

TWO-STAGE STOCHASTIC PROGRAMMING WITH LINEARLY BI-PARAMETERIZED QUADRATIC RECOURSE*

JUNYI LIU[†], YING CUI[‡], JONG-SHI PANG[†], AND SUVRAJEET SEN[†]

Abstract. This paper studies the class of two-stage stochastic programs with a linearly bi-parameterized recourse function defined by a convex quadratic program. A distinguishing feature of this new class of nonconvex stochastic programs is that the objective function in the second stage is linearly parameterized by the first-stage decision variable, in addition to the standard linear parameterization in the constraints. While a recent result has established that the resulting recourse function is of the difference-of-convex (dc) kind, the associated dc decomposition of the recourse function does not provide an easy way to compute a directional stationary solution of the two-stage stochastic program. Based on an implicit convex-concave property of the bi-parameterized recourse function, we introduce the concept of a generalized critical point of such a recourse function and provide a sufficient condition for such a point to be a directional stationary point of the stochastic program. We describe an iterative algorithm that combines regularization, convexification, and sampling and establish the subsequential convergence of the algorithm to a generalized critical point, with probability 1.

Key words. two-stage stochastic programming, difference-of-convex, directional stationarity

AMS subject classifications. 90C15, 90C26

DOI. 10.1137/19M1276819

1. Introduction. Since the mid 1950s, due to the large number of real-world decision-making problems in the presence of uncertainty, the field of stochastic programming (SP) has flourished with significant theoretical and algorithmic advances. To name a few texts, we refer the reader to Birge and Louveaux [6] for SP models and algorithms, Higle and Sen [15] for the stochastic decomposition solution approach, and Shapiro, Dentcheva, and Ruszczyński [35] for comprehensive mathematical theory.

To date, there are two overwhelming features of two-stage SP models in practical applications. One is the setting where the first-stage decisions affect only the constraints, and only linearly, and do not affect the objective of the second-stage recourse function. In the second feature, the recourse function is defined as the value function of a linear or quadratic program with a parameterized right-hand side. These two features are accompanied by a standing assumption that the recourse function is finite for both computational reasons and analytical simplifications. In view of the voluminous advances of the practical solution of nonlinear programs of many kinds in the past several decades, one is led to the question of why the state of the art of computational two-stage SP has remained largely restricted under these two features. There is perhaps a very good justification of this “constraint-only linearly parameterized recourse” paradigm in practice; namely, the resulting recourse function is convex and piecewise affine, thereby readily enabling the use of powerful linear programming

*Received by the editors July 23, 2019; accepted for publication (in revised form) June 15, 2020; published electronically September 21, 2020.

<https://doi.org/10.1137/19M1276819>

Funding: This work is based on research that was supported by the U.S. National Science Foundation under grant CMMI-1538605. The third author was additionally supported by the U.S. Air Force Office of Sponsored Research under grant FA9550-18-1-0382.

[†]Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-0193 (junyiliu@usc.edu, jongship@usc.edu, s.sen@usc.edu).

[‡]Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455 (yingcui@umn.edu).

advances for solving two-stage stochastic programs and easing the treatment of other complications, such as the generation of scenarios (i.e., the discretization/aggregation of randomness) and the large size of discretized linear programs to be solved. Among the reasons for this restricted setting are the possible loss of convexity and piecewise linearity when one deviates from the setting. The former is a serious handicap when one is committed to the computation of a globally optimal solution of the problem. However, if one is willing to trade global optimality for model fidelity, then one may be interested in an extended modeling paradigm that goes beyond the traditional linear recourse with sole right-hand side parameterization. An example is the case when both the cost vector and the right-hand constraint vector in the second-stage problem are simultaneously parameterized by the first-stage decision and uncertainty. In this case, we call the value function associated with the second-stage problem a *linearly bi-parameterized recourse function*. This paper aims to address this class of two-stage stochastic programs where, in addition, the objective function of the second-stage optimization problem is defined by a deterministic convex quadratic function.

2. The setting and literature review. This paper studies the problem defined in (2.1) and (2.2) below:

$$(2.1) \quad \underset{x}{\text{minimize}} \quad \zeta(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\xi}} \left[\psi(x, \tilde{\xi}) \right] \quad \text{subject to} \quad x \in X \subseteq \mathbb{R}^{n_1},$$

where the recourse function $\psi(x, \xi)$ is the optimal objective value of the following quadratic program (QP):

$$(2.2) \quad \begin{aligned} \psi(x, \xi) \triangleq \quad & \underset{y}{\text{minimum}} \quad [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Q y \\ & \text{subject to} \quad y \in Y(x, \xi) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\xi)x + Dy \geq b(\xi)\}. \end{aligned}$$

In this setting, $\tilde{\xi}$ is a random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with Ω the sample space whose elements are denoted by ω , \mathcal{A} the σ -algebra generated by subsets of Ω , and \mathbb{P} a probability measure defined on \mathcal{A} ; that is, $\tilde{\xi}$ is a measurable mapping from Ω into the image space $\Xi \subseteq \mathbb{R}^m$. The tilde on $\tilde{\xi}$ signifies this mapping, whereas ξ without the tilde refers to a realization of the random variable, that is, $\xi = \tilde{\xi}(\omega)$ for some $\omega \in \Omega$. The first-stage objective function φ is assumed to be convex on X , which is compact and convex. The functions in (2.2) are as follows: $f: \Xi \rightarrow \mathbb{R}^{n_2}$, $G: \Xi \rightarrow \mathbb{R}^{n_2 \times n_1}$; $A: \Xi \mapsto \mathbb{R}^{\ell \times n_1}$, and $b: \Xi \rightarrow \mathbb{R}^\ell$; thus $f(\xi)$, $G(\xi)$, $A(\xi)$, and $b(\xi)$ are the realizations of these functions composed with the random variable $\tilde{\xi}$, respectively. The deterministic constants are as follows: $Q \in \mathbb{R}^{n_2 \times n_2}$ is symmetric positive semidefinite and $D \in \mathbb{R}^{\ell \times n_2}$. We let Q and D be deterministic matrices as an extension of a standard recourse function in traditional two-stage SP in which both Q and G equal zero. For ease of reference, we summarize this setup in the following blanket assumption:

(A) *The set X is a compact convex set, and φ is a convex function on X ; moreover, Q is a symmetric positive semidefinite matrix.*

It should be pointed out that since the first-stage objective φ is assumed convex and the set X is compact, it follows that there exists a constant $\text{Lip}_\varphi > 0$ such that

$$(2.3) \quad |\varphi(x) - \varphi(x')| \leq \text{Lip}_\varphi \|x - x'\| \quad \text{for all } x, x' \in X,$$

where $\|\bullet\|$ denotes (throughout the paper) the Euclidean norm of vectors (and matrices). To avoid technical complications and to follow standard treatment in two-stage stochastic programming, we assume the following:

(B) *The second-stage problem satisfies the relatively complete recourse property on $X \times \Xi$; i.e., the recourse function $\psi(x, \xi)$ is finite for all $x \in X$ and almost all realizations $\xi \in \Xi$ of the random variable ξ (more precisely, for all $\xi \in \Xi' \triangleq \tilde{\xi}(\Omega') \subseteq \Xi$ with Ω' a subset of Ω satisfying $\mathbb{P}(\omega \in \Omega') = 1$).*

With Q being positive semidefinite, assumption (B) is equivalent to the following: for all $x \in X$ and almost all realizations $\xi \in \Xi$ of the random variable $\tilde{\xi}$, the following assumptions hold:

(B1) **Feasibility:** *The set $Y(x, \xi) \neq \emptyset$; i.e., $b(\xi) - A(\xi)x \in \text{Range } D - \mathbb{R}_+^\ell$.*

(B2) **Finiteness:** *The objective function of the recourse function is bounded below on its feasible set; i.e., $[Dv \geq 0, Qv = 0] \Rightarrow [f(\xi) + G(\xi)x]^\top v \geq 0$.*

We refer the reader to the monograph [23] for a comprehensive study of the qualitative properties of solutions to (convex) quadratic programs and to [28, section 4.1] for a supplement to the reference. We say that a random variable Z is *essentially bounded* if its essential supremum $\text{essup}(Z) \triangleq \inf \{ \alpha \mid \mathbb{P}\{Z \mid |Z| > \alpha\} = 0 \}$ is finite. To avoid some technical issues, we further assume the following:

(C) *The given random functions $f(\tilde{\xi})$, $G(\tilde{\xi})$, $A(\tilde{\xi})$, and $b(\tilde{\xi})$ are all essentially bounded; i.e., their norms are essentially bounded; hence, in particular, they have finite moments.*

A question that needs to be noted at the outset is the well-definedness of the expected recourse function $\mathbf{E}_{\tilde{\xi}}[\psi(x, \tilde{\xi})]$ under the above setting, particularly because of the unusual bi-parameterization therein. For a general treatment of this issue, the reader can consult [35, section 2.3], where properties of such an expected-value function are addressed for an abstract optimization-based recourse function; see, in particular, Theorem 7.37 of the cited reference. For our bi-parameterized quadratic recourse function, we can apply [28, Lemma 1] to deduce several properties of the recourse function $\psi(\bullet, \xi)$ and its expectation. Let $M(x, \xi)$ be the argmin (i.e., set of optimal solutions) associated with $\psi(x, \xi)$. It follows that the sets $QM(x, \xi)$ and $[f(\xi) + G(\xi)x]^\top M(x, \xi)$ are singletons for all $(x, \xi) \in X \times \Xi'$ under the relatively complete recourse assumption; moreover, for all such ξ , the unique elements in these sets are Lipschitz continuous functions on X , which is compact by assumption; moreover, by the fact that Q and D are deterministic matrices and by assumption (C), it follows that the expected recourse function $\mathbf{E}_{\tilde{\xi}}[\psi(x, \tilde{\xi})]$ is well defined, finite, and continuous [35, Theorem 7.43] for all $x \in X$.

As early as the mid 1990s, quadratic recourse functions in SP have been studied [8, 22]; section 6.2 in [6] discusses multistage stochastic programs with quadratic recourse. The recent reference [24] presents asymptotic results for two-stage stochastic programs with quadratic recourse. In these references, the second-stage problems are all of the singly parameterized kind with the first-stage variable x appearing linearly in the constraint only. In the recent paper [26], it is shown that the value function $\psi(\bullet, \xi)$ of the bi-parameterized recourse is a difference-of-convex (dc) and thus a directionally differentiable function for fixed ξ . This class of nonconvex functions has a long history; an early work is the unpublished manuscript [37]; a brief history is documented in the former reference. Thus the recourse function $\mathbf{E}_{\tilde{\xi}}[\psi(\bullet, \tilde{\xi})]$ and the combined objective function ζ are also dc. Hence, in principle, the dc algorithm (DCA) [38, 2] could be applied to solve the two-stage stochastic program (2.1). Nevertheless, there are several major difficulties in a direct application of this solution approach. First, the dc decomposition of the value function $\psi(\bullet, \xi)$ given in [26] is only of conceptual value and practically not suitable for computation. Second, some kind of discretization is needed to approximate the expectation operator. In this vein, advances in SP methodologies

can deal with the latter task; techniques such as sample average approximation (SAA), stochastic decomposition, and stochastic approximation can all be applied. However, these techniques should be combined with some convex approximations of (2.1) so that the resulting solution procedure becomes practically implementable and not just of theoretical interest. Third, the convergence of the DCA pertains to a “critical point” of the dc function to be minimized. In turn, the definition of such a point depends on a given dc decomposition of the function, in the absence of which it is not clear what kind of limit one can expect of an iterative method. For a recent reference related to this discussion, see [39]. The main goal of this work is to develop an implementable successive sampling and convex programming based algorithm to address these issues. In doing so, we acknowledge that the proposed SAA-based procedure may not be the most efficient way to solve the problem (2.1) in practice; nevertheless, in light of the lack of previous attempts to rigorously study a bi-parameterized second-stage recourse, our hope is that this paper will stimulate further research on this problem, which is significantly different from much of the existing research in computational two-stage SP that pertains largely to singly parameterized problems.

Before delving into specific details, it is best to highlight the contributions of our work. While SAA methods have been studied extensively for stochastic programs [3, 17, 18, 19, 21, 29, 32, 34, 36, 42], when applied to nonconvex stochastic programs, these methods require solving nonconvex subproblems; cf., e.g., the smooth SAA method for computing stationary points presented in [42]. Moreover, for non-differentiable problems, the limit points to which the iterates converge are stationary in a general sense that may be fairly relaxed. Most importantly, for the specific problem (2.1), none of its detailed structure is exploited by the methods available in the existing literature. Ideally, we would want to design a convergent method that can provably approximate a directional stationary point of the problem because, as noted in [27], such stationary points are the “sharpest” among stationary points of all types. Nevertheless, in spite of the work in the cited reference and its extension [25], this does not appear to be straightforward because, for one thing, the problem (2.1) lacks the kind of structure required in these references even when the objective function ζ is treated as a deterministic function. In practice, the expected value needs to be discretized via sampling when faced with general distributions; this further invalidates the deterministic approach in the cited references. Lastly, for reasons already mentioned, existing SAA-type methods do not ensure such sharp stationarity of the computed solutions. Without achieving the stated goal directly, we ask the following question: if we can design a practically implementable convex-programming based algorithm for (2.1), what stationarity property can one expect a limit point of the generated sequence to have, and when will such a point be directionally stationary? In a nutshell, the contributions of this paper are twofold: (a) to identify an implicit convex-concave property of the recourse function $\psi(\bullet, \xi)$ based on which the concept of a “generalized critical point” is defined; a sufficient condition is presented for such a point to be a directional stationary point that highlights the role of multipliers of the second-stage constraints; and (b) to present an algorithm that combines regularization, sampling, and convexification for computing such a point with probability 1.

2.1. Motivating applications. An example of a linearly bi-parameterized stochastic program can be illustrated by the two-stage shipment planning with pricing discussed in [5]. In the first stage, the decision involves the price and the number of units for production in multiple warehouses. In the second stage, after the demands

at multiple retail stores are realized, there is the option of the last-minute backup production at a higher cost and the choice of the number of units to ship from the warehouses to the retail stores. The second stage can be modeled as an optimization problem to minimize the shipping and subcontracting costs where the pricing decision has an approximately linear effect on the demand in the constraint vector and also has a linear effect on the costs appearing in the objective function; both effects occur in the second-stage decision problem. In this case, the value function associated with the second-stage problem is a linearly bi-parameterized recourse function. The SP formulation of this problem is as follows. We consider one product in a network of M factory/warehouse combination and N retail stores. The first-stage decision variables are the product's price p and the amount x_i to produce and store at each warehouse i for $i = 1, \dots, M$ at the cost of c_{1i} per unit produced. After the demand d_j at each location j for $j = 1, \dots, N$ realized in the second stage, there is an option of last-minute production y_i in factory i at an affine cost $c_{2i} + q_{2i}y_i$ per unit produced, with $q_{2i} \geq 0$, followed by a decision of the units z_{ij} to be shipped from warehouse i to the location j with the cost s_{ij} per unit shipped. Suppose that the random demand d_j at the location j is approximately linearly dependent on the price p : $d_j = \alpha_j(\tilde{\xi})p + \beta_j(\tilde{\xi})$ where $\{\alpha_j(\tilde{\xi})\}$ and $\{\beta_j(\tilde{\xi})\}$ are random coefficients independent of p . The resulting two-stage stochastic program is

$$(2.4) \quad \underset{\mathbf{x}, p \geq 0}{\text{minimize}} \quad \mathbf{c}_1^\top \mathbf{x} + \mathbf{E}_{\tilde{\xi}} [h(\mathbf{x}, p, \tilde{\xi})],$$

where $h(\mathbf{x}, p, \xi)$ is a recourse function satisfying

$$(2.5) \quad \begin{aligned} h(\mathbf{x}, p, \xi) = & \underset{\mathbf{y}, \mathbf{z} \geq 0}{\text{minimize}} \quad \sum_{i=1}^M (c_{2i} + q_{2i}y_i) y_i + \sum_{i=1}^M \sum_{j=1}^N (s_{ij} - p) z_{ij} \\ & \text{subject to} \quad \sum_{i=1}^M z_{ij} \leq \alpha_j(\xi)p + \beta_j(\xi), \quad j = 1, \dots, N \\ & \text{and} \quad \sum_{j=1}^N z_{ij} \leq x_i + y_i, \quad i = 1, \dots, M. \end{aligned}$$

In general, this kind of bi-parameterization arises naturally if the first-stage decisions have an impact on the objective of the second-stage decision-making.

Omitting the mathematical formulation, we describe another instance of how bi-parameterization may arise in power systems planning, where there is a significant proportion of renewable energy. For such systems, the two-stage paradigm of stochastic programming reflects a planning process in which electricity production from traditional thermal plants is preplanned in the first stage, whereas backup fast-ramping generators are to supplement production after the production from the renewable energy (wind and solar) is observed. Production costs in the second stage pertain to fuel costs of fast-ramping generators, and these unit costs depend on the rate of production that is necessary to meet demand. In power system planning, it is customary to go beyond constant (or even random) unit costs to affine unit costs which reflect the ramp-rates required to meet demand. Because ramping rates depend not only on the observed renewable production but also on the shortfall due to the first-stage thermal power decisions, the second-stage unit costs are dependent on the random renewable generation as well as the first stage thermal generation. With the growth of renewable energy in many jurisdictions, such modeling tools will become more appropriate in future generations of power system planning problems.

A third instance of a bi-parameterized recourse function occurs even in a traditional two-stage stochastic program where the constraint parameterization is of the kind

$$Dy \geq \min(a, b(\xi) - A(\xi)x)$$

for some constant vector a , and the “min” operator is the componentwise minimum of two vectors. The key point here is that the right-hand side is a concave, piecewise affine parameterization which we illustrate by a simple min-function. This kind of piecewise parameterization can be used to model upper bounds of the second-stage variable, such as $y \leq \max(0, a - x)$ where the prescribed bound a is diminished by the first-stage variable x . For examples of this sort of piecewise parameterization in certain network interdiction models, see [13]. As the first step to replace the piecewise right-hand side by a smooth function, we introduce the auxiliary (nonnegative) slack variable $s \triangleq Dy - \min(a, b(\xi) - A(\xi)x)$ and formulate the recourse function as

$$\begin{aligned} \psi(x, \xi) \triangleq & \underset{s \geq 0; y}{\text{minimum}} \quad [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Qy \\ & \text{subject to} \quad 0 \leq b(\xi) - A(\xi)x + s - Dy \perp a + s - Dy \geq 0, \end{aligned}$$

where the last constraint is the complementarity formulation of the piecewise affine equation $Dy - s = \min(a, b(\xi) - A(\xi)x)$, with \perp denoting the perpendicularity notation, which in this context describes the complementarity relation of two nonnegative vectors. Following well-known approximations of complementarity constraints, we may consider an approximated recourse function by employing a penalty formulation of ψ , obtaining

$$\begin{aligned} \psi_\gamma(x, \xi) \triangleq & \underset{s \geq 0; y}{\text{minimum}} \quad [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Qy \\ & + \gamma [b(\xi) - A(\xi)x + s - Dy]^\top [a + s - Dy] \\ & \text{subject to} \quad 0 \leq b(\xi) - A(\xi)x + s - Dy \quad \text{and} \quad a + s - Dy \geq 0, \end{aligned}$$

where $\gamma > 0$ is a penalty parameter. The latter function $\psi_\gamma(x, \xi)$ is clearly a bi-parameterized recourse function in the variables (s, y) with x remaining the first-stage variable for fixed γ . As a referee correctly noted, it is not immediately clear that, due to the product term, the parameter γ is independent of the realizations ξ for this penalty formulation to be exactly equivalent to the complementarity constraint. Nevertheless, under the above setting, it is reasonable for the penalty formulation to provide a relaxation, if not an exact equivalence, of the complementarity constraint. The detailed treatment of this issue is beyond the scope of this paper.

Lastly, we mention the recent paper [14] which introduced a class of stochastic programs of the traditional singly parameterized kind but with the uncertainty dependent on the first-stage variable x . Letting $p(x, \bullet)$ be the decision-dependent density function of the random variable $\tilde{\xi}$ and considering a standard linear-programming based recourse function, we may write this stochastic program as

$$(2.6) \quad \underset{x \in X}{\text{minimize}} \quad \varphi(x) + \int_{\Xi} p(x, \xi) \psi(x, \xi) d\xi, \quad \text{where } \psi(x, \xi) \triangleq \underset{y \in Y(x, \xi)}{\text{minimum}} \quad f(\xi)^\top y,$$

which is equivalent to the following bi-parameterized formulation:

$$\underset{x \in X}{\text{minimize}} \quad \varphi(x) + \int_{\Xi} \hat{\psi}(x, \xi) d\xi, \quad \text{where} \quad \hat{\psi}(x, \xi) \triangleq \underset{y \in Y(x, \xi)}{\text{minimum}} \quad p(x, \xi) f(\xi)^\top y.$$

Focusing on the case of finite scenarios, the reference [14] recognized that the problem (2.6) is nonconvex. As such, the authors apply a global optimization solver to a host of test problems and report computational results. In contrast, our treatment allows the original problem (2.1), and thus (2.6) in particular, to have general distributions, and we propose an iterative method for computing a stationary point.

Before proceeding to the main development of this paper, we note that in the case of finite scenarios, there is a complete deterministic equivalent formulation of (2.1) that is like its analogue of the singly parameterized problem (i.e., when $G(\xi) = 0$); since this formulation is fairly straightforward, we omit the details except to note that the resulting deterministic equivalent formulation is a potentially very large-scale nonconvex program with the nonconvexity being the result of the product $[G(\xi)x]^\top y$ in the joint variables x and y , and this size is proportional to the number of scenarios.

3. The combined RCS approach. In designing algorithms for solving the stochastic program (2.1) with general distributions of the random variable $\tilde{\xi}$, we are faced with several basic challenges. Besides the bi-parameterization in the recourse function and the nonconvexity and nondifferentiability of this function, the positive semidefiniteness of the matrix Q is a major concern because it could readily lead to the multiplicity of optimal solutions of the recourse function and possibly even to their unboundedness, and the same concerns hold for the constraint multipliers. The evaluation of the expected value poses a computational challenge in practical implementation. We overcome these challenges by employing regularization (R), convexification (C), and sampling (S); hence the term “RCS” approach. Overall, these three maneuvers lead to a convex-programming based, regularized, and sampled method for solving (2.1).

3.1. Regularization. Regularization of the second-stage QP is a key step in defining the algorithm. Specifically, for a given scalar $\alpha \geq 0$ such that $Q_\alpha \triangleq Q + \alpha \mathbb{I}$ is positive definite, where \mathbb{I} is the identity matrix, we consider the following Tikhonov regularization of the recourse function:

$$(3.1) \quad \begin{aligned} \psi_\alpha(x, \xi) \triangleq & \underset{y}{\text{minimum}} \quad [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Q_\alpha y \\ & \text{subject to} \quad y \in Y(x, \xi) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\xi)x + Dy \geq b(\xi)\}. \end{aligned}$$

If Q happens to be positive definite, we may take α to be zero. Throughout the treatment below, the matrix Q_α is always positive definite; thus $\alpha + \rho_{\min}(Q)$ is always positive, where $\rho_{\min}(Q)$ is the smallest eigenvalue of Q . While the reference [26] demonstrated the dc property of the value function of a bi-parameterized convex quadratic program, a dc decomposition of $\psi(x, \xi)$ is quite involved in general. Nevertheless, for the regularized recourse function $\psi_\alpha(x, \xi)$, we can obtain a dc de-

composition rather easily. Indeed, we have

$$\begin{aligned}
 & \psi_\alpha(x, \xi) \\
 &= \underset{y \in Y(x, \xi)}{\text{minimum}} \begin{bmatrix} \frac{1}{2} \left[y + [Q_\alpha]^{-1} (f(\xi) + G(\xi)x) \right]^\top Q_\alpha \left[y + [Q_\alpha]^{-1} (f(\xi) + G(\xi)x) \right] \\ -\frac{1}{2} [f(\xi) + G(\xi)x]^\top [Q_\alpha]^{-1} [f(\xi) + G(\xi)x] \end{bmatrix} \\
 &= \underset{\lambda \geq 0}{\text{maximum}} \begin{bmatrix} -\frac{1}{2} \lambda^\top D [Q_\alpha]^{-1} D^\top \lambda \\ -\frac{1}{2} [f(\xi) + G(\xi)x]^\top [Q_\alpha]^{-1} [f(\xi) + G(\xi)x] \\ +\lambda^\top [b(\xi) - A(\xi)x + D [Q_\alpha]^{-1} (f(\xi) + G(\xi)x)] \end{bmatrix} \quad \text{by duality} \\
 &= \psi_{\alpha,1}(x, \xi) - \psi_{\alpha,2}(x, \xi),
 \end{aligned}$$

where both $\psi_{\alpha,1}(\bullet, \xi)$ and $\psi_{\alpha,2}(\bullet, \xi)$ are convex functions for fixed ξ and given by (3.2)

$$\begin{aligned}
 \psi_{\alpha,1}(x, \xi) &\triangleq \underset{\lambda \geq 0}{\text{maximum}} \begin{bmatrix} -\frac{1}{2} \lambda^\top D [Q_\alpha]^{-1} D^\top \lambda \\ +\lambda^\top [b(\xi) - A(\xi)x + D [Q_\alpha]^{-1} (f(\xi) + G(\xi)x)] \end{bmatrix}, \\
 \psi_{\alpha,2}(x, \xi) &\triangleq \frac{1}{2} [f(\xi) + G(\xi)x]^\top [Q_\alpha]^{-1} [f(\xi) + G(\xi)x].
 \end{aligned}$$

Thus we have obtained a dc decomposition of the regularized value function $\psi_\alpha(\bullet, \xi)$. Note that for fixed α , $\psi_{\alpha,1}(\bullet, \xi)$ is piecewise linear-quadratic and thus generally not differentiable, while $\psi_{\alpha,2}(\bullet, \xi)$ is quadratic and thus differentiable. The two-stage stochastic program (2.1) may then be approximated by the following stochastic dc program:

$$(3.3) \quad \underset{x \in X}{\text{minimize}} \quad \zeta_\alpha(x) = \varphi(x) + \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha,1}(x, \tilde{\xi})] - \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha,2}(x, \tilde{\xi})].$$

For a given vector $x' \in X$, using the gradient $\nabla_x \psi_{\alpha,2}(x', \xi)$ to define the “linearization” of the function $\psi_{\alpha,2}(\bullet, \xi)$ at the vector x' , we obtain the following semilinearization of the function ζ_α at x' :

$$\begin{aligned}
 (3.4) \quad \hat{\zeta}_\alpha(x; x') &\triangleq \varphi(x) + \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha,1}(x, \tilde{\xi})] - \mathbf{E}_{\tilde{\xi}} [\hat{\psi}_{\alpha,2}(x, \tilde{\xi}; x')] \\
 &\text{with } \hat{\psi}_{\alpha,2}(x, \xi; x') \triangleq \psi_{\alpha,2}(x', \xi) + \nabla_x \psi_{\alpha,2}(x', \xi)^\top (x - x').
 \end{aligned}$$

Note that $\hat{\zeta}_\alpha(\bullet; x')$ is a convex function on X ; moreover, we have $\zeta_\alpha(x') = \hat{\zeta}_\alpha(x'; x')$ and $\zeta_\alpha(x) \leq \hat{\zeta}_\alpha(x; x')$ for all $x \in X$.

For solving (2.1), we employ independent and identically distributed (iid) samples of the random variable $\tilde{\xi}$ to approximate the expectation in the above convex majorization $\hat{\zeta}_\alpha(x; x')$ by their sample averages. Combining regularization, convexification, and sampling, we may now state the proposed algorithm for solving problem (2.1). In the algorithm, a sequence of iterates is generated by solving convex sub-programs. We leave open how this is done, with the understanding that this can be accomplished by a host of state-of-the-art convex-programming algorithms.

The RCS Algorithm

• **(Initialization)** Let $\{L_\nu\}_{\nu=0}^\infty \uparrow \infty$ be a sequence of positive integers, and let $\{\alpha_\nu\}_{\nu=0}^\infty \downarrow 0$ be a sequence of nonnegative scalars such that Q_{α_ν} is positive definite for all ν . Let $\gamma > 0$ be a given scalar. Let an initial feasible vector $x^0 \in X$ be given. Set $\nu = 0$.

• **(Main iteration)** At iteration ν , generate iid samples $\{\xi^{\nu,i}\}_{i=1}^{L_\nu}$ that are also independent from those in the past iterations. Generate the next iterate $x^{\nu+1}$ by solving a convex program:

$$(3.5) \quad x^{\nu+1} = \underset{x \in X}{\operatorname{argmin}} \left[\underbrace{\varphi(x) + \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \psi_{\alpha_\nu,1}(x, \xi^{\nu,i}) - \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \hat{\psi}_{\alpha_\nu,2}(x, \xi^{\nu,i}; x^\nu)}_{\text{convex in } x; \text{ denoted } \hat{\zeta}_{\alpha_\nu}(x; x^\nu)} + \frac{1}{2\gamma} \|x - x^\nu\|^2 \rightarrow \text{proximal term} \right].$$

In classical stochastic gradient algorithms for convex optimization problems, either we could choose a fixed batch size (i.e., the integer L_ν at each iteration) with decreasing step sizes (given by iteration-dependent γ_ν) similar to Robbins–Monro stochastic approximation [30], or we could choose an increasing batch size with constant step sizes. However, for a nonconvex problem, the choice is not so clear cut. According to our subsequent analysis, diminishing step sizes could lead to the accumulated error becoming unbounded, thus making it difficult to establish the convergence of the algorithm. Thus here the step size is assumed to be a fixed positive constant γ , and the sample sizes $\{L_\nu\}$ are chosen from an unbounded sequence satisfying a summability condition. In this setting, the convergence of the algorithm (see Theorem 5.6) requires conditions on the two sequences $\{\alpha_\nu\}$ and $\{L_\nu\}$ but not on γ which can be arbitrary. The plots in Figure 4 in section 6 show performance of the RCS algorithm relative to γ . The proximal term strongly convexifies the convex function $\hat{\zeta}_{\alpha_\nu}(\bullet; x^\nu)$. Dependent on the selected samples and the immediate past iterate, each new iterate $x^{\nu+1}$ is a measurable random vector [11]. In what follows, we aim to establish an almost sure stationarity property of every accumulation point of the sequence $\{x^{\nu+1}\}_{\nu=0}^\infty$ of such iterates. Starting from the pioneering work of Dupačova and Wets [11], the almost-sure convergence of the SAA-type scheme has been studied extensively; some references include [21, 17, 18, 19, 29, 32, 34, 36, 35, 42] and, more recently, [3]. Nevertheless, none of the results is applicable to establishing the convergence of the RCS algorithm, mainly because this algorithm pertains to a particular sequence generated by a combination of the SAA scheme with variable sample sizes along with convexification and regularization. The last two requirements (convexification and regularization) are introduced for the sake of algorithmic realizability.

In spite of the above-mentioned differences with a classical SAA-type scheme applied directly to the stochastic program (2.1)–(2.2), given the iterative x^ν , the next iterate $x^{\nu+1}$ can be interpreted as the result of applying one step of SAA approximation to

$$\underset{x \in X}{\operatorname{minimize}} \varphi(x) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_\nu,1}(x, \tilde{\xi}) - \hat{\psi}_{\alpha_\nu,2}(x, \tilde{\xi}; x^\nu) \right] + \frac{1}{2\gamma} \|x - x^\nu\|^2$$

that is derived from the dc-based convexification of the nonconvex regularized recourse function $\mathbf{E}_{\tilde{\xi}} \psi_{\alpha_\nu}(x, \tilde{\xi})$. This identification plays an important role in the proof of convergence of the RCS algorithm; see (5.13).

It is possible to consider two variants of the RCS algorithm. The first variant is that we leave the expectation operator alone and do not explicitly convexify the regularized recourse function. This results in the following vanilla version of a regularized method for solving (2.1):

• (Sequential regularization only) Here we have the deterministic iterates $\{\hat{x}^\nu\}$, where each \hat{x}^ν is a d-stationary solution of the dc stochastic program

$$(3.6) \quad \underset{x \in X}{\text{minimize}} \quad \zeta_{\alpha_\nu}(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_\nu}(x, \tilde{\xi}) \right].$$

This variant gives the option of using whatever algorithm is desired for generating the sequence $\{\hat{x}^\nu\}$. The drawback of this much simplified algorithm is that each subproblem (3.6) is a dc program involving the expected value of a bi-parameterized strictly convex recourse function. Thus, for practical purposes, we do not assume that each \hat{x}^ν is an optimal solution of (3.6) but rather only a directional stationary point. The other variant of the RCS algorithm is a little more specific and employs the DCA but still leaves the expectation operator alone. In this version, strictly convex SP subproblems are involved:

• (Simultaneous regularization and convexification) Here we have the iterates $\{\tilde{x}^\nu\}$, where each $\tilde{x}^{\nu+1}$, given \tilde{x}^ν , is the unique minimizer of the strongly convex stochastic program with a strongly convexified recourse function:

$$\underset{x \in X}{\text{minimize}} \quad \hat{\zeta}_{\alpha_\nu}(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_{\nu,1}}(x, \tilde{\xi}) \right] - \mathbf{E}_{\tilde{\xi}} \left[\hat{\psi}_{\alpha_{\nu,2}}(x, \tilde{\xi}; \tilde{x}^\nu) \right] + \frac{1}{2\gamma} \|x - \tilde{x}^\nu\|^2.$$

In the rest of the paper, we present a detailed convergence analysis of the RCS algorithm. With suitable modifications, the same analysis can be applied to the above two variations. The first question to be addressed is what stationarity property of the stochastic program (2.1) produced by an accumulation point of the sequence is satisfied. In the next section, we introduce a new stationarity concept that provides the answer. A key technical step in the analysis is that we need to derive uniform bounds for various function values, subgradients, gradients, and error estimates so that probabilistic convergence results can be applied. This requires significant preparations which we will provide as part of the convergence analysis in section 5.

4. Implicit dc functions and generalized critical points. This section is divided into three subsections. Subsection 4.1 reviews the definitions of some known stationarity concepts in preparation for the introduction of a new criticality property in subsection 4.3. In between, we present in subsection 4.2 a directional derivative formula for the recourse function $\psi_\alpha(\bullet, \xi)$ (for $\alpha \geq 0$) as a key to connecting the new criticality property with the familiar stationarity concepts.

4.1. Known concepts of stationarity. In the literature of dc programming, three basic kinds of stationary solutions have received the most attention: the directional derivative based stationarity [27], the convex-analysis based Clarke stationarity [9], and the dc-decomposition based critical point [38], with the first kind of stationary solutions being “sharpest” as a necessary condition for a local minimizer; i.e., a local minimizer must be directionally stationary, and a directional stationary solution must be stationary in the latter two senses but not conversely. The ultimate aim of the analysis of the RCS is to establish its (subsequential) convergence to a directional stationary solution of (2.1) under some additional assumptions. This is done by way of a novel kind of stationarity concept to be introduced in the next subsection. Here, we summarize some well-known derivative concepts of nonsmooth functions and associated definitions of stationarity.

DEFINITION 4.1. Let $g : Z \rightarrow \mathbb{R}$ be a function defined on an open set $Z \subseteq \mathbb{R}^n$.

(a) The one-sided directional derivative of g at $\bar{z} \in Z$ along the direction $d \in \mathbb{R}^n$ is

$$g'(\bar{z}; d) \triangleq \lim_{\tau \downarrow 0} \frac{g(\bar{z} + \tau d) - g(\bar{z})}{\tau}$$

if the limit exists; g is said to be directionally differentiable at $\bar{z} \in Z$ if $g'(\bar{z}; d)$ exists for all $d \in \mathbb{R}^n$.

(b) The Clarke directional derivative of g at $\bar{z} \in Z$ along the direction $d \in \mathbb{R}^n$ is

$$g^0(\bar{z}; d) \triangleq \limsup_{\substack{z \rightarrow \bar{z} \\ \tau \downarrow 0}} \frac{g(z + \tau d) - g(z)}{\tau},$$

which is finite when g is Lipschitz continuous near \bar{z} .

(c) We say that the function g is Clarke regular at $\bar{z} \in Z$ if g is directionally differentiable at \bar{z} and $g^0(\bar{z}; d) = g'(\bar{z}; d)$ for all $d \in \mathbb{R}^n$.

Clearly, $g^0(\bar{z}; d) \geq g'(\bar{z}; d)$ for all $d \in \mathbb{R}^n$. The Clarke subdifferential of g at \bar{z} is the set $\partial_C g(\bar{z}) \triangleq \{v \mid g^0(\bar{z}; d) \geq v^T d \text{ for all } d \in \mathbb{R}^n\}$. For a convex function g , $\partial_C g$ coincides with the subgradient ∂g in convex analysis. Based on Definition 4.1, we define the d(irectional)-stationarity and C(larke)-stationarity as follows. We also include the concept of a critical point of a dc function.

DEFINITION 4.2. Let $f : Z \rightarrow \mathbb{R}$ be a locally Lipschitz and directionally differentiable function defined on an open set $Z \subseteq \mathbb{R}^n$ containing the closed convex set X . A vector $\bar{x} \in X$ is said to be a

- d(irectional)-stationary point of f on X if $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$;
- C(larke)-stationary point of f on X if $f^0(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$ or, equivalently, if $0 \in \partial_C f(\bar{x}) + \mathcal{N}(\bar{x}; X)$ where $\mathcal{N}(\bar{x}; X)$ is the normal cone of X at \bar{x} .

If $f = g - h$ is a dc function with g and h convex, then $\bar{x} \in X$ is said to be a critical point of f on X if $0 \in [\partial g(\bar{x}) - \partial h(\bar{x})] + \mathcal{N}(\bar{x}; X)$.

Clearly, d-stationarity \Rightarrow C-stationarity \Rightarrow criticality; the latter is for a dc function.

4.2. A directional derivative formula. In the following, we provide an explicit directional derivative formula for the recourse function $\psi(\bullet, \xi)$. While such formulas have been obtained for the value function of linear programs (see the early papers [41, 16]) and general convex programs under some regularity conditions, such as solution boundedness and constraint qualifications, a comprehensive perturbation (including directional) analysis of parametric nonlinear programs can be found in [7]. The most one relevant to a bi-parameterized convex quadratic program is [20, Corollary 3.3] which gives a directional derivative formula for the value function of a parametric nonlinear program under the constant rank constraint qualification (CRCQ) and certain abstract stability conditions. While the CRCQ is immediately satisfied by linear constraints, the stability conditions can be verified to hold under the relatively complete recourse assumption. Interestingly, the directional derivative formula in question can also be proved by using a sum property of the total directional derivative of the bivariate function $\psi(\bullet, \bullet, \xi)$ (see (4.2)) and the partial directional derivatives of the two partial functions $\psi(\bullet, z, \xi)$ and $\psi(x, \bullet, \xi)$; see [31] and [12, Exercise 3.7.4]. In what follows, we present this formula for the regularized recourse function $\psi_\alpha(x, \xi)$ for an arbitrary $\alpha \geq 0$.

To prepare for this formula, we let $M^\alpha(x, \xi)$ and $\Lambda^\alpha(x, \xi)$ denote, respectively, the sets of optimal primal and optimal dual solutions of the second-stage regularized value function $\psi_\alpha(x, \xi)$ given by (3.1). Using the fact that $QM^\alpha(x, \xi)$ and

$[f(\xi) + G(\xi)x]^\top M^\alpha(x, \xi)$ are singletons [28, Lemma 1], we can use any $\bar{y} \in M^\alpha(x, \xi)$ to represent the dual optimal set as follows:

$$\Lambda^\alpha(x, \xi) = \left\{ \lambda \geq 0 \mid \begin{array}{ll} D^\top \lambda & = f(\xi) + G(\xi)x + Q_\alpha \bar{y} \\ \lambda^\top [b(\xi) - A(\xi)x] & = 2\psi_\alpha(x, \xi) - \bar{y}^\top [f(\xi) + G(\xi)x] \end{array} \right\},$$

which shows, in particular, that $\Lambda^\alpha(x, \xi)$ is a polyhedral set dependent on the pair (x, ξ) only and independent of the primal optimal solution \bar{y} . The primal optimal set has the following polyhedral representation: define the index set

$$\mathcal{I}^\alpha(x, \xi) \triangleq \{i \mid \exists \lambda \in \Lambda^\alpha(x, \xi) \text{ with } \lambda_i > 0\};$$

we then have, for any $\bar{\lambda} \in \Lambda^\alpha(x, \xi)$,

$$(4.1) \quad M^\alpha(x, \xi) = \left\{ y \in Y(x, \xi) \mid \begin{array}{l} f(\xi) + G(\xi)x + Q_\alpha y = D^\top \bar{\lambda} \\ [A(\xi)x + Dy = b(\xi)]_i \quad \text{for all } i \in \mathcal{I}^\alpha(x, \xi) \end{array} \right\}.$$

Let $L_\alpha(x, \xi; y, \lambda) \triangleq [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Q_\alpha y + \lambda^\top [b(\xi) - A(\xi)x - Dy]$ denote the Lagrangian function of the QP associated with $\psi_\alpha(x, \xi)$, with λ the constraint multiplier. We note that $\nabla_x L_\alpha(x, \xi; y, \lambda) = G(\xi)^\top y - A(\xi)^\top \lambda$. Let

$$(4.2) \quad \begin{aligned} \bar{\psi}_\alpha(x, z, \xi) &\triangleq \underset{y}{\text{minimum}} \quad [f(\xi) + G(\xi)z]^\top y + \frac{1}{2} y^\top [Q + \alpha \mathbb{I}] y \\ &\text{subject to } y \in Y(x, \xi) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\xi)x + Dy \geq b(\xi)\} \end{aligned}$$

be the lifted regularized recourse function corresponding to the regularized recourse $\psi_\alpha(x, \xi)$. The proof of the following result can be found in the above-cited references.

PROPOSITION 4.3 ([7] and [20, Corollary 3.3]). *Under assumptions (A) and (B), for any $\alpha \geq 0$, the directional derivatives of $\psi_\alpha(\bullet, \xi)$ exist on X , and for all (\bar{x}, d) in $X \times \mathbb{R}^{n_1}$,*

$$(4.3) \quad \begin{aligned} \psi_\alpha(\bullet, \xi)'(\bar{x}; d) &= \min_{y \in M^\alpha(\bar{x}, \xi)} \max_{\lambda \in \Lambda^\alpha(\bar{x}, \xi)} [G(\xi)^\top y - A(\xi)^\top \lambda]^\top d \\ &= \max_{\lambda \in \Lambda^\alpha(\bar{x}, \xi)} [-A(\xi)^\top \lambda]^\top d + \min_{y \in M^\alpha(\bar{x}, \xi)} [G(\xi)^\top y]^\top d \\ &= \bar{\psi}_\alpha(\bar{x}, \bullet, \xi)'(\bar{x}; d) + \bar{\psi}_\alpha(\bullet, \bar{x}, \xi)'(\bar{x}; d). \end{aligned}$$

4.3. Generalized criticality. A dc function is the sum of a convex function and a concave function. It turns out that the recourse function $\psi(\bullet, \xi)$ also has such a convexity-concavity feature, although this is not explicit. Specifically, we say that a function $g : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, where X is a convex set, satisfies the *convexity-concavity* property if there exists a bivariate function $h : X \times X \rightarrow \mathbb{R}$ such that $g(x) = h(x, x)$ for any $x \in X$, $h(\bullet, x)$ is convex, and $h(x, \bullet)$ is concave. To see that the second-stage recourse function $\psi(\bullet, \xi)$ satisfies the convex-concave property for fixed ξ , define a lifted recourse function

$$(4.4) \quad \begin{aligned} \bar{\psi}(x, z, \xi) &\triangleq \underset{y}{\text{minimum}} \quad [f(\xi) + G(\xi)z]^\top y + \frac{1}{2} y^\top Q y \\ &\text{subject to } y \in Y(x, \xi) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\xi)x + Dy \geq b(\xi)\}, \end{aligned}$$

whose optimal solution set is denoted by $\bar{M}(x, z, \xi)$. Clearly, $\psi(x, \xi) = \bar{\psi}(x, x, \xi)$, and $\bar{\psi}(\bullet, \bullet, \xi)$ has the desired convexity-concavity property; by [28, Lemma 1], this

lifted recourse function $\bar{\psi}(\bullet, \bullet, \xi)$ is continuous on its domain of finiteness. The so-defined convexity-concavity property can be thought of as an implicit dc property. As such, we may define a generalized critical point for a univariate function satisfying the convex-concave property.

DEFINITION 4.4. Let $S \subseteq \mathbb{R}^n$ be a convex set, and let a function $g : S \rightarrow \mathbb{R}$ satisfy the convex-concave property with the associated bivariate convex-concave function h . We say that $\bar{x} \in S$ is a generalized critical point of g on S if

$$(4.5) \quad 0 \in \partial_x h(\bar{x}, \bar{x}) - \partial_z(-h)(\bar{x}, \bar{x}) + \mathcal{N}(\bar{x}; S),$$

where $\partial_x h(\bar{x}, \bar{x})$ is the subgradient of the convex function $h(\bullet, \bar{x})$ at \bar{x} , and $\partial_z(-h)(\bar{x}, \bar{x})$ is the subgradient of the convex function $-h(\bar{x}, \bullet)$ at \bar{x} .

The term “generalized critical point” stems from the special case of a dc function $f = g_1 - g_2$ where both g_1 and g_2 are convex. For this function f , we can associate the bivariate function $h(x, z) = g_1(x) - g_2(z)$. With this association, it is clear that the expression (4.5) becomes $0 \in \partial g_1(\bar{x}) - \partial g_2(\bar{x}) + \mathcal{N}(\bar{x}; S)$, which is precisely the definition of a critical point in dc programming. Note that, like the latter, a generalized critical point of the function g depends on the bivariate convex-concave function h .

Specializing Definition 4.4 to the objective function of (2.1), we say that $\bar{x} \in X$ is a *generalized critical point* of this stochastic program if

$$(4.6) \quad 0 \in \partial \varphi(\bar{x}) + \partial_x \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})] - \partial_z \mathbf{E}_{\tilde{\xi}} [-\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})] + \mathcal{N}(\bar{x}; X),$$

where $\partial_x \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})]$ is the subdifferential of the convex function $\mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bullet, \bar{x}, \tilde{\xi})]$ at \bar{x} ; this holds similarly for $\partial_z \mathbf{E}_{\tilde{\xi}} [-\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})]$.

As an intermediate step to relate the new criticality definition to known stationarity concepts, we need the following estimate of the Clarke directional derivative of the value function ψ based on its lifted counterpart $\bar{\psi}$.

LEMMA 4.5. Suppose that assumptions (B) and (C) hold. Then for any $x, \bar{x} \in X$,

$$\mathbf{E}_{\tilde{\xi}} [\psi(\bullet, \tilde{\xi})^0(\bar{x}; x - \bar{x})] \leq \max_{u \in \partial_x \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})]} u^\top (x - \bar{x}) + \max_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}).$$

Proof. Let $L(x, \xi; y, \lambda) \triangleq [f(\xi) + G(\xi)x]^\top y + \frac{1}{2} y^\top Q y + \lambda^\top [b(\xi) - A(\xi)x - Dy]$. By using a proof similar to that for Theorem 2 in [16] which only assumes the closedness and convexity of the feasible set (see also [33, Theorem 3] that assumes, in addition, the tameness of the parametric problem), we can deduce the following upper bound of Clarke’s directional derivative of the value function $\psi(\bullet, \xi)$:

$$\begin{aligned} \psi(\bullet, \xi)^0(\bar{x}; x - \bar{x}) &\leq \max_{y \in M(\bar{x}, \xi)} \max_{\lambda \in \Lambda(\bar{x}, \xi)} \nabla_x L(x, \xi; y, \lambda)^\top (x - \bar{x}) \\ &= \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)^\top \lambda]^\top (x - \bar{x}) + \max_{y \in M(\bar{x}, \xi)} [G(\xi)^\top y]^\top (x - \bar{x}). \end{aligned}$$

It follows from Danskin’s theorem that (see, e.g., [12, Theorem 10.2.1])

$$\partial_x \bar{\psi}(\bar{x}, \bar{x}, \xi) = -A(\xi)^\top \Lambda(\bar{x}, \xi) \quad \text{and} \quad \partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \xi) = -G(\xi)^\top M(\bar{x}, \xi),$$

which yields

$$\begin{cases} \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)^\top \lambda]^\top (x - \bar{x}) = \max_{u \in \partial_x \bar{\psi}(\bar{x}, \bar{x}, \xi)} u^\top (x - \bar{x}), \\ \max_{y \in M(\bar{x}, \xi)} [G(\xi)^\top y]^\top (x - \bar{x}) = \max_{v \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \xi)} v^\top (x - \bar{x}). \end{cases}$$

Therefore, to prove the stated inequality of this lemma, it suffices to show that for any $x, \bar{x} \in X$,

$$(4.7) \quad \begin{cases} \mathbf{E}_{\tilde{\xi}} \left[\max_{u \in \partial_x \bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})} u^\top (x - \bar{x}) \right] = \max_{u \in \partial_x \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})]} u^\top (x - \bar{x}), \\ \mathbf{E}_{\tilde{\xi}} \left[\max_{v \in -\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})} v^\top (x - \bar{x}) \right] = \max_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}). \end{cases}$$

Below we shall prove the second equation; the first can be obtained in a similar manner. Let

$$\bar{v}(\bar{x}, x) \in \operatorname{argmax}_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}).$$

The convexity of $(-\bar{\psi})(\bar{x}, \bullet, \xi)$ implies that $\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})] = \mathbf{E}_{\tilde{\xi}} [\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]$; see [1] and [35, Theorem 7.47]. Then $v(\bar{x}, x, \xi) \in -\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \xi)$ exists such that $\mathbf{E}_{\tilde{\xi}} [v(\bar{x}, x, \tilde{\xi})] = \bar{v}(\bar{x}, x)$. Therefore,

$$(4.8) \quad \begin{aligned} & \max_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}) = \bar{v}(\bar{x}, x)^\top (x - \bar{x}) \\ & = \mathbf{E}_{\tilde{\xi}} [v(\bar{x}, x, \tilde{\xi})^\top (x - \bar{x})] \leq \mathbf{E}_{\tilde{\xi}} \left[\max_{v \in -\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})} v^\top (x - \bar{x}) \right]. \end{aligned}$$

Conversely, it follows from [35, Theorems 7.34 and 7.37] that there exists a measurable selection

$$v(\bar{x}, x, \xi) \in \operatorname{argmax}_{v \in -\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \xi)} v^\top (x - \bar{x}).$$

Then

$$\mathbf{E}_{\tilde{\xi}} [v(\bar{x}, x, \tilde{\xi})] \in -\mathbf{E}_{\tilde{\xi}} [\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})] = -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})],$$

which implies

$$(4.9) \quad \begin{aligned} & \mathbf{E}_{\tilde{\xi}} \left[\max_{v \in -\partial_z (-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})} v^\top (x - \bar{x}) \right] = \mathbf{E}_{\tilde{\xi}} [v(\bar{x}, x, \tilde{\xi})^\top (x - \bar{x})] \\ & = \left(\mathbf{E}_{\tilde{\xi}} [v(\bar{x}, x, \tilde{\xi})] \right)^\top (x - \bar{x}) \leq \max_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}). \end{aligned}$$

The second equation in (4.7) follows by combining the inequalities (4.8) and (4.9). \square

The following proposition relates three sets: S_{gc} of generalized critical points, S_C of Clarke stationary points, and S_d of directional stationary points, all pertaining to the stochastic program (2.1).

PROPOSITION 4.6. *Under assumptions (A), (B), and (C), it holds that*

- (a) $S_d \subseteq S_C \subseteq S_{gc}$;
- (b) *if for almost all $x \in X$, $G(\xi)^\top M(x, \xi)$ is a singleton, then the function $\psi(\bullet, \xi)$ is Clarke regular on X , i.e., $\psi(\bullet, \xi)^0(x; d) = \psi(\bullet, \xi)'(x; d)$ for all $x \in X$ and $d \in \mathbb{R}^{n_1}$. Thus, $S_d = S_C = S_{gc}$.*

Proof. (a) From Definition 4.2, we clearly have $S_d \subseteq S_C$. To prove $S_C \subseteq S_{gc}$, we have from [40, Proposition 2.12] and Lemma 4.5 that for any $\bar{x} \in S_C$ and all $x \in X$,

$$\begin{aligned} 0 &\leq \varphi'(\bar{x}; x - \bar{x}) + \left(\mathbf{E}_{\tilde{\xi}} \left[\psi(\bullet, \tilde{\xi}) \right] \right)^0(\bar{x}; x - \bar{x}) \\ &\leq \varphi'(\bar{x}; x - \bar{x}) + \mathbf{E}_{\tilde{\xi}} \left[\psi(\bullet, \tilde{\xi})^0(\bar{x}; x - \bar{x}) \right] \\ &\leq \varphi'(\bar{x}; x - \bar{x}) + \max_{u \in \partial_x \mathbf{E}_{\tilde{\xi}} [\psi(\bar{x}, \bar{x}, \tilde{\xi})]} u^\top (x - \bar{x}) + \max_{v \in -\partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})]} v^\top (x - \bar{x}). \end{aligned}$$

The above inequality yields $0 \in \partial\varphi(\bar{x}) + \partial_x \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})] - \partial_z \mathbf{E}_{\tilde{\xi}} [(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\xi})] + \mathcal{N}(\bar{x}; X)$. Hence \bar{x} is a generalized critical point, and part (a) follows.

(b) Suppose for almost all $\xi \in \Xi$ that $G(\xi)^\top M(x, \xi)$ is a singleton. By Proposition 4.3, for any $d \in \mathbb{R}^{n_1}$, we have

$$\begin{aligned} \psi(\bullet, \xi)'(\bar{x}; d) &= \min_{y \in M(\bar{x}, \xi)} [G(\xi)d]^\top y + \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)d]^\top \lambda \\ &= [G(\xi)d]^\top y + \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)d]^\top \lambda \quad \text{for any } y \in M(\bar{x}, \xi). \end{aligned}$$

As before, we have

$$\begin{aligned} \psi(\bullet, \xi)^0(\bar{x}; d) &\leq \max_{y \in M(\bar{x}, \xi)} [G(\xi)d]^\top y + \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)d]^\top \lambda \\ &= [G(\xi)d]^\top y + \max_{\lambda \in \Lambda(\bar{x}, \xi)} [-A(\xi)d]^\top \lambda \quad \text{for any } y \in M(\bar{x}, \xi) \\ &= \psi(\bullet, \xi)'(\bar{x}; d). \end{aligned}$$

Thus $\psi(\bullet, \xi)$ is Clarke regular. The last assertion follows by the interchangeability of expectation with the directional derivatives and subdifferentials. \square

5. Convergence analysis. We begin the analysis of the RCS algorithm by making several matrix-theoretic assumptions:

(D) $[Dv \geq 0 \text{ and } Qv = 0] \Rightarrow v = 0$.

(E) For almost all $\xi \in \Xi$,

$$(5.1) \quad [D^\top \lambda = 0 \text{ and } \lambda \geq 0] \Rightarrow A(\xi)^\top \lambda = 0.$$

(F) For almost all $\xi \in \Xi$, $\text{Range } G(\xi) \subseteq \text{Range } Q$.

Conditions (D) and (F) are obviously valid when Q is positive definite. Thus these two conditions are needed only for the case of a positive semidefinite Q . Conditions (D) and (E) are needed to ensure certain uniform boundedness properties of the solutions and multipliers of the regularized QPs in $\psi_\alpha(x, \xi)$ and also of the subgradients of the convex function $\psi_{\alpha,1}(\bullet, \xi)$. There are some simple sufficient conditions for (E) to hold: (i) the Slater condition holds for the sets $Y(x, \xi)$; i.e., there exists $D\bar{y} > 0$; and (ii) $\text{Range } A(\xi) \subseteq \text{Range } D$ for almost all $\xi \in \Xi$. Both sufficient conditions are fairly obvious. Note that the condition (ii) is the counterpart of assumption (F). Specifically, the latter condition pertains to the pair $(G(\xi), Q)$ which appears in the objective function of the recourse function $\psi(x, \xi)$, whereas the former condition pertains to the pair $(A(\xi), D)$ which appears in the constraint of the same recourse function $\psi(x, \xi)$.

5.1. Some lemmas. We present several consequences of assumptions (A)–(E) in the lemmas below.

LEMMA 5.1. *Let $Q \in \mathbb{R}^{n_2 \times n_2}$ be a symmetric positive semidefinite matrix, and let $D \in \mathbb{R}^{\ell \times n_2}$. If assumption (D) holds, then there exist positive constants $\bar{\beta}$ and $\bar{\alpha}$ such that for all $\alpha \in [0, \bar{\alpha}]$ and all pairs of vectors (q, b) for which the QP,*

$$(5.2) \quad \underset{y}{\text{minimize}} \quad q^\top y + \frac{1}{2} y^\top [Q + \alpha \mathbb{I}] y \quad \text{subject to} \quad Dy \geq b$$

has an optimal solution, say $y^\alpha(q, b)$; it holds that $\|y^\alpha(q, b)\| \leq \bar{\beta} \|(q, b)\|$.

Proof. The proof is by contradiction and follows by considering the Karush–Kuhn–Tuckers (KKT) conditions of the QP (5.2) as a linear complementarity problem and using a standard normalization followed by a limiting argument; see [10]. Details are omitted. \square

Our next result occurs under assumption (E) deriving various important bounds. To motivate these bounds, we note that since we need to work with the directional derivative of the convex regularized function $\psi_{\alpha_{\nu,1}}(\bullet, \xi^{\nu,i})$ at the iterate $x^{\nu+1}$, which would involve the optimal solution(s) $\lambda^{\alpha_{\nu,1}}(x^{\nu+1}, \xi)$ of this function in the dual form (cf. (3.2)), we need some boundedness property of such solutions. However, we do not need the dual solutions themselves to be bounded. Instead, as we shall see, the key is to obtain a uniform bound on $A(\xi)^\top \lambda^{\alpha_{\nu,1}}(x^{\nu+1}, \xi)$. This turns out to require a nontrivial argument which employs the following lemma.

LEMMA 5.2. *Let $\Upsilon \triangleq \{A \in \mathbb{R}^{\ell \times n_1} \mid [D^\top \lambda = 0, \lambda \geq 0] \Rightarrow A^\top \lambda = 0\}$. The following two statements hold for this convex family of matrices:*

- (a) *For every $A \in \Upsilon$, the scalar $\gamma(A) \triangleq \max\{\|A^\top \lambda\|_1 \mid \|D^\top \lambda\|_1 = 1, \lambda \geq 0\} < \infty$;*
- (b) *for every compact subset $\hat{\Upsilon}$ of Υ , the scalar $\max_{A \in \hat{\Upsilon}} \gamma(A) < \infty$. Thus for any such subset $\hat{\Upsilon}$, there exists a constant $\hat{\gamma} > 0$ such that for every $A \in \hat{\Upsilon}$, it holds that $\|A^\top \lambda\| \leq \hat{\gamma} \|D^\top \lambda\|$ for all $\lambda \geq 0$.*

Proof. By the property of any matrix $A \in \Upsilon$, the quantity $\|A^\top \lambda\|_1$ is bounded above on the set $\{\lambda \mid \|D^\top \lambda\|_1 = 1, \lambda \geq 0\}$ which is the union of finitely many polyhedra. By linear programming theory, this observation is enough to establish statement (a). It is not difficult to show that the scalar function $\gamma(\bullet)$ is continuous on the set Υ . Thus the finiteness of the scalar $\max_{A \in \hat{\Upsilon}} \gamma(A)$ follows. The last statement in part (b) follows by a simple normalization argument. \square

Recall that $\Lambda^\alpha(x, \xi)$ is the optimal dual solution set of the regularized recourse function $\psi_\alpha(x, \xi)$. For each pair (x, ξ) , let $u^\alpha(x, \xi) \in \partial_x \psi_{\alpha,1}(x, \xi)$ be such that $\mathbf{E}_{\tilde{\xi}}[u^\alpha(x, \tilde{\xi})] \in \partial \mathbf{E}_{\tilde{\xi}}[\psi_{\alpha,1}(x, \tilde{\xi})]$. The subgradient $u^\alpha(x, \xi)$ is a convex combination of finitely many vectors where each is equal to

$$(5.3) \quad \tilde{u}^\alpha(x, \lambda, \xi) \triangleq \left[-A(\xi) + D(Q + \alpha \mathbb{I})^{-1} G(\xi) \right]^\top \lambda,$$

where $\lambda \in \Lambda^\alpha(x, \xi)$. The next lemma gives a bound for $\mathbf{E}_{\tilde{\xi}}[\|u^\alpha(x, \tilde{\xi})\|]$ via such a convex combination.

LEMMA 5.3. *Under assumptions (A), (B), and (C), there exist positive constants θ_i for $i = 1, 2, 3$ such that for all $x \in X$, $\alpha \geq 0$, with Q_α positive definite, and almost all $\xi \in \Xi$ and all $\lambda \in \Lambda^\alpha(x, \xi)$,*

$$(5.4) \quad \max \left(\left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\|, \|D^\top \lambda\| \right) \leq \frac{\theta_1}{\alpha + \rho_{\min}(Q)} + \theta_2 + \theta_3 \alpha.$$

If, in addition, assumption (E) holds, then there exist constants θ'_i for $i = 1, 2, 3$ such that for all $x \in X$,

$$(5.5) \quad \mathbf{E}_{\tilde{\xi}} \left[\left\| u^\alpha(x, \tilde{\xi}) \right\| \right] \leq \left[\mathbf{E}_{\tilde{\xi}} \left[\left\| u^\alpha(x, \tilde{\xi}) \right\|^2 \right] \right]^{1/2} \leq \frac{\theta'_1}{\alpha + \rho_{\min}(Q)} + \theta'_2 + \theta'_3 \alpha.$$

Proof. By Hoffman's error bound applied to the linear inequality systems $Dy \geq r$ defined by the matrix D with variable right-hand side r for which the system is feasible, it follows that there exists a constant $c^D > 0$ dependent on the matrix D only such that for all $x \in X$ and almost all $\xi \in \Xi$, a vector $\hat{y}(x, \xi)$ exists satisfying $A(\xi)x + D\hat{y}(x, \xi) \geq b(\xi)$ and

$$(5.6) \quad \|\hat{y}(x, \xi)\| \leq c^D \| [b(\xi) - A(\xi)x]_+ \|,$$

where $[\bullet]_+ \triangleq \max(\bullet, 0)$ is the plus operator of vectors. For any $\lambda \in \Lambda^\alpha(x, \xi)$, it satisfies the following complementarity conditions:

$$0 \leq \lambda \perp \begin{bmatrix} D(Q + \alpha \mathbb{I})^{-1} D^\top \lambda \\ -[b(\xi) - A(\xi)x + D(Q + \alpha \mathbb{I})^{-1}(f(\xi) + G(\xi)x)] \end{bmatrix} \geq 0.$$

Letting $s(x, \xi) \triangleq A(\xi)x + D\hat{y}(x, \xi) - b(\xi) \geq 0$ be the slack variable associated with the feasible vector $\hat{y}(x, \xi)$, we deduce from the above complementarity conditions that

$$\begin{aligned} & (\alpha + \rho_{\min}(Q)) \left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\|^2 \\ & \leq \left[(Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right]^\top [Q + \alpha \mathbb{I}] \left[(Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right] \\ & = \lambda^\top D(Q + \alpha \mathbb{I})^{-1} D^\top \lambda = \lambda^\top \left[b(\xi) - A(\xi)x + D(Q + \alpha \mathbb{I})^{-1}(f(\xi) + G(\xi)x) \right] \\ & = \lambda^\top [D\hat{y}(x, \xi) - s(x, \xi)] + \left[(Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right]^\top (f(\xi) + G(\xi)x) \\ & \leq \left[(Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right]^\top [(Q + \alpha \mathbb{I})\hat{y}(x, \xi) + (f(\xi) + G(\xi)x)], \end{aligned}$$

which yields

$$\left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\| \leq \frac{1}{\alpha + \rho_{\min}(Q)} \{ \|Q + \alpha \mathbb{I}\| \|\hat{y}(x, \xi)\| + \|f(\xi) + G(\xi)x\| \}.$$

Using the bound (5.6), the inequality $\|D^\top \lambda\| \leq \|Q + \alpha \mathbb{I}\| \|(Q + \alpha \mathbb{I})^{-1} (D^\top \lambda)\|$, and the essential boundedness assumption (C), we easily deduce from the above inequalities the desired bound (5.4) for some constants θ_i , $i = 1, 2, 3$.

To prove the last assertion of the lemma, we note that by the essential boundedness of $\|A(\xi)\|$, there exists a constant $M > 0$ such that $\mathbb{P}[\hat{\Xi}] = 1$, where $\hat{\Xi} \triangleq \{\xi \mid \|A(\xi)\| \leq M\}$. Without loss of generality, we may assume that $A(\xi)$ satisfies the implication (5.1) for all $\xi \in \hat{\Xi}$. Let $\hat{\Upsilon}$ be the convex hull of these matrices $A(\xi)$ for $\xi \in \hat{\Xi}$. Since the family of matrices Υ in Lemma 5.2 is convex, it follows that $\hat{\Upsilon}$ is a compact subset of Υ . Hence, there exists a constant $\hat{\gamma} > 0$ such that for all $\xi \in \hat{\Xi}$, $\|A(\xi)^\top \lambda\| \leq \hat{\gamma} \|D^\top \lambda\|$ for all $\lambda \geq 0$. By the definition (5.3), we obtain, for any $\lambda \in \Lambda^\alpha(x, \xi)$,

$$\|\tilde{u}^\alpha(x, \lambda, \xi)\| \leq \|A(\xi)^\top \lambda\| + \|G(\xi)\| \left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\|.$$

By (5.4) and the bound on $\|A(\xi)^\top \lambda\|$ in terms of $\|D^\top \lambda\|$ obtained in Lemma 5.2, the bound (5.5) follows readily for some constants θ'_i , $i = 1, 2, 3$. \square

An immediate consequence of the bound (5.5) is the Lipschitz continuity of the regularized recourse function $\psi_\alpha(x, \xi)$ with a Lipschitz constant of the order

$$\frac{1}{\alpha + \rho_{\min}(Q)}$$

for $\alpha > 0$ sufficiently small.

COROLLARY 5.4. *There exists a constant $\text{Lip}_\psi > 0$ such that for all $\alpha \geq 0$ sufficiently small with Q_α positive definite, and all x and x' in X ,*

$$(5.7) \quad \left| \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha,1}(x, \tilde{\xi})] - \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha,1}(x', \tilde{\xi})] \right| \leq \frac{\text{Lip}_\psi}{\alpha + \rho_{\min}(Q)} \|x - x'\|.$$

Proof. Since $\psi_{\alpha,1}(\bullet, \xi)$ is convex, we have

$$\psi_{\alpha,1}(x, \xi) - \psi_{\alpha,1}(x', \xi) \geq u^\alpha(x', \xi)^\top (x - x') \geq -\|u^\alpha(x', \xi)\| \|x - x'\|;$$

interchanging x and x' yields

$$|\psi_{\alpha,1}(x, \xi) - \psi_{\alpha,1}(x', \xi)| \leq \sup_{z \in X} \|u^\alpha(z, \xi)\| \|x - x'\|.$$

Thus (5.7) follows from (5.5) because α either equals zero (Q is positive definite) or is positive (Q is positive semidefinite) and sufficiently small. \square

Since $\nabla_x \psi_{\alpha,2}(x, \xi) = G(\xi)^\top (Q + \alpha \mathbb{I})^{-1} [f(\xi) + G(\xi)x]$, we have

$$\|\nabla_x \psi_{\alpha,2}(x, \xi)\| \leq (\alpha_\nu + \rho_{\min}(Q))^{-1} \|G(\xi)\| \|f(\xi) + G(\xi)x\|.$$

Thus without loss of generality, we may take the constant Lip_ψ in the above corollary so that

$$(5.8) \quad \mathbf{E}_{\tilde{\xi}} \left[\left\| \nabla_x \psi_{\alpha,2}(x, \tilde{\xi}) \right\| \right] \leq \frac{\text{Lip}_\psi}{\alpha + \rho_{\min}(Q)} \quad \text{for all } x \in X.$$

We recall the lifted regularized recourse function $\bar{\psi}_\alpha(x, z, \xi)$ defined in (4.2) and the directional derivative formula (4.3) for the regularized recourse function $\psi_\alpha(x, \xi)$. Let $y^\alpha(x, z, \xi)$ be the unique minimizer of the QP associated with $\bar{\psi}_\alpha(x, z, \xi)$ for a given triplet (x, z, ξ) . The following lemma pertains to the limit of this lifted regularized recourse function as $\alpha \downarrow 0$.

LEMMA 5.5. *Under assumptions (A)–(D), if $\{(x^k, z^k)\} \subset X \times X$ is a sequence converging to (\bar{x}, \bar{z}) , and $\{\alpha_k\}$ is a sequence of nonnegative scalars converging to zero with $Q + \alpha_k \mathbb{I}$ positive definite for all k , the following three statements hold:*

(a) *The sequence $\{y^{\alpha_k}(x^k, z^k, \xi)\}$ is bounded, and every accumulation point belongs to $\bar{M}(\bar{x}, \bar{z}, \xi)$, i.e., is an optimal solution of the problem*

$$\underset{y \in Y(\bar{x}, \bar{z})}{\text{minimize}} (f(\xi) + G(\xi)\bar{z})^\top y + \frac{1}{2} y^\top Q y;$$

(b) *the following limits hold uniformly for all $(x, z) \in X \times X$:*

$$(5.9) \quad \lim_{k \rightarrow \infty} \bar{\psi}_{\alpha_k}(x, z, \xi) = \bar{\psi}(x, z, \xi), \quad \lim_{k \rightarrow \infty} \mathbf{E}_{\tilde{\xi}} [\bar{\psi}_{\alpha_k}(x, z, \tilde{\xi})] = \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(x, z, \tilde{\xi})];$$

moreover,

$$(5.10) \quad \lim_{k \rightarrow \infty} \bar{\psi}_{\alpha_k}(x^k, z^k, \xi) = \bar{\psi}(\bar{x}, \bar{z}, \xi), \quad \lim_{k \rightarrow \infty} \mathbf{E}_{\tilde{\xi}} [\bar{\psi}_{\alpha_k}(x^k, z^k, \tilde{\xi})] = \mathbf{E}_{\tilde{\xi}} [\psi(\bar{x}, \bar{z}, \tilde{\xi})];$$

(c) the sequence $\{\mathbf{E}_{\tilde{\xi}}[G(\tilde{\xi})^\top y^{\alpha_k}(x^k, x^k, \tilde{\xi})]\}$ is bounded; moreover, every accumulation point belongs to $-\partial_z \mathbf{E}_{\tilde{\xi}}[-\bar{\psi}(\bar{x}, \bar{x}, \tilde{\xi})]$.

Proof. By Lemma 5.1, the sequence $\{y^k \triangleq y^{\alpha_k}(x^k, z^k, \xi)\}$ is bounded. Without loss of generality, assume that this sequence converges to a limit y^∞ which clearly belongs to $Y(\bar{x}, \xi)$. To show that y^∞ has the claimed minimizing property, consider the following KKT conditions: for some λ^k ,

$$\begin{aligned} f(\xi) + G(\xi)z^k + [Q + \alpha_k \mathbb{I}]y^k - D^\top \lambda^k &= 0, \\ 0 \leq \lambda^k \perp A(\xi)x^k + Dy^k - b(\xi) &\geq 0. \end{aligned}$$

Passing to the limit $k \rightarrow \infty$, we easily deduce that $y^\infty \in \bar{M}(\bar{x}, \bar{z}, \xi)$.

For (b), it suffices to prove the two limits in (5.9); those in (5.10) follow from the uniform convergence of the latter limits and the fact that for fixed ξ , the function $\bar{\psi}(\bullet, \bullet, \xi)$ is continuous on $X \times X$. For any $\bar{y} \in \bar{M}(x, z, \xi)$, which we can choose by Lemma 5.1 to be uniformly bounded for all $(x, z) \in X \times X$ and almost all $\xi \in \Xi$, we have

$$\begin{aligned} \bar{\psi}_{\alpha_k}(x, z, \xi) &= [f(\xi) + G(\xi)z]^\top y^{\alpha_k}(x, z, \xi) + \frac{1}{2} (y^{\alpha_k}(x, z, \xi))^\top [Q + \alpha_k \mathbb{I}] y^{\alpha_k}(x, z, \xi) \\ &\leq [f(\xi) + G(\xi)z]^\top \bar{y} + \frac{1}{2} \bar{y}^\top [Q + \alpha_k \mathbb{I}] \bar{y} = \bar{\psi}(x, z, \xi) + \frac{\alpha_k}{2} \|\bar{y}\|^2. \end{aligned}$$

Conversely, we have

$$\begin{aligned} \bar{\psi}(x, z, \xi) &= [f(\xi) + G(\xi)z]^\top \bar{y} + \frac{1}{2} \bar{y}^\top Q \bar{y} \\ &\leq [f(\xi) + G(\xi)z]^\top y^{\alpha_k}(x, z, \xi) + \frac{1}{2} y^{\alpha_k}(x, z, \xi)^\top Q y^{\alpha_k}(x, z, \xi) \leq \bar{\psi}_{\alpha_k}(x, z, \xi). \end{aligned}$$

Consequently, the first limit in (5.9) holds uniformly in (x, z) . The second limit follows by the dominated convergence theorem as $\{\bar{\psi}_{\alpha_k}(x, z, \xi)\}$ is uniformly bounded by a constant independent of almost all $\xi \in \Xi$.

To prove (c), for every $\alpha_k > 0$, letting

$$a^k \triangleq \mathbf{E}_{\tilde{\xi}} [G(\tilde{\xi})^\top y^{\alpha_k}(x^k, x^k, \tilde{\xi})] = -\nabla_z \mathbf{E}_{\tilde{\xi}} [-\bar{\psi}_{\alpha_k}(x^k, x^k, \tilde{\xi})],$$

we have, for any z ,

$$\mathbf{E}_{\tilde{\xi}} [-\bar{\psi}_{\alpha_k}(x^k, z, \tilde{\xi})] \geq \mathbf{E}_{\tilde{\xi}} [-\bar{\psi}_{\alpha_k}(x^k, x^k, \tilde{\xi})] - (a^k)^\top (z - x^k).$$

The sequence $\{a^k\}$ is bounded; if \bar{a} is the limit of a convergent subsequence $\{a^k\}_{k \in \kappa}$, then passing to the limit $k(\in \kappa) \rightarrow \infty$, the above inequality yields, using the second limit in (5.10),

$$\mathbf{E}_{\tilde{\xi}} [-\psi(\bar{x}, z, \tilde{\xi})] \geq \mathbf{E}_{\tilde{\xi}} [-\psi(\bar{x}, \bar{x}, \tilde{\xi})] - \bar{a}^\top (z - \bar{x}).$$

Hence $-\bar{a}$ is a subgradient of the convex function $\mathbf{E}_{\tilde{\xi}} [-\psi(\bar{x}, \bullet, \tilde{\xi})]$ at \bar{x} . \square

5.2. Convergence theorems. Before formally stating the convergence results of the RCS algorithm and its two variations, we need to set down the precise notion of convergence, i.e., the measure-theoretic setting of the convergence. This is needed due to the incremental samples in the algorithm. Adopting the setup in [17], we let Ω^{L_k} denote the L_k -fold Cartesian product of the sample space Ω ; let \mathbb{P}_ν be a probability measure on $\hat{\Omega}^\nu \triangleq \prod_{k=0}^\nu \Omega^{L_k}$, and let \mathbf{E}_ν be the expectation operator induced by \mathbb{P}_ν . Let \mathcal{F}^ν be the σ -algebra generated by the tuple of random variables $(\{\tilde{\xi}^{0,i}\}_{i=1}^{L_0}, \dots, \{\tilde{\xi}^{\nu,i}\}_{i=1}^{L_\nu})$, so that the family $\{\mathcal{F}^\nu\}$ is a filtration on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\hat{\Omega}^\infty \triangleq \prod_{k=0}^\infty \Omega^{L_k}$, and let \mathbb{P}_∞ denote the corresponding probability distribution on $\hat{\Omega}^\infty$. Let \mathbf{E}_∞ be the expectation operator induced by \mathbb{P}_∞ . Note that if \mathcal{Z} is a random variable dependent only on events in the iterations up to ν , i.e., defined on the σ -algebra \mathcal{F}^ν , then $\mathbf{E}_\infty[\mathcal{Z}] = \mathbf{E}_\nu[\mathcal{Z}]$. Finally, let $\mathbf{E}_\nu[\bullet \mid \mathcal{F}^{\nu-1}]$ denote the conditional expectation on the probability space $(\hat{\Omega}^\nu, \mathcal{F}^\nu, \mathbb{P}_\nu)$ given the σ -algebra $\mathcal{F}^{\nu-1}$.

In the convergence theorem for the main RCS algorithm, assumptions (A)–(E) are needed, together with conditions on the sequences of sample sizes $\{L_\nu\}_{\nu=0}^\infty$ and regularization parameters $\{\alpha_\nu\}_{\nu=0}^\infty$. The conclusion is subsequential convergence to a generalized critical point, which becomes a directional stationary point under the further assumption (F). For the modified sequence $\{\hat{x}^\nu\}$, where both convexification and sampling are left out, and $\{\tilde{x}^\nu\}$, where only sampling is left out, only assumptions (A)–(D) are needed along with a condition on $\{\alpha_\nu\}_{\nu=0}^\infty$. In the case where Q is positive definite, we can take $\alpha_\nu = 0$ for all ν .

THEOREM 5.6 (the RCS algorithm). *Under assumptions (A)–(E), let $\{\alpha_\nu\}$ be a nonincreasing sequence of nonnegative scalars satisfying $\{\alpha_\nu\} \downarrow 0$, let Q_{α_ν} be positive definite for all ν , and let*

$$(5.11) \quad \liminf_{\nu \rightarrow \infty} (\alpha_\nu + \rho_{\min}(Q)) \sqrt{\nu} > 0.$$

Let $\{L_\nu\}$ satisfy

$$(5.12) \quad \sum_{\nu=0}^{\infty} \frac{1}{(\alpha_\nu + \rho_{\min}(Q))^2 \sqrt{L_\nu}} < \infty.$$

Then every accumulation point of the sequence $\{x^\nu\}$ produced by the RCS algorithm is a generalized critical point of the stochastic program (2.1), with \mathbb{P}_∞ -probability 1. If, in addition, (F) holds, then such a point is a d -stationary point or, equivalently, a C -stationary point.

In the two simplified versions of the RCS algorithm, the two sequences $\{\hat{x}^\nu\}$ and $\{\tilde{x}^\nu\}$ are taken to be deterministic sequences. Thus no “probability 1” requirement is attached to the conclusions. Proofs of these two simplified cases are omitted.

THEOREM 5.7 (regularization only). *Under assumptions (A)–(D), let $\{\alpha_\nu\} \downarrow 0$ be a nonincreasing sequence of nonnegative scalars satisfying (5.11) and such that Q_{α_ν} is positive definite for all ν . Then every accumulation point of the sequence $\{\hat{x}^\nu\}$ is a generalized critical point of the stochastic program (2.1). If, in addition, (F) holds, then such a point is a d -stationary point or, equivalently, a C -stationary point.*

THEOREM 5.8 (regularization + convexification). *Under assumptions (A)–(D), let $\{\alpha_\nu\} \downarrow 0$ be a nonincreasing sequence of nonnegative scalars satisfying (5.11) such that Q_{α_ν} is positive definite for all ν . The conclusions of Theorem 5.7 hold for the sequence $\{\tilde{x}^\nu\}$.*

5.3. Proof of Theorem 5.6. In the RCS algorithm, each subproblem uses two types of approximations for the original objective function: one is the linear approximation of the concave summand, and the other is the sample average approximation for both expectations. Based on this observation, we analyze the connection of two consecutive iterates $x^{\nu+1}$ and x^ν in several steps by constructing an intermediate iterate $\tilde{x}^{\nu+1/2}$ by using only the linear approximation of the concave summand while maintaining the following expectation functionals:

$$(5.13) \quad \tilde{x}^{\nu+1/2} \triangleq \operatorname{argmin}_{x \in X} \varphi(x) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_{\nu,1}}(x, \tilde{\xi}) - \hat{\psi}_{\alpha_{\nu,2}}(x, \tilde{\xi}; x^\nu) \right] + \frac{1}{2\gamma} \|x - x^\nu\|^2.$$

The rest of the proof is organized as follows.

1. Using the definition of $\tilde{x}^{\nu+1/2}$, we give the relation of two iterates, x^ν and $\tilde{x}^{\nu+1/2}$.

2. We relate $\tilde{x}^{\nu+1/2}$ with its SAA approximation $x^{\nu+1}$.

3. Combining the above two relations, we derive an (inexact) descent inequality of the objective values $\zeta_{\alpha_\nu}(x^{\nu+1})$ with the error $\zeta_{\alpha_\nu}(x^{\nu+1}) - \zeta_{\alpha_\nu}(x^\nu)$ dependent on the sample sizes. By [4, Lemma 5.31], we deduce that the sequence of objective values $\{\zeta_{\alpha_\nu}(x^{\nu+1})\}$ converges and also $\sum_{\nu=0}^{\infty} \|x^{\nu+1} - x^\nu\|^2$ is finite with \mathbb{P}_∞ -probability 1.

4. Finally, we prove the desired convergence asserted by the theorem by analyzing the limit property of the optimality condition in the update (3.5).

The following derivations present the details of the above steps.

Step 1. By [4, section 27.1], we can derive

$$(5.14) \quad \begin{aligned} & \left(\varphi(\tilde{x}^{\nu+1/2}) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \tilde{\xi}) \right] \right) - \left(\varphi(x^\nu) + \mathbf{E}_{\tilde{\xi}} \left[\psi_{\alpha_{\nu,1}}(x^\nu, \tilde{\xi}) \right] \right) \\ & \leq -\frac{1}{\gamma} \left(\tilde{x}^{\nu+1/2} - x^\nu \right)^\top \left(\tilde{x}^{\nu+1/2} - x^\nu - \gamma \mathbf{E}_{\tilde{\xi}} \left[\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) \right] \right) \\ & = -\frac{1}{\gamma} \left\| \tilde{x}^{\nu+1/2} - x^\nu \right\|^2 + \left(\tilde{x}^{\nu+1/2} - x^\nu \right)^\top \mathbf{E}_{\tilde{\xi}} \left[\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) \right]. \end{aligned}$$

Step 2. For the second relation, given the point x^ν and the samples $\{\{\xi^{k,i}\}_{i=1}^{L_k}\}_{k=0}^{\nu-1}$ up to iteration $\nu-1$, $x^{\nu+1}$ is an SAA approximation of $\tilde{x}^{\nu+1/2}$ using the samples $\{\xi^{\nu,i}\}_{i=1}^{L_\nu}$ at iteration ν . By the respective optimalities of these two iterates, $x^{\nu+1}$ and $\tilde{x}^{\nu+1/2}$, there exist $\tilde{a}^{\nu+1/2} \in \partial\varphi(\tilde{x}^{\nu+1/2})$ and $v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \in \partial\mathbf{E}_{\tilde{\xi}}[\psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \tilde{\xi})]$ so that by using the interchangeability of gradient and expectation, we have

$$(x^{\nu+1} - \tilde{x}^{\nu+1/2})^\top \left[\tilde{a}^{\nu+1/2} + v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) - \mathbf{E}_{\tilde{\xi}} \left[\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) \right] + \frac{1}{\gamma} (\tilde{x}^{\nu+1/2} - x^\nu) \right] \geq 0,$$

and for some $a^{\nu+1} \in \partial\varphi(x^{\nu+1})$ and $u^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i}) \in \partial_x \psi_{\alpha_{\nu,1}}(x^{\nu+1}, \xi^{\nu,i})$,

$$(\tilde{x}^{\nu+1/2} - x^{\nu+1})^\top \left[a^{\nu+1} + \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i}) - \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \xi^{\nu,i}) + \frac{1}{\gamma} (x^{\nu+1} - x^\nu) \right] \geq 0.$$

Let $u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi) \in \partial_x \psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \xi)$ satisfy $\mathbf{E}_{\tilde{\xi}}[u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\xi})] = v^{\alpha_\nu}(\tilde{x}^{\nu+1/2})$. It follows from the monotonicity of $\partial\varphi$ and $\partial_x \psi_{\alpha_{\nu,1}}(\bullet, \xi)$ that

$$\begin{cases} (x^{\nu+1} - \tilde{x}^{\nu+1/2})^\top (a^{\nu+1} - \tilde{a}^{\nu+1/2}) \geq 0, \\ \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} (x^{\nu+1} - \tilde{x}^{\nu+1/2})^\top [u^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i}) - u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi^{\nu,i})] \geq 0. \end{cases}$$

Adding the above four inequalities together, we deduce that

$$\begin{aligned} 0 \leq & (\tilde{x}^{\nu+1/2} - x^{\nu+1})^\top \left[\frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi^{\nu,i}) - \mathbf{E}_{\tilde{\xi}}[u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\xi})] \right. \\ & \left. - \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \xi^{\nu,i}) + \mathbf{E}_{\tilde{\xi}}[\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi})] \right] \\ & + \frac{1}{\gamma} (\tilde{x}^{\nu+1/2} - x^{\nu+1})^\top (x^{\nu+1} - \tilde{x}^{\nu+1/2}). \end{aligned}$$

Hence, with

$$\begin{cases} \tilde{B}^\nu \triangleq \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi^{\nu,i}) - \mathbf{E}_{\tilde{\xi}}[u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\xi})], \\ \tilde{C}^\nu \triangleq \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \xi^{\nu,i}) - \mathbf{E}_{\tilde{\xi}}[\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi})], \end{cases}$$

we deduce that $\|\tilde{x}^{\nu+1/2} - x^{\nu+1}\| \leq \gamma[\|\tilde{B}^\nu\| + \|\tilde{C}^\nu\|]$. In what follows, we derive bounds for $\mathbf{E}_\nu[\|\tilde{B}^\nu\|]$ and $\mathbf{E}_\nu[\|\tilde{C}^\nu\|]$. There exists a constant $\tilde{V}_1 > 0$ such that for all ν with α_ν sufficiently small,

$$\begin{aligned} \mathbf{E}_\nu[\|\tilde{B}^\nu\|] &= \mathbf{E}_{\nu-1} \left[\mathbf{E}_\nu \left[\left\| \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi^{\nu,i}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\| \mid \mathcal{F}^{\nu-1} \right] \right] \\ &\leq \mathbf{E}_{\nu-1} \left[\mathbf{E}_\nu \left[\left\| \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \xi^{\nu,i}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\ &= \mathbf{E}_{\nu-1} \left[\frac{1}{L_\nu} \mathbf{E}_{\tilde{\xi}} \left[\left\| u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\xi}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\ &= \mathbf{E}_{\nu-1} \left[\frac{1}{L_\nu} \left[\mathbf{E}_{\tilde{\xi}} \left[\left\| u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\xi}) \right\|^2 \right] - \left\| v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\ &\leq \frac{\tilde{V}_1}{(\alpha_\nu + \rho_{\min}(Q)) L_\nu^{1/2}} \quad \text{by Lemma 5.3.} \end{aligned}$$

Similarly, (5.8) yields the existence of a constant $\tilde{V}_2 > 0$ such that

$$\begin{aligned} & \mathbf{E}_\nu \left[\left\| \tilde{C}^\nu \right\| \right] \\ & \leq \mathbf{E}_{\nu-1} \left[\frac{1}{L_\nu} \mathbf{E}_{\tilde{\xi}} \left[\left\| \nabla_x \psi_{\alpha_{\nu,2}}(\tilde{x}^{\nu+1/2}, \tilde{\xi}) - \mathbf{E}_{\tilde{\xi}} [\nabla_x \psi_{\alpha_{\nu,2}}(\tilde{x}^{\nu+1/2}, \tilde{\xi})] \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\ & \leq \frac{\tilde{V}_2}{(\alpha_\nu + \rho_{\min}(Q)) L_\nu^{1/2}}. \end{aligned}$$

Recalling the Lipschitz constant Lip_φ from (2.3), Lip_ψ from (5.7) in Corollary 5.4, and also (5.8), we have

$$\begin{aligned} & \hat{\zeta}_{\alpha_\nu}(x^{\nu+1}; x^\nu) - \hat{\zeta}_{\alpha_\nu}(\tilde{x}^{\nu+1/2}; x^\nu) + \frac{1}{2\gamma} \|x^{\nu+1} - x^\nu\|^2 - \frac{1}{2\gamma} \|\tilde{x}^{\nu+1/2} - x^\nu\|^2 \\ & = \varphi(x^{\nu+1}) + \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,1}}(x^{\nu+1}, \tilde{\xi})] - \varphi(\tilde{x}^{\nu+1/2}) - \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \tilde{\xi})] \\ & \quad - \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) + \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi})^\top (x^{\nu+1} - x^\nu)] \\ & \quad + \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) + \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi})^\top (\tilde{x}^{\nu+1/2} - x^\nu)] \\ & \quad + \frac{1}{2\gamma} (x^{\nu+1} - \tilde{x}^{\nu+1/2})^\top (x^{\nu+1} + \tilde{x}^{\nu+1/2} - 2x^\nu) \\ & \leq \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) \|x^{\nu+1} - \tilde{x}^{\nu+1/2}\| \quad \text{for some constant } \tilde{\Upsilon} > 0 \\ & \leq \gamma \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) [\|\tilde{B}^\nu\| + \|\tilde{C}^\nu\|]. \end{aligned}$$

Step 3. By (3.4), we have

$$\begin{aligned} & \hat{\zeta}_{\alpha_\nu}(\tilde{x}^{\nu+1/2}; x^\nu) = \varphi(\tilde{x}^{\nu+1/2}) \\ & \quad + \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \tilde{\xi})] - \mathbf{E}_{\tilde{\xi}} [\psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi}) + \nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\xi})^\top (\tilde{x}^{\nu+1/2} - x^\nu)]. \end{aligned}$$

Using the inequalities $\zeta_{\alpha_{\nu+1}}(x^{\nu+1}) \leq \zeta_{\alpha_\nu}(x^{\nu+1}) \leq \hat{\zeta}_{\alpha_\nu}(\tilde{x}^{\nu+1}; x^\nu)$ and combining those in Steps 1 and 2, we may deduce that

$$\begin{aligned} \zeta_{\alpha_{\nu+1}}(x^{\nu+1}) - \zeta_{\alpha_\nu}(x^\nu) & \leq \gamma \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) [\|\tilde{B}^\nu\| + \|\tilde{C}^\nu\|] \\ & \quad - \frac{1}{2\gamma} \|x^{\nu+1} - x^\nu\|^2 - \frac{1}{2\gamma} \|\tilde{x}^{\nu+1/2} - x^\nu\|^2. \end{aligned}$$

By the bounds of $\mathbf{E}_\nu [\|\tilde{B}^\nu\|]$ and $\mathbf{E}_\nu [\|\tilde{C}^\nu\|]$ in Step 2, we have

$$\begin{aligned} & \sum_{\nu=0}^{\infty} \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) \mathbf{E}_\nu [\|\tilde{B}^\nu\| + \|\tilde{C}^\nu\|] \\ & \leq \sum_{\nu=0}^{\infty} \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) \left[\frac{\tilde{V}_1}{(\alpha_\nu + \rho_{\min}(Q)) L_\nu^{1/2}} + \frac{\tilde{V}_2}{(\alpha_\nu + \rho_{\min}(Q)) L_\nu^{1/2}} \right] \end{aligned}$$

with the right-hand sum finite by assumption. Hence we have that with probability 1,

$$\sum_{\nu=0}^{\infty} \left(\text{Lip}_{\varphi} + \frac{\text{Lip}_{\psi}}{\alpha_{\nu} + \rho_{\min}(Q)} + \tilde{\Upsilon} \right) \left[\|\tilde{B}^{\nu}\| + \|\tilde{C}^{\nu}\| \right]$$

is finite. By [4, Lemma 5.31], it follows that the sequence $\{\zeta_{\alpha_{\nu}}(x^{\nu})\}$ converges. Moreover, the two sums

$$\sum_{\nu=0}^{\infty} \left\| \tilde{x}^{\nu+1/2} - x^{\nu} \right\|^2 \quad \text{and} \quad \sum_{\nu=0}^{\infty} \left\| x^{\nu+1} - x^{\nu} \right\|^2$$

are finite with probability 1.

Step 4. Writing out the variational condition of the optimality of $x^{\nu+1}$ for the problem (3.5), we can readily show that $x^{\nu+1}$ is optimal for

$$\underset{x \in X}{\text{minimize}} \left[\begin{aligned} & \varphi(x) + \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \left[G(\xi^{\nu,i})^{\top} y^{\alpha_{\nu}}(x^{\nu+1}, \xi^{\nu,i}) \right]^{\top} (x - x^{\nu+1}) \\ & + \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \bar{\psi}_{\alpha_{\nu}}(x, x^{\nu+1}, \xi^{\nu,i}) + \frac{1}{2\gamma} \|x - x^{\nu}\|^2 \\ & + \left\{ \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \left[\nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu+1}, \xi^{\nu,i}) - \nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu}, \xi^{\nu,i}) \right] \right\}^{\top} (x - x^{\nu+1}) \end{aligned} \right].$$

Let \tilde{x}^{∞} be the limit of a convergence subsequence $\{x^{\nu+1}\}_{\nu \in \kappa}$. To complete the proof, it remains to show the following limits:

- every limit point of the sequence $\left\{ \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \left[G(\xi^{\nu,i})^{\top} y^{\alpha_{\nu}}(x^{\nu+1}, \xi^{\nu,i}) \right] \right\}_{\nu \in \kappa}$, one of which must exist, belongs to $-\partial_z \mathbf{E}_{\tilde{\xi}}[-\bar{\psi}(\tilde{x}^{\infty}, \tilde{x}^{\infty}, \tilde{\xi})]$;
- the sequence $\left\{ \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \bar{\psi}_{\alpha_{\nu}}(x, x^{\nu+1}, \xi^{\nu,i}) \right\}$ converges uniformly to $\mathbf{E}_{\tilde{\xi}}[\bar{\psi}(x, \tilde{x}^{\infty}, \tilde{\xi})]$ for $x \in X$;
- $\lim_{\nu(\in \kappa) \rightarrow \infty} \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \left[\nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu+1}, \xi^{\nu,i}) - \nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu}, \xi^{\nu,i}) \right] = 0$.

All these limits can be proved by invoking Lemma 5.5 and some suitable bounds of the respective summands in the above limits. Specifically, the convergence of the second sequence follows easily from (5.9); that of the third sequence follows from the identity

$$\nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu+1}, \xi) - \nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu}, \xi) = G(\xi)^{\top} [Q + \alpha_{\nu} \mathbb{I}]^{-1} G(\xi) [x^{\nu+1} - x^{\nu}]$$

that yields the bound

$$\left\| \nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu+1}, \xi) - \nabla_x \psi_{\alpha_{\nu,2}}(x^{\nu}, \xi) \right\| \leq \frac{\text{constant}}{\alpha_{\nu} + \rho_{\min}(Q)} \|x^{\nu+1} - x^{\nu}\|.$$

Since the sum $\sum_{\nu=0}^{\infty} \|x^{\nu} - x^{\nu+1}\|^2 < \infty$, we must have $\lim_{\nu \rightarrow \infty} \sqrt{\nu} \|x^{\nu+1} - x^{\nu}\| = 0$. Consequently, if (5.11) is assumed, then the third limit holds. Finally, for the convergence of the first sequence, we first note the following two probability 1 limits:

$$(5.15) \quad \lim_{\nu(\in \kappa) \rightarrow \infty} \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \bar{\psi}_{\alpha_{\nu}}(x^{\nu+1}, x^{\nu+1}, \xi^{\nu,i}) = \mathbf{E}_{\tilde{\xi}} \left[\bar{\psi}(\tilde{x}^{\infty}, \tilde{x}^{\infty}, \tilde{\xi}) \right],$$

and for all $x \in X$,

$$(5.16) \quad \lim_{\nu(\in \kappa) \rightarrow \infty} \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x, \xi^{\nu,i}) = \mathbf{E}_{\tilde{\xi}} \left[\bar{\psi}(\tilde{x}^\infty, x, \tilde{\xi}) \right].$$

Since $\bar{\psi}_{\alpha_\nu}(x^{\nu+1}, \bullet, \xi^{\nu,i})$ is concave differentiable with gradient $\nabla_z \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x^{\nu+1}, \xi^{\nu,i})$ equal to $G(\xi^{\nu,i})^\top y^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i})$, we have

$$\begin{aligned} \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x, \xi^{\nu,i}) &\leq \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x^{\nu+1}, \xi^{\nu,i}) \\ &\quad + \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} [G(\xi^{\nu,i})^\top y^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i})]^\top (x - x^{\nu+1}). \end{aligned}$$

For the subsequence of $\{\frac{1}{L_\nu} \sum_{i=1}^{L_\nu} [G(\xi^{\nu,i})^\top y^{\alpha_\nu}(x^{\nu+1}, \xi^{\nu,i})]\}_{\nu \in \kappa}$, let a^∞ be an accumulation point. Then by the limits (5.15) and (5.16), we deduce

$$\mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\tilde{x}^\infty, x, \tilde{\xi})] \leq \mathbf{E}_{\tilde{\xi}} [\bar{\psi}(\tilde{x}^\infty, \tilde{x}^\infty, \tilde{\xi})] + (a^\infty)^\top (x - \tilde{x}^\infty) \quad \text{with probability 1.}$$

Hence $-a^\infty \in \partial_z \mathbf{E}_{\tilde{\xi}} [-\bar{\psi}(\tilde{x}^\infty, \tilde{x}^\infty, \tilde{\xi})]$ with probability 1. \square

6. Numerical experiments. In this section, we report the results of numerical experiments which test the effectiveness of the RCS algorithm by solving the joint production, pricing, and shipment problem presented in subsection 2.1 with the number of factories $M = 5$ and the number of stores $N = 5$. All the computations were conducted in Python on Mac OS X with 2.7 GHz Intel Core i5 8GB RAM. The parameters of the problem are as follows: the first-stage production costs per unit $\mathbf{c}_1 = (2, 3, 3, 4, 2)^\top$, the last-minute production costs per unit $\mathbf{c}_2 = (2.2, 3.2, 3.3, 4.2, 2.4)^\top$, and the shipment cost per unit from the factory i to the store j , $s_{i,j} = 2$ for all $i = 1, \dots, 5$ and $j = 1, \dots, 5$. Each scenario of the random slope α in the linear demand function at each store is generated from truncated normal distributions on the closed intervals $[-1.5, -0.5]$, $[-2, -1]$, $[-2.5, -1.5]$, $[-3, -2]$, and $[-2.5, -1.5]$, respectively; the scenarios of the random intercepts are generated from truncated normal distributions on $[15, 17]$, $[20, 22]$, $[25, 27]$, $[30, 32]$, and $[25, 27]$. The price variable is constrained in a closed interval $[1, 10]$. With a finite scenario set $\{\xi^1, \dots, \xi^K\}$, the original problem (2.4)–(2.5) is equivalent to the following optimization problem in the price variable alone:

$$(6.1) \quad \underset{1 \leq p \leq 10}{\text{minimize}} \quad v(p), \quad \text{where } v(p) \text{ is the value function of a linear program:}$$

$$\begin{aligned} v(p) &\triangleq \underset{\mathbf{x}, \{y^k\}, \{z^k\} \geq 0}{\text{minimize}} \quad \mathbf{c}_1^\top \mathbf{x} + \frac{1}{K} \sum_{k=1}^K \left(\sum_{i=1}^M c_{2i} y_i^k + \sum_{i=1}^M \sum_{j=1}^N (s_{ij} - p) z_{ij}^k \right) \\ &\text{subject to} \quad \sum_{i=1}^M z_{ij}^k \leq \alpha_j(\xi^k) p + \beta_j(\xi^k), \quad j = 1, \dots, N, \quad k = 1, \dots, K \\ &\text{and} \quad \sum_{j=1}^N z_{ij}^k \leq x_i + y_i^k, \quad i = 1, \dots, M, \quad k = 1, \dots, K. \end{aligned}$$

In Figure 1, we compute values of $v(p)$ at price values evenly discretized on $[0, 10]$. From the figure, we can deduce that $v(p)$ is a nonconvex function, yet we are able to

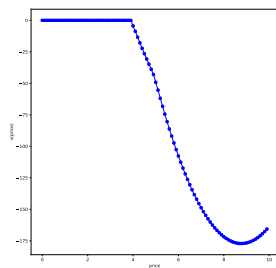


FIG. 1. Value function of the price.

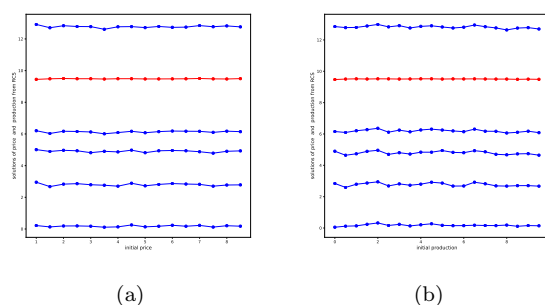


FIG. 2. Solutions of (6.1) with different initializations.

determine an approximately optimal price by a discrete search, which is possible in this case of a scalar price variable. However, when the price variable is a vector associated with multiple products, obtaining the optimal price solution is not straightforward.

The RCS algorithm iteratively generates the iid samples of size L_ν from the scenario set. At each iteration ν , each subproblem is a convex quadratic program with $(M + L_\nu * M + L_\nu * M * N + 1)$ decision variables and $L_\nu * (M + N)$ constraints (excluding the bound constraints); it is solved by the default setting in CPLEX using the barrier optimizer. In Figure 2, under the setting of $L_\nu = \lceil \nu^\tau \rceil$ where $\tau = 2.3$ and the regularization parameter $\alpha_\nu = \nu^{-\kappa}$ where $\kappa = 0.05$, the solutions of price and production provided by the RCS algorithm after 15 iterations are provided under the different initial price solutions and production solutions. This implies that the limit points of the sequence may not be affected by the initial points.

We compare the performance of the RCS algorithm for six configurations of $\tau \in \{2.3, 2.5\}$ and $\kappa \in \{0.05, 0.3, 0.5\}$ in Figure 3 with the same initial iterate, the same scenario set of size 1000, and the same proximal parameter $1/\gamma = 0.1$ by running 30 replications for each configuration and 10 iterations for each replication. The condition (5.11) in Theorem 5.6 is satisfied with these choices of κ . However, the condition (5.12) in the theorem which involves $\{L_\nu\}$ and $\{\alpha_\nu\}$ together is not satisfied under several combinations of (τ, κ) . This is specified in Table 1 in which F represents that the condition (5.12) is not satisfied, and T means it is satisfied.

Observing that the variance of the solution value is decreasing to zero as the iteration number increases, we can deduce that the solution sequence might have the

TABLE 1

τ	2.3	2.3	2.3	2.5	2.5	2.5
κ	0.05	0.3	0.5	0.05	0.3	0.5
condition (5.12)	T	F	F	T	F	F

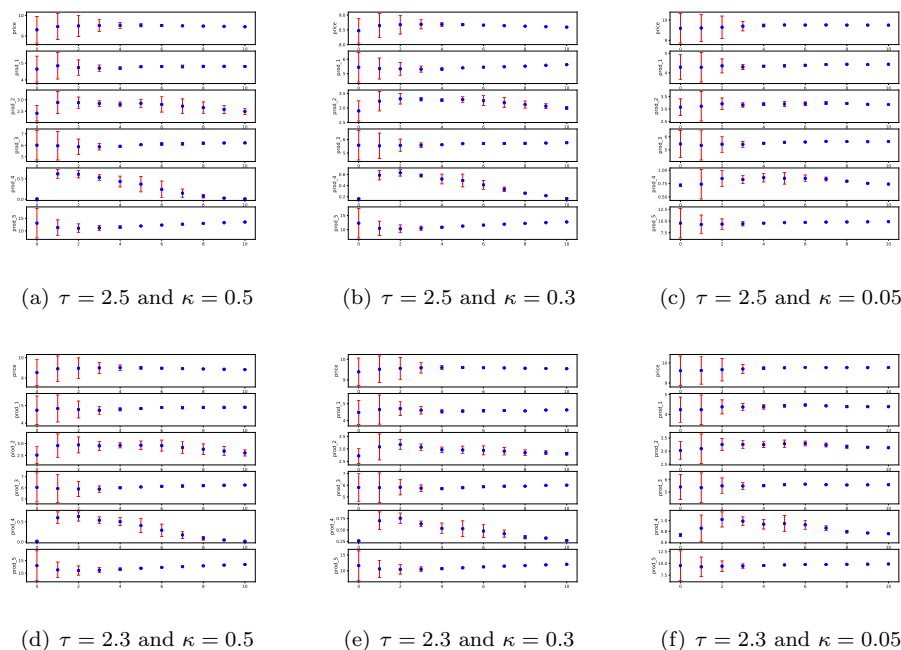


FIG. 3. Performance of the RCS algorithm with six types of parameter configuration.

same limit points as long as the initial solutions are the same. Figure 3 also shows that the variance of the solution of production quantities, under which the theoretical condition (5.12) is not satisfied, is larger than the variance of those in settings of (τ, κ) satisfying the condition (5.12) as the iteration number increases.

In Figure 4, we terminated the RCS algorithm when the distance between the current solution and the previous solution was below 0.05. We compare the number of iterations of the RCS algorithm for problems with different proximal parameters $1/\gamma = 0.01, 0.1, 0.5$, respectively. As we can see from the figure, the limit points of the solution sequence could be different by setting different values of γ . We can also deduce that the RCS algorithm converges faster with a smaller $1/\gamma$ value.

7. Conclusions. We have studied a linearly bi-parameterized stochastic program with convex quadratic recourse, and have established the almost sure sub-sequential convergence of a combined RCS algorithmic framework for computing a generalized critical point. Such a point becomes a directional stationary point under a regularity condition on the second-stage recourse function. The newly introduced concepts of implicit convex-concave functions and generalized critical points enrich the foundation of modern nonconvex nondifferentiable optimization problems. We anticipate increased application of these concepts in other deterministic and stochastic programs. The work done in this paper on this class of linearly bi-parametrized

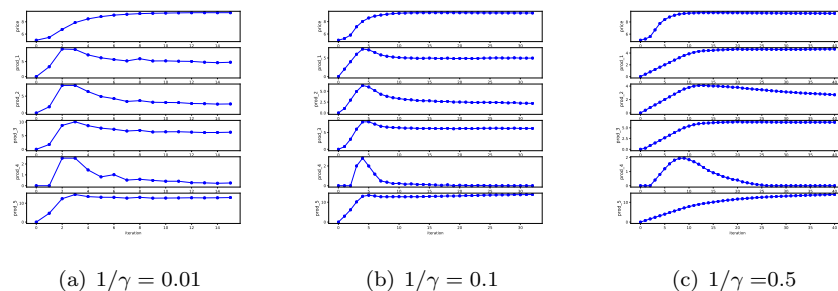


FIG. 4. Performance of the RCS algorithm with different proximal parameters γ .

stochastic programs is by no means complete. The sequential convergence and the convergence rate of the proposed algorithm, as well as the efficient computation of the sample-based RCS subproblems, are waiting to be further explored.

REFERENCES

- [1] R. J. AUMANN, *Integrals of set-valued functions*, J. Math. Anal. Appl., 12 (1965), pp. 1–12.
- [2] L. T. H. AN AND P. D. TAO, *The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems*, Ann. Oper. Res., 133 (2005), pp. 23–46.
- [3] D. BANHOLZER, J. FLIEGE, AND R. WERNER, *On Almost-Sure Rates of Convergence for Sample Average Approximations*, eprint, Optimization Online, 2017, http://www.optimization-online.org/DB_FILE/2017/01/5834.pdf.
- [4] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [5] D. BERTSIMAS AND N. KALLUS, *From Predictive to Prescriptive Analytics*, preprint, <https://arxiv.org/abs/1402.5481>, 2018.
- [6] J. R. BIRGE AND F. LOUVEAUX, *Introduction to Stochastic Programming*, Springer, 2011.
- [7] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer, 2000.
- [8] X. CHEN, L. QI, AND R. S. WOMERSLEY, *Newton's method for quadratic stochastic programs with recourse*, J. Comput. Appl. Math., 60 (1995), pp. 29–46.
- [9] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, 1990, <https://doi.org/10.1137/1.9781611971309>.
- [10] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, reprint of the 1992 edition published by Academic Press, Classics Appl. Math. 60, SIAM, 2009, <https://doi.org/10.1137/1.9780898719000>.
- [11] J. DUPAČOVÁ AND R. J. WETS, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Statist., 16 (1988), pp. 1517–1549.
- [12] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, 2003.
- [13] T. HAO AND J.-S. PANG, *Piecewise affine parameterized value-function based bilevel non-cooperative games*, Math. Program., 180 (2020), pp. 33–73, <https://doi.org/10.1007/s10107-018-1344-7>.
- [14] L. HELLEMO, P. I. BARTON, AND A. TOMASGARD, *Decision-dependent probabilities in stochastic programs with recourse*, Comput. Management Sci., 15 (2018), pp. 369–395.
- [15] J. L. HIGLE AND S. SEN, *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Nonconvex Optim. Appl. 8, Springer, 1996.
- [16] W. HOGAN, *Directional derivatives for extremal-value functions with applications to the completely convex case*, Oper. Res., 21 (1973), pp. 188–209.
- [17] T. HOMEM-DE-MELLO, *Variable-sample methods for stochastic optimization*, ACM Trans. Model. Comput. Simul., 13 (2003), pp. 108–133.
- [18] T. HOMEM-DE-MELLO, *On rates of convergence for stochastic optimization problems under non-*

- independent and identically distributed sampling, *SIAM J. Optim.*, 19 (2008), pp. 524–551, <https://doi.org/10.1137/060657418>.
- [19] T. HOMEM-DE-MELLO AND G. BAYRAKSAN, *Monte Carlo sampling-based methods for stochastic optimization*, *Surveys Oper. Res. Management Sci.*, 19 (2014), pp. 56–85.
 - [20] R. JANIN, *Directional derivative of the marginal function in nonlinear programming*, *Math. Programming Stud.*, (1984), pp. 110–126.
 - [21] A. KING AND R. T. ROCKAFELLAR, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, *Math. Oper. Res.*, 18 (1993), pp. 148–162.
 - [22] K. K. LAU AND R. S. WOMERSLEY, *Multistage quadratic stochastic programming*, *J. Comput. Appl. Math.*, 129 (2001), pp. 105–138.
 - [23] G. LEE, N. TAM, AND N. YEN, *Quadratic Programming and Affine Variational Inequalities: A Qualitative Study*, Springer, 2005.
 - [24] J. LIU AND S. SEN, *Asymptotic Results of Stochastic Decomposition for Two-Stage Stochastic Quadratic Programming*, eprint, Optimization Online, 2019, http://www.optimization-online.org/DB_FILE/2018/08/6787.pdf.
 - [25] Z. LU, Z. ZHOU, AND Z. SUN, *Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization*, *Math. Program.*, 176 (2019), pp. 369–401, <https://doi.org/10.1007/s10107-018-1318-9>.
 - [26] M. NOUIEHED, J.-S. PANG, AND M. RAZAVIYAYN, *On the pervasiveness of difference-convexity in optimization and statistics*, *Math. Program.*, 174 (2019), pp. 195–222.
 - [27] J.-S. PANG, M. RAZAVIYAYN, AND A. ALVARADO, *Computing B-stationary points of nonsmooth DC programs*, *Math. Oper. Res.*, 42 (2016), pp. 95–118.
 - [28] J.-S. PANG, S. SEN, AND U. V. SHANBHAG, *Two-stage non-cooperative games with risk-averse players*, *Math. Program.*, 165 (2017), pp. 235–290.
 - [29] D. RALPH AND H. XU, *Convergence of stationary points of sample average two-stage stochastic programs: A generalized equation approach*, *Math. Oper. Res.*, 36 (2011), pp. 568–592.
 - [30] H. ROBBINS AND S. MONRO, *A stochastic approximation method*, *Ann. Math. Statist.*, 22 (1951), pp. 400–407.
 - [31] S. M. ROBINSON, *An implicit-function theorem for a class of nonsmooth functions*, *Math. Oper. Res.*, 16 (1991), pp. 292–309.
 - [32] S. M. ROBINSON, *Analysis of sample-path optimization*, *Math. Oper. Res.*, 21 (1996), pp. 513–528.
 - [33] R. T. ROCKAFELLAR, *Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming*, in *Nondifferential and Variational Techniques in Optimization*, Springer, 1982, pp. 28–66.
 - [34] A. SHAPIRO, *Monte Carlo sampling methods*, in *Stochastic Programming*, *Handbooks Oper. Res. Management Sci.* 10, Elsevier Sci. B. V., 2003, pp. 353–425.
 - [35] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed., *MOS-SIAM Ser. Optim.* 16, SIAM, 2014, <https://doi.org/10.1137/1.9781611973433>.
 - [36] A. SHAPIRO AND T. HOMEM-DE-MELLO, *On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, *SIAM J. Optim.*, 11 (2000), pp. 70–86, <https://doi.org/10.1137/S1052623498349541>.
 - [37] A. SHAPIRO AND Y. YOMDIN, *On Functions, Representable as a Difference of Two Convex Functions, and Necessary Conditions in a Constrained Optimization*, preprint, Beer-Sheva, 1981.
 - [38] P. D. TAO AND L. T. H. AN, *Convex analysis approach to DC programming: Theory, algorithms and applications*, *Acta Math. Vietnam.*, 22 (1997), pp. 289–355.
 - [39] W. VAN ACKOOLJ AND W. DE OLIVEIRA, *Nonsmooth and nonconvex optimization via approximate difference-of-convex decompositions*, *J. Optim. Theory Appl.*, 189 (2019), pp. 49–80.
 - [40] R. J.-B. WETS, *Stochastic programming*, in *Optimization*, *Handbooks Oper. Res. Management Sci.* 1, North-Holland, 1989, pp. 573–629.
 - [41] A. C. WILLIAMS, *Marginal values in linear programming*, *J. Soc. Indust. Appl. Math.*, 11 (1963), pp. 82–94, <https://doi.org/10.1137/0111006>.
 - [42] H. XU AND D. ZHANG, *Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications*, *Math. Program.*, 119 (2009), pp. 371–401.