A SIGMOIDAL APPROXIMATION FOR CHANCE-CONSTRAINED NONLINEAR PROGRAMS

YANKAI CAO AND VICTOR M. ZAVALA *

Abstract. We propose a sigmoidal approximation (SigVaR) for the value-at-risk (VaR) and we use this approximation to tackle nonlinear programming problems (NLPs) with chance constraints. We prove that the approximation is conservative and that the level of conservatism can be made arbitrarily small for limiting parameter values. The SigVar approximation brings computational benefits over exact mixed-integer and difference of convex functions reformulations because its sample average approximation can be cast as a standard NLP. Unfortunately, as with any sigmoidal function, SigVaR becomes numerically unstable in the limit of its parameter values. To ameliorate this issue, we propose a scheme that solves a sequence of approximations of increasing quality. We also establish conditions under which SigVaR is less conservative than the well-known conditional value at risk (CVaR) and Bernstein approximations and we use this result to initialize the proposed scheme. We conduct small- and large-scale numerical studies to demonstrate the benefits and limitations of the proposed approximation.

Key words. nonlinear optimization, chance constraints, large-scale, approximation

AMS subject classifications. 90C15, 90C30, 90C55

1. Problem Definition and Setting.

We study the chance-constrained nonlinear program (CC-P):

$$\min_{x \in \mathcal{X}} \ \varphi(x) \tag{1.1a}$$

$$\text{s.t.} \ \mathbb{P}\left(f(x,\Xi) \le 0\right) \ge 1-\alpha. \tag{1.1b}$$

Here, $x \in \mathbb{R}^n$ are decision variables and the objective function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and potentially nonconvex. The set $\mathcal{X} := \{\, x \,|g(x) \ge 0\}$ is assumed to be compact and non-empty and is comprised of twice differentiable and potentially nonconvex constraints $g : \mathbb{R}^n \to \mathbb{R}^m$. We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we assume that Ω is a measurable space equipped with σ-algebra $\mathcal{F}$ of subsets of Ω, and that $\bar{\Xi}$ is a linear space of $\mathcal{F}$-measurable functions $\Xi : \Omega \to \mathbb{R}^d$ (random variables). The probability measure function is given by $\mathbb{P} : \mathcal{F} \to [0,1]$ and we use $\xi \in \mathbb{R}^d$ to denote realizations of Ξ. The scalar constraint function $f : \mathbb{R}^n \times \bar{\Xi} \to \mathbb{R}$ is also assumed to be twice continuously differentiable and potentially nonconvex. We define the scalar random variable $Z := f(x,\Xi)$ with realizations $z \in \mathbb{R}$. When appropriate, we use the notation $Z(x)$ to highlight the dependence of the random variable Z on the decision x. We make the blanket assumption that Ξ and $Z(x)$ (for all $x \in \mathcal{X}$) are continuous random variables. We highlight special considerations for discrete random variables when appropriate. We use $\mathbb{P}(Z \in D)$ to denote the probability of the event $Z \in D$ and recall that $\mathbb{P}(Z \in D) = \int_D p_Z(z)dz$ where $D \subseteq \mathbb{R}$ and $\mathbb{P}(Z \in (-\infty, t]) = \int_{-\infty}^t p_Z(z)dz = F_Z(t)$ for $t \in \mathbb{R}$. Here, $p_Z : \mathbb{R} \to [0,\infty)$ and $F_Z : \mathbb{R} \to [0,1]$ are the density and cumulative density functions of Z, respectively.

The CC (1.1b) requires that the event $\{f(x,\Xi) \in (-\infty, 0]\}$ occurs with probability of at least $1-\alpha$, where $\alpha \in (0,1]$. Since $\mathbb{P}(Z \le 0) = F_Z(0)$, the CC can also be written as $F_{f(x,\Xi)}(0) \ge 1-\alpha$ or $1 - F_{f(x,\Xi)}(0) \le \alpha$. We recall that the $(1-\alpha)$-quantile of Z is $Q_Z(1-\alpha) = \mathrm{VaR}_{1-\alpha}(Z) := \arg\min_{t \in \mathbb{R}}\{F_Z(t) \ge 1-\alpha\}$ (where

*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Dr, Madison, WI 53706, USA. This work was partially supported by the U.S. Department of Energy grant DE-SC0014114. (victor.zavala@wisc.edu).

VaR is known as the value-at-risk). Consequently, the CC can also be written as $\mathrm{VaR}_{1-\alpha}(f(x,\Xi)) \leq 0$. Another important observation is that $\mathbb{E}[1_D(Z)] = \mathbb{P}(Z \in D)$ holds, where $1_D : \mathbb{R} \to \{0,1\}$ denotes the indicator function of set D (i.e., $1_D(Z) = 1$ if $Z \in D$ and $1_D(Z) = 0$ if $Z \notin D$). Consequently, (1.1b) can be written as $\mathbb{E}[1_{(-\infty,0]}(f(x,\Xi))] \geq 1-\alpha$ or, equivalently, as $\mathbb{E}[1_{(0,\infty)}(f(x,\Xi))] \leq \alpha$. We define the feasible set of CC-P as $\mathcal{X}(\alpha) := \mathcal{X} \cap \mathcal{P}(\alpha)$, where $\mathcal{P}(\alpha) := \{\, x \,|\, \mathbb{P}\,(f(x,\Xi) \leq 0) \geq 1-\alpha\}$ and we assume $\mathcal{X}(\alpha)$ to be compact and non-empty for all $\alpha \in (0,1]$. We denote an optimal solution and objective value of (1.1) as $x^*(\alpha)$ and $\varphi^*(\alpha)$, respectively. We focus our attention on NLPs with a single CCs but the concepts discussed can also be applied to multiple single CCs of the form $\mathbb{P}\,(f_i(x,\Xi) \leq 0) \geq 1-\alpha_i,\ i = 1,...,r$.

A distinguishing and challenging feature of CC-P is that it cannot be solved exactly (except for certain simplified settings). Settings that admit exact solutions include those in which the quantile $Q_{f(x,\Xi)}(1-\alpha)$ can be expressed in algebraic form (e.g., the constraint function $f(\cdot,\cdot)$ is linear in both arguments and the random data vector is Gaussian [2]) or cases in which the cumulative density $F_{f(x,\Xi)}(\cdot)$ and its derivatives can be computed explicitly [16]. Exact reformulations with integer variables, originally proposed in [12], use an indicator function representation of the CC. Unfortunately, in the context of CC-P, the integer reformulation would lead to large-scale and nonconvex mixed-integer nonlinear programs (MINLPs). Conservative and computationally more tractable approximations of CC-P can be used to avoid the need for solving MINLPs. A conservative approximation can be obtained by using the so-called scenario-based approach [14, 4]. In this approach, we solve a stochastic NLP that enforces $f(x,\Xi) \leq 0$ with probability one (almost surely). Such an approach leads to structured NLPs, which can in turn be solved using parallel interior-point solvers [10]. A drawback of the scenario approach is that it can be overly conservative and does not offer direct control on the probability level of CC. Alternative conservative approximations include the conditional value-at-risk (CVaR) approximation and the Bernstein approximation, which use convex approximations of the indicator function [13]. The authors in [8] propose a difference of convex functions (DC) approximation for the indicator function and they show that the approximation can be made equivalent to CC-P. This approach, however, requires of specialized solution algorithms that are not guaranteed to work in a general nonconvex NLP setting.

In this work, we propose an approximation for CC (1.1) that uses a tailored sigmoidal function to outer-approximate the indicator function. We use this sigmoidal function to construct a risk measure, that we call SigVaR, and show that this is a conservative approximation of the value at risk (VaR). We prove that the SigVaR approximation is always conservative and that it can be made equivalent to CC-P for limiting parameter values. We also show that the approximation can be made less conservative than the CVaR and Bernstein approximations and we establish a connection with the DC approximation. A benefit of the SigVaR approximation is that it can be handled by using standard NLP solvers, thus offering parallel solution capabilities. As with most sigmoidal functions, however, a drawback of SigVaR is that numerical stability is encountered as the approximation approaches the indicator function. To ameliorate this issue, we propose a scheme that solves a sequence of approximations of increasing quality. Small and large case studies are used to illustrate the concepts and demonstrate performance.

The paper is organized as follows. Section 2 introduces basic nomenclature and reviews CVaR, Bernstein, and DC approximations. Section 3 introduces the SigVaR approximation and establishes properties. Section 4 outlines a numerical scheme to solve a sequence of SIgVaR approximations. Section 5 provides numerical studies. Final remarks are provided in section 6.

2. Review on CC Approximations.

We review approaches to deal with CC-P (1.1b) in order to introduce some basic concepts.

2.1. CVaR Approximation.

Because $\mathbb{P}(Z > 0) = \mathbb{E}[1_{(0,\infty)}(Z)]$, the CC can be expressed as $\mathbb{P}(f(x,\Xi) > 0) \leq \alpha$, and we can use the equivalent formulation:

$$\mathbb{E}[1_{(0,\infty)}(f(x,\Xi))] \geq \alpha. \tag{2.2}$$

A computationally practical approach to approximate the CC is to find a *conservative approximation*. This is done by finding an approximating function $\psi : \mathbb{R} \to \mathbb{R}$ satisfying $\psi(z) \geq 1_{[0,\infty)}(z) \geq 1_{(0,\infty)}(z)$ for any $z \in \mathbb{R}$. For such a function we have that $\psi(t^{-1}z) \geq 1_{[0,\infty)}(t^{-1}z) = 1_{(0,\infty)}(z)$ for any parameter $t > 0$. Consequently,

$$\mathbb{E}[\psi(t^{-1}Z)] \geq \mathbb{P}(Z > 0). \tag{2.3}$$

We can thus conclude that the satisfaction of the constraint:

$$\mathbb{E}[\psi(t^{-1}Z)] \leq \alpha, \tag{2.4}$$

implies that $\mathbb{P}(Z > 0) \leq \alpha$ is satisfied (and so does $\mathbb{P}(f(x,\Xi) \leq 0) \geq 1-\alpha$). Because (2.4) is valid for all $t > 0$ we also have, if $\psi(\cdot)$ is convex, that:

$$\inf_{t>0}\{t\,\mathbb{E}[\psi(t^{-1}Z)] - t\,\alpha\} \leq 0 \tag{2.5}$$

implies $\mathbb{P}(Z > 0) \leq \alpha$. The quality of the conservative approximation depends on the choice of the approximating function $\psi(\cdot)$. The choice $\psi(z) := [1+z]_+$ with $[z]_+ := \max\{z, 0\}$ leads to the approximation:

$$\inf_{t>0}\{\mathbb{E}\left[[Z+t]_+\right] - t\alpha\} \leq 0. \tag{2.6}$$

It can be shown that $\inf_{t>0}$ can be replaced with $\inf_t$ to obtain:

$$\inf_{t\in\mathbb{R}}\left\{\alpha^{-1}\mathbb{E}\left[[Z+t]_+\right] - t\right\} \leq 0. \tag{2.7}$$

By redefining $t \leftarrow -t$ and recalling that $\mathrm{CVaR}_{1-\alpha}(Z) := \inf_{t\in\mathbb{R}}\left\{t + \alpha^{-1}\mathbb{E}\left[[Z-t]_+\right]\right\}$, we can see that (2.7) can be used to derive a conservative approximation of CC-P (1.1) of the form:

$$\min_{x\in\mathcal{X}} \; \varphi(x) \tag{2.8a}$$

$$\text{s.t.} \; \mathrm{CVaR}_{1-\alpha}(f(x,\Xi)) \leq 0. \tag{2.8b}$$

We denote an optimal objective value and solution of this problem (which we call CVaR-P) as $\varphi_c(\alpha)$ and $x_c(\alpha)$, respectively. We define the feasible set of CVaR-P $\mathcal{X}_c(\alpha)$ and note, because CVaR provides a conservative approximation, that $\mathcal{X}_c(\alpha) \subseteq \mathcal{X}(\alpha)$. Consequently, any feasible solution $x_c(\alpha)$ of CVaR-P is feasible for CC-P (1.1). This also implies that $\varphi_c(\alpha) \geq \varphi(\alpha)$ for all $\alpha \in (0, 1]$.

We define $Z_c(\alpha) := f(x_c(\alpha), \Xi)$ and recall that [15]:

$$\mathrm{VaR}_{1-\alpha}(Z_c(\alpha)) = \arg\min_t \left\{t + \alpha^{-1}\mathbb{E}\left[[Z_c(t) - t]_+\right]\right\}, \tag{2.9}$$

and thus $\mathrm{VaR}_{1-\alpha}(Z_c(\alpha)) \leq \mathrm{CVaR}_{1-\alpha}(Z_c(\alpha))$. This observation also highlights that CVaR provides a conservative approximation for the CC.

Crucial to our results is the constant:

$$\gamma_\alpha := -t_c(\alpha)^{-1}. \tag{2.10}$$

with $t_c(\alpha) \in \arg\min_t\{t + \alpha^{-1}\mathbb{E}[Z_c(t) - t]_+\}$. From (2.5) with $\psi(z) = [1 + z]_+$, we have that (2.8b) implies $\mathbb{E}[[\gamma_\alpha Z_c(\alpha) + 1]_+] \leq \alpha$. We now show that we can always find a $t_c(\alpha) < 0$ (equivalently $\gamma_\alpha > 0$) at any $x_c(\alpha)$. Since (2.8b) is satisfied at $x_c(\alpha)$, we have that either $\mathrm{CVaR}_{1-\alpha}(Z_c(\alpha)) < 0$, which implies $\mathrm{VaR}_{1-\alpha}(Z_c(\alpha)) < 0$ it follows that $\gamma_\alpha > 0$ with $t_c(\alpha) = \mathrm{VaR}_{1-\alpha}(Z_c(\alpha))$. If $\mathrm{CVaR}_{1-\alpha}(Z_c(\alpha)) = 0$ we can have $\mathrm{VaR}_{1-\alpha}(Z_c(\alpha)) < 0$ (for which we have already established that $\gamma_\alpha > 0$ exists) or $\mathrm{VaR}_{1-\alpha}(Z_c(\alpha)) = 0$. In the later case we have $\mathbb{E}\left[[Z_c(\alpha)]_+\right] = 0$ and thus $\mathrm{argmin}_t\{t + \alpha^{-1}\mathrm{E}[[Z_c(\alpha) - t]_+]\} = \mathbb{R}$ and thus one can pick any $t_c(\alpha) < 0$ such that $\gamma_\alpha > 0$.

A key advantage of the CVaR approximation is that it can be cast as a standard NLP. Moreover, if $f(x, \xi)$ is convex in x for given ξ, CVaR is also convex in x. One can also prove that the function $\psi(z) = [1 + z]_+$ is the *tightest convex approximation* of $1_{[0,\infty)}(z)$. Despite these benefits, the CVaR approximation can be quite conservative. Moreover, the CVaR approximation does not offer a mechanism to enforce convergence to a solution of CC-P.

2.2. Bernstein Approximation. If we use the function $\psi(z) = e^z$, (2.4) takes the form $\mathbb{E}[e^{t^{-1}Z}] \leq \alpha$. For $t > 0$ this is equivalent to,

$$t \log(\mathbb{E}[e^{t^{-1}Z}]) \leq t \log(\alpha). \tag{2.11}$$

Because this relationship is valid for all $t > 0$, we can also conclude that:

$$\inf_{t>0}\{t \log(\mathbb{E}[e^{t^{-1}Z}]) - t \log(\alpha)\} \leq 0 \tag{2.12}$$

which is called Bernstein approximation. From the definition of entropic value-at-risk (EVaR) [1]:

$$\mathrm{EVaR}_{1-\alpha}(Z) := \inf_{t>0} \left\{t^{-1}\log\left(\alpha^{-1}\mathbb{E}[e^{tZ}]\right)\right\}, \tag{2.13}$$

and it is thus easy to see that we can use (2.12) is equivalent to:

$$\mathrm{EVaR}_{1-\alpha}(Z) \leq 0. \tag{2.14}$$

This conservative approximation can be handled using standard NLP techniques. Moreover, if $f(x, \Xi)$ is convex in x for given ξ, EVaR is also convex in x. Unfortunately, one can prove that EVaR is even more conservative than CVaR. This follows from $\mathrm{VaR}_{1-\alpha}(Z) \leq \mathrm{CVaR}_{1-\alpha}(Z) \leq \mathrm{EVaR}_{1-\alpha}(Z)$ [1].

2.3. DC Approximation. In [8] it is shown that the indicator function can be approximated by using a difference of convex functions. The DC approximation of $\mathbb{P}(f(x, \Xi) > 0) \leq \alpha$ has the form:

$$\inf_{t>0} \ \frac{1}{t}\mathbb{E}\left[\psi(f(x, \Xi), t) - \psi(f(x, \Xi), 0)\right] \leq \alpha \tag{2.15}$$

where $\psi(z,t) := [z+t]_+$. This DC approximation is exact in the limit $t \to 0$ and one can approximate the limit by using a small perturbation of the form:

$$\epsilon^{-1}\mathbb{E}\left[\psi(f(x,\Xi),\epsilon) - \psi(f(x,\Xi),0)\right] \leq \alpha. \tag{2.16}$$

By using approximation (2.16) instead of (1.1b), we obtain problem DC-P. In [8] it is shown that DC-P is equivalent to CC-P for $\epsilon \to 0$. A practical limitation of DC-P is its sample average approximation (SAA) cannot be cast as a standard NLP, due to the difference of max functions. Consequently, tailored algorithms are needed [8].

3. Sigmoidal Approximation.

Our work is motivated by the observation that the indicator function can be approximated by using a *sigmoid function* of the form:

$$\psi_s^{\mu,\tau}(z) := \frac{1+\mu}{\mu + e^{-\tau z}}, \tag{3.17}$$

where $\mu, \tau \in \mathbb{R}_+$ are parameters. We note that this sigmoid function is a special case of the generalized logistic function, which is a standard approximation function for the indicator function [6]. The associated approximate CC constraint takes the form:

$$\mathbb{E}\left[\psi_s^{\mu,\tau}(f(x,\Xi))\right] \leq \alpha. \tag{3.18}$$

In this work, we consider a *tailored variant of the above sigmoid function* of the form:

$$\psi_{ss}^{\mu,\tau}(z) := \left[2\frac{1+\mu}{\mu + e^{-\tau z}} - 1\right]_+. \tag{3.19}$$

This gives the approximate CC constraint:

$$\mathbb{E}\left[\psi_{ss}^{\mu,\tau}(f(x,\Xi))\right] \leq \alpha. \tag{3.20}$$

The motivation behind the tailored variant is illustrated in Figure 1, where we can see that the variant is more accurate than the standard counterpart. This is because the max function sets $\psi_{ss}^{\mu,\tau}(z) = 0$ for all $z \leq -\delta$ and where $\delta := \frac{1}{\tau}\log(2+\mu)$.

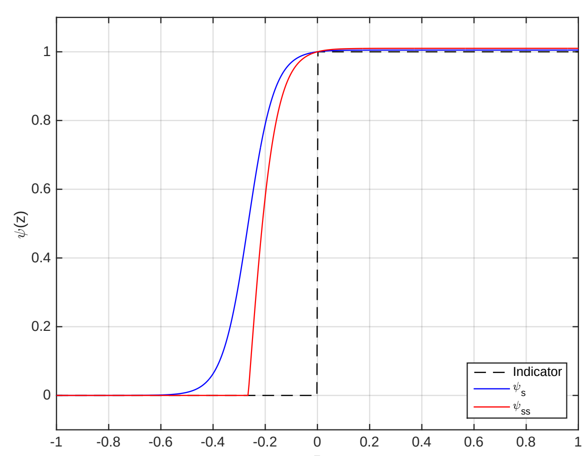

Fig. 1. *Comparison of standard and tailored sigmoid functions.*

We now show that sigmoid functions provide natural conservative approximations for chance constraints.

THEOREM 1. *The constraints* (3.18) *and* (3.20) *are conservative approximations of the CC* (1.1b) *for any* $\mu, \tau \in \mathbb{R}_+$, $\alpha \in (0, 1]$, *and* $x \in \mathcal{X}$.

Proof. Consider the random variable $Z = f(x, \Xi)$ with instances $z \in \mathbb{R}$. Since $e^{-\tau z} \geq 0$ holds for $z \in \mathbb{R}$ and $e^{-\tau z} \leq 1$ holds for $z \in \mathbb{R}_+$ we have that $\frac{1+\mu}{\mu+e^{-\tau z}} \geq 0$ for any $z \in \mathbb{R}$ and $\frac{1+\mu}{\mu+e^{-\tau z}} \geq 1$ holds for $z \in \mathbb{R}_+$. We thus have that $2\frac{1+\mu}{\mu+e^{-\tau z}} - 1 \geq 1$ holds for $z \in \mathbb{R}_+$. Therefore, $\frac{1+\mu}{\mu+e^{-\tau z}} \geq \mathbb{1}_{[0,\infty)}(z)$ and $\psi_{ss}^{\mu,\tau}(z) \geq \mathbb{1}_{[0,\infty)}(z)$ for any $z \in \mathbb{R}_+$. Consequently, $\mathbb{E}[\frac{1+\mu}{\mu+e^{-\tau Z}}] \geq \mathbb{E}[\mathbb{1}_{[0,\infty)}(Z)] \geq \mathbb{P}(Z > 0)$ and $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z)] \geq \mathbb{E}[\mathbb{1}_{[0,\infty)}(Z)] \geq \mathbb{P}(Z > 0)$. The result follows. ∎

We use the tailored sigmoid function to define the Sigmoidal Value-at-Risk (SigVaR):

$$
\begin{aligned}
\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) &:= \inf_{t\in\mathbb{R}} \left\{\mathbb{E}\left[\psi_{ss}^{\mu,\tau}(Z-t)\right] \leq \alpha\right\} \\
&= \inf_{t\in\mathbb{R}} \left\{\mathbb{E}\left[\left[2\frac{1+\mu}{\mu+e^{-\tau(Z-t)}} - 1\right]_+\right] \leq \alpha\right\}. \qquad (3.21)
\end{aligned}
$$

and we use this to formulate the problem:

$$
\begin{aligned}
&\min_{x\in\mathcal{X}} \; \varphi(x) && (3.22\text{a})\\
&\text{s.t. } \mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(f(x,\Xi)) \leq 0. && (3.22\text{b})
\end{aligned}
$$

We define an optimal objective and solution of (3.22) as $\varphi_{ss}^{\mu,\tau}(\alpha)$ and $x_{ss}^{\mu,\tau}(\alpha)$. We also define the feasible set of (3.22) as $\mathcal{X}_{ss}^{\mu,\tau}(\alpha)$. From Theorem 1, it is clear that $\mathcal{X}_{ss}^{\mu,\tau}(\alpha) \subseteq \mathcal{X}(\alpha)$ for all $\mu, \tau \in \mathbb{R}_+$. This implies that $\varphi_{ss}^{\mu,\tau}(\alpha) \geq \varphi(\alpha)$ for all $\alpha \in (0, 1]$ and $\mu, \tau \in \mathbb{R}_+$.

The definition of SigVaR is motivated by the observation that $\mathrm{VaR}_{1-\alpha}(Z) = \arg\min_{t\in\mathbb{R}}\{\mathbb{P}(Z-t > 0) \leq \alpha\}$ can also be expressed in terms of the indicator function:

$$
\mathrm{VaR}_{1-\alpha}(Z) = \arg\min_{t\in\mathbb{R}} \{\mathbb{E}\left[\mathbb{1}_{(0,\infty)}(Z-t)\right] \leq \alpha\}. \qquad (3.23)
$$

Because we have established that the sigmoid function $\psi_{ss}^{\mu,\tau}(\cdot)$ is a conservative approximation of $\mathbb{1}_{(0,\infty)}(\cdot)$, we have that $\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) \geq \mathrm{VaR}_{1-\alpha}(Z)$. Consequently, SigVaR can be interpreted as an approximate quantile and (3.22) is an conservative VaR representation of CC-P. As in the case of the VaR representation of CC-P, problem (3.22) is not particularly attractive for computation. However, this problem also has the following equivalent representation (we call this SigVaR-P):

$$
\begin{aligned}
&\min_{x\in\mathcal{X}} \; \varphi(x) && (3.24\text{a})\\
&\text{s.t. } \mathbb{E}\left[\psi_{ss}^{\mu,\tau}(f(x,\Xi))\right] \leq \alpha. && (3.24\text{b})
\end{aligned}
$$

In Section 4 we will show that the SAA approximation of SigVaR-P can be cast as a standard NLP. To show that (3.24) and (3.22) are equivalent, we make the following observations. If $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z)] \leq \alpha$ is satisfied then it implies that $t = 0$ satisfies $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z-t)] \leq \alpha$, and since $\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z)$ is the smallest t satisfying $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z-t)] \leq \alpha$, then $\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) \leq 0$. On the other hand, if $\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) \leq 0$ is satisfied, according to the definition, $t = \mathrm{SigVaR}_{1-\alpha}^{\mu,\tau}(Z)$ satisfies $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z-t)] \leq \alpha$. Since $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z-t)]$ is a decreasing function of t, then $t = 0$ also satisfies $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z-t)] \leq \alpha$ and thus $\mathbb{E}[\psi_{ss}^{\mu,\tau}(Z)] \leq \alpha$.

3.1. Relationship with CC-P. We now proceed to show that SigVaR-P becomes an exact approximation of DC-P in the limit of its parameter values. For the random variable $Z(x) = f(x, \Xi)$ with $x \in \mathcal{X}$, we define the SigVaR-CC approximation error is defined as:

$$\epsilon_{\mu,\tau}(x) := \mathbb{E}\left[\psi_{ss}^{\mu,\tau}(Z(x))\right] - \mathbb{P}(Z(x) > 0). \tag{3.25}$$

From Theorem 1 we have that $\epsilon_{\mu,\tau}(x) \geq 0$ for all $\mu, \tau \in \mathbb{R}_+$ because SigVaR is a conservative approximation of CC for all $x \in \mathcal{X}$.

We proceed to establish a bound for the SigVaR-CC approximation error. We note that, since $Z(x)$ is a continuous random variable, its density is measurable and bounded (i.e., $p_{Z(x)} : \mathbb{R} \to [0, \infty)$). We thus have that a constant $L(x) := \sup_{z\in\mathbb{R}}\{p_{Z(x)}(z)\} \in (0, \infty)$ exists and satisfies $\mathbb{P}(-\delta \leq Z(x) \leq 0) = \int_{-\delta}^{0} p_{Z(x)}(z)dz \leq \int_{-\delta}^{0} L(x)dz = L(x)\delta$ for all $x \in \mathcal{X}$.

LEMMA 3.1. *The SigVaR-CC error is bounded as* $\epsilon_{\mu,\tau}(x) \leq \frac{\log(2+\mu)L(x)}{\tau} + \frac{2}{\mu}$ *for all* $x \in \mathcal{X}$.

Proof. We can establish the following sequence of implications:

$$\begin{aligned}
\epsilon_{\mu,\tau}(x) &= \int_{-\infty}^{\infty} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz - \int_{0}^{\infty} p_Z(z)dz \\
&= \int_{-\infty}^{0} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz + \int_{0}^{\infty} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz - \int_{0}^{\infty} p_Z(z)dz \\
&= \int_{-\infty}^{0} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz + \int_{0}^{\infty} \left(2\frac{1+\mu}{\mu+e^{-\tau Z}} - 1\right) p_Z(z)dz - \int_{0}^{\infty} p_Z(z)dz \\
&= \int_{-\infty}^{0} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz + \int_{0}^{\infty} \left(2\frac{1+\mu}{\mu+e^{-\tau Z}} - 2\right) p_Z(z)dz \\
&= \int_{-\frac{1}{\tau}\log(2+\mu)}^{0} \psi_{ss}^{\mu,\tau}(z)p_Z(z)dz + \int_{0}^{\infty} \left(\frac{2+2\mu}{\mu+e^{-\tau Z}} - 2\right) p_Z(z)dz \\
&\leq \int_{-\frac{1}{\tau}\log(2+\mu)}^{0} p_Z(z)dz + \int_{0}^{\infty} \frac{2}{\mu}p_Z(z)dz \\
&= \mathbb{P}\left(-\frac{1}{\tau}\log(2+\mu) \leq Z < 0\right) + \frac{2}{\mu}\mathbb{P}(Z > 0) \\
&\leq \frac{1}{\tau}\log(2+\mu)L(x) + \frac{2}{\mu}.
\end{aligned}$$

Here, the last inequality follows from $\mathbb{P}(Z > 0) \leq 1$. ∎

THEOREM 2. *Let* $\tau(\mu) := (1 + \mu)\theta$ *with* $\theta > 0$*. Then SigVaR-P is equivalent to CC-P as* $\lim \mu \to \infty$.

Proof. From Lemma 3.1 we can establish the bound $\epsilon_{\mu,\tau} \leq \tau(\mu)^{-1}\log(2+\mu)L + 2\mu^{-1}$ with $L := \sup_{x\in\mathcal{X}} L(x)$. The result follows. □ ∎

This result implies that $\lim_{\mu\to\infty} \mathcal{X}_{ss}^{\mu,\tau(\mu)}(\alpha) = \mathcal{X}(\alpha)$ and $\lim_{\mu\to\infty} \varphi_{ss}^{\mu,\tau(\mu)}(\alpha) = \varphi(\alpha)$. The following result shows that we can construct a sequence of SigVaR approximations of increasing quality by progressively increasing μ.

THEOREM 3. *Let* $\tau(\mu) := (1 + \mu)\theta$ *with* $\theta > 0$*. We have that* $\mathcal{X}_{ss}^{\mu^+,\tau(\mu^+)}(\alpha) \supset \mathcal{X}_{ss}^{\mu,\tau(\mu)}(\alpha)$ *and* $\varphi_{ss}^{\mu^+,\tau(\mu^+)}(\alpha) \leq \varphi_{ss}^{\mu,\tau(\mu)}(\alpha)$ *for* $\mu^+ > \mu > 0$ *and for all* $\alpha \in (0, 1]$.

Proof. We show that $\psi_{ss}^{\mu,\tau}(z) < \psi_{ss}^{\mu^+,\tau}(z)$ for any $z \in \mathbb{R} \setminus \{0\}$ (for $z = 0$, we have $\psi_{ss}^{\mu,\tau}(z) = 1$ for any μ). To proceed, it suffices to show that the kernel function $\frac{1+\mu}{\mu+e^{-\tau(\mu)z}}$ is a strictly decreasing function of μ for all $z \in \mathbb{R} \setminus \{0\}$. We establish this by showing that the derivative of of the kernel function is negative:

$$\begin{aligned}
\frac{d}{d\mu}\left(\frac{1+\mu}{\mu+e^{-\tau(\mu)z}}\right) &= \frac{\mu+e^{-\tau(\mu)z}-(1+\mu)(1-\theta z e^{-\tau(\mu)z})}{(\mu+e^{-\tau(\mu)z})^2} \\
&= \frac{-1+(1+(1+\mu)\theta z)e^{-\tau(\mu)z}}{(\mu+e^{-\tau(\mu)z})^2} \\
&= \frac{-1+(1+\tau(\mu)z)e^{-\tau(\mu)z}}{(\mu+e^{-\tau(\mu)z})^2} \\
&< 0.
\end{aligned}$$

The last step follows from $1+\tau(\mu)z < e^{\tau(\mu)z}$, for any $z \in \mathbb{R} \setminus \{0\}$ (from Taylor's theorem and from the convexity of the exponential function). ∎

Remark: When $Z(x)$ is a discrete random variable, we can establish the error bound of Lemma 3.1 if $Z(x)$ has finite outcomes and we have that $\mathbb{P}(Z(x)=0)=0$. In particular, if $Z(x)$ has m finite possible outcomes $z_1 < z_2 < \cdots < z_{m'} < 0 < z_{m'+1} < \cdots < z_m$ and define the corresponding probabilities as p_i, $i = 1, ..., m$. A bounding constant $L(x)$ can be found in this case by noticing that $\mathbb{P}(-\delta \le Z(x) \le 0) = \sum_{i=1}^{m'} p_i$ if $-\delta \le z_1$, $\mathbb{P}(-\delta \le Z(x) \le 0) = \sum_{i=k}^{m'} p_i$, if $z_{k-1} < -\delta \le z_k$, and $\mathbb{P}(-\delta \le Z(x) \le 0) = 0$ if $z_{m'} < -\delta$. We thus have that $L(x) := \max_{k\in\{1,...,m'\}} \left\{\sum_{i=k}^{m'} p_i/z_k\right\}$ satisfies $\mathbb{P}(-\delta \le Z(x) \le 0) \le L(x)\delta$. By using this property we can establish that:

$$\begin{aligned}
\epsilon_{\mu,\tau}(x) &= \sum_{i=1}^{m} \psi_{ss}^{\mu,\tau}(z_i)p_i - \sum_{i=m'+1}^{m} p_i \\
&= \sum_{i=1}^{m'-1} \psi_{ss}^{\mu,\tau}(z_i)p_i + \sum_{i=m'}^{m} \psi_{ss}^{\mu,\tau}(z_i)p_i - \sum_{i=m'}^{m} p_i \\
&= \sum_{i=1}^{m'-1} \psi_{ss}^{\mu,\tau}(z_i)p_i + \sum_{i=m'}^{m} \left(2\frac{1+\mu}{\mu+e^{-\tau z_i}} - 1\right) p_i - \sum_{i=m'}^{m} p_i \\
&= \sum_{i=1}^{m'-1} \psi_{ss}^{\mu,\tau}(z_i)p_i + \sum_{i=m'}^{m} \left(2\frac{1+\mu}{\mu+e^{-\tau z_i}} - 2\right) p_i \\
&= \sum_{-\frac{1}{\tau}\log(2+\mu)\le z_i<0} \left(2\frac{1+\mu}{\mu+e^{-\tau z_i}} - 1\right) p_i + \sum_{i=m'+1}^{m} \left(2\frac{1+\mu}{\mu+e^{-\tau z_i}} - 2\right) p_i \\
&\le \sum_{-\frac{1}{\tau}\log(2+\mu)\le z_i<0} p_i + \sum_{i=m'+1}^{m} \frac{2}{\mu} p_i \\
&= \mathbb{P}\left(-\frac{1}{\tau}\log(2+\mu) \le Z < 0\right) + \frac{2}{\mu}\mathbb{P}(Z>0) \\
&\le \frac{\log(2+\mu)L(x)}{\tau} + \frac{2}{\mu}.
\end{aligned}$$

Consequently, the results of Theorem 3 hold. This result is of relevance because we are often interested in solving discrete approximations of SigVar-P (e.g., by using SAA).

3.2. Relationship with CVaR-P. We now proceed to show that the parameters of SigVaR-P can be selected in such a way that it provides an approximation of CC-P that is at least as good as that of CVaR-P.

PROPOSITION 3.2. *Assume a fixed* $\alpha \in (0,1]$ *and that* $\mu, \tau_\alpha \in \mathbb{R}_+$ *satisfy* $\mu \geq \bar{\mu}$ *(where* $\bar{\mu} \in \mathbb{R}_+$ *is the positive root of* $\bar{\mu} - \log(2+\bar{\mu}) = 1$*),* $\tau_\alpha := \frac{\mu+1}{2}\gamma_\alpha$*, and* γ_α *defined in* (2.10)*. We have that* $\mathcal{X}_c(\alpha) \subseteq \mathcal{X}_{ss}^{\mu,\tau}(\alpha)$ *and* $\varphi_{ss}^{\mu,\tau_\alpha}(\alpha) \leq \varphi_c(\alpha)$*.*

Proof. For simplicity, we omit dependence on α for $x_c(\alpha)$, γ_α, and τ_α (we simply write x_c, γ, τ). We proceed by proving that any solution x_c of CVaR-P is a feasible solution for SigVaR-P provided that μ, τ satisfy the conditions of the theorem. Since $x_c \in \mathcal{X}$, this would imply that we can always find μ, τ such that $\mathcal{X}_c(\alpha) \subseteq \mathcal{X}_{ss}^{\mu,\tau_\alpha}(\alpha)$ and $\varphi_{ss}^{\mu,\tau}(\alpha) \leq \varphi_c(\alpha)$. We define the random variable $Z_c = f(x_c, \Xi)$ with realizations $z_c \in \mathbb{R}$; the constraint (2.8b) evaluated at x_c, γ can be written as $\mathbb{E}[[\gamma Z_c + 1]_+] \leq \alpha$. It suffices to show that $[\gamma z_c + 1]_+ \geq [2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1]_+$ holds for any $z_c \in \mathbb{R}$. If $z_c < -\delta$ we have that $2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1 < 0$ and, consequently, $[\gamma z_c + 1]_+ \geq [2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1]_+$. For $z_c \geq -\delta$ we have that,

$$\begin{aligned} \gamma z_c + 1 \geq & 1 - \frac{\gamma}{\tau}\log(2+\mu) \\ \geq & 1 - \frac{2\log(2+\mu)}{\mu+1} \\ > & 0. \end{aligned} \tag{3.26}$$

This inequality follows because $\frac{2\log(2+\mu)}{\mu+1}$ is a monotonically decreasing function for $\mu \in \mathbb{R}_+$. We also observe that, for $2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1 \geq 0$,

$$\begin{aligned} [\gamma z_c + 1]_+ - \left[2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1\right]_+ &= [\gamma z_c + 1] - \left[2\frac{1+\mu}{\mu+e^{-\tau z_c}} - 1\right] \\ &= \frac{(\gamma z_c + 2)(\mu + e^{-\tau z_c}) - 2 - 2\mu}{\mu + e^{-\tau z_c}}. \end{aligned} \tag{3.27}$$

We now define $h(z_c) := (\gamma z_c + 2)(\mu + e^{-\tau z_c}) - 2 - 2\mu$ and proceed to show that $h(z_c) \geq 0$ holds for $0 \geq z_c \geq -\delta$. This is established from the following sequence of implications:

$$h(z_c) = (\gamma z_c + 2)(\mu + e^{-\tau z_c}) - 2 - 2\mu \tag{3.28a}$$

$$= (\gamma z_c + 2)\left(\mu + \sum_{n=0}^{\infty}\frac{(-\tau z_c)^n}{n!}\right) - 2 - 2\mu \tag{3.28b}$$

$$\geq (\gamma z_c + 2)\left(\mu + 1 - \tau z_c + \frac{(\tau z_c)^2}{2}\right) - 2 - 2\mu \tag{3.28c}$$

$$= \gamma z_c\left(\mu + 1 - \frac{2c}{a} - \tau z_c + \frac{\tau^2 z_c}{a} + \frac{\tau^2 z_c^2}{2}\right) \tag{3.28d}$$

$$= \gamma\tau z_c^2\left(\frac{\mu+1}{2} - 1 + \frac{\tau z_c}{2}\right) \tag{3.28e}$$

$$\geq \gamma\tau z_c^2\left(\frac{\mu - 1 - \log(2+\mu)}{2}\right) \tag{3.28f}$$

$$\geq 0. \tag{3.28g}$$

Equation (3.28f) follows because $\mu - 1 - \log(2+\mu)$ is a monotonically increasing function for $\mu \geq 0$. Equation (3.28g) follows for $\mu \geq \bar{\mu}$. For $z_c \geq 0$ we have,

$$\begin{aligned} h'(z_c) &= \gamma\mu + (\gamma - \tau\gamma z_c - 2\tau)e^{-\tau z_c} \\ &= \gamma\mu + 2(\gamma-\tau)e^{-\tau z_c} - \gamma\left(\frac{\tau z_c + 1}{e^{\tau z_c}}\right) \\ &\geq \gamma\mu + 2(\gamma - \tau) - \gamma \\ &= 0. \end{aligned} \tag{3.29}$$

This follows because $\gamma - \tau < 0$, $0 < e^{-\tau z_c} \leq 1$, and $\frac{\tau z_c + 1}{e^{\tau z_c}} \leq 1$. Since $h(0) = 0$ we have that $h(z_c) \geq 0$ for $z_c \geq 0$. We thus have that $\text{SigVaR}^{\mu,\tau}_{1-\alpha}(f(x_c(\alpha), \Xi)) \leq 0$ holds for μ, τ satisfying the conditions of the theorem. ∎

As discussed in Section 4, Proposition 3.2 is of practical relevance because it indicates that we can use the solution of CVaR-P (which is a computationally attractive formulation) to find an initial guess for SigVar-P. We also note that Proposition 3.2 implies that SigVaR provides an approximation that is at least as good as that of EVaR.

3.3. Relationship with DC-P. The following results compare the solutions of SigVaR-P and DC-P. To establish these results, we define the SigVaR-DC error:

$$d_{\mu,\tau} := \mathbb{E}[\psi^{\mu,\tau}_{ss}(Z)] - \epsilon^{-1}\mathbb{E}\left[[Z+\epsilon]_+ - [Z]_+\right]. \tag{3.30}$$

In addition, we define $d_{\mu,\tau}(z) := \psi^{\mu,\tau}_{ss}(z) - \epsilon^{-1}\left[[z+\epsilon]_+ - [z]_+\right]$ for all $z \in \mathbb{R}$. Consequently, $d_{\mu,\tau} = \mathbb{E}[d_{\mu,\tau(Z)}]$.

We now establish an upper bound for the SigVar-DC error.

PROPOSITION 3.3. *Assume that $\tau \in \mathbb{R}_+$ satisfies $\tau \leq \frac{1}{2}\epsilon^{-1}$. We have that $d_{\mu,\tau} \geq 0$ for any $\mu \in \mathbb{R}_+$.*

Proof. We proceed by proving that $d_{\mu,\tau}(z) \geq 0$ holds for any $z \in \mathbb{R}$. If $z < -\epsilon$ we have that $\epsilon^{-1}\left[[z+\epsilon]_+ - [z]_+\right] = 0$ and, consequently, $d_{\mu,\tau} \geq 0$. For $z \geq -\epsilon$ we have that,

$$2\frac{1+\mu}{\mu + e^{-\tau z}} - 1 \geq 2\frac{1+\mu}{\mu + e^{\tau\epsilon}} - 1 \geq 0. \tag{3.31}$$

We also observe that, for $-\epsilon \leq z \leq 0$,

$$d_{\mu,\tau} = \left[2\frac{1+\mu}{\mu+e^{-\tau z}} - 1\right] - \left[\epsilon^{-1}z + 1\right] = \frac{-\hat{h}(z)}{\mu + e^{-\tau z}}. \tag{3.32}$$

We proceed to show that $\hat{h}(z) := (\epsilon^{-1}z+2)(\mu+e^{-\tau z}) - 2 - 2\mu \leq 0$ holds for $-\epsilon \leq z \leq 0$. This is established from the following sequence of implications:

$$\begin{aligned} \hat{h}'(z) &= \epsilon^{-1}\mu + \left(\epsilon^{-1} - \epsilon^{-1}\tau z - 2\tau\right)e^{-\tau z} \\ &\geq \epsilon^{-1}\mu + \left(\epsilon^{-1} - 2\tau\right)e^{-\tau z} \\ &\geq \epsilon^{-1}\mu. \end{aligned} \tag{3.33}$$

Since $\hat{h}(0) = 0$, we have that $\hat{h}(z) \leq 0$ for$-\epsilon \leq z \leq 0$. For $z \geq 0$ we have, $d_{\mu,\tau} = \left[2\frac{1+\mu}{\mu+e^{-\tau z}} - 1\right] - 1 \geq 0$. ∎

This result shows that as $\epsilon \to 0$, the range of feasible τ that make SigVaR-P more conservative increases. We now establish a lower bound for the SigVar-DC error.

PROPOSITION 3.4. *Assume $\mu, \tau \in \mathbb{R}_+$ satisfy $\mu \geq \bar{\mu}$ where $\bar{\mu}$ is the positive root of $\bar{\mu} - \log(2+\bar{\mu}) = 1$ and $\tau \geq \frac{1}{2}\epsilon^{-1}(\mu+1)$. We have that $d_{\mu,\tau} \leq \frac{2}{\mu}$.*

Proof. We proceed by proving that $d_{\mu,\tau} \leq \frac{2}{\mu}$ holds for any $z \in \mathbb{R}$ if μ, τ satisfy the conditions of the theorem. If $z < -\delta$ we have that $2\frac{1+\mu}{\mu+e^{-\tau z}} - 1 < 0$ and, consequently, $d_{\mu,\tau} \leq 0$. For $z \geq -\delta$ we have that,

$$\begin{aligned}
\epsilon^{-1}z + 1 \geq & 1 - \frac{1}{\epsilon\tau}\log(2+\mu) \\
\geq & 1 - \frac{2\log(2+\mu)}{\mu+1} \\
\geq & 0. \qquad (3.34)
\end{aligned}$$

Equation (3.34) follows because $\frac{2\log(2+\mu)}{\mu+1}$ is a monotonically decreasing function for $\mu \geq 0$. We also observe that, for $-\delta \leq z \leq 0$,

$$\begin{aligned}
d_{\mu,\tau} &= \left[2\frac{1+\mu}{\mu+e^{-\tau z}} - 1\right] - \left[\epsilon^{-1}z + 1\right] \\
&= \frac{-\hat{h}(z)}{\mu+e^{-\tau z}}. \qquad (3.35)
\end{aligned}$$

We define $\hat{h}(z) := (\epsilon^{-1}z+2)(\mu+e^{-\tau z})-2-2\mu$ and proceed to show that $\hat{h}(z) \geq 0$ holds for $-\delta \leq z \leq 0$. This is established from the following sequence of implications:

$$\begin{aligned}
\hat{h}(z) &= \left(\epsilon^{-1}z+2\right)\left(\mu+e^{-\tau z}\right) - 2 - 2\mu && (3.36a)\\
&= \left(\epsilon^{-1}z+2\right)\left(\mu + \sum_{n=0}^{\infty}\frac{(-\tau z)^n}{n!}\right) - 2 - 2\mu && (3.36b)\\
&\geq \left(\epsilon^{-1}z+2\right)\left(\mu + 1 - \tau z + \frac{(\tau z)^2}{2}\right) - 2 - 2\mu && (3.36c)\\
&= \epsilon^{-1}z\left(\mu + 1 - 2\epsilon\tau - \tau z + \epsilon\tau^2 z + \frac{\tau^2 z^2}{2}\right) && (3.36d)\\
&\geq \epsilon^{-1}\tau z^2\left(\frac{\mu+1}{2} - 1 + \frac{\tau z}{2}\right) && (3.36e)\\
&\geq \epsilon^{-1}\tau z^2\left(\frac{\mu - 1 - \log(2+\mu)}{2}\right) && (3.36f)\\
&\geq 0. && (3.36g)
\end{aligned}$$

Equation (3.36b) uses a Taylor series expansion of the exponential function and (3.36f) follows because $\mu-1-\log(2+\mu)$ is a monotonically increasing function of μ for $\mu \geq 0$. Equation (3.36g) follows for $\mu \geq \bar{\mu}$. For $z \geq 0$ we have that $d_{\mu,\tau} = \left[2\frac{1+\mu}{\mu+e^{-\tau z}} - 1\right] - 1 \leq \frac{2}{\mu}$. The result follows. ∎

This result shows that improving the quality of the DC-P approximation (by setting $\epsilon \to 0$) corresponds to setting $\mu, \tau \to \infty$ for SigVaR-P (e.g., by using $\tau(\mu) = \theta(\mu+1)$ with $\theta = \frac{1}{2}\epsilon^{-1}$). The results also indicate that it is possible to overcome the computational limitations associated to DC-P.

4. Computational Implementation. We use SAA to convert SigVar-P into a finite-dimensional problem [11]. We generate a set of realizations $\xi \in \Omega$ from p_Ξ. The SAA approximation is given by:

$$\min_{x\in\mathcal{X}, z_\xi\in\mathbb{R}, \phi_\xi\in\mathbb{R}_+} \; \varphi(x) \tag{4.37a}$$
$$\text{s.t.} \; z_\xi = f(x,\xi), \;\; \xi\in\Omega \tag{4.37b}$$
$$\phi_\xi \geq 2\frac{1+\mu}{\mu+e^{-\tau z_\xi}} - 1, \;\; \xi\in\Omega \tag{4.37c}$$
$$\frac{1}{|\Omega|}\sum_{\xi\in\Omega}\phi_\xi \leq \alpha. \tag{4.37d}$$

Large values of τ will cause difficulty for the NLP solver due to the high nonlinearity of the sigmoid function. For example, the first derivative of $2\frac{1+\mu}{\mu+e^{-\tau z_\xi}}$ with respect to z_ξ is $\mathcal{O}(\tau)$ and thus becomes increasingly steep as τ is increased. Moreover, the second derivative is $\mathcal{O}(\tau^2)$. Consequently, we propose a scheme to solve a sequence of SigVaR approximations of increasing quality and with this achieve more robustness. The scheme (called `SigVaR-Alg`) begins by finding a solution of the SAA approximation of the CVaR-P. The SAA approximation of CVaR-P is:

$$\min_{x\in\mathcal{X}, z_\xi\in\mathbb{R}, \phi_\xi\in\mathbb{R}_+, t\in\mathbb{R}} \; \varphi(x) \tag{4.38a}$$
$$\text{s.t.} \; z_\xi = f(x,\xi), \;\; \xi\in\Omega \tag{4.38b}$$
$$\phi_\xi \geq z_\xi - t, \;\; \xi\in\Omega \tag{4.38c}$$
$$\frac{1}{|\Omega|}\sum_{\xi\in\Omega}\phi_\xi \leq -t\alpha. \tag{4.38d}$$

Algorithm `SigVaR-Alg`

---

1. Initialize
Given $\lambda > 1$, $\alpha \in (0,1]$, and target $\mu^* \in \mathbb{R}_+$.
Initialize iteration index $\ell \leftarrow 0$.
Solve CVaR problem (4.38) and set $\gamma \leftarrow -\frac{1}{t_c(\alpha)}$, $x^*_\ell \leftarrow x_c(\alpha)$, and $\varphi^*_\ell \leftarrow \varphi_c(\alpha)$.
Set $\mu_\ell \leftarrow \bar{\mu}$, $\tau_\ell \leftarrow \frac{\mu_\ell+1}{2}\gamma$, where $\bar{\mu}$ is positive root of $\bar{\mu} - \log(2+\bar{\mu}) = 1$.
Update iteration index $\ell \leftarrow \ell + 1$.

2. Solve SigVar-P
Use $x^*_{\ell-1}$ as initial guess and solve SigVaR-P (4.37) with μ_ℓ, τ_ℓ.
Set $x^*_\ell \leftarrow x^{\mu_\ell,\tau_\ell}_{ss}(\alpha)$ and $\varphi^*_\ell \leftarrow \varphi^{\mu_\ell,\tau_\ell}_{ss}(\alpha)$.
if $\mu_\ell \geq \mu^*$ **then**
  Go to Step **4**.
else
  Go to Step **3**.
end if

3. Update parameters
Set $\mu_{\ell+1} \leftarrow \lambda \cdot \mu_\ell$ and $\tau_{\ell+1} \leftarrow \frac{\mu_{\ell+1}+1}{2}\gamma$.
Update iteration index $\ell \leftarrow \ell + 1$ and return to Step **2**.

4. Stop with x^*_ℓ

From Proposition 3.2, we have that $\varphi^*_1 \leq \varphi^*_0$ holds and from Theorem 3 we have that $\varphi^*_{\ell+1} \leq \varphi^*_\ell$ holds for all $\ell \geq 1$ (provided that the NLPs are solved to global optimality).

5. Numerical Studies.

We use a couple of small-scale studies to illustrate the theoretical properties of SigVaR and a large-scale wind turbine optimization study to demonstrate its practical benefits.

5.1. Analytical Example.

Consider the following CC-P:

$$\min_{x \in \mathbb{R}} \; x \tag{5.39a}$$

$$\text{s.t.} \;\; \mathbb{P}(\Xi \leq x) \geq 1 - \alpha, \tag{5.39b}$$

with $\Xi \sim \mathcal{U}(0,1)$. The optimal objective value and solution are $\varphi(\alpha) = x^*(\alpha) = 1-\alpha$ and we note that $\mathbb{P}(\Xi \leq x^*(\alpha)) = 1-\alpha$. This implies $1-\alpha = F(x^*(\alpha)) = Q_{1-\alpha}(\Xi) = x^*(\alpha)$. We handle the CC (5.39b) using the VaR formulation (**??**) (exact), the CVaR approximation (2.8b), the EVaR approximation (2.14), and the SigVaR approximation (3.22b). It can be shown that the optimal solution and objective values obtained with these approaches are, respectively, $\text{VaR}_{1-\alpha}(\Xi) = Q_{1-\alpha}(\Xi)$, $\text{CVaR}_{1-\alpha}(\Xi)$, $\text{EVaR}_{1-\alpha}(\Xi)$, and $\text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi)$. Moreover, $\text{VaR}_{1-\alpha}(\Xi) = 1-\alpha$, $\text{CVaR}_{1-\alpha}(\Xi) = \frac{1}{2}(2-\alpha)$, and $\text{EVaR}_{1-\alpha}(\Xi) = \inf_{t>0}\{t\log(te^{t^{-1}} - t) - t\log\alpha\}$. For the case of SigVar we have that, for $\alpha \geq \frac{2+2\mu}{\mu\tau}\log(\frac{2+\mu+\mu e^\tau}{2+2\mu}) - 1$,

$$\text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi) = \tau^{-1}\log\left(\frac{\mu e^\tau - \mu\beta}{\beta - 1}\right) \tag{5.40}$$

where $\beta = e^{\frac{(\alpha+1]\mu\tau}{2+2\mu}}$; and, otherwise, we have that,

$$\text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi) = \inf_{t\in\mathbb{R}}\left\{\frac{2+\mu}{\mu\tau}\log(2+\mu) + \frac{2+2\mu}{\mu\tau}\log\left(\frac{\mu e^{\tau(1-t)}+1}{2+2\mu}\right) + t - 1 \leq \alpha\right\}. \tag{5.41}$$

The optimal objective values for all approaches as a function of α are shown in Figure 2. For the SigVaR approximation we use $\mu = 100, \tau = (\mu+1) = 101$. As predicted by the properties of SigVaR, we have that $\text{VaR}_{1-\alpha}(\Xi) \leq \text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi)$ for all α.

We have that $Z(x) = \Xi - x \sim \mathcal{U}(-x, 1-x)$ for $x \in \mathbb{X}$. Consequently, the constant $L = \sup_{x\in\mathcal{X}} L(x) = \sup_z\{p_{Z(x)}(z)\} = 1$ satisfies $\mathbb{P}(-\delta \leq Z(x) < 0) \leq L\delta$ for all $x \in \mathcal{X}$. From Lemma 3.1, the approximation error of the SigVaR function is bounded as $\epsilon_{\mu,\tau} \leq \frac{\log(2+\mu)L}{\tau} + \frac{2}{\mu} = \frac{\log(102)}{101} + \frac{2}{100} = 0.066$. We note that this is an upper bound of the empirical error $\epsilon^{\mu,\tau} = 0.052$ observed in Figure 2 and computed by $\epsilon_{\mu,\tau} = \text{SVaR}^{\mu,\tau}_{1-\alpha}(\Xi) - \text{VaR}_{1-\alpha}(\Xi)$ (vertical distance at each $x^*(\alpha) = 1-\alpha$).

From the solution of the CVaR approximation we obtain that $t_c(\alpha) = -\frac{\alpha}{2} < 0$ and this $\gamma_\alpha = -\frac{1}{t_c(\alpha)} = \frac{2}{\alpha}$. The conditions of Proposition 3.2 are satisfied for $\mu = 100, \tau = 101$ for all $\alpha \geq 0.07$ and thus $\text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi) \leq \text{CVaR}_{1-\alpha}(\Xi) \leq \text{EVaR}_{1-\alpha}(\Xi)$, which is verified in Figure 2 (for $\alpha < 0.07$ the conditions of Proposition 3.2 are not satisfied and SigVaR becomes more conservative). The extreme conservatism of CVaR and EVaR becomes obvious at large values of α. In particular, at $\alpha = 1$ we see that $\text{SigVaR}^{\mu,\tau}_{1-\alpha}(\Xi) = 0.052$ and $\text{CVaR}_{1-\alpha}(\Xi)$=0.5, which illustrates that the quality of the approximation can be substantially improved.

Problem 5.39 can be solved numerically by using SAA (in our experiments we use 1000 scenarios). The CC-P in this case can be cast as an MILP, CVaR-P is an LP, and SigVaR-P is a standard NLP. The MILPs are solved with the solver `SCIP`

This manuscript is for review purposes only.

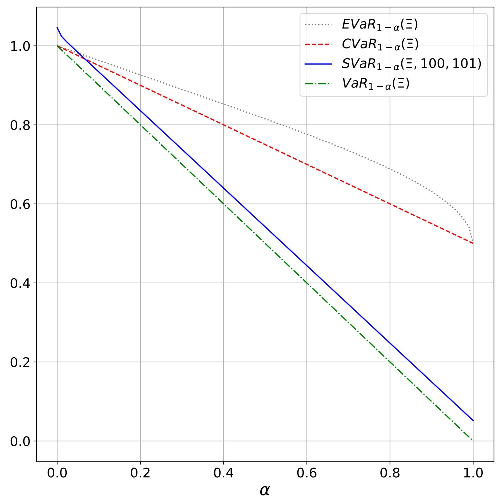

FIG. 2. *Optimal objectives obtained with VaR, CVaR, EVaR, and SigVaR for analytical example.*

and the LPs and NLPs are solved with `IPOPT`. Theorem 3 shows that $\mathrm{SigVaR}_{1-\alpha}^{\mu,\tau(\mu)}(\Xi)$ becomes less conservative for increasing μ, which is verified in Figure 3 for $\alpha = 0.9$ and $\alpha = 0.1$. For $\alpha = 0.9$, the solution of the MILP formulation is 0.09, which is close to the analytical solution of 0.1. `SigVaR-Alg` first finds the solution of CVaR approximation, which is 0.547. In iteration 1, we solve with SigVaR approximation with $\mu = 2.5$ and $\tau = 3.8$, and find a solution of 0.49. After 9 iterations, we solve a SigVaR approximation with $\mu = 641$ and $\tau = 702$ and find a a solution of 0.099. The gap between MILP formulation and SigVaR is only 2% of the gap between MILP formulation and CVaR. For $\alpha = 0.9$, the gap is 20% but we also see that the gap is more difficult to close.

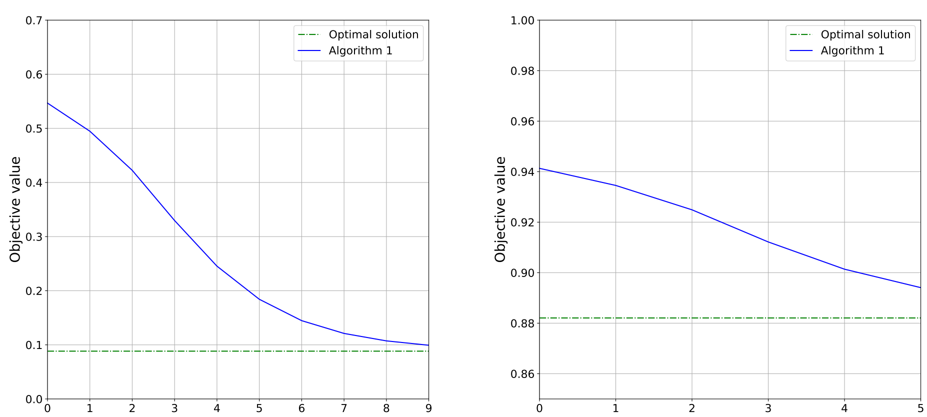

FIG. 3. *Performance of SigVar on analytical example with* $\alpha = 0.9$ *(left) and* $\alpha = 0.1$ *(right).*

5.2. Farmer Problem. We consider modified version of the classical farmer problem [3]. In this problem, the farmer needs to decide how much land to allocate to grow wheat, corn, and beets while considering the uncertainty on crop yields. The farmer has the option to buy/sell crops to satisfy contracts and maximize revenue (minimize cost). The formulation is given by:

$$\min_{x, y_j(\cdot), w_j(\cdot)} \quad \mathbb{E}\left[f(\Xi)\right] \tag{5.42a}$$

$$\text{s.t.} \sum_{j \in \mathcal{P}} x_j \leq \bar{x} \tag{5.42b}$$

$$\tau_j(\Xi) x_j + y_j(\Xi) - w_j(\Xi) \geq \beta_j, j \in \mathcal{P} \quad \text{a.s.} \tag{5.42c}$$

$$f(\Xi) = \sum_{j \in \mathcal{P}} \left(\gamma_j^x x_j + \gamma_j^y y_j(\Xi) - \gamma_j^w w_j(\Xi)\right) \quad \text{a.s.} \tag{5.42d}$$

$$\mathbb{P}\left(f(\Xi) \leq \bar{f}\right) \geq 1 - \alpha \tag{5.42e}$$

$$0 \leq w_j(\Xi) \leq \bar{w}_j, \ 0 \leq y_j(\Xi) \leq \bar{y}_j, j \in \mathcal{P} \quad \text{a.s.} \tag{5.42f}$$

where x_j denotes the land allocated to each crop at cost γ_j^x, $y_j(\xi)$ represents the crops bought at price γ_j^y, $w_j(\xi)$ denotes the crops sold at price γ_j^w, $\mathcal{P}$ denotes the set of crops {wheat,corn,beets}, $\tau_j(\xi)$ is the yield of crops, β_j denotes demand contracts and $\bar{x}, \bar{y}_\ell, \bar{w}_\ell$ represents capacities. Constraint (5.42e) requires that the cost $f(\cdot)$ is lower than the threshold $\bar{f}$ with probability of at least $1 - \alpha$. We assume that the yield of wheat and corn is constant, while the yield of beets follows a normal distribution $\mathcal{N}(20, 5)$. We generate 1000 scenarios from this distribution and we set $1 - \alpha = 0.9$ and $\bar{f} = \$53,000$.

The performance of `SigVaR-Alg` is summarized in Table 1. CC-P denotes the solution of the MILP formulation. As can be seen, the expected cost of the MILP formulation is \$-100248. `SigVaR-Alg` first finds the solution of CVaR approximation. The expected cost of CVaR approximation is \$-77535 (which is around 23% higher than the optimal MILP cost). This is because, although Equation (5.42e) only requires the cost to be lower than the threshold with probability equal to larger than 0.9, the solution of CVaR formulation satisfies the constraint with probability 0.967. Figure 4 shows the histogram of the cost obtained with CVaR, SigVaR, and MILP formulations. Here, it becomes obvious that CVaR can significantly distort the cost distribution due to high conservatism. From the solution of the CVaR approximation we obtain $t_c(\alpha) = -6250 < 0$ and $\gamma = 0.00016 > 0$. After 5 iterations, `SigVaR-Alg` solves the SigVaR approximation with $\mu = 40$ and $\tau = 0.00329$ and finds a solution with expected cost equal to \$-93292 (which is is around 7% higher than the optimal MILP cost). The gap between the MILP and SigVaR formulations is only 30% of the gap between the MILP and CVaR formulations. We also observe that, as the iterations proceed, the objective value of SigVaR-P decreases monotonically, $\mathbb{P}(f(\Xi) \leq \bar{f})$ decreases, and $\text{VaR}_{1-\alpha}(f(\Xi))$ increases. We can thus see that the SigVar formulation can significantly reduce the conservatism of the CVaR solution. We acknowledge, however, that we are unable to close the gap further due to numerical instability of the NLP solver.

5.3. Wind Turbine Optimization. We now solve a CC-P that seeks to find optimal pitch and torque control policies for a wind turbine given uncertainty in wind speed conditions. The formulation seeks to maximize expected power and to satisfy a CC on the maximum mechanical load experienced by the wind turbine. We represent

TABLE 1
Performance of `SigVaR-Alg` *on farmer problem with* $\alpha = 0.1$.

ℓ	μ	τ	$\mathbb{E}[f(\Xi)]$	$\mathrm{VaR}_{1-\alpha}(f(\Xi))$	$\mathbb{P}\left(f(x,\Xi) \leq f\right)$
CVaR-P($\ell = 0$)	-	-	-77535	-59250	0.967
1	2.5	0.00028	-80807	-58378	0.952
2	5.0	0.00048	-83765	-57589	0.937
3	10.0	0.00088	-87360	-56630	0.925
4	20.0	0.00168	-90758	-55724	0.918
5	40.0	0.00329	-93292	-55043	0.911
CC-P	-	-	-100248	-53000	0.9

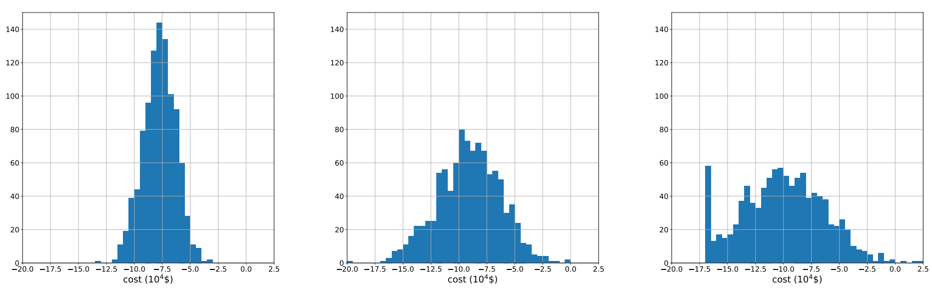

FIG. 4. *Cost distribution using CVaR-P (left), SigVaR-P (middle) and CC-P (right) formulation.*

this problem in the following abstract form:

$$\max_{u \in \mathcal{U}} \mathbb{E}\left[\frac{1}{T}\int_{\mathcal{T}} y_P(t,\Xi)dt\right] \tag{5.43a}$$

$$\text{s.t. } (y_P(\Xi,t), y_L(\Xi,t)) = \mathcal{M}(u(t), u(t,\Xi), V(\Xi,t)),\; t \in \mathcal{T} \tag{5.43b}$$

$$\mathbb{P}\left\{y_L^{max}(\Xi) \leq \bar{y}_L\right\} \geq 1 - \alpha \tag{5.43c}$$

$$y_L^{max}(\Xi) = \max_{t \in \mathcal{T}} y_L(t,\Xi) \tag{5.43d}$$

$$y_L(\Xi,t) \leq \hat{y}_L,\; t \in \mathcal{T} \tag{5.43e}$$

where $t \in \mathcal{T} := [0,T]$, $V(\Xi,t)$ is the wind speed, $y_P(\Xi,t)$ is the wind turbine power, $y_L(\Xi,t)$ is the mechanical load with associated threshold $\bar{y}_L$. For a time horizon of ten minutes, we set the control actions for the first 10 seconds to be first stage variables $u(t)$ (the implemented control actions) and the rest to be second stage variables $u(t,\Xi)$ (the recourse control actions). Equation (5.43b) is an abstract representation of a wind turbine model (comprises nonlinear differential and algebraic equations). The model details are presented in [5].

An important practical problem is that power maximization conflicts with the mechanical load experienced by the turbine (i.e., the higher the power extracted the higher the load). Consequently, it is important to carefully trade-off these metrics so as to prevent putting the turbine at extreme mechanical risk. The probabilistic constraint (5.43c) enforces that the probability that the peak load $y_L^{max}(\Xi)$ exceeds the threshold $\bar{y}_L$ is no more than α. Constraint (5.43e) enforces that the peak load never exceeds another (less conservative) threshold $\hat{y}_L$. In our experiments we set $\alpha = 0.5, \bar{y}_L = 60$ MNm, and $\hat{y}_L = 200$ MNm.

To solve this problem, we discretize the dynamic model by using a Radau collocation scheme [18]. To accurately capture extreme loads we have found that it is necessary to discretize the model using a resolution of 0.5 seconds over 10 minutes, giving rise to 1,200 time steps. For an NLP with 230 scenarios, the total number of variables in this problem is 5.5 million. The NLPs arising in this application were implemented in `PLASMO` [9] and solved with the parallel solver `PIPS-NLP` [7] (which exploits the structure of the stochastic program at the linear algebra level) and with the off-the-shelf serial solver `IPOPT`[17] (which treats the problem as a general NLP). Because of the size of the problem and because the wind turbine model is nonconvex, MINLP formulations of CC-P are computationally intractable.

Table 2 summarizes the performance of `SigVaR-Alg`. The serial solver `IPOPT` takes 1.3 hours to solve the CVaR-P while the parallel solver `PIPS-NLP` take only 30 minutes using 23 computing cores. The expected power obtained with CVaR-P is 3.548 MW and we have found this performance to be too conservative. In particular, although the CC (5.43c) only requires $\max_{t\in\mathcal{T}} y_L(t,\Xi) \leq \bar{y}_L$ to hold with a probability of 0.5, the CVaR-P solution satisfies it with probability 0.748. From the solution CVaR-P we obtain $\gamma_\alpha = 0.822$. From Table 2 we also see that the SigVaR approximation becomes less conservative as we increase μ, τ and that the objective value is progressively improved (power is maximized). After three iterations, `SigVaR-Alg` solves SigVaR-P with $\mu = 10$ and $\tau = 4.52$ and achieves an expected power of 3.865 MW (an improvement of 8.9% over CVaR-P). The probability of satisfying the maximum load threshold is reduced to 0.583. At a price of electricity of 30 \$/MWh, these cost savings obtained with SigVaR-P translate to around \$83,000 per year (for a single 5 MW wind turbine). We can thus see that the economic benefits of reducing conservatism can be quite significant.

TABLE 2
Performance of `SigVaR-Alg` *on wind turbine optimization problem with* $\alpha = 0.5$.

ℓ	μ	τ	$\mathbb{E}\left[\frac{1}{T}\int_{\mathcal{T}} y_P(t,\Xi)dt\right]$ (MW)	$\text{VaR}_{1-\alpha}(y_L^{max}(\Xi))$ (MNm)	$\mathbb{P}\left\{y_L^{max}(\Xi) \leq \bar{y}_L\right\}$ (MNm)
CVaR-P($\ell = 0$)	-	-	3.548	47.85	0.748
1	2.5	1.44	3.766	49.56	0.726
2	5.0	2.47	3.835	52.12	0.643
3	10.0	4.52	3.865	54.52	0.583

Figure 5 shows the cost distribution for the maximum load obtained with the CVaR-P and SigVaR-P. It is clear that CVaR is significantly more conservative and pushes the mechanical load towards small values. SigVaR, on the other hand, allows for an equal proportion of load violations and with this it can extract more power. This is illustrated in Figure 6, where we show that SigVaR achieves a larger proportion of scenarios with a large power output.

6. Concluding Remarks. We have proposed a sigmoidal approximation for chance constraints that we call SigVaR. We prove that SigVaR is conservative and that the level of conservatism can be made arbitrarily small for limiting values of the approximation parameters. We also provide conditions for the parameters guaranteeing that the SigVaR approximation is less conservative than the conditional value at risk (CVaR) approximation. The SigVar approximation brings computational benefits over mixed-integer and difference of convex functions reformulations because it can be formulated as a standard nonlinear program. We also conduct numerical experiments to demonstrate that it can significantly reduce the conservatism of CVaR. A limitation of SigVaR, however, is that numerical instability is encountered for limiting parameter values. To ameliorate this issue, we propose an algorithmic scheme that

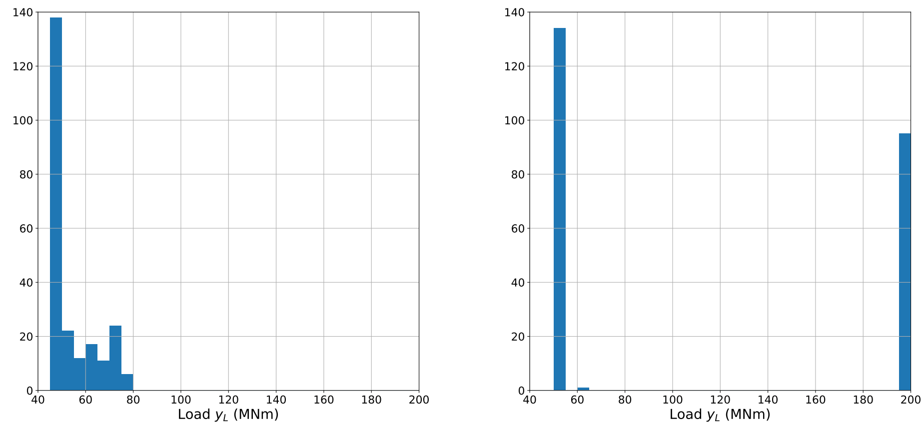

FIG. 5. *Histogram of mechanical load using CVaR (left) and SigVaR (right) formulation.*

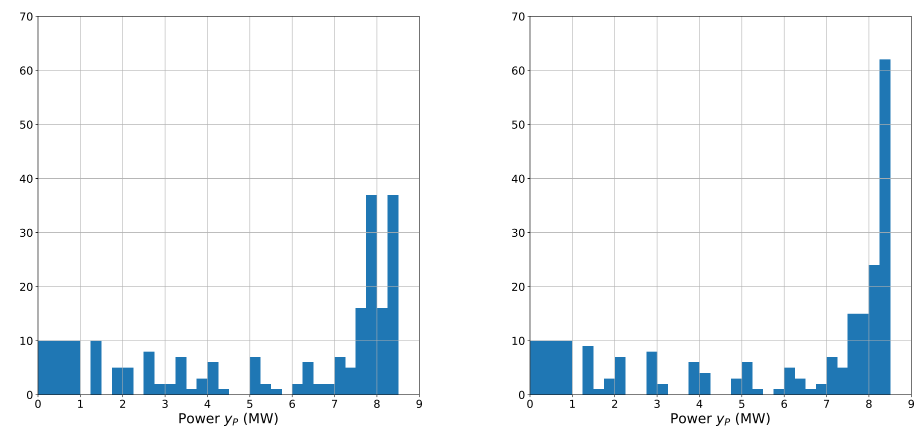

FIG. 6. *Histogram of wind turbine power using CVaR (left) and SigVaR (right) formulation.*

solves a sequence of approximations of increasing quality. As part of future work, we are interested in studying more closely the behavior of the sigmoidal approximation from an algorithmic stand-point. In particular, while the proposed scheme does improve numerical performance, extreme sensitivity of the sigmoidal function for large parameter values remains an issue.

REFERENCES

[1] A. AHMADI-JAVID, *Entropic value-at-risk: A new coherent risk measure*, Journal of Optimization Theory and Applications, 155 (2012), pp. 1105–1123.

[2] D. BIENSTOCK, M. CHERTKOV, AND S. HARNETT, *Chance-constrained optimal power flow: Risk-*

aware network control under uncertainty, SIAM Review, 56 (2014), pp. 461–495.

[3] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer Science & Business Media, 2011.

[4] G. Calafiore and M. C. Campi, *Uncertain convex programs: randomized solutions and confidence levels*, Mathematical Programming, 102 (2005), pp. 25–46.

[5] Y. Cao, F. D'Amato, and V. M. Zavala, *Stochastic optimization formulations for wind turbine power maximization and extreme load mitigation*, Under Review (2017).

[6] C. Chen and O. L. Mangasarian, *Smoothing methods for convex inequalities and linear complementarity problems*, Mathematical programming, 71 (1995), pp. 51–69.

[7] N.-Y. Chiang and V. M. Zavala, *An inertia-free filter line-search algorithm for large-scale nonlinear programming*, Computational Optimization and Applications, 64 (2016), pp. 327–354.

[8] L. J. Hong, Y. Yang, and L. Zhang, *Sequential convex approximations to joint chance constrained programs: A monte carlo approach*, Operations Research, 59 (2011), pp. 617–630.

[9] J. Jalving, S. Abhyankar, K. Kim, M. Hereld, and V. M. Zavala, *A graph-based computational framework for simulation and optimization of coupled infrastructure networks*, Undr Review, (2016).

[10] J. Kang, N. Chiang, C. D. Laird, and V. M. Zavala, *Nonlinear programming strategies on high-performance computers*, in Decision and Control (CDC), 2015 IEEE 54th Annual Conference on, IEEE, 2015, pp. 4612–4620.

[11] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM Journal on Optimization, 12 (2002), pp. 479–502.

[12] J. Luedtke, S. Ahmed, and G. L. Nemhauser, *An integer programming approach for linear programs with probabilistic constraints*, Mathematical Programming, 122 (2010), pp. 247–272.

[13] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM Journal on Optimization, 17 (2006), pp. 969–996.

[14] A. Nemirovski and A. Shapiro, *Scenario approximations of chance constraints*, in Probabilistic and randomized methods for design under uncertainty, Springer, 2006, pp. 3–47.

[15] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of risk, 2 (2000), pp. 21–42.

[16] W. Van Ackooij and R. Henrion, *Gradient formulae for nonlinear probabilistic constraints with gaussian and gaussian-like distributions*, SIAM Journal on Optimization, 24 (2014), pp. 1864–1889.

[17] A. Wächter and L. T. Biegler, *On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming*, Mathematical Programming, 106 (2006), pp. 25–57.

[18] V. M. Zavala, *Computational strategies for the optimal operation of large-scale chemical processes*, ProQuest, 2008.