sigma > 0$ is the penalty parameter. Note that Problem (20) is the ℓ_1 exact penalty formulation [24] of (19). In particular, the term $V(\mathbf{y}, \mathbf{z})$ can be used directly as an exact penalty function without any positive-part truncation since $y_s \geq 0$ and $z_s \geq 0$ imply $y_s z_s \geq 0$ for all $s \in [S]$.

With this equivalence in hand, we take (20) as the starting point of our analysis since the penalty term only appears once in the objective function. In particular, we propose an algorithmic framework that simultaneously addresses the limitations of both Algorithm 1 and ALSO-X+.

2.3 Algorithm in the Lifted Space

In this subsection, we develop an alternative exact penalty framework for solving the lifted MPCC formulation (19) through its penalized counterpart (20). For any fixed penalty parameter $\sigma > 0$, the penalized model (20) is still nonconvex because of the bilinear term $\sum_{s=1}^S y_s z_s$ appearing in the objective. Nevertheless, from Proposition 2.3, we observe that once the remaining variables (\mathbf{x}, \mathbf{y}) are minimized out, this nonconvexity induces a concave minimization structure in the dual variable \mathbf{z} .

Before proceeding, we introduce several pieces of notation used throughout the remainder of this paper. Using the scenario-wise representation $g_s(\mathbf{x}) = \max_{i \in [I]} h_{s,i}(\mathbf{x})$, we write the feasible region of Problem (20) as

$$\Omega_0 := \left\{ (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathcal{X} \times \mathbb{R}_+^S \times \mathcal{C} : h_{s,i}(\mathbf{x}) \leq y_s, \forall s \in [S], \forall i \in [I] \right\}.$$

The objective function of Problem (20) is denoted by

$$F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + \sigma V(\mathbf{y}, \mathbf{z}).$$

Similarly, the feasible region of Problem (19) is

$$\Omega := \left\{ (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega_0 : y_s z_s = 0, \forall s \in [S] \right\}.$$

We next isolate the dependence on \mathbf{z} by introducing the following value function.

Proposition 2.5. *Let $\Psi : \mathbb{R}^S \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ be defined by*

$$\Psi(\mathbf{z}, \sigma) := \max_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_+^S} \left\{ -f(\mathbf{x}) - \sigma \sum_{s=1}^S y_s z_s : g_s(\mathbf{x}) \leq y_s, \forall s \in [S] \right\}.$$

Under Assumptions 1 and 2, the function $\Psi(\mathbf{z}, \sigma)$ is convex in \mathbf{z} for every fixed $\sigma > 0$.

Proof. For each fixed feasible pair (\mathbf{x}, \mathbf{y}) , $-F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is affine in \mathbf{z} . By definition, $\Psi(\cdot, \sigma)$ is the pointwise supremum of $-F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z})$ on Ω_0 . Thus $\Psi(\cdot, \sigma)$ is convex. \square

We remark here that the convexity of Ψ in \mathbf{z} does not depend on the convexity of the function $f(\mathbf{x})$ or of \mathcal{X} as long as the maximum is attained. Therefore, the following analysis can be naturally extended to the discrete setting, e.g., when $\mathcal{X} = \{0, 1\}^d$. With this notation, we can rewrite the penalized problem (20) as an optimization problem in \mathbf{z} only. Indeed, for any fixed \mathbf{z} , the inner minimization over (\mathbf{x}, \mathbf{y}) in (20) equals $-\Psi(\mathbf{z}, \sigma)$, and hence Problem (20) is equivalent to the following DC program

$$\min_{\mathbf{z} \in \mathcal{C}} -\Psi(\mathbf{z}, \sigma). \quad (21)$$

To implement a proximal DC method for (21), we need a computable subgradient of $\Psi(\cdot, \sigma)$ to linearize the concave part. This can be done by Danskin's theorem, where we can restrict the variable \mathbf{y} to a compact set under Assumptions 1 and 2 without loss of generality.

Proposition 2.6. [7, proposition B.25] *For any $\mathbf{z} \in \mathcal{C}$ and $\sigma > 0$, let $(\mathbf{x}^*(\mathbf{z}, \sigma), \mathbf{y}^*(\mathbf{z}, \sigma))$ be an optimal solution to the following subproblem:*

$$(\mathbf{x}^*(\mathbf{z}, \sigma), \mathbf{y}^*(\mathbf{z}, \sigma)) \in \arg \min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathbb{R}_+^S} \left\{ f(\mathbf{x}) + \sigma \sum_{s=1}^S y_s z_s : g_s(\mathbf{x}) \leq y_s, \forall s \in [S] \right\}. \quad (22)$$

Define

$$\mathbf{n}(\mathbf{z}, \sigma) := -\sigma(y_1^*(\mathbf{z}, \sigma), \dots, y_S^*(\mathbf{z}, \sigma))^\top.$$

Then $\mathbf{n}(\mathbf{z}, \sigma) \in \partial_{\mathbf{z}} \Psi(\mathbf{z}, \sigma)$.

We now apply a DC algorithm to solve the DC program (21) with fixed $\sigma > 0$. Given the current iterate $\mathbf{z}^k \in \mathcal{C}$ and $\sigma > 0$, we first choose a subgradient $\mathbf{n}^k \in \partial_{\mathbf{z}} \Psi(\mathbf{z}^k, \sigma)$, which can be computed via Proposition 2.6. Linearizing the concave term $-\Psi(\cdot, \sigma)$ at \mathbf{z}^k and adding a proximal regularizer yields the convex subproblem

$$\mathbf{z}^{k+1} = \arg \min_{\mathbf{z} \in \mathcal{C}} \left\{ -\langle \mathbf{n}^k, \mathbf{z} \rangle + \frac{\rho}{2} \|\mathbf{z} - \mathbf{z}^k\|_2^2 \right\} = \Pi_{\mathcal{C}}(\mathbf{z}^k - \frac{1}{\rho} \mathbf{n}^k),$$

where $\rho > 0$ is a proximal parameter to enhance numerical stability and avoid potential oscillation. Hence, each proximal DC update in the \mathbf{z} -space is reduced to a projection onto the simple polyhedron \mathcal{C} . This projection can be computed efficiently via sorting and scales well for large S .

The complete implementation of the penalty method is summarized in Algorithm 2. Similar to Algorithm 1, we start with a relatively small penalty parameter σ^0 and increase it geometrically. The penalty update continues until the violation of the chance constraint becomes sufficiently small, and further increasing σ will have little effect. Moreover, since the objective function in (20) is differentiable, the sequence generated by Algorithm 2 converges subsequentially to a stationary point of Problem (20), which can be defined using the standard first-order stationarity condition. Let $\mathbf{w} := (\mathbf{x}, \mathbf{y}, \mathbf{z})$. We say $\bar{\mathbf{w}} \in \Omega_0$ is a stationary point of (20) if the following equation holds:

$$\mathbf{0} \in \nabla F_\sigma(\bar{\mathbf{w}}) + \mathcal{N}_{\Omega_0}(\bar{\mathbf{w}}),$$

where $\mathcal{N}_{\Omega_0}(\bar{\mathbf{w}})$ is the normal cone of Ω_0 at $\bar{\mathbf{w}}$.

We make the following remarks on Algorithm 2:

- (i) For the fixed σ , the inner loop of Algorithm 2 can be viewed as the special case of the proximal point method for DC program proposed in [5] by letting the primal proximal step-size be 0.

Algorithm 2 Penalty Based Proximal DC Algorithm in the Lifted Space

Require: Initial $\mathbf{z}^{0,0} \in \mathcal{C}$, initial penalty $\sigma^0 > 0$, growth factor $\beta > 1$.

```
1: for  $t = 0, 1, 2, \dots$  do
2:   for  $k = 0, 1, 2, \dots$  do
3:     Find  $\mathbf{n}^{k,t} \in \partial_{\mathbf{z}}\Psi(\mathbf{z}^{k,t}, \sigma^t)$  via solving subproblem (22)
4:     Update  $\mathbf{z}^{k+1,t} = \Pi_{\mathcal{C}}(\mathbf{z}^{k,t} - \frac{1}{\rho}\mathbf{n}^{k,t})$ 
5:   end for
6:   Update  $\sigma^{t+1} = \beta \sigma^t$ .
7: end for
```

When $\rho = 0$, the inner loop for solving (21) converges in finite iterations. This is because the inner update reduces to the classical DCA step

$$\mathbf{z}^{k+1} \in \arg \min_{\mathbf{z} \in \mathcal{C}} \{-\langle \mathbf{n}^k, \mathbf{z} \rangle\},$$

which is a linear program over the polyhedron \mathcal{C} . Hence, an optimal solution can always be chosen at an extreme point of \mathcal{C} . The algorithm terminates in a finite number of iterations if the objective function is no longer strictly decreasing, since the number of extreme points of \mathcal{C} is finite. This property is structural and does not rely on the specific form of Ψ beyond Assumptions 1 and 2.

- (ii) Similar to Algorithm 1, we set ρ to be a small constant ($10^{-4} - 10^{-3}$) to improve numerical stability and mitigate oscillations. A larger ρ reduces the effective step-size $1/\rho$, leading to more conservative updates and noticeably longer runtime.
- (iii) In [30], the authors proposed a bisection-based approximation of Problem (2) named ALSO-X+ via the lifted formulation (19). Moreover, they found that, when solving the subproblems, the alternating minimization (AM) approach yields better solutions than a different DC approach (see, e.g., [29]). Specifically, their DC method is built on the following quadratic DC decomposition:

$$\mathbf{y}^\top \mathbf{z} = \frac{1}{4} \{ \|\mathbf{y} + \mathbf{z}\|^2 - \|\mathbf{y} - \mathbf{z}\|^2 \}.$$

We remark that Algorithm 2 can also be interpreted as an AM approach, but crucially without relying on the inefficient bisection procedure. In particular, for a fixed penalty parameter σ , Algorithm 2 proceeds by alternately performing two steps: (a) solving a convex subproblem in (\mathbf{x}, \mathbf{y}) with \mathbf{z} fixed; and (b) updating \mathbf{z} through a proximal projection step.

- (iv) The proposed framework admits effective warm starts in both the outer and inner loops.
 - (a) *Warm start for \mathbf{z} across outer penalty iterations.* In Algorithm 2, we reuse the final iterate $\mathbf{z}^{*,t}$ obtained at penalty level σ^t as the initializer $\mathbf{z}^{0,t+1}$ for the next penalty level σ^{t+1} .
 - (b) *Warm start for (\mathbf{x}, \mathbf{y}) across inner iterations.* Within Algorithm 2, the subproblem in Step 2 is solved repeatedly with changing \mathbf{z}^k , while the feasible set remains unchanged. Therefore, one can warm-start the solver for the (\mathbf{x}, \mathbf{y}) -subproblem using the previous solution $(\mathbf{x}^{k-1,t}, \mathbf{y}^{k-1,t})$ as an initial point when computing $(\mathbf{x}^{k,t}, \mathbf{y}^{k,t})$. This is particularly effective because only the objective coefficients associated with \mathbf{y} vary through \mathbf{z} ,

whereas all constraints stay fixed. For instance, we can use the simplex method to solve the subproblems when all the constraints and the objective function are linear.

- (c) *Early termination for small penalties.* For early penalty levels, it is often unnecessary to solve the fixed- σ subproblem to high accuracy, since constraint enforcement is weak when σ is small. In practice, one may terminate Algorithm 2 early for small σ , and gradually tighten the stopping criterion as σ increases.
- (v) In Algorithm 2, every (\mathbf{x}, \mathbf{y}) -subproblem (22) is convex, and can be solved efficiently by modern solvers such as Gurobi. Importantly, across iterations, the constraints are unchanged and only the linear objective coefficients in \mathbf{y} vary through $\mathbf{z}^{k,t}$. This structure makes warm-starting particularly effective in practice. When the constraints admit special forms (e.g., polyhedral constraints or linearly-constrained quadratic programs), the solver can amortize most of the presolve and factorization costs: the expensive preprocessing is essentially performed once, and subsequent solves are accelerated significantly.

3 Exact Penalty Frameworks

In this section, we aim to develop an exact penalty framework for both algorithms proposed in Section 2. To this end, we provide two complementary theoretical guarantees. First, under a constraint qualification, we show that the proposed penalty for both the primal and lifted formulation is globally exact, in the sense that all global minimizers of the original problem can be recovered by solving the penalized problem with a sufficiently large penalty parameter. Second, we focus on the local exactness. For the primal formulation, we prove the local exactness of the penalty and provide sufficient conditions to verify local optimality. For the lifted problem, instead of pursuing local exactness in terms of local minimizers, which typically requires strong regularity conditions (e.g., strict complementary slackness), we establish an equivalence between strongly stationary points of the original and the stationary points of the penalized formulation under a MPCC-tailored constraint qualification. This equivalence implies local optimality in the lifted space and serves as a theoretical guarantee for the developed algorithm.

3.1 Global Exact Penalty

In this section, we establish a global exact penalty relationship for the primal problem (10) and the lifted counterpart (19). The main challenge is that Problem (19) is an MPCC, for which standard constraint qualifications typically fail. To circumvent this difficulty, we adopt a recursive argument that exploits the fact that Problem (19) and the DC formulation (10) are equivalent at the level of global minimizers. Specifically, we first prove a global exact penalty result between Problem (10) and its penalized form (12) under mild and verifiable assumptions. We then transfer this result back to Problem (19) and (20) by leveraging their equivalence.

Since Problem (10) minimizes a convex function under a convex set with an additional DC constraint under Assumptions 1 and 2, we now introduce the Mangasarian–Fromovitz constraint qualification (MFCQ) based on the Clarke subdifferential.

Assumption 3. [10, theorem 2.4], [8, theorem 3.2] *For every $\bar{\mathbf{x}} \in \mathcal{X}$ with $\phi(\bar{\mathbf{x}}) = 0$, there exists*

a direction $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}})$ such that

$$\phi^\circ(\bar{\mathbf{x}}; \mathbf{d}) = \sup_{\mathbf{v} \in \partial\phi(\bar{\mathbf{x}})} \langle \mathbf{v}, \mathbf{d} \rangle < 0.$$

We remark that Assumption 3 is weaker than [50, Assumption 2], which imposes a generalized MFCQ condition for a DC constraint of the form $G_1(\mathbf{x}) - G_2(\mathbf{x}) \leq 0$. By [17, proposition 2.3.1, 2.3.3], we have

$$\partial(G_1 - G_2)(\mathbf{x}) \subseteq \partial G_1(\mathbf{x}) + \partial(-G_2)(\mathbf{x}) = \partial G_1(\mathbf{x}) - \partial G_2(\mathbf{x}).$$

Therefore, we have

$$\phi^\circ(\mathbf{x}; \mathbf{d}) = \sup_{\mathbf{v} \in \partial\phi(\mathbf{x})} \langle \mathbf{v}, \mathbf{d} \rangle \leq \sup_{\substack{\mathbf{s}_{G_1} \in \partial G_1(\mathbf{x}), \\ \mathbf{s}_{G_2} \in \partial G_2(\mathbf{x})}} \langle \mathbf{s}_{G_1} - \mathbf{s}_{G_2}, \mathbf{d} \rangle = \sup_{\mathbf{s}_{G_1} \in \partial G_1(\mathbf{x})} \langle \mathbf{s}_{G_1}, \mathbf{d} \rangle - \inf_{\mathbf{s}_{G_2} \in \partial G_2(\mathbf{x})} \langle \mathbf{s}_{G_2}, \mathbf{d} \rangle$$

for any $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\mathbf{x})$. The first equality is from [17, proposition 2.1.2(b)]. A sufficient condition for the equivalence between the two assumptions is that G_2 is differentiable at \mathbf{x} . The reader may refer to [50, remark 2] for additional verifiable sufficient conditions under which both constraint qualifications hold. We are now able to present the theorem establishing the global exact penalty between the primal Problem (10) and (12).

Theorem 3.1. *Suppose that Assumption 1, 2 and 3 hold. Let $\bar{\mathbf{x}} \in \mathcal{X}$ be given. Then the following statements are equivalent:*

- (a) $\bar{\mathbf{x}}$ is a global minimizer for Problem (10);
- (b) $\phi(\bar{\mathbf{x}}) = 0$ and there exists $\bar{\sigma} > 0$ such that $\bar{\mathbf{x}}$ is a global minimizer for Problem (12) for any $\sigma > \bar{\sigma}$.

Proof. By Theorem 1.3, it suffices to establish the global error bound

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \kappa [\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X}, \quad (23)$$

for some constant $\kappa > 0$, where $\hat{\mathcal{X}} := \{\mathbf{x} \in \mathcal{X} : \phi(\mathbf{x}) \leq 0\}$. By Proposition 1.1 and Assumption 3, for every boundary point $\bar{\mathbf{x}} \in \mathcal{X}$ with $\phi(\bar{\mathbf{x}}) = 0$, there exist constants $\kappa_{\bar{\mathbf{x}}} > 0$ and $\rho_{\bar{\mathbf{x}}} > 0$ such that

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \kappa_{\bar{\mathbf{x}}} [\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, \rho_{\bar{\mathbf{x}}}). \quad (24)$$

Let $\mathcal{B} := \mathcal{X} \cap \{\mathbf{x} : \phi(\mathbf{x}) = 0\}$. Since \mathcal{X} is compact and ϕ is continuous, \mathcal{B} is compact. The open cover $\{B(\bar{\mathbf{x}}, \rho_{\bar{\mathbf{x}}})\}_{\bar{\mathbf{x}} \in \mathcal{B}}$ admits a finite subcover $\{B(\bar{\mathbf{x}}^j, \rho_j)\}_{j=1}^J$. Define

$$\kappa_0 := \max_{j \in [J]} \kappa_{\bar{\mathbf{x}}^j}, \quad \mathcal{U} := \mathcal{X} \cap \bigcup_{j=1}^J B(\bar{\mathbf{x}}^j, \rho_j).$$

Then (24) implies the error bound $\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \kappa_0 [\phi(\mathbf{x})]_+$ holds for all $\mathbf{x} \in \mathcal{U}$.

Let $\mathcal{V} := \mathcal{X} \setminus \mathcal{U}$. Then \mathcal{V} is compact and, since $\mathcal{B} \subseteq \mathcal{U}$, we have $\mathcal{V} \cap \{\mathbf{x} : \phi(\mathbf{x}) = 0\} = \emptyset$. Thus $\pi := \min_{\mathbf{x} \in \mathcal{V}} |\phi(\mathbf{x})| > 0$. For any $\mathbf{x} \in \mathcal{V}$, either $\phi(\mathbf{x}) \leq 0$ with $\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) = 0$ or $\phi(\mathbf{x}) > 0$ with

$[\phi(\mathbf{x})]_+ \geq \pi$. In both cases, using $\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \text{diam}(\mathcal{X})$ for $\mathbf{x} \in \mathcal{X}$, we obtain

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \text{diam}(\mathcal{X}) \leq \frac{\text{diam}(\mathcal{X})}{\pi} [\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{V}.$$

Combining this estimate on \mathcal{V} with the error bound on \mathcal{U} proves the global bound (23). \square

Before proceeding, we note that the global exact penalty equivalence can also be established in a widely encountered structured piecewise-linear setting. This structure allows us to derive the same exact penalty result by exploiting polyhedral geometry and the finite combinatorial nature of active pieces, without invoking a generalized MFCQ condition. Such a setting arises in many practical applications, including our numerical experiments. We summarize this result in the following proposition.

Proposition 3.2. *Suppose that Assumption 2 holds, \mathcal{X} is a polyhedron and that each $h_{s,i}(\mathbf{x})$ is affine in \mathbf{x} . Then the following statements are equivalent:*

- (a) $\bar{\mathbf{x}}$ is a global minimizer for Problem (10);
- (b) $\phi(\bar{\mathbf{x}}) = 0$ and there exists $\bar{\sigma} > 0$ such that $\bar{\mathbf{x}}$ is a global minimizer for Problem (12) for any $\sigma > \bar{\sigma}$.

Proof. As in Theorem 1.3, it suffices to show that there exists $\kappa > 0$ such that

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \kappa [\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X}. \quad (25)$$

We divide the proof into two steps.

Step 1. A linear error bound for the union of $g_s(\mathbf{x}) \leq 0$ on \mathcal{X} . For any $T \subseteq [S]$, define

$$\mathcal{J}_T := \{\mathbf{x} \in \mathcal{X} : g_s(\mathbf{x}) \leq 0 \forall s \in T\}.$$

Since \mathcal{X} is a polyhedron and each $h_{i,s}(\mathbf{x})$ is affine, $g_s(\mathbf{x}) \leq 0$ is equivalent to the linear system $h_{i,s}(\mathbf{x}) \leq 0$ for all $i \in [I]$. Fix any linear inequality description $\mathcal{A}_T \mathbf{x} \leq \mathbf{c}_T$ of \mathcal{J}_T . By Hoffman's bound (see, e.g., [25]), there exists $\gamma_T > 0$ such that

$$\text{dist}(\mathbf{x}, \mathcal{J}_T) \leq \gamma_T \|(\mathcal{A}_T \mathbf{x} - \mathbf{c}_T)_+\|, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Since each inequality in $\mathcal{A}_T \mathbf{x} \leq \mathbf{c}_T$ is either from \mathcal{X} or of the form $h_{i,s}(\mathbf{x}) \leq 0$, there exists $C_T > 0$ such that $\|(\mathcal{A}_T \mathbf{x} - \mathbf{c}_T)_+\| \leq C_T \max_{s \in T} [g_s(\mathbf{x})]_+$ holds for all $\mathbf{x} \in \mathcal{X}$. Hence, with $\kappa_T := \gamma_T C_T$,

$$\text{dist}(\mathbf{x}, \mathcal{J}_T) \leq \kappa_T \max_{s \in T} [g_s(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X}.$$

Step 2. A global linear error bound for $\phi(\mathbf{x})$ on \mathcal{X} . The chance constraint in (2) is equivalent to $\hat{\mathcal{X}} = \bigcup_{|T|=S-m} \mathcal{J}_T$. Given $\mathbf{x} \in \mathcal{X}$, let $T(\mathbf{x}) \subseteq [S]$ be an index set of size $(S-m)$ attaining the $(S-m)$ smallest values among $\{g_s(\mathbf{x})\}_{s=1}^S$. Since $\text{dist}(\mathbf{x}, \bigcup_T \mathcal{J}_T) = \min_T \text{dist}(\mathbf{x}, \mathcal{J}_T)$ for a finite union of closed sets, we have $\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \text{dist}(\mathbf{x}, \mathcal{J}_{T(\mathbf{x})})$. By Remark (i) of Proposition 2.2, we have

$$\max_{s \in T(\mathbf{x})} [g_s(\mathbf{x})]_+ = [g_{(S-m)}(\mathbf{x})]_+ = [\phi(\mathbf{x})]_+.$$

Therefore,

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \text{dist}(\mathbf{x}, \mathcal{J}_{T(\mathbf{x})}) \leq \kappa_{T(\mathbf{x})}[\phi(\mathbf{x})]_+.$$

Let $\kappa := \max_{|T|=S-m} \kappa_T$. κ is finite since there are finitely many such T and this yields (25). \square

The exact penalty property established for the DC model (10) naturally suggests a parallel result for the lifted bilinear formulation (19). We finally show that the penalized lifted problem (20) is globally exact: for sufficiently large σ , minimizing (20) recovers the global minimizers of (19).

Theorem 3.3. *Suppose that Assumption 1, 2 and 3 hold. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ be feasible for Problem (20). Then the following two statements are equivalent:*

- (a) $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of Problem (19);
- (b) $V(\bar{\mathbf{y}}, \bar{\mathbf{z}}) = 0$ and there exists $\bar{\sigma} > 0$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of Problem (20) for all $\sigma > \bar{\sigma}$.

Proof. We divide the proof into three steps.

Step 1: A lower bound linking V and $[\phi]_+$. Let $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ be feasible for Problem (20). Since $y_s \geq 0$ and $g_s(\mathbf{x}) \leq y_s$, we have $y_s \geq [g_s(\mathbf{x})]_+$ for all s . Let $\psi_s(\mathbf{x}) := [g_s(\mathbf{x})]_+$ and let $\psi_{(1)}(\mathbf{x}) \leq \dots \leq \psi_{(S)}(\mathbf{x})$ denote the sorted values. Then

$$V(\mathbf{y}, \mathbf{z}) = \sum_{s=1}^S y_s z_s \geq \sum_{s=1}^S \psi_s(\mathbf{x}) z_s.$$

Let $\mathcal{C} = \{\mathbf{z} \in \mathbb{R}^S : \sum_{s=1}^S z_s \geq S - m, 0 \leq z_s \leq 1, \forall s\}$. The minimum of $\sum_{s=1}^S \psi_s(\mathbf{x}) z_s$ over $\mathbf{z} \in \mathcal{C}$ is attained by assigning weight 1 to the $S - m$ smallest components. Hence, for every $\mathbf{z} \in \mathcal{C}$,

$$\sum_{s=1}^S \psi_s(\mathbf{x}) z_s \geq \sum_{s=1}^{S-m} \psi_{(s)}(\mathbf{x}) \geq \psi_{(S-m)}(\mathbf{x}).$$

By Remark (i) of Proposition 2.2, $\phi(\mathbf{x}) = G_1(\mathbf{x}) - G_2(\mathbf{x}) = g_{(S-m)}(\mathbf{x})$, so $[\phi(\mathbf{x})]_+ = [g_{(S-m)}(\mathbf{x})]_+ = \psi_{(S-m)}(\mathbf{x})$. Therefore,

$$V(\mathbf{y}, \mathbf{z}) \geq [\phi(\mathbf{x})]_+, \quad \forall (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega_0. \quad (26)$$

Step 2: (a) \Rightarrow (b). Assume $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of (19). Then $V(\bar{\mathbf{y}}, \bar{\mathbf{z}}) = 0$ and $\bar{\mathbf{x}}$ is a global minimizer of (10) since the two problems are equivalent and coincide in the objective function. By Theorem 3.1, there exists $\bar{\sigma} > 0$ such that, for every $\sigma > \bar{\sigma}$,

$$f(\bar{\mathbf{x}}) \leq f(\mathbf{x}) + \sigma[\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X}. \quad (27)$$

Now fix any $\sigma > \bar{\sigma}$ and any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ feasible for (20). Using (27) and (26), we obtain

$$F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + \sigma V(\mathbf{y}, \mathbf{z}) \geq f(\mathbf{x}) + \sigma[\phi(\mathbf{x})]_+ \geq f(\bar{\mathbf{x}}) = F_\sigma(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}),$$

which shows that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of (20) for all $\sigma > \bar{\sigma}$.

Step 3: (b) \Rightarrow (a). Assume $V(\bar{\mathbf{y}}, \bar{\mathbf{z}}) = 0$ and $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of (20) for some $\sigma > 0$. Then $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ and is feasible for (19). For any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ feasible for (19), we have $V(\mathbf{y}, \mathbf{z}) = 0$

and thus $F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x})$. Global optimality of $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ for (20) yields

$$f(\bar{\mathbf{x}}) = F_\sigma(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \leq F_\sigma(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}), \quad \forall (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega,$$

so $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a global minimizer of (19). \square

3.2 Local Exact Penalty of the Primal Formulation

In this section, we aim to establish the local exact penalty relationship for the primal formulation (10) and its penalized counterpart (12). The result is identical to Theorem 1.2 under Assumption 3 and thus we omit the proof.

Theorem 3.4. *Suppose that Assumptions 1, 2 and 3 hold. Let $\bar{\mathbf{x}} \in \mathcal{X}$ be given. Then the following statements are equivalent:*

(a) $\bar{\mathbf{x}}$ is a local minimizer of Problem (10), i.e., there exists $r_0 > 0$ such that

$$f(\bar{\mathbf{x}}) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in \hat{\mathcal{X}} \cap B(\bar{\mathbf{x}}, r_0).$$

(b) $\phi(\bar{\mathbf{x}}) \leq 0$ and there exists $\bar{\sigma} > 0$ such that, for every $\sigma > \bar{\sigma}$, $\bar{\mathbf{x}}$ is a local minimizer of the penalized problem (12), i.e., there exists $r_\sigma > 0$ such that

$$f(\bar{\mathbf{x}}) + \sigma[\phi(\bar{\mathbf{x}})]_+ \leq f(\mathbf{x}) + \sigma[\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, r_\sigma).$$

We point out that Assumption 3 is not necessary in the structured piecewise-linear setting. Specifically, if \mathcal{X} is a polyhedron and each $h_{s,i}(\mathbf{x})$ is affine in \mathbf{x} , then one can derive a *global* linear error bound of the form

$$\text{dist}(\mathbf{x}, \hat{\mathcal{X}}) \leq \kappa [\phi(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X},$$

by the argument in Proposition 3.2. This global error bound immediately implies the above local exact penalty property. Moreover, since Algorithm 1 can only guarantee convergence to a critical point, in the above piecewise-linear setting, we can provide an equivalent characterization of local minimizers together with a simple verifiable sufficient condition.

Proposition 3.5. [18, proposition 6.1.1.] *Suppose Assumptions 1 and 2 hold. In addition, assume that each $h_{s,i}(\mathbf{x})$ is affine in \mathbf{x} . Then $\bar{\mathbf{x}} \in \mathcal{X}$ is a d-stationary point of Problem (12) if and only if it is a local minimizer of Problem (12).*

A sufficient condition for a critical point of a DC program to be d-stationary is that the concave part is differentiable at this point [18, proposition 6.1.10]. In the nonsmooth setting, this is equivalent to the fact that its Clarke subdifferential is a singleton. In our structured piecewise-linear setting, this differentiability requirement can be checked directly using the definition of G_2 in (8). G_2 is differentiable at $\bar{\mathbf{x}} \in \mathcal{X}$ provided that the following two uniqueness conditions hold:

(i) (*Unique top- m scenarios*) The m largest components of $(g_1(\bar{\mathbf{x}}), \dots, g_S(\bar{\mathbf{x}}))$ are uniquely identified, i.e.,

$$g_{(S-m)}(\bar{\mathbf{x}}) < g_{(S-m+1)}(\bar{\mathbf{x}}),$$

so that the index set of the top- m scenarios

$$\mathcal{S}_m(\bar{\mathbf{x}}) := \left\{ s \in [S] : g_s(\bar{\mathbf{x}}) \geq g_{(S-m+1)}(\bar{\mathbf{x}}) \right\}$$

satisfies $|\mathcal{S}_m(\bar{\mathbf{x}})| = m$;

- (ii) (*Unique active piece within each selected scenario*) For every $s \in \mathcal{S}_m(\bar{\mathbf{x}})$, the maximizer in $g_s(\bar{\mathbf{x}}) = \max_{i \in [I]} h_{s,i}(\bar{\mathbf{x}})$ is unique, i.e.,

$$\left| \arg \max_{i \in [I]} h_{s,i}(\bar{\mathbf{x}}) \right| = 1.$$

3.3 Equivalence between (Strongly) Stationary Points of the Lifted Formulation

In this section, we investigate the local connection between the bilinear model Problem (19) and its smooth penalized counterpart Problem (20). A well-known subtlety in MPCC is that even when the penalty parameter is sufficiently large, the set of local minimizers of an MPCC generally does not admit a one-to-one correspondence with that of its penalty counterpart [45, 21, 29].

Nevertheless, any accumulation point of the sequence generated by Algorithm 2 is a stationary point of Problem (20). Therefore, instead of seeking a direct equivalence between local minimizers, we adopt a more practical perspective by relating stationary points of the smooth penalized problem (20) to a suitable notion of generalized stationarity for Problem (19).

We now introduce the definition of strong stationarity for Problem (19). Let $\bar{\mathbf{w}} = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ be given. Define the index sets

$$\begin{aligned} \mathcal{I}_y(\bar{\mathbf{w}}) &:= \{s \in [S] : \bar{y}_s = 0, \bar{z}_s > 0\}, \\ \mathcal{I}_z(\bar{\mathbf{w}}) &:= \{s \in [S] : \bar{z}_s = 0, \bar{y}_s > 0\}, \\ \mathcal{I}_0(\bar{\mathbf{w}}) &:= \{s \in [S] : \bar{y}_s = \bar{z}_s = 0\}. \end{aligned}$$

Using these index sets, we define the *relaxed program* at $\bar{\mathbf{w}}$ by fixing the active-side variables to maintain the same complementarity pattern.

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Omega_0} \{f(\mathbf{x}) : y_s = 0, \forall s \in \mathcal{I}_y(\bar{\mathbf{w}}), z_s = 0, \forall s \in \mathcal{I}_z(\bar{\mathbf{w}})\}. \quad (28)$$

Definition 3.6 (Strong stationarity [29]). *Under Assumptions 1 and 2, a point $\bar{\mathbf{w}} \in \Omega$ is called a strongly stationary point of Problem (19) if it is an optimal solution to the convex relaxed program (28).*

Denote the feasible region of Problem (28) by

$$\Omega_{\text{re}}(\bar{\mathbf{w}}) := \left\{ (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega_0 : y_s = 0 \forall s \in \mathcal{I}_y(\bar{\mathbf{w}}), z_s = 0 \forall s \in \mathcal{I}_z(\bar{\mathbf{w}}) \right\}.$$

Since (28) is a convex program, $\bar{\mathbf{w}}$ solves it if and only if

$$\mathbf{0} \in \nabla f(\bar{\mathbf{x}}) \times \{\mathbf{0}\} \times \{\mathbf{0}\} + \mathcal{N}_{\Omega_{\text{re}}(\bar{\mathbf{w}})}(\bar{\mathbf{w}}). \quad (29)$$

The next theorem shows that every strongly stationary point is a local minimizer of Problem (19), which is known in the MPCC literature.

Theorem 3.7. [29, proposition 1] *Let $\bar{\mathbf{w}} = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ be a strongly stationary point of Problem (19). Then $\bar{\mathbf{w}}$ is a local minimizer of Problem (19).*

We next establish the connection between stationary points of the penalized problem (20) and strongly stationary points of (19). For this purpose, we adopt a standard constraint qualification tailored to MPCC models.

Definition 3.8 (MPCC-MFCQ [26, 29]). *Let $\bar{\mathbf{w}} \in \Omega$ be given. We say that MPCC-MFCQ holds at $\bar{\mathbf{w}}$ if the classical MFCQ holds at $\bar{\mathbf{w}}$ for the convex program (28).*

Finally with this constraint qualification at hand, we are able to establish the equivalence between the stationary points of the two problems.

Theorem 3.9. [29, proposition 6] *Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ be given. Suppose that Assumptions 1 and 2 hold.*

- (a) *If $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ is a stationary point for Problem (20) for some $\sigma > 0$ and is feasible for Problem (19), then it is a strongly stationary point of Problem (19).*
- (b) *If $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ is a strongly stationary point for Problem (19) and the MPCC-MFCQ holds, then it is a stationary point of Problem (20) for all σ sufficiently large.*

4 Identifying Spurious Local Minimizers

In the previous section, we have established the local exact penalty for the primal formulation and the fact that strong stationarity for the lifted formulation (19) implies local optimality in the lifted space. A natural question is whether such a local minimizer of the lifted formulation (19) also yields a meaningful local optimality guarantee for the primal formulation (10) in the \mathbf{x} -space. More precisely, given a local minimizer $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ of (19), whether the projection $\bar{\mathbf{x}}$ is a local minimizer of the DC formulation (10) remains unclear.

Our aim in this section is to answer this question. We show that such an implication from the \mathbf{x} -space to the lifted space always holds: every local minimizer $\bar{\mathbf{x}}$ of (10) can be lifted to a local minimizer of (19). However, the reverse does not hold in general: (19) may admit spurious local minimizers whose \mathbf{x} -components are not locally optimal for (10). We then identify a mild regularity condition under which the reverse statement holds.

Define $\psi_s(\mathbf{x}) := [g_s(\mathbf{x})]_+$ for all $s \in [S]$ and the set-valued mapping

$$Z(\mathbf{x}) = \arg \min \left\{ \sum_{s=1}^S \psi_s(\mathbf{x}) z_s : \mathbf{z} \in \mathcal{C} \right\}.$$

Since \mathcal{C} is nonempty, compact, and convex and $(\psi_1(\mathbf{x}), \dots, \psi_S(\mathbf{x}))$ is continuous in \mathbf{x} , Berge's maximum theorem [6] implies that $Z(\mathbf{x})$ is upper semicontinuous. We first show that every local minimum of the DC formulation (10) yields a local minimum of the lifted formulation (19).

Theorem 4.1. *Let $\bar{\mathbf{x}} \in \hat{\mathcal{X}}$ be a local minimizer of Problem (10). Then there exists $\bar{\mathbf{y}} \in \mathbb{R}_+^S$ and $\bar{\mathbf{z}} \in \mathcal{C}$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ is a local minimizer of Problem (19).*

Proof. Since $\bar{\mathbf{x}} \in \hat{\mathcal{X}}$, Proposition 2.1 implies that there exists an index set $\bar{J} \subseteq [S]$ with $|\bar{J}| = S - m$ such that $g_s(\bar{\mathbf{x}}) \leq 0$ for all $s \in \bar{J}$. Define $\bar{\mathbf{z}} \in \mathcal{C}$ by $\bar{z}_s = 1$ if $s \in \bar{J}$ and $\bar{z}_s := 0$ otherwise. Define $\bar{\mathbf{y}} \in \mathbb{R}_+^S$ by $\bar{y}_s = 0$ if $s \in \bar{J}$ and $\bar{y}_s := [g_s(\bar{\mathbf{x}})]_+$ otherwise. Then $\bar{\mathbf{y}} \geq 0$, $g_s(\bar{\mathbf{x}}) \leq \bar{y}_s$ for all s , and $\sum_{s=1}^S \bar{y}_s \bar{z}_s = 0$, hence $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$.

Let $\varepsilon > 0$ be such that $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for all $\mathbf{x} \in \hat{\mathcal{X}} \cap B(\bar{\mathbf{x}}, \varepsilon)$. Take any $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega$ with $\|(\mathbf{x}, \mathbf{y}, \mathbf{z}) - (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})\| < \varepsilon$. Therefore, we have $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for every $\|(\mathbf{x}, \mathbf{y}, \mathbf{z}) - (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})\| < \varepsilon$ and $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega$ since Problem (10) and (19) have the same objective function. \square

However, the inverse of Theorem 4.1 does not hold without additional assumptions. This can be demonstrated using the following one-dimensional example.

Example 4.2. Consider the one-dimensional setting with $S = 2$, $m = 1$, $\mathcal{X} = [-1, +\infty)$. Let $f(x) = x$, the sample based constraints $g_1(x) = [x]_+$ and $g_2(x) = [-x]_+$. Since $g_s(x) \geq 0$, we have

$$G_1(x) - G_2(x) = \min\{[x]_+, [-x]_+\} = 0, \quad \forall x \in \mathcal{X}.$$

Therefore, the DC formulation (10) reduces to $\min\{x : x \geq -1\}$, whose unique global minimizer is $x = -1$.

We now show that $x = 0$ can nevertheless appear as the \mathbf{x} -component of a local minimizer of the lifted formulation, which can be written as follows

$$\min_{x, \mathbf{y} \geq 0, \mathbf{z} \in \mathcal{C}} \{x : g_1(x) \leq y_1, g_2(x) \leq y_2, y_1 z_1 + y_2 z_2 = 0\}.$$

Take $\bar{x} = 0$, $\bar{\mathbf{y}} = (0, 0)$, and $\bar{\mathbf{z}} = (1, 1)$. Then $(\bar{x}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ since $g_1(0) = g_2(0) = 0$, $\sum_{s=1}^2 \bar{z}_s = 2 \geq S - m = 1$, and $\sum_{s=1}^2 \bar{y}_s \bar{z}_s = 0$. Moreover, there exists $r \in (0, 1/4)$ such that for any feasible $(x, \mathbf{y}, \mathbf{z}) \in \Omega$ with $\|(x, \mathbf{y}, \mathbf{z}) - (\bar{x}, \bar{\mathbf{y}}, \bar{\mathbf{z}})\| < r$, we have $z_1 > 1/2$ and $z_2 > 1/2$, which together with $\mathbf{y} \geq 0$ and $\sum_{s=1}^2 y_s z_s = 0$ implies $y_1 = y_2 = 0$. Hence feasibility forces $g_1(x) \leq 0$ and $g_2(x) \leq 0$, i.e., $[x]_+ = [-x]_+ = 0$, so necessarily $x = 0$. Consequently, every feasible point in a sufficiently small neighborhood of $(\bar{x}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ has the same objective value $f(x) = 0$, showing that $(\bar{x}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is a local minimizer of (19) while $\bar{x} = 0$ is not a local minimizer of (10).

In the above example, $\mathcal{C} = \{(z_1, z_2) \in [0, 1]^2 : z_1 + z_2 \geq S - m = 1\}$ and one can verify that

$$Z(x) = \begin{cases} \{(1, 0)\}, & x < 0, \\ \mathcal{C}, & x = 0, \\ \{(0, 1)\}, & x > 0. \end{cases}$$

Example 4.2 arises because, at $\bar{x} = 0$, the set-valued mapping $Z(\mathbf{x})$ contains isolated elements. In particular, for any sequence $x^n \downarrow 0$ we have $Z(x^n) = \{(0, 1)\}$. Similarly, for any sequence $x^n \uparrow 0$ we have $Z(x^n) = \{(1, 0)\}$ and thus no choice $\mathbf{z}^n \in Z(x^n)$ can converge to $(1, 1)$. This is precisely the pathology that allows $(x, \mathbf{y}, \mathbf{z}) = (0, (0, 0), (1, 1))$ to be a local minimizer of (19) even though 0 is not a local minimizer of the DC formulation (10). We next introduce an assumption that can eliminate such spurious isolated local minimizers.

Theorem 4.3. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ be a local minimizer of Problem (19). Suppose that $Z(\mathbf{x})$ is lower semicontinuous at $\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}}$ is a local minimizer of Problem (10).

Proof. Define $\tilde{\mathbf{y}} \in \mathbb{R}^S$ by $\tilde{y}_s = \psi_s(\bar{\mathbf{x}})$ for all $s \in [S]$. We first show that $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$ and is also a local minimizer of Problem (19).

Since $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$, we have $g_s(\bar{\mathbf{x}}) \leq \bar{y}_s$, $\bar{y}_s \geq 0$, and $\bar{y}_s \bar{z}_s = 0$ for all $s \in [S]$. Hence $\tilde{y}_s = [g_s(\bar{\mathbf{x}})]_+ \leq \bar{y}_s$, so $g_s(\bar{\mathbf{x}}) \leq \tilde{y}_s$ and $\tilde{y}_s \geq 0$ for all s . Since $\tilde{\mathbf{y}} \leq \bar{\mathbf{y}}$, $\tilde{\mathbf{y}} \geq \mathbf{0}$, and $\bar{\mathbf{y}}^\top \bar{\mathbf{z}} = 0$, we get $\tilde{\mathbf{y}}^\top \bar{\mathbf{z}} = 0$. Thus $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$.

Suppose, to the contrary, that $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})$ is not a local minimizer of Problem (19). Then there exists a sequence $\{(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\} \subset \Omega$ such that $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k) \rightarrow (\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})$ and $f(\mathbf{x}^k) < f(\bar{\mathbf{x}})$ for all k .

Partition $[S]$ into

$$I^+ := \{s : \bar{z}_s > 0\}, \quad I_{>}^{0,+} := \{s : \bar{z}_s = 0, \bar{y}_s > \psi_s(\bar{\mathbf{x}})\},$$

$$I_{\leq}^{0,+} := \{s : \bar{z}_s = 0, \bar{y}_s = \psi_s(\bar{\mathbf{x}}) > 0\}, \quad I^{0,0} := \{s : \bar{z}_s = 0, \bar{y}_s = 0\}.$$

For each k , define $\hat{\mathbf{y}}^k \in \mathbb{R}^S$ by

$$\hat{y}_s^k := \begin{cases} 0, & s \in I^+, \\ \bar{y}_s, & s \in I_{>}^{0,+}, \\ [g_s(\mathbf{x}^k)]_+, & s \in I_{\leq}^{0,+} \cup I^{0,0}. \end{cases}$$

Then $(\mathbf{x}^k, \hat{\mathbf{y}}^k, \bar{\mathbf{z}}) \in \Omega$ for all sufficiently large k : if $s \in I^+$, then $\bar{z}_s > 0$, so $z_s^k > 0$ for all sufficiently large k ; since $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k) \in \Omega$, $y_s^k \geq 0$, $z_s^k \geq 0$, and $\sum_{t=1}^S y_t^k z_t^k = 0$, hence $y_s^k z_s^k = 0$ and therefore $y_s^k = 0$. Thus $g_s(\mathbf{x}^k) \leq y_s^k = 0 = \hat{y}_s^k$. If $s \in I_{>}^{0,+}$, then $\bar{y}_s > \psi_s(\bar{\mathbf{x}}) = [g_s(\bar{\mathbf{x}})]_+ \geq g_s(\bar{\mathbf{x}})$, so by continuity of g_s , $g_s(\mathbf{x}^k) < \bar{y}_s = \hat{y}_s^k$ for all sufficiently large k . If $s \in I_{\leq}^{0,+} \cup I^{0,0}$, then by definition $\hat{y}_s^k = [g_s(\mathbf{x}^k)]_+ \geq g_s(\mathbf{x}^k)$ and $\hat{y}_s^k \geq 0$. In all cases, since $\bar{z}_s = 0$ for $s \notin I^+$, we have $\hat{y}_s^k \bar{z}_s = 0$ for every s .

Moreover, $\hat{\mathbf{y}}^k \rightarrow \bar{\mathbf{y}}$: for $s \in I^+$, both are 0; for $s \in I_{>}^{0,+}$, $\hat{y}_s^k = \bar{y}_s$; and for $s \in I_{\leq}^{0,+} \cup I^{0,0}$,

$$\hat{y}_s^k = [g_s(\mathbf{x}^k)]_+ \rightarrow [g_s(\bar{\mathbf{x}})]_+ = \psi_s(\bar{\mathbf{x}}) = \bar{y}_s.$$

Thus $(\mathbf{x}^k, \hat{\mathbf{y}}^k, \bar{\mathbf{z}}) \rightarrow (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$. Since the objective of Problem (19) depends only on \mathbf{x} , we have $f(\mathbf{x}^k) = f(\mathbf{x}^k, \hat{\mathbf{y}}^k, \bar{\mathbf{z}}) < f(\bar{\mathbf{x}})$ for all k , contradicting the local minimality of $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$. Therefore, $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})$ is a local minimizer of Problem (19).

Now let $r > 0$ be such that

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) \quad \forall (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Omega \text{ with } \|(\mathbf{x}, \mathbf{y}, \mathbf{z}) - (\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})\| < r.$$

Since $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}}) \in \Omega$, we have $\bar{\mathbf{z}} \in \mathcal{C}$ and $\sum_{s=1}^S \tilde{y}_s \bar{z}_s = 0$. As $\tilde{y}_s = \psi_s(\bar{\mathbf{x}}) \geq 0$ for all s , it follows that $\sum_{s=1}^S \psi_s(\bar{\mathbf{x}}) \bar{z}_s = 0$, so $\bar{\mathbf{z}}$ attains the minimum of

$$\min \left\{ \sum_{s=1}^S \psi_s(\bar{\mathbf{x}}) z_s : \mathbf{z} \in \mathcal{C} \right\}.$$

Hence $\bar{\mathbf{z}} \in Z(\bar{\mathbf{x}})$.

Fix $\varepsilon := r/3$. Since Z is lower semicontinuous at $\bar{\mathbf{x}}$ and $\bar{\mathbf{z}} \in Z(\bar{\mathbf{x}})$, there exists $\rho_1 > 0$ such that for every $\mathbf{x} \in \mathcal{X}$ with $\|\mathbf{x} - \bar{\mathbf{x}}\| < \rho_1$, one can choose $\mathbf{z}(\mathbf{x}) \in Z(\mathbf{x})$ satisfying $\|\mathbf{z}(\mathbf{x}) - \bar{\mathbf{z}}\| < \varepsilon$. Since

ψ is continuous, there exists $\rho_2 > 0$ such that $\|\psi(\mathbf{x}) - \psi(\bar{\mathbf{x}})\| < \varepsilon$ whenever $\|\mathbf{x} - \bar{\mathbf{x}}\| < \rho_2$.

Set $\rho := \min\{\rho_1, \rho_2, \varepsilon\}$ and take any $\mathbf{x} \in \hat{\mathcal{X}} \cap B(\bar{\mathbf{x}}, \rho)$. Let $\mathbf{y}(\mathbf{x}) := \psi(\mathbf{x})$ and choose $\mathbf{z}(\mathbf{x}) \in Z(\mathbf{x})$ as above. Since $\mathbf{x} \in \hat{\mathcal{X}}$, the optimal value of

$$\min\left\{\sum_{s=1}^S \psi_s(\mathbf{x})z_s : \mathbf{z} \in \mathcal{C}\right\}$$

is 0, so $\sum_{s=1}^S y_s(\mathbf{x})z_s(\mathbf{x}) = 0$. Also, $g_s(\mathbf{x}) \leq [g_s(\mathbf{x})]_+ = y_s(\mathbf{x})$ and $y_s(\mathbf{x}) \geq 0$ for all s , hence $(\mathbf{x}, \mathbf{y}(\mathbf{x}), \mathbf{z}(\mathbf{x})) \in \Omega$.

Furthermore,

$$\|(\mathbf{x}, \mathbf{y}(\mathbf{x}), \mathbf{z}(\mathbf{x})) - (\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})\| \leq \|\mathbf{x} - \bar{\mathbf{x}}\| + \|\mathbf{y}(\mathbf{x}) - \tilde{\mathbf{y}}\| + \|\mathbf{z}(\mathbf{x}) - \bar{\mathbf{z}}\| < \varepsilon + \varepsilon + \varepsilon = r.$$

Therefore, by the local minimality of $(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}})$,

$$f(\mathbf{x}) = f(\mathbf{x}, \mathbf{y}(\mathbf{x}), \mathbf{z}(\mathbf{x})) \geq f(\bar{\mathbf{x}}, \tilde{\mathbf{y}}, \bar{\mathbf{z}}) = f(\bar{\mathbf{x}}).$$

Since $\mathbf{x} \in \hat{\mathcal{X}} \cap B(\bar{\mathbf{x}}, \rho)$ was arbitrary, $\bar{\mathbf{x}}$ is a local minimizer of Problem (10). \square

A simple sufficient condition for the lower semicontinuity of $Z(\mathbf{x})$ at $\bar{\mathbf{x}}$ is that it is a singleton, which has been proved in [9]. In our setting, a standard sufficient condition ensuring that $Z(\bar{\mathbf{x}})$ is a singleton is that the threshold exhibits a strict gap, namely

$$\psi_{(S-m)}(\bar{\mathbf{x}}) = 0 < \psi_{(S-m+1)}(\bar{\mathbf{x}}),$$

which implies that the set of $(S-m)$ smallest components of $\{\psi_1(\bar{\mathbf{x}})\}_{s=1}^S$ is uniquely determined and thus the set-valued mapping $Z(\bar{\mathbf{x}})$ is a singleton.

Finally, we emphasize that the lower semicontinuity of Z is only a sufficient condition: spurious local minima may occur at isolated lifted points, but not every isolated lifted local minimizer is spurious in the sense of having an \mathbf{x} -component that fails to be locally optimal for the DC formulation.

Example 4.4. *Let us revisit Example 4.2 but replace the objective by $f(x) = x^2$. Since $\phi(x) \equiv 0$ on \mathcal{X} , the DC formulation (10) becomes $\min\{x^2 : x \geq -1\}$, so $x = 0$ is a global, hence local minimizer. On the other hand, the lifted point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) = (0, (0, 0), (1, 1))$ remains feasible for (19) and is again an isolated local minimizer in the lifted space by the same neighborhood argument as in Example 4.2. In this case, however, the lifted local minimizer is not spurious: its \mathbf{x} -component $\bar{\mathbf{x}} = 0$ is also a local minimizer of the primal formulation, even though $Z(\cdot)$ is still not lower semicontinuous at 0.*

5 Numerical Experiments

In this section, we conduct experiments to test the performance of the proposed algorithms in Section 2 on both real and synthetic datasets. For ease of reference, we denote our proposed Algorithm 1 by PenDC-P(riimal) and Algorithm 2 by PenDC-L(ifted). We compare our algorithms

with some other state-of-the-art methods, including the CVaR approximation [43, 40], mixed-integer formulation (30) (MIP) proposed in [2], a DC based algorithm (DCA) in [50], and the bisection-based approximation algorithm ALSO-X+ proposed in [30]. For the MIP formulation, we directly solve the following mixed-integer program:

$$v^* = \min_{\mathbf{x} \in \mathcal{X}, \mathbf{z} \in \{0,1\}^S} \left\{ f(\mathbf{x}) : \sum_{s=1}^S z_s \geq S - m, g_s(\mathbf{x}) \leq (1 - z_s)M_s, \forall s \in [S] \right\}, \quad (30)$$

In particular, we use Gurobi (v12.0.3) to solve all linear, quadratic, and mixed-integer (sub)problems. We set a time limit of 600 seconds for all the MIP runs with the default optimality gap tolerance of 0.01%. For instances that cannot be solved to optimality within the time limit, we use “gap” to denote averaged optimality gap as $\text{gap}(\%) = (|\text{UB} - \text{LB}|)/(|\text{LB}|) \times 100$ among the instances. We use “fval” to denote averaged returned objective values of the instances, “time” to denote the averaged CPU time (in seconds) and “prob” to denote the average probability of the chance constraint. We use “/” to indicate that the algorithm fails to provide a valid feasible solution for at least one of the instances. For each setting, we generate five random instances and report the average performance over these instances. The row “solved” reports the number of instances returned with a feasible solution among those five instances.

For the outer loop of the proposed method, we terminate the algorithm when the sample based chance constraint is satisfied. For the inner loop of the proposed methods and all the test methods except ALSO-X+, we terminate the algorithm when $|f^k - f^{k+1}| / \max\{1, |f^{k+1}|\} \leq 10^{-6}$ for $k = 0, 1, 2, \dots$. For PenDC-P and PenDC-L, the initial points are chosen randomly. For the first two outer iterations of the two penalty methods, we terminate the inner loop after 1 and 2 iterations, respectively, as a warm start. For ALSO-X+, we terminate the algorithm when the difference between the upper bound and the lower bound of the objective value $|t^U - t^L| \leq 10^{-6}$. All numerical experiments in this section are implemented in Python 3.9.21 and executed on a Linux server equipped with 256 GB RAM and a 96-core AMD EPYC 7402 CPU running at 2.8 GHz.

5.1 A VaR-constrained Portfolio Optimization Problem

In this section, we consider a Value-at-Risk (VaR) constrained mean–variance portfolio selection model studied in [4, 50]. Let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ denote the estimated mean vector and covariance matrix of returns for n risky assets, and let $\gamma > 0$ be a risk-aversion parameter. Let $\mathbf{x} \in \mathbb{R}_+^n$ denote the portfolio weight vector. The problem can be formulated as follows:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \gamma \mathbf{x}^\top \boldsymbol{\Sigma} \mathbf{x} - \boldsymbol{\mu}^\top \mathbf{x} : \Pr\{\tilde{\boldsymbol{\xi}}^\top \mathbf{x} \geq R\} \geq 1 - \alpha, \sum_{i=1}^n x_i = 1, 0 \leq x_i \leq u_i, i \in [n] \right\}.$$

The problem can be interpreted as follows: the objective minimizes a quadratic variance penalty minus expected return, subject to a chance constraint requiring that the realized portfolio return $\boldsymbol{\xi}^\top \mathbf{x}$ exceeds a prespecified target R with probability at least $1 - \alpha$. In addition, we impose the budget constraint $\sum_{i=1}^n x_i = 1$ and box constraints $0 \leq x_i \leq u_i$ for each $i \in [n]$ to avoid overconcentration.

To construct test instances from real data, we use a dataset of 2523 daily returns of 435

stocks in the S&P 500 index over March 2006 to March 2016, which can be downloaded from <https://github.com/INFORMSJoC/2024.0648>. Specifically, we consider four problem sizes $n \in \{100, 200, 300, 400\}$ and set the sample size to $S = 3n$. For each pair (n, S) , we run the algorithms on five independent instances and report the average performance metrics over these five instances. For the PenDC-L method, we set $\sigma_0 = 5 \cdot 10^{-3}$, $\beta = 4.0$, and $\rho = 10^{-4}$. For the PenDC-P method, we set $\sigma_0 = 3 \cdot 10^{-3}$, $\beta = 1.5$, and $\rho = 0$.

Table 1: Comparisons of the portfolio optimization problem.

(α, S)		MIP	CVaR	PenDC-L	PenDC-P	DCA	ALSO-X+
$(0.05, 300)$	fval	-0.013550	-0.011785	-0.013398	-0.013053	-0.012125	-0.013097
	time(gap)	207.1164 (0.0%)	0.1093	0.1109	0.1936	0.4343	3.3849
	prob	0.9500	1.0000	0.9500	0.9507	0.9653	0.9500
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.05, 600)$	fval	-0.013508	-0.011790	-0.013379	-0.012992	-0.012527	-0.013208
	time(gap)	600.0085 (4.8%)	0.6746	0.3961	0.9973	5.4026	18.6961
	prob	0.9500	1.0000	0.9507	0.9500	0.9620	0.9507
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.05, 900)$	fval	-0.013428	-0.011697	-0.013382	-0.012817	-0.012850	-0.013144
	time(gap)	600.0144 (8.7%)	1.9169	0.9956	2.5666	33.2572	58.1248
	prob	0.9500	1.0000	0.9504	0.9500	0.9182	0.9504
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.05, 1200)$	fval	-0.013557	-0.011825	-0.013597	-0.013247	-0.013090	-0.013273
	time(gap)	600.0488 (10.6%)	4.7453	2.2373	7.1035	83.0278	155.2660
	prob	0.9504	1.0000	0.9507	0.9505	0.9248	0.9503
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.10, 300)$	fval	-0.014429	-0.011785	-0.014281	-0.014138	-0.013522	-0.014108
	time(gap)	36.5684 (0.0%)	0.1085	0.1002	0.1781	1.1241	3.3367
	prob	0.9000	1.0000	0.9060	0.9007	0.9020	0.9007
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.10, 600)$	fval	-0.014213	-0.011790	-0.014151	-0.013986	-0.013718	-0.014073
	time(gap)	600.0077 (1.4%)	0.6771	0.3166	0.8218	11.4020	17.5378
	prob	0.9003	1.0000	0.9057	0.9000	0.8953	0.9000
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.10, 900)$	fval	-0.014347	-0.011697	-0.014272	-0.013964	-0.013831	-0.014196
	time(gap)	600.0097 (3.1%)	1.9063	0.8954	2.3002	43.9097	54.7558
	prob	0.8998	1.0000	0.9058	0.9000	0.9042	0.9007
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$(0.10, 1200)$	fval	-0.014558	-0.011825	-0.014455	-0.014448	-0.014253	-0.014453
	time(gap)	600.0161 (3.6%)	4.7528	1.9307	6.4855	154.4464	141.1504
	prob	0.9015	1.0000	0.9107	0.9000	0.8973	0.9013
	solved	5/5	5/5	5/5	5/5	5/5	5/5

The results in Table 1 demonstrate that PenDC-L performs competitively in terms of solution quality and computational efficiency on the test problems. Across all tested combinations of (α, S) , PenDC-L produces objective values that are nearly identical to the MIP benchmark whenever the MIP is solved to optimality, while requiring orders-of-magnitude less runtime and remaining stable as n increases. In contrast, the CVaR approximation is fast but systematically more conservative, as reflected by its empirical satisfaction probabilities being close to 1 in all instances. This conservatism directly translates into a noticeable objective gap. Compared with the DCA, PenDC-P can improve the performance of DCA by producing a feasible solution with lower objective value and shorter running time. For the ALSO-X+ method, we can see that it can return a feasible solution with the lowest objective value among the remaining methods. However, this comes at the cost of substantially longer runtime.

5.2 A Probabilistic Resource Planning Problem

We next consider a chance constrained linear resource planning problem, which has been studied in [38, 4]. The model concerns shipping goods from n suppliers to m customers at minimum total transportation cost, where customer demands are uncertain. Specifically, the demand of customer $j \in [m]$ is modeled by a random variable ξ_j , while each supplier $i \in [n]$ is subject to a deterministic capacity limit $\theta_i > 0$. Let $c_{ij} \geq 0$ denote the unit shipping cost from supplier i to customer j , and let $x_{ij} \geq 0$ be the amount shipped along arc (i, j) . The shipment plan is decided in advance of demand realization, and feasibility is enforced in a probabilistic sense by requiring that all customer demands are met simultaneously with probability at least $(1 - \alpha)$. The resulting formulation is

$$\min_{\mathbf{x} \in \mathbb{R}^{n \times m}} \left\{ \begin{array}{l} \sum_{i=1}^n \sum_{j=1}^m c_{ij} x_{ij} : \\ \Pr \left\{ \sum_{i=1}^n x_{ij} \geq \xi_j, \forall j \in [m] \right\} \geq 1 - \alpha, \\ \sum_{j=1}^m x_{ij} \leq \theta_i, \forall i \in [n], \quad x_{ij} \geq 0, \quad \forall i \in [n], j \in [m] \end{array} \right\}.$$

We use the public dataset provided at <http://homepages.cae.wisc.edu/~luedtkej/>. Throughout, we let $(n, m) \in \{(40, 100), (40, 200)\}$ and consider three sample sizes $S \in \{1000, 2000, 3000\}$. For both the PenDC-L and PenDC-P methods, we set $\sigma_0 = 5$, $\beta = 4.5$, and $\rho = 10^{-3}$.

Table 2: Comparisons of the probabilistic resource planning problem.

(α, n, S)		MIP	CVaR	PenDC-L	PenDC-P	DCA	ALSO-X+
$\left(\begin{array}{l} 0.05 \\ 100 \\ 1000 \end{array} \right)$	fval	4.1309	/	4.1422	4.2554	4.2560	4.1754
	time(gap)	143.8471 (0.0%)	/	19.2711	25.2524	22.7656	67.8204
	prob	0.9500	/	0.9502	0.9500	0.9500	0.9500
	solved	5/5	4/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{l} 0.05 \\ 100 \\ 2000 \end{array} \right)$	fval	4.4168	4.6644	4.4264	4.6088	4.6083	4.4578
	time(gap)	429.6553 (0.0%)	1.8925	54.8647	59.5165	46.3589	189.1209
	prob	0.9500	1.0000	0.9500	0.9500	0.9500	0.9503
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{l} 0.05 \\ 200 \\ 2000 \end{array} \right)$	fval	8.5613	/	8.3748	8.6736	/	8.4288
	time(gap)	600.2509 (6.4%)	/	204.4876	457.3059	/	645.4984
	prob	0.9638	/	0.9500	0.9278	/	0.9501
	solved	5/5	4/5	5/5	5/5	4/5	5/5
$\left(\begin{array}{l} 0.05 \\ 200 \\ 3000 \end{array} \right)$	fval	8.8224	/	8.5996	8.9440	8.9440	8.6534
	time(gap)	600.3817 (12.6%)	/	334.5961	220.8843	165.6295	1302.5201
	prob	0.9654	/	0.9500	0.9500	0.9500	0.9500
	solved	5/5	4/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{l} 0.10 \\ 100 \\ 1000 \end{array} \right)$	fval	4.0540	/	4.0663	4.2306	4.2306	4.0847
	time(gap)	406.5311 (0.0%)	/	19.3168	25.8939	25.8532	81.7118
	prob	0.9000	/	0.9000	0.9000	0.9000	0.9000
	solved	5/5	4/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{l} 0.10 \\ 100 \\ 2000 \end{array} \right)$	fval	4.3221	4.6644	4.3318	4.5787	4.5800	4.3581
	time(gap)	600.3751 (0.6%)	1.8589	52.9770	62.6593	50.3822	235.6824
	prob	0.9004	1.0000	0.9002	0.9000	0.9000	0.9000
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{l} 0.10 \\ 200 \\ 2000 \end{array} \right)$	fval	8.4548	/	8.2194	/	/	8.2649
	time(gap)	600.2532 (12.4%)	/	220.6448	/	/	844.3599
	prob	0.9079	/	0.9002	/	/	0.9001
	solved	5/5	4/5	5/5	4/5	4/5	5/5
$\left(\begin{array}{l} 0.10 \\ 200 \\ 3000 \end{array} \right)$	fval	8.7236	/	8.4273	8.8973	8.8974	8.4733
	time(gap)	602.9494 (18.0%)	/	352.0631	235.6632	191.6438	1388.0507
	prob	0.9171	/	0.9001	0.9000	0.9000	0.9001
	solved	5/5	4/5	5/5	5/5	5/5	5/5

*The magnitude of fval is 10^7 .

Table 2 highlights several distinctive computational features of the compared methods on the resource planning problem. Among the iterative approaches, DCA is often the fastest whenever it succeeds, because DCA has a finite termination guarantee in the polyhedral setting, and the number of iterations is smaller than that of the penalty based methods. In this numerical setting, the DCA method usually converges within only 3-5 iterations. However, Table 2 also shows that both DCA and the CVaR approximation fail to provide a valid solution under some instances, which is consistent with the fact that both reformulations may become overly conservative and therefore infeasible. In comparison, the PenDC-P method can sometimes alleviate infeasibility caused by overly conservative subproblems, and can often return a lower objective value with slightly longer running time compared with the DCA method, despite the fact that the objective value is still larger than the ALSO-X+. Finally, we can see that PenDC-L outperforms all the other methods in terms of the objective value since it returns the lowest objective, even compared with ALSO-X+. Moreover, its running time remains comparable to the DCA due to an efficient warm-start strategy.

5.3 Linear Objective with Joint Quadratic Chance Constraint

We further test our algorithm on a problem with a linear objective and a joint nonlinear (convex) chance constraint, which has been used as a benchmark in [27, 50]. The decision variable is $\mathbf{x} \in \mathbb{R}_+^d$ and the objective is to minimize the negative sum of components. The feasibility requirement is imposed through a joint quadratic chance constraint of the form

$$\min_{\mathbf{x} \in \mathbb{R}_+^d} \left\{ - \sum_{i=1}^d x_i : \Pr \left\{ \sum_{i=1}^d \xi_{ij}^2 x_i^2 \leq \theta, \forall j \in [m] \right\} \geq 1 - \alpha \right\},$$

where the random coefficients $\{\xi_{ij}\}_{i \in [d], j \in [m]}$ follow a dependent Gaussian structure.

In particular, for each j , the vector $(\xi_{1j}, \dots, \xi_{dj})$ is multivariate normal with $\mathbb{E}[\xi_{ij}] = j/d$ and $\text{Var}(\xi_{ij}) = 1$, and the within- j correlations satisfy $\text{Cov}(\xi_{ij}, \xi_{i'j}) = 0.5$ for $i \neq i'$. Across different indices $j \neq j'$, the corresponding random variables are independent, i.e., $\text{Cov}(\xi_{ij}, \xi_{i'j'}) = 0$. In the numerical study, we fix $d = 20$, $m = 20$, and $\theta = 100$, and we vary the sample size S . Specifically, we consider $S \in \{500, 1000, 2000\}$. For the PenDC-P method, we let $\sigma^0 = 4 \cdot 10^{-3}$, $\beta = 15$ and $\rho = 0$. For the PenDC-L method, we let $\sigma^0 = 8 \cdot 10^{-5}$, $\beta = 10$ and $\rho = 10^{-3}$. For the first two outer iterations of the two penalty methods, we terminate the inner loop after 1 and 2 iterations, respectively, as a warm start. We also set a time limit of 600 seconds for the DCA method.

From Table 3, we can see that both the DCA and the PenDC-P methods exhibit long running times due to the severe oscillation in the updates. Moreover, the PenDC-P method fails to provide a feasible solution in most of the higher-dimensional instances, and the DCA sometimes violates the empirical chance constraint, which reflects their limitations in such settings. In contrast, the PenDC-L maintains stability in runtimes while achieving the lowest objective values in most settings among the iterative methods. Compared with the ALSO-X+, the PenDC-L method returns a lower objective value with considerably shorter running time, highlighting its scalability.

Table 3: Comparisons of the norm optimization problem.

(α, S)		MIP	CVaR	PenDC-L	PenDC-P	DCA	ALSO-X+
$\left(\begin{array}{c} 0.05 \\ 500 \end{array} \right)$	fval	-16.5403	-14.9226	-16.5263	-15.4856	/	-16.5106
	time(gap)	545.4053 (2.2%)	8.6787	43.0250	607.0177	/	220.0617
	prob	0.9500	0.9792	0.9500	0.9500	/	0.9500
	solved	5/5	5/5	5/5	5/5	4/5	5/5
$\left(\begin{array}{c} 0.05 \\ 1000 \end{array} \right)$	fval	-16.3881	-15.0186	-16.3725	/	/	-16.3646
	time(gap)	600.1338 (24.6%)	20.4617	93.8265	/	/	559.3773
	prob	0.9500	0.9820	0.9500	/	/	0.9500
	solved	5/5	5/5	5/5	0/5	4/5	5/5
$\left(\begin{array}{c} 0.05 \\ 2000 \end{array} \right)$	fval	-16.0655	-14.8409	-16.1215	/	-16.0585	-16.1146
	time(gap)	600.7272 (50.1%)	48.9682	252.3422	/	626.9383	1289.8408
	prob	0.9506	0.9808	0.9502	/	0.9500	0.9501
	solved	5/5	5/5	5/5	0/5	5/5	5/5
$\left(\begin{array}{c} 0.10 \\ 500 \end{array} \right)$	fval	-17.4619	-15.7353	-17.4138	-15.6822	-17.3876	-17.4291
	time(gap)	600.0095 (22.9%)	7.6620	42.2942	621.9502	590.9415	236.4814
	prob	0.9000	0.9620	0.9000	0.8992	0.9000	0.9000
	solved	5/5	5/5	5/5	5/5	5/5	5/5
$\left(\begin{array}{c} 0.10 \\ 1000 \end{array} \right)$	fval	-17.3869	-15.7922	-17.3578	/	-17.2949	-17.3490
	time(gap)	601.3015 (55.3%)	18.5244	94.8007	/	614.2368	611.8454
	prob	0.9000	0.9638	0.9006	/	0.9000	0.9004
	solved	5/5	5/5	5/5	0/5	5/5	5/5
$\left(\begin{array}{c} 0.10 \\ 2000 \end{array} \right)$	fval	-17.1944	-15.6308	-17.1653	/	-17.1227	-17.1643
	time(gap)	608.1971 (63.3%)	50.8092	258.9725	/	634.5459	1453.3697
	prob	0.9001	0.9636	0.9003	/	0.9000	0.9002
	solved	5/5	5/5	5/5	0/5	5/5	5/5

6 Conclusion

In this paper, we developed penalty based DC algorithms for the SAA approximation of chance constrained programs in a convex setting. Using a rank-based DC representation of the empirical chance constraint, we proposed a primal-space method that avoids feasible initialization and a numerically more stable lifted formulation with a finite termination guarantee. We established exact penalty and stationarity guarantees for both formulations under mild constraint qualification assumptions. Numerical studies demonstrated the efficiency of the proposed methods. Future works include extending the method to distributionally robust chance constrained programs (DR-CCPs) under Wasserstein ambiguity sets [52, 14], and chance constrained programs with linear conic inequality constraints [49].

References

- [1] Shabbir Ahmed and Dimitri J Papageorgiou. Probabilistic set covering with correlations. *Operations Research*, 61(2):438–452, 2013.
- [2] Shabbir Ahmed and Alexander Shapiro. Solving chance-constrained stochastic programs via sampling and integer programming. In *State-of-the-art decision-making tools in the information-intensive age*, pages 261–269. Informs, 2008.
- [3] Shabbir Ahmed and Weijun Xie. Relaxations and approximations of chance constraints under finite distributions. *Mathematical Programming*, 170(1):43–65, 2018.
- [4] Xiaodi Bai, Jie Sun, and Xiaojin Zheng. An augmented lagrangian decomposition method for chance-constrained optimization problems. *INFORMS Journal on Computing*, 33(3):1056–1069, 2021.

- [5] Sebastian Banert and Radu Ioan Boț. A general double-proximal gradient algorithm for dc programming. *Mathematical programming*, 178(1):301–326, 2019.
- [6] Claude Berge. *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces and Convexity*. Oliver & Boyd, Edinburgh, 1963. Translated by E. M. Patterson. See Ch. 6, § 3 (Maximum Theorem), p. 116.
- [7] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.
- [8] Jonathan M Borwein. Stability and regular points of inequality systems. *Journal of optimization theory and applications*, 48(1):9–52, 1986.
- [9] Oleg P Burdakov, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016.
- [10] James V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29(4):968–998, 1991.
- [11] Abraham Charnes and William W Cooper. Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations research*, 11(1):18–39, 1963.
- [12] Abraham Charnes, William W Cooper, and Gifford H Symonds. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management science*, 4(3):235–263, 1958.
- [13] Pengyu Chen, Xu Shi, Rujun Jiang, and Jiulin Wang. Penalty-based methods for simple bilevel optimization under hölderian error bounds. *Advances in Neural Information Processing Systems*, 37:140731–140765, 2024.
- [14] Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over wasserstein balls. *Operations Research*, 72(1):410–424, 2024.
- [15] Sin-Shuen Cheung, Anthony Man-Cho So, and Kuncheng Wang. Linear matrix inequalities with stochastically dependent perturbations and applications to chance-constrained semidefinite optimization. *SIAM Journal on Optimization*, 22(4):1394–1430, 2012.
- [16] Jehum Cho and Anthony Papavasiliou. Exact mixed-integer programming approach for chance-constrained multi-area reserve sizing. *IEEE Transactions on Power Systems*, 39(2):3310–3323, 2023.
- [17] Frank H Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.
- [18] Ying Cui and Jong-Shi Pang. *Modern nonconvex nondifferentiable optimization*. SIAM, 2021.
- [19] Yan Deng, Huiwen Jia, Shabbir Ahmed, Jon Lee, and Siqian Shen. Scenario grouping and decomposition algorithms for chance-constrained programs. *INFORMS Journal on Computing*, 33(2):757–773, 2021.

- [20] Thai Dinh, Ricardo Fukasawa, and James Luedtke. Exact algorithms for the chance-constrained vehicle routing problem. *Mathematical Programming*, 172(1):105–138, 2018.
- [21] Haw-ren Fang, Sven Leyffer, and Todd Munson. A pivoting algorithm for linear programming with linear complementarity constraints. *Optimization Methods and Software*, 27(1):89–114, 2012.
- [22] Shubhechyya Ghosal and Wolfram Wiesemann. The distributionally robust chance-constrained vehicle routing problem. *Operations Research*, 68(3):716–732, 2020.
- [23] Jun-ya Gotoh, Akiko Takeda, and Katsuya Tono. Dc formulations and algorithms for sparse optimization problems. *Mathematical Programming*, 169(1):141–176, 2018.
- [24] S-P Han and Olvi L Mangasarian. Exact penalty functions in nonlinear programming. *Mathematical programming*, 17(1):251–269, 1979.
- [25] Alan J Hoffman. On approximate solutions of systems of linear inequalities. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 174–176. World Scientific, 2003.
- [26] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1):257–288, 2013.
- [27] L Jeff Hong, Yi Yang, and Liwei Zhang. Sequential convex approximations to joint chance constrained programs: A monte carlo approach. *Operations Research*, 59(3):617–630, 2011.
- [28] Reiner Horst and Nguyen V Thoai. Dc programming: overview. *Journal of Optimization Theory and Applications*, 103(1):1–43, 1999.
- [29] Francisco Jara-Moroni, Jong-Shi Pang, and Andreas Wächter. A study of the difference-of-convex approach for solving linear programs with complementarity constraints. *Mathematical Programming*, 169(1):221–254, 2018.
- [30] Nan Jiang and Weijun Xie. Also- x and also- $x+$: Better convex approximations for chance constrained programs. *Operations Research*, 70(6):3581–3600, 2022.
- [31] Nan Jiang and Weijun Xie. Also- $x\#$: Better convex approximations for distributionally robust chance constrained programs. *Mathematical Programming*, 213:575–638, 2025.
- [32] Simge Kucukyavuz and Ruiwei Jiang. Chance-constrained optimization under limited distributional information: A review of reformulations based on sampling and distributional robustness. *EURO Journal on Computational Optimization*, 10:100030, 2022.
- [33] Hoai An Le Thi, Tao Pham Dinh, and Huynh Van Ngai. Exact penalty and error bounds in dc programming. *Journal of Global Optimization*, 52(3):509–535, 2012.
- [34] Sven Leyffer. Mathematical programs with complementarity constraints. *SIAG/OPT Views-and-News*, 14(1):15–18, 2003.
- [35] Thomas Lipp and Stephen Boyd. Variations and extension of the convex–concave procedure. *Optimization and Engineering*, 17(2):263–287, 2016.

- [36] Zhaosong Lu. Sequential convex programming methods for a class of structured nonlinear programming. *arXiv preprint arXiv:1210.3039*, 2012.
- [37] James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.
- [38] James Luedtke, Shabbir Ahmed, and George L Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming*, 122(2):247–272, 2010.
- [39] Zhi-Quan Luo, Jong-Shi Pang, Daniel Ralph, and Shi-Quan Wu. Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. *Mathematical Programming*, 75(1):19–76, 1996.
- [40] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.
- [41] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.
- [42] Álvaro Porras, Line Roald, Juan Miguel Morales, and Salvador Pineda. Unifying chance-constrained and robust optimal power flow for resilient network operations. *IEEE Transactions on Control of Network Systems*, 12(1):1052–1061, 2025.
- [43] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.
- [44] Andrzej Ruszczyński. Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming*, 93(2):195–215, 2002.
- [45] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.
- [46] Qianhao Sun, Yao Zhang, Hanting Zhao, Wei Huo, Jian Liao, and Jianxue Wang. Network-side carbon emission reduction via dispatching power electronic devices in ac-dc hybrid distribution systems. *IEEE Transactions on Automation Science and Engineering*, 2025.
- [47] Pham Dinh Tao and LT Hoai An. Convex analysis approach to dc programming: theory, algorithms and applications. *Acta mathematica vietnamica*, 22(1):289–355, 1997.
- [48] John F Toland. Duality in nonconvex optimization. *Journal of Mathematical Analysis and Applications*, 66(2):399–415, 1978.
- [49] Wim van Ackooij, Pedro Pérez-Aros, Claudia Soto, and Emilio Vilches. Inner moreau envelope of nonsmooth conic chance-constrained optimization problems. *Mathematics of Operations Research*, 49(3):1419–1451, 2024.
- [50] Peng Wang, Rujun Jiang, Qingyuan Kong, and Laura Balzano. A proximal difference-of-convex algorithm for sample average approximation of chance constrained programming. *INFORMS Journal on Computing*, 38(1):315–339, 2026.

- [51] Yilin Wen, Yi Guo, Zechun Hu, and Gabriela Hug. Stochastic modeling for the aggregated flexibility of distributed energy resources. *Electric Power Systems Research*, 234:110628, 2024.
- [52] Weijun Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, 186(1):115–155, 2021.
- [53] Weijun Xie and Shabbir Ahmed. Bicriteria approximation of chance-constrained covering problems. *Operations Research*, 68(2):516–533, 2020.
- [54] Liang Xu, Chao Zhang, Zhou Xu, and Daniel Zhuoyu Long. A nonparametric robust optimization approach for chance-constrained knapsack problem. *SIAM Journal on Optimization*, 35(2):739–766, 2025.
- [55] Shenglong Zhou, Lili Pan, Naihua Xiu, and Geoffrey Ye Li. A 0/1 constrained optimization solving sample average approximation for chance constrained programming. *Mathematics of Operations Research*, 50(4):2688–2716, 2025.

A Omitted Proofs in Section 1.3

Proof of Proposition 1.1

Proof. We start by recalling the standard representation of the Clarke directional derivative (see, e.g., [17, Proposition 2.1.2]):

$$g^\circ(\bar{\mathbf{x}}; \mathbf{d}) = \max_{\mathbf{v} \in \partial g(\bar{\mathbf{x}})} \langle \mathbf{v}, \mathbf{d} \rangle. \quad (31)$$

Step 1: GMFCQ $\implies 0 \notin \partial g(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$. Assume GMFCQ holds at $\bar{\mathbf{x}}$, i.e., there exists $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}})$ such that $g^\circ(\bar{\mathbf{x}}; \mathbf{d}) < 0$. By (31), this implies $\langle \mathbf{v}, \mathbf{d} \rangle < 0$ for all $\mathbf{v} \in \partial g(\bar{\mathbf{x}})$. If $0 \in \partial g(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$, then there exist $\mathbf{v} \in \partial g(\bar{\mathbf{x}})$ and $\mathbf{z} \in \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$ with $\mathbf{v} + \mathbf{z} = 0$. Since $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}(\bar{\mathbf{x}})$ and $\mathbf{z} \in \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$, we have $\langle \mathbf{z}, \mathbf{d} \rangle \leq 0$ and hence $\langle \mathbf{v}, \mathbf{d} \rangle = -\langle \mathbf{z}, \mathbf{d} \rangle \geq 0$, a contradiction. Therefore $0 \notin \partial g(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})$.

Step 2: $0 \notin \partial g(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}}) \implies$ metric regularity. Define

$$\ker([\partial g(\bar{\mathbf{x}}), I_d]) := \{(\lambda, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}^d : 0 \in \lambda \partial g(\bar{\mathbf{x}}) + \mathbf{z}\}.$$

Since $g(\bar{\mathbf{x}}) = 0$, we have $\mathcal{N}_{\mathbb{R}_-}(g(\bar{\mathbf{x}})) = \mathcal{N}_{\mathbb{R}_-}(0) = \mathbb{R}_+$. A direct verification shows that

$$0 \notin \partial g(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}}) \iff \ker([\partial g(\bar{\mathbf{x}}), I_d]) \cap (\mathcal{N}_{\mathbb{R}_-}(g(\bar{\mathbf{x}})) \times \mathcal{N}_{\mathcal{X}}(\bar{\mathbf{x}})) = \{(0, \mathbf{0})\},$$

which is the regularity condition [10, Eq. (2.8)] (see also [8, Eq. (33)]). Applying [10, Thm. 2.4] (see also [8, Thm. 3.2]) to the set-valued mapping $\mathcal{G}(\mathbf{x}) = g(\mathbf{x}) + \mathbb{R}_+$, we conclude that $\mathcal{F} = \mathcal{G}^{-1}(0)$ is metrically regular at $\bar{\mathbf{x}}$.

Finally, metric regularity at $(\bar{\mathbf{x}}, 0)$ yields constants $\kappa > 0$ and $\varepsilon > 0$ such that

$$\text{dist}(\mathbf{x}, \mathcal{G}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{G}(\mathbf{x})) = \kappa [g(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, \varepsilon),$$

which is exactly (5). □

Proof of Theorem 1.2

Proof. (i) Let $\sigma > 0$ and assume that $\bar{\mathbf{x}}$ is a local minimizer of (6) with $g(\bar{\mathbf{x}}) \leq 0$. Then $[g(\bar{\mathbf{x}})]_+ = 0$. Take $r > 0$ such that

$$f(\bar{\mathbf{x}}) + \sigma[g(\bar{\mathbf{x}})]_+ \leq f(\mathbf{x}) + \sigma[g(\mathbf{x})]_+, \quad \forall \mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, r).$$

For any $\mathbf{x} \in \mathcal{F} \cap B(\bar{\mathbf{x}}, r)$, we have $[g(\mathbf{x})]_+ = 0$ and thus $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$. Hence $\bar{\mathbf{x}}$ is a local minimizer of (3).

(ii) Assume that $\bar{\mathbf{x}}$ is a local minimizer of (3). Let κ, ε be the constants in the error bound (5). Fix any $\sigma > L_f \kappa$. We show that $\bar{\mathbf{x}}$ locally minimizes (6).

Take any $\mathbf{x} \in \mathcal{X} \cap B(\bar{\mathbf{x}}, \varepsilon)$ and choose $\tilde{\mathbf{x}} \in \mathcal{F}$ such that $\|\mathbf{x} - \tilde{\mathbf{x}}\| = \text{dist}(\mathbf{x}, \mathcal{F})$. By Lipschitzness of f and the error bound,

$$f(\tilde{\mathbf{x}}) \leq f(\mathbf{x}) + L_f \|\tilde{\mathbf{x}} - \mathbf{x}\| = f(\mathbf{x}) + L_f \text{dist}(\mathbf{x}, \mathcal{F}) \leq f(\mathbf{x}) + L_f \kappa [g(\mathbf{x})]_+.$$

Moreover, since $\bar{\mathbf{x}}$ is a local minimizer of (3), we have $f(\bar{\mathbf{x}}) \leq f(\tilde{\mathbf{x}})$ for all \mathbf{x} sufficiently close to $\bar{\mathbf{x}}$. Therefore, for such \mathbf{x} ,

$$f(\bar{\mathbf{x}}) \leq f(\mathbf{x}) + L_f \kappa [g(\mathbf{x})]_+ \leq f(\mathbf{x}) + \sigma [g(\mathbf{x})]_+.$$

Since $[g(\bar{\mathbf{x}})]_+ = 0$, this shows that $\bar{\mathbf{x}}$ is a local minimizer of (6). Taking $\bar{\sigma} := L_f \kappa$ completes the proof. \square