

Two-stage Stochastic Programming with Linearly Bi-parameterized Quadratic Recourse*

Junyi Liu [†] Ying Cui[‡] Jong-Shi Pang Suvrajeet Sen

Original December 2018

Abstract

This paper studies the class of two-stage stochastic programs (SPs) 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-convexity (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 SP. 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 SP. 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.

1 Introduction

Since the mid-1950's, 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 monographs, we refer to Birge and Louveaux [5] for SP models and algorithms, Hige and Sen [13] for the stochastic decomposition solution approach, and Shapiro, Dentcheva, and Ruszczyński [34] for a comprehensive mathematical theory.

To date, there are two overwhelming features of two-stage SP models in practical applications. One is the setting that the first-stage decisions affect only the constraints, and only linearly, and not the objective of the second-stage recourse function. The second feature is that the recourse function is defined as the value function of a linear or quadratic program with a parameterized right-hand side. 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 under these two restrictions. There is perhaps a very

*All authors are affiliated with University of Southern California. Email: junyiliu@usc.edu, yingcui@usc.edu, jongship@usc.edu, s.sen@usc.edu.

[†]The work of this and the fourth author is supported by the U.S. National Science Foundation under grant CMMI-1822327 and the U.S. Air Force Office of Sponsored Research under grant FA9550-15-1-0267.

[‡]The work of this and the third author is supported by the U.S. National Science Foundation under grant CMMI-1538605 and the U.S. Air Force Office of Sponsored Research under grant FA9550-18-1-0382.

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 employment of powerful linear programming advances for solving two-stage SPs and easing the treatment of other complications such as the generation of scenarios (i.e., the discretization of the randomness) and the large size of discretized linear programs to be solved. Among reasons for this restricted setting is 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*. The paper aims to address this class of two-stage SPs 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 (1) and (2) below:

$$\underset{x}{\text{minimize}} \quad \zeta(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\omega}}[\psi(x, \tilde{\omega})] \quad \text{subject to} \quad x \in X \subseteq \mathbb{R}^{n_1}, \quad (1)$$

where the recourse function $\psi(x, \omega)$ is given by:

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

In this setting, $\tilde{\omega}$ is a random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with Ω being the sample space, \mathcal{A} being the σ -algebra generated by subsets of Ω and \mathbb{P} being a probability measure defined on \mathcal{A} . The tilde on $\tilde{\omega}$ signifies a random variable, whereas ω without the tilde refers to a realization of the random variable. The first-stage objective function φ is assumed to be convex on an open set Ξ containing the set X , which is compact and convex; the random data is as follows: $f(\omega) \in \mathbb{R}^{n_2}$, $G(\omega) \in \mathbb{R}^{n_2 \times n_1}$; $A(\omega) \in \mathbb{R}^{\ell \times n_1}$, and $\xi(\omega) \in \mathbb{R}^\ell$; the deterministic constants are: $Q \in \mathbb{R}^{n_2 \times n_2}$ 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 which has Q equal to zero (and $G(\omega)$ equal to zero). For ease of reference, we summarize this blanket set-up in the assumption below:

(A) The set X is a compact convex set and φ is a convex function on an open set Ξ containing 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

$$|\varphi(x) - \varphi(x')| \leq \text{Lip}_\varphi \|x - x'\| \quad \text{for all } x, x' \in X, \quad (3)$$

where $\|\bullet\|$ denotes (throughout the paper) the ℓ_2 -norm of vectors (and matrices). To avoid some technical complications, we assume that

(B) the second-stage problem satisfies the relatively complete recourse property on Ξ ; i.e. the recourse function $\psi(x, \omega)$ is finite for all $x \in \Xi$ and almost all $\omega \in \Omega$.

We refer to the monograph [21] for a contemporary comprehensive study of the qualitative properties of solutions to (convex) quadratic programs and to [26, Section 4.1] for a supplement to the reference. In particular, with Q being symmetric positive semidefinite, the relatively complete recourse assumption stipulates the validity [26, Proposition 2] of the following two conditions for all $x \in \Xi$ and almost all $\omega \in \Omega$; i.e., for all ω in a subset $\widehat{\Omega}$ of Ω satisfying $\mathbb{P}(\omega \in \widehat{\Omega}) = 1$:

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

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

We say that a random variable \mathcal{Z} is *essentially bounded* if its essential supremum

$$\text{esssup}(\mathcal{Z}) \triangleq \inf \{ \alpha \mid \mathbb{P}\{\mathcal{Z} \mid |\mathcal{Z}| > \alpha\} = 0 \}$$

is finite. To avoid some technical issues, we further assume that

(C) the given random functions $f(\tilde{\omega})$, $G(\tilde{\omega})$, $A(\tilde{\omega})$, and $\xi(\tilde{\omega})$ 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 $\mathbf{E}_{\tilde{\omega}}[\psi(x, \tilde{\omega})]$ under the above setting, particularly because of the unusual bi-parameterization therein. For a general treatment of this issue, the reader can consult [34, 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 [26, Lemma 1] to deduce several properties of the resource function $\psi(\bullet, \omega)$ and its expectation. Let $M(x, \omega)$ be the argmin (i.e., set of optimal solutions) associated with $\psi(x, \omega)$. It follows that the sets $QM(x, \omega)$ and $[f(\omega) + G(\omega)x]M(x, \omega)$ are singletons for all $(x, \omega) \in \Xi \times \widehat{\Omega}$ under the relatively completely recourse assumption; moreover, for all $\omega \in \widehat{\Omega}$, the unique elements in these sets are Lipschitz continuous functions on X which is compact by assumption; moreover the expectation of the ω -dependent Lipschitz constant is finite by assumption (C). The same Lipschitz property holds for $\psi(\bullet, \omega)$. Therefore, throughout the rest of the paper, it is justified to take the expected recourse function $\mathbf{E}_{\tilde{\omega}}[\psi(x, \tilde{\omega})]$ to be well defined, finite, and continuous [34, Theorem 7.43] for all $x \in X$.

As early as in the mid-1990's, quadratic recourse functions in SP have been studied in [7],[20]; Section 6.2 in [5] discusses multi-stage SPs with quadratic recourse. The recent reference [22] presents asymptotic results for two-stage SPs 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 [24], it is shown that the value function $\psi(\bullet, \omega)$ of the bi-parameterized recourse is a difference-of-convex (dc), thus directionally differentiable function for fixed ω . This class of nonconvex functions has a long history; an early work is the unpublished manuscript [32]; a brief history is documented in the former reference. Thus the recourse function $\mathbf{E}_{\tilde{\omega}}[\psi(\bullet, \tilde{\omega})]$ and the combined objective function ζ are also dc. Hence, in principle, the difference-of-convex algorithm (DCA) [36, 1] could be applied to solve the two-stage SP (1). Nevertheless, there are several major difficulties in a direct application of this solution approach. First: the dc decomposition of the value function $\psi(\bullet, \omega)$ given in [24] 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 (1) so that the resulting solution procedure would become practically implementable and not only 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 [37]. 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 sample-average-approximation based procedure may not be the most efficient way to solve the problem (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 [2, 16, 15, 17, 19, 27, 30, 33, 35, 40], when applied to nonconvex SPs, these methods require solving nonconvex subproblems; cf. e.g. the smooth sample average approximation method for computing stationary points presented in [40]. Moreover, for nondifferentiable 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 (1) on hand, none of its detailed structure is exploited by these methods existing in the 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 [25], such a kind of stationary points is the “sharpest” among stationary points of all types. Nevertheless, in spite of the work in the cited reference and its extension [23], this does not appear easy because for one thing, the problem (1) lacks the kind of structure required in these references even if 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 yield such kind of sharp stationary solutions. Short of achieving the stated goal directly, the next question is the following: if we can design a practically implementable convex-programming based algorithm for (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) identify an implicit convex-concave property of the recourse function $\psi(\bullet, \omega)$ based on which the concept of a “generalized critical point” is defined; a sufficient condition is presented for a such a point to be a directional stationary point that highlights the role of multipliers of the second-stage constraints; and (b) an iterative algorithm that combines regularization, sampling, and convexification is described for computing such a point.

2.1 Motivating instances

An example of a linearly bi-parameterized SP can be illustrated by the two-stage shipment planning with pricing discussed in [4]. In the first stage, the decision is the price and the number of units for production in multiple warehouses. In the second stage, after the demands at multiple retailer stores are realized, there is the option of the last-minute backup production at a higher cost and determine the number of units to ship from the warehouses to the retailer 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 a linear effect on the costs appearing in the objective function, both 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 problems is as follows.

We consider one product in a network of M factory/warehouse combination and N retailer 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 $j = 1, \dots, N$ realizes in the second stage, there is an option of last-minute production y_i in factory i at a cost $c_{2i} > c_{1i}$ per unit produced 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. The lost demands suffer a penalty c_3 per unit and the leftover products have a storage cost c_4 per unit. Suppose that the random demand d_j at the location j is approximately linearly dependent on the price p : $d_j = \alpha_j(\tilde{\omega})p + \beta_j(\tilde{\omega})$ where $\{\alpha_j(\tilde{\omega})\}$ and $\{\beta_j(\tilde{\omega})\}$ are random coefficients independent of p . The resulting two-stage SP is as follows:

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

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

$$\begin{aligned} h(\mathbf{x}, p, \tilde{\omega}) = & \underset{\mathbf{y}, \mathbf{z} \geq 0}{\text{minimize}} \quad \mathbf{c}_2^\top \mathbf{y} + \sum_{i=1}^M \sum_{j=1}^N (s_{ij} - p) z_{ij} \\ & \text{subject to} \quad \sum_{i=1}^M z_{ij} \geq \alpha_j(\omega)p + \beta_j(\omega), j = 1, \dots, N \\ & \text{and} \quad \sum_{j=1}^N z_{ij} \leq x_i + y_i, i = 1, \dots, M. \end{aligned}$$

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 in the system. For such systems, the two-stage paradigm of stochastic programming reflects a planning process in which electricity production from traditional thermal plants are pre-planned in the first stage, whereas, backup fast-ramping generators are to supplement production after the production from the renewable energy (wind and solar) are 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 approximated by the parameters which 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 SP where the constraint parameterization is of the kind

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

for some constant vector a and the "min" operator is the component-wise 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 [11]. As the first step to replace the piecewise right-hand side by a smooth function, we introduce the auxiliary (nonnegative) slack variable $s \triangleq Dy - \min(a, \xi(\omega) - A(\omega)x)$, and formulate the recourse function as:

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

where the last constraint is the complementarity formulation of the piecewise affine equation $Dy - s = \min(a, \xi(\omega) - A(\omega)x)$, 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

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

where $\gamma > 0$ is a penalty parameter. The latter function $\psi_\gamma(x, \omega)$ is clearly a bi-parameterized recourse function in the variables (s, y) with x remaining the first-stage variable, for fixed γ .

Lastly, we mention the recent paper [12] in which the authors introduced a class of SPs 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 $\tilde{\omega}$ and considering a standard linear-programming based recourse function, we may write this SP as:

$$\underset{x \in X}{\text{minimize}} \quad \varphi(x) + \int_{\Omega} p(x, \omega) \psi(x, \omega) d\omega, \quad \text{where} \quad \psi(x, \omega) \triangleq \underset{y \in Y(x, \omega)}{\text{minimum}} \quad f(\omega)^\top y, \quad (4)$$

which is equivalent to the following bi-parameterized formulation:

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

Focusing on the case of finite scenarios, the reference [12] recognized that the problem (4) is nonconvex. As such, the authors apply a global optimization solver to a host of test problems and report computational results. In contrast to this reference, our treatment allows the original problem (1), and thus (4) 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 (1) that is like its analog of the singly-parameterized problem (i.e., when $G(\omega) = 0$); since this formulation is fairly straightforward, we omit the details except to note that the resulting deterministic equivalent is a potentially very large-scale non-convex program with the non-convexity being the result of the product $[G(\omega)x]^\top y$ 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 SP (1) with general distributions of the random variable $\tilde{\omega}$, we are faced with several basic challenges of this problem. 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 their unboundedness and the same 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). Overall, these three maneuvers lead to a convex programming based, regularized and sampled method for solving (1), which we call the RCS algorithm. Some additional assumptions will also be introduced to deal with the boundedness issue of the intermediate solutions.

3.1 Regularization

Regularization of the second-stage QP is a key step to define 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:

$$\begin{aligned} \psi_\alpha(x, \omega) \triangleq & \underset{y}{\text{minimum}} \quad [f(\omega) + G(\omega)x]^\top y + \frac{1}{2} y^\top Q^\alpha y \\ & \text{subject to} \quad y \in Y(x, \omega) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega)\}. \end{aligned} \quad (5)$$

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 [24] has demonstrated the dc property of the value function of a bi-parameterized convex quadratic program, a dc decomposition of this function is quite involved in general. Nevertheless for the regularized recourse function $\psi_\alpha(x, \omega)$, we can obtain a dc decomposition rather easily. Indeed, we have

$$\begin{aligned} \psi_\alpha(x, \omega) &= \\ & \underset{y \in Y(x, \omega)}{\text{minimum}} \left[\begin{array}{l} \frac{1}{2} [y + [Q^\alpha]^{-1} (f(\omega) + G(\omega)x)]^\top Q^\alpha [y + [Q^\alpha]^{-1} (f(\omega) + G(\omega)x)] \\ -\frac{1}{2} [f(\omega) + G(\omega)x]^\top [Q^\alpha]^{-1} [f(\omega) + G(\omega)x] \end{array} \right] \\ &= \underset{\lambda \geq 0}{\text{maximum}} \left[\begin{array}{l} -\frac{1}{2} \lambda^\top D [Q^\alpha]^{-1} D^\top \lambda \\ -\frac{1}{2} [f(\omega) + G(\omega)x]^\top [Q^\alpha]^{-1} [f(\omega) + G(\omega)x] \\ +\lambda^\top [\xi(\omega) - A(\omega)x + D [Q^\alpha]^{-1} (f(\omega) + G(\omega)x)] \end{array} \right], \quad \text{by duality} \\ &= \psi_{\alpha,1}(x, \omega) - \psi_{\alpha,2}(x, \omega), \quad \text{where} \\ \psi_{\alpha,1}(x, \omega) &\triangleq \underset{\lambda \geq 0}{\text{maximum}} \left[\begin{array}{l} -\frac{1}{2} \lambda^\top D [Q^\alpha]^{-1} D^\top \lambda + \\ \lambda^\top [\xi(\omega) - A(\omega)x + D [Q^\alpha]^{-1} (f(\omega) + G(\omega)x)] \end{array} \right] \\ \psi_{\alpha,2}(x, \omega) &\triangleq \frac{1}{2} [f(\omega) + G(\omega)x]^\top [Q^\alpha]^{-1} [f(\omega) + G(\omega)x], \end{aligned} \quad (6)$$

are both convex functions in x for fixed ω . Thus we have obtained a dc decomposition of the regularized value function $\psi_\alpha(\bullet, \omega)$. Note that for fixed α , $\psi_{\alpha,1}(\bullet, \omega)$ is piecewise linear-quadratic

thus generally not differentiable while $\psi_{\alpha,2}(\bullet, \omega)$ is quadratic and thus differentiable. The two-stage SP (1) may then be approximated by the following stochastic dc program:

$$\underset{x \in X}{\text{minimize}} \zeta_{\alpha}(x) = \varphi(x) + \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha,1}(x, \tilde{\omega})] - \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha,2}(x, \tilde{\omega})]. \quad (7)$$

For a given vector $x' \in X$, using the gradient $\nabla_x \psi_{\alpha,2}(x', \omega)$ to define the “linearization” of the function $\psi_{\alpha,2}(\bullet, \omega)$ at the vector x' , we obtain the semi-linearization of the function ζ_{α} at x' :

$$\begin{aligned} \widehat{\zeta}_{\alpha}(x; x') &\triangleq \varphi(x) + \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha,1}(x, \tilde{\omega})] - \mathbf{E}_{\tilde{\omega}} [\widehat{\psi}_{\alpha,2}(x, \tilde{\omega}; x')] \\ \text{with } \widehat{\psi}_{\alpha,2}(x, \omega; x') &\triangleq \psi_{\alpha,2}(x', \omega) + \nabla_x \psi_{\alpha,2}(x', \omega)^{\top} (x - x'). \end{aligned} \quad (8)$$

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

For solving (1), we employ independent and identically distributed (iid) samples of the random variable $\tilde{\omega}$ to approximate the expectation in the above convex majorization $\widehat{\zeta}_{\alpha}(x; x')$ by their sample averages. Combining regularization, convexification, and sampling, we may now state the proposed algorithm for solving problem (1). In the algorithm a sequence of iterates is generated by solving convex subprograms. We leave open how this is done with the understanding 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 $\{\alpha_{\nu}\}_{\nu=0}^{\infty} \downarrow 0$ be a sequence of nonnegative scalars such that $Q^{\alpha_{\nu}}$ 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 $\{\omega^{\nu,i}\}_{i=1}^{L_{\nu}}$ that are also independent from those in the past iterations. Generate the next iterate $x^{\nu+1}$ by solving a convex program:

$$x^{\nu+1} = \underset{x \in X}{\text{argmin}} \left[\begin{array}{l} \underbrace{\varphi(x) + \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \psi_{\alpha_{\nu},1}(x, \omega^{\nu,i}) - \frac{1}{L_{\nu}} \sum_{i=1}^{L_{\nu}} \widehat{\psi}_{\alpha_{\nu},2}(x, \omega^{\nu,i}; x^{\nu})}_{\text{convex in } x; \text{ denoted } \widehat{\zeta}_{\alpha_{\nu}}(x; x^{\nu})} \\ + \underbrace{\frac{1}{2\gamma} \|x - x^{\nu}\|^2}_{\text{proximal term}} \end{array} \right] \quad (9)$$

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 [28], 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. The last term $\frac{1}{2\gamma} \|x - x^{\nu}\|^2$ in (9) is a proximal term added to strongly convexify the convex function $\widehat{\zeta}_{\alpha_{\nu}}(\bullet; x^{\nu})$. Dependent on the selected samples and the immediate past iterate, each new iterate $x^{\nu+1}$ is a

measurable random vector [9]. 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 Dupačová and Wets [9], the almost-sure convergence of the SAA-type scheme has been studied extensively; some references include [19, 16, 15, 17, 27, 30, 33, 35, 34, 40] and the most recent [2]. Nevertheless, none of the results are applicable to establish 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 imposed for the sake of algorithmic practicality. In addition, the convergence of the algorithm depends on conditions on the sequence of integers $\{L_\nu\}_{\nu=0}^{\infty}$ and scalars $\{\alpha_\nu\}_{\nu=0}^{\infty}$ which are given in Theorem 11.

It is possible to consider two variants of the RCS. 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 (1).

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

$$\underset{x \in X}{\text{minimize}} \zeta_{\alpha_\nu}(x) \triangleq \varphi(x) + \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha_\nu}(x, \tilde{\omega})]. \quad (10)$$

This variant gives the user the option of whatever algorithm is desired to be used for generating the sequence $\{\hat{x}^\nu\}$. The drawback of this much simplified algorithm is that each subproblem (10) 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 (10); but rather, only a directional stationary point. The other variant of the RCS Algorithm is a little more specific and employs the difference-of-convex algorithm but still leaves the expectation operator alone. In this version, strictly convex SP subproblems are involved.

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

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

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 an accumulation point of the sequence produced has with reference to the SP (1). 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 $\psi_\alpha(\bullet, \omega)$ (for $\alpha \geq 0$) as a key to connect 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 [25], the convex-analysis based Clarke stationarity [8], and the dc-decomposition based critical point [36], 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 (1), under some additional assumptions to be presented in Section 5.1. 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 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 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 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 being convex, then $\bar{x} \in X$ is said to be a critical point of f on X if $0 \in [\partial g(\bar{x}) - \partial h(\bar{x})] + \mathcal{N}(\bar{x}; X)$.

It is clear that 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 formula for the directional derivative formula for the recourse function $\psi(\bullet, \omega)$. While such formulas have been obtained for the value function of linear programs (see the early papers [39, 14]) 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 [6]. The most relevant to a bi-parameterized convex quadratic program is [18, Corollary 3.3] that 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 $\bar{\psi}(\bullet, \bullet, \omega)$ and the partial directional derivatives of the two partial functions $\bar{\psi}(\bullet, z, \omega)$ and $\bar{\psi}(x, \bullet, \omega)$; see [29] and [10, Exercise 3.7.4]. In what follows, we present this formula for the regularized recourse function $\psi_\alpha(x, \omega)$ for an arbitrary $\alpha \geq 0$.

To prepare for this formula, we let $M^\alpha(x, \omega)$ and $\Lambda^\alpha(x, \omega)$ denote respectively the sets of optimal primal and optimal dual solutions of the second-stage regularized value function $\psi_\alpha(x, \omega)$ given by (5). Using the fact that $QM^\alpha(x, \omega)$ and $[f(\omega) + G(\omega)x]M^\alpha(x, \omega)$ are singletons [26, Lemma 1], we can use any $\bar{y} \in M^\alpha(x, \omega)$ to represent the dual optimal set as follows:

$$\Lambda^\alpha(x, \omega) = \left\{ \lambda \geq 0 \left| \begin{array}{l} D^\top \lambda = f(\omega) + G(\omega)x + Q^\alpha \bar{y} \\ \lambda^\top [\xi(\omega) - A(\omega)x] = 2\psi_\alpha(x, \omega) - \bar{y}^\top [f(\omega) + G(\omega)x] \end{array} \right. \right\},$$

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

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

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

$$M^\alpha(x, \omega) = \left\{ y \in Y(x, \omega) \left| \begin{array}{l} f(\omega) + G(\omega)x + Q^\alpha y = D^\top \bar{\lambda} \\ [A(\omega)x + Dy = \xi(\omega)]_i, \quad \forall i \in \mathcal{I}^\alpha(x, \omega) \end{array} \right. \right\}. \quad (11)$$

Let $L_\alpha(x, \omega; y, \lambda) \triangleq [f(\omega) + G(\omega)x]^\top y + \frac{1}{2} y^\top Q^\alpha y + \lambda^\top [\xi(\omega) - A(\omega)x - Dy]$ denote the Lagrangian function of the quadratic program associated with $\psi_\alpha(x, \omega)$ with λ being the constraint multiplier. Let

$$\begin{aligned} \bar{\psi}_\alpha(x, z, \omega) &\triangleq \underset{y}{\text{minimum}} [f(\omega) + G(\omega)z]^\top y + \frac{1}{2} y^\top [Q + \alpha \mathbb{I}] y \\ &\text{subject to } y \in Y(x, \omega) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega)\} \end{aligned} \quad (12)$$

be the lifted regularized recourse function corresponding to the regularized recourse $\psi_\alpha(x, \omega)$. We have the following result whose proof can be found in the above-cited references.

Proposition 3. ([6] and [18, Corollary 3.3]) *Under assumptions (A) and (B), for any $\alpha \geq 0$, the directional derivative of $\psi_\alpha(\bullet, \omega)$ exists on Ξ and for all $\bar{x} \in X$ and $d \in \mathbb{R}^{n_1}$,*

$$\begin{aligned} \psi_\alpha(\bullet, \omega)'(\bar{x}; d) &= \min_{y \in M^\alpha(\bar{x}, \omega)} \max_{\lambda \in \Lambda^\alpha(\bar{x}, \omega)} \left[\underbrace{G(\omega)^\top y - A(\omega)^\top \lambda}_{= \nabla_x L_\alpha(x, \omega; y, \lambda)} \right]^\top d \\ &= \max_{\lambda \in \Lambda^\alpha(\bar{x}, \omega)} \left[-A(\omega)^\top \lambda \right]^\top d + \min_{y \in M^\alpha(\bar{x}, \omega)} \left[G(\omega)^\top y \right]^\top d \\ &= \bar{\psi}_\alpha(\bar{x}, \bullet, \omega)'(\bar{x}; d) + \bar{\psi}_\alpha(\bullet, \bar{x}, \omega)'(\bar{x}; d). \end{aligned} \quad (13)$$

4.3 Generalized criticality

A difference-of-convex function is the sum of a convex and a concave function. It turns out that the recourse function $\psi(\bullet, \omega)$ also has such a convexity-concavity feature, although 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 $x \in X$ and $h(\bullet, x)$ is convex and $h(x, \bullet)$ is concave. To see that the second-stage recourse function $\psi(\bullet, \omega)$ satisfies the convex-concave property for fixed ω , define a lifted recourse function

$$\begin{aligned} \bar{\psi}(x, z, \omega) \triangleq & \underset{y}{\text{minimum}} \quad [f(\omega) + G(\omega)z]^\top y + \frac{1}{2} y^\top Q y \\ & \text{subject to} \quad y \in Y(x, \omega) \triangleq \{y \in \mathbb{R}^{n_2} \mid A(\omega)x + Dy \geq \xi(\omega)\}, \end{aligned} \quad (14)$$

whose optimal solution set we denote $\bar{M}(x, z, \omega)$. Clearly, $\psi(x, \omega) = \bar{\psi}(x, x, \omega)$ and $\bar{\psi}(\bullet, \bullet, \omega)$ has the desired convexity-concavity property; by [26, Lemma 1], this lifted recourse function $\bar{\psi}(\bullet, \bullet, \omega)$ 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. Let $S \subseteq \mathbb{R}^n$ be a convex set and 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

$$0 \in \partial_x h(\bar{x}, \bar{x}) - \partial_z (-h)(\bar{x}, \bar{x}) + \mathcal{N}(\bar{x}; S), \quad (15)$$

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

The term ‘‘generalized critical point’’ stems from the special case of a dc function $f = g_1 - g_2$ where g_1 and g_2 are both 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 (15) 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 to the objective function of (1), we say that $\bar{x} \in X$ is a *generalized critical point* of this SP if

$$0 \in \partial \varphi(\bar{x}) + \partial_x \mathbf{E}_{\tilde{\omega}} [\bar{\psi}(\bar{x}, \bar{x}, \tilde{\omega})] - \partial_z \mathbf{E}_{\tilde{\omega}} [-\bar{\psi}(\bar{x}, \bar{x}, \tilde{\omega})] + \mathcal{N}(\bar{x}; X), \quad (16)$$

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

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

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

Proof. By using a similar proof as that for Theorem 2 in [14] which only assumes the closedness and convexity of the feasible set (see also [31, 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 $\psi(\bullet, \omega)$:

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

It follows from Danskin's Theorem that (see, e.g., [10, Theorem 10.2.1])

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

which yields

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

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

$$\begin{cases} \mathbf{E}_{\tilde{\omega}} \left[\max_{u \in \partial_x \bar{\psi}(\bar{x}, \bar{x}, \omega)} u^\top (x - \bar{x}) \right] = \max_{u \in \partial_z \mathbf{E}_{\tilde{\omega}}[\bar{\psi}(\bar{x}, \bar{x}, \tilde{\omega})]} u^\top (x - \bar{x}), \\ \mathbf{E}_{\tilde{\omega}} \left[\max_{v \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)} v^\top (x - \bar{x}) \right] = \max_{v \in -\partial_z \mathbf{E}_{\tilde{\omega}}[(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\omega})]} v^\top (x - \bar{x}). \end{cases} \quad (17)$$

Below we shall prove the second equation; the first one can be obtained in a similar manner. Let $\bar{v}(\bar{x}, x) \in \operatorname{argmax}_{v \in -\partial_z \mathbf{E}_{\tilde{\omega}}[(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\omega})]} v^\top (x - \bar{x})$. The convexity of $(-\bar{\psi})(\bar{x}, \bullet, \tilde{\omega})$ implies that $\partial_z \mathbf{E}_{\tilde{\omega}}[(-\bar{\psi})(\bar{x}, \bar{x}, \omega)] = \mathbf{E}_{\tilde{\omega}}[\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)]$ (cf. [34, Theorem 7.47]). Then there exists $v(\bar{x}, x, \omega) \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)$ such that $\mathbf{E}_{\tilde{\omega}}[v(\bar{x}, x, \tilde{\omega})] = \bar{v}(\bar{x}, x)$. Therefore,

$$\begin{aligned} \max_{v \in -\partial_z \mathbf{E}_{\tilde{\omega}}[(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\omega})]} v^\top (x - \bar{x}) &= \bar{v}(\bar{x}, x)^\top (x - \bar{x}) \\ &= \mathbf{E}_{\tilde{\omega}}[v(\bar{x}, x, \omega)^\top (x - \bar{x})] \leq \mathbf{E}_{\tilde{\omega}} \left[\max_{v \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)} v^\top (x - \bar{x}) \right]. \end{aligned} \quad (18)$$

Conversely, it follows from [34, Theorems 7.34 and 7.37] that there exists a measurable selection $v(\bar{x}, x, \omega) \in \operatorname{argmax}_{v \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)} v^\top (x - \bar{x})$. Then

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

which implies

$$\begin{aligned} \mathbf{E}_{\tilde{\omega}} \left[\max_{v \in -\partial_z(-\bar{\psi})(\bar{x}, \bar{x}, \omega)} v^\top (x - \bar{x}) \right] &= \mathbf{E}_{\tilde{\omega}}[v(\bar{x}, x, \omega)^\top (x - \bar{x})] \\ &= (\mathbf{E}_{\tilde{\omega}}[v(\bar{x}, x, \omega)])^\top (x - \bar{x}) \leq \max_{v \in -\partial_z \mathbf{E}_{\tilde{\omega}}[(-\bar{\psi})(\bar{x}, \bar{x}, \tilde{\omega})]} v^\top (x - \bar{x}). \end{aligned} \quad (19)$$

The second equation in (17) follows by combining the inequalities (18) and (19). \square

The main result of this subsection is given by the following proposition, in which we relate 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 SP (1).

Proposition 6. *Under assumptions (A), (B) and (C), it holds that*

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

Proof. (a) From Definition 2, we clearly have $S_{\text{d}} \subseteq S_{\text{C}}$. To prove $S_{\text{C}} \subseteq S_{\text{gc}}$, we have, from [38, Proposition 2.12] and Lemma 5, for any $\bar{x} \in S_{\text{C}}$ and all $x \in X$,

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

The above inequality yields

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

By [34, Theorem 7.47], we can interchange the expectation and the subdifferentials; hence \bar{x} is a generalized critical point and part (a) follows.

(b) Suppose for almost all $\omega \in \Omega$, $G(\omega)^\top M(x, \omega)$ is a singleton. By Proposition 3, for any $d \in \mathbb{R}^{n_1}$, we have

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

As before, we have

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

This shows that $\psi(\bullet, \omega)$ is Clarke regular. The last part follows readily by the interchangeability of expectation with the directional derivatives and subdifferentials. Details are omitted. \square

5 Convergence Analysis

The main convergence results of the proposed algorithms are presented in this section. Before stating the theorems, we first make several matrix-theoretic assumptions under which the desired uniform α -dependent boundedness properties of the optimal solutions and multipliers of the regularized recourse function $\psi_\alpha(x, \omega)$ can be established.

5.1 Matrix-theoretic assumptions

Besides the basic assumptions (A), (B), and (C), several additional assumptions are needed for various purposes. These are:

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

(E) For almost all $\omega \in \Omega$,

$$\left[D^\top \lambda = 0 \text{ and } \lambda \geq 0 \right] \Rightarrow A(\omega)^\top \lambda = 0; \quad (20)$$

(F) For almost all $\omega \in \Omega$, $\text{Range } G(\omega) \subseteq \text{Range } Q$;

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 of the regularized QPs in $\psi_\alpha(x, \omega)$ and also of the subgradients of the convex function $\psi_{\alpha,1}(\bullet, \omega)$. We first present the following consequence of condition (D).

Lemma 7. *Let $Q \in \mathbb{R}^{n_2 \times n_2}$ be a symmetric positive semidefinite matrix and $D \in \mathbb{R}^{\ell \times n_2}$. 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:*

$$\underset{y}{\text{minimum}} \quad q^\top y + \frac{1}{2} y^\top [Q + \alpha \mathbb{I}] y \quad \text{subject to } Dy \geq b \quad (21)$$

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

Proof. The proof is by contradiction. Suppose that no such pair of scalars $(\bar{\beta}, \bar{\alpha})$ exists. Then there exist two sequences of vectors, $\{q^k\}$ and $\{b^k\}$, and a sequence of nonnegative scalars $\{\alpha_k\}$ converging to zero such that the sequence of optimal solutions $\{y^k \triangleq y^{\alpha_k}(q^k, b^k)\}$ (for $\alpha_k = 0$, the corresponding solution $y^{\alpha_k}(q^k, b^k)$ is arbitrary) satisfies $\|y^k\| > k \left\| \begin{pmatrix} q^k \\ b^k \end{pmatrix} \right\|$. For each k , there exists a multiplier λ^k such that the Karush-Kuhn-Tucker (KKT) conditions hold:

$$\begin{aligned} q^k + [Q + \alpha_k \mathbb{I}] y^k - D^\top \lambda^k &= 0 \\ 0 \leq \lambda^k \perp Dy^k - b^k &\geq 0. \end{aligned}$$

By considering the normalized sequence $\left\{ \frac{y^k}{\|y^k\|} \right\}$, which must have a nonzero accumulation point, say \bar{y} , dividing the above complementarity conditions by $\|y^k\|$, and letting $k \rightarrow \infty$ along the subsequence that yields \bar{y} , we can deduce that $D\bar{y} \geq 0$ and $Q\bar{y} = 0$. This is a contradiction. \square

There are some simple sufficient conditions for (E) to hold: (i) the Slater condition holds for the sets $Y(x, \omega)$; i.e., there exists $D\bar{y} > 0$; and (ii) $\text{Range } A(\omega) \subseteq \text{Range } D$ for almost all $\omega \in \Omega$. Both sufficient conditions are fairly obvious. Note that the second condition (ii) is the counterpart of assumption (F). Specifically, the latter condition pertains to the pair $(G(\omega), Q)$ which appears in the objective function of the recourse function $\psi(x, \omega)$, whereas the former condition pertains to the pair $(A(\omega), D)$ which appears in the constraint of the same recourse function $\psi(x, \omega)$. In what follows, we will see how assumption (E) can be used to derive various important bounds. To motivate these bounds, we note that since we need to work with the directional derivative of

the convex regularized function $\psi_{\alpha\nu,1}(\bullet, \omega^{\nu,i})$ at the iterate $x^{\nu+1}$, which would involve the optimal solution(s) $\lambda^{\alpha\nu,1}(x^{\nu+1}, \omega)$ of this function in the dual form (cf. (6)), 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 $A(\omega)^\top \lambda^{\alpha\nu,1}(x^{\nu+1}, \omega)$. This turns out to require a non-trivial argument which employs the following lemma.

Lemma 8. *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 $\Lambda^\alpha(x, \omega)$ is the optimal dual solution set of the regularized recourse function $\psi_\alpha(x, \omega)$. For each pair (x, ω) , let $u^\alpha(x, \omega) \in \partial_x \psi_{\alpha,1}(x, \omega)$ be such that $\mathbf{E}_{\tilde{\omega}}[u^\alpha(x, \tilde{\omega})] \in \partial \mathbf{E}_{\tilde{\omega}}[\psi_{\alpha,1}(x, \tilde{\omega})]$. The subgradient $u^\alpha(x, \omega)$ is a convex combination of finitely many vectors each equal to

$$\tilde{u}^\alpha(x, \lambda, \omega) \triangleq \left[-A(\omega)x + D(Q + \alpha \mathbb{I})^{-1} G(\omega)x \right]^\top \lambda, \quad (22)$$

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

Lemma 9. *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 being positive definite, and almost all $\omega \in \Omega$ and all $\lambda \in \Lambda^\alpha(x, \omega)$,*

$$\max \left(\left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\|, \left\| D^\top \lambda \right\| \right) \leq \frac{\theta_1}{\alpha + \rho_{\min}(Q)} + \theta_2 + \theta_3 \alpha. \quad (23)$$

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

$$\mathbf{E}_{\tilde{\omega}}[\|u^\alpha(x, \tilde{\omega})\|] \leq \left[\mathbf{E}_{\tilde{\omega}}[\|u^\alpha(x, \tilde{\omega})\|^2] \right]^{1/2} \leq \frac{\theta'_1}{\alpha + \rho_{\min}(Q)} + \theta'_2 + \theta'_3 \alpha. \quad (24)$$

Proof. By Hoffman's error bound for polyhedral sets, there exists a constant $c^D > 0$ dependent on the matrix D only such that for all $x \in X$ and almost all $\omega \in \Omega$, there exists $\hat{y}(x, \omega)$ satisfying $A(\omega)x + D\hat{y}(x, \omega) \geq \xi(\omega)$ and

$$\|\hat{y}(x, \omega)\| \leq c^D \left\| [\xi(\omega) - A(\omega)x]_+ \right\|, \quad (25)$$

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

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

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

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

which yields

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

Using the bound (25), the inequality $\|D^\top \lambda\| \leq \|Q + \alpha \mathbb{I}\| \left\| (Q + \alpha \mathbb{I})^{-1} (D^\top \lambda) \right\|$, and the essential boundedness assumption (C), we easily deduce from the above inequalities the desired bound (23) 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 $\|A(\omega)\|$, there exists a constant $M > 0$ such that $\mathbb{P}[\hat{\Omega}] = 1$, where $\hat{\Omega} \triangleq \{\omega \mid \|A(\omega)\| \leq M\}$. Without loss of generality, we may assume that $A(\omega)$ satisfies the implication (20) for all $\omega \in \hat{\Omega}$. Let $\hat{\Upsilon}$ be the convex hull of these matrices $A(\omega)$ for $\omega \in \hat{\Omega}$. Since the family of matrices Υ in Lemma 8 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 $\omega \in \hat{\Omega}$, $\|A(\omega)^\top \lambda\| \leq \hat{\gamma} \|D^\top \lambda\|$ for all $\lambda \geq 0$. By the definition (22), we obtain, for any $\lambda \in \Lambda^\alpha(x, \omega)$,

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

By (23) and the bound on $\|A(\omega)^\top \lambda\|$ in terms of $\|D^\top \lambda\|$ obtained in Lemma 8, the bound (24) follows readily for some constants θ'_i , $i = 1, 2, 3$. \square

An immediate consequence of the bound (24) is the Lipschitz continuity of the regularized recourse function $\psi_\alpha(x, \omega)$ with a Lipschitz constant of the order $\frac{1}{\alpha + \rho_{\min}(Q)}$ for $\alpha > 0$ sufficiently small.

Corollary 10. *There exists a constant $Lip_\psi > 0$ such that for all $\alpha \geq 0$ sufficiently small with Q^α being positive definite, and all x and x' in X ,*

$$\left| \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha,1}(x, \tilde{\omega})] - \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha,1}(x', \tilde{\omega})] \right| \leq \frac{Lip_\psi}{\alpha + \rho_{\min}(Q)} \|x - x'\|. \quad (26)$$

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

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

interchanging x and x' yields

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

Thus (26) follows from (24) because α either equals zero (Q is positive definite) or is positive (Q is positive semidefinite) and sufficiently small. \square

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

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

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

$$\mathbf{E}_{\tilde{\omega}} [\|\nabla_x \psi_{\alpha,2}(x, \tilde{\omega})\|] \leq \frac{\text{Lip}_\psi}{\alpha + \rho_{\min}(Q)}, \quad \forall x \in X. \quad (27)$$

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 set-up in [16], 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 $\widehat{\Omega}^\nu \triangleq \prod_{k=1}^\nu \Omega^{L_k}$ and \mathbf{E}_ν be the expectation operator induced by \mathbb{P}_ν . Let \mathcal{F}^ν denote the sigma-algebra generated by subsets of $\widehat{\Omega}^\nu$ so that the family $\{\mathcal{F}^\nu\}_\nu$ is a filtration on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\widehat{\Omega}^\infty \triangleq \prod_{k=1}^\infty \Omega^{L_k}$ and \mathbb{P}_∞ denote the corresponding probability distribution on $\widehat{\Omega}^\infty$. Let \mathbf{E}_∞ be the expectation operator induced by \mathbb{P}_∞ . Notice if \mathcal{Z} is a random variable dependent only on events in the iterations up to ν , i.e., defined on the sigma-algebra \mathcal{F}^ν , then $\mathbf{E}_\infty[\mathcal{Z}] = \mathbf{E}_\nu[\mathcal{Z}]$. Finally, let $\mathbf{E}_\nu[\bullet | \mathcal{F}^{\nu-1}]$ denote the conditional expectation on the probability space $(\widehat{\Omega}^\nu, \mathcal{F}^\nu, \mathbb{P}_\nu)$ given the σ -algebra $\mathcal{F}^{\nu-1}$.

In the convergence theorem for the main RCS algorithm, the 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 $\{\widehat{x}^\nu\}$ where convexification and sampling are both left out, and $\{\widetilde{x}^\nu\}$ 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 11. The RCS Algorithm. *Under assumptions (A)–(E), let $\{\alpha_\nu\}$ be a nonincreasing sequence of nonnegative scalars satisfying $\{\alpha_\nu\} \downarrow 0$, Q^{α_ν} is positive definite for all ν , and*

$$\liminf_{\nu \rightarrow \infty} (\alpha_\nu + \rho_{\min}(Q)) \sqrt{L_\nu} > 0. \quad (28)$$

Let $\{L_\nu\}$ satisfy $\sum_{\nu=1}^\infty \frac{1}{(\alpha_\nu + \rho_{\min}(Q))^2 \sqrt{L_\nu}} < \infty$. Then every accumulation point of the sequence $\{x^\nu\}$ produced by the RCS algorithm is a generalized critical point of the SP (1), with 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 $\{\widehat{x}^\nu\}$ and $\{\widetilde{x}^\nu\}$ are taken to be deterministic sequences. Thus no “probability 1” requirement is attached to the conclusion.

Theorem 12. Regularization only. *Under assumptions (A)–(D), let $\{\alpha_\nu\} \downarrow 0$ be a nonincreasing sequence of nonnegative scalars satisfying (28) and such that Q^{α_ν} is positive definite for all ν . Then every accumulation point of the sequence $\{\widehat{x}^\nu\}$ is a generalized critical point of the SP (1). If in addition (F) holds, then such a point is a d -stationary point, or equivalently, a C -stationary point.*

Theorem 13. Regularization + convexification. *Under assumptions (A)–(D), let $\{\alpha_\nu\} \downarrow 0$ be a nonincreasing sequence of nonnegative scalars such that Q^{α_ν} is positive definite for all ν . The conclusions of Theorem 12 hold for the sequence $\{\tilde{x}^\nu\}$.*

In the next section, we give the detailed proof of Theorem 11, omitting that of the two simplified cases. Here, we recall the lifted regularized recourse function $\bar{\psi}_\alpha(x, z, \omega)$ defined in (12) and the directional derivative formula (13) for the regularized recourse function $\psi_\alpha(x, \omega)$. Let $y^\alpha(x, z, \omega)$ be the unique minimizer of the QP associated with $\bar{\psi}_\alpha(x, z, \omega)$ for a given triplet (x, z, ω) . The following lemma pertains to the limit of this lifted regularized recourse function as $\alpha \downarrow 0$.

Lemma 14. *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 $Q + \alpha_k \mathbb{I}$ being positive definite for all k , the following three statements hold:*

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

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

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

$$\lim_{k \rightarrow \infty} \bar{\psi}_{\alpha_k}(x, z, \omega) = \bar{\psi}(x, z, \omega), \quad \lim_{k \rightarrow \infty} \mathbf{E}_{\tilde{\omega}} [\bar{\psi}_{\alpha_k}(x, z, \tilde{\omega})] = \mathbf{E}_{\tilde{\omega}} [\bar{\psi}(x, z, \tilde{\omega})]; \quad (29)$$

moreover,

$$\lim_{k \rightarrow \infty} \bar{\psi}_{\alpha_k}(x^k, z^k, \omega) = \bar{\psi}(\bar{x}, \bar{z}, \omega), \quad \lim_{k \rightarrow \infty} \mathbf{E}_{\tilde{\omega}} [\bar{\psi}_{\alpha_k}(x^k, z^k, \tilde{\omega})] = \mathbf{E}_{\tilde{\omega}} [\bar{\psi}(\bar{x}, \bar{z}, \tilde{\omega})]; \quad (30)$$

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

Proof. By Lemma 7, the sequence $\{y^k \triangleq y^{\alpha_k}(x^k, z^k, \omega)\}$ is bounded. Without loss of generality, assume that this sequence converges to a limit y^∞ which clearly belongs to $Y(\bar{x}, \omega)$. To show that y^∞ has the claimed minimizing property, consider the KKT conditions: for some λ^k :

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

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

For (b), it suffices to prove the two limits in (29); those in (30) follow from the uniform convergence of latter limits and the fact that for fixed ω , the function $\bar{\psi}(\bullet, \bullet, \omega)$ is continuous on $X \times X$. We have, for any $\bar{y} \in M(x, z, \omega)$ which we can choose, by Lemma 7, to be uniformly bounded for all $(x, z) \in X \times X$ and almost all $\omega \in \Omega$,

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

Conversely, we have

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

Consequently, the first limit in (29) holds uniformly in (x, z) . The second limit follows by the dominated convergence theorem as $\{\bar{\psi}_{\alpha_k}(x, z, \omega)\}$ is uniformly bounded by a constant independently of almost all $\omega \in \Omega$.

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

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

we have, for any z ,

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

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 (30),

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

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

5.3 Proof of Theorem 11

Throughout the proof, it is understood that the convergence probability is with respect to \mathbb{P}_∞ unless otherwise specified. 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 expectation functionals:

$$\tilde{x}^{\nu+1/2} \triangleq \operatorname{argmin}_{x \in X} \varphi(x) + \mathbf{E}_{\tilde{\omega}} \left[\psi_{\alpha_{\nu,1}}(x, \tilde{\omega}) - \hat{\psi}_{\alpha_{\nu,2}}(x, \tilde{\omega}; x^\nu) \right] + \frac{1}{2\gamma} \|x - x^\nu\|^2. \quad (31)$$

On one hand, $\tilde{x}^{\nu+1/2}$ is connected with x^ν in two ways: the linearization at x^ν and regularization via the proximal map. On the other hand, $x^{\nu+1}$ is the unique optimal solution of the strongly convex program (9), which is the SAA approximation of the two-stage convex stochastic program (31) that defines $\tilde{x}^{\nu+1/2}$.

With the above preparation, 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 [3, Lemma 5.31], we deduce that the sequence of objective values $\{\zeta_{\alpha_\nu}(x^{\nu+1})\}$ converges and also $\sum_{\nu=1}^{\infty} \|x^{\nu+1} - x^\nu\|^2$ 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 (9).

The following derivations present the details of the above steps.

Step 1. By [3, Section 27.1], we can derive

$$\begin{aligned}
& (\varphi(\tilde{x}^{\nu+1/2}) + \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha_{\nu,1}}(\tilde{x}^{\nu+1/2}, \tilde{\omega})]) - (\varphi(x^\nu) + \mathbf{E}_{\tilde{\omega}} [\psi_{\alpha_{\nu,1}}(x^\nu, \tilde{\omega})]) \\
& \leq -\frac{1}{\gamma} \left(\tilde{x}^{\nu+1/2} - x^\nu \right)^\top \left(\tilde{x}^{\nu+1/2} - x^\nu - \gamma \mathbf{E}_{\tilde{\omega}} [\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\omega})] \right) \\
& = -\frac{1}{\gamma} \left\| \tilde{x}^{\nu+1/2} - x^\nu \right\|^2 + \left(\tilde{x}^{\nu+1/2} - x^\nu \right)^\top \mathbf{E}_{\tilde{\omega}} [\nabla_x \psi_{\alpha_{\nu,2}}(x^\nu, \tilde{\omega})].
\end{aligned} \tag{32}$$

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

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

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

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

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

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

Adding the above four inequalities together, we deduce:

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

Hence, with

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

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

$$\begin{aligned}
\mathbf{E}_\nu [\|B^\nu\|] &= \mathbf{E}_{\nu-1} \left[\mathbf{E}_\nu \left[\left\| \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \omega^{\nu,i}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\| \mid \mathcal{F}^{\nu-1} \right] \right] \\
&\leq \mathbf{E}_{\nu-1} \left[\mathbf{E}_\nu \left[\left\| \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \omega^{\nu,i}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\
&= \mathbf{E}_{\nu-1} \left[\frac{1}{L_\nu} \mathbf{E}_{\tilde{\omega}} \left[\left\| u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\omega}) - v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\
&= \mathbf{E}_{\nu-1} \left[\frac{1}{L_\nu} \left[\mathbf{E}_{\tilde{\omega}} \left[\left\| u^{\alpha_\nu}(\tilde{x}^{\nu+1/2}, \tilde{\omega}) \right\|^2 \right] - \left\| v^{\alpha_\nu}(\tilde{x}^{\nu+1/2}) \right\|^2 \mid \mathcal{F}^{\nu-1} \right] \right]^{1/2} \\
&\leq \frac{\tilde{V}_1}{(\alpha_\nu + \rho_{\min}(Q)) L_\nu^{1/2}} \quad \text{by Lemma 9.}
\end{aligned}$$

Similarly, (27) yields the existence of a constant $\tilde{V}_2 > 0$ such that

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

Recalling the Lipschitz constant Lip_φ from (3), Lip_ψ from (26) in Corollary 10, and also (26), we have

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

Step 3. Using the inequality $\zeta_{\alpha_{\nu+1}}(x^{\nu+1}) \leq \zeta_{\alpha_\nu}(x^{\nu+1})$, we deduce that

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

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

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

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

$\sum_{\nu=1}^{\infty} \left(\text{Lip}_\varphi + \frac{\text{Lip}_\psi}{\alpha_\nu} + \tilde{\Upsilon} \right) \left[\|\tilde{B}^\nu\| + \|\tilde{C}^\nu\| \right]$ is finite. By [3, Lemma 5.31], it follows that the sequence

$\{\zeta_{\alpha_\nu}(x^\nu)\}$ converges. Moreover, the two sums $\sum_{\nu=1}^{\infty} \|\tilde{x}^{\nu+1/2} - x^\nu\|^2$ and $\sum_{\nu=1}^{\infty} \|x^{\nu+1} - x^\nu\|^2$ are finite

with probability 1.

Step 4. Writing out the variational condition of the optimality of $x^{\nu+1}$ for the problem (9), we can readily show that $x^{\nu+1}$ is optimal for

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

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

All these limits can be proved by invoking Lemma 14 and some suitable bounds of the respective summands in the above limits. Specifically, the convergence of the second sequence follows easily

from (29); that of the third sequence follows from the identity

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

that yields the bound

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

Since the sum $\sum_{\nu=1}^{\infty} \|x^\nu - x^{\nu+1}\|^2 < \infty$, we must have $\lim_{\nu \rightarrow \infty} \sqrt{\nu} \|x^{\nu+1} - x^\nu\| = 0$. Consequently, if (28) is assumed, then the third limit holds. Finally, for the convergence of the first sequence, we first note the following two probability-one limits:

$$\lim_{k(\in \kappa) \rightarrow \infty} \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x^{\nu+1}, \omega^{\nu,i}) = \mathbf{E}_{\tilde{\omega}} [\bar{\psi}(\tilde{x}^\infty, \tilde{x}^\infty, \tilde{\omega})], \quad (33)$$

and for all $x \in X$,

$$\lim_{k(\in \kappa) \rightarrow \infty} \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \bar{\psi}_{\alpha_\nu}(x^{\nu+1}, x, \omega^{\nu,i}) = \mathbf{E}_{\tilde{\omega}} [\bar{\psi}(\tilde{x}^\infty, x, \tilde{\omega})]. \quad (34)$$

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

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

For the subsequence of $\left\{ \frac{1}{L_\nu} \sum_{i=1}^{L_\nu} \left[G(\omega^{\nu,i})^\top y^{\alpha_\nu}(x^{\nu+1}, \omega^{\nu,i}) \right] \right\}_{\nu \in \kappa}$, let a^∞ be an accumulation point.

Then by the limits (33) and (34), we deduce

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

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

6 Conclusions

We have studied a linearly bi-parameterized stochastic program with convex quadratic recourse. We have established the almost sure subsequential convergence of a combined RCS algorithmic framework for computing a generalized critical point, which 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 more of their applications in other deterministic and stochastic programs. The work done in this paper on this class of linearly bi-parametrized 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] Le Thi Hoai An and Pham Dinh Tao. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of Operations Research*, pages 133(1–4):23–46, 2005.
- [2] Dirk Banholzer, Jörg Fliege, and Ralf Werner. On almost-sure rates of convergence for sample average approximations. *SIAM Journal on Optimization*, 2017.
- [3] Heinz H. Bauschke and Patrick Louis Combettes. Convex analysis and monotone operator theory in hilbert spaces. *Springer*, 2011.
- [4] Dimitri Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *arXiv:1402.5481v4*, 2018.
- [5] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.
- [6] Joseph Frédéric Bonnans and Alexander Shapiro. Perturbation analysis of optimization problems. *Springer Science & Business Media*, 2013.
- [7] Xiaojun Chen, Liqun Qi, and Robert S Womersley. Newton’s method for quadratic stochastic programs with recourse. *Journal of Computational and Applied Mathematics*, 60(1-2):29–46, 1995.
- [8] Frank H. Clarke. Optimization and nonsmooth analysis. *Classics in Applied Mathematics, Society of Industrial and Applied Mathematics*, 5, 1990.
- [9] Jitka Dupačová and Roger J.B. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics*, pages 16(4): 1517–1549, 1988.
- [10] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity problems*. Springer Science & Business Media, 2003.
- [11] Tianyu Hao and Jong-Shi Pang. Piecewise affine parameterized value-function based bilevel non-cooperative games. *Mathematical Programming*, pages 1–41.
- [12] Lars Hellemo, Paul I Barton, and Asgeir Tomasgard. Stochastic programming with decision dependent probabilities, 2015.
- [13] Julia L. Higle and Suvrajeet Sen. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, volume 8. Springer Science & Business Media, 2013.
- [14] William Hogan. Directional derivatives for extremal-value functions with applications to the completely convex case. *Operations Research*, pages 21(1): 188–209, 1973.
- [15] Tito Homem-de Mello. On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization*.
- [16] Tito Homem-de Mello. Variable-sample methods for stochastic optimization. *ACM Transactions on Modeling and Computer Simulation*, pages 13(2): 108–133, 2003.

- [17] Tito Homem-de Mello and G. Bayraksan. Monte-carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, pages 19(1): 56–85, 2014.
- [18] Robert Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*.
- [19] Alan King and R. Tyrrell Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, pages 18(1): 148–162, 1993.
- [20] Karen K Lau and Robert S Womersley. Multistage quadratic stochastic programming. *Journal of computational and applied mathematics*, 129(1-2):105–138, 2001.
- [21] G.M. Lee, N.N. Tam, and N.D. Yen. *Quadratic Programming and Affine Variational Inequalities: A Qualitative Study*. Springer, New York.
- [22] Junyi Liu and Suvrajeet Sen. Asymptotic results of stochastic decomposition for two-stage stochastic quadratic programming. *SIAM Journal on Optimization (submitted)*, 2018.
- [23] Zhaosong Lu, Zirui Zhou, and Zhe Sun. Enhanced proximal dc algorithms with extrapolation for a class of structured nonsmooth dc minimization. *Mathematical Programming*, 176(1-2):369–401, 2019.
- [24] Maher Nouiehed, Jong-Shi Pang, and Meisam Razaviyayn. On the pervasiveness of difference-convexity in optimization and statistics. *Mathematical Programming*, pages 1–28, 2017.
- [25] Jong-Shi Pang, Meisam Razaviyayn, and Alberth Alvarado. Computing b-stationary points of nonsmooth dc programs. *Mathematics of Operations Research*, 42(1):95–118, 2016.
- [26] Jong-Shi Pang, Suvrajeet Sen, and Uday V Shanbhag. Two-stage non-cooperative games with risk-averse players. *Mathematical Programming*, 165(1):235–290, 2017.
- [27] Daniel Ralph and Huifu Xu. Convergence of stationary points of sample average two-stage stochastic programs: A generalized equation approach. *Mathematics of Operations Research*, 36(3):568–592, 2011.
- [28] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.
- [29] Stephen M Robinson. An implicit-function theorem for a class of nonsmooth functions. *Mathematics of operations research*, 16(2):292–309, 1991.
- [30] Stephen M Robinson. Analysis of sample-path optimization. *Mathematics of Operations Research*, 21(3):513–528, 1996.
- [31] R Tyrrell Rockafellar. Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming. In *Nondifferential and variational techniques in optimization*, pages 28–66. Springer, 1982.
- [32] 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.

- [33] Alexander Shapiro. Monte carlo sampling methods. *Handbooks in operations research and management science*, 10:353–425, 2003.
- [34] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.
- [35] Alexander Shapiro and Tito Homem-de Mello. On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11(1):70–86, 2000.
- [36] Pham Dinh Tao and Le Thi Hoai An. Convex analysis approach to dc programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997.
- [37] Wim Van Ackooij and Welington De Oliveira. Nonsmooth and nonconvex optimization via approximate difference-of-convex decompositions. *Journal of Optimization Theory and Applications*, pages 1–32, 2019.
- [38] Roger J-B Wets. Chapter viii stochastic programming. *Handbooks in operations research and management science*, 1:573–629, 1989.
- [39] AC Williams. Marginal values in linear programming. *Journal of the Society for Industrial and Applied Mathematics*, 11(1):82–94, 1963.
- [40] Huifu Xu and Dali Zhang. Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming*, 119(2):371–401, 2009.