SIAM J. OPTIM.
Vol. 34, No. 1, pp. 306–335

A DECOMPOSITION ALGORITHM FOR TWO-STAGE STOCHASTIC PROGRAMS WITH NONCONVEX RECOURSE FUNCTIONS*

HANYANG LI† AND YING CUI†

Abstract. In this paper, we have studied a decomposition method for solving a class of nonconvex two-stage stochastic programs, where both the objective and constraints of the second-stage problem are nonlinearly parameterized by the first-stage variables. Due to the failure of the Clarke regularity of the resulting nonconvex recourse function, classical decomposition approaches such as Benders decomposition and (augmented) Lagrangian-based algorithms cannot be directly generalized to solve such models. By exploring an implicitly convex-concave structure of the recourse function, we introduce a novel decomposition framework based on the so-called partial Moreau envelope. The algorithm successively generates strongly convex quadratic approximations of the recourse function based on the solutions of the second-stage convex subproblems and adds them to the first-stage master problem. Convergence has been established for both a fixed number of scenarios and a sequential internal sampling strategy. Numerical experiments are conducted to demonstrate the effectiveness of the proposed algorithm.

Key words. two-stage stochastic program, nonconvex recourse, decomposition

MSC codes. 90C15, 90C26

DOI. 10.1137/22M1488533

1. Introduction.

Stochastic programming (SP) is a mathematical framework to model decision making in the presence of uncertainty [4, 54]. Two-stage SPs constitute a special class of this paradigm where partial decisions have to be made before the observation of the entire information, while the rest of the decisions are determined after the full information is revealed. Most existing computational studies of continuous two-stage SPs are devoted to convex problems, especially linear problems [51, 4, 28, 54].

However, there are many emerging applications in operations research that call for complex nonlinear two-stage SP models and computational methods. Let us first introduce the mathematical formulation of such problems before discussing the applications. The central optimization problem under consideration in this paper takes the following form:

$$\underset{x \in X}{\text{minimize}}\ \zeta(x) \triangleq \varphi(x) + \mathbb{E}_{\tilde{\xi}}\left[\psi(x;\tilde{\xi})\right], \tag{1.1}$$

where $\psi(x;\xi)$ is the second-stage recourse function that is given by

$$\psi(x;\xi) \triangleq \begin{cases} \underset{y}{\text{minimum}}\,[\, f(x,y;\xi) \text{ subject to } G(x,y;\xi) \le 0\,] & \text{if } x \in \overline{X}, \\ +\infty & \text{if } x \notin \overline{X}. \end{cases} \tag{1.2}$$

In the above formulation, X and $\overline{X}$ are nonempty convex compact subsets in $\mathbb{R}^{n_1}$ with $X \subseteq \operatorname{int}(\overline{X})$, $\varphi : \mathbb{R}^{n_1} \to \mathbb{R}$ is a deterministic convex function that only depends on the

*Received by the editors April 4, 2022; accepted for publication (in revised form) June 27, 2023; published electronically January 19, 2024.
https://doi.org/10.1137/22M1488533
Funding: The authors are partially supported by the National Science Foundation under grants CCF-2153352 and DMS-2309729.
†Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA 94720 USA (hanyang_li@berkeley.edu, yingcui@berkeley.edu).

first-stage decision x; $\tilde{\xi}:\Omega\to\Xi$ is a random vector on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with $\Xi\subseteq\mathbb{R}^m$ being a measurable closed set; $\xi=\tilde{\xi}(\omega)$ for some $\omega\in\Omega$ represents a realization of the random vector $\tilde{\xi}$; and $f:\mathbb{R}^{n_1+n_2}\times\Xi\to\mathbb{R}$ and $G\triangleq(g_1,\ldots,g_\ell)^\top:\mathbb{R}^{n_1+n_2}\times\Xi\to\mathbb{R}^\ell$ are two Carathéodory functions (i.e., $f(\bullet,\bullet;\xi)$ and $G(\bullet,\bullet;\xi)$ are continuous for almost any $\xi\in\Xi$; $f(x,y;\bullet)$ and $G(x,y;\bullet)$ are measurable for any $(x,y)\in\mathbb{R}^{n_1+n_2}$) that are jointly determined by the first-stage variable x and the second-stage variable y. We assume that, for almost any $\xi\in\Xi$, the function $f(\bullet,\bullet;\xi)$ is concave-convex (i.e., $f(\bullet,y;\xi)$ is concave for $y\in\mathbb{R}^{n_2}$ and $f(x,\bullet;\xi)$ is convex for $x\in\mathbb{R}^{n_1}$), $g_j(\bullet,\bullet;\xi)$ is jointly convex for each $j=1,\ldots,\ell$, and $\mathbb{E}_{\tilde{\xi}}[\psi(x;\tilde{\xi})]$ is well-defined.

An example of concave-convex $f(\bullet,\bullet;\xi)$ is a bilinear function $x^\top D(\xi)y$ for some random matrix $D(\xi)\in\mathbb{R}^{n_1\times n_2}$. The above settings notably extend the classical paradigm for continuous two-stage SPs [54, Chapter 2.3] in the following directions:

(i) The first-stage variable x appears not only in the constraints of the second-stage problem, but also in the objective f. The recourse function $\psi(\bullet;\xi)$ is nonconvex since $f(\bullet,\bullet;\xi)$ is not jointly convex. This is fundamentally different from the recent papers [25, 24] that have assumed the joint convexity of $f(\bullet,\bullet;\xi)$.

(ii) Both the objective function f and the constraint map G can be nonsmooth.

These two features together lead to a complex nonconvex and nonsmooth recourse function $\psi(\bullet;\xi)$ (see Figure 1), which constitutes the major challenge for designing rigorous and efficient numerical methods to solve problem (1.1).

Recourse functions in the form of (1.2) arise from many applications. One important source of the nonconvex recourse function in (1.2) comes from the decision-dependent/influenced uncertainty [31, 20, 27, 37], where the probability distribution of the random vector $\tilde{\xi}$ is dependent on the first-stage variable x. This is in contrast to the classical SP paradigm under exogenous uncertainty, where the distribution of $\tilde{\xi}$ is not affected by the first-stage decisions. There is growing interest in the endogenous uncertainty in the recent literature on stochastic and robust programs [21, 22, 56, 27]. A typical example where the random parameters can be altered by a decision is that the price (as a first-stage variable) may affect the distribution of the product demand. Assume that the probability distribution of $\tilde{\xi}$ is given by $\mathbb{P}_x$ that depends on the first-stage variable x. We consider the corresponding two-stage SP model:

$$\underset{x\in X}{\text{minimize}}\ \left\{\varphi(x)+\mathbb{E}_{\tilde{\xi}\sim\mathbb{P}_x}[\psi(x;\tilde{\xi})]=\varphi(x)+\int_\Xi\psi(x;\tilde{\xi})\mathrm{d}\mathbb{P}_x(\tilde{\xi})\right\}.$$

If there exists a decision-independent distribution $\widehat{\mathbb{P}}$ such that $\mathbb{P}_x$ is absolutely continuous with respect to $\widehat{\mathbb{P}}$ for any $x\in X$, we can apply [18, Proposition 3.9] to reformulate the above problem into the following one:

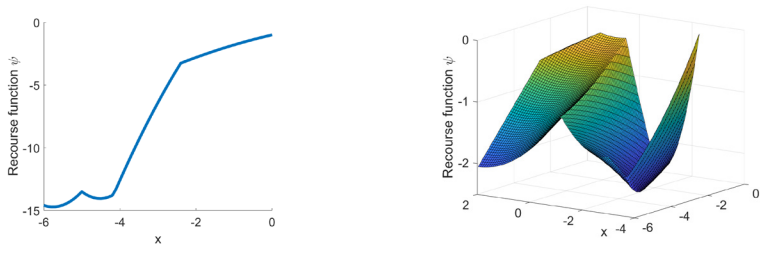

FIG. 1. *The nonconvex nonsmooth recourse functions for fixed ξ's. Left: $x\in\mathbb{R}$. Right: $x\in\mathbb{R}^2$.*

$$\underset{x\in X}{\text{minimize}} \quad \varphi(x)+\mathbb{E}_{\tilde{\xi}\sim\widehat{\mathbb{P}}}[\widehat{\psi}(x;\tilde{\xi})] \text{ with } \widehat{\psi}(x;\xi)\triangleq\psi(x;\xi)\frac{\mathrm{d}\mathbb{P}_x(\xi)}{\mathrm{d}\widehat{\mathbb{P}}(\xi)}, \tag{1.3}$$

where $\mathrm{d}\mathbb{P}_x(\xi)/\mathrm{d}\widehat{\mathbb{P}}(\xi)$ is the Radon–Nikodym derivative of $\mathbb{P}_x$ with respect to $\widehat{\mathbb{P}}$. Even if originally the first-stage decision x only appears in the constraints of the second-stage problem in $\psi(x;\xi)$, the above transformation would make x also appear in the second-stage objective function through the multiplication of $\mathrm{d}\mathbb{P}_x(\xi)/\mathrm{d}\widehat{\mathbb{P}}(\xi)$.

A specific example of the decision-dependent uncertainty in SPs is a power system planning problem originated in [33] and expanded in [27]. Assume that $\tilde{\xi}=(\{\tilde{d}_j\}_{j\in\mathcal{J}},\{\tilde{\pi}_j\}_{j\in\mathcal{J}},\{\tilde{q}_i\}_{i\in\mathcal{I}})$ follows a discrete distribution with the support $\{\xi^s\}_{s=1}^S$, where $\tilde{d}_j$ and $\tilde{\pi}_j$ represent the demand and the price of electricity in the location $j\in\mathcal{J}$, and $\tilde{q}_i$ is the unit production cost of the power plant $i\in\mathcal{I}$. The distribution of $\tilde{\xi}$ is a convex combination of $|\mathcal{G}|$ given discrete distributions (each with probability p_{sg} for the scenario ξ^s), whose weights $\{x_g\}_{g\in\mathcal{G}}$ are parts of the first-stage decisions, i.e., $\mathbb{P}_x(\tilde{\xi}=\xi^s)=\sum_{g\in\mathcal{G}}p_{sg}\,x_g$ for each s. The capacity of each power plant $\{x_i\}_{i\in\mathcal{I}}$ also needs to be determined in the first stage. The second-stage decisions are the production y_{ij} from the power plant i to the location j for each s ($y=(y_{ij})_{i\in\mathcal{I},j\in\mathcal{J}}$ is bounded between ℓ_y and u_y). By letting $\widehat{\mathbb{P}}(\tilde{\xi}=\xi^s)=1/S$ for each s in (1.3), we can rewrite the recourse function as

$$\begin{aligned}\widehat{\psi}(\{x_i\}_{i\in\mathcal{I}},\{x_g\}_{g\in\mathcal{G}};\xi^s)=\underset{\ell_y\le y\le u_y}{\text{minimum}}\ & S\sum_{g\in\mathcal{G}}p_{sg}\,x_g\sum_{i\in\mathcal{I},\,j\in\mathcal{J}}(q_{is}-\pi_{js})\,y_{ij}\\ \text{subject to}\ & \sum_{j\in\mathcal{J}}y_{ij}\le x_i,\ i\in\mathcal{I};\ \sum_{i\in\mathcal{I}}y_{ij}=d_{js},\ j\in\mathcal{J},\end{aligned} \tag{1.4}$$

and obtain a decision-independent SP with the recourse function $\widehat{\psi}$. Observe that both the objective and constraints depend on the first-stage variables. In particular, the objective function is convex in x_g and concave in y, which fits our problem setting. Later, we will apply our proposed algorithms to solve a two-stage SP with the above recourse function in section 6.

The second example of the nonconvex recourse function in (1.2) is the stochastic interdiction problem [14, 26], where the defender may want to maximize the second-stage objective function instead of minimizing it. Even for the simple linear second-stage problem with only x appearing in the constraints, the recourse function

$$\begin{aligned}\widetilde{\psi}(x;\xi)\triangleq\underset{y}{\text{maximum}}\quad & c(\xi)^\top y\\ \text{subject to}\quad & T(\xi)x+W(\xi)y=h(\xi)\end{aligned}$$

is not convex in x. One may take the dual of the second-stage maximization problem so that the recourse function is a parametrized minimization problem

$$\begin{aligned}\widetilde{\psi}(x;\xi)=\underset{\lambda}{\text{minimum}}\quad & \lambda^\top T(\xi)x-\lambda^\top h(\xi)\\ \text{subject to}\quad & W(\xi)^\top\lambda+c(\xi)=0.\end{aligned}$$

However, this dualization would bring a bilinear term $\lambda^\top T(\xi)x$ of the first-stage variable x and the second-stage variable λ to the objective function that necessities the concave-convex structure of $f(\bullet,\bullet;\xi)$. To the best of our knowledge, there is no known rigorous decomposition method to solve a general nonconvex two-stage min-max stochastic program even when the second stage is a linear maximization program biparametrized by the first-stage variable in both the objective and constraints.

When the distribution of $\tilde{\xi}$ is taken as the empirical distribution of observed realizations $\xi^1, \ldots, \xi^S$, the simplest way to tackle the problem (1.1) is to simultaneously solve the first-stage variable x and second-stage variables $y^1, \ldots, y^S$ (each y^s is attached to one scenario ξ^s) via the sample average approximation [54]:

$$\begin{array}{ll} \underset{x \in X,\, y^1, \ldots, y^S}{\text{minimize}} & \varphi(x) + \dfrac{1}{S} \displaystyle\sum_{s=1}^{S} f(x, y^s; \xi^s) \\ \text{subject to} & G(x, y^s; \xi^s) \leq 0, \quad s = 1, \ldots, S. \end{array} \tag{1.5}$$

However, this approach can be prohibitive when the number of scenarios S is large since the dimension of the unknown variables is $n_1 + n_2 S$. Even if S is small or moderate, the above formulation may still be difficult to handle under our setting as the function $f(\bullet, \bullet; \xi)$ is not jointly convex (for example, when $f(\bullet, \bullet; \xi)$ is bilinear). It is also challenging to apply stochastic approximation methods [43, 40, 36, 19, 52, 16, 5] to solve (1.1), since it is not clear how to compute a (Clarke) subdifferential of the nonconvex recourse function $\psi(\bullet; \xi)$. Without strong assumptions like the uniqueness of the second-stage solutions, only a superset of the subdifferential $\partial\psi(\bullet; \xi)$ at given x is computable [7, Chapter 4]. When f and G are twice continuously differentiable, the authors in [8] have adopted a smoothing method to deal with the possibly nonconvex recourse function by adding the Tikhonov-regularized barrier of the inequality constraints to the second-stage objective function. For a special class of two-stage nonconvex quadratic SPs under the simplex constraint, the paper [6] has derived upper and lower approximations of the objective values via copositive programs.

Notice that the constraints in (1.5) are in fact blockwise separable in $y^1, \ldots, y^S$ so that there is a block-angular structure between the first- and second-stage variables. Decomposition algorithms of two-stage SPs take advantage of this special structure to efficiently handle a large number of scenarios via solving S numbers of low-dimensional subproblems [51]. Two classical decomposition algorithms for two-stage SPs are (augmented) Lagrangian decomposition and Benders decomposition. *(Augmented) Lagrangian decompositions* (including the progressive hedging algorithm) copy the first-stage variable S times and attach one to each scenario [23, 48]. In order to force the nonanticipativity of the first-stage decision, one has to add equality constraints among all copies to ensure that x is the same across different realizations of the uncertainty. However, there are two major bottlenecks to applying this kind of dual-based algorithm to solve the problem (1.1). One, each subproblem pertaining to one pair of variables (x^s, y^s) is still nonconvex if f is not a jointly convex function, so that it is in general not easy to obtain its global optimal solution. Two, the convergence of these dual approaches is largely restricted to the convex problems or special integer problems [48, 10]. Although there are some recent advances for the convergence study of the progressive hedging algorithm for solving nonconvex SPs under the local convexity conditions [46, 47], it is not clear whether the problem (1.1) satisfies those conditions without further assumptions on f and g. *Benders decomposition* (or L-shaped methods) [3, 57, 59] alternatively updates the first-stage and second-stage variables, where the second-stage subproblem can be solved in parallel to save the computational time and reduce the storage burden. In order to derive valid inequalities of x and add them to the first-stage master problem, one usually uses subgradient inequalities of the (convex) recourse function to generate a sequence of lower approximations. However, when the recourse function is associated with the complex nonconvex function in (1.1), it is challenging to derive its lower bounds based on the computed second-stage solutions. In fact, for the recourse functions in Figure 1, there seems not to

exist a convex function that touches one of the downward cusps and at the same time approximates the original function from below.

In this paper, we tackle the two-stage stochastic program (1.1) by a novel lifting technique that transforms the complex nonconvex and nondifferentiable recourse function (1.2) in the original space to a structured convex-concave function in a lifted space. The reveal of this latent structure enables us to construct convex surrogations of the recourse function at the latest first-stage iterate, whose evaluations are decomposable across different scenarios. Such surrogate functions are then added to the master problem to generate the next first-stage iterate. We shall prove that repeating the above procedure, the sequence of the first-stage iterates converges to a properly defined stationary solution of (1.1). In order to further reduce the computational cost per step when the number of scenarios S is large as well as to handle the case where $\tilde{\xi}$ is continuously distributed, we also propose a framework that incorporates sequential sampling into the surrogation algorithm. The sequential sampling method gradually adds scenarios and generates cuts along the iterations, which has the advantage that one may obtain satisfactory descent progress in the early iterations with relatively small sample sizes to accelerate the overall procedure.

The paper is organized as follows. Section 2 introduces notation and provides preliminary knowledge. In section 3, we discuss the implicitly convex-concave structure of the recourse function and derive its computationally tractable approximations. A decomposition algorithm for solving problem (1.1) with a fixed number of scenarios is proposed and analyzed in section 4. To further handle the continuously distributed random vectors as well as to reduce the computational cost of the decomposition algorithm in its early stage, we provide an internal sampling version of the algorithm in section 5 and show the almost sure convergence of the iterative sequence. In section 6, we conduct extensive numerical experiments to show the effectiveness of our proposed frameworks. The paper ends with a concluding section.

2. Preliminaries. We first summarize the notation used throughout the paper. We write $\mathbb{Z}_+$ as the set of all nonnegative integers and $\mathbb{R}^n$ as the n-dimensional Euclidean space equipped with the inner product $\langle x, y\rangle = x^\top y$ and the induced norm $\|x\| \triangleq \sqrt{x^\top x}$. The symbol $\mathbb{B}(x, \delta)$ is used to denote the closed ball of radius $\delta > 0$ centered at a vector $x \in \mathbb{R}^n$. Let A and C be two nonempty subsets of $\mathbb{R}^n$. The diameter of A is defined as $R(A) \triangleq \sup_{x,y\in A}\|x - y\|$, and the distance from a vector $x \in \mathbb{R}^n$ to A is defined as $\operatorname{dist}(x, A) \triangleq \inf_{y\in A}\|y - x\|$. The one-sided deviation of A from C is defined as $\mathbb{D}(A, C) \triangleq \sup_{x\in A} \operatorname{dist}(x, C)$.

We next introduce the concepts of generalized derivatives and subdifferentials for nonsmooth functions. Interested readers are referred to the monographs [12, 49, 35] for thorough discussions on these subjects. Consider a function $f : \mathcal{O} \to \mathbb{R}$ defined on an open set $\mathcal{O} \subseteq \mathbb{R}^n$. The classical one-sided directional derivative and the Clarke directional derivative of f at $\bar{x} \in \mathcal{O}$ along the direction $d \in \mathbb{R}^n$ are defined as

$$f'(\bar{x}; d) \triangleq \lim_{t\downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} \quad \text{and} \quad f^\circ(\bar{x}; d) \triangleq \limsup_{x\to\bar{x},\, t\downarrow 0} \frac{f(x + td) - f(x)}{t}$$

if these two limits exist. f is said to be directionally differentiable at $\bar{x} \in \mathcal{O}$ if it is directionally differentiable along any direction $d \in \mathbb{R}^n$. The Clarke directional derivative $f^\circ(\bar{x}; d)$ is finite for any direction d when f is locally Lipschitz continuous at $\bar{x}$. The Clarke subdifferential of f at $\bar{x}$ is the set $\partial_C f(\bar{x}) \triangleq \{v \in \mathbb{R}^n \mid f^\circ(\bar{x}; d) \geq v^\top d \text{ for all } d \in \mathbb{R}^n\}$, which coincides with the usual subdifferential in convex analysis for a convex function. If f is strictly differentiable at $\bar{x}$, then $\partial_C f(\bar{x}) = \{\nabla f(\bar{x})\}$.

We say that f is Clarke regular at $\bar{x} \in \mathcal{O}$ if f is directionally differentiable at $\bar{x}$ and $f^\circ(\bar{x}; d) = f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$. This Clarke regularity at $\bar{x}$ is equivalent to having $f(x) \geq f(\bar{x}) + \bar{v}^\top (x - \bar{x}) + o(\|x - \bar{x}\|)$ for any $\bar{v} \in \partial_C f(\bar{x})$. Therefore, if a function fails to satisfy the Clarke regularity at $\bar{x}$ (for example, at the downward cusp in the left panel of Figure 1), there does not exist an approximate linear lower bound of the original function based on the Clarke subdifferentials with small o error locally.

Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function that is directionally differentiable. We say $\bar{x} \in X$ is a directional stationary point of f on X if $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$, and a Clarke stationary point if $f^\circ(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$; the latter is equivalent to $0 \in \partial_C f(\bar{x}) + \mathcal{N}_X(\bar{x})$ with $\mathcal{N}_X(\bar{x})$ being the normal cone of X.

Let $\mathcal{F} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. Its outer limit at $x \in \mathbb{R}^n$ is defined as

$$\limsup_{x \to \bar{x}} \mathcal{F}(x) \triangleq \bigcup_{x^\nu \to \bar{x}} \limsup_{\nu \to \infty} \mathcal{F}(x^\nu) = \big\{ u \mid \exists\, x^\nu \to \bar{x}, \exists\, u^\nu \to u \text{ with } u^\nu \in \mathcal{F}(x^\nu) \big\}.$$

We say $\mathcal{F}$ is outer semicontinuous (osc) at $\bar{x} \in \mathbb{R}^n$ if $\limsup_{x \to \bar{x}} \mathcal{F}(x) \subseteq \mathcal{F}(\bar{x})$.

3. The implicit convexity-concavity of the recourse functions. A key ingredient in designing a decomposition method for solving the two-stage SP (1.1) is to derive a computationally friendly approximation of the nonconvex recourse function (1.2) at any given $x \in X$ and $\xi \in \Xi$. This is the main content of the present section.

For simplicity, we omit ξ in (1.2) throughout this section and rewrite the recourse function as, for $x \in \mathbb{R}^{n_1}$,

$$\psi(x) \triangleq \begin{cases} \underset{y}{\text{minimum}} \, [\, f(x,y) \text{ subject to } G(x,y) \leq 0 \,] & \text{if } x \in \overline{X}, \\ +\infty & \text{if } x \notin \overline{X}, \end{cases} \tag{3.1}$$

where $f(\bullet, \bullet)$ is concave-convex and each $g_j(\bullet, \bullet)$ is jointly convex for $j = 1, \ldots, \ell$. We assume that for any $x \in \overline{X}$, the minimization problem of y in (3.1) has an optimal solution, which implies the finiteness of $\psi(x)$ on $\overline{X}$. In the following, we show that the above function, although generally nonconvex and nondifferentiable in $\mathbb{R}^{n_1}$, has a benign structure in a lifted space. Leveraging this structure, we then derive an approximate difference-of-convex (dc) decomposition of the recourse function that is computationally tractable. Such an approximation is the cornerstone of the decomposition method to be presented in the next two sections.

3.1. The implicit convexity-concavity of ψ. As mentioned in the first section, the difficulty in designing a decomposition method for solving (1.1) is due to the lack of a valid inequality of the recourse function, which is partially because x appears in both the objective and the constraints of the parametric problem in (3.1). However, if x either in the objective or in the constraints is fixed, the resulting functions are relatively easy to analyze. Specifically, for any fixed $\bar{x} \in \overline{X}$, consider the functions

$$\psi_{\text{cvx}}(x) \triangleq \begin{bmatrix} \underset{y}{\text{minimum}} & f(\bar{x}, y) \\ \text{subject to} & G(x,y) \leq 0 \end{bmatrix} \text{ and } \psi_{\text{cve}}(x) \triangleq \begin{bmatrix} \underset{y}{\text{minimum}} & f(x,y) \\ \text{subject to} & G(\bar{x},y) \leq 0 \end{bmatrix}.$$

We can easily derive the following structural properties of ψ_{cvx} and ψ_{cve} by using the fact that the inf-projection of a jointly convex function is convex and the infimum of a family of concave functions is concave.

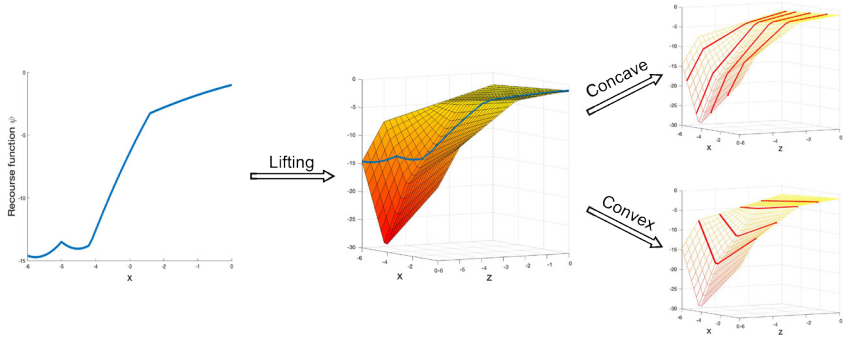

FIG. 2. *An illustration of the icc structure of the nonconvex recourse function. Left: the original* one-*dimensional recourse function that is neither convex nor concave. Middle: the lifted bivariate counterpart in* $\mathbb{R}^2$ *that is convex in* x *and concave in* z. *Right: the exposure of the concave component in the* z *coordinate and the convex component in the* x *coordinate.*

LEMMA 3.1. *Let* $\bar{x} \in \overline{X} \subseteq \mathbb{R}^{n_1}$ *be fixed. Suppose that the minimization problems in defining* $\psi_{\rm cvx}$ *and* $\psi_{\rm cve}$ *both have nonempty solution sets for any* $x \in \overline{X}$. *Then the function* $\psi_{\rm cvx}$ *is convex and* $\psi_{\rm cve}$ *is concave on* $\overline{X}$.

Lemma 3.1 suggests that the recourse function (3.1) has a hidden convex-concave structure. Indeed, such a function ψ belongs to a special class of nonconvex functions coined implicitly convex-concave (icc) functions that are formally defined below (See Figure 2). For an extended-real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, the effective domain of f is defined as $\operatorname{dom} f \triangleq \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$.

DEFINITION 3.2 (see [15, Definition 4.4.4]). *A function* $\theta : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *with* $\operatorname{dom}\theta$ *being a convex set is said to be icc if there exists a function* $\overline{\theta} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ *satisfying*

(a) $\overline{\theta}(x,z) = +\infty$ *if* $x \notin \operatorname{dom}\theta, z \in \mathbb{R}^n$, *and* $\overline{\theta}(x,z) = -\infty$ *if* $x \in \operatorname{dom}\theta, z \notin \operatorname{dom}\theta$;
(b) $\overline{\theta}(\bullet, z)$ *is convex for any fixed* $z \in \operatorname{dom}\theta$;
(c) $\overline{\theta}(x, \bullet)$ *is concave for any fixed* $x \in \operatorname{dom}\theta$;
(d) $\theta(x) = \overline{\theta}(x,x)$ *for any* $x \in \operatorname{dom}\theta$.

The above concept is first introduced in [32] to analyze the convergence property of a dc algorithm to solve two-stage convex biparametric quadratic SPs. More properties of icc functions are studied in the recent monograph [15]. In fact, the term "icc" suggests that this class of functions is a generalization of the dc functions, as the latter is "explicitly convex-concave," i.e., for any dc function $\theta(x) = \theta_1(x) - \theta_2(x)$ with both θ_1 and θ_2 convex, one can always associate it with the bivariate function $\overline{\theta}(x,y) = \theta_1(x) - \theta_2(y)$ to explicitly expose the convexity-concavity of θ in the lifted pair (x,y). Back to the recourse function ψ (3.1), we consider its lifted bivariate counterpart

$$\overline{\psi}(x,z) \triangleq \begin{cases} \underset{y}{\text{minimum}} \{ f(z,y) \mid G(x,y) \le 0 \} & \text{if } x, z \in \overline{X}, \\ +\infty & \text{if } x \notin \overline{X}, \\ -\infty & \text{if } x \in \overline{X} \text{ and } z \notin \overline{X}. \end{cases} \tag{3.2}$$

If the minimization problem of y in (3.2) has a nonempty solution set for any $(x,z) \in \overline{X} \times \overline{X}$, it is not difficult to see that the assumption in Lemma 3.1 holds. Henceforth, the following result is a direct consequence of Lemma 3.1. No proof is needed.

PROPOSITION 3.3. *Assume that for any $(x,z)\in\overline{X}\times\overline{X}$, the minimization problem of y in* (3.2) *has a nonempty solution set. Then ψ in* (3.1) *is an icc function associated with the lifted function $\overline{\psi}$ in* (3.2).

One shall see from the subsequent sections that the derived icc property of the recourse function is useful to study the two-stage SP (1.1). On one hand, we can leverage this structure to construct an approximation of the nonconvex recourse function ψ at any given x; on the other hand, it enables us to define a stationary point of (1.1) that is provably computable by our later designed algorithms. To fulfill these tasks, we first derive a superset of the Clarke subdifferential of ψ. To proceed, we denote $\partial_1\overline{\psi}(x,z)$ as the subdifferential of the convex function $\overline{\psi}(\bullet,z)$ at x for any $z\in\overline{X}$, and $\partial_2(-\overline{\psi})(x,z)$ as the subdifferential of the convex function $(-\overline{\psi})(x,\bullet)$ at z for any $x\in\overline{X}$. We also write $\overline{Y}(x,z)$ as the set of all optimal solutions of problem (3.2).

LEMMA 3.4. *Assume that for any $(x,z)\in\overline{X}\times\overline{X}$, the minimization problem of y in* (3.2) *has a nonempty solution set. Then for all $(x,z)\in\overline{X}\times\overline{X}$:*

$$\{\partial_1(-f)(z,y)\,|\,y\in\overline{Y}(x,z)\}\subseteq\partial_2(-\overline{\psi})(x,z).$$

Proof. For any $(x,z)\in\overline{X}\times\overline{X}$, we take any $y\in\overline{Y}(x,z)$ and any $c\in\partial_1(-f)(z,y)$ to obtain

$$\overline{\psi}(x,z')\le f(z',y)\le f(z,y)+(-c)^\top(z'-z)=\overline{\psi}(x,z)+c^\top(z-z')\quad\forall\,z'\in\mathbb{R}^n,$$

where the first inequality holds because $\overline{\psi}(x,z')=-\infty$ if $z'\notin\overline{X}$ and $y\in\overline{Y}(x,z)$ must be feasible to the constraint $G(x,y)\le 0$ if $z'\in\overline{X}$, and the second inequality is due to the concavity of $f(\bullet,y)$. By applying [45, Theorem 23.5], we have $c\in\partial_2(-\overline{\psi})(x,z)$ and part (a) is proved. ∎

Part (a) of the above lemma can be viewed as a weaker version of Danskin's theorem [42, 41, 11]. Instead of a complete characterization of the subdifferential of an optimal value function in the aforementioned papers, we only need to obtain one element from this subdifferential to design our algorithms later. Therefore, only a one-sided inclusion as in part (a) is needed, which holds without the compactness of the feasible set $\{y\,|\,G(x,y)\le 0\}$ for any given $x\in\overline{X}$.

Next we summarize several results regarding icc functions that will be used in the subsequent analysis.

LEMMA 3.5. *Consider any icc function ψ associated with a lifted function $\overline{\psi}$ that is continuous relative to* $\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$. *The following properties of the set-valued mappings $\partial_1\overline{\psi}$ and $\partial_2(-\overline{\psi})$ hold:*

(a) *$\partial_1\overline{\psi}$ and $\partial_2(-\overline{\psi})$ are osc relative to* $\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$;

(b) *$\partial_1\overline{\psi}$ and $\partial_2(-\overline{\psi})$ are locally bounded on* $\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$, *and $\overline{\psi}$ is locally Lipschitz continuous relative to* $\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$;

(c) *$\partial_C\psi(x)\subseteq\partial_1\overline{\psi}(x,x)-\partial_2(-\overline{\psi})(x,x)$ for any $x\in\mathrm{int}(\mathrm{dom}\,\psi)$.*

Proof. (a) To prove that $\partial_1\overline{\psi}$ is osc relative to $\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$, we first notice that for any fixed $d\in\mathbb{R}^{n_1}$, the directional derivative $\overline{\psi}_1'((\bullet,\bullet);d)$ is upper semicontinuous jointly at $(\bar{x},\bar{z})\in\mathrm{int}(\mathrm{dom}\,\psi)\times\mathrm{int}(\mathrm{dom}\,\psi)$ [15, Proposition 4.4.26(a)]. Since $\overline{\psi}_1'((x,z);\bullet)$ is the support function of the partial subgradient $\partial_1\overline{\psi}(x,z)$, we have

$$\limsup_{(x,z)\to(\bar{x},\bar{z})}\left(\sup_{a\in\partial_1\overline{\psi}(x,z)} a^\top d\right)\le \sup_{a'\in\partial_1\overline{\psi}(\bar{x},\bar{z})}(a')^\top d \qquad \forall\, d\in\mathbb{R}^{n_1},$$

which implies $\limsup_{(x,z)\to(\bar{x},\bar{z})}\partial_1\overline{\psi}(x,z)\subseteq\partial_1\overline{\psi}(\bar{x},\bar{z})$ by [45, Corollary 13.1.1]. The outer semicontinuity of $\partial_2(-\overline{\psi})$ relative to $\operatorname{int}(\operatorname{dom}\psi)\times\operatorname{int}(\operatorname{dom}\psi)$ can be proved similarly.

(b) Suppose for the sake of contradiction that $\partial_1\overline{\psi}$ is not locally bounded at some $(\bar{x},\bar{z})\in\operatorname{int}(\operatorname{dom}\psi)\times\operatorname{int}(\operatorname{dom}\psi)$. Then there exists a sequence of subgradients $c^k\in\partial_1\overline{\psi}(x^k,z^k)$ with $(x^k,z^k)\to(\bar{x},\bar{z})$, and $\|c^k\|\to+\infty$. By taking a subsequence if necessary, we assume that the normalized subgradient $d^k=c^k/\|c^k\|$ converges to some d of unit length. Since $x^k\to\bar{x}\in\operatorname{int}(\operatorname{dom}\psi)$, there exists a positive scalar $t>0$ such that $\mathbb{B}(\bar{x},t)\subseteq\operatorname{int}(\operatorname{dom}\psi)$ and $\mathbb{B}(x^k,t)\subseteq\operatorname{int}(\operatorname{dom}\psi)$ for all sufficiently large k. Using the convexity of $\overline{\psi}(\bullet,z^k)$, we obtain $\overline{\psi}(x^k+td^k,z^k)-\overline{\psi}(x^k,z^k)\ge(c^k)^\top td^k=t\|c^k\|$. Taking limits on both sides and using the continuity of $\overline{\psi}(\bullet,\bullet)$ relative to $\operatorname{int}(\operatorname{dom}\psi)\times\operatorname{int}(\operatorname{dom}\psi)$, we have

$$+\infty>\overline{\psi}(\bar{x}+td,\bar{z})-\overline{\psi}(\bar{x},\bar{z})\ge t\lim_{k\to\infty}\|c^k\|,$$

which is a contradiction. The local Lipschitz continuity of $\overline{\psi}$ relative to $\operatorname{int}(\operatorname{dom}\psi)\times\operatorname{int}(\operatorname{dom}\psi)$ can then be easily proved. Part (c) is a consequence of [15, Proposition 4.4.26(c)] on the relationship between the subdifferentials of an icc function and its lifted counterpart. ∎

3.2. The partial Moreau envelope. Equipped with the lifted function $\overline{\psi}$, one may be able to construct computationally friendly surrogations of the recourse function in (3.1) via a modification of the usual Moreau envelope. Let us first recall the definition of the classical Moreau envelope. An extended-real-valued function f is said to be proper if $f(x)<+\infty$ for some $x\in\mathbb{R}^n$, and $f(x)>-\infty$ for all $x\in\mathbb{R}^n$. Given a proper, lower semicontinuous (lsc) function $\theta:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and a positive scalar γ, its *Moreau envelope* is

$$e_\gamma^{\mathrm{ori}}\theta(x)\triangleq\inf_{z\in\mathbb{R}^n}\left\{\theta(z)+\frac{1}{2\gamma}\|x-z\|^2\right\},\quad x\in\mathbb{R}^n.$$

We use the superscript "ori" to emphasize that this is the original definition of the Moreau envelope and is different from our later modification. The function θ is said to be prox-bounded if there exists $\gamma>0$ such that $e_\gamma^{\mathrm{ori}}\theta(x)>-\infty$ for some $x\in\mathbb{R}^n$. It is known that a proper, lsc, and convex function is always prox-bounded and its Moreau envelope is continuously differentiable (cf. [49, Theorem 2.26]). In general, for any proper, lsc, and prox-bounded function θ, the parametric functions $e_\gamma^{\mathrm{ori}}\theta(x)\uparrow\theta(x)$ as $\gamma\downarrow 0$ for all $x\in\mathbb{R}^n$. Therefore, one can view the Moreau envelope as a lower approximation of the original function. However, if θ is nonconvex and nonsmooth, the function $e_\gamma^{\mathrm{ori}}\theta$ may be neither convex nor smooth. Nevertheless, for any $x\in\mathbb{R}^n$, it holds that (see, e.g., [2, 34])

$$e_\gamma^{\mathrm{ori}}\theta(x)=\frac{1}{2\gamma}\|x\|^2-\underbrace{\sup_{z\in\mathbb{R}^n}\left\{-\theta(z)-\frac{1}{2\gamma}\|z\|^2+\frac{1}{\gamma}z^\top x\right\}}_{\text{convex in } x \text{ even if } \theta \text{ is nonconvex}},\tag{3.3}$$

which indicates that one can always obtain a dc decomposition of $e_\gamma^{\mathrm{ori}}\theta$ whether θ is convex or not. The only trouble brought by the nonconvexity of θ is that the inner

sup problem for z may not be concave (especially if θ is not weakly convex), thus one may not be able to evaluate the subgradient of the second term at a given x when using the dc algorithm to minimize the function $e_\gamma^{\rm ori}\theta$. Specifically, in the context of the recourse function (3.1), its associated Moreau envelope is

$$\begin{aligned} e_\gamma^{\rm ori}\psi(x) &= \inf_{z\in\mathbb{R}^n}\left\{\psi(z)+\frac{1}{2\gamma}\|x-z\|^2\right\} \\ &= \frac{1}{2\gamma}\|x\|^2-\sup_{z\in\overline{X},y}\left\{-f(z,y)-\frac{1}{2\gamma}\|z\|^2+\frac{1}{\gamma}z^\top y\,\middle|\,G(z,y)\le 0\right\}, \end{aligned}$$

where the inner sup problem is not jointly concave in (z,y) since f is not assumed to be jointly convex. This issue motivates us to introduce the following new type of envelopes tailored to icc functions that is more computationally tractable:

$$e_\gamma\theta(z)\triangleq\inf_{x\in\mathbb{R}^n}\left\{\overline{\theta}(x,z)+\frac{1}{2\gamma}\|x-z\|^2\right\}, \tag{3.4}$$

where $\theta:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is any icc function and $\overline{\theta}:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{\pm\infty\}$ is its lifted counterpart as in Definition 3.2. When $\overline{\theta}(x,z)$ is independent of z (so that this function only has the convex part), the above definition reduces to the usual Moreau envelope. Hence, we term the new regularization of θ in (3.4) its *partial Moreau* envelope. Similarly as in (3.3), the newly defined partial Moreau has the following explicit dc decomposition:

$$e_\gamma\theta(z)=\underbrace{\frac{1}{2\gamma}\|z\|^2}_{\text{strongly convex}}\ \underbrace{-\sup_{x\in\mathbb{R}^n}\left\{-\overline{\theta}(x,z)-\frac{1}{2\gamma}\|x\|^2+\frac{1}{\gamma}z^\top x\right\}}_{\text{denoted as } g_\gamma(z),\ \text{convex}}. \tag{3.5}$$

We denote the optimal solution mapping of the minimization problem in (3.4) as

$$P_\gamma\theta(z)\triangleq\operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{\overline{\theta}(x,z)+\frac{1}{2\gamma}\|x-z\|^2\right\},\quad z\in\mathbb{R}^n.$$

For any $z\in\mathbb{R}^n$, it holds that $\emptyset\neq P_\gamma\theta(z)\subseteq\operatorname{dom}\theta$. When $z\in\operatorname{dom}\theta$, the mapping is single-valued since the inner objective function is strongly convex in x; for this case, we follow the terminology in the literature to call $P_\gamma\theta(z)$ the proximal point of $\overline{\theta}$ at z. Similar to the classical Moreau envelope, the partial Moreau envelope approximates the original function from below. The following lemma establishes the gap between the partial Moreau envelope and the original function under the Lipschitz continuity of $\overline{\theta}(\bullet,z)$ relative to $\operatorname{dom}\theta$ for any fixed z from a compact set contained in $\operatorname{int}(\operatorname{dom}\theta)$. The latter Lipschitz continuity is stronger than the one derived in Lemma 3.5(b) that is only relative to $\operatorname{int}(\operatorname{dom}\psi)\times\operatorname{int}(\operatorname{dom}\psi)$. The proof is adapted from [39, Proposition 3.4] on a similar property regarding the classical Moreau envelope.

LEMMA 3.6. *Consider an icc function $\theta:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and its lifted counterpart $\overline{\theta}:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{\pm\infty\}$. Let X be a compact subset of* $\operatorname{int}(\operatorname{dom}\theta)$. *Assume that $\overline{\theta}(\bullet,z)$ is Lipschitz continuous relative to* $\operatorname{dom}\theta$ *with Lipschitz constant κ for every $z\in X$, i.e.,*

$$\left|\overline{\theta}(x_1,z)-\overline{\theta}(x_2,z)\right|\le\kappa\|x_1-x_2\|\quad\forall\,x_1,x_2\in\operatorname{dom}\theta,\ z\in X,$$

then $0\le\theta(z)-e_\gamma\theta(z)\le\gamma\kappa^2/2$ for any $z\in X$.

Proof. For any $z \in X$, it holds that

$$\begin{aligned} 0 \le \theta(z) - e_\gamma\theta(z) &= \overline{\theta}(z,z) - \overline{\theta}\left(P_\gamma\theta(z), z\right) - \left\|P_\gamma\theta(z) - z\right\|^2/(2\gamma) \\ &\le \kappa\left\|P_\gamma\theta(z) - z\right\| - \left\|P_\gamma\theta(z) - z\right\|^2/(2\gamma) \le \gamma\kappa^2/2, \end{aligned}$$

where the second inequality follows from the Lipschitz continuity of $\overline{\theta}(\bullet, z)$ and the last inequality uses the fact that $\max_{t\ge 0}\left[\kappa t - t^2/(2\gamma)\right] = \gamma\kappa^2/2$. ∎

With g_γ defined in (3.5), it follows from similar arguments in the proof of Lemma 3.4 that for any $z \in \operatorname{dom}\theta$,

$$\frac{1}{\gamma}P_\gamma\theta(z) + \partial_2(-\overline{\theta})(P_\gamma\theta(z), z) \subseteq \partial g_\gamma(z).$$

One can then obtain the following convex majorization of $e_\gamma\theta(z)$ at any given point $z' \in \operatorname{dom}\theta$ based on the subgradient inequality of the convex function g_γ:

$$e_\gamma\theta(z) \le \widehat{e}_\gamma\theta(z;z') \triangleq \frac{1}{2\gamma}\|z\|^2 - g_\gamma(z') - \left(P_\gamma\theta(z')/\gamma + c\right)^\top(z - z') \ \forall\, z \in \operatorname{dom}\theta, \tag{3.6}$$

where $c \in \partial_2(-\overline{\theta})\left(P_\gamma\theta(z'), z'\right)$.

4. The decomposition algorithm and its convergence. Based on the discussion in the last section, we are now ready to present the decomposition algorithm for solving the nonconvex two-stage SP (1.1) and analyze its convergence. In this section, we focus on the case where there are fixed scenarios $\{\xi^1, \ldots, \xi^S\}$, each realized with probability $1/S$. The problem (1.1) then reduces to

$$\underset{x\in X\subseteq\mathbb{R}^{n_1}}{\text{minimize}} \left\{\varphi(x) + \frac{1}{S}\sum_{s=1}^{S}\psi(x;\xi^s)\right\}, \tag{4.1}$$

where each $\psi(x;\xi^s)$ is given by (1.2). The above problem can be viewed as a sample average approximation of the two-stage SP (1.1) under a prescribed sample size S. All the discussions in this section can be easily adapted to the case where the distribution of $\tilde{\xi}$ has finite support (but unequal probability mass for different ξ^s). We will work on the internal sampling scheme for continuously distributed $\tilde{\xi}$ in the next section.

4.1. The algorithmic framework. Our goal is to solve the nonconvex problem (4.1) via a successive approximation scheme. For any $\xi \in \Xi$ and $z \in X$, the partial Moreau envelope of the recourse function (1.2) associated with the bivariate function (3.2) is

$$\begin{aligned} e_\gamma\psi(z;\xi) &\triangleq \underset{x\in\mathbb{R}^{n_1}}{\text{minimum}} \left\{\overline{\psi}(x,z;\xi) + \frac{1}{2\gamma}\|x - z\|^2\right\} \\ &= \left[\begin{array}{ll} \underset{x\in X,\, y\in\mathbb{R}^{n_2}}{\text{minimum}} & f(z,y;\xi) + \dfrac{1}{2\gamma}\|x - z\|^2 \\ \text{subject to} & G(x,y;\xi) \le 0 \end{array}\right]. \end{aligned} \tag{4.2}$$

We consider a double-loop algorithm where the outer loop updates the parameter γ in the partial Moreau envelope and the inner loop solves the nonconvex problem $\text{minimize}_{x\in X}[\,\varphi(x) + \sum_{s=1}^{S} e_{\gamma_\nu}\psi(x;\xi^s)/S\,]$ to stationarity for a fixed γ_ν. To solve the latter nonconvex problem during the νth inner loop, we replace the Moreau-regularized recourse function $e_{\gamma_\nu}\psi(x;\xi)$ with its upper approximation constructed

at the latest iterate $x_{\nu,i}$, where i denotes the inner iterate index. For each s, let $\left(x^s_{\nu,i}, y^s_{\nu,i}\right)$ be one of the optimal solutions of (4.2) at $(z,\xi) = (x_{\nu,i}, \xi^s)$, which can be computed by solving a convex optimization problem. Notice that $x^s_{\nu,i} = P_{\gamma_\nu}\psi(x_{\nu,i};\xi^s) \in \overline{X}$. One may then derive from (3.6) the following upper approximating function of $e_{\gamma_\nu}\psi(x;\xi^s)$:

$$\widehat{e}_{\gamma_\nu}\psi(x;\xi^s;x_{\nu,i}) \triangleq \frac{1}{2\gamma_\nu}\|x\|^2 - g_{\gamma_\nu}(x_{\nu,i};\xi^s) - \left(x^s_{\nu,i}/\gamma_\nu + c^s_{\nu,i}\right)^\top (x - x_{\nu,i}), \tag{4.3}$$

where $g_{\gamma_\nu}(x_{\nu,i};\xi^s) \triangleq \|x_{\nu,i}\|^2/(2\gamma_\nu) - e_{\gamma_\nu}\psi(x_{\nu,i};\xi^s)$ and $c^s_{\nu,i} \in \partial_2(-\overline{\psi})(x^s_{\nu,i}, x_{\nu,i};\xi^s)$. Due to Lemma 3.4, a particular way to choose $c^s_{\nu,i}$ is to take an element from $\partial_1(-f)\left(x^s_{\nu,i}, y^s_{\nu,i};\xi^s\right)$. The resulting master problem to generate the next first-stage iterate $x_{\nu,i+1}$ is

$$\underset{x\in X}{\text{minimize}} \left\{\varphi(x) + \frac{1}{S}\sum_{s=1}^{S} \widehat{e}_{\gamma_\nu}\psi(x;\xi^s;x_{\nu,i})\right\}. \tag{4.4}$$

The inner iteration continues until the distance between two consecutive iterates $x_{\nu,i}$ and $x_{\nu,i+1}$ is sufficiently close. We summarize the procedure of the decomposition algorithm below. When $S = 1$, it reduces to the algorithm in [15, Algorithm 7.2.1] to minimize an icc function (without decomposition).

Notice that each $\widehat{e}_{\gamma_\nu}\psi(x;\xi^s;x_{\nu,i})$ is a strongly convex quadratic function in x. Therefore, the master problem (4.4) is a strongly convex optimization problem with n_1 number of variables, which is usually easy to solve. In practical implementation, solutions to the subproblem (4.2) and the master problem (4.4) may be calculated inaccurately. For the sake of concise analysis, we have chosen not to include these potential errors in our study.

4.2. Convergence analysis. This subsection is about the global convergence of the sequence generated by Algorithm 1. We begin with several technical assumptions that will be used in our convergence analysis. Since our focus of this section is to solve problem (4.1) with fixed scenarios $\{\xi^1,\ldots,\xi^S\}$, one should interpret Assumptions A–B as requirements on all realizations $\{\xi^1,\ldots,\xi^S\}$. The "almost surely" part will be used in the next section where a generally distributed $\tilde{\xi}$ is considered.

Algorithm 1 A decomposition algorithm for the nonconvex two-stage SP (4.1)

Input: Initial point $x_0 \in X$, and two scalar sequences $\{\gamma_\nu\}_{\nu\geq 0} \downarrow 0$ and $\{\varepsilon_\nu\}_{\nu\geq 0} \downarrow 0$.

Outer loop: Set $\nu = 0$.

1: Execute the inner loop with the initial point x_ν, and parameters γ_ν and ε_ν.

2: Set $\nu \leftarrow \nu + 1$ and repeat step 1 until a prescribed stopping criterion is satisfied.

Inner loop: Set $i = 0$ and $x_{\nu,0} = x_\nu$.

1: Solve the subproblem (4.2) at $(z,\xi)=(x_{\nu,i},\xi^s)$ for all s to get solutions $(x^s_{\nu,i}, y^s_{\nu,i})$.

2: Select $c^s_{\nu,i} \in \partial_1(-f)(x^s_{\nu,i}, y^s_{\nu,i};\xi^s) \subseteq \partial_2(-\overline{\psi})(x^s_{\nu,i}, x_{\nu,i};\xi^s)$ for each s.

3: Solve the master problem (4.4) to obtain $x_{\nu,i+1}$.

4: Set $i \leftarrow i + 1$ and repeat the above steps if $\|x_{\nu,i+1} - x_{\nu,i}\| > \varepsilon_\nu\gamma_\nu$. Otherwise, break the inner loop with $x_{\nu+1} \triangleq x_{\nu,i+1}$.

Assumption A. The minimization problem of y in defining $\overline{\psi}(x,z;\tilde{\xi})$ in (3.2) has an optimal solution for any $(x,z)\in\overline{X}\times\overline{X}$ almost surely.

Assumption B. There exists a measurable function $\kappa_1:\Xi\to\mathbb{R}_+$ such that $\mathbb{E}_{\tilde{\xi}}[\kappa_1(\tilde{\xi})]<\infty$ and the following condition holds almost surely:

$$\left|\overline{\psi}(x_1,z;\tilde{\xi})-\overline{\psi}(x_2,z;\tilde{\xi})\right|\le\kappa_1(\tilde{\xi})\,\|x_1-x_2\|\quad\forall\,(x_1,x_2,z)\in\overline{X}\times\overline{X}\times X.$$

Some remarks are in order. Assumption A guarantees that for any $z\in\overline{X}$, the optimal solution of (4.2) in terms of y exists. This assumption also implies the relatively complete recourse of the original problem (1.1) that $\psi(x;\tilde{\xi})$ is finite for all feasible $x\in X$ almost surely. Assumption B is a stochastic version of the Lipschitz condition in Lemma 3.6.

For convenience, we denote

$$\overline{\zeta}_S(x)\triangleq\varphi(x)+\frac{1}{S}\sum_{s=1}^{S}\psi(x;\xi^s)\quad\text{and}\quad\widehat{\zeta}_{S,\gamma_\nu}(x)\triangleq\varphi(x)+\frac{1}{S}\sum_{s=1}^{S}e_{\gamma_\nu}\psi(x;\xi^s).\tag{4.5}$$

In the proposition below, we show that for any prescribed positive scalar ε_ν, the νth inner loop of Algorithm 1 terminates in finite steps.

PROPOSITION 4.1 (convergence of the inner loop for Algorithm 1). *Suppose that Assumptions* A–B *hold. Then the following statements hold for any νth inner loop.*

(a) $\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i+1})\le\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i})-\|x_{\nu,i}-x_{\nu,i+1}\|^2/(2\gamma_\nu)$ *for any* $i\ge0$.

(b) $\lim_{i\to+\infty}\|x_{\nu,i}-x_{\nu,i+1}\|=0$ *and the stopping criterion* $\|x_{\nu,i+1}-x_{\nu,i}\|\le\varepsilon_\nu\gamma_\nu$ *is achievable in finite steps, i.e.,*

$$i_\nu\triangleq\min\{\,i\in\mathbb{Z}_+\mid\|x_{\nu,i+1}-x_{\nu,i}\|\le\varepsilon_\nu\gamma_\nu\,\}<+\infty.$$

In addition, we have

$$\operatorname{dist}\left(0,\ \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)-\partial_2(-\overline{\psi})(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)+\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu})\right]\right.$$
$$\left.+\,\partial\varphi(x_{\nu,i_\nu+1})+\mathcal{N}_X(x_{\nu,i_\nu+1})\right)\le\varepsilon_\nu.$$

Proof. Consider the νth inner loop of Algorithm 1. We first show that the sequence $\{\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i})\}_{i\ge0}$ is nonincreasing. Since the function $\widehat{e}_{\gamma_\nu}\psi(\bullet;\xi^s;x_{\nu,i})$ defined in (4.3) is quadratic, we have, by writing $a^s_{\nu,i}\triangleq\nabla_x\widehat{e}_{\gamma_\nu}\psi(\bullet;\xi^s;x_{\nu,i})(x_{\nu,i+1})$,

$$\begin{aligned}\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i})&=\varphi(x_{\nu,i})+\frac{1}{S}\sum_{s=1}^{S}\widehat{e}_{\gamma_\nu}\psi(x_{\nu,i};\xi^s;x_{\nu,i})\\&=\varphi(x_{\nu,i})+\frac{1}{S}\sum_{s=1}^{S}\widehat{e}_{\gamma_\nu}\psi(x_{\nu,i+1};\xi^s;x_{\nu,i})\\&\quad+\frac{1}{S}\sum_{s=1}^{S}(a^s_{\nu,i})^\top(x_{\nu,i}-x_{\nu,i+1})+\frac{1}{2\gamma_\nu}\|x_{\nu,i}-x_{\nu,i+1}\|^2,\end{aligned}$$

where the first equality is because $\widehat{e}_{\gamma_\nu}\psi(x_{\nu,i};\xi^s;x_{\nu,i})=e_{\gamma_\nu}\psi(x_{\nu,i};\xi^s)$. Since $x_{\nu,i+1}$ is the optimal solution of the master problem (4.4), one may obtain that

$$\left(b_{\nu,i}+\frac{1}{S}\sum_{s=1}^{S}a^s_{\nu,i}\right)^\top(x_{\nu,i}-x_{\nu,i+1})\ge0\quad\text{for some } b_{\nu,i}\in\partial\varphi(x_{\nu,i+1}).$$

The convexity of φ and the above inequalities imply that

$$\begin{aligned}\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i}) &\geq \varphi(x_{\nu,i+1}) + \frac{1}{S}\sum_{s=1}^{S}\widehat{e}_{\gamma_\nu}\psi(x_{\nu,i+1};\xi^s;x_{\nu,i}) + \frac{1}{2\gamma_\nu}\|x_{\nu,i}-x_{\nu,i+1}\|^2 \\ &\geq \varphi(x_{\nu,i+1}) + \frac{1}{S}\sum_{s=1}^{S}e_{\gamma_\nu}\psi(x_{\nu,i+1};\xi^s) + \frac{1}{2\gamma_\nu}\|x_{\nu,i}-x_{\nu,i+1}\|^2 \\ &= \widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i+1}) + \frac{1}{2\gamma_\nu}\|x_{\nu,i}-x_{\nu,i+1}\|^2,\end{aligned} \tag{4.6}$$

where the second inequality uses the fact that $\widehat{e}_\gamma\,\psi(x;\xi^s;\bar{x}) \geq e_\gamma\psi(x;\xi^s)$ for any x, $\bar{x}$, and ξ^s. We thus prove part (a). Using the compactness of X and Assumption B, one may further derive from Lemma 3.6 that

$$\inf_{x\in X}\widehat{\zeta}_{S,\gamma_\nu}(x) \geq \left[\inf_{x\in X}\varphi(x)\right] + \frac{1}{S}\sum_{s=1}^{S}\left[\inf_{x\in X}\psi(x;\xi^s) - \frac{\gamma_\nu}{2}\cdot\kappa_1(\xi^s)^2\right] > -\infty.$$

Hence, the sequence $\{\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i})\}_{i\geq 0}$ is bounded below, which further yields that $\{\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu,i})\}_{i\geq 0}$ converges and $\|x_{\nu,i+1}-x_{\nu,i}\|$ converges to 0 as $i\to\infty$. The latter convergence indicates that the inner iterations terminate after finite steps.

To show the rest of the statement of this proposition, we first obtain from the optimality condition of the master problem (4.4) at $x_{\nu,i_\nu+1}$ that

$$\begin{aligned}0 &\in \partial\left(\varphi(x_{\nu,i_\nu+1}) + \frac{1}{S}\sum_{s=1}^{S}\widehat{e}_{\gamma_\nu}\psi(x_{\nu,i_\nu+1};\xi^s;x_{\nu,i_\nu})\right) + \mathcal{N}_X(x_{\nu,i_\nu+1}) \\ &= \partial\varphi(x_{\nu,i_\nu+1}) + \frac{1}{S}\sum_{s=1}^{S}\left(\frac{x_{\nu,i_\nu+1}-x^s_{\nu,i_\nu}}{\gamma_\nu} - c^s_{\nu,i_\nu}\right) + \mathcal{N}_X(x_{\nu,i_\nu+1}),\end{aligned} \tag{4.7}$$

where the equation is due to the sum rule of the subdifferentials for convex functions [45, Theorem 23.8]. From the optimality condition of the subproblem (4.2), we obtain

$$0 \in \partial_1\,\overline{\psi}\left(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s\right) + \left(x^s_{\nu,i_\nu}-x_{\nu,i_\nu}\right)/\gamma_\nu + \mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu}) \quad \forall\, s=1,\ldots,S.$$

Taking the sum over (4.7) and the above inclusions from $s=1,\ldots,S$, we get

$$\begin{aligned}0 &\in \frac{1}{S}\sum_{s=1}^{S}\partial_1\,\overline{\psi}\left(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s\right) + \left[\partial\varphi(x_{\nu,i_\nu+1}) - \frac{1}{S}\sum_{s=1}^{S}c^s_{\nu,i_\nu} + \mathcal{N}_X(x_{\nu,i_\nu+1})\right] \\ &\quad + (x_{\nu,i_\nu+1}-x_{\nu,i_\nu})/\gamma_\nu + \frac{1}{S}\sum_{s=1}^{S}\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu}) \\ &\subseteq \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s) - \partial_2(-\overline{\psi})(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)\right] + \partial\varphi(x_{\nu,i_\nu+1}) \\ &\quad + (x_{\nu,i_\nu+1}-x_{\nu,i_\nu})/\gamma_\nu + \mathcal{N}_X(x_{\nu,i_\nu+1}) + \frac{1}{S}\sum_{s=1}^{S}\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu}),\end{aligned}$$

where the last inclusion is due to the definition of c^s_{ν,i_ν}. Consequently, we derive

$$\begin{aligned}&\operatorname{dist}\left(0,\ \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s) - \partial_2(-\overline{\psi})(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s) + \mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu})\right]\right. \\ &\qquad\qquad \left. + \partial\varphi(x_{\nu,i_\nu+1}) + \mathcal{N}_X(x_{\nu,i_\nu+1})\right) \\ &\leq \|x_{\nu,i_\nu+1}-x_{\nu,i_\nu}\|/\gamma_\nu \leq \varepsilon_\nu,\end{aligned}$$

where the last inequality is due to the stopping rule of the inner loop. ∎

We are now ready to present the global convergence of the sequence generated by Theorem 4.2. We shall prove that every accumulation point $\bar{x}$ of $\{x_\nu\}_{\nu\geq 0}$ is a critical point of problem (4.1) satisfying

$$0 \in \partial\varphi(\bar{x}) + \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}\,(\bar{x},\bar{x};\xi^s) - \partial_2(-\overline{\psi})\,(\bar{x},\bar{x};\xi^s)\right] + \mathcal{N}_X(\bar{x}). \tag{4.8}$$

It has been shown in Lemma 3.5 that under Assumptions A and B, we have $\partial_C\psi(x;\xi) \subseteq \partial_1\overline{\psi}\,(x,x;\xi) - \partial_2(-\overline{\psi})(x,x;\xi)$. Hence, the condition (4.8) is weaker than the Clarke stationarity of problem (4.1) pertaining to $0 \in \partial\varphi(\bar{x}) + \sum_{s=1}^{S}\partial_C\psi(\bar{x};\xi^s)/S + \mathcal{N}_X(\bar{x})$. The term "critical point" is adapted from the result of the dc algorithm to solve a dc problem $\text{minimize}_{x\in X}\,[\theta_1(x) - \theta_2(x)]$, where the accumulation point satisfies $0 \in \partial\theta_1(x) - \partial\theta_2(x) + \mathcal{N}_X(x)$.

THEOREM 4.2 (subsequential convergence for Algorithm 1). *Let Assumptions* A–B *hold. Then any accumulation point of the sequence* $\{x_\nu\}_{\nu\geq 0}$ *generated by Algorithm* 1 *is a critical point of* (4.1) *satisfying* (4.8).

Proof. Let $\bar{x}$ be the limit of a convergent subsequence $\{x_{\nu+1}\}_{\nu\in N}$, where N is a subset of $\mathbb{Z}_+$. We first show that $\{x^s_{\nu,i_\nu}\}_{\nu\in N}$ also converges to $\bar{x}$. By Proposition 4.1, we know that $x_{\nu,i_\nu} \to \bar{x}$ as $\nu(\in N) \to \infty$. For each $s = 1,\ldots,S$ and $\nu \in N$, we have, by the definition of $x^s_{\nu,i_\nu} = P_{\gamma_\nu}\psi(x_{\nu,i_\nu};\xi^s) \in \overline{X}$,

$$\begin{aligned}
&\frac{1}{2\gamma_\nu}\left\|x^s_{\nu,i_\nu} - x_{\nu,i_\nu}\right\|^2 + \inf_{x\in\overline{X}}\overline{\psi}(x,x_{\nu,i_\nu};\xi^s)\\
&\quad\le \frac{1}{2\gamma_\nu}\left\|x^s_{\nu,i_\nu} - x_{\nu,i_\nu}\right\|^2 + \overline{\psi}\left(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s\right)\\
&\quad= e_{\gamma_\nu}\psi\left(x_{\nu,i_\nu};\xi^s\right) \le \frac{1}{2\gamma_\nu}\|\bar{x} - x_{\nu,i_\nu}\|^2 + \overline{\psi}(\bar{x},x_{\nu,i_\nu};\xi^s).
\end{aligned}$$

The above inequality yields that

$$\begin{aligned}
\left\|x^s_{\nu,i_\nu} - x_{\nu,i_\nu}\right\| &\le \sqrt{\|\bar{x} - x_{\nu,i_\nu}\|^2 + 2\gamma_\nu\left(\overline{\psi}(\bar{x},x_{\nu,i_\nu};\xi^s) - \inf_{x\in\overline{X}}\overline{\psi}(x,x_{\nu,i_\nu};\xi^s)\right)}\\
&\le \|\bar{x} - x_{\nu,i_\nu}\| + \sqrt{2\gamma_\nu\left(\overline{\psi}(\bar{x},x_{\nu,i_\nu};\xi^s) - \inf_{x\in\overline{X}}\overline{\psi}(x,x_{\nu,i_\nu};\xi^s)\right)}\\
&\le \|\bar{x} - x_{\nu,i_\nu}\| + \sqrt{2\gamma_\nu\,\kappa_1(\xi^s)\,R(\overline{X})} \quad \to 0 \text{ as } \nu(\in N)\to\infty,
\end{aligned}$$

where $R(\overline{X})$ denotes the diameter of the compact set $\overline{X}$, and the last inequality follows from the uniform Lipschitz continuity of $\overline{\psi}(\bullet,z;\xi)$ on $\overline{X}$ in Assumption B. Since $\kappa_1(\xi^s) < \infty$ for each s and $\gamma_\nu \downarrow 0$, we derive that

$$\left\|x^s_{\nu,i_\nu} - \bar{x}\right\| \le \left\|x^s_{\nu,i_\nu} - x_{\nu,i_\nu}\right\| + \|x_{\nu,i_\nu} - \bar{x}\| \ \to 0 \text{ as } \nu(\in N)\to\infty\,.$$

Hence, we have obtained the convergence of $\{x^s_{\nu,i_\nu}\}_{\nu\in N}$ to $\bar{x}$ for all s. Using the triangle inequality of the distance function, we have that

$$\begin{aligned}
&\operatorname{dist}\left(0,\ \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}\,(\bar{x},\bar{x};\xi^s)-\partial_2(-\overline{\psi})\,(\bar{x},\bar{x};\xi^s)\right]+\ \partial\varphi(\bar{x})+\mathcal{N}_X(\bar{x})\ \right)\\
&\le \underbrace{\operatorname{dist}\left(0,\ \frac{1}{S}\sum_{s=1}^{S}\left[\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)-\partial_2(-\overline{\psi})(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)+\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu})\right]\right.}_{\text{(i)}}\\
&\qquad\left.+\ \partial\varphi(x_{\nu,i_\nu+1})+w_\nu\right)\\
&\quad+\ \frac{1}{S}\sum_{s=1}^{S}\underbrace{\mathbb{D}\left(\begin{bmatrix}\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)\\ -\partial_2(-\overline{\psi})(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)\end{bmatrix},\begin{bmatrix}\partial_1\overline{\psi}\,(\bar{x},\bar{x};\xi^s)\\ -\partial_2(-\overline{\psi})\,(\bar{x},\bar{x};\xi^s)\end{bmatrix}\right)}_{\text{(ii)}}\\
&\quad+\ \frac{1}{S}\sum_{s=1}^{S}\underbrace{\mathbb{D}\left(\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu}),\{0\}\right)}_{\text{(iii)}}+\underbrace{\mathbb{D}\left(\partial\varphi(x_{\nu,i_\nu+1}),\,\partial\varphi(\bar{x})\right)}_{\text{(iv)}}+\underbrace{\operatorname{dist}\left(w_\nu,\,\mathcal{N}_X(\bar{x})\right)}_{\text{(v)}},
\end{aligned}$$

where w_ν can be any element in $\mathcal{N}_X(x_{\nu,i_\nu+1})$. By Proposition 4.1, there is a sequence $\{w_\nu\}_{\nu\ge0}$ with $w_\nu\in\mathcal{N}_X(x_{\nu,i_\nu+1})$ such that (i) converges to 0. Since for all s, $\lim_{\nu(\in N)\to\infty}x^s_{\nu,i_\nu}=\bar{x}\in X\subseteq\operatorname{int}(\overline{X})$, we thus obtain $\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu})=\{0\}$ for sufficiently large ν and any s. Then (iii) $\to0$ as $\nu(\in N)\to\infty$. Next we show that the terms (ii), (iv), and (v) converge to 0. It is known that the following terms are osc: $\partial_1\overline{\psi}$, $\partial_2(-\overline{\psi})$ (see Lemma 3.5(a)), $\partial\varphi$, and $\mathcal{N}_X(\bar{x})$ [49, Propositions 6.6 and 8.7]. Based on [49, Proposition 5.12], we only need to prove that for each s, the sequences

$$\{\partial_1\overline{\psi}(x^s_{\nu,i_\nu},x_{\nu,i_\nu};\xi^s)\}_{\nu(\in N)\ge\nu_0},\ \{\partial\varphi(x_{\nu,i_\nu+1})\}_{\nu(\in N)\ge\nu_0}\ \text{ and }\{w_\nu\}_{\nu(\in N)\ge\nu_0}\tag{4.9}$$

are uniformly bounded for sufficiently large ν_0. Indeed, Lemma 3.5(b) implies that for $\nu\in N$ sufficiently large and any s, the first sequence in (4.9) is uniformly bounded. The uniform boundedness of the second sequence in (4.9) is a direct consequence of [45, Theorem 24.7] since φ is real-valued and convex, and $x_{\nu,i_\nu+1}\to\bar{x}$. Last, $\{w_\nu\}_{\nu\ge0}$ must be bounded because (i) converges to 0 and all sequences in (i) except w_ν have proven to be uniformly bounded. Henceforth, we have proved that any accumulation point $\bar{x}$ is a critical point of (4.1) satisfying (4.8). ∎

In the following, we establish the convergence to a stronger type of stationarity under additional assumptions. Suppose that $f(\bullet,\bullet;\xi^s)$ and $G(\bullet,\bullet;\xi^s)$ are continuously differentiable for all ξ^s. At $x=\bar{x}$, we say $\bar{y}^s$ is an optimal solution of the convex second-stage problem with $\xi=\xi^s$ and $\bar{\lambda}^s$ being the corresponding multiplier if the following Karush–Kuhn–Tucker (KKT) condition is satisfied:

$$0\in\nabla_y f(\bar{x},\bar{y}^s;\xi^s)+\sum_{j=1}^{\ell}\bar{\lambda}^s_j\nabla_y g_j(\bar{x},\bar{y}^s;\xi^s)\ \text{ and }\ \bar{\lambda}^s\in N_{\mathbb{R}^\ell_-}\left(G(\bar{x},\bar{y}^s;\xi^s)\right).\tag{4.10}$$

We use $Y(\bar{x},\xi^s)$ and $M(\bar{x},\xi^s)$ to denote the set of all optimal solutions and multipliers satisfying the above condition, respectively. When $M(\bar{x},\xi^s)$ is nonempty, one may write the critical cone of the second-stage problem at $\bar{y}^s\in Y(\bar{x},\xi^s)$ as

$$C_{\bar{x}}(\bar{y}^s,\xi^s)\triangleq\left\{d\in\mathbb{R}^{n_2}\ \middle|\ \begin{array}{c}\nabla_y f(\bar{x},\bar{y}^s;\xi^s)^\top d=0,\\ \nabla_y g_j(\bar{x},\bar{y}^s;\xi^s)^\top d\in\mathcal{T}_{\mathbb{R}_-}(g_j(\bar{x},y;\xi)),\ j=1,\dots,\ell\end{array}\right\},$$

where $\mathcal{T}_D(x)$ denotes the tangent cone of a closed convex set D. In the following, we show that if the second-stage solutions $\{\bar{y}^s\}_{s=1}^S$ are unique at the accumulation point $\bar{x}$ for each s, then $\bar{x}$ in fact satisfies a stronger condition.

COROLLARY 4.3 (convergence to a directional stationary point). *Let $\bar{x}$ be an accumulation point of the sequence $\{x_\nu\}_{\nu\geq 0}$ generated by Algorithm* 1*. In addition to the assumptions in Theorem* 4.2*, if the conditions*

(a) *the feasible set $\{y\in\mathbb{R}^{n_2}\mid G(x,y;\xi^s)\leq 0\}$ is bounded, uniformly for $x\in X$;*
(b) *$f(\bullet,\bullet;\xi^s)$ and $G(\bullet,\bullet;\xi^s)$ are twice continuously differentiable;*
(c) *the set of multipliers $M(\bar{x},\xi^s)$ is nonempty and there exists $\bar{y}^s\in Y(\bar{x},\xi^s)$ satisfying the second order sufficient condition that for all $d\in C_{\bar{x}}(\bar{y}^s,\xi^s)\setminus\{0\}$,*

$$\sup_{\lambda\in M(\bar{y}^s,\xi^s)} d^\top\nabla^2_{yy}\left[f(\bar{x},\bar{y}^s;\xi^s)+\sum_{j=1}^{\ell}\lambda_j g_j(\bar{x},\bar{y}^s;\xi^s)\right]d>0,$$

are satisfied for each s, then $\bar{x}$ is a directional stationary point of problem (4.1)*, i.e.,*

$$\varphi'(\bar{x};d)+\frac{1}{S}\sum_{s=1}^{S}\psi'(\bar{x};d)\geq 0\quad\forall\, d\in\mathcal{T}_X(\bar{x}).$$

Proof. We first prove that $\partial_2(-\overline{\psi})(\bar{x},\bar{x};\xi^s)$ is a singleton under given assumptions. By condition (a) and the convexity of $(-f)(\bullet,z)$, we can apply the Danskin theorem [11, Theorem 2.1] to get

$$\partial_2(-\overline{\psi})(\bar{x},\bar{x};\xi^s)=\operatorname{conv}\Big\{-\nabla_x f(\bar{x},y;\xi^s)\mid y\in Y(\bar{x};\xi^s)\Big\}.$$

Since the second order sufficient condition of the second-stage problem holds at $\bar{x}$ for any ξ^s, we have that $Y(\bar{x};\xi^s)$ is a singleton [49, Example 13.25], which further implies that $\partial_2(-\overline{\psi})(\bar{x},\bar{x};\xi^s)$ is a singleton. The desired directional stationarity of $\bar{x}$ then follows from [15, Proposition 6.1.11]. ∎

Theorem 4.2 and Corollary 4.3 pertain to the subsequential convergence of the iterative sequence generated by Algorithm 1. In the following, we show that the full sequence of the objective values along the iterations converges if the sequence of the Moreau parameters $\{\gamma_\nu\}_{\nu\geq 0}$ is summable. This result particularly indicates that although the sequence $\{x_\nu\}_{\nu\geq 0}$ may have multiple accumulation points, the objective values at the accumulation points are the same. To proceed, we remind the readers of the definition of $\overline{\zeta}_S$ in (4.5).

THEOREM 4.4 (convergence of objective values for Algorithm 1). *Suppose that Assumptions* A–B *hold. Let $\{x_\nu\}_{\nu\geq 0}$ be the sequence generated by Algorithm* 1 *under the additional condition that $\sum_{\nu=0}^{\infty}\gamma_\nu<\infty$. Then $\lim_{\nu\to\infty}\overline{\zeta}_S(x_\nu)=\overline{\zeta}_S(\bar{x})$, where $\bar{x}$ is any accumulation point of the iterative sequence $\{x_\nu\}_{\nu\geq 0}$.*

Proof. One may derive that

$$\begin{aligned}\overline{\zeta}_S(x_{\nu+1})-\overline{\zeta}_S(x_\nu)=&\Big[\overline{\zeta}_S(x_{\nu+1})-\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu+1})\Big]\\&+\Big[\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu+1})-\widehat{\zeta}_{S,\gamma_\nu}(x_\nu)\Big]+\Big[\widehat{\zeta}_{S,\gamma_\nu}(x_\nu)-\overline{\zeta}_S(x_\nu)\Big],\end{aligned}\tag{4.11}$$

where the first and last terms on the right side are gaps between the partial Moreau envelopes $\widehat{\zeta}_{S,\gamma_\nu}$ and original functions $\overline{\zeta}_S$ at $x_{\nu+1}$ and x_ν, respectively. By Lemma 3.6 and Assumption B, we may obtain that

$$\overline{\zeta}_S(x_{\nu+1})-\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu+1})\leq\frac{1}{S}\sum_{s=1}^{S}\gamma_\nu\,\kappa_1(\xi^s)^2/2\quad\text{and}\quad\widehat{\zeta}_{S,\gamma_\nu}(x_\nu)-\overline{\zeta}_S(x_\nu)\leq 0.\tag{4.12}$$

Recall that $x_{\nu+1}=x_{\nu,i_\nu+1}$ and $x_\nu=x_{\nu,0}$. Then the second term on the right side of (4.11) can be bounded above based on Proposition 4.1(a) that

$$\widehat{\zeta}_{S,\gamma_\nu}(x_{\nu+1}) - \widehat{\zeta}_{S,\gamma_\nu}(x_\nu) \le -\frac{1}{2\gamma_\nu}\sum_{i=0}^{i_\nu}\|x_{\nu,i+1}-x_{\nu,i}\|^2. \tag{4.13}$$

Substituting (4.12) and (4.13) into the inequality (4.11), we have

$$\overline{\zeta}_S(x_{\nu+1}) - \overline{\zeta}_S(x_\nu) \le -\frac{1}{2\gamma_\nu}\sum_{i=0}^{i_\nu}\|x_{\nu,i+1}-x_{\nu,i}\|^2 + \frac{\gamma_\nu}{2}\left[\frac{1}{S}\sum_{s=1}^{S}\kappa_1(\xi^s)^2\right].$$

Since the sequence $\{\overline{\zeta}_S(x_\nu)\}_{\nu\ge0}$ must be bounded below due to Assumption A and $\sum_{\nu=1}^{\infty}\gamma_\nu\sum_{s=1}^{S}\kappa_1(\xi^s)^2/S<\infty$, one may easily obtain the convergence of $\{\overline{\zeta}_S(x_\nu)\}_{\nu\ge0}$ that is a so-called quasi-Fejér monotone sequence; see, e.g., [13, Lemma 3.1]. For any convergent subsequence $\{x_{\nu+1}\}_{\nu\in N}$ and its limit $\bar{x}$, by the continuity of $\psi(x;\xi^s)$ on X for each s, we have $\overline{\zeta}_S(x_{\nu+1})\to\overline{\zeta}_S(\bar{x})$ as $\nu(\in N)\to\infty$. Therefore, the full sequence $\{\overline{\zeta}_S(x_\nu)\}_{\nu\ge0}$ converges to $\overline{\zeta}_S(\bar{x})$ for any accumulation point $\bar{x}$. ∎

5. A sampling-based decomposition algorithm. In this section, we consider a generally distributed random vector $\tilde{\xi}$ with a known distribution. Instead of the approach in the previous section that deals with a fixed batch of samples throughout the algorithm, we incorporate the sampling strategy into the outer loop to progressively enlarge the problem size. In general, there are two ways to do the sampling for solving SPs. One is to use the sample average approximation to select a subset of data before the execution of the numerical algorithm [53, 54, 30]. The other is to adopt a sequential sampling technique [28, 29, 38, 50] where scenarios are gradually added along the iterations. Our method falls into the latter category.

We rely on the law of large numbers (LLN) for convex subdifferentials to establish the almost sure convergence of $\{x_\nu\}_{\nu\ge0}$ generated by Algorithm 2. To facilitate this tool, the Lipschitz continuity of the original function is needed. We have already assumed in Assumption B that $\overline{\psi}(\bullet,z;\tilde{\xi})$ is Lipschitz continuous relative to $\overline{X}$ for $z\in X$ almost surely. In the following, we further assume that $\overline{\psi}(x,\bullet;\tilde{\xi})$ is Lipschitz continuous relative to X for any $x\in\overline{X}$ almost surely. We remind readers that based on Lemma 3.5, for any fixed ξ, this Lipschitz continuity automatically holds relative to $\mathrm{int}(\overline{X})$.

Assumption C. There exists a measurable function $\kappa_2:\Xi\to\mathbb{R}_+$ such that $\mathbb{E}_{\tilde{\xi}}[\kappa_2(\tilde{\xi})]<\infty$ and the following inequality holds almost surely:

$$\left|\overline{\psi}(x,x_1;\tilde{\xi}) - \overline{\psi}(x,x_2;\tilde{\xi})\right| \le \kappa_2(\tilde{\xi})\,\|x_1-x_2\| \quad \forall\,(x,x_1,x_2)\in\overline{X}\times X\times X.$$

Algorithm 2 A sampling-based decomposition algorithm for the SP (1.1)

Input: Initial point $x_0\in X$, two positive scalar sequences $\{\gamma_\nu\}_{\nu\ge0}\downarrow0$, $\{\varepsilon_\nu\}_{\nu\ge0}\downarrow0$, and a sequence of incremental sample size $\{S_\nu\}_{\nu\ge0}$.

Outer loop: Set $S_{-1}=0$ and $\nu=0$.

1: Generate independent and identically distributed samples $\{\xi^{S_{\nu-1}+\triangle}\}_{\triangle=1}^{S_\nu-S_{\nu-1}}$ from the distribution of $\tilde{\xi}$ that are independent of previous samples.

2: Execute the inner loop of Algorithm 1 with the initial point x_ν, samples $\{\xi^s\}_{s=1}^{S_\nu}\triangleq\{\xi^s\}_{s=1}^{S_{\nu-1}}\cup\{\xi^{S_{\nu-1}+\triangle}\}_{\triangle=1}^{S_\nu-S_{\nu-1}}$, and parameters γ_ν and ε_ν.

3: Set $\nu\leftarrow\nu+1$ and repeat step 1 until a prescribed stopping criterion is satisfied.

For any $r > 0$, we denote

$$\partial_1^r \overline{\psi}(\bar{x}, \bar{x}; \xi) \triangleq \bigcup_{x,z \in \mathbb{B}(\bar{x}, r)} \partial_1 \overline{\psi}(x, z; \xi),$$

and similarly for $\partial_2^r(-\overline{\psi})$. We have the following result on the LLN for the subdifferentials of icc functions, which is a consequence of the LLN for random set-valued mappings [1]. In fact, the result can be viewed as a pointwise version of [55, Theorem 2].

LEMMA 5.1. *Suppose that Assumptions* A–C *hold. For any fixed* $x \in X$ *and any* $r > r' \geq 0$ *such that* $\mathbb{B}(x, r) \subseteq \operatorname{int}(\overline{X})$*, the following limit holds almost surely:*

$$\lim_{\nu \to \infty} \mathbb{D}\left(\frac{1}{S_\nu} \sum_{s=1}^{S_\nu} \begin{bmatrix} \partial_1^{r'} \overline{\psi}(x, x; \xi^s) \\ -\partial_2^{r'}(-\overline{\psi})(x, x; \xi^s) \end{bmatrix}, \begin{bmatrix} \mathbb{E}_{\tilde{\xi}} \left[\partial_1^r \overline{\psi}(x, x; \tilde{\xi}) \right] \\ -\mathbb{E}_{\tilde{\xi}} \left[\partial_2^r(-\overline{\psi})(x, x; \tilde{\xi}) \right] \end{bmatrix} \right) = 0,$$

where the expectation of a random set-valued mapping $\mathbb{E}_{\tilde{\xi}}[\mathcal{A}(x; \tilde{\xi})]$ *is defined as the set of* $\mathbb{E}_{\tilde{\xi}}[a(x; \tilde{\xi})]$ *for all measurable selections* $a(x; \tilde{\xi}) \in \mathcal{A}(x; \tilde{\xi})$*.*

A noteworthy remark about the preceding lemma is that we can interchange the partial subdifferential and expectation in the right-hand side of the distance, i.e.,

$$\mathbb{E}_{\tilde{\xi}} \left[\partial_1 \overline{\psi}(x, x; \tilde{\xi}) \right] = \partial_1 \mathbb{E}_{\tilde{\xi}} \left[\overline{\psi}(x, x; \tilde{\xi}) \right] \quad \forall x \in X.$$

This is because a convex function is Clarke regular at x if it is finite-valued in a neighborhood of x, and the Clarke regularity ensures the interchangeability of the subdifferential and the expectation [12, Proposition 2.3.6 and Theorem 2.7.2]. Below is the main theorem of this section on the almost surely subsequential convergence of the iterative sequence generated by Algorithm 2.

THEOREM 5.2 (subsequential convergence of Algorithm 2). *Suppose that Assumptions* A–C *hold and* $\kappa_1(\tilde{\xi})$ *in Assumption* B *is essentially bounded, i.e.,* $\inf\{t \mid \mathbb{P}(\kappa_1(\tilde{\xi}) > t) = 0\} < +\infty$*. Let* $\{x_\nu\}_{\nu \geq 0}$ *be the sequence generated by Algorithm* 2 *and* $\bar{x}$ *be any accumulation point. For any* $r > 0$ *such that* $\mathbb{B}(\bar{x}, r) \subseteq \operatorname{int}(\overline{X})$*, the following inclusion holds almost surely:*

$$0 \in \partial \varphi(\bar{x}) + \partial_1^r \mathbb{E}_{\tilde{\xi}} \left[\overline{\psi}(\bar{x}, \bar{x}; \tilde{\xi}) \right] - \partial_2^r \mathbb{E}_{\tilde{\xi}} \left[(-\overline{\psi})(\bar{x}, \bar{x}; \tilde{\xi}) \right] + \mathcal{N}_X(\bar{x}).$$

In addition, if the set-valued mapping $\partial_1 \mathbb{E}_{\tilde{\xi}}[\overline{\psi}(\bullet, \bullet; \tilde{\xi})] - \partial_2 \mathbb{E}_{\tilde{\xi}}[(-\overline{\psi})(\bullet, \bullet; \tilde{\xi})]$ *is continuous at* $(\bar{x}, \bar{x})$*, then almost surely*

$$0 \in \partial \varphi(\bar{x}) + \partial_1 \mathbb{E}_{\tilde{\xi}} \left[\overline{\psi}(\bar{x}, \bar{x}; \tilde{\xi}) \right] - \partial_2 \mathbb{E}_{\tilde{\xi}} \left[(-\overline{\psi})(\bar{x}, \bar{x}; \tilde{\xi}) \right] + \mathcal{N}_X(\bar{x}),$$

i.e., every accumulation point $\bar{x}$ *is a critical point of problem* (1.1) *almost surely.*

Proof. Consider any subsequence $\{x_{\nu+1}\}_{\nu \in N}$ that converges to $\bar{x}$. Using similar derivations as in the proof of Theorem 4.2, one can obtain the almost sure convergence of $\{x_{\nu, i_\nu}\}_{\nu \in N}$ to $\bar{x}$ and the following inequalities for each $\nu \in N$ and $s = 1, \ldots, S_\nu$:

$$\begin{aligned} \left\| x_{\nu, i_\nu}^s - \bar{x} \right\| &\leq \left\| x_{\nu, i_\nu}^s - x_{\nu, i_\nu} \right\| + \left\| x_{\nu, i_\nu} - \bar{x} \right\| \\ &\leq 2 \left\| x_{\nu, i_\nu} - \bar{x} \right\| + \sqrt{2 \gamma_\nu \kappa_1(\xi^s) R(\overline{X})}. \end{aligned}$$

Notice that $\gamma_\nu \downarrow 0$ and $\kappa_1(\xi^s)$ is almost surely bounded by a constant independent of s. Thus, for any given $r' > 0$, the ball $\mathbb{B}(\bar{x}, r')$ almost surely contains x_{ν,i_ν} and the proximal points $\{x^s_{\nu,i_\nu}\}_{s=1}^{S_\nu}$ for all $\nu \in N$ sufficiently large. Consequently,

$$\partial_1 \overline{\psi}(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) \subseteq \partial_1^{r'} \overline{\psi}(\bar{x}, \bar{x}; \xi^s) \quad \text{almost surely for all } \nu \in N \text{ sufficiently large.}$$

We then obtain from Lemma 5.1 that, for any $r > 0$ such that $\mathbb{B}(\bar{x}, r) \subseteq \mathrm{int}(\overline{X})$, the following limit holds almost surely:

(5.1)
$$\lim_{\nu(\in N)\to\infty} \mathbb{D}\left(\frac{1}{S_\nu} \sum_{s=1}^{S_\nu} \begin{bmatrix} \partial_1 \overline{\psi}(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) \\ -\partial_2(-\overline{\psi})(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) \end{bmatrix}, \begin{matrix} \partial_1^r \mathbb{E}_{\tilde{\xi}}\left[\overline{\psi}(\bar{x}, \bar{x}; \tilde{\xi})\right] \\ -\partial_2^r \mathbb{E}_{\tilde{\xi}}\left[(-\overline{\psi})(\bar{x}, \bar{x}; \tilde{\xi})\right] \end{matrix} \right) = 0.$$

Thus, the following estimation follows almost surely:

$$\begin{aligned}
&\mathrm{dist}\left(0,\, \partial\varphi(\bar{x}) + \partial_1^r \mathbb{E}_{\tilde{\xi}}\left[\overline{\psi}(\bar{x}, \bar{x}; \tilde{\xi})\right] - \partial_2^r \mathbb{E}_{\tilde{\xi}}\left[(-\overline{\psi})(\bar{x}, \bar{x}; \tilde{\xi})\right] + \mathcal{N}_X(\bar{x})\right) \\
&\le \underbrace{\mathrm{dist}\left(0,\ \frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\left[\partial_1 \overline{\psi}(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) - \partial_2(-\overline{\psi})(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) + \mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu})\right] + \partial\varphi(x_{\nu,i_\nu+1}) + w_\nu\right)}_{\text{(i}')} \\
&\quad + \underbrace{\mathbb{D}\left(\frac{1}{S_\nu} \sum_{s=1}^{S_\nu} \begin{bmatrix} \partial_1 \overline{\psi}(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) \\ -\partial_2(-\overline{\psi})(x^s_{\nu,i_\nu}, x_{\nu,i_\nu}; \xi^s) \end{bmatrix}, \begin{matrix} \partial_1^r \mathbb{E}_{\tilde{\xi}}\left[\overline{\psi}(\bar{x}, \bar{x}; \tilde{\xi})\right] \\ -\partial_2^r \mathbb{E}_{\tilde{\xi}}\left[(-\overline{\psi})(\bar{x}, \bar{x}; \tilde{\xi})\right] \end{matrix} \right)}_{\text{(ii}')} \\
&\quad + \underbrace{\frac{1}{S_\nu}\sum_{s=1}^{S_\nu} \mathbb{D}\left(\mathcal{N}_{\overline{X}}(x^s_{\nu,i_\nu}), \{0\}\right)}_{\text{(iii}')} + \underbrace{\mathbb{D}\left(\partial\varphi(x_{\nu,i_\nu+1}), \partial\varphi(\bar{x})\right)}_{\text{(iv}')} + \underbrace{\mathrm{dist}\left(w_\nu, \mathcal{N}_X(\bar{x})\right)}_{\text{(v}')},
\end{aligned}$$

where w_ν can be any element in $\mathcal{N}_X(x_{\nu,i_\nu+1})$. By Proposition 4.1 (with sample size S_ν instead of S), there is a sequence $\{w_\nu\}_{\nu\ge 0}$ with $w_\nu \in \mathcal{N}_X(x_{\nu,i_\nu+1})$ such that (i′) converges to 0. As shown in (5.1), we have (ii′) $\to 0$. The term (iii′) $\to 0$ because $\{x^s_{\nu,i_\nu}\}_{s=1}^{S_\nu} \subseteq \mathbb{B}(\bar{x}, r) \subseteq \mathrm{int}(\overline{X})$ holds almost surely for sufficiently large ν. The convergence of the last two terms (iv′) and (v′) to 0 can be derived based on similar arguments to their counterparts in the proof of Theorem 4.2. Finally, if the set-valued mapping $\partial_1 \mathbb{E}_{\tilde{\xi}}[\overline{\psi}(\bullet, \bullet; \tilde{\xi})] - \partial_2 \mathbb{E}_{\tilde{\xi}}[(-\overline{\psi})(\bullet, \bullet; \tilde{\xi})]$ is continuous at $(\bar{x}, \bar{x})$, one may adopt similar arguments as in the proof of [55, Theorem 3] to derive the almost sure convergence to a critical point. ∎

One can further derive an analogous result of Corollary 4.3 for the sequence generated by Algorithm 2 by strengthening the conditions (a), (b), and (c) in the former corollary to almost any $\xi \in \Xi$ so that $\partial_2 \mathbb{E}_{\tilde{\xi}}[(-\overline{\psi})(\bar{x}, \bar{x}; \xi)]$ is a singleton. We omit the details here for brevity. The last result of this section is the almost sure convergence of the objective values of $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\ge 0}$ under proper assumptions on the sample sizes S_ν and Moreau parameters γ_ν. To proceed, we first present a lemma on the convergence rate of the sample average approximation in expectation. This result

is obtained by using the Rademacher average of the random function $\psi(x;\tilde{\xi})$, which has its source in [17, Corollary 3.2]; see also [15, Theorem 10.1.5].

LEMMA 5.3. *Let X be a compact set in $\mathbb{R}^n$ and $\tilde{\xi}:\Omega\to\Xi\subseteq\mathbb{R}^m$ be a random vector defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. Let $\psi:X\times\Xi\to\mathbb{R}$ be a Carathéodory function. Suppose that ψ is uniformly bounded on $X\times\Xi$ and Lipschitz continuous in x with modulus independent of ξ. Let $\{\xi^s\}_{s=1}^S$ be independent and identically distributed random vectors following the distribution of $\tilde{\xi}$. Then there exists a constant C such that for any $\eta\in(0,1/2)$, we have*

$$\mathbb{E}\left[\sup_{x\in X}\left|\frac{1}{S}\sum_{s=1}^{S}\psi(x;\xi^s)-\mathbb{E}_{\tilde{\xi}}\left[\psi(x;\tilde{\xi})\right]\right|\right]\le\frac{C\sqrt{1-2\eta}}{S^{\eta}}\quad\forall\, S>0.$$

We make a remark about the above lemma. For an icc function $\psi(x;\xi)$ associated with the lifted counterpart $\overline{\psi}(x,z;\xi)$, the uniform Lipschitz continuity of $\psi(\bullet;\xi)$ holds on X when $\overline{\psi}(\bullet,z;\xi)$ is uniformly Lipschitz continuous over $(z,\xi)\in X\times\Xi$ and $\overline{\psi}(x,\bullet;\xi)$ is uniformly Lipschitz continuous over $(x,\xi)\in X\times\Xi$. Indeed, one can deduce this uniform Lipschitz continuity from Assumptions B and C with $\sup_{\xi\in\Xi}[\max(\kappa_1(\xi),\kappa_2(\xi))]<\infty$, which also implies the essential boundedness of $\kappa_1(\tilde{\xi})$ assumed in Theorem 5.2. We are now ready to present the almost surely sequential convergence of the objective values generated by the internal sampling scheme.

THEOREM 5.4 (sequential convergence of objective values for Algorithm 2). *Suppose that assumptions in Theorem 5.2 and conditions for ψ in Lemma 5.3 hold. Let $\{x_\nu\}_{\nu\ge0}$ be the sequence generated by Algorithm 2. Assume that the parameter of the partial Moreau envelope γ_ν and the sample size S_ν satisfy*

$$\sum_{\nu=1}^{\infty}\gamma_\nu<\infty,\qquad\sum_{\nu=1}^{\infty}\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}\,(S_\nu)^{\eta}}<\infty\quad\text{for some }\eta\in(0,1/2).$$

Then $\lim_{\nu\to\infty}\overline{\zeta}_{S_\nu}(x_\nu)=\overline{\zeta}(\bar{x})$ almost surely, where $\bar{x}$ is any accumulation point of the iterative sequence $\{x_\nu\}_{\nu\ge0}$.

Proof. We first prove the almost sure convergence of $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\ge0}$. We have

$$\begin{aligned}&\overline{\zeta}_{S_{\nu+1}}(x_{\nu+1})-\overline{\zeta}_{S_\nu}(x_\nu)\\&=\underbrace{\left[\overline{\zeta}_{S_{\nu+1}}(x_{\nu+1})-\overline{\zeta}_{S_\nu}(x_{\nu+1})\right]}_{\triangleq R_{\nu,1}}+\underbrace{\left[\overline{\zeta}_{S_\nu}(x_{\nu+1})-\widehat{\zeta}_{S_\nu,\gamma_\nu}(x_{\nu+1})\right]}_{\triangleq R_{\nu,2}}\\&\quad+\underbrace{\left[\widehat{\zeta}_{S_\nu,\gamma_\nu}(x_{\nu+1})-\widehat{\zeta}_{S_\nu,\gamma_\nu}(x_\nu)\right]}_{\triangleq R_{\nu,3}}+\underbrace{\left[\widehat{\zeta}_{S_\nu,\gamma_\nu}(x_\nu)-\overline{\zeta}_{S_\nu}(x_\nu)\right]}_{\triangleq R_{\nu,4}}.\end{aligned}\tag{5.2}$$

Using results of (4.12) and (4.13), we obtain

$$R_{\nu,2}\le\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\frac{\gamma_\nu\,\kappa_1(\xi^s)^2}{2},\quad R_{\nu,3}\le-\frac{1}{2\gamma_\nu}\sum_{i=0}^{i_\nu}\|x_{\nu,i+1}-x_{\nu,i}\|^2,\quad\text{and}\quad R_{\nu,4}\le0.$$

Next we compute $R_{\nu,1}$ that is the error of the sample augmentation. It holds that

$$\begin{aligned} R_{\nu,1} &= \frac{1}{S_{\nu+1}}\left[\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)+\sum_{s=S_\nu+1}^{S_{\nu+1}}\psi(x_{\nu+1};\xi^s)\right]-\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)\\ &= \left(\frac{S_\nu}{S_{\nu+1}}-1\right)\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)+\frac{1}{S_{\nu+1}}\sum_{s=S_\nu+1}^{S_{\nu+1}}\psi(x_{\nu+1};\xi^s)\\ &= \left(\frac{S_\nu}{S_{\nu+1}}-1\right)\left[\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\frac{1}{S_{\nu+1}-S_\nu}\sum_{s=S_\nu+1}^{S_{\nu+1}}\psi(x_{\nu+1};\xi^s)\right]. \end{aligned}$$

Let $\mathcal{F}_\nu \triangleq \sigma(\xi^1,\xi^2,\dots,\xi^{S_\nu})$ be a filtration, i.e., an increasing sequence of σ-fields generated by samples used in outer iterations. Obviously $x_{\nu+1}$ is adapted to $\mathcal{F}_\nu$ and $\{\xi^s\}_{s=S_\nu+1}^{S_{\nu+1}}$ are independent of $\mathcal{F}_\nu$. Therefore, by taking conditional expectation of $R_{\nu,1}$ given $\mathcal{F}_\nu$, we obtain

$$\mathbb{E}[\,R_{\nu,1}\,|\,\mathcal{F}_\nu\,]=\left(\frac{S_\nu}{S_{\nu+1}}-1\right)\left[\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x_{\nu+1};\tilde\xi)\right]\right],$$

where $\{\xi^s\}_{s=1}^{S_\nu}$ and $\tilde\xi$ are independent and identically distributed. Based on the estimations of the terms $R_{\nu,1}, R_{\nu,2}, R_{\nu,3}$, and $R_{\nu,4}$, we have, by taking conditional expectation of (5.2) given $\mathcal{F}_\nu$,

$$\begin{aligned} &\mathbb{E}\Big[\,\overline{\zeta}_{S_{\nu+1}}(x_{\nu+1})\,|\,\mathcal{F}_\nu\Big]-\overline{\zeta}_{S_\nu}(x_\nu)-\frac{1}{2\gamma_\nu}\sum_{i=0}^{i_\nu}\|x_{\nu,i+1}-x_{\nu,i}\|^2\\ &\quad\le\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}}\left|\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x_{\nu+1};\tilde\xi)\right]\right|+\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\frac{\gamma_\nu\,\kappa_1(\xi^s)^2}{2}. \end{aligned}\tag{5.3}$$

In order to show the almost sure convergence of $\{\overline{\zeta}_{s_\nu}(x_\nu)\}_{\nu\ge0}$, we need to verify that the right side of the preceding inequality is summable over ν almost surely and the sequence $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\ge0}$ is bounded below almost surely. We have

$$\begin{aligned} &\mathbb{E}\left[\sum_{\nu=1}^{\infty}\left(\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}}\right)\left|\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x_{\nu+1};\tilde\xi)\right]\right|\right]\\ &=\sum_{\nu=1}^{\infty}\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}}\,\mathbb{E}\left[\left|\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x_{\nu+1};\tilde\xi)\right]\right|\right]\\ &\le\sum_{\nu=1}^{\infty}\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}}\,\mathbb{E}\left[\sup_{x\in X}\left|\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x;\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x;\tilde\xi)\right]\right|\right]\\ &\le\sum_{\nu=1}^{\infty}\frac{(S_{\nu+1}-S_\nu)\,C\sqrt{1-2\eta}}{S_{\nu+1}\,(S_\nu)^\eta}<\infty,\ \text{for some } \eta\in(0,1/2),\ \text{by Lemma 5.3.} \end{aligned}$$

Hence, we derive that

$$\sum_{\nu=1}^{\infty}\left(\frac{S_{\nu+1}-S_\nu}{S_{\nu+1}}\right)\left|\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\psi(x_{\nu+1};\xi^s)-\mathbb{E}_{\tilde\xi}\left[\psi(x_{\nu+1};\tilde\xi)\right]\right|<\infty\quad\text{almost surely.}$$

Since $\{\gamma_\nu\}_{\nu\geq 0}$ is assumed to be summable, we can obtain

$$\mathbb{E}\left[\sum_{\nu=0}^{\infty}\left(\frac{1}{S_\nu}\sum_{s=1}^{S_\nu}\frac{\gamma_\nu\,\kappa_1(\xi^s)^2}{2}\right)\right]=\sum_{\nu=0}^{\infty}\mathbb{E}\left[\frac{\gamma_\nu}{S_\nu}\sum_{s=1}^{S_\nu}\frac{\kappa_1(\xi^s)^2}{2}\right]=\sum_{\nu=0}^{\infty}\frac{\gamma_\nu\,\mathbb{E}_{\tilde{\xi}}\left[\kappa_1(\tilde{\xi})^2\right]}{2}<\infty.$$

Consequently, $\sum_{\nu=1}^{\infty}\sum_{s=1}^{S_\nu}\gamma_\nu\,\kappa(\xi^s)^2/(2S_\nu)<\infty$ almost surely. We have thus proved that the right side of (5.3) is summable over ν almost surely. Next, we show that $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\geq 0}$ is bounded below almost surely. To see this, note that

$$\sup_{x\in X}\left|\overline{\zeta}_{S_\nu}(x)\right|\leq\sup_{x\in X}\left|\overline{\zeta}_{S_\nu}(x)-\zeta(x)\right|+\sup_{x\in X}\left|\zeta(x)\right|,$$

where the first term converges to 0 almost surely by the uniform LLN (cf. [54, Theorem 9.60]) and the second one is bounded due to the continuity of $\zeta(x)=\varphi(x)+\mathbb{E}_{\tilde{\xi}}[\psi(x;\tilde{\xi})]$ on the compact set X. Therefore, there exists a constant M such that $\overline{\zeta}_{S_\nu}(x_\nu)$ is bounded below by M almost surely for any ν. Applying the Robbins–Siegmund nonnegative almost supermartingale convergence lemma (cf. [44, Theorem 1]), we have $\sum_{\nu=1}^{\infty}\sum_{i=0}^{i_\nu}\|x_{\nu,i+1}-x_{\nu,i}\|^2/(2\gamma_\nu)<\infty$ almost surely and the sequence $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\geq 0}$ converges almost surely. Finally, let $\bar{x}$ be the limit of a convergent subsequence $\{x_\nu\}_{\nu\in N}$. Using the uniform convergence of $\overline{\zeta}_{S_\nu}$ to ζ and the continuity of ζ on the compact set X, it follows from [54, Proposition 5.1] that $\overline{\zeta}_{S_\nu}(x_\nu)$ converges to $\zeta(\bar{x})$ almost surely as $\nu(\in N)\to\infty$. This argument, together with the convergence of the full sequence $\{\overline{\zeta}_{S_\nu}(x_\nu)\}_{\nu\geq 0}$, completes the proof of this theorem. ∎

6. Numerical experiments. In this section, we present numerical results for a power system planning problem with the recourse function in (1.4) and a linear first-stage objective in terms of $x=(\{x_i\}_{i\in\mathcal{I}},\{x_g\}_{g\in\mathcal{G}})$. The overall deterministic equivalent formulation to minimize the total cost is given by

$$\begin{aligned}
&\underset{\substack{\ell_x\leq x\leq u_x,\\ \{\ell_y\leq y_s\leq u_y\}_{s=1}^{S}}}{\text{minimize}} && \sum_{i\in\mathcal{I}}c_i\,x_i+\sum_{g\in\mathcal{G}}c_g\,x_g+\sum_{s=1}^{S}\left(\sum_{g\in\mathcal{G}}p_{sg}\,x_g\right)\left[\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}(q_{is}-\pi_{js})\,y_{ijs}\right]\\
&\text{subject to} && \sum_{i\in\mathcal{I}}c_i\,x_i+\sum_{g\in\mathcal{G}}c_g\,x_g\leq B\ \text{(budget constraint)},\qquad \sum_{g\in\mathcal{G}}x_g=1,\\
&&& \sum_{j\in\mathcal{J}}y_{ijs}\leq x_i,\quad i\in\mathcal{I},\ s=1,\ldots,S\quad\text{(capacity constraints)},\\
&&& \sum_{i\in\mathcal{I}}y_{ijs}=d_{js},\quad j\in\mathcal{J},\ s=1,\ldots,S\quad\text{(demand constraints)}.
\end{aligned}\tag{6.1}$$

In our experiments, we set $|\mathcal{I}|=|\mathcal{G}|=5$ and $|\mathcal{J}|=8$. The box constraints of x and y_s are $[8,15]^5\times[0,1]^5$ and $[0,2]^{5\times 8}$ for each $s=1,\ldots,S$. The unit costs in the first-stage $\{c_i\}_{i\in\mathcal{I}}\cup\{c_g\}_{g\in\mathcal{G}}$ are independently generated from a uniform distribution on $[0,2]$. For each scenario, $\{q_{is}\}_{i\in\mathcal{I}},\{\pi_{js}\}_{j\in\mathcal{J}}$, and $\{d_{js}\}_{j\in\mathcal{J}}$ are generated from truncated normal distributions $\mathcal{N}(1,2^2)$ on $[2,3],[4,5]$, and $[1,4]$, respectively. To construct a set of probabilities $\cup_{s=1}^{S}\{p_{sg}\}_{g\in\mathcal{G}}$ satisfying $\sum_{s=1}^{S}p_{sg}=1$, we first randomly generate $S\times|\mathcal{G}|$ values from uniform distributions on $[0,1]$, and then group every S value and normalize them such that the sum of values in each group is 1. All the experiments are conducted in MATLAB 2021a on a desktop (Intel Core i7-10700 CPU at 2.90 GHz processors and 16 GB of RAM).

6.1. Fixed scenarios. Since the test example with fixed scenarios is in fact a large-scale nonconvex quadratic problem, it can also be directly solved by off-the-shelf nonlinear programming solvers. We compare the performance of our proposed decomposition algorithm based on the partial Moreau envelope (DPME) with the interior-point-based solvers Knitro [9] and IPOPT [58], both of which run with linear solver MUMPS 5.4.1. The absolute and relative feasibility and optimality errors are computed according to the termination criteria of Knitro.[1]

The quantities $\mathrm{KKT}_{\mathrm{abs}}$ and $\mathrm{KKT}_{\mathrm{rel}}$ are defined as the maximum of absolute and relative feasibility and optimality errors, respectively. The initial points are chosen to be the same for all algorithms. Although this may not necessarily force all algorithms to converge to the same objective values, we do observe such a phenomenon in the experiments. Further implementation details of these algorithms are provided below.

Knitro (version 13.0.0). The "*knitro_qp*" function is called in our numerical experiments to solve nonconvex quadratic programs from the MATLAB environment. We have set "hessopt = 0" to compute the exact Hessian in the interior point method instead of using the (L)BFGS approximations, as we have observed that the former choice is faster for all the problems tested here. We have directly set "convex = 0" to declare our problems are nonconvex so that the solver does not need to spend time on checking the convexity of the problems. In order to fairly compare the KKT errors at the computed solutions by different methods, we have disabled problem scaling and presolve options by setting "scale = 0" and "presolve = 0." We report the results based on three different settings:

1. Knitro-direct: Set "algorithm = 1" so that the direct solver is used to solve linear equations. For the termination options, the KKT relative and absolute tolerances are set to be 10^{-4} and 10^{-2}, respectively, i.e., "feastol = opttol = 10^{-4}" and "feastol_abs = opttol_abs = 10^{-2}."
2. Knitro-CG-1: Set "algorithm = 2" so that the KKT system is solved using a projected conjugate gradient method. Stopping criteria are the same as above.
3. Knitro-CG-2: All are the same as Knitro-CG-1 except that the KKT relative and absolute tolerances are set to be 10^{-6} and 10^{-3}, respectively.

IPOPT (version 3.14.4). Due to different scaling strategies and reformulations, the termination criteria of Knitro and IPOPT are not directly comparable. We set "ipopt.tol = 5×10^{-2}" in our experiments as we find the computed solutions based on this tolerance are about the same quality in terms of $\mathrm{KKT}_{\mathrm{abs}}$ and $\mathrm{KKT}_{\mathrm{rel}}$ as those provided by Knitro. We also set "ipopt.hessian_constant = 'yes'" to use exact Hessian in the interior point method, and we have not adopted the (L)BFGS method for the same reason as mentioned above.

DPME. Each master problem for the first stage and the subproblem of the second stage are convex quadratic programs, which we have called Gurobi to solve. In the same way as Knitro, we compute the relative and absolute feasibility errors (denoted as $\mathrm{feas}_{\mathrm{rel}}$ and $\mathrm{feas}_{\mathrm{abs}}$, respectively) and optimality errors (denoted as $\mathrm{opt}_{\mathrm{rel}}$ and $\mathrm{opt}_{\mathrm{abs}}$, respectively). Let the overall objective value at the νth outer loop be obj_ν. We terminate the algorithm if (6.2a) holds, and either (6.2b) or (6.2c) is true.

[1] Knitro user guide: https://www.artelys.com/docs/knitro/2_userGuide/termination.html.

$$\text{feas}_{\text{rel}} \le 10^{-4}, \ \text{feas}_{\text{abs}} \le 10^{-2}, \tag{6.2a}$$

$$\text{opt}_{\text{rel}} \le 10^{-4}, \ \text{opt}_{\text{abs}} \le 10^{-2}, \tag{6.2b}$$

$$\frac{|\text{obj}_{\nu-1} - \text{obj}_{\nu}|}{\max\{1, |\text{obj}_{\nu-1}|\}} \le 10^{-3}. \tag{6.2c}$$

Table 2 and Figure 3 summarize the performance of different algorithms when the number of scenarios S varies from 1,000 to 110,000 over 50 independent replications (the sizes of the deterministic equivalent problems are listed in Table 1). For each algorithm, we report the mean and the standard deviation of the total iteration numbers, the absolute and relative KKT errors, objective values, and the wall-clock time. One may find that for small-sized problems (such as when $S \le 10{,}000$), the interior point method that is implemented by either Knitro or IPOPT can solve the problem faster than our DPME, which may be due to two reasons: one is that the gain of the decomposition cannot compensate for the overhead of the communication between the master problem and the subproblems; the other is that we have not used the second order information as in the interior point method. However, for the cases where S is large (such as when $S \ge 30{,}000$), DPME is the fastest method and scales linearly in terms of S, which shows the power of the decomposition.

6.2. Sampling-based decomposition. We test the sampling-based DPME proposed in Algorithm 2 for the same test problem with the total number of scenarios $S = 40{,}000$. Instead of using all scenarios at each iteration, we gradually add them to reduce the computational cost especially at the early stage. In our experiments, the sample size S_ν is taken as $\eta\nu$ for different values of η. In order to understand how

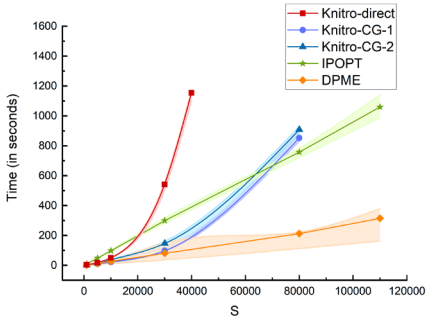

FIG. 3. *Performance of all algorithms with different numbers of scenarios S over 50 independent replications. Shaded areas: the tubes between 25% and 75% quantiles of the running time. Solid lines: means of the running time.*

TABLE 1
Dimensions of test problems. "Rows" stands for the number of constraints; "Columns" stands for the number of variables.

	Problem sizes	
S	Rows	Columns
1,000	13,000	40,010
5,000	65,000	200,010
10,000	130,000	400,010
30,000	390,000	1,200,010
80,000	1,040,000	3,200,010
110,000	1,430,000	4,400,010

TABLE 2

The performance of Knitro-direct, Knitro-CG, IPOPT, and DPME. In the table, "S" is the number of scenarios; the numbers without parentheses are the means over 100 *replications and the numbers in parentheses are the standard deviations; "t" means the method exceeds the preset time limit, which is* 1,200*s; "a" stands for Knitro-direct; "b" stands for Knitro-CG-*1*; "c" stands for Knitro-CG-*2*; "d" stands for IPOPT and "e" stands for DPME.*

	Iterations	KKT_{abs}	KKT_{rel}	Objective values	Time (in seconds)
S	a | b | c | d | e	a | b | c | d | e	a | b | c | d | e	a | b | c | d | e	a | b | c | d | e
1,000	23| 29| 38| 30| 15	2.1-3| 1.4-3| 2.4-5| 1.4-5| 2.0-3	5.5-5| 3.9-5| 6.2-7| 3.7-7| 9.8-5	**1.018**| 1.051| **1.018**| 1.020| **1.018**	**2**| **2**| 3| 11| 3
	(3| 2| 3| 3| 13)	(1.3-3| 1.2-3| 9.9-6| 1.4-6| 2.8-3)	(3.3-5| 3.2-5| 2.6-7| 3.6-8| 5.5-5)		(0| 0| 0| 1| 2)
5,000	26| 28| 38| 29| 15	2.4-3| 1.5-3| 2.6-5| 1.4-5| 2.5-3	6.3-5| 4.1-5| 6.7-7| 3.6-7| 1.1-4	y**1.033**| 1.141| 1.034| 1.041| 1.034	16| **11**| 17| 50| 13
	(6| 2| 2| 2| 13)	(9.3-4| 1.2-3| 9.2-6| 1.2-6| 3.5-3)	(2.4-5| 3.3-5| 2.4-7| 3.0-8| 7.1-5)		(2| 1| 2| 3| 10)
10,000	28| 28| 38| 29| 16	2.2-3| 1.3-3| 2.8-5| 1.7-5| 2.4-3	5.7-5| 3.7-5| 7.2-7| 4.6-7| 1.1-4	**1.024**| 1.928| 1.025| 1.033| **1.024**	48| **23**| 39| 99| 28
	(7| 3| 3| 2| 14)	(1.3-3| 9.8-4| 7.5-6| 1.1-5| 3.4-3)	(3.3-5| 2.8-5| 2.0-7| 2.8-7| 6.8-5)		(5| 2| 5| 7| 21)
30,000	29| 28| 40| 29| 17	2.0-3| 9.3-4| 1.8-5| 1.5-5| 2.6-3	5.1-5| 2.7-5| 4.8-7| 4.2-7| 1.1-4	**1.026**| 7.699| 1.027| 1.027| **1.026**	539| 99| 148| 301| **82**
	(10| 4| 6| 3| 14)	(8.5-4| 8.7-4| 9.7-6| 1.2-5| 3.4-3)	(2.2-5| 2.5-5| 2.6-7| 3.4-7| 6.8-5)		(77| 8| 32| 25| 68)
80,000	–| 26| 36| 26| 16	–| 1.0-3| 1.2-5| 1.6-5| 2.4-3	–| 2.9-5| 3.2-7| 4.5-7| 1.1-4	–| 16.376| 1.027| 1.053| **1.024**	**t**| 854| 910| 759| **214**
	(–| 4| 3| 2| 13)	(–| 9.6-4| 1.2-5| 1.5-5| 3.4-3)	(–| 2.7-5| 3.2-7| 4.2-7| 6.8-5)		(–| 133| 39| 72| 173)
110,000	–| –| –| 26| 16	–| –| –| 1.8-5| 2.5-3	–| –| –| 5.0-7| 1.1-4	–| –| –| 1.1074| **1.032**	**t**| **t**| **t**| 1061| **316**
	(–| –| –| 3| 13)	(–| –| –| 1.6-5| 3.3-3)	(–| –| –| 4.6-7| 6.7-5)		(–| –| –| 99| 253)

TABLE 3

The performance of the sampling-based DPME. In the table, "σ" is the standard deviation of the normal distribution from which we generate the data; "η" represents the linear growth rate of sample size such that the number of scenarios used in νth outer iteration $S_\nu = \eta\nu$; "–" in the column of η stands for the benchmark of DPME using full scenarios.

σ	η	Iterations outer | total	KKT_{abs}	KKT_{rel}	Objective values	Time (s)
0.5	100	2 (0) | 9 (12)	6.7-2 (1.4-1)	5.8-3 (1.2-2)	28.4047	28 (31)
	400	2 (0) | 8 (10)	2.3-2 (5.5-2)	2.0-3 (4.8-3)	28.3349	38 (48)
	800	2 (0) | 7 (9)	1.2-2 (4.1-2)	1.0-3 (3.6-3)	28.3176	40 (41)
	1600	1 (0) | 6 (8)	6.7-3 (2.3-2)	5.8-4 (2.0-3)	28.3113	34 (37)
	3200	1 (0) | 6 (8)	1.2-3 (3.4-3)	1.0-4 (3.0-4)	28.3002	32 (37)
	–	1 (0) | 6 (8)	1.1-3 (3.4-3)	9.9-5 (2.9-4)	28.3001	41 (47)
1	200	2 (0) | 8 (10)	4.6-2 (7.0-2)	2.8-3 (4.2-3)	23.2485	33 (37)
	800	2 (0) | 8 (10)	2.1-2 (5.4-2)	1.3-3 (3.3-3)	23.1989	42 (45)
	1600	2 (0) | 6 (8)	1.3-2 (3.7-2)	7.8-4 (2.3-3)	23.1871	34 (35)
	3200	2 (0) | 6 (8)	8.1-3 (2.8-2)	4.8-4 (1.7-3)	23.1778	36 (40)
	6400	1 (0) | 6 (8)	1.7-3 (3.2-3)	9.9-5 (1.9-4)	23.1648	35 (40)
	–	1 (0) | 5 (6)	1.3-3 (2.8-3)	7.8-5 (1.6-4)	23.1644	38 (37)
2	400	2 (0) | 14 (13)	8.2-2 (1.3-1)	3.1-3 (4.9-3)	13.5780	69 (58)
	1600	2 (0) | 11 (11)	4.0-2 (6.1-2)	1.5-3 (2.3-3)	13.5039	58 (53)
	3200	2 (0) | 11 (13)	2.8-2 (4.8-2)	1.1-3 (1.8-3)	13.4851	63 (65)
	6400	2 (0) | 11 (12)	2.3-2 (3.4-2)	8.5-4 (1.3-3)	13.4685	64 (65)
	12800	2 (0) | 10 (11)	1.4-2 (2.0-2)	5.4-4 (7.5-4)	13.4515	65 (69)
	–	2 (0) | 10 (12)	1.4-2 (1.9-2)	5.1-4 (7.4-4)	13.4494	69 (73)

the growth rate η depends on the distributions of the random scenarios, we generate $\{q_{is}\}_{i\in\mathcal{I}}, \{\pi_{js}\}_{j\in\mathcal{J}}$, and $\{d_{js}\}_{j\in\mathcal{J}}$ from truncated normal distributions $\mathcal{N}(1,\sigma^2)$ on $[2,13], [3,10]$, and $[1,4]$ with $\sigma \in \{0.5,1,2\}$, and vary the values of η.

The stopping criteria for the sampling-based DPME are the same as the ones for fixed scenarios, where we check the violation of the KKT system for the deterministic equivalent problem formulated using all scenarios $S = 40{,}000$. However, unlike the case for fixed scenarios, we do not have all the second-stage solutions $\{y_s\}_{s=1}^S$ to compute the KKT residual since some samples may not have been used yet. To resolve this issue, we compute all $\{y_s\}_{s=1}^S$ at every νth outer iteration, and then estimate the multipliers corresponding to the first-stage budget and box constraints by minimizing the current KKT residual.

In Table 3, we summarize the performance of Algorithm 2 for different combinations of (σ,η), where we also provide the results obtained from Algorithm 1 without sampling for benchmarks. It can be observed from the table that problems with larger variability may need a faster growth rate of the batch size to retain the same level of solution quality. If the growth rate is properly chosen, the sampling-based DPME can outperform the fixed-scenario version in the computational time with comparable solution qualities.

7. Conclusion. Compared with the extensive research on the algorithms for convex (especially linear) two-stage SPs, efficient computational algorithms for solving continuous nonconvex two-stage SPs have been much less explored. In this paper, we have made a first attempt in developing the decomposition scheme for a special class of latter problems. The key of the proposed algorithm is the derivation of successive strongly convex approximations of the nonconvex recourse functions. We hope the work done in the paper will stimulate researchers' interests in a broader paradigm of

two-stage SPs that goes beyond the classical convex settings. There are a lot of open questions that deserve future investigations, such as how to combine the stochastic dual dynamic programming approach with the tools developed in the current paper to solve nonconvex multistage SPs, as well as how to design rigorous stopping criteria for the general nonconvex SPs with continuous distributions.

REFERENCES

[1] Z. Artstein and R. A. Vitale, *A strong law of large numbers for random compact sets*, Ann. Probab., 3 (1975), pp. 879–882.

[2] E. Asplund, *Differentiability of the metric projection in finite-dimensional Euclidean space*, Proc. Amer. Math. Soc., 38 (1973), pp. 218–219.

[3] J. F. Benders, *Partitioning procedures for solving mixed-variables programming problems*, Numer. Math., 4 (1962), pp. 238–252.

[4] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, 2nd ed., Springer, New York, 2011.

[5] J. Bolte, T. Le, and E. Pauwels, *Subgradient Sampling for Nonsmooth Nonconvex Minimization*, arXiv:2202.13744, 2022.

[6] I. M. Bomze, M. Gabl, F. Maggioni, and G. C. Pflug, *Two-stage stochastic standard quadratic optimization*, European J. Oper. Res., 299 (2022), pp. 21–34.

[7] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2013.

[8] P. Borges, C. Sagastizábal, and M. Solodov, *A regularized smoothing method for fully parameterized convex problems with applications to convex and nonconvex two-stage stochastic programming*, Math. Program., 189 (2022), pp. 117–149.

[9] R. H. Byrd, J. Nocedal, and R. A. Waltz, *Knitro: An integrated package for nonlinear optimization*, in Large Scale Nonlinear Optimization, Springer, New York, 2006, pp. 35–59.

[10] B. J. Chun and S. M. Robinson, *Scenario analysis via bundle decomposition*, Ann. Oper. Res., 56 (1995), pp. 39–63.

[11] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.

[12] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[13] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, Stud. Comput. Math. 8, Elsevier, Amsterdam, 2001, pp. 115–152.

[14] K. J. Cormican, D. P. Morton, and R. K. Wood, *Stochastic network interdiction*, Oper. Res., 46 (1998), pp. 184–197.

[15] Y. Cui and J.-S. Pang, *Modern Nonconvex Nondifferentiable Optimization*, SIAM, Philadelphia, 2021.

[16] D. Davis, D. Drusvyatskiy, S. Kakade, and J. D. Lee, *Stochastic subgradient method converges on tame functions*, Found. Comput. Math., 20 (2020), pp. 119–154.

[17] Y. M. Ermoliev and V. I. Norkin, *Sample average approximation method for compound stochastic optimization problems*, SIAM J. Optim., 23 (2013), pp. 2231–2263.

[18] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, Pure Appl. Math. 40, John Wiley & Sons, New York, 1999.

[19] S. Ghadimi and G. Lan, *Stochastic first-and zeroth-order methods for nonconvex stochastic programming*, SIAM J. Optim., 23 (2013), pp. 2341–2368.

[20] V. Goel and I. E. Grossmann, *A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves*, Comput. Chem. Eng., 28 (2004), pp. 1409–1429.

[21] V. Goel and I. E. Grossmann, *A Lagrangian duality based branch and bound for solving linear stochastic programs with decision dependent uncertainty*, in European Symposium on Computer-Aided Process Engineering, Computer Aided Chemical Engineering 20, Elsevier, Amsterdam, 2005, pp. 55–60.

[22] V. Goel and I. E. Grossmann, *A class of stochastic programs with decision dependent uncertainty*, Math. Program., 108 (2006), pp. 355–394.

[23] M. Guignard, *Lagrangean relaxation*, TOP, 11 (2003), pp. 151–200.

[24] V. Guigues, *Inexact stochastic mirror descent for two-stage nonlinear stochastic programs*, Math. Program., 187 (2021), pp. 533–577.

[25] V. Guigues, R. Monteiro, and B. Svaiter, *Inexact cuts in stochastic dual dynamic programming applied to multistage stochastic nondifferentiable problems*, SIAM J. Optim., 31 (2021), pp. 2084–2110.
[26] T. Hao and J.-S. Pang, *Piecewise affine parameterized value-function based bilevel noncooperative games*, Math. Program., 180 (2020), pp. 33–73.
[27] L. Hellemo, P. I. Barton, and A. Tomasgard, *Decision-dependent probabilities in stochastic programs with recourse*, Comput. Manag. Sci., 15 (2018), pp. 369–395.
[28] J. L. Higle and S. Sen, *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Nonconvex Optim. Appl. 8, Springer, New York, 2013.
[29] T. Homem-de Mello, *Variable-sample methods for stochastic optimization*, ACM Trans. Model. Comput. Simul., 13 (2003), pp. 10–133.
[30] T. Homem-de Mello and G. Bayraksan, *Monte Carlo sampling-based methods for stochastic optimization*, Surv. Oper. Res. Manage. Sci., 19 (2014), pp. 56–85.
[31] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff, *A class of stochastic programs with decision dependent random elements*, Ann. Oper. Res., 82 (1998), pp. 83–106.
[32] J. Liu, Y. Cui, J.-S. Pang, and S. Sen, *Two-stage stochastic programming with linearly bi-parameterized quadratic recourse*, SIAM J. Optim., 30 (2020), pp. 2530–2558.
[33] F. V. Louveaux and Y. Smeers, *Optimal investments for electricity generation: A stochastic model and a test problem*, in Numerical Techniques for Stochastic Optimization Problems, Y. Ermoliev and R. J.-B. Wets, eds., Springer-Verlag, Berlin, 1988, pp. 445–452.
[34] Y. Lucet, *Fast Moreau envelope computation* I*: Numerical algorithms*, Numer. Algorithms, 43 (2006), pp. 235–249.
[35] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I*: Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.
[36] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, *Robust stochastic approximation approach to stochastic programming*, SIAM J. Optim., 19 (2009), pp. 1574–1609.
[37] O. Nohadani and K. Sharma, *Optimization under decision-dependent uncertainty*, SIAM J. Optim., 28 (2018), pp. 1773–1795.
[38] R. Pasupathy and Y. Song, *Adaptive sequential sample average approximation for solving two-stage stochastic linear programs*, SIAM J. Optim., 31 (2021), pp. 1017–1048.
[39] C. Planiden and X. Wang, *Proximal mappings and Moreau envelopes of single-variable convex piecewise cubic functions and multivariable gauge functions*, in Nonsmooth Optimization and Its Applications, Springer, New York, 2019, pp. 89–130.
[40] B. T. Polyak, *New stochastic approximation type procedures*, Automat. i Telemekh., 7 (1990).
[41] D. P. Bertsekas, *Control of Uncertain Systems with a Set-Membership Description of the Uncertainty*, Ph.D. thesis, Massachusetts Institute of Technology, 1971.
[42] J. M. Danskin, *The Theory of Max-Min and Its Application to Weapons Allocation Problems*, Economics and Operations Research 5, Springer, Berlin, 1967.
[43] H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statist., 22 (1951), pp. 400–407.
[44] H. Robbins and D. Siegmund, *A convergence theorem for nonnegative almost supermartingales and some applications*, in Optimizing Methods in Statistics, Academic Press, New York, 1971, pp. 233–257.
[45] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[46] R. T. Rockafellar, *Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity*, Set-Valued Var. Anal., 27 (2019), pp. 863–893.
[47] R. T. Rockafellar, *Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality*, Math. Program., 198 (2022), pp. 159–194.
[48] R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Math. Oper. Res., 16 (1991), pp. 119–147.
[49] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 2009.
[50] J. Royset and E. Polak, *Extensions of stochastic optimization results to problems with system failure probability functions*, J. Optim. Theory Appl., 133 (2007), pp. 1–18.
[51] A. Ruszczyński, *Decomposition methods in stochastic programming*, Math. Program., 79 (1997), pp. 333–353.
[52] A. Ruszczyński, *Convergence of a stochastic subgradient method with averaging for nonsmooth nonconvex constrained optimization*, Optim. Lett., 14 (2020), pp. 1615–1625.
[53] A. Shapiro, *Monte Carlo sampling methods*, in Stochastic Programming, Handbooks in Operations Research and Management Science 10, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 2003, pp. 353–425.

[54] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, 3rd ed., SIAM, Philadelphia, 2021.
[55] A. Shapiro and H. Xu, *Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions*, J. Math. Anal. Appl., 325 (2007), pp. 1390–1399.
[56] B. Tarhan, I. E. Grossmann, and V. Goel, *Stochastic programming approach for the planning of offshore oil or gas field infrastructure under decision-dependent uncertainty*, Ind. Eng. Chem. Res., 48 (2009), pp. 3078–3097.
[57] R. M. Van Slyke and R. J.-B, *Wets, L-shaped linear programs with applications to optimal control and stochastic programming*, SIAM J. Appl. Math., 17 (1969), pp. 638–663.
[58] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57.
[59] R. J.-B. Wets, *Large scale linear programming techniques in stochastic programming*, in Numerical Methods in Stochastic Programming, Yu. Ermoliev and R. J.-B. Wets, eds., Springer, Berlin, 1988, pp. 65–94.