EFFICIENT OPTIMIZATION PROCEDURES FOR CVAR-CONSTRAINED OPTIMIZATION WITH REGULARLY VARYING RISK FACTORS

Anish Senapati[1], Jose Blanchet[2], Fan Zhang[2], and Bert Zwart[3]

[1]Institute of Computational and Mathematical Eng., Stanford University, Stanford, CA, USA
[2]Dept. of Management Science and Eng, Stanford University, Stanford, CA, USA
[3] Center of Mathematics and Computer Science, Amsterdam, and Eindhoven University of Technology, Eindhoven, NETHERLANDS

ABSTRACT

We study chance-constrained optimization problems (CC-OPT) with regularly varying distributions where risk is measured through Conditional Value-at-Risk. The usual probabilistic constraints within CCOPTs have limitations in modeling and tractability, motivating a less constrained conditional value-at-risk CC-OPT. We design a stochastic gradient descent-type algorithm to solve this relaxation, combining techniques and theory from the optimization and rare-event simulation literature. Rare-event simulation techniques and a precise preconditioning motivated through an epi-convergence argument were employed to find the optimal solution as the chance constraints become tighter. We show that our method does not depend on the constraints' rarity for regularly varying distributions. Theoretical and numerical results concerning two chance-constrained problems illustrate the advantages of our new method over classical stochastic gradient descent methods with a near-constant runtime complexity as a function of the rarity parameter.

1 INTRODUCTION

Many systems face the challenge of achieving optimal utility while also satisfying risk constraints with high probability. Such problems can be formulated into chance-constrained optimization problems whose objective is to satisfy and solve the following optimization problem.

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \text{Prob}\{\phi(\boldsymbol{x},\boldsymbol{\xi})>0\}\leq\delta \\ & \boldsymbol{x}\in\mathbb{R}^m_{++}, \end{array} \qquad (CCP_\delta)$$

where $\boldsymbol{x}\in\mathbb{R}^m$ is an m-dimensional decision variable, and $\boldsymbol{\xi}\in\mathbb{R}^d$ is an d-dimensional random vector. The elements of $\boldsymbol{\xi}$ are often referred to as risk factors; the function $\phi:\mathbb{R}^m\times\mathbb{R}^d\to\mathbb{R}$ is often assumed to be convex in $\boldsymbol{x}$ and models a cost constraint; the function $\boldsymbol{c}^T\boldsymbol{x}$ represents the cost associated with the decision $\boldsymbol{x}$; the parameter $\delta>0$ is the risk tolerance level.

One of the main challenges of the chance-constrained optimization (CCP_δ) is in the case of extreme events (corresponding to when $\delta\to 0$). Many real-world applications require such fine levels of precision within the probabilistic constraint e.g., airlines requiring probabilistic constraints on the order of $10^{-5}\sim 10^{-10}$. Therefore, beyond solving general chance-constrained optimization problems, it is paramount to solve rare event chance-constrained problems with high computational efficiency.

Significant effort has been put into methods to find the solution to these chance-constrained optimization problems. In complete generality, these problems are provably NP-hard (Luedtke et al. 2010). The works of Tong et al. (2022) use large deviation principles to construct convex tractable analytic approximations. However, such approaches tend to be limited to risk factors with Gaussian or elliptical distributions. Other major approaches include the scenario approach which approximates the chance-constrained problem with deterministic constraints $\phi(\boldsymbol{x},\boldsymbol{\xi}^{(i)})<0$ (Calafiore and Campi 2005; Nemirovski and Shapiro 2006) and

979-8-3315-8726-0/25/$31.00 ©2025 IEEE

the sample-average approximation approach which approximates $\boldsymbol{\xi}$ with a sampled empirical distribution (Luedtke and Ahmed 2008).

While chance-constrained problems such as (CCP_δ) offer terrific insights from a modeling perspective, the non-convexity of the probabilistic constraint makes the optimization problem computationally challenging. Blanchet et al. (2024) propose a sampling procedure for regularly varying distributions to augment and reduce the sample complexity of the scenario approach for CCP_δ as $\delta \to 0$. However, even in problem instances where computational complexity can be decreased, the heavy restrictions from constraint compliance result in impractical feasible sets and subsequently unreasonable optimization values as the probabilistic constraint becomes tighter. For that reason, a conditional value-at-risk (CVaR) relaxation scheme (Rockafellar and Uryasev 2000) is commonly used to alleviate such problems. The conditional value-at-risk (CVaR) at level $\alpha \in (0,1)$ for a loss random variable X with density is defined as $\mathrm{CVaR}_\alpha\{X\} = \mathbb{E}[X|X \geq \mathrm{VaR}_\alpha\{X\}]$ where $\mathrm{VaR}_{1-\delta}\{X\}$ is the $(1-\delta)$ quantile of X. The CVaR constraint is more conservative than the chance constraint problem, however this optimization problem is convex and more tractable (Nemirovski and Shapiro 2007). Lemma 1 allows for the reformulation of optimization problems involving CVaR (Rockafellar and Uryasev 2002).

Lemma 1 For $\delta \in (0,1)$, define $h_\delta : \mathbb{R} \to \mathbb{R}$ as $h_\delta(z) = z + \frac{1}{\delta}\mathbb{E}\left[(X-z)^+\right]$, where $(t)^+ = \max(0,t)$. Then we have h_δ is finite and convex and

$$\mathrm{CVaR}_\delta\{X\} = \min_{z\in\mathbb{R}} h_\delta(z), \qquad \mathrm{VaR}_\delta\{X\} = \min\{z \in \mathbb{R} : h_\delta(z) = \mathrm{CVaR}_\delta\{X\}\}.$$

We study the constrained CVaR optimization problem whose epigraphical reformulation is:

$$f^*(\delta) = \min_{\boldsymbol{x}\in\mathbb{R}^m_{++}} \left\{\boldsymbol{c}^T\boldsymbol{x} \,\middle|\, \mathrm{CVaR}_{1-\delta}\{\phi(\boldsymbol{x},\boldsymbol{\xi})\} \leq 0\right\} = \min_{\substack{\boldsymbol{x}\in\mathbb{R}^m_{++}\\ z\in\mathbb{R}}} \left\{\boldsymbol{c}^T\boldsymbol{x} \,\middle|\, z + \frac{1}{\delta}\mathbb{E}\left[(\phi(\boldsymbol{x},\boldsymbol{\xi}) - z)^+\right] \leq 0\right\} \tag{HC-CVaR$_\delta$}$$

We present an algorithm based on importance sampling which efficiently solves (HC-CVaR$_\delta$) to relative accuracy for regularly varying random variables as $\delta \to 0$. Blanchet, Jorritsma, and Zwart (2024) derive asymptotic relationships between the solutions of (CCP_δ) and the solutions of both the scenario approach and CVaR relaxation as $\delta \to 0$. However, their paper does not address optimization procedures to solve the problem, but only the optimality gap present from both relaxations of program (CCP_δ).

Formally, for sufficiently small δ, our algorithm outputs a solution $f^{(n)}(\delta)$ such that:

$$\frac{|f^{(n)}(\delta) - f^*(\delta)|}{f^*(\delta)} < \varepsilon$$

with a non-asymptotic runtime independent of δ. Here, n denotes the iteration count; that is $f^{(n)}(\delta)$ represents the solution obtained after n steps of the iterative scheme that will be developed in the Section 4. Section 2 introduces assumptions for our CVaR problem and two illustrative running examples. Section 3 gives a convergence argument to determine the optimal critical scaling of both the optimal decision $\boldsymbol{x}^*$ and the optimal Lagrange multiplier λ^* for the hard-constrained problem. Section 4 uses the scalings derived in Section 2 to define and iteratively solve a preconditioned optimization problem which can be solved with constant sample complexity as a function of δ. Doing so requires oracle access to a state-dependent importance sampler to construct samples for our optimization procedure. Section 5 presents numerical experiments for our algorithm.

1.1 Notation

We use $\mathbb{R}_{++}$ to denote the set of positive real numbers. We use $\mathbf{1}$ to denote a column vector with all ones in the entries. For a matrix $\boldsymbol{A}$, we use $\boldsymbol{A}^T$ to denote its transpose. The identity matrix is denoted by $\boldsymbol{I}$. We use $\bar{F}_X^{-1}(\delta)$ as shorthand for $\inf\{x \in \mathbb{R} | \mathbb{P}(X > x) \leq \delta\}$ i.e the $1-\delta$ quantile of the random variable X. We use $\mathbb{I}(\cdot)$ to denote the indicator function of an event.

2 ASSUMPTIONS AND RUNNING EXAMPLES

We assume that the distribution of the risk factor $\boldsymbol{\xi}$ is a regularly varying variable.

Definition 1 A d-dimensional random vector $\boldsymbol{\xi} \in \mathbb{R}^d$ is regularly varying with index α if there exists a random vector $\Theta \in \mathbb{S}^{d-1}$ such that for all $t > 0$, we have

$$\frac{\mathbb{P}(|\boldsymbol{\xi}| > tu, \boldsymbol{\xi}/|\boldsymbol{\xi}| \in \cdot)}{\mathbb{P}(|\boldsymbol{\xi}| > u)} \to t^{-\alpha}\mathbb{P}(\Theta \in \cdot) \text{ as } u \to \infty. \tag{1}$$

We refer readers to (Resnick 1987) for more information on multivariate regularly variation. The intuition behind the representation above is that for sufficiently large values of $||\boldsymbol{\xi}||$, $\boldsymbol{\xi}$, the random variable can be separated into a radial component that follows 1-dimensional Pareto distribution and a random angular component on the unit sphere independent of the radial component.

Assumption 1 The risk factors $\boldsymbol{\xi}$ are drawn from a regularly varying distribution with limit measure Θ and tail index $\alpha > 2$.

Assumption 2 Suppose that $\phi(\boldsymbol{x}, \boldsymbol{\xi})$ is convex is $\boldsymbol{x}$ and satisfies $\sup_{\boldsymbol{x} \in \mathbb{R}_{++}} \mathbb{E}[||\nabla_x \phi(\boldsymbol{x}, \boldsymbol{\xi})||^2] \leq L^2$.

While the constraint of $\sup_{\boldsymbol{x} \in \mathbb{R}_{++}} \mathbb{E}[||\nabla_x \phi(\boldsymbol{x}, \boldsymbol{\xi})||^2] \leq L^2$ may seem restrictive, we focus on 2 different examples for our chance-constrained programs, defined by risk factors $\phi(\boldsymbol{x}, \boldsymbol{\xi})$ which satisfy this condition. Additionally, as this assumption is used in bounding the algorithmic runtime, it can be relaxed to be $\sup_{\boldsymbol{x} \in O_\delta} \mathbb{E}[||\nabla_x \phi(\boldsymbol{x}, \boldsymbol{\xi})||^2] \leq L^2$ where O_δ is defined in the sequel.

2.1 Portfolio Optimization

Suppose that there are m different assets to invest in. Let $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_m)$ denote the amount of dollars invested in different assets. The investments have mean return μ_i and a random loss $\boldsymbol{\xi}_i$. The portfolio optimizer's goal is the maximize the mean returns $\boldsymbol{\mu}^T \boldsymbol{x}$ with a portfolio risk constraint $\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{x}^T \boldsymbol{\xi}\right\} \leq \eta$ for a constant $\eta > 0$. We have the equivalent problem of

$$\begin{array}{ll} \min\limits_{\boldsymbol{x} \in \mathbb{R}^d_{++}} & -\boldsymbol{\mu}^T \boldsymbol{x} \\ \text{subject to} & \mathrm{CVaR}_{1-\delta}\left\{\phi(\boldsymbol{x}, \boldsymbol{\xi})\right\} \leq 0, \end{array} \tag{2}$$

where $\phi(\boldsymbol{x}, \boldsymbol{\xi}) = \boldsymbol{x}^T \boldsymbol{\xi} - \eta$. Without loss of generality, we assume that $\eta = 1$ for this problem.

2.2 Salvage Fund

Suppose that there are m firms in a financial system. Let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_m)$ be the losses that each firm incurs and is responsible to pay. Let $\mathbf{Q} = (\mathbf{Q}_{i,j} : i, j \in \{1, ..., m\})$ be a deterministic matrix where $\boldsymbol{Q}_{i,j}$ denotes the amount of money received by firm j when firm i pays one dollar. We assume that $\mathbf{Q}_{i,j} \geq 0$ and $\sum_j \mathbf{Q}_{i,j} < 1$. Let $\boldsymbol{x}$ be the vector that a separate entity, the salvage fund, allocates to each firm. Let $\boldsymbol{y}$ denote the final settlement amounts that each firm holds after all interactions between firm payments have happened. This settlement should satisfy that $\boldsymbol{y} \leq \boldsymbol{\xi}$ (no firm has more than what they owe) and $(\mathbf{I} - \mathbf{Q}^T)\boldsymbol{y} \leq \boldsymbol{x}$ (no firm pays more than what they were given by the salvage fund). Let $y^*(\boldsymbol{x}, \boldsymbol{\xi})$ be the maximal value of $\mathbf{1}^T \boldsymbol{y}$ such that both conditions are satisfied. So,

$$y^*(\boldsymbol{x}, \boldsymbol{\xi}) = \begin{array}{ll} \underset{y}{\text{argmax}} & \mathbf{1}^T \boldsymbol{y} \\ \text{subject to} & (\boldsymbol{I} - \boldsymbol{Q}^T)\boldsymbol{y} \leq \boldsymbol{x}, \boldsymbol{y} \geq 0, \boldsymbol{y} \leq \boldsymbol{\xi}. \end{array}$$

We say firm i bankrupts if $\boldsymbol{\xi}_i - y^*_i \geq m_i$ where $\boldsymbol{m} \in \mathbb{R}^d_+$ is a given vector. We wish to ensure that bankruptcy does not happen with probability $1 - \delta$ while minimizing the costs of the salvage fund. This reduces to the chance-constrained problem of

$$\begin{array}{ll}\min\limits_{\boldsymbol{x}} & \mathbf{1}^T\boldsymbol{x}\\ \text{subject to} & \mathbb{P}(\boldsymbol{\xi}-y^*(\boldsymbol{x},\boldsymbol{\xi})\leq\boldsymbol{m})\geq 1-\delta.\end{array}\tag{3}$$

Now, consider the following function:

$$\phi(\boldsymbol{x},\boldsymbol{\xi})=\begin{array}{ll}\min\limits_{b,\boldsymbol{y}} & b\\ \text{subject to} & (\boldsymbol{\xi}-\boldsymbol{y}-\boldsymbol{m})\leq b\mathbf{1},(\boldsymbol{I}-\boldsymbol{Q}^T)\boldsymbol{y}\leq\boldsymbol{x},\boldsymbol{y}\geq 0.\end{array}\tag{4}$$

Notice that $\phi(\boldsymbol{x},\boldsymbol{\xi})\leq 0$ iff $\boldsymbol{\xi}-y^*(\boldsymbol{x},\boldsymbol{\xi})\leq\boldsymbol{m}$. Then, the problem of

$$\begin{array}{ll}\min\limits_{\boldsymbol{x}\in\mathbb{R}^d_{++}} & \mathbf{1}^T\boldsymbol{x}\\ \text{subject to} & \mathrm{CVaR}_{1-\delta}\left\{\phi(\boldsymbol{x},\boldsymbol{\xi})\right\}\leq 0\end{array}\tag{5}$$

is the CVaR relaxation of finding the minimum amount the salvage fund must pay.

3 SCALING LAWS AND ASSOCIATED PRECONDITIONING

The goal of this section is to construct a reformulation of (HC-CVaR$_\delta$) whose solution is independent of δ when δ is sufficiently small. We first aim to construct a set O_δ that satisfies three key properties:

1. It contains the feasible set, specifically $F_\delta:=\{\boldsymbol{x}\mid\mathrm{CVaR}_{1-\delta}\left\{\phi(\boldsymbol{x},\boldsymbol{\xi})\right\}\leq 0\}\subseteq O_\delta$.
2. It is a compact subset of $[0,\infty)^m$.
3. It admits a scaling structure $O_\delta=\beta_\delta\bar{O}$, where β_δ is a tight scaling parameter and $\bar{O}\subset\mathbb{R}^m$ is a fixed, compact set with non-empty interior that is independent of δ.

In the context of the portfolio optimization problem, we observe that for each i, the inequality $\eta\geq\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{x}^T\boldsymbol{\xi}\right\}\geq\boldsymbol{x}_i\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{\xi}_i\right\}$ holds. This motivates the introduction of the set

$$O_\delta=\left\{\boldsymbol{x}\in\mathbb{R}^d_{++}\;\middle|\;\boldsymbol{x}_i\leq\eta/\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{\xi}_i\right\}\quad\forall i\in[d]\right\}.$$

By construction, $F_\delta\subseteq O_\delta$. Furthermore, O_δ is compact and admits the scaling representation $O_\delta=\beta_\delta\bar{O}$, where the scaling factor is given by $\beta_\delta=\mathrm{CVaR}_{1-\delta}\left\{\mathbf{1}^T\boldsymbol{\xi}\right\}^{-1}\approx\delta^{1/\alpha}$ since the $(1-\delta)$-quantile and CVaR for regularly varying distributions scale like $\delta^{-1/\alpha}$ as $\delta\to 0$.

For the salvage fund problem, we first employ a result from (Blanchet et al. 2024) to represent the risk constraint $\phi(\boldsymbol{x},\boldsymbol{\xi})$ as $\max_{i=1,\ldots,d}\boldsymbol{\xi}_i-\boldsymbol{e}_i^T(\boldsymbol{I}-\boldsymbol{Q}^T)^{-1}\boldsymbol{x}-\boldsymbol{m}_i$ where $\boldsymbol{e}_i$ denotes the unit vector on the ith coordinate. We use the following lemma to construct the superset to the feasible set F_δ.

Lemma 2 Let $\boldsymbol{\xi}$ be a random variable with density, we have that $\mathrm{CVaR}_{1-\delta}\left\{\max_{i=1,..,N}\boldsymbol{a}_i^T\boldsymbol{\xi}+\boldsymbol{b}_i^T\boldsymbol{x}+c_i\right\}\geq\max_{i=1,..,N}\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{a}_i^T\boldsymbol{\xi}+\boldsymbol{b}_i^T\boldsymbol{x}+c_i\right\}$

This follows directly from the optimization formulation of CVaR (Rockafellar and Uryasev 2002), together with weak duality. This lemma along with further calculations show that if $x\in F_\delta$, $-\boldsymbol{e}_i^T(\boldsymbol{I}-\boldsymbol{Q}^T)^{-1}\boldsymbol{x}+\boldsymbol{m}_i+\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{\xi}_i\right\}\leq 0$ for all $i\in[m]$. While such an outer set is not compact, it has a lower bound which scales on the order of $\mathrm{CVaR}_{1-\delta}\left\{\mathbf{1}^T\boldsymbol{\xi}\right\}$. To ensure compactness while also maintaining the uniform critical scaling of O_δ, we introduce the additional constraint $||\boldsymbol{x}||_\infty\leq C_h\mathrm{CVaR}_{1-\delta}\left\{\mathbf{1}^T\boldsymbol{\xi}\right\}$. Therefore, we have the set

$$O_\delta=\bigcap_{i=1}^{m}\{\boldsymbol{x}|\boldsymbol{e}_i^T(\boldsymbol{I}-\boldsymbol{Q}^T)^{-1}\boldsymbol{x}>\boldsymbol{m}_i+\mathrm{CVaR}_{1-\delta}\left\{\boldsymbol{\xi}_i\right\},||\boldsymbol{x}||_\infty\leq C_h\mathrm{CVaR}_{1-\delta}\left\{\mathbf{1}^T\boldsymbol{\xi}\right\}\}.\tag{6}$$

In practice, this additional constraint is reasonable as the modeler should have a limitation on the budget spent after being given an order of magnitude of the expected money to be spent. The critical scaling is then $\beta_\delta = \text{CVaR}_{1-\delta}\left\{\mathbf{1}^T\boldsymbol{\xi}\right\} \approx \delta^{-1/\alpha}$.

Constructing the Lagrangian function of (HC-CVaR$_\delta$), we have the equivalent problem of

$$f^\star := \inf_{\boldsymbol{x}\in\mathbb{R}^d_{++}} \sup_{\lambda\geq 0} \underbrace{\boldsymbol{c}^\top\boldsymbol{x} + \lambda\ \text{CVaR}_{1-\delta}\big(\phi(\boldsymbol{x},\boldsymbol{\xi})\big)}_{f(\boldsymbol{x},\lambda)} = \inf_{\boldsymbol{x}\in O_\delta} \sup_{\lambda\geq 0} \boldsymbol{c}^\top\boldsymbol{x} + \lambda\ \text{CVaR}_{1-\delta}\big(\phi(\boldsymbol{x},\boldsymbol{\xi})\big) \tag{7}$$

Since O_δ is compact and the inner optimization problem is convex in x and concave in λ, Sion's minimax theorem (Sion 1958) applies allowing us to solve the dual problem of $\max_{\lambda\geq 0}\min_{\boldsymbol{x}\in O_\delta} f(\boldsymbol{x},\lambda)$. We first focus on the inner problem minimization problem of $\min_{\boldsymbol{x}\in O_\delta} f(\boldsymbol{x},\lambda)$. As O_δ scales with rate β_δ for sufficiently small δ, we introduce the substitution $\boldsymbol{x} = \bar{\boldsymbol{x}}\beta_\delta$ and instead solve

$$f^\star = \beta_\delta g^\star = \beta_\delta \max_{\lambda\geq 0}\min_{\bar{\boldsymbol{x}}\in\bar{O}} \left[\boldsymbol{c}^T\bar{\boldsymbol{x}} + \frac{\lambda}{\beta_\delta}\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\}\right] := \max_{\lambda\geq 0}\min_{\bar{\boldsymbol{x}}\in\bar{O}} g_\delta(\bar{\boldsymbol{x}},\lambda) \tag{8}$$

where $\bar{\boldsymbol{x}}$ is now in a compact domain $\bar{O}$ independent of δ. We now focus on the characterizing the tight scaling factor of the term $\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\}$ as $\delta\to 0$. For the portfolio optimization problem, the following limit holds under Assumption 1: $\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\} \to \frac{\alpha}{\alpha-1}\mathbb{E}[\phi(\bar{\boldsymbol{x}},\Theta)^\alpha]^{\frac{1}{\alpha}} - 1$ (Blanchet, Jorritsma, and Zwart 2024) as $\delta\to 0$. A similar proof technique in the context of the salvage fund problem shows that $\beta_\delta^{-1}\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\} \to \frac{\alpha}{\alpha-1}\cdot\int_1^\infty \mathbb{P}\left(\phi(\bar{\boldsymbol{x}},\Theta u) > 1\right)\alpha u^{-\alpha-1}du$ as $\delta\to 0$. In both cases, after finding the appropriate scaling, the CVaR risk terms uniformly converge to a stable limit independent of δ for any subset C that is a compact subset of $\mathbb{R}^m$.

With the scaling of the optimal budget $\boldsymbol{x}^*$ and the CVaR term characterized, we finally analyze the scaling of the optimal Lagrange multiplier λ^* as a function of δ, denoting the scaling factor of λ^* by κ_δ. For now, we focus in the portfolio optimization case, but the salvage fund case follows with similar reasoning. For notational simplicity, denote $K(\bar{\boldsymbol{x}}) = \frac{\alpha}{\alpha-1}\mathbb{E}[\phi(\bar{\boldsymbol{x}},\Theta)^\alpha]^{\frac{1}{\alpha}} - 1$. In the limit as $\delta\to 0$, the two components in the optimization program in (8) can only balance each other if $\boldsymbol{c}^T\bar{\boldsymbol{x}}$ and $\frac{\lambda}{\beta_\delta}K(\bar{\boldsymbol{x}})$ are of the same order. Therefore, we anticipate the scaling of $\lambda = \beta_\delta\bar{\lambda}$. For technical and computational purposes, we limit our optimization problem to $\bar{\lambda}\in[M_l,M_h]$ for some finite values M_l and M_h.

Due to the uniform convergence of $\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\}$ and the finiteness of $g_\delta(\bar{\boldsymbol{x}},\lambda)$ on $\bar{O}\times[M_l,M_h]$, our optimization problem converges uniformly and therefore epiconverges on this compact domain (Rockafellar and Wets 1998):

$$g_\delta(\bar{\boldsymbol{x}},\bar{\lambda}) = \boldsymbol{c}^T\bar{\boldsymbol{x}} + \bar{\lambda}\text{CVaR}_{1-\delta}\left\{\phi(\bar{\boldsymbol{x}}\beta_\delta,\boldsymbol{\xi})\right\} \xrightarrow{epi} \boldsymbol{c}^T\bar{\boldsymbol{x}} + \bar{\lambda}K(\bar{\boldsymbol{x}}) := g_0(\bar{\boldsymbol{x}},\bar{\lambda}) \tag{9}$$

Defining the functions $\psi_\delta(\bar{\lambda}) = \min_{\bar{\boldsymbol{x}}\in\bar{O}} g_\delta(\bar{\boldsymbol{x}},\bar{\lambda})$ and $\psi_0(\bar{\lambda}) = \min_{\bar{\boldsymbol{x}}\in\bar{O}} g_0(\bar{\boldsymbol{x}},\bar{\lambda})$, epi-convergence of the function $g_\delta(\bar{\boldsymbol{x}},\bar{\lambda})$ also implies uniform convergence and hence epi-convergence of the functions $\{\psi_\delta(\bar{\lambda})\}$ to $\psi_0(\bar{\lambda})$. Additionally, since the domain of $\psi_\delta(\bar{\lambda})$ is limited to $\bar{\lambda}\in[M_l,M_h]$, $\psi_\delta(\bar{\lambda})$ is equi-coercive.

Epi-convergence combined with the equi-coerciveness of $\psi_\delta(\lambda)$ allows us to deduce the convergence of maximizers implying that $\bar{\lambda}^*_\delta \to \bar{\lambda}^*_0$ where $\bar{\lambda}^*_0$ is the maximizer of limiting problem $\max_{\bar{\lambda}\in[M_l,M_h]}\psi_0(\bar{\lambda}) = \max_{\bar{\lambda}\in[M_l,M_h]}\min_{x\in\bar{O}} \boldsymbol{c}^T\bar{\boldsymbol{x}} + \bar{\lambda}K(\bar{\boldsymbol{x}})$. Returning back to the scaling, we then have that $\lambda^*_\delta = \beta_\delta\bar{\lambda}_\delta = \bar{\lambda}_0\beta_\delta + o(\beta_\delta)$; the scaling of the Lagrange multiplier is $\kappa_\delta = \beta_\delta$. The same logic applied can also be applied to the salvage fund problem to find the critical scaling of λ^* to be $\kappa_\delta = 1$.

The additional extra constraint of $\bar{\lambda}\in[M_l,M_h]$ may seem counterintuitive as we are arguing that $\bar{\lambda}$ is independent of δ. However, the epi-convergence argument reveals an insightful observation: provided that the optimal solution $\bar{\lambda}^*_0$ of the limiting problem $\max_{\bar{\lambda}\geq 0}\psi_0(\bar{\lambda})$ is finite, unique and independent of δ, M_l and M_h can be chosen quite conservatively apriori to ensure that $\lambda^*_0\in[M_l,M_h]$ and the convergence of maximizers then guarantees that the optimal Lagrange multiplier $\bar{\lambda}^*_\delta \in [M_l, M_h]$ for small enough δ. Noticing that $\max_{\lambda \geq 0} \psi_0(\bar{\lambda})$ is equivalent to the program $\{\min\limits_{\boldsymbol{x} \in \bar{O}} \boldsymbol{c}^T \bar{\boldsymbol{x}}$ subject to $K(\bar{\boldsymbol{x}}) \leq 0\}$, a sufficient condition to ensure that $\bar{\lambda}^*_0$ are finite and unique are verification of Slaters condition for this new optimization program. Independence from δ is immediate since this new program has constraints which are independent of δ.

4 STOCHASTIC GRADIENT DESCENT PROCEDURES

The prior section provided a scaling method for λ^* and $\boldsymbol{x}^*$ through the multipliers κ_δ and β_δ respectively. Subsequently, an equivalent preconditioned problem $g(\bar{\boldsymbol{x}}, \bar{\lambda})$ efficiently was introduced. We now aim to solve the maxmin problem of $\max_{\bar{\lambda} \in [M_l, M_h]} \min_{\boldsymbol{x} \in \bar{O}} \boldsymbol{c}^T \bar{\boldsymbol{x}} + \bar{\lambda} \mathrm{CVaR}_{1-\delta} \{\phi(\bar{\boldsymbol{x}}, \boldsymbol{\xi})\}$ with a stochastic gradient descent procedure (Robbins and Monro 1951). Analysis of SGD shows a convergence rate of $\frac{1}{\sqrt{n}}$ for non-smooth objectives and $\frac{1}{n}$ for smooth functions or strongly convex functions to the optimal value (Moulines and Bach 2011). Such descent methods have a dependence on (i) the distance of the initial iterate to an optimal point and (ii) a bound on $\mathbb{E}_{\boldsymbol{\xi}}[||\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{\xi})||_2^2]$ where f is the function we wish to optimize. Looking at the preconditioned problem with the optimization reformulation of CVaR, we have

$$\min_{\bar{\boldsymbol{x}} \in O_\delta, z \in \mathbb{R}} \boldsymbol{c}^T \bar{\boldsymbol{x}} + \frac{\bar{\lambda} \kappa_\delta}{\beta_\delta} \left(z + \frac{1}{\delta} \mathbb{E}[(\phi(\beta_\delta \bar{\boldsymbol{x}}, \boldsymbol{\xi}) - z)^+] \right), \tag{10}$$

we can also precondition $z = h_\delta \bar{z}$ where h_δ is a scaling factor that allows the CVaR term to converge to a stable limit. The intuition here relies on the fact that for regularly varying distributions, we have that the CVaR and VaR values have the same dependence as a function of δ. Therefore, for the portfolio optimization problem, we have that $h_\delta = 1$, and for the salvage fund problem, we have that $h_\delta = \beta_\delta$. In both cases, one arrives at the final preconditioned problem of:

$$\min_{\bar{\boldsymbol{x}} \in O_\delta, z \in \mathbb{R}} \boldsymbol{c}^T \bar{\boldsymbol{x}} + \bar{\lambda} \left(\bar{z} + \frac{1}{\delta} \mathbb{E}[(\frac{1}{h_\delta} \phi(\beta_\delta \bar{\boldsymbol{x}}, \boldsymbol{\xi}) - \bar{z})^+] \right) := \mathbb{E}[G(\bar{\boldsymbol{x}}, \bar{z}, \bar{\lambda}, \boldsymbol{\xi})]. \tag{11}$$

Due to the preconditioning, the distance from the inital iterate to the optimal point is of constant order as $\delta \to 0$. However, bounding the variance with $\mathbb{E}_{\boldsymbol{\xi}}[||\nabla_{(\bar{\boldsymbol{x}}, \bar{z})} G||_2^2] \leq 2||\boldsymbol{c}||^2 + \left(\frac{\bar{\lambda}}{\delta}\right)^2 (2 \left(\frac{\beta_\delta}{h_\delta}\right)^2 L^2 + 1)$ reveals a prohibitive $\frac{1}{\delta^2}$ term. Such a term limits the convergence rate of stochastic gradient descent procedures as $\delta \to 0$.

To resolve this issue, we employ importance sampling. As a review of importance sampling, consider calculating an expectation of function $f(\boldsymbol{x})$, where $\boldsymbol{x} \sim P$. if the function $f(\boldsymbol{x})$ has high dependence on $\boldsymbol{x}$ values which are near the "tails" of the distribution, the number of samples needed to get an accurate representation grows rapidly. To alleviate this problem, we can sample random variables $\boldsymbol{y}^{(i)}$ from a separate distribution Q and calculate $\mathbb{E}_P[f(\boldsymbol{x})]$ using the transformation: $\mathbb{E}_P[f(\boldsymbol{x})] = \int f(\boldsymbol{x}) P(dx) = \int f(\boldsymbol{x}) \frac{P(dx)}{Q(dx)} Q(dx)$. Properly tuning the alternative distribution Q can lead to a reduction in the variance of our measurement. For a basic review on importance sampling methods, see (Asmussen and Glynn 2007; Shortle and L'Ecuyer 2011). Importance sampling in the context of stochastic gradient descent has been used to improve the optimization of finite sum optimization (Needell et al. 2014). We introduce the following notion of efficiency for importance samplers.

Definition 2 An alternative measure Q is said to have bounded relative error with respect to the scaling parameter δ and event c_δ if

$$\limsup_{\delta \to 0} \frac{\mathbb{E}^Q(R^2 \mathbb{I}(c_\delta))}{\mathbb{E}^Q(R \mathbb{I}(c_\delta))^2} = C_I < \infty$$

where $R(\boldsymbol{x})$ is the likelihood ratio associated with Q for $\boldsymbol{x}$.

Algorithm 1: Stochastic Gradient Descent of $g(\bar{\boldsymbol{x}},\bar{z}) = \mathbb{E}[G(\bar{\boldsymbol{x}},\bar{z},\boldsymbol{\xi},\bar{\lambda})]$ with importance sampling

```
Initialize $\bar{\boldsymbol{x}}_0 \in \mathbb{R}^d_+, \bar{\boldsymbol{z}}_0 \in \mathbb{R}$ independent of $\delta$;
for $t = 1,2,...,T$ do
    Sample $\boldsymbol{\xi}_i$ through an importance sampler $Q$ that satisfies Assumptions 3;
    Calculate likelihood ratio $R(\boldsymbol{\xi}_i;\bar{\boldsymbol{x}}_k,\boldsymbol{z}_k,\delta)$;
    Calculate subgradients according to formulas;
        • $\tilde{\partial}_{\bar{\boldsymbol{x}}}G(\bar{\boldsymbol{x}},\bar{z},\boldsymbol{\xi};\delta,\lambda) = c + \frac{\lambda}{\delta}\frac{\beta_\delta}{h_\delta}\mathbb{I}(\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}},\xi) > \bar{z}) \cdot \nabla_{\bar{\boldsymbol{x}}}\phi(\beta_\delta\bar{\boldsymbol{x}},\xi) \cdot R(\boldsymbol{\xi}_i;\bar{\boldsymbol{x}}_k,\bar{z}_k,\delta)$
        • $\tilde{\partial}_{\bar{z}}G(\bar{\boldsymbol{x}},\bar{z},\boldsymbol{\xi};\delta,\lambda) = \lambda - \frac{\lambda}{\delta}\mathbb{I}(\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}},\boldsymbol{\xi}) > \bar{z}) \cdot R(\boldsymbol{\xi}_i;\bar{\boldsymbol{x}}_k,\bar{z}_k,\delta)$

    Set $\bar{\boldsymbol{x}}_k = \Pi_{\bar{O}}(\bar{\boldsymbol{x}}_{k-1} - \eta_t\tilde{\partial}_{\bar{\boldsymbol{x}}}G(\bar{\boldsymbol{x}}_k,\bar{z}_k,\boldsymbol{\xi}_{\boldsymbol{i}},\lambda)$ and $\bar{z}_k = \bar{z}_{k-1} - \eta_t\tilde{\partial}_{\bar{z}}G(\bar{\boldsymbol{x}}_k,\bar{z}_k,\boldsymbol{\xi}_{\boldsymbol{i}},\lambda)$;
Return $\hat{\bar{\boldsymbol{x}}} = \frac{1}{T}\sum_{k=1}^T\bar{\boldsymbol{x}}_k$, $\hat{\bar{z}} = \frac{1}{T}\sum_{k=1}^T\bar{z}_k$ and $\hat{f} = f(\beta_\delta\hat{\boldsymbol{x}})$
```

We begin by introducing the following assumption on our desired importance sampler.

Assumption 3 Suppose that for any points $\bar{\boldsymbol{x}} \in \mathbb{R}^m$ and $\boldsymbol{z} \in \mathbb{R}$, we can construct an importance sampling distribution $Q(\bar{\boldsymbol{x}}_k,\bar{z},\delta)$ and likelihood parameter $R(\boldsymbol{\xi}_{k+1};\bar{\boldsymbol{x}}_k,\bar{z},\delta)$ with bounded relative error for the event $\{\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}},\boldsymbol{\xi}) > \bar{z}\}$ as $\delta \to 0$.

Importantly, our importance sampler is allowed to be state-dependent; it depends on the current position $\boldsymbol{x}_k$, z_k. We can incorporate the importance sampler into our stochastic subgradient approach as shown in Algorithm 1. The results by He et al. (2024) analyze algorithms akin to Algorithm 1 by showing almost sure convergence to the optimal solution along with asymptotic normality that is independent of δ assuming bounded relative error.

To complete the theoretical analysis of our gradient descent procedure, we introduce a typical assumption of decaying step sizes:

Assumption 4 The step sizes in the stochastic gradient descent algorithm are of the form $\eta_t = \eta_0 t^{-\tau}$ for a constant η_0 and $0 \leq \tau < 1$.

For technical reasons, we also need to ensure that our solution is bounded away from 0 for any δ.

Assumption 5 For any $\delta > 0$ and $\lambda > 0$, the optimal value to the optimization program (10) is finite and positive.

With the assumptions set, we have the following theorem demonstrating that our preconditioned importance sampling based algorithm has no dependence on the rarity parameter δ in convergence.

Theorem 1 Assume that Assumptions 1-5 are enforced. Further suppose that β_δ and h_δ scale such that $\lim_{\delta\to 0}\frac{\beta_\delta}{h_\delta} < \infty$ Then, we have that there exists a computable δ_1 such that for all $\delta < \delta_1$ and all $\lambda > 0$, Algorithm 1 has the following relative error guarantee: For $f(\boldsymbol{x};\lambda) = \min_{\boldsymbol{x}\in O_\delta}\boldsymbol{c}^T\boldsymbol{x} + \lambda\mathrm{CVaR}_{1-\delta}\{\phi(\boldsymbol{x},\boldsymbol{\xi})\}$, we have $\frac{f(\beta_\delta\frac{1}{T}\sum_{i=1}^n\bar{\boldsymbol{x}}_k;\lambda) - f(\boldsymbol{x}_*,\lambda)}{f(\boldsymbol{x}_*,\lambda)} \leq \frac{K}{n^\tau}$ for a constant K independent of δ

Proof Sketch: There are three main ingredients in the proof method provided. Firstly, we must provide a new upper bound to

$$\mathbb{E}_Q[||\nabla_{(\bar{\boldsymbol{x}},\bar{z})}G(\bar{\boldsymbol{x}},\bar{z},\bar{\lambda},\boldsymbol{\xi})||^2] \leq 2||\boldsymbol{c}||^2 + 2\frac{\lambda^2}{\delta^2}\left[\left(\frac{\beta_\delta}{h_\delta}\right)^2 L^2 + 1\right]\mathbb{E}_Q[R(\boldsymbol{\xi},\bar{\boldsymbol{x}}_k,\bar{z}_k,\delta)^2\mathbb{I}(\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}}_k,\boldsymbol{\xi}) > \bar{z}_k)]. \tag{12}$$

Due to the assumption on the ratio of $\frac{\beta_\delta}{h_\delta}$, that term can be bounded by a finite constant for sufficiently small δ. With the bounded relative error assumption, we have that the $\mathbb{E}_Q[R(\boldsymbol{\xi},\bar{\boldsymbol{x}}_k,\bar{z}_k,\delta)^2\mathbb{I}(\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}},\boldsymbol{\xi}) > \bar{z})] \leq (C_I + 1)\mathbb{P}(\frac{1}{h_\delta}\phi(\beta_\delta\bar{\boldsymbol{x}}_k,\boldsymbol{\xi}) > \bar{z}_k)^2$ for sufficiently small δ. We next have to argue that for any possible iterate $\{\bar{\boldsymbol{x}}_k, z_k\}$, we have that $\mathbb{P}(\frac{1}{h_\delta}\phi(\beta_\delta \bar{\boldsymbol{x}}_k, \boldsymbol{\xi}) > \bar{z}_k) \leq C\delta$ for some constant C. Due to the scaling chosen for $\bar{\boldsymbol{x}}$ and $\bar{z}$ along with the compact nature of the set $\bar{O}$, this can be verified for both running examples. So, we have that $\mathbb{E}_Q[||\nabla_{(\bar{\boldsymbol{x}},\bar{z})} G(\bar{\boldsymbol{x}}, \bar{z}, \bar{\lambda}, \boldsymbol{\xi})||^2] \leq 2||\boldsymbol{c}||^2 + 2\lambda^2(L^2+1)(C_I+1)C^2 := G_T$ a bound which does not blow up as $\delta \to 0$. Using standard projected subgradient descent procedures (Duchi 2018), we have that

$$\begin{aligned}
\mathbb{E}[G(\frac{1}{n}\sum \bar{\boldsymbol{x}}_k, \frac{1}{n}\sum \bar{z}_k, \bar{\lambda}, \xi)] - \mathbb{E}[G(\bar{\boldsymbol{x}}^*, \bar{z}^*, \bar{\lambda}, \xi)] &\leq \frac{1}{2n}\left[\frac{||(\bar{\boldsymbol{x}}_0, \bar{z}_0) - (\bar{\boldsymbol{x}}_*, \bar{z}_*)||^2 + G_T \sum_{k=1}^n \eta_k^2}{\eta_n} + G_T^2 \sum_{i=1}^n \eta_k\right] \\
&\leq \frac{1}{2n}\left[\frac{||(\bar{\boldsymbol{x}}_0, \bar{z}_0) - (\bar{\boldsymbol{x}}_*, \bar{z}_*)||^2 + G_T^2 \int_0^n \eta_0 x^{-2\tau} dx}{\eta_n} + G_T^2 \int_0^n \eta_0 x^{-\tau} dx\right] \\
&\leq n^{-\tau} \underbrace{\left[||(\bar{\boldsymbol{x}}_0, \bar{z}_0) - (\bar{\boldsymbol{x}}_*, \bar{z}_*)||^2 + \frac{G_T^2 \eta_0}{1-2\tau} + G_T^2 \frac{\eta_0}{1-\tau}\right]}_{K}.
\end{aligned}$$

Due to the relation $f(\boldsymbol{x}, \lambda) = \beta_\delta \cdot \mathbb{E}[G(\bar{\boldsymbol{x}}, \bar{z}, \bar{\lambda}, \boldsymbol{\xi})]$, we have that

$$\frac{f(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i, \lambda) - f(\boldsymbol{x}_*, \lambda)}{f(\boldsymbol{x}_*, \lambda)} = \frac{\mathbb{E}[G(\frac{1}{n}\sum_{i=1}^n \bar{\boldsymbol{x}}_i, \bar{z}, \bar{\lambda}, \boldsymbol{\xi}) - G(\bar{\boldsymbol{x}}_i, \bar{z}_*, \bar{\lambda}, \boldsymbol{\xi})]}{\mathbb{E}[G(\bar{\boldsymbol{x}}_*, \bar{z}_*, \bar{\lambda}, \boldsymbol{\xi})]} \leq \frac{K}{n^\tau \mathbb{E}[G(\bar{\boldsymbol{x}}_*, \bar{z}_*, \bar{\lambda}, \boldsymbol{\xi})]}. \tag{13}$$

Finally, we argue that $\mathbb{E}[G(\bar{\boldsymbol{x}}_*, \bar{z}_*, \bar{\lambda}, \boldsymbol{\xi})]$ which is the optimal solution to the program in (11) has no dependence on δ. Due to the optimal critical scalings chosen in the previous section and Assumption 5, all terms in the minimization program are of constant order and the result follows.

4.1 Ascent Procedure For Optimal Lagrange Multiplier

The prior section provided an algorithm to optimize the minimization problem in (11) which is the inner minimization problem. The focus of this section is to combine this algorithm with a secondary ascent procedure to find the optimal value of $\bar{\lambda}$ that solves the maxmin game. By the envelope theorem, we have that $\nabla\psi(\bar{\lambda})$ where $\psi(\bar{\lambda}) = \min_{\bar{\boldsymbol{x}} \in \bar{O}} c^T \bar{\boldsymbol{x}} + \bar{\lambda}\mathrm{CVaR}_{1-\delta}\left\{\frac{1}{h_\delta}\phi(\beta_\delta \bar{\boldsymbol{x}}, \boldsymbol{\xi})\right\}$ is

$$\bar{z}^*(\lambda) + \frac{1}{\delta}\mathbb{E}\left[(\frac{1}{h_\delta}\phi(\beta_\delta \bar{\boldsymbol{x}}^*, \lambda) - \bar{z}^*(\lambda))_+\right] = \mathrm{CVaR}_{1-\delta}\left\{\frac{1}{h_\delta}\phi(\beta_\delta \bar{\boldsymbol{x}}^*, \boldsymbol{\xi})\right\}.$$

So, we can construct a projected sub-gradient ascent algorithm on $\psi(\bar{\lambda})$ to converge to the optimal $\bar{\lambda}^*$. This algorithmic procedure is highlighted in Algorithm 2. Importantly, to once again reduce the variance of the measurement of the CVaR term that guides the gradient ascent, we employ the same state dependent importance sampler from Assumption 3 at the convergence point achieved from Algorithm 1. The output of this procedure is a optimal solution to the *preconditioned* parameters $\bar{\boldsymbol{x}}$ and $\bar{\lambda}$ and an estimate of $f(\boldsymbol{x})$ which is computed through the critical scaling β_δ.

Algorithm 2: Stochastic Gradient Ascent of $\psi(\bar{\lambda})$ with importance sampling

```
Input $\delta$, $\lambda_0 > 0$, $M_l$, and $M_h$;
for $k = 0, 1, ..., T-1$ do
    Solve the inner minimization problem using Algorithm 1 with $\bar{\lambda} = \lambda_k$ to get $(\bar{\boldsymbol{x}}_k, \bar{z}_k)$;
    Sample $N$ points $\boldsymbol{\xi}_i$ from an importance sampler $Q$ that satisfies Assumption 3 at point $(\bar{\boldsymbol{x}}_k, \bar{z}_k)$;
    Calculate their associated likelihood parameters $R(\boldsymbol{\xi}_i, \bar{\boldsymbol{x}}_k, \boldsymbol{z}_k, \delta)$;
    Compute $\Delta_k = \bar{z}_k + \frac{1}{\delta}\frac{1}{N}\sum_{i=1}^N (\frac{1}{h_\delta}\phi(\beta_\delta \bar{\boldsymbol{x}}_k, \boldsymbol{\xi}) - \bar{z}_k)^+$;
    Set $\lambda_{k+1} = \Pi_{[M_l, M_h]}[\lambda_k + \eta_k \Delta_k]$
Return $\lambda_T$, $\bar{\boldsymbol{x}}_T$, and $\beta_\delta \mathbb{E}[G(\bar{\boldsymbol{x}}_T, \bar{z}_T, \bar{\lambda}, \xi)]$
```

5 NUMERICAL EXPERIMENTS

Due to space constraints, the numerical experiments presented in this version of the paper focus on the salvage fund problem. All computations were implemented in Python and the linear optimization program which defines $\phi(\boldsymbol{x},\boldsymbol{\xi})$ was solved using the CVXPY optimization package. The parameters of the problem were chosen as follows. The matrix $\boldsymbol{Q}$ was defined to be a $m*m$ matrix where $\boldsymbol{Q}_{ij}=0$ if $i=j$ and $1/m$ otherwise. The vector $\boldsymbol{m}$ has defined to be 1 for all $i\in[m]$. The losses $\boldsymbol{\xi}_i$ were set to be i.i.d Pareto random variables with scale parameter $\alpha=3$. For step sizes, η_0 was chosen to be $1/2$ and τ was chosen to be $2/3$. From the discussion of the critical scaling, we have $h_\delta=\beta_\delta\approx\delta^{-1/a}$ and $\kappa_\delta=1$ for the salvage fund problem. To be overly conservative, the upper bound on the feasible set O had C_h set to be 100.

5.1 Construction of Importance Sampler

To deploy our algorithm, we develop an importance sampling scheme for the event of interest: $\{\phi(\frac{\boldsymbol{x}}{\delta^{\frac{1}{a}}},\boldsymbol{\xi})>\frac{z}{\delta^{\frac{1}{a}}}\}$ that satisfies the conditions in Assumption 3. Since $\phi(\frac{\bar{\boldsymbol{x}}}{\delta^{1/a}},\boldsymbol{\xi})$ is a linear minimization problem, any feasible solution to the constraints given in $\phi(\frac{\bar{\boldsymbol{x}}}{\delta^{1/a}},\boldsymbol{\xi})$ would be an upper bound to the function itself. A necessary condition for the inequality $\phi(\frac{\bar{\boldsymbol{x}}}{\delta^{1/a}},\boldsymbol{\xi})>\frac{\bar{z}}{\delta^{1/a}}$ is that any feasible solution to the linear programming problem is also greater than $\frac{\bar{z}}{\delta^{1/a}}$. Specifically, the values $b=\max(\boldsymbol{\xi}_i-\boldsymbol{m}_i)$ and $y=0$ will always be a feasible solution to the constraints of the linear program presented in (4) with an objective value of $\max(\boldsymbol{\xi}_i-\boldsymbol{m}_i)$. So, we sample such that $\max(\boldsymbol{\xi}_i-\boldsymbol{m}_i)\geq\frac{\bar{z}}{\delta^{1/a}}$ to ensure all samples satisfy the necessary condition. We have that

$$\left\{\max(\boldsymbol{\xi}_i-\boldsymbol{m}_i)\geq\frac{\bar{z}}{\delta^{1/a}}\right\}\subset\bigcup_{i=1}^{d}\left\{\xi_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i\right\}. \tag{14}$$

With this defined, our importance sampler uses a mixture probability. Specifically, consider the probability distribution where with probability $\frac{P\left(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i\right)}{\sum_{i=1}^{d}P\left(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i\right)}$, we sample with $\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i$ through inversion sampling while all other components are drawn from the Pareto distribution. Conditioned on $\boldsymbol{\xi}_i$ being chosen, the random vector $\boldsymbol{\xi}$ has probability measure $\frac{\mathbb{I}(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i)P(dl)}{\mathbb{P}(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i)}$. Therefore the full proposal distribution is

$$Q(dl)=\sum_{i=1}^{d}\frac{\mathbb{I}(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i)P(dl)}{\sum_{i=1}^{d}\mathbb{P}(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i)}. \tag{15}$$

For this importance sampler, the likelihood ratio $R(\boldsymbol{\xi},w_k,\delta)$ is $\frac{\sum_{i=1}^{d}P\left(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i\right)}{\sum_{i=1}^{d}I\left(\boldsymbol{\xi}_i>\frac{\bar{z}}{\delta^{1/a}}+\boldsymbol{m}_i\right)}$. The sampler $Q(dl)$ and the associated likelihood ratio can be shown to be unbiased and have bounded relative error as $\delta\to 0$ thereby satisfying Assumption 3. Blanchet et al. (2019) contains the associated calculations similar sampling distributions albeit for normal distributions. Proving bounded relative error of such samplers for regularly varying distributions follow the same arguments with minor calculations differences.

5.2 Numerical Results for Inner Minimization

To test the efficiency of Algorithm 1 and its lack of dependence on δ, the projected subgradient descent algorithm was run with initial iterates $\bar{\boldsymbol{x}}_0$ and $\bar{z}_0$ to be $\mathbf{1}$. To first find an estimate of $\boldsymbol{x}^*$ and z^*, the algorithm was run until two consecutive iterates satisfied $|\bar{\boldsymbol{x}}_n-\bar{\boldsymbol{x}}_{n-1}|\leq 10^{-3}$ and $|z_n-z_{n-1}|\leq 10^{-3}$ for a fixed value of $\lambda=.8$. Once these optimal values were stabilized, 30000 importance samples were drawn to get a more accurate representation of the minimal objective function $\psi_\delta(\lambda)$. These values are recorded in the second column of Table 1.

To test relative error convergence, both algorithms were run until the optimization values reached within 5% of the optimal value recorded in Table 1. The number of iterations and time per iteration were recorded for both our preconditioned importance sampling method and a standard Monte Carlo SGD approach. To further reduce variance for small δ, batch gradient descent was employed for both algorithms. The number of samples for each batch was increased appropriately to allow for the standard SGD approach to converge. To estimate the CVaR risk value when testing convergence, an empirical estimate using 20000 importance samples was used if the function value was suspected to be close to the objective value (the samples used in the gradient descent iteration indicated an empirical objective function close to the $\psi_\delta(\lambda)$). The required iteration and computation complexity (time/samples) for each method are shown in Table 1.

Table 1: Iteration and time measurements to reach 5% error for the naive gradient descent algorithm and the preconditioned scaled importance sampling algorithm.

δ	$\psi_\delta(\lambda)$	Samples	Iterations (Importance)	Time (Importance)	Iterations (Naive)	Time (Naive)
10^{-2}	12.577	2000	38	36.2 sec	2	49.3 sec
10^{-3}	27.813	4000	33	82.3 sec	156	79.4 sec
10^{-4}	62.938	7500	32	100.6 sec	3681	94.5 sec
10^{-5}	139.377	15000	21	243.7 sec	22421	231.7 sec

The constant number of iterations in the convergence of the importance sampling scaled SGD method is evident from the table unlike the naive method which has a massive level of growth in the iterations needed as $\delta \to 0$. Both SGD methods take similar levels of time per iteration indicating no disadvantage in computation efficiency in the importance sampling scaled method used. This numerical results affirm our approach of efficiently optimizing the CVaR minimization problem as the importance sampling critically scaled SGD method displays a constant number of samples used to arrive at a solution which has a desired ε relative error. As a final comparison, the scaled importance sampling based SGD was compared to the naive unscaled SGD solution for a fixed $\delta = 10^{-4}$. Both methods were run until 5% relative error. The left plot of Figure 1 shows a comparison of the convergence of the relative values while the right plot shows an empirical CDF for the number of iterations needed for our scaled importance sampling SGD approach. We once again see the rapid convergence with our optimization procedure while a naive stochastic gradient descent approach faces two separate but important problems of taking too long to reach the neighborhood of the optimal solution and having significant noise when approaching the optimal solution. This matches our discussion about the runtime analysis regarding theorems projected gradient descent. The preconditioning allows us to immediately begin our optimization procedure in an appropriate neighborhood around the optimal budget and the importance sampling reduces the large variance in the gradients around the optimum. The right plot shows a empirical CDF of the number of iterations needed to reach 5% accuracy for 400 different runs of the scaled importance sampling based SGD. We see the majority of runs take on the order of 50-150 iterations of batch stochastic gradient descent while all runs took less than 250 iterations significantly less than the batch iterations needed for a naive SGD method.

5.3 Numerical Results of Ascent Procedure for Optimal Lagrange Multiplier

With numerical evidence of the convergence of the inner minimization problem of the Lagrangian problem for a fixed λ, we employ Algorithm 2 to converge to an optimal λ. The outer ascent procedure was run for $T = 200$ iterations and the inner optimization problem was run for 150 iterations. To speed up convergence, the values of $(\bar{\boldsymbol{x}}_k, \bar{z}_k)$ were used as the initial condition for the next run of the outer procedure. Table 2 shows the convergence to the optimal λ for varying values of δ along with an estimate of the optimal value of $\mathbf{1}^T \boldsymbol{x}_* + \lambda_* \mathrm{CVaR}_{1-\delta}\{\phi(\boldsymbol{x}_*, \boldsymbol{\xi})\} := f^{(n)}(\boldsymbol{x}, \delta)$. For all δ, we used 3000 samples to estimate Δ_k in Algorithm 2. We set $M_l = .1$ and $M_h = 2$ and initialized with $\lambda_0 = 1$.

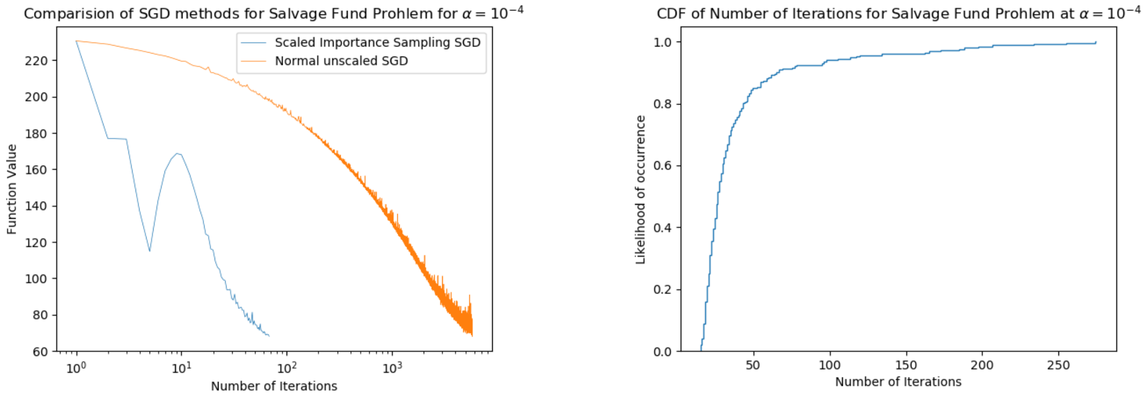

Figure 1: (left) A graph of the comparison of an example run of the number of iterations needed to reach 5% accuracy for a batch stochastic gradient descent method and our importance sampling based SGD method for the salvage fund problem. The number of iterations required to reach the desired ε accuracy is orders of magnitude smaller. (right) An empirical CDF of the number of iterations needed to each 5% accuracy of the true solution for the salvage fund problem with $\delta = 10^{-4}$.

Table 2: Numerical Results for Ascent Procedure to optimize λ for Salvage Fund Problem.

δ	Optimal λ	$f^{(n)}(\boldsymbol{x}, \delta)$
10^{-2}	.726	18.424
10^{-3}	.883	41.339
10^{-4}	.904	87.316
10^{-5}	.876	218.366

As δ approaches 0, the optimal λ seemed to converge to same neighborhood. This matches the epi-convergence observation that the limiting optimization problem had a unique solution under the conditions, the optimal Lagrange multiplier will converge to that value.

6 CONCLUSION AND DISCUSSION

We develop a critically scaled importance sampling based stochastic gradient descent approach to solve CVaR constrained optimization problems within the rare event regime efficiently. Through our two guiding examples, we demonstrate a end-to-end procedure to develop both the necessary critical scalings needed for the algorithm through epi-convergence arguments and the required importance sampling assumptions. A rather interesting research avenue includes using this initial optimization procedure for the CVaR constrained problem to achieve optimization values and feasible solutions for the original probabilistic CC-OPT.

ACKNOWLEDGMENTS

The authors are grateful to the comments and suggestions from the reviewers. This research is partly supported by the Air Force Office of Scientific Research under award number FA9550-20-1-0397 and ONR N000142412655. Support from NSF 2229012, 2312204, 2403007 is also gratefully acknowledged. The authors would also like to thank The Burt and Deedee McMurtry Fellowship for supporting their research.

REFERENCES

Asmussen, S. and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Stochastic Modelling and Applied Probability. Springer New York.

Blanchet, J., J. Jorritsma, and B. Zwart. 2024. "Optimization under Rare Events: Scaling Laws for Linear Chance-constrained Programs". *arXiv preprint arXiv: 2407.11825*.

Blanchet, J., J. Li, and M. K. Nakayama. 2019. "Rare-Event Simulation for Distribution Networks". *Operations Research* 67(5):1383–1396.

Blanchet, J., F. Zhang, and B. Zwart. 2024. "Efficient Scenario Generation for Heavy-Tailed Chance Constrained Optimization". *Stochastic Systems* 14(1):22–46.

Calafiore, G. C. and M. C. Campi. 2005. "Uncertain convex programs: randomized solutions and confidence levels". *Mathematical Programming* 102:25–46.

Duchi, J. C. 2018, November. "Introductory lectures on stochastic optimization". In *IAS/Park City Mathematics Series*, edited by M. Mahoney, J. C. Duchi, and A. Gilbert, Volume 25, 99–185. Providence, Rhode Island: American Mathematical Society.

He, S., G. Jiang, H. Lam, and M. C. Fu. 2024. "Adaptive Importance Sampling for Efficient Stochastic Root Finding and Quantile Estimation". *Operations Research* 72(6):2612–2630.

Luedtke, J. and S. Ahmed. 2008. "A Sample Approximation Approach for Optimization with Probabilistic Constraints". *SIAM Journal on Optimization* 19(2):674–699.

Luedtke, J., S. Ahmed, and G. L. Nemhauser. 2010, Apr. "An integer programming approach for linear programs with probabilistic constraints". *Mathematical Programming* 122(2):247–272.

Moulines, E. and F. Bach. 2011. "Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning". In *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Volume 24: Curran Associates, Inc.

Needell, D., R. Ward, and N. Srebro. 2014. "Stochastic Gradient Descent, Weighted Sampling, and the Randomized Kaczmarz algorithm". In *Advances in Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Volume 27: Curran Associates, Inc.

Nemirovski, A. and A. Shapiro. 2006. "Scenario Approximations of Chance Constraints". In *Probabilistic and Randomized Methods for Design under Uncertainty*, edited by G. Calafiore and F. Dabbene, 3–47. London: Springer London.

Nemirovski, A. and A. Shapiro. 2007. "Convex approximations of chance constrained programs". *SIAM Journal on Optimization* 17(4):969–996.

Resnick, S. I. 1987. *Extreme Values, Regular Variation and Point Processes*. New York, NY: Springer New York.

Robbins, H. and S. Monro. 1951. "A Stochastic Approximation Method". *The Annals of Mathematical Statistics* 22(3):400 – 407.

Rockafellar, R. T. and S. Uryasev. 2000. "Optimization of conditional value-at-risk". *Journal of Risk* 2:21–42.

Rockafellar, R. T. and S. Uryasev. 2002. "Conditional value-at-risk for general loss distributions". *Journal of Banking & Finance* 26(7):1443–1471.

Rockafellar, R. T. and R. J.-B. Wets. 1998. *Variational Analysis*. Heidelberg: Springer-Verlag.

Shortle, J. F. and P. L'Ecuyer. 2011. *Introduction to Rare-Event Simulation*. John Wiley & Sons, Ltd.

Sion, M. 1958. "On general minimax theorems". *Pacific J. Math.* 8:171–176.

Tong, S., A. Subramanyam, and V. Rao. 2022. "Optimization under Rare Chance Constraints". *SIAM Journal on Optimization* 32(2):930–958.

AUTHOR BIOGRAPHIES

ANISH SENAPATI is Ph.D. student at the Institute of Computational and Mathematical Enginnering at Stanford University. His research interests include stochastic optimization, stochastic simulation, and rare-event simulation. His email address is asenapat@stanford.edu

FAN ZHANG was a Ph.D. candidate in the Department of Management Science and Engineering at Stanford University. His research interests include applied probability, stochastic simulation and robust optimization. He now works in the finance industry. His email is fzh@stanford.edu.

JOSE BLANCHET is a Professor of MS&E at Stanford University. He holds a Ph.D. in Operations Research from Stanford University. He has research interests in applied probability and Monte Carlo methods. His email address is jose.blanchet@stanford.edu and his website is https://web.stanford.edu/jblanche/.

BERT ZWART is a group leader at the Center for Mathematics and Computer Science (CWI) in Amsterdam, and professor at Eindhoven University of Technology. Bert's research expertise is in applied probability and stochastic networks, and applications in communication and energy networks. His email is bert.zwart@cwi.nl and his website is https://www.cwi.nl/en/people/bert-zwart/.