2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Global Resolution of Chance-Constrained Optimization Problems: Minkowski Functionals and Monotone Inclusions

Peixuan Zhang[1] Uday V. Shanbhag[1] Constantino M. Lagoa[2] Ibrahim E. Bardakci[3]

***Abstract*— Chance-constrained optimization problems, an important subclass of stochastic optimization problems, are often complicated by nonsmoothness, and nonconvexity. Thus far, non-asymptotic rates and complexity guarantees for computing an ϵ-global minimizer remain open questions. We consider a subclass of problems in which the probability is defined as $\mathbb{P}\{\zeta \mid \zeta \in \mathbf{K}(\mathbf{x})\}$, where $\mathbf{K}$ is a set defined as $\mathbf{K}(\mathbf{x}) = \{\zeta \in \mathcal{K} \mid c(\mathbf{x},\zeta) \le 1\}$, $c(\mathbf{x},\bullet)$ is a positively homogeneous function for any $\mathbf{x} \in \mathcal{X}$, and $\mathcal{K}$ is a nonempty and convex set, symmetric about the origin. We make two contributions in this context. (i) First, when ζ admits a log-concave density on $\mathcal{K}$, the probability function is equivalent to an expectation of a nonsmooth Clarke-regular integrand, allowing for the chance-constrained problem to be restated as a convex program. Under a suitable regularity condition, the necessary and sufficient conditions of this problem are given by a monotone inclusion with a compositional expectation-valued operator. (ii) Second, when ζ admits a uniform density, we present a variance-reduced proximal scheme and provide amongst the first rate and complexity guarantees for resolving chance-constrained optimization problems.**

I. INTRODUCTION

The chance-constrained optimization problem has been studied extensively over the last 70 years [1]. A prototypical instance, denoted by (CCOPT), is defined as

$$\min_{\mathbf{x} \in \mathcal{X}} \{ f(\mathbf{x}) \mid \mathbb{P}\{ \zeta \mid c(\mathbf{x},\zeta) \le 1 \} \ge (1-\epsilon) \}, \quad \text{(CCOPT)}$$

where $f : \mathcal{X} \to \mathbb{R}$ is a continuous function, $c : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}^m$ is a vector function, $\mathbb{P}$ is the given probability measure, and ϵ is a positive scalar. Chance-constrained programming [1] has found utility in a breadth of planning, operational, and financial settings (cf. [2]) as well as control and decision theory [3]. Computational resolution has been roughly partitioned into three sub-areas: (i) *Sequential unconstrained minimization (SUMT)* techniques utilize the nonlinear programming approach as represented by SUMT [4] and require gradients of the probability function [5]. (ii) *Monte-Carlo sampling methods* rely on recasting the probability of interest as an expectation of a (discontinuous) indicator function [6]. Recent efforts have employed smoothing to address the discontinuity [7] while convergence guarantees to Clarke-stationary points have been proven via variational analysis. (iii) *Integer programming techniques* [8] have led to a sample-average approximation (SAA) framework for chance-constrained optimization, where the SAA problem requires solving an integer program; such estimators converge a.s. to the global minimizers of the original problem [8]. Yet a key gap persists.

Gap. To the best of our knowledge, there are no non-asymptotic rate and overall complexity guarantees for computing an ϵ-global minimizer of (CCOPT).

The above gap motivates the present work, which represents a comprehensive generalization of our prior work [9] focusing on probability maximization problems. The main contributions of the present work are as follows.

(I) When $c(\bullet,\zeta)$ abides by a suitable algebraic structure and ζ admits a log-concave density on a convex set $\mathcal{K}$ symmetric about the origin, by leveraging a *layer-cake representation*, we show that (CCOPT) is equivalent to a convex stochastic optimization problem with a compositional expectation-valued constraint. In particular, this expectation is with respect to a suitably defined Gaussian density.

(II) Under a suitable regularity condition, the necessary and sufficient conditions of the aforementioned optimization problem can be viewed as a monotone stochastic inclusion. A variance-reduced inexact stochastic proximal-point framework is presented for resolving such a problem when ζ admits a uniform density and is supported by rate and complexity guarantees.

The remainder of this paper is organized into five sections. Section II provides some preliminary background while in Section III, we prove that under a log-concavity assumption on the density, the necessary and sufficient conditions of (CCOPT) are given by a monotone inclusion with a compositional expectation-valued operator. A variance-reduced proximal scheme with rate and complexity guarantees is derived in Section IV.

II. PRELIMINARIES

A. Background

Our approach relies on extending the framework presented in [9] for the probability maximization problem, defined as

$$\max_{\mathbf{x} \in \mathcal{X}} \mathbb{P}\{ \zeta \mid \zeta \in \mathbf{K}(\mathbf{x}) \}, \text{ where} \tag{1}$$

$$\mathbf{K}(\mathbf{x}) \triangleq \{ \zeta \mid c(\mathbf{x},\zeta) \le 1 \} \tag{2}$$

and $c : \mathcal{X} \times \mathbb{R}^n \to \mathbb{R}^{n_1}$ is a real-valued map. Our prior research [9] relied on assuming that ζ was uniformly distributed on a compact and convex set, symmetric about the origin (such as a sphere or an ellipsoid). In this paper, we

[1]Industrial and Manufacturing Engineering, Pennsylvania State University, Univ. Park, (pqz5090,udaybag@psu.edu); [2]Electrical Engineering and Computer Science, Pennsylvania State University, Univ. Park (cml18@psu.edu); [3] Electrical-Electronics Engineering, Bartin University, Turkey (ibardakci@bartin.edu.tr)

introduce a crucial *generalization* to symmetric log-concave densities from uniform densities.

Definition 1 (Sym. log-concavity). *A function $h : \mathbb{R}^n \to [0,\infty)$ is sym. log-concave if for any $u, v \in \mathbb{R}^n, \lambda \in [0,1]$, $h(u) = h(-u)$ and $h((1-\lambda)u+\lambda v) \geq [h(u)]^{1-\lambda}[h(v)]^{\lambda}$.* □

This generalization complicates matters significantly since a key requirement of our prior work is the positive homogeneity of $c(\mathbf{x},\bullet)$ for any $\mathbf{x}$. This requirement is crucial in leveraging our claim building our equivalence claim. Before continuing, we recall the definition of positive homogeneity.

Definition 2 (Positive Homogeneity). *A continuous function $h : \mathbb{R}^n \to \mathbb{R}$ is called positively homogeneous function (PHF) with degree $p \in \mathbb{R}$ if it is a nonnegative function and $h(\alpha \mathbf{x}) = \alpha^p h(\mathbf{x})$ for all $\alpha > 0$ and all $\mathbf{x} \in \mathbb{R}^n$.* □

Our analysis closely relies on leveraging the Minkowski functional of $\mathcal{K}$, defined next.

Definition 3 (Minkowski Functional). *Let the set $\mathcal{K} \subset \mathbb{R}^n$. Then, the Minkowski functional associated with the set $\mathcal{K}$, denoted by $\| \bullet \|_{\mathcal{K}}$, is defined as follows for any $\zeta \in \mathcal{K}$.*

$$\|\zeta\|_{\mathcal{K}} \triangleq \inf \left\{ t > 0 \mid \tfrac{\zeta}{t} \in \mathcal{K} \right\}. \qquad \square$$

Recall that $\| \bullet \|_{\mathcal{K}}$ defines a norm when $\mathcal{K}$ is compact, convex and symmetric about the origin. For instance, if $\mathcal{K}$ is the unit ball in $\mathbb{R}^n$, then the Minkowski functional reduces to $\| \bullet \|_2$ in $\mathbb{R}^n$ i.e. $\|\zeta\|_{\mathcal{K}} = \|\zeta\|_2$. The class of symmetric log-concave distributions is a large class that includes many commonly used distributions such as Gaussian, uniform over convex symmetric sets, Laplace, and Logistic. Next, we discuss the avenue adopted for probability maximization.

B. Probability maximization

In this subsection, we define the set $\mathbf{K}(\mathbf{x})$ as

$$\mathbf{K}(\mathbf{x}) \triangleq \{ \zeta \mid c(\mathbf{x},\zeta) \leq 1 \}, \tag{3}$$

where for $i = 1, \cdots, n_1$, $\tilde{c}_i(\mathbf{x},\zeta)$ is a positively homogeneous function with degree p for any $\mathbf{x} \in \mathcal{X}$ and c is defined as $c(\mathbf{x},\zeta) \triangleq \max_{i \in \{1,\ldots,n_1\}} \tilde{c}_i(\mathbf{x},\zeta)$. By the definition of c, we have that $c(\mathbf{x},\bullet)$ is a positively homogeneous function of degree p for any $\mathbf{x} \in \mathcal{X}$. In prior work [9], we considered the following probability maximization problem.

$$\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}), \text{ where } f(\mathbf{x}) \triangleq \mathbb{P}\{ \zeta \in \mathcal{K} \mid \zeta \in \mathbf{K}(\mathbf{x}) \}. \tag{4}$$

The above probability can be expressed as an expectation with respect to a particular density by leveraging a result relating the integral over a set defined by the intersection of inequalities specified by PHFs to a distinct integral.

Lemma 1. [10, Cor. 2.3] *Let h be a positively homogeneous function of degree p and let $r_1, \ldots, r_\ell$ be PHFs of degree $0 \neq t \in \mathbb{R}$. Let Ψ be a bounded set defined as $\Psi \triangleq \{\zeta \mid r_k(\zeta) \leq 1, k = 1, \cdots, \ell\}$. If $\int_{\mathbb{R}^n} |h(\xi)|\, e^{-\max\{r_1(\xi),\ldots,r_\ell(\xi)\}}\, d\xi < \infty$, then*

$$\int_\Psi h(\zeta)\, d\zeta = \tfrac{1}{\Gamma(1+(n+p)/t)} \int_{\mathbb{R}^n} h(\xi) e^{-\max\{r_1(\xi),\ldots,r_\ell(\xi)\}}\, d\xi.$$

In [9], when ζ is uniformly distributed on $\mathcal{K}$, the above result allowed for relating the probability defined in (4) to an expectation of a continuous (Clarke-regular) integrand with respect to a suitably density, as specified next. *This contrasts with standard approaches where probabilities can be cast as expectations of (discontinuous) indicator functions.*

Theorem 1. Let ζ be uniformly distributed on the set $\mathcal{K}$, where $\mathcal{K}$ is a closed, convex, and compact set, symmetric about the origin. Let ξ be a random vector whose support is the whole space $\mathbb{R}^n$, i.e. $p_\xi(\xi) > 0$ for all $\xi \in \mathbb{R}^n$. Define the continuous function $\tilde{F}_{\text{unif}}(\bullet,\xi)$ as

$$\tilde{F}_{\text{unif}}(\mathbf{x},\xi) \triangleq \tfrac{1}{\text{vol}(\mathcal{K})\Gamma(1+d/p)} \tfrac{1}{p_\xi(\xi)} e^{-\max\{c(\mathbf{x},\xi), \|\xi\|_{\mathcal{K}}^p\}}, \tag{5}$$

where $c(\mathbf{x},\zeta) \triangleq \max_{i=1,2,\ldots,n_1} \tilde{c}_i(\mathbf{x},\zeta)$. Then $\mathbb{P}\{\zeta \mid \zeta \in \mathbf{K}(\mathbf{x})\} = \mathbb{E}_\xi \left[\tilde{F}_{\text{unif}}(\mathbf{x},\xi) \right]$. □

We now articulate the gaps in our prior work.

(a) *Distributional assumptions.* Prior work required ζ to be uniformly distributed on $\mathcal{K}$ and no direct extension to generalizations (such as log-concave distributions) was unavailable.
(b) *Probabilistic constraints.* Prior work focused on maximizing the probability, defined in (4). By observing that a global maximizer of (4) can be obtained by minimizing a convex composition of $\mathbb{E}\left[\tilde{F}_{\text{unif}}(x,\xi)\right]$, a regularized variance-reduced scheme is developed. Our interest is in obtaining a global minimizer of (CCOPT), a probabilistically constrained problem and a generalization of (4).

III. LOG-CONCAVE GENERALIZATIONS

In this section, we develop a convex representation for chance-constrained optimization problems characterized by log-concave, rather than uniform, densities. The necessary and sufficient conditions of optimality of the chance-constrained problem can be cast as a monotone stochastic inclusion. Throughout this section, we impose the following requirement on the density of the random variable ζ.

Assumption 1. *The density of ζ, denoted by p_ζ, is log-concave and symmetric about the origin. Furthermore, let $\beta \triangleq \max_\zeta p_\zeta(\zeta) = p_\zeta(0)$.*

Let the super-level set of p_ζ be defined as $\mathcal{S}(\tau) \triangleq \{ \zeta \mid p_\zeta(\zeta) \geq \tau \}$ for some $\tau > 0$. For $\tau \in (0,\beta)$, $\mathcal{S}(\tau)$ is compact, convex and symmetric about the origin. Suppose the indicator function of the set $\mathcal{S}(\tau)$ is defined as

$$\mathbf{1}_{\mathcal{S}(\tau)}(\zeta) \triangleq \begin{cases} 1 & \text{if} \quad \zeta \in \mathcal{S}(\tau) \\ 0 & \text{if} \quad \zeta \notin \mathcal{S}(\tau). \end{cases}$$

We first show that the probability $\mathbb{P}[\,\zeta \mid \zeta \in \mathbf{K}(\mathbf{x})\,\}$ can be expressed as an expectation of a continuous integrand when ζ has a log-concave density. We begin by assuming that $m = 1$ in the definition of c, i.e. $c(\mathbf{x}, \zeta) = \tilde{c}_1(\mathbf{x}, \zeta)$, where $\tilde{c}_1(\mathbf{x}, \zeta)$ is defined as

$$\tilde{c}_1(\mathbf{x}, \zeta) \triangleq |\zeta^\top \mathbf{x}|. \tag{6}$$

Remark 1. *We should note at this point that a broad range of probabilistic constraints involving linear functions can be formulated using modifications of the function above.*

Theorem 2. Suppose Assumption 1 holds and $\tilde{c}$ is defined as (6). Then $\mathbb{P}\{\,\zeta \mid \zeta \in \mathbf{K}(\mathbf{x})\,\} = \mathbb{E}_{\xi,t}\left[\tilde{F}(\mathbf{x}, \xi, \tau)\right]$, where $\tilde{F}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\tilde{F}(\mathbf{x}, \xi, \tau) \triangleq \mathcal{C}_\tau (2\pi)^{n/2} e^{-\max\{|\xi^T\mathbf{x}|^2, \|\xi\|^2_{\mathcal{S}(\tau)}\} + \frac{\|\xi\|^2_{\mathcal{S}(\tau)}}{2}}$$

and $\mathbb{E}\left[\tilde{F}(\mathbf{x}, \xi, \tau)\right]$ denotes the expectation of $\tilde{F}(\mathbf{x}; \xi, \tau)$ with respect to joint distribution of ξ and τ, denoted by $\tilde{p}_{\xi,\tau}$ and defined as $\tilde{p}_{\xi,\tau}(\xi, \tau) \triangleq \frac{1}{\beta(2\pi)^{n/2}D_\tau} e^{\frac{-\|\xi\|^2_{\mathcal{S}(\tau)}}{2}}$, where D_τ is a positive scalar such that $\int_{(0,\beta]}\int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}D_\tau} e^{\frac{-\|\xi\|^2_{\mathcal{S}(\tau)}}{2}} = 1$ and $\mathcal{C}_\tau \triangleq \mathcal{C} D_\tau$, where $\mathcal{C} \triangleq \frac{1}{\Gamma(1+n/2)}$, and the random parameter τ has a density given by $\tilde{p}_\tau(\tau) \triangleq \frac{1}{\beta}$, and $\beta > 0$.

Proof. Since $p_\zeta(\mathbf{z}) \geq 0$ for any $\mathbf{z}$, attaining its maximum at $\beta \triangleq \left(\max_{\mathbf{z}\in\mathbb{R}^n} p(\mathbf{z})\right) = p(0)$ as a result of the symmetric nature of the density function $p_{\mathbf{z}}$. Since $\mathbf{K}$ is defined as $\mathbf{K}(\mathbf{x}) \triangleq \{\zeta \mid 1 - |\zeta^\top \mathbf{x}| \geq 0\}$, it follows that

$$\begin{aligned}
&\mathbb{P}\{\,\zeta \mid \zeta \in \mathbf{K}(\mathbf{x})\,\} = \int_{\mathbf{K}(\mathbf{x})} p_\zeta(\zeta) d\zeta \\
&= \int_{\mathbf{K}(\mathbf{x})}\int_{[0,\beta]} \mathbf{1}_{\mathcal{S}(\tau)}(\zeta)\; d\tau\; d\zeta = \int_{[0,\beta]}\int_{\mathbf{K}(\mathbf{x})} \mathbf{1}_{\mathcal{S}(\tau)}(\zeta)\; d\zeta\; d\tau \\
&= \int_{[0,\beta]}\int_{\mathbf{K}(\mathbf{x})\cap\mathcal{S}(\tau)} 1\; d\zeta\; d\tau, = \int_{(0,\beta]}\int_{\mathbf{K}(\mathbf{x})\cap\mathcal{S}(\tau)} 1\; d\zeta\; d\tau,
\end{aligned}$$

where second equality follows from the Fubini-Tonelli Theorem for interchanging the order of the integration. Since the super-level sets $\mathcal{S}(\tau)$ are compact (since $\tau > 0$), convex and symmetric, the Minkowski functional $\|\bullet\|_{\mathcal{S}(\tau)}$ associated with these super-level sets defines a norm. Therefore, it is a PHF and by the definition of the Minkowski functional,

$$\zeta \in \mathcal{S}(\tau) \iff \|\zeta\|_{\mathcal{S}(\tau)} \leq 1,$$

for some $\tau > 0$. Then it follows that

$$\begin{aligned}
\mathbf{K}(\mathbf{x}) \cap \mathcal{S}(\tau) &= \{\,\zeta \mid |\zeta^\top \mathbf{x}| \leq 1\} \cap \{\,\zeta \mid \zeta \in \mathcal{S}(\tau)\,\} \\
&= \{\,\zeta \mid |\zeta^\top \mathbf{x}| \leq 1\} \cap \{\zeta \mid \|\zeta\|_{\mathcal{S}(\tau)} \leq 1\,\} \\
&= \{\,\zeta \mid |\zeta^\top \mathbf{x}|^2 \leq 1\,\} \cap \left\{\zeta \mid \|\zeta\|^2_{\mathcal{S}(\tau)} \leq 1\right\} \\
&= \left\{\,\zeta \mid \max\left\{|\zeta^\top \mathbf{x}|^2, \|\zeta\|^2_{\mathcal{S}(\tau)}\,\right\} \leq 1\right\}.
\end{aligned}$$

If g is defined as $g(\mathbf{x}, \zeta) \triangleq \max\left\{|\zeta^\top \mathbf{x}|^2, \|\zeta\|^2_{\mathcal{S}(\tau)}\right\}$ and $\Lambda(\mathbf{x}, \tau) = \mathbf{K}(\mathbf{x}) \cap \mathcal{S}(\tau)$, we may invoke Lemma 1 by noting that $\Lambda(\mathbf{x}, \tau)$ is a bounded set since $\tau \in (0, \beta)$, which implies that $\mathcal{S}(\tau)$ is a bounded set. Consequently,

$$\begin{aligned}
&\int_{(0,\beta]}\int_{\mathbf{K}(\mathbf{x})\cap\mathcal{S}(\tau)} d\zeta\; d\tau = \frac{\int_{(0,\beta]}\int_{\mathbb{R}^n} e^{-g(\mathbf{x},\xi)} d\xi\; d\tau}{\Gamma(1+n/2))} \\
&= \int_{(0,\beta]}\int_{\mathbb{R}^n} \left(\mathcal{C}_\tau (2\pi)^{n/2} e^{-\max\left\{|\xi^\top\mathbf{x}|^2, \|\xi\|^2_{\mathcal{S}(\tau)}\right\} + \frac{\|\xi\|^2_{\mathcal{S}(\tau)}}{2}}\right) \\
&\times \left(\frac{1}{\beta}\frac{1}{D_{\tau,\beta}(2\pi)^{n/2}} e^{-\frac{\|\xi\|^2_{\mathcal{S}(\tau)}}{2}}\right)\; d\xi\; d\tau \\
&= \int_{(0,\beta]}\int_{\mathbb{R}^n} \tilde{F}(\mathbf{x}, \xi, \tau)\tilde{p}_{\xi,\tau}(\xi, \tau) d\xi d\tau = \mathbb{E}_{\xi,\tau}\left[\tilde{F}(\mathbf{x}, \xi, \tau)\right],
\end{aligned}$$

where D_τ is chosen such that $\tilde{p}_{\xi,\tau}$ is a density, i.e.

$$\int_{(0,\beta]}\int_{\mathbb{R}^n} \tilde{p}_{\xi,\tau}(\xi, \tau) d\xi d\tau = 1.$$

□

We observe that $F(\bullet, \xi, \tau)$ can be proven to be Clarke-regular as done in our prior work [9], which in turn allows for claiming the interchange $\partial\mathbb{E}[\tilde{F}(\mathbf{x}, \xi, \tau)] = \mathbb{E}[\partial\tilde{F}(\mathbf{x}, \xi, \tau)]$. We now investigate the development of a convex representation of the chance-constrained problem by first recalling a result from the study of convex measures [11, Lemma 6.2].

Lemma 2. *Consider an α-concave symmetric probability measure $\mathbb{P}$ and let $\mathbf{K}$ be defined as* (2)*. Then for any $\alpha \geq -1$, d is convex on $\mathbb{R}^n$, where $d(\mathbf{x}) \triangleq \frac{1}{\mathbb{P}\{\,\zeta\,|\,\zeta\,\in\,\mathbf{K}(\mathbf{x})\,\}}$.*□

Consequently, a reformulation of (CCOPT), given by (CCP_1), is indeed a convex optimization problem when f is convex on $\mathcal{X}$ and $h(\mathbf{x}) = \frac{1}{\mathbb{E}_{\xi,\tau}\left[\tilde{F}(\mathbf{x},\xi,\tau)\right]} - \frac{1}{1-\epsilon}$.

$$\min_{x\,\in\,\mathcal{X}} \{f(\mathbf{x}) \mid h(\mathbf{x}) \leq 0\} \tag{CCP_1}$$

As a result, under a suitable regularity condition, $\mathbf{x}^*$ is the optimal solution of (CCP_1) if and only if $(\mathbf{x}^*, \lambda^*)$ is primal-dual solution of the following system, where h is a scalar-valued function.

$$\begin{aligned}
&0 \in \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda\partial_{\mathbf{x}} h(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x}) \\
&0 \leq \lambda \perp h(\mathbf{x}) \leq 0.
\end{aligned} \tag{7}$$

In fact, (7) can be cast as an inclusion, defined as

$$\begin{aligned}
&0 \in \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda\partial_{\mathbf{x}} h(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x}) \\
&0 \in -h(\mathbf{x}) + \mathcal{N}_{\mathbb{R}_+}(\lambda),
\end{aligned} \tag{SMI}$$

and compactly representable as $0 \in T(\mathbf{z})$, where

$$\begin{aligned}
\mathbf{z} = (\mathbf{x}; \lambda) \text{ and } T(\mathbf{z}) &\triangleq \{\nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda\partial_{\mathbf{x}} h(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x})\} \\
&\times \left\{-h(\mathbf{x}) + \mathcal{N}_{\mathbb{R}_+}(\lambda)\right\}.
\end{aligned} \tag{8}$$

Note that the interchangeability result allows us to claim that $\partial h(\mathbf{x}) = -\frac{\mathbb{E}[\tilde{G}(\mathbf{x},\xi,\tau)]}{(\mathbb{E}[\tilde{F}(\mathbf{x},\xi,\tau)])^2}$ where $\tilde{G}(\mathbf{x}, \xi, \tau) \in \partial\tilde{F}(\mathbf{x}, \xi, \tau)$ and $\partial[\bullet]$ represents the Clarke subdifferential when the argument is not necessarily convex. Our next result formalizes the relationship between (CCOPT), (CCP_1), and the necessary and sufficient optimality conditions of the latter.

Theorem 3. Consider (CCOPT) where $\mathcal{X}$ is a closed convex set, $f : \mathcal{X} \to \mathbb{R}$ is a smooth convex function, $c(\mathbf{x}, \zeta) \triangleq$

$\left|\zeta^\top \mathbf{x}\right|$, and $\mathbf{K}$ is defined as (2). Suppose Assumption 1 holds and there exists an $\widehat{\mathbf{x}}$ such that $\mathbb{P}\left\{\zeta \mid \left|\zeta^\top \mathbf{x}\right| \leq 1\right\} > (1-\epsilon)$. Then (CCOPT) is equivalent to (CCP_1) and $\bar{\mathbf{x}}$ is a solution of (CCP_1) if and only if $(\bar{\mathbf{x}}, \bar{\lambda})$ is a solution of (7). □

Proof sketch. It can be seen that (CCP_1) is a simple reformulation of (CCOPT) while convexity of (CCP_1) follows from Lemma 2. Under the prescribed Slater regularity condition, the KKT conditions are necessary and sufficient. □

Next, we present a scheme for resolving (CCOPT), but restricted to settings where ζ is uniformly distributed on $\mathcal{K}$. Consequently, $h(\mathbf{x}) \triangleq 1/\mathbb{E}[\tilde{F}(\mathbf{x}, \xi)]$.

IV. A VARIANCE-REDUCED PROXIMAL SCHEME

A. Proximal-point framework

One approach for resolving a deterministic monotone inclusion is the proximal-point algorithm (PPA) proposed by Rockafellar [12]. First, we observe that the map T, defined in (8), is monotone (proof omitted). It may be recalled that the resolvent operator of a monotone operator T, defined as $J_\alpha^T \triangleq (I + \alpha T)^{-1}$, satisfies

$$\left[\,\tilde{\mathbf{z}} = J_\alpha^T(\mathbf{z})\,\right] \quad \equiv \quad \left[\,0 \in T(\tilde{\mathbf{z}}) + \tfrac{1}{\alpha}(\tilde{\mathbf{z}} - \mathbf{z})\,\right]. \tag{9}$$

Challenges arising in employing a proximal-point framework include the computation of the following in finite time: (i) the resolvent $J_\alpha^T(\mathbf{z}) = (I + \alpha T)^{-1}(\mathbf{z})$; (ii) an unbiased evaluation and subgradient of h where $h(\mathbf{x}) = \frac{1}{\mathbb{E}\left[\tilde{F}(\mathbf{x},\xi)\right]} - \frac{1}{1-\epsilon}$. To this end, a variance-reduced inexact variant proximal-point framework (**VR-IPP**) is proposed, necessitating the generation of a sequence $\{\mathbf{z}_k\}$ such that each iterate is an $\mathbf{e}_k$-approximate evaluation of the resolvent operator, leading to the following update rule, given $\mathbf{z}_0$.

$$\mathbf{z}_{k+1} := J_{\alpha_k}^T(\mathbf{z}_k) + \mathbf{e}_k, \quad k \geq 0. \tag{\textbf{VR-IPP}}$$

If $\mathbf{e}_k \equiv 0$ for all k, (**VR-IPP**) reduces to the exact proximal-point method. Such a framework has been proposed in [13] to resolve monotone inclusion problems when T is an expectation-valued set-valued operator and $J_{\alpha_k}^T(\mathbf{z}_k)$ is approximated via Monte-Carlo sampling, contributing to the random error $\mathbf{e}_k$. In this particular setting, T, as defined in (8), is a compositional expectation-valued map. We compute a an approximate solution of $J_{\alpha_k}^T(\mathbf{z}_k)$ (or equivalently a zero of $T(\bullet) + \frac{1}{\alpha_k}(\mathbf{z} - \mathbf{z}_k)$ by generating a sequence $\{\mathbf{y}_k^j\}_{j=1}^{M_k}$, as per the following update rule for $j = 1, \cdots, M_k$, where $\psi(r) \triangleq 1/r$, $\mathbf{y}_k^j = (\mathbf{x}_k^j; \lambda_k^j)$, and $\mathbf{y}_k^0 = \mathbf{z}_{k-1}$.

$$\mathbf{x}_k^{j+1} := \Pi_{\mathcal{X}}\left[\mathbf{x}_k^j - \gamma_j\left(\Delta \mathbf{x}_j^k + \tfrac{\mathbf{x}_j^k - \mathbf{x}_k}{\alpha_k}\right)\right], \tag{10}$$

$$\lambda_k^{j+1} := \Pi_{+}\left[\lambda_k^j - \gamma_j\left(\Delta \lambda_j^k + \tfrac{\lambda_k^j - \lambda_k}{\alpha_k}\right)\right], \tag{11}$$

$$\text{where } \tilde{\mathbf{u}}_k^j \triangleq \tilde{\mathbf{v}}_k^j + \tfrac{1}{\alpha_k}(\mathbf{z}_{k,j} - \mathbf{z}_k), \tilde{\mathbf{v}}_k^j \triangleq \begin{bmatrix}\Delta \mathbf{x}_k^j \\ \Delta \lambda_k^j\end{bmatrix} \tag{12}$$

$$= \begin{bmatrix}\nabla_{\mathbf{x}} f(\mathbf{x}_k^j) + \lambda_k^j \frac{\left(\sum_{\ell=1}^{N_j} \tilde{G}(\mathbf{x}_k^j, \xi_{k,j}^\ell)\right)\psi'_{\epsilon_j}(\bar{F}_{N_j}(\mathbf{x}_k^j))}{N_j} \\ -h_{N_j,\epsilon_j}(\mathbf{x}_k^j)\end{bmatrix}, \tag{13}$$

$$\tilde{G}(\mathbf{x}_k^j, \xi_{k,j}^\ell) \in \partial_{\mathbf{x}} \tilde{F}(\mathbf{x}_j^k, \xi_j^k), \psi'_\epsilon(r) = \tfrac{-1}{r^2+\epsilon}, \tag{14}$$

$$\bar{h}_{N_j,\epsilon_j}(\mathbf{x}_k^j) = \left(\tfrac{1}{\bar{F}_{N_j}(\mathbf{x}_k^j) + \epsilon_j} - \tfrac{1}{1-\epsilon}\right), \tag{15}$$

$$\text{and } \bar{F}_{N_j}(\mathbf{x}_k^j) = \tfrac{\sum_{\ell=1}^{N_j} \tilde{F}(\mathbf{x}_k^j, \xi_{k,j}^\ell)}{N_j}.$$

Moreover, the strongly monotone inclusion problem, given by (9), can be formulated as

$$\begin{aligned} 0 &\in \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \partial_{\mathbf{x}} h(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x}) + \tfrac{1}{\alpha_k}(\mathbf{x} - \mathbf{x}_k) \\ 0 &\in -h(\mathbf{x}) + \mathcal{N}_{\mathbb{R}_+}(\lambda) + \tfrac{1}{\alpha_k}(\lambda - \lambda_k), \end{aligned} \tag{SMI($\mathbf{z}_k$)}$$

is equivalent to the variational inequality problem $\mathrm{VI}(\mathcal{Z}, H(\bullet, \mathbf{z}_k))$ where $\mathcal{Z} \triangleq \mathcal{X} \times \mathbb{R}_+$ and

$$H(\bullet, \mathbf{z}_k) \triangleq \begin{bmatrix}\nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \partial_{\mathbf{x}} h(\mathbf{x}) \\ -h(\mathbf{x})\end{bmatrix} + \tfrac{1}{\alpha_k}\begin{bmatrix}\mathbf{x} - \mathbf{x}_k \\ \lambda - \lambda_k\end{bmatrix}. \tag{16}$$

Recall that $\mathrm{VI}(\mathcal{Z}, H(\bullet, \mathbf{z}_k))$ requires $(\mathbf{u}_k^*, \mathbf{z}_k^*)$ such that

$$(\mathbf{z} - \mathbf{z}_k^*)^\top \mathbf{u}_k^* \geq 0, \qquad \forall\, \mathbf{z} \in \mathcal{Z}, \tag{17}$$

where $\mathbf{u}_k^* \in H(\mathbf{z}_k^*, \mathbf{z}_k)$. However, $\tilde{\mathbf{u}}_k^j$ is not an unbiased evaluation of $H(\mathbf{y}_k^j, \mathbf{z}_k)$ since $\tilde{\mathbf{v}}_k^j$ is not an unbiased evaluation of $T(\mathbf{y}_k^j)$, where

$$\tilde{\mathbf{v}}_k^j = \bar{\mathbf{v}}_k^j + \mathbf{w}_k^j = \underbrace{\begin{bmatrix}\nabla_{\mathbf{x}} f(\mathbf{x}_k^j) + \lambda_k^j G(\mathbf{x}_k^j)\psi'(F(\mathbf{x}_k^j)) \\ -h(\mathbf{x}_k^j)\end{bmatrix}}_{\triangleq \bar{\mathbf{v}}_k^j} \tag{18}$$

$$+ \underbrace{\begin{bmatrix}\lambda_k^j\left(\frac{\sum_{\ell=1}^{N_j} \tilde{G}(\mathbf{x}_k^j, \xi_{k,j}^\ell)}{N_j}\psi'_{\epsilon_j}(\bar{F}_{N_j}(\mathbf{x}_k^j)) - G(\mathbf{x}_k^j)\psi'(F(\mathbf{x}_k^j))\right) \\ -\bar{h}_{N_j,\epsilon_j}(\mathbf{x}_k^j) + h(\mathbf{x}_k^j)\end{bmatrix}}_{\triangleq \mathbf{w}_k^j},$$

$$F(\mathbf{x}) = \mathbb{E}\left[\tilde{F}(\mathbf{x}, \xi)\right], G(\mathbf{x}) \in \partial_{\mathbf{x}} \mathbb{E}\left[\tilde{F}(\mathbf{x}, \xi)\right]. \tag{19}$$

We close by providing a formal statement in Algorithm 1.

Algorithm 1 Variance reduced proximal-point (**VR-SPP**)

Require: Given K, $\{\alpha_k\}_{k=0}^{K-1}$, $\{M_k\}_{k=0}^{K-1}$, $\{\gamma_j\}_{j=1}^{M_k}$, $k := 0, \mathbf{z}_0 := (\mathbf{x}_0; \lambda_0)$ where $\mathbf{x}_0 \in \mathcal{X}$, $\lambda_0 \geq 0$.

while $k < K$ **do** ▷ Step 1
  Let $\mathbf{y}_k^0 = \mathbf{z}_k$.
  Generate $\{\mathbf{y}_k^j\}_{j=1}^{M_k - 1}$ by (10)–(11)
  Let $\mathbf{z}_{k+1} = \mathbf{y}_k$.
  Set $k : k+1$ and go to step 1.
end while

B. Analysis of moments

In this subsection, we analyze the moment properties of $\mathbf{w}_k^j$. By invoking (18),

$$\begin{aligned} \|\mathbf{w}_k^j\|^2 &\leq (\lambda_k^j)^2 \|\mathbf{w}_{k,G}^j\|^2 + \|\mathbf{w}_{k,h}^j\|^2 \\ &\leq \|\mathbf{y}_k^j\|^2 \|\mathbf{w}_{k,G}^j\|^2 + \|\mathbf{w}_{k,h}^j\|^2, \text{ where} \end{aligned} \tag{20}$$

$$\mathbf{w}_{k,G}^j \triangleq \bar{G}_{N_j}(\mathbf{x}_k^j)\psi'_{\epsilon_j}(\bar{F}_{N_j}(\mathbf{x}_k^j)) - G(\mathbf{x}_k^j)\psi'(F(\mathbf{x}_k^j)), \tag{21}$$

$$\mathbf{w}_{k,h}^j \triangleq -\bar{h}_{N_j,\epsilon_j}(\mathbf{x}_k^j) + h(\mathbf{x}_k^j), \tag{22}$$

$\bar{h}_{N_j,\epsilon_j}(\mathbf{x}_k^j)$ is defined in (15) and $\bar{G}_{N_j}(\mathbf{x}_k^j) = \frac{\sum_{\ell=1}^{N_j}\left(\tilde{G}(\mathbf{x}_k^j, \xi_{k,j}^\ell)\right)}{N_j}$. Prior to deriving a bound on the conditional second moments, we define the σ-algebra $\mathcal{F}_k$

for $k \geq 1$ as the history up to iteration k as $\mathcal{F}_k = \mathcal{F}_{k-1} \cup \left\{ \{\tilde{v}^\ell_{k-1,0}\}_{\ell=0}^{N_0-1}, \cdots, \{\tilde{v}^\ell_{k-1,N_{k-1}-1}\}_{j=0}^{N_{k-1}-1} \right\}$, where $\mathcal{F}_0 = \{\mathbf{x}_0, \lambda_0\}$ and $\tilde{v}^{j,\ell}_k \triangleq \left\{ \tilde{F}(\mathbf{x}^j_k, \xi^\ell_{k,j})\} \cup \{\tilde{G}(\mathbf{x}^j_k, \xi^\ell_{k,j}) \right\}$. Moreover, the σ-algebra $\mathcal{F}_{k,j}$ at inner iteration $j \geq 1$ is defined as $\mathcal{F}_{k,j} \triangleq \mathcal{F}_{k-1} \cup \left\{ \{\tilde{v}^\ell_{0,k}\}_{\ell=0}^{N_0-1}, \cdots, \{\tilde{v}^\ell_{j-1,k}\}_{\ell=0}^{N_{j-1}-1} \right\}$.

Lemma 3. *(Bounds on $\mathbf{w}^j_{k,G}$ and $\mathbf{w}^j_{k,h}$)* Suppose $\mathbf{w}^j_{k,G}$ and $\mathbf{w}^j_{k,h}$ are defined as (21) and (22), respectively. Suppose $\mathbb{E}\left[\|G(\mathbf{x}^j_k, \xi)\| \, |\mathcal{F}_{k,j}\right] \leq M_G^2$ and $|\tilde{F}(\mathbf{x}, \xi)| \leq M_F$ for any $\mathbf{x} \in \mathcal{X}, \xi \in \Xi$. Suppose $\epsilon_j = N_j^{-\frac{1}{4}}$ and $F(\mathbf{x}) \geq \epsilon_{\tilde{F}}$ for any $\mathbf{x} \in \mathcal{X}$. Then for any k, j,

$$\mathbb{E}\left[\|\mathbf{w}^j_{k,G}\|^2 \mid \mathcal{F}_{k,j}\right] \leq \frac{\nu_G^2}{\sqrt{N_j}} \text{ and } \mathbb{E}\left[\|\mathbf{w}^j_{k,h}\|^2 \mid \mathcal{F}_{k,j}\right] \leq \frac{\nu_h^2}{\sqrt{N_j}}$$

hold almost surely, where $\nu_g^2 \triangleq \frac{\mathcal{C}^2_{\mathcal{K}}(2\pi)^n \mathbb{E}_{\tilde{p}}[\|\xi\|^2]}{e}$, $\nu_h^2 = \frac{\nu^2_{\tilde{F}}}{\epsilon^2_{\tilde{F}}} + \frac{1}{\epsilon^4_{\tilde{F}}}$, $\nu^2_{\tilde{F}} \triangleq 2(\mathcal{C}^2_{\mathcal{K}}(2\pi)^n + 1)$, and $\nu_G^2 \triangleq \frac{3\nu_g^2}{\epsilon_h^2} + M_G^2 \frac{24\nu_h^2}{\epsilon^4_{\tilde{F}}} + \frac{6(M_F^2+1)\nu_f^2}{\epsilon^4_{\tilde{F}}} + \frac{M_G^2}{\epsilon^8_{\tilde{F}}}$. □

Lemma 4. Consider $\bar{\mathbf{v}}^j_k$ as defined in (18). Then there exist positive scalars C_v, D_v such that for any $k, j \geq 1$,

$$\mathbb{E}\left[\|\bar{\mathbf{v}}^j_k\|^2 \mid \mathcal{F}_{k,j}\right] \leq C_v \|\mathbf{y}^j_k\|^2 + D_v, \text{ a.s.} \quad \square \tag{23}$$

Suppose $\mathbf{y}^*_k$ denotes a solution of $\mathrm{VI}(F_k(\bullet, \mathbf{z}_k), \mathcal{Z})$. Then we may derive the following conditional bounds on $\|\tilde{\mathbf{u}}^j_k\|^2$ and $\|\mathbf{y}^*_k - \mathbf{z}_k\|^2$, respectively for any $k, j \geq 1$.

Lemma 5. Consider $\tilde{\mathbf{u}}^j_k$ and $\|\mathbf{y}^*_k - \mathbf{z}_k\|^2$ for any $k, j \geq 1$. Then the following hold almost surely.

$$\mathbb{E}\left[\|\tilde{\mathbf{u}}^j_k\|^2 \mid \mathcal{F}_{k,j}\right] \leq \left(6C_v + \frac{3}{\alpha_k^2} + \frac{24\nu_G^2}{\sqrt{N_j}}\right) \left\|\mathbf{y}^j_k - \mathbf{z}_k\right\|^2 \tag{24}$$
$$+ \left(6C_v + \frac{24\nu_G^2}{\sqrt{N_j}}\right) \|\mathbf{z}_k\|^2 + \frac{18\nu_G^2}{\sqrt{N_j}} \|\mathbf{y}^*_k\|^2 + 3D_v + 3\frac{\nu_h^2}{\sqrt{N_j}},$$
$$\mathbb{E}\left[\|\mathbf{y}^*_k - \mathbf{z}_k\|^2 \mid \mathcal{F}_k\right] \leq 8\|\mathbf{z}_k\|^2 + 8\|\mathbf{z}^*\|^2. \quad \square \tag{25}$$

We now derive a recursion for $\mathbb{E}[\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 \mid \mathcal{F}_k]$, allowing us to obtain an error bound for the resolvent.

Proposition 1. Consider a sequence $\{\mathbf{y}^j_k\}$ generated for computing an approximate solution of $\mathrm{VI}(\mathcal{Z}, F_k(\bullet, \mathbf{z}_k))$ with $\epsilon_j = N_j^{-1/4}$ and $N_j = \lceil \gamma_j^{-2} \rceil$ for any j. Then for any j, the following holds a.s. .

$$\mathbb{E}\left[\left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2 \mid \mathcal{F}_k\right] \leq (1 - \beta_j)\mathbb{E}\left[\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 \mid \mathcal{F}_k\right]$$
$$+ (8\delta_j + \varepsilon_j)\|\mathbf{z}_k\|^2 + \varphi_j + 8\delta_j\|\mathbf{z}^*\|^2,$$

where $\beta_j, \delta_j, \varepsilon_j$, and φ_j are defined as (27)–(32). □

Proof. By definition of the update rule (10)–(11), we may derive a bound on $\left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2$ by invoking non-expansivity of the projection operator, strong monotonicity, and the property that $\mathbf{y}^*_k$ is a solution of $\mathrm{VI}(\mathcal{Z}, F_k(\bullet, \mathbf{z}_k))$.

$$\left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2 = \left\|\Pi_{\mathcal{Y}}\left[\mathbf{y}^j_k - \gamma_j \tilde{\mathbf{u}}^j_k\right] - \mathbf{y}^*_k\right\|^2$$
$$\leq \left\|\mathbf{y}^j_k - \gamma_j \tilde{\mathbf{u}}^j_k - \mathbf{y}^*_k\right\|^2 \leq \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \gamma_j^2 \|\tilde{\mathbf{u}}^j_k\|^2$$
$$- 2\gamma_j \underbrace{(\mathbf{y}^j_k - \mathbf{y}^*_k)^\top (\bar{\mathbf{u}}^j_k - \mathbf{u}^*_k)}_{\geq \frac{1}{\alpha_k}\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2} - 2\gamma_j \underbrace{(\mathbf{y}^j_k - \mathbf{y}^*_k)^\top \mathbf{u}^*_k}_{\geq 0}$$
$$+ \frac{\gamma_j}{\alpha_k}\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \alpha_k \gamma_j \|\mathbf{w}^j_k\|^2.$$
$$\implies \left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2 \overset{(20)}{\leq} \left(1 - \frac{2\gamma_j}{\alpha_k} + \frac{\gamma_j}{\alpha_k}\right) \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2$$
$$+ \gamma_j^2 \|\tilde{\mathbf{u}}^j_k\|^2 + \alpha_k \gamma_j \|\mathbf{y}^j_k\|^2 \|\mathbf{w}^j_{k,G}\|^2 + \alpha_k \gamma_j \|\mathbf{w}^j_{k,h}\|^2$$
$$\leq \left(1 - \frac{\gamma_j}{\alpha_k}\right) \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \gamma_j^2 \|\tilde{\mathbf{u}}^j_k\|^2 + 2\alpha_k \gamma_j \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2$$
$$\times \|\mathbf{w}^j_{k,G}\|^2 + 2\alpha_k \gamma_j \|\mathbf{y}^*_k\|^2 \|\mathbf{w}^j_{k,G}\|^2 + \alpha_k \gamma_j \|\mathbf{w}^j_{k,h}\|^2$$

where $\tilde{\mathbf{u}}^j_k = \bar{\mathbf{u}}^j_k + \mathbf{w}^j_k$. By taking conditional expectations with respect to $\mathcal{F}_{k,j}$, we obtain the following sequence of inequalities.

$$\mathbb{E}\left[\left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2 \mid \mathcal{F}_{k,j}\right] \leq \left(1 - \frac{\gamma_j}{\alpha_k}\right) \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2$$
$$+ \gamma_j^2 \mathbb{E}\left[\|\tilde{\mathbf{u}}^j_k\|^2 \mid \mathcal{F}_{k,j}\right] + \alpha_k \gamma_j \mathbb{E}\left[\|\mathbf{w}^j_{k,h}\|^2 \mid \mathcal{F}_{k,j}\right]$$
$$+ 2\alpha_k \gamma_j \left(\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \|\mathbf{y}^*_k\|^2\right) \mathbb{E}\left[\|\mathbf{w}^j_{k,G}\|^2 \mid \mathcal{F}_{k,j}\right]$$
$$\overset{\text{Lemma 3}}{\leq} \left(1 - \frac{\gamma_j}{\alpha_k}\right) \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \gamma_j^2 \mathbb{E}\left[\|\tilde{\mathbf{u}}^j_k\|^2 \mid \mathcal{F}_{k,j}\right]$$
$$+ \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} \|\mathbf{y}^*_k\|^2 + \frac{\alpha_k \gamma_j \nu_h^2}{\sqrt{N_j}}$$
$$= \left(1 - \frac{\gamma_j}{\alpha_k} + \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}}\right) \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2$$
$$+ \gamma_j^2 \mathbb{E}\left[\|\tilde{\mathbf{u}}^j_k\|^2 \mid \mathcal{F}_{k,j}\right] + \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} \|\mathbf{y}^*_k\|^2 + \frac{\alpha_k \gamma_j \nu_h^2}{\sqrt{N_j}}. \tag{26}$$

Consequently, we have the following bound.

$$\mathbb{E}\left[\left\|\mathbf{y}^{j+1}_k - \mathbf{y}^*_k\right\|^2 \mid \mathcal{F}_{k,j}\right]$$
$$\overset{(26),(24)}{\leq} \left(1 - \frac{\gamma_j}{\alpha_k} + \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} + 2\gamma_j^2 \left(6C_v + \frac{3}{\alpha_k^2} + \frac{24\nu_G^2}{\sqrt{N_j}}\right)\right)$$
$$\times \|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \frac{\alpha_k \gamma_j \nu_h^2}{\sqrt{N_j}} + \gamma_j^2 \left(3D_v + 3\frac{\nu_h^2}{\sqrt{N_j}}\right)$$
$$+ \left(2\gamma_j^2 \left(6C_v + \frac{3}{\alpha_k^2} + \frac{24\nu_G^2}{\sqrt{N_j}}\right) + 2\left(\frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} + \frac{18\gamma_j^2 \nu_G^2}{\sqrt{N_j}}\right)\right)$$
$$\times \|\mathbf{y}^*_k - \mathbf{z}_k\|^2$$
$$+ \left(\gamma_j^2 \left(6C_v + \frac{24\nu_G^2}{\sqrt{N_j}}\right) + 2\left(\frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} + \frac{18\gamma_j^2 \nu_G^2}{\sqrt{N_j}}\right)\right) \|\mathbf{z}_k\|^2$$
$$\leq (1 - \beta_j)\|\mathbf{y}^j_k - \mathbf{y}^*_k\|^2 + \delta_j \|\mathbf{y}^*_k - \mathbf{z}_k\|^2 + \varepsilon_j \|\mathbf{z}_k\|^2 + \varphi_j,$$

where the last inequality follows from $N_j = \lceil \gamma_j^{-2} \rceil$, $\alpha_k = \alpha$ for all k, and defining $\beta_j, \delta_j, \varphi_j$, and ε_j as

$$\beta_j \triangleq \frac{\gamma_j}{\alpha} - 2\gamma_j^2 \left(\alpha \nu_G^2 + 6C_v + \frac{3}{\alpha^2} + 24\nu_G^2 \gamma_j\right) \tag{27}$$
$$\delta_j \triangleq 2\gamma_j^2 \left(\left(6C_v + \frac{3}{\alpha^2} + 24\gamma_j \nu_G^2\right) + (2\alpha + 18\gamma_j)\nu_G^2\right) \tag{28}$$
$$+ \frac{2\alpha_k \gamma_j \nu_G^2}{\sqrt{N_j}} \|\mathbf{y}^*_k\|^2 + \gamma_j^2 \left(6C_v + \frac{24\nu_G^2}{\sqrt{N_j}}\right) \|\mathbf{z}_k\|^2 \tag{29}$$

$$+\gamma_j^2\left(\frac{18\nu_G^2}{\sqrt{N_j}}\|\mathbf{y}_k^*\|^2\right)+\gamma_j^2\left(3D_v+\frac{3\nu_h^2}{\sqrt{N_j}}\right)+\frac{\alpha_k\gamma_j\nu_h^2}{\sqrt{N_j}} \quad (30)$$

$$\varepsilon_j \triangleq 2\gamma_j^2\left(\left(6C_v+24\gamma_j\nu_G^2\right)+2\left(2\alpha+18\gamma_j\right)\nu_G^2\right) \quad (31)$$

$$\varphi_j \triangleq \gamma_j^2\left(3D_v+3\nu_h^2\gamma_j\right)+\alpha\gamma_j^2\nu_h^2, \quad (32)$$

respectively. Consequently, we have that

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathbf{y}_k^{j+1}-\mathbf{y}_k^*\right\|^2 \mid \mathcal{F}_k\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left\|\mathbf{y}_k^{j+1}-\mathbf{y}_k^*\right\|^2 \mid \mathcal{F}_{k,j}\right] \mid \mathcal{F}_k\right] \\
&\leq (1-\beta_j)\mathbb{E}\left[\|\mathbf{y}_k^j-\mathbf{y}_k^*\|^2 \mid \mathcal{F}_k\right]+\delta_j\mathbb{E}\left[\|\mathbf{y}_k^*-\mathbf{z}_k\|^2 \mid \mathcal{F}_k\right] \\
&+\varepsilon_j\|\mathbf{z}_k\|^2+\varphi_j \overset{(25)}{\leq} (1-\beta_j)\mathbb{E}\left[\|\mathbf{y}_k^j-\mathbf{y}_k^*\|^2 \mid \mathcal{F}_k\right] \\
&+(8\delta_j+\varepsilon_j)\|\mathbf{z}_k\|^2+\varphi_j+8\delta_j\|\mathbf{z}^*\|^2.
\end{aligned}$$

□

We now derive a rate statement for $\mathbb{E}[\,\|\mathbf{y}_k^j-\mathbf{y}_k^*\|^2 \mid \mathcal{F}_k\,]$ using the above recursion under the caveat that the solution set of the original inclusion is bounded. Note that boundedness of the solution set of a monotone inclusion with a maximal monotone operator has been examined in [14].

Proposition 2. Suppose $\{\mathbf{y}_k^j\}$ generated for computing an approximate solution of $\mathrm{VI}(\mathcal{Z},F_k(\bullet,\mathbf{z}_k))$. Suppose $\gamma_j=\frac{\theta}{j}$, $N_j=\lceil\gamma_j^{-2}\rceil$, and $\epsilon_j=N_j^{-1/4}$ for $j\geq 0$. Suppose $\|\mathbf{z}^*\|\leq B$ for any solution $\mathbf{z}^*$ of (SMI). Then there exist positive scalars ν_1,ν_2 such that for any sufficiently large j and $k\geq 0$, $\mathbb{E}\left[\left\|\mathbf{y}_k^j-\mathbf{y}_k^*\right\|^2 \mid \mathcal{F}_k\right]\leq\frac{\nu_1^2+\nu_2^2\|\mathbf{z}_k\|^2}{2j}$ holds a.s. □

By the above error bound, we may now claim a.s. convergence for the sequence $\{\mathbf{z}_k\}$ generated by (**VR-SPP**) by appealing to a result proven in our prior work [13, Prop. 6].

Theorem 4. Consider a sequence $\{\mathbf{z}_k\}$ generated by Algorithm 1. Suppose $M_k=\lceil(k+1)^{2a}\rceil$, $\gamma_j=\frac{\theta}{j}$, $N_j=\lceil\gamma_j^{-2}\rceil$, $\epsilon_j=N_j^{-1/4}$ for $j,k\geq 1$. Furthermore, $\alpha>0$, $\theta>\alpha/2$, and $a>1$. For any $\mathbf{z}_0\in\mathcal{Z}$, $\mathbf{z}_k \xrightarrow[k\to\infty]{a.s.} \mathbf{z}^*$. □

We now provide some properties of J_α^T and T_α.

Lemma 6 (Properties of T_α and J_α^T). [12], [15] For a set-valued maximal monotone operator $T:\mathbb{R}^n\to\mathbb{R}^n$ and for $\alpha>0$, the Yosida approximation operator is defined as $T_\alpha\triangleq\frac{1}{\alpha}(I-J_\lambda^T)$. Then the following hold.

(a) $0\in T(\mathbf{x}) \iff T_\alpha(\mathbf{x})=0$.
(b) T_α is a single-valued and $\frac{1}{\alpha}$-Lipschitz continuous map.
(c) J_α^T is a single-valued and non-expansive map. □

We now derive rate and complexity guarantees associated with computing an ϵ-solution of T_α.

Proposition 3 (Rate of convergence of (VR-SPP) under maximal monotonicity). Consider a sequence $\{\mathbf{z}_k\}$ generated by (**VR-SPP**). Suppose $M_k=\lceil(k+1)^{2a}\rceil$, $\gamma_j=\frac{\theta}{j}$, $N_j=\lceil\gamma_j^{-2}\rceil$, $\epsilon_j=N_j^{-1/4}$ for $j,k\geq 1$. Furthermore, $\alpha>0$, $\theta>\alpha/2$, and $a>1$.

(a) For any $k\geq 0$, we have that $\mathbb{E}[\|T_\alpha(\mathbf{z}^k)\|^2]=\mathcal{O}\left(\frac{1}{k+1}\right)$.
(b) Suppose $\mathbf{x}^{K+1}$ satisfies $\mathbb{E}[\|T_\alpha(\mathbf{z}^{K+1})\|^2]\leq\epsilon$. Then the oracle complexity of computing such an $\mathbf{z}^{K+1}$ satisfies $\sum_{k=0}^{K}\sum_{j=1}^{M_k}N_j\leq\frac{\tilde{C}}{\epsilon^{6a+1}}$. □

V. CONCLUDING REMARKS

Chance-constrained optimization problems assume relevance in decision and control settings. Yet, there is a glaring lacuna in terms of providing non-asymptotic rate and complexity guarantees for even subclasses of such problems. Under a log-concavity assumption on the density, we show that the chance-constrained problem with a prescribed probabilistic constraint is equivalent to a convex optimization problem with a compositional expectation-valued constraint. We then present amongst the first methods equipped with rate and complexity guarantees for such a problem in the form of a variance-reduced proximal-point method.

REFERENCES

[1] A. Charnes, W. W. Cooper, and G. H. Symonds, "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil," *Management science*, vol. 4, no. 3, pp. 235–263, 1958.
[2] A. Shapiro, D. Dentcheva, and A. Ruszczyn'ski, *Lectures on Stochastic Programming, SIAM, Philadelphia*, 2009.
[3] G. C. Calafiore and L. Fagiano, "Robust model predictive control via scenario optimization," *IEEE Trans. Automat. Control*, vol. 58, no. 1, pp. 219–224, 2013.
[4] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM, 1990.
[5] W. van Ackooij and R. Henrion, "Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions," *SIAM J. Optim.*, vol. 24, no. 4, pp. 1864–1889, 2014.
[6] V. I. Norkin, "The analysis and optimization of probability functions," 1993.
[7] Y. Cui, J. Liu, and J.-S. Pang, "Nonconvex and nonsmooth approaches for affine chance-constrained stochastic programs," *Set-Valued and Variational Analysis*, vol. 30, no. 3, pp. 1149–1211, 2022.
[8] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, "Sample average approximation method for chance constrained programming: theory and applications," *Journal of optimization theory and applications*, vol. 142, no. 2, pp. 399–416, 2009.
[9] I. E. Bardakci, A. Jalilzadeh, C. Lagoa, and U. V. Shanbhag, "Probability maximization via minkowski functionals: convex representations and tractable resolution," *Mathematical programming*, pp. 1–43, 2022.
[10] J. B. Lasserre, "Level sets and nongaussian integrals of positively homogeneous functions," *IGTR*, vol. 17, no. 1, 2015.
[11] S. G. Bobkov, "Convex bodies and norms associated to convex measures," *Probability theory and related fields*, vol. 147, no. 1-2, pp. 303–332, 2010.
[12] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM journal on control and optimization*, vol. 14, no. 5, pp. 877–898, 1976.
[13] S. Cui and U. V. Shanbhag, "On the computation of equilibria in monotone and potential stochastic hierarchical games," *Mathematical Programming*, pp. 1–59, 2022.
[14] Y. Zhang, Y. He, and Y. Jiang, "Existence and boundedness of solutions to maximal monotone inclusion problem," *Optim. Lett.*, vol. 11, no. 8, pp. 1565–1570, 2017.
[15] E. Corman and X. Yuan, "A generalized proximal point algorithm and its convergence rate," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1614–1638, 2014.