Scalable Solution Strategies for
Chance-Constrained Nonlinear Programs

Javier Tovar-Facio†, Yankai Cao‡, Jose M. Ponce-Ortega†, and Victor M. Zavala‡

† Department of Chemical and Biological Engineering, University of Wisconsin-Madison
1415 Engineering Dr, Madison, WI 53706, USA
‡ Department of Chemical Engineering, Universidad Michoacana de San Nicolas de Hidalgo
Francisco J. Mugica S/N, Morelia, Michoacan 58030, Mexico

Abstract

Probabilistic (chance) constraints are a powerful modeling paradigm that helps decision-makers control risk. Unfortunately, chance constraints (CCs) cannot be handled directly by off-the-shelf optimization solvers and specialized reformulations and solution procedures are often needed. In this work, we review different strategies to tackle large-scale nonlinear programs (NLPs) with CCs. In particular, we use moment matching when the algebraic structure of the moments and of the quantile function of the CC distribution are known. To tackle more general settings with arbitrary distributions we use a sigmoidal approximation, which provides a mechanism to achieve exact solutions. We demonstrate that this approach significantly reduces the conservatism of popular approximations such as the conditional value at risk and the scenario (almost surely) approach. A flare system design study is used to illustrate the developments.

Keywords: chance constraints, risk, optimization, moments.

1 Introduction and Background Information

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

\[ \min_{d \in \mathcal{D}} \varphi(d) \]  
\[ \text{s.t. } P \left( f(d, \Xi) \leq \bar{f} \right) \geq 1 - \alpha. \]

Here, \( d \in \mathbb{R}^n \) are decision variables and the objective function \( \varphi : \mathbb{R}^n \rightarrow \mathbb{R} \) is twice continuously differentiable. The set \( \mathcal{D} := \{ d \mid g(d) \geq 0 \} \) is assumed to be compact and non-empty and is comprised of twice differentiable constraints \( g : \mathbb{R}^n \rightarrow \mathbb{R}^m \). We consider the probability space \((\Omega, \mathcal{F}, P)\) and we assume that \( \Omega \) is a measurable space equipped with \( \sigma \)-algebra \( \mathcal{F} \) of subsets of \( \Omega \), and that \( \Xi \) is a linear space of \( \mathcal{F} \)-measurable functions \( \Xi : \Omega \rightarrow \mathbb{R}^d \) (random variables representing input data). The probability measure function is given by \( P : \mathcal{F} \rightarrow [0, 1] \) and we use \( \xi \in \mathbb{R}^d \) to denote realizations of the random data vector \( \Xi \). The scalar constraint function \( f : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R} \) is also assumed to be twice
continuously differentiable and potentially nonconvex and \( f \in \mathbb{R} \) is a threshold value. We define the scalar random variable \( Z := f(d, \Xi) \) with realizations \( z \in \mathbb{R} \). We make the blanket assumption that \( \Xi \) and \( Z = f(d, \Xi) \) (for all \( d \in \mathcal{D} \)) are continuous random variables. This guarantees the existence of a probability density function (pdf) \( p_Z : \mathbb{R} \to [0, \infty) \) and of a cumulative density function (cdf) \( F_Z : \mathbb{R} \to [0, 1] \) of \( Z \) satisfying \( \mathbb{P}(Z \in (-\infty, \bar{f})) = \int_{-\infty}^{\bar{f}} p_Z(z)dz = F_Z(\bar{f}) \).

The CC (1.1b) requires that the event \( \{ f(d, \Xi) \in (-\infty, \bar{f}) \} \) occurs with probability of at least \( 1 - \alpha \). Since \( \mathbb{P}(Z \leq \bar{f}) = F_Z(\bar{f}) \), the CC can also be written as

\[
F_{f(d,\Xi)}(\bar{f}) \geq 1 - \alpha \tag{1.2}
\]

or, equivalently, \( 1 - F_{Z}(\bar{f}) = \mathbb{P}(Z > \bar{f}) \leq \alpha \). We also recall that the \((1 - \alpha)\)-quantile of \( Z \) is defined as:

\[
Q_{Z}(1 - \alpha) \in \operatorname{argmin}_{t \in \mathbb{R}} \{ F_{Z}(t) \geq 1 - \alpha \}. \tag{1.3}
\]

In the stochastic programming literature, the quantile is also known as the value-at-risk (VaR), which we define as \( \text{VaR}_{1-\alpha}(Z) := Q_Z(1-\alpha) \). The definition of the quantile reveals that this can be expressed as the inverse of the cdf and thus CC (1.1b) can be written as \( \bar{f} \geq F_{Z}^{-1}(1 - \alpha) = Q_{Z}(1 - \alpha) \) or:

\[
Q_{f(d,\Xi)}(1 - \alpha) \leq \bar{f}. \tag{1.4}
\]

or \( \text{VaR}_{1-\alpha}(f(d,\Xi)) \leq \bar{f} \).

Another important observation is that \( \mathbb{E}[1_{\mathcal{W}}(Z)] = \mathbb{P}(Z \in \mathcal{W}) \) holds, where \( 1_{\mathcal{W}} : \mathbb{R} \to \{0, 1\} \) denotes the indicator function of set \( \mathcal{W} \) (i.e., \( 1_{\mathcal{W}}(Z) = 1 \) if \( Z \in \mathcal{W} \) and \( 1_{\mathcal{W}}(Z) = 0 \) if \( Z \notin \mathcal{D} \)). As a result, CC (1.1b) can be written as

\[
\mathbb{E}[1_{(-\infty, \bar{f})}(f(d, \Xi))] \geq 1 - \alpha, \tag{1.5}
\]

or, equivalently, \( \mathbb{E}[1_{(f, \infty)}(f(d, \Xi))] \leq \alpha \).

We define the feasible set of CC-P as \( \mathcal{D}(\alpha) := \mathcal{D} \cap \mathcal{P}(\alpha) \), where \( \mathcal{P}(\alpha) := \{ d \mid \mathbb{P}(f(d, \Xi) \leq \bar{f}) \geq 1 - \alpha \} \) and we assume \( \mathcal{D}(\alpha) \) to be compact and non-empty for all \( \alpha \in (0, 1) \). We denote an optimal solution and objective value of CC-P as \( d^*(\alpha) \) and \( \varphi^*(\alpha) \), respectively. In this work, we focus our attention on problems with a single CC but the concepts discussed can also be applied to multiple single CCs of the form \( \mathbb{P}(f_i(d, \Xi) \leq \bar{f}_i) \geq 1 - \alpha_i, \ i = 1, ..., r \). Formulations with joint chance constraints are significantly more complex and are left as a topic of future work.

A distinguishing and challenging feature of the CC (1.1b) is that it cannot be handled directly by off-the-shelf optimization solvers. The CC can be reformulated into a standard NLP when the quantile function \( Q_{f(d,\Xi)}(1 - \alpha) \) can be expressed in algebraic form. This property has been exploited extensively in the special case in which the mapping \( f(d, \Xi) \) is linear in both arguments and the random data vector \( \Xi \) is Gaussian [1]. Under these conditions, the random variable \( f(d, \Xi) \) is Gaussian for all \( d \in \mathcal{D} \) and its quantile can be expressed as a weighted sum of its two moments (the expectation and the variance). Exact solutions can also be obtained numerically when the cdf of \( f(d, \Xi) \) and its derivatives can be computed explicitly [2]. For more general settings it is possible to derive exact reformulations using integer variables, as originally proposed in [3]. Unfortunately, in the context
of CC-P, integer reformulations would lead to large-scale and nonconvex mixed-integer nonlinear programs (MINLPs).

Tractable but conservative approximations of CC-P can be used to avoid the need for solving MINLPs. A straightforward conservative approximation can be obtained by using the so-called scenario-based approach [4, 5]. In this approach, we solve a stochastic NLP that enforces $f(d, \Xi) \leq 0$ with probability one (for all scenarios). Such an approach leads to structured NLPs, which can in turn be solved using parallel interior-point solvers [6]. 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 CC [7]. The authors in [8] propose a difference of convex functions (DC) approximation and they show that the approximation can be made equivalent to CC-P. This approach, however, requires specialized solution algorithms that are not guaranteed to work in a general nonconvex NLP setting.

In this work, we review exact approaches to handle CC-P. We consider the special case in which the algebraic form of the quantile function is known (or approximately known) and propose to use moment matching to compute its parameters. We argue that this procedure can be applied to a wide range of distributions that go beyond Gaussian such as the uniform, log-normal, generalized extreme value (Weibull, Frechet, Gumbel), Laplace, exponential, and logistic distributions. The reason is that random phenomena often observed in science and engineering can be explained using asymptotic results such as the central limit and the extreme value (Fisher–Tippett) theorem and/or because basic transformations and fundamental relationships between distributions can be exploited. While this approach cannot be applied to general settings, we believe that there is value in studying the structure of the problem at hand to explore if suitable algebraic approximations emerge. To handle more general settings, we consider the use of a recently proposed sigmoidal approximation approach, which provides a mechanism to find exact solutions for CC-P. We demonstrate that this approach significantly outperforms existing conservative approximations such as CVaR and the scenario approach.

## 2 Moment Matching

A fundamental complication that arises in handling CCs is that the shape of the pdf of $Z$ is a complex function of the decision variables $d$ and of the physical model $f(d, \Xi)$. Consequently, it is in general difficult to anticipate the shape of the pdf (and thus of its associated cdf and quantile function). In certain settings, however, one might be able to anticipate the shape of the pdf by analyzing how uncertainty in $\Xi$ propagates through the model and by exploiting fundamental properties of random variables and asymptotic statistical results. Unfortunately, even if the shape of pdf is known, its parameters (e.g., location, scale, rate) are also complex functions of the decision variables of the model and thus cannot be pre-computed. The key observation that we make is that such parameters can in fact be computed on-the-fly by using moment matching techniques.

The first and second moments of the random variable $Z$ are the expectation $\mathbb{E}[Z]$ and the variance $\mathbb{V}[Z]$, respectively. These moments can often be expressed as algebraic functions of the parameters of
the pdf of $Z$ and thus the pdf parameters can be inferred from empirical observations of the moments. This is a standard approach used for fitting pdfs to empirical data.

For the Gaussian variable $Z \sim \mathcal{N}(a, b)$ with location $a$ and scale parameter $b$ we have that $\mathbb{E}[Z] = a$ and $\mathbb{V}[Z] = b^2$. As discussed previously, one can guarantee that $Z = f(x, \Xi)$ is Gaussian if the constraint mapping is linear and the random input vector $\Xi$ is Gaussian. For instance, we have that the linear transformation $m_1 + m_2 Z$ satisfies $m_1 + m_2 Z \sim \mathcal{N}(m_1 + am_2, bm_2)$. Consequently, if one has empirical data (observations) on the expectation and variance of $Z$ then one can infer the parameters $a, b$. To see how moment matching can be applied to reformulate CCs, we recall that the quantile function of a Gaussian variable is $Q_Z(1 - \alpha) = a + b \Phi^{-1}(1 - \alpha)$, where $\Phi^{-1}(1 - \alpha)$ is the $1 - \alpha$ quantile of the standard Gaussian variable (which can be pre-computed). Consequently, CC-P (1.1) can be reformulated as:

$$\begin{align*}
\min_{d \in \mathcal{D}, a, b} & \quad \varphi(d) \\
\text{s.t.} & \quad \mathbb{E}[f(d, \Xi)] = a \\
& \quad \mathbb{V}[f(d, \Xi)] = b^2 \\
& \quad a + b \Phi^{-1}(1 - \alpha) \leq \bar{f},
\end{align*}$$

(2.6a)

Note that $a, b$ are decision variables in the problem because the moments are functions of the decision variable $d$. Empirical data for the expectation and variance can be obtained by using Monte Carlo sampling (this approach is known as a sample average approximation or SAA). This gives the NLP:

$$\begin{align*}
\min_{d \in \mathcal{D}, a, b} & \quad \varphi(d) \\
\text{s.t.} & \quad \frac{1}{|\Omega|} \sum_{\xi \in \Omega} f(d, \xi) = a \\
& \quad \frac{1}{|\Omega| - 1} \sum_{\xi \in \Omega} (f(d, \xi) - a)^2 = b^2 \\
& \quad a + b \Phi^{-1}(1 - \alpha) \leq \bar{f},
\end{align*}$$

(2.7a)

(2.7b)

(2.7c)

(2.7d)

where $\xi \in \Omega$ are independent samples of $\Xi$.

For a uniform random variable $Z \sim \mathcal{U}(a, b)$ we have that $\mathbb{E}[Z] = \frac{1}{2}(a + b)$ and $\mathbb{V}[Z] = \frac{1}{12}(b - a)^2$, and thus having data on the expectation and variance of $Z$, one can solve a system of two equations to determine the parameters $a$ and $b$. Linear transformations of uniform random variables are also uniform. The quantile function for a uniform random variable is $Q_Z(1 - \alpha) = a + (1 - \alpha)(b - a)$ for $\alpha \in (0, 1)$. The CC (1.1b) can thus be reformulated using the following set of constraints:

$$\begin{align*}
\mathbb{E}[f(d, \Xi)] & = \frac{1}{2}(a + b) \\
\mathbb{V}[f(d, \Xi)] & = \frac{1}{12}(b - a)^2 \\
a + (1 - \alpha)(b - a) & \leq \bar{f},
\end{align*}$$

(2.8a)

(2.8b)

(2.8c)

Gaussian random variables are of great practical interest because many random phenomena observed in scientific and engineering applications follow this behavior. This is also in part the result of one of the most fundamental results in statistics known as the central limit theorem.
states that the sum of \( n \) identically distributed and independent random variables (of any form) \( S_n = \frac{1}{n}(Z_1 + Z_2 + \cdots + Z_n) \) is a Gaussian variable as \( n \to \infty \):

\[
\lim_{n \to \infty} S_n \sim \mathcal{N}(a, b/\sqrt{n}). \tag{2.9}
\]

An analogous and often overlooked result in statistics, known as the extreme value theorem (Fisher-Tippett theorem), states that the maximum of independent and identically distributed random variables (of any form) \( M_n = \max\{Z_1, Z_2, ..., Z_n\} \) follows a generalized extreme value (GEV) distribution as \( n \to \infty \):

\[
\lim_{n \to \infty} M_n \sim \text{GEV}(a, b, c). \tag{2.10}
\]

The quantile function of the GEV distribution is:

\[
Q_Z(1 - \alpha) = \begin{cases} 
    a - b(1 + \log(1 - \alpha)^{-c})/c & c < 0 \text{ or } c > 0 \\
    a - b\log(-\log(1 - \alpha)) & c = 0 
\end{cases} \tag{2.11}
\]

Here, \( a, b, c \) are the location, scale, and shape parameters of the distribution. A random variable that follows a GEV distribution is often denoted as \( Z \sim \text{GEV}(a, b, c) \). The shape parameter \( c \) defines different types of distributions. In particular, when \( c = 0 \) the GEV distribution becomes the Gumbel distribution and the CC can be formulated using the following set of constraints:

\[
\begin{align*}
\mathbb{E}[f(d, \Xi)] &= a + \gamma b \tag{2.12a} \\
\mathbb{V}[f(d, \Xi)] &= \frac{\pi^2}{6}b^2 \tag{2.12b} \\
a - b\log(-\log(1 - \alpha)) &\leq \bar{f} \tag{2.12c}
\end{align*}
\]

where \( x, a, b \) are decision variables. Interestingly, linear transformations of Gumbel random variables (e.g., if \( Z \sim \text{GEV}(a, b, 0) \) then \( m_1 + m_2Z \sim \text{GEV}(m_1 + am_2, bm_2, 0) \)). GEV variables are often used to capture extreme and rare events such as flooding and earthquake events. This observation is important because the design of many engineered systems such as infrastructures is driven by distributions of extreme events. In the context of NLPs with CCs, GEV variables have been recently used to design control systems for wind turbines, because extreme mechanical stress events limit the controller actuation region \[9\]. In particular, it is observed that extreme (peak) mechanical loads observed over log time horizons follow a Gumbel distribution (as suggested by the extreme value theorem) \[10\].

There exist many relationships between pdfs and functional transformations that can be exploited to infer the structural form of the density of \( f(d, \Xi) \). Logarithmic transformations are often encountered in physical models and, of particular interest is the transformation of Gaussian (normal) random variables to log-normal ones. This is done by noticing that, if \( \log(Z) \sim \mathcal{N}(a, b) \) then \( Z \) is log-normal (denoted as \( Z \sim \text{LogN}(a, b) \)). Another way of expressing this is that, if \( Z \sim \mathcal{N}(a, b) \), then \( \exp(X) \sim \text{LogN}(a, b) \). The quantile function of a log-normal variable is:

\[
Q_Z(1 - \alpha) = \exp\left(a + b\Phi^{-1}(1 - \alpha)\right). \tag{2.13}
\]
The log-normal distribution is often used to model variables that are skewed and that can only take positive values (e.g., flows, concentrations). Log-normal variables also have interesting transformation properties; for instance, the product of log-normal variables is log normal, powers of a log-normal variable are log-normal, and the inverse of a log-normal variable is log-normal. It is also known that the sum of Gumbel random variables follows a logistic distribution. The quantile function of the logistic distribution is:

$$Q_Z(1 - \alpha) = a + b \log((1 - \alpha)/\alpha)$$  \hspace{1cm} (2.14)

while the moments of this distribution are $\mathbb{E}[Z] = a$ and $\mathbb{V}[Z] = b^2 \pi^2 /3$. Linear transformations also preserve the structure of logistic random variables. Moreover, an exponential transformation of a Logistic random variable $\exp(Z)$ follows a log-Logistic distribution (an analog of normal and log-normal variables), which has a quantile function of the form:

$$Q_Z(1 - \alpha) = a \left( \frac{1 - \alpha}{\alpha} \right)^{1/b}$$  \hspace{1cm} (2.15)

and the moments are algebraic functions of $a$ and $b$. It is also known that if $Z$ follows an exponential distribution (denoted as $Z \sim \text{Exp}(a)$) then $\log Z$ is a Gumbel distribution. The quantile function of an exponential distribution is given by:

$$Q_Z(1 - \alpha) = -\frac{1}{a} \log(\alpha),$$  \hspace{1cm} (2.16)

and the first moment is $\mathbb{E}[Z] = 1/a$.

The Weibull distribution is another special case of the GEV distribution, which is obtained for $c < 0$. This distribution is used to model non-negative random variables and commonly arises in reliability analysis. The quantile of a Weibull distribution has the algebraic form:

$$Q_Z(1 - \alpha) = a \left( -\log(\alpha) \right)^{1/b}.$$  \hspace{1cm} (2.17)

Unfortunately, the moments of a Weibull random variable involve complex gamma functions of $a, b$ and thus moment matching is difficult to implement computationally. Nevertheless, the Weibull distribution is of practical interest because it often arises from transformations of random variables. For instance, the logarithm of a uniform random variable yields a Weibull variable and powers of an exponential random variable are Weibull.

The Rayleigh distribution is another useful way of representing non-negative random variables. This distribution has a quantile function:

$$Q_Z(1 - \alpha) = a \sqrt{-2 \log(\alpha)},$$  \hspace{1cm} (2.18)

and we note that this distribution only has one parameter. Such parameter can be obtained by matching the first moment, which is given by $\mathbb{E}[Z] = a \sqrt{\pi/2}$. Rayleigh random variables have interesting connections with other random variables. For instance, if $Z_1, Z_2$ are independent normal variables then $Z = \sqrt{Z_1^2 + Z_2^2}$ is Rayleigh distributed (denoted as $Z \sim R(a)$). This indicates that Rayleigh variables are Euclidean norms of random vectors with independent entries. It is also known that
if $Z$ follows an exponential distribution then $\sqrt{Z}$ is Rayleigh and if $Z$ is uniform then $a\sqrt{-2\log Z}$ is Rayleigh. A useful summary on properties of transformations of random variables can be found in [11].

It is in general difficult to infer the exact structure of the distribution of $Z = f(d, \Xi)$ due to complex nonlinear transformations arising in physical models and the presence of inequality constraints. When structure analysis is not possible one can resort to Monte Carlo analysis to determine the shape of $f(d, \Xi)$ at different values of $x$ and check if the shape persists and if it follows any well-known distributions. If the shape does not persist then one can also resort to general solution techniques, as we describe next.

2.1 Approximations for General Settings

We review approaches to deal with CC-P (1.1b) in a general setting in which no prior knowledge on the shape of $f(d, \Xi)$ is available.

2.1.1 CVaR Approximation

To facilitate notation we redefine $Z \leftarrow Z - \bar{f}$ so that the CC constraint becomes $\mathbb{P}(Z \leq 0) \geq 1 - \alpha$, which is equivalent to $\mathbb{P}(Z > 0) \leq \alpha$. Because $\mathbb{P}(Z > 0) = \mathbb{E}[1_{(0, \infty)}(Z)]$, the CC can be expressed as:

$$\mathbb{E}[1_{(0, \infty)}(f(d, \Xi) - \bar{f})] \leq \alpha.$$ (2.19)

A computationally practical approach to approximate the CC is to find a conservative approximation. This is done by finding an approximating function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ 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,

$$\mathbb{E}[\psi(t^{-1}Z)] \geq \mathbb{P}(Z > 0).$$ (2.20)

We can thus conclude that the satisfaction of the constraint:

$$\mathbb{E}[\psi(t^{-1}Z)] \leq \alpha,$$ (2.21)

implies that $\mathbb{P}(Z > 0) \leq \alpha$ is satisfied (and so does $\mathbb{P}(f(d, \Xi) \leq \bar{f}) \geq 1 - \alpha$). Because (2.21) is valid for all $t > 0$ we also have, if $\psi(\cdot)$ is convex, then:

$$\inf_{t > 0} \left\{t \mathbb{E}[\psi(t^{-1}Z)] - t\alpha \right\} \leq 0$$ (2.22)

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:

$$\inf_{t > 0} \{\mathbb{E}[[Z + t]_+] - t\alpha\} \leq 0.$$ (2.23)

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

$$\inf_{t \in \mathbb{R}} \{\alpha^{-1}\mathbb{E}[[Z + t]_+] - t\} \leq 0.$$ (2.24)
By redefining \( t \leftarrow -t \) and recalling that
\[
\text{CVaR}_{1-\alpha}(Z) := \inf_{t \in \mathbb{R}} \left\{ t + \alpha^{-1}\mathbb{E}[\max\{Z - t, 0\}] \right\}, \tag{2.25}
\]
we can see that (2.24) can be used to derive a conservative approximation of CC-P (1.1) of the form:
\[
\min_{d \in D} \varphi(d) \tag{2.26a}
\]
\[
\text{s.t. } \text{CVaR}_{1-\alpha}(f(d, \Xi) - \bar{f}) \leq 0. \tag{2.26b}
\]
We denote an optimal objective value and solution of this problem (denoted as CVaR-P) as \( \varphi_c(\alpha) \) and \( d_c(\alpha) \), respectively. We define the feasible set of CVaR-P as \( D_c(\alpha) \) and note, because CVaR provides a conservative approximation, that \( D_c(\alpha) \subseteq D(\alpha) \). Consequently, any feasible solution \( d_c(\alpha) \) 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 \( Z_c(\alpha) := f(d_c(\alpha), \Xi) \) and recall that [12]:
\[
\text{VaR}_{1-\alpha}(Z_c(\alpha)) = \arg\min_{t} \left\{ t + \alpha^{-1}\mathbb{E}[\max\{Z_c(t) - t, 0\}] \right\}, \tag{2.27}
\]
and thus \( \text{VaR}_{1-\alpha}(Z_c(\alpha)) \leq \text{CVaR}_{1-\alpha}(Z_c(\alpha)) \). This observation also highlights that CVaR provides a conservative approximation for the CC. We will find it convenient to define the constant:
\[
\gamma_\alpha := -t_c(\alpha)^{-1}, \tag{2.28}
\]
with \( t_c(\alpha) \in \arg\min_{t} \{ t + \alpha^{-1}\mathbb{E}[Z_c(t) - t]_+ \} \).

A key advantage of the CVaR approximation is that it can be cast as a standard NLP. Moreover, if \( f(d, \xi) \) is convex in \( d \) for given \( \xi \), CVaR is also convex in \( d \). 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.

2.1.2 Sigmoidal Approximation

An important observation is that the indicator function can be outer-approximated by using a sigmoidal function of the form:
\[
\psi_{\mu,\tau}(z) := \frac{1 + \mu}{\mu + e^{-\tau z}}, \tag{2.29}
\]
where \( \mu, \tau \in \mathbb{R}_+ \) are parameters. This sigmoidal function is a special case of the generalized logistic function, which is a standard smooth (and thus differentiable) approximation of the indicator function [13]. The associated approximate CC constraint takes the form:
\[
\mathbb{E}[\psi_{\mu,\tau}^\mu(f(d, \Xi))] \leq \alpha. \tag{2.30}
\]
In this work, we consider a tailored variant of the above sigmoidal function of the form:
\[
\psi_{ss}^{\mu,\tau}(z) := \left[ 2 \frac{1 + \mu}{\mu + e^{-\tau z}} - 1 \right]_+. \tag{2.31}
\]
This gives the approximate CC constraint:
\[
E \left[ \psi_{ss}^{\mu,\tau}(f(d, \Xi)) \right] \leq \alpha.
\] (2.32)

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 \( \delta := \frac{1}{\tau} \log(2+\mu) \). In [14] it is shown that sigmoid functions provide conservative approximations for CC (1.1b) for any \( \mu, \tau \in \mathbb{R}_+ \), \( \alpha \in (0, 1] \), and \( d \in D \).

![Figure 1: Conservative approximation of the indicator function using sigmoidal functions.](http://zavalab.engr.wisc.edu)

We use the tailored sigmoidal function to define the Sigmoidal Value-at-Risk (SigVaR):
\[
\text{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) := \inf_{t \in \mathbb{R}} \{ E \left[ \psi_{ss}^{\mu,\tau}(Z - t) \right] \leq \alpha \}
\]
\[
= \inf_{t \in \mathbb{R}} \left\{ E \left[ \left[ \frac{1 + \mu}{\mu + e^{-\tau(Z-t)}} - 1 \right]_+ \right] \leq \alpha \right\}. \tag{2.33}
\]

and we use this to formulate the approximation of CC-P:
\[
\min_{d \in D} \varphi(d) \tag{2.34a}
\]
\[
\text{s.t. } \text{SigVaR}_{1-\alpha}^{\mu,\tau}(f(d, \Xi)) \leq \bar{f}. \tag{2.34b}
\]

We define an optimal objective and solution of (2.34) as \( \varphi_{ss}^{\mu,\tau}(\alpha) \) and \( \bar{x}_{ss}^{\mu,\tau}(\alpha) \), respectively. We also define the feasible set of (2.34) as \( D_{ss}^{\mu,\tau}(\alpha) \). Because the sigmoidal function is a conservative approximation, we have that \( D_{ss}^{\mu,\tau}(\alpha) \subseteq D(\alpha) \) 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 \( \text{VaR}_{1-\alpha}(Z) = \arg\min_{t \in \mathbb{R}} \{ P(Z - t > 0) \leq \alpha \} \) can also be expressed in terms of the indicator function:
\[
\text{VaR}_{1-\alpha}(Z) = \arg\min_{t \in \mathbb{R}} \{ E \left[ 1_{(0,\infty)}(Z - t) \right] \leq \alpha \}. \tag{2.35}
\]

Because the sigmoid function \( \psi_{ss}^{\mu,\tau}(\cdot) \) is a conservative approximation of \( 1_{(0,\infty)}(\cdot) \), we have that
\[
\text{SigVaR}_{1-\alpha}^{\mu,\tau}(Z) \geq \text{VaR}_{1-\alpha}(Z), \tag{2.36}
\]
for all $\alpha \in (0, 1]$ and $\mu, \tau \in \mathbb{R}_+$. Consequently, SigVaR can be interpreted as an approximate quantile and (2.34) is a conservative representation of CC-P. As in the case of the VaR representation of CC-P, problem (2.34) is not particularly attractive for computation. However, this problem also has the following equivalent representation (denoted as SigVar-P):

$$\min_{d \in D} \varphi(d) \quad \text{(2.37a)}$$
$$\text{s.t. } \mathbb{E} \left[ \psi^{\mu,\tau}_{ss}(f(d, \Xi) - \bar{f}) \right] \leq \alpha. \quad \text{(2.37b)}$$

The SigVar-P problem has the following interesting properties, established in [14]:

- Let $\tau(\mu) := (1 + \mu)\theta$ with $\theta > 0$. Then $\lim_{\mu \to \infty} \lambda^{\mu,\tau(\mu)}(\alpha) = \lambda(\alpha)$ and $\lim_{\mu \to \infty} \varphi^{\mu,\tau(\mu)}(\alpha) = \varphi(\alpha)$. This result indicates that SigVaR-P approaches CC-P as $\mu \to \infty$.

- Let $\tau(\mu) := (1 + \mu)\theta$ with $\theta > 0$. We have that $\lambda^{\mu,\tau^+(\mu)}(\alpha) \supset \lambda^{\mu,\tau(\mu)}(\alpha)$ and $\varphi^{\mu,\tau^+(\mu)}(\alpha) \leq \varphi^{\mu,\tau(\mu)}(\alpha)$ for $\mu^+ > \mu > 0$ and for all $\alpha \in (0, 1]$. This result indicates that approximations of increasing quality can be constructed by increasing $\mu$ (and $\tau$).

- 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 $\bar{\mu} - \log(2 + \bar{\mu}) = 1$) and $\tau_\alpha := \frac{2 + 1}{2} \gamma c$ (with $\gamma c$ obtained from (2.28)). We have that the feasible set and optimal objective values for CVaR-P and SigVaR-P satisfy $\lambda_c(\alpha) \subseteq \lambda^{\mu,\tau(\mu)}(\alpha)$ and $\varphi^{\mu,\tau(\mu)}(\alpha) \leq \varphi_c(\alpha)$. This result indicates that the parameters of SigVaR-P can be selected in such a way that the formulation provides an approximation of CC-P that is at least as good as that of CVaR-P. Moreover, it indicates that we can use the solution of CVaR-P to find an initial guess for SigVar-P (by using the initial parameter estimate $\tau = \tau_\alpha$).

- The SAA approximation of SigVaR-P can be cast as a standard NLP of the form:

$$\min_{d \in D, \xi \in \mathbb{R}, \phi \in \mathbb{R}_+} \varphi(d) \quad \text{(2.38a)}$$
$$\text{s.t. } z_\xi = f(d, \Xi) - \bar{f}, \ \xi \in \Omega \quad \text{(2.38b)}$$
$$\phi_\xi \geq \frac{1 + \mu}{\mu + e^{-\tau_\xi} - 1}, \ \xi \in \Omega \quad \text{(2.38c)}$$
$$\frac{1}{|\Omega|} \sum_{\xi \in \Omega} \phi_\xi \leq \alpha. \quad \text{(2.38d)}$$

This indicates that SigVaR can be used to solve large-scale NLPs with chance constraints by using powerful serial and parallel solvers such as Ipopt and PIPS-NLP [15, 16].

### 2.2 Almost-Surely (Scenario) Approximation

A straightforward approach to obtain a conservative approximation of CC is to enforce the constraint $\mathbb{P}(f(d, \Xi) \leq \bar{f}) \geq 1 - \alpha$ with $1 - \alpha = 1$ (with probability one or almost surely). This is equivalent to enforce the constraint $f(x, \Xi) \leq \bar{f}$ for all possible realizations of $\Xi$. The SAA approximation is given by the standard scenario-based stochastic programming formulation:
The AS approximation is popular because it is easy to implement and can be solved existing serial and parallel NLP solvers.

\[
\begin{align*}
\min_{d \in D} & \quad \varphi(d) \\
\text{s.t.} & \quad f(d, \Xi) \leq \bar{f}, \quad \xi \in \Omega.
\end{align*}
\]

Figure 2: Schematic representation of flare stack system.

3 Case Study: Flare System Design

We consider the design of a flare stack system that combusts a waste fuel gas flow (see Figure 2). This study is used to compare the effectiveness of different solution approaches for CC-P. Gas flares are used as safety (relief) devices to manage abnormal situations in infrastructure systems (natural gas and oil processing plants and pipelines), manufacturing facilities (chemical plants, offshore rigs), and power generation facilities. Abnormal situations include equipment failures, off-specification products, and excess materials in start-up/shutdown procedures. In particular, flares prevent over-pressuring of equipment and use combustion to convert flammable, toxic or corrosive vapors to less dangerous compounds [17]. Flare design is influenced by several uncertain factors such as the amount and composition of the waste flow stream to be combusted and the ambient conditions. These systems are currently designed based on typical historical values for waste fuel gases and ambient conditions [17,18]. Consequently, an improperly designed flare can be susceptible to extreme events not experienced before. Here, we propose to use stochastic programming formulations to systematically capture uncertain conditions in the design procedure.
3.1 Physical Model

In this section, we present a mathematical model for sizing a flare system. The design goals are to minimize capital cost while controlling the radiation level at ground level (which is a function of the input waste flow to be combusted). The model is derived from the American Petroleum Institute (API) standard 521 [18].

The heat released by combustion $H$ (BTU/h) is a function of the random input waste flow $Q$ (lb/h) and the heat of combustion $h_c$ (BTU/lb):

$$H = h_c Q$$  \hspace{1cm} (3.40)

The flame length $L$ (ft) can be calculated as a function of the released heat using an approximation of the form:

$$\log L = a_1 \log H - a_2$$  \hspace{1cm} (3.41)

The flare stack diameter $t$ (ft) is sized on a velocity basis. This is done by relating this to the Mach number $M$ and the waste flow as:

$$M^2 = \frac{a_3}{t^2} Q^2.$$  \hspace{1cm} (3.42)

The flare tip exit velocity $U$ (ft/s) is function of the flow and the diameter:

$$U = a_4 \frac{Q}{t^2}$$ \hspace{1cm} (3.43)

The wind speed $w$ (ft/s) is an important environmental factor that affects the tilting of the flame and the distance from the centre of the flame. The following correlations capture the flame distortion as a result of the wind speed and the exit velocity:

$$\log \Delta X = \log(a_5 L) + a_6 (\log w - \log U)$$ \hspace{1cm} (3.44)

$$\log \Delta Y = \log(a_7 L) - a_8 (\log w - \log U)$$ \hspace{1cm} (3.45)

Here, $\Delta X$ and $\Delta Y$ (ft) are the horizontal and vertical distortions. The distortions are used to compute the horizontal $X$, vertical $Y$, and total distance $D$ (ft) to a given ground-level safe point $(r, 0)$ as:

$$X = r - \frac{1}{2} \Delta X$$ \hspace{1cm} (3.46)

$$Y = h + \frac{1}{2} \Delta Y$$ \hspace{1cm} (3.47)

$$D^2 = X^2 + Y^2.$$ \hspace{1cm} (3.48)

Here, $h$ (ft) is the flare height. The flame radiation $K$ (BTU/h ft$^2$) is a function of the heat released and the total distance:

$$K = a_9 \frac{H}{D^2}. $$ \hspace{1cm} (3.49)

A primary safety goal in the flare stack design problem is to control the risk that the radiation exceeds a certain threshold value $\bar{k}$ (BTU/h ft$^2$) at the ground-level reference point $(r, 0)$. This is modeled using the CC:

$$\mathbb{P}(K \leq \bar{k}) \geq 1 - \alpha.$$ \hspace{1cm} (3.50)
The objective function is the cost (USD), which is a function of height and diameter:

\[ \varphi(t, h) = (a_{10} + a_{11} \cdot t + a_{12} \cdot h)^2. \] (3.51)

The height and the diameter play a key role in controlling the radiation at the reference point (i.e., a higher and wider flare reduces the radiation intensity). As a result, there is an inherent trade-off between capital cost and safety that needs to be carefully handled. In particular, an unnecessary flare over design can be quite expensive.

We highlight that some of the above model equations are obtained by applying logarithmic transformations to the original physical equations while some equations have been left in their original form. This is done in order to facilitate the exposition of the structural model analysis presented in the next section. In particular, the model seeks to highlight how uncertainty in the waste flow \( Q \) propagates through the model leading to uncertainty in the radiation \( K \), which defines the chance constraint. The parameters \( a_{1}, ..., a_{12} \) can be derived from physical and economic parameters (e.g., emissivity, molecular weight, and temperature). We do not provide the original model equations and detailed expressions for the parameters in order to simplify the presentation. The interested reader can find the detailed model (in its original form) and associated parameters in https://github.com/zavalab/JuliaBox/tree/master/FlareDesign.

### 3.2 Structural Model Analysis

In order to determine the shape of the pdf of \( K \), it is necessary to understand how the waste flow \( Q \) (corresponding to the input random data \( \Xi \)) propagates through the physical model to obtain \( K \) (captured by the function \( f(d, \Xi) \), where \( d \) are the design variables corresponding to the diameter \( t \) and height \( h \)). We explore a couple of cases to illustrate how this can be done.

Assume that the waste flow follows a log-normal distribution \( Q \sim \text{LogN} \). From (3.40), we have that \( H \sim \text{LogN} \). We know that \( \log H \sim \mathcal{N} \) and from (3.41) we thus have that \( \log D^2 \sim \mathcal{N} \). From (3.43) we have that, for fixed \( t, U \sim \log N \) and thus \( \log U \sim \mathcal{N} \). From (3.44) and (3.45), we have that \( \log \Delta X \sim \mathcal{N} \) and \( \log \Delta Y \sim \mathcal{N} \) and from (3.46), we have that \( X \sim \log N \) and \( Y \sim \log N \). It is also known that \( X^2 \sim \log N \) and \( Y^2 \sim \log N \) but the sum of log-normal variables cannot be guaranteed to be log-normal. It has been reported, however, that a log-normal distribution provides an accurate approximation on the right tail of the distribution for the sum of log-normal variables (as needed in CCs) [19]. Consequently, it is safe to assume that \( D^2 \sim \log N \). Now note that (3.49) can be transformed as \( \log K = \log a_9 + \log H - \log D^2 \) because \( K, H, \) and \( D^2 \) are nonnegative variables. Since \( \log H \sim \mathcal{N} \) and \( \log D^2 \sim \mathcal{N} \), we have that \( \log K \sim \mathcal{N} \) and \( K \sim \log N \). Consequently, the CC (3.50) can be formulated as:

\[ \mathbb{E}[K] = a \] (3.52a)
\[ \mathbb{V}[K] = b^2 \] (3.52b)
\[ a + b \Phi^{-1}(1 - \alpha) \leq \log \bar{k} \] (3.52c)

The above discussion illustrates that the structure of the CC can be inferred in some non-obvious cases by exploiting fundamental properties of random variables. We highlight that, because we as-
sumed a structure for $D^2$ that is not guaranteed in general, it is important to validate that the distribution of $K$ is log-normal and that the CC is satisfied at the desired probability level at the solution of the problem.

To see how one might run into obstacles when inferring the underlying distribution of $K$ from the model structure, assume now that the random input flow follows an exponential distribution $Q \sim \text{Exp}$. From (3.40) we have that $H \sim \text{Exp}$. We know that the logarithm of an exponential random variable is Gumbel and thus $\log Q \sim \text{GEV}$. The linear transformation of a Gumbel variable is Gumbel and thus $\log H \sim \text{GEV}$. Also, note that $\log L$ and $\log U$ can both be expressed in terms of $\log Q$ alone and thus $\log \Delta X \sim \text{GEV}$, and $\log \Delta Y \sim \text{GEV}$. Consequently, $\Delta X$ and $\Delta Y$ are exponential. We now note that $D^2$ can be expressed as a sum of $X$, $X^2$, $Y$, and $Y^2$. It is known that any positive power of an exponential random variable is Weibull and it is also known that the exponential random variable is a special case of a Weibull variable and thus $D^2$ is a sum of Weibull variables. Unfortunately, nothing is known about the sum of Weibull variables so the structural analysis breaks down. Interestingly, numerical experiments shown in the next section show that the radiation indeed follows an exponential distribution when the input flow is exponential as well. This observation can, in fact, be attributed to a different structural aspect of the model. In particular, by inspecting (3.49) more closely, one can find that the term $a_9/D^2$ is a random variable with variance $O(10^{-10})$ for the range of design variables of interest. Given the small variance, this random variable effectively behaves as a constant and thus the radiation $K$ is just a linear transformation of $H$, which is in turn a linear transformation of $Q$. Consequently, we can expect that $K$ approximately follows an exponential distribution. This observation also indicates that any distribution of $Q$ that is not affected by a linear transformation will yield the same distribution for $K$ (e.g., normal, uniform, logistic, log-normal, and exponential). We thus highlight that, while it would be ideal to guarantee that the shape of the pdf always has a certain form, in practice, it might be sufficient to ensure that this shape is preserved in a given domain of interest.

### 3.3 Results

Our goal is to design a flare system that minimizes cost and satisfies the CC on the radiation using the AS, CVaR, SigVar, and MM approaches. In order to validate the observations made in the previous structural analysis, we consider a case in which the input flow to the flare follows a log-normal distribution (with mean $a = 10,000$ lb/h and standard deviation $b = 3,000$ lb/h) and a case in which it follows an exponential distribution (with a rate parameter $a = 21,000$ lb/h). All formulations were solved using one thousand random samples for the inlet flow (we use the same samples for all of approaches in order to achieve consistency). The flare design must satisfy a chance constraint on the radiation with a maximum threshold of $\bar{k}=2000$ Btu/(hr ft$^2$) and with a probability of $1 - \alpha =0.95$. The optimization formulations are all standard NLPs that were implemented using the open-source modeling language JuMP [20] and solved with Ipopt [15].
3.3.1 Optimal Design Specifications

Tables 1 and 2 present the optimal cost, diameter, and height for the flare stack under the different approaches. The results show that, for both distributions, the optimal flare stack diameter is the same for CVaR, SigVaR, and MM. This is due to the fact that the diameter sizing cost (constant $a_{11}$) is twelve times greater than height sizing cost (constant $a_{12}$). Consequently, it is cheaper to increase the height than to increase the diameter in order to satisfy the radiation restrictions.

The results also show that the AS approach leads to extremely conservative designs. For the log-normal case the AS design is 64% more expensive than the MM design while for the exponential case it is 44% more expensive. This is due to the fact that the AS approach does not allow for explicit control of the probability of constraint satisfaction. The CVaR approximation reduces this extreme conservatism but still over designs the system in a perceptible manner. For the log-normal distribution, the optimal height achieved with CVaR is 67.70% larger than that obtained with MM. For the exponential distribution, the optimal height is 49.76% larger than that obtained with MM. We also see that the CVaR design is 20% more expensive for the log-normal case and 17% more expensive for exponential case, both relative to the MM designs. We thus conclude that, while the CVaR approximation can help mitigate extreme conservatism of the AS approach and solve the problem without any knowledge of the radiation pdf, this approach is still quite conservative.

The results also show that the solutions of SigVaR and MM are close (the SigVaR design is 1% more expensive for the log-normal case and 3% more expensive for the exponential case). This reinforces the observations of the structural model analysis, which indicate that the radiation follows the same distribution of the waste flow. Moreover, the results highlight the fact that the SigVaR solution is conservative.

<table>
<thead>
<tr>
<th></th>
<th>AS</th>
<th>CVaR</th>
<th>SigVaR</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost (USD)</td>
<td>233,179</td>
<td>172,329</td>
<td>143,609</td>
<td>142,128</td>
</tr>
<tr>
<td>Diameter (ft)</td>
<td>1.70</td>
<td>1.70</td>
<td>1.70</td>
<td>1.70</td>
</tr>
<tr>
<td>Height (ft)</td>
<td>179.87</td>
<td>105.08</td>
<td>65.16</td>
<td>63.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AS</th>
<th>CVaR</th>
<th>SigVaR</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost (USD)</td>
<td>146,470</td>
<td>118,072</td>
<td>105,589</td>
<td>101,258</td>
</tr>
<tr>
<td>Diameter (ft)</td>
<td>1.30</td>
<td>1.30</td>
<td>1.30</td>
<td>1.30</td>
</tr>
<tr>
<td>Height (ft)</td>
<td>127.54</td>
<td>84.39</td>
<td>63.78</td>
<td>56.35</td>
</tr>
</tbody>
</table>

Tables 3 and 4 present the convergence history of the SigVaR approach. Here, each iteration solves a SigVaR problem with increasing parameter values $\mu$, $\tau$, leading to approximations of increasing quality. The initial iteration for the SigVaR scheme is the CVaR approximation. For the log-normal case, we can see that, after 10 iterations of the SigVaR scheme, the approximation is solved with $\mu = 2566.14$ and $\tau = 1.8181$ and the height and cost are reduced by 61.26% and 16.67%, respectively,
Table 3: SigVaR convergence history when waste flow follows a log-normal distribution.

<table>
<thead>
<tr>
<th>SigVaR Iter</th>
<th>Cost (USD)</th>
<th>Diameter (ft)</th>
<th>Height (ft)</th>
<th>μ</th>
<th>τ</th>
<th>Ipopt Iter</th>
</tr>
</thead>
<tbody>
<tr>
<td>–</td>
<td>172,329</td>
<td>1.7</td>
<td>105.08</td>
<td>2.506</td>
<td>0.0024</td>
<td>106</td>
</tr>
<tr>
<td>1</td>
<td>165,522</td>
<td>1.7</td>
<td>95.95</td>
<td>5.012</td>
<td>0.0043</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>160,548</td>
<td>1.7</td>
<td>89.15</td>
<td>10.024</td>
<td>0.0078</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>155,363</td>
<td>1.7</td>
<td>81.95</td>
<td>20.048</td>
<td>0.0149</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>150,458</td>
<td>1.7</td>
<td>75.02</td>
<td>40.096</td>
<td>0.0291</td>
<td>82</td>
</tr>
<tr>
<td>5</td>
<td>147,083</td>
<td>1.7</td>
<td>70.19</td>
<td>80.192</td>
<td>0.0575</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>145,324</td>
<td>1.7</td>
<td>67.66</td>
<td>160.384</td>
<td>0.1143</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>144,407</td>
<td>1.7</td>
<td>66.33</td>
<td>320.768</td>
<td>0.2279</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>143,912</td>
<td>1.7</td>
<td>65.16</td>
<td>641.536</td>
<td>0.4550</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>143,609</td>
<td>1.7</td>
<td>65.16</td>
<td>1283.07</td>
<td>0.9094</td>
<td>36</td>
</tr>
<tr>
<td>10</td>
<td>143,609</td>
<td>1.7</td>
<td>65.16</td>
<td>2566.14</td>
<td>1.8181</td>
<td>19</td>
</tr>
</tbody>
</table>

relative to the CVaR approximation. We also see diminishing cost returns as the parameters are increased and the cost eventually settles, indicating that the SigVaR approximation is close to optimal. We also see that the number of iterations needed by Ipopt is initially large but is significantly reduced at subsequent iterations. This is because we exploit warm-start information and the solution of the problems become increasingly closer as the iterations progress. For the exponential distribution case we see a similar behavior but the SigVaR approximation cannot be solved with Ipopt for large parameter values. Consequently, the best approximation that we could find is a conservative approximation of the MM solution (but this is significantly less conservative than that of CVaR). This result highlights the computational advantage provided by the MM approach (if the shape of the pdf can be anticipated).

Table 4: SigVaR convergence history when flow follows an exponential distribution.

<table>
<thead>
<tr>
<th>iter</th>
<th>Cost (USD)</th>
<th>Diameter (ft)</th>
<th>Height (ft)</th>
<th>μ</th>
<th>τ</th>
<th>Ipopt Iter</th>
</tr>
</thead>
<tbody>
<tr>
<td>–</td>
<td>118,073</td>
<td>1.3</td>
<td>84.39</td>
<td>2.506</td>
<td>0.0043</td>
<td>114</td>
</tr>
<tr>
<td>1</td>
<td>114,361</td>
<td>1.3</td>
<td>78.38</td>
<td>5.012</td>
<td>0.0074</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>112,156</td>
<td>1.3</td>
<td>74.76</td>
<td>10.024</td>
<td>0.0137</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>110,321</td>
<td>1.3</td>
<td>71.73</td>
<td>20.048</td>
<td>0.0261</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>108,772</td>
<td>1.3</td>
<td>69.15</td>
<td>40.096</td>
<td>0.0509</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>107,369</td>
<td>1.3</td>
<td>66.79</td>
<td>80.192</td>
<td>0.1006</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>106,427</td>
<td>1.3</td>
<td>65.20</td>
<td>160.384</td>
<td>0.2000</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>105,941</td>
<td>1.3</td>
<td>64.38</td>
<td>320.768</td>
<td>0.3987</td>
<td>66</td>
</tr>
<tr>
<td>8</td>
<td>105,705</td>
<td>1.3</td>
<td>63.98</td>
<td>641.536</td>
<td>0.7961</td>
<td>53</td>
</tr>
<tr>
<td>9</td>
<td>105,589</td>
<td>1.3</td>
<td>63.78</td>
<td>1283.07</td>
<td>1.5910</td>
<td>54</td>
</tr>
</tbody>
</table>
3.3.2 Probability Density Functions

The MM approach also provides the parameters for the pdf of the radiation at the optimal solution. For the exponential case, we know that the rate is the reciprocal of the mean radiation at the solution \( \mathbb{E}[K] = 1/a \) (we obtain an actual value at the solution of \( a = 0.0014978 \)). For the log-normal case, we have that the location parameter is the mean log radiation \( \mathbb{E}[\log(K)] = a \) (at the solution we obtain \( a = 5.66 \)) and we have that the scale parameter is the standard deviation \( \mathbb{V}[\log(K)] = b^2 \) (at the solution we obtain \( b = 1.178 \)).

Figure 3 shows the empirical pdfs for the input flow (assuming a log-normal distribution) and for the radiation at the optimal solution of the CVaR, SigVaR, and MM approaches. The histogram of the MM approach is also compared against the corresponding model pdf. We first observe that, for the MM approach, the distribution of the radiation is indeed log-normal. This indicates that the structure of the flare model preserves the log-normal shape of the input flow. We also observe that...
the histograms of SigVaR and MM are quite similar, with SigVaR being slightly more conservative. In particular, the tail of the SigVaR distribution is very similar to that of MM (the tail reaches values of 6,000 Btu/(hr ft$^2$) for SigVaR compared to 6,100 Btu/(hr ft$^2$) with MM). The histogram of CVaR further validates the observation that this approach is very conservative (the tail reaches values of 4000 Btu/(hr ft$^2$)).

Figure 4 shows the empirical distributions for the input flow (assuming an exponential distribution) and for the radiation at the optimal solution of the CVaR, SigVaR, and MM approaches. Here, we again confirm that the model structure preserves exponential shape of the input flow. For this case, the tail of the SigVaR distribution deviates from that of MM in a more perceptible manner (the tail reaches values of 3,900 Btu/(hr ft$^2$) for SigVaR compared to 4,200 Btu/(hr ft$^2$) with MM). The tail of CVaR reaches values of only 3000 Btu/(hr ft$^2$).

![Probability distributions](image)

Figure 4: Top-Left: Flow pdf assuming an exponential distribution, Top-Right: Radiation pdf using moment matching, Bottom-Left: Radiation pdf using CVaR, Bottom-Right: Radiation pdf using SigVaR.
Figure 5: Top: Radiation cdf when flow follows a log-normal distribution, Bottom: Radiation cdf when flow follows an exponential distribution (moment matching results).

3.3.3 Cumulative Density Functions

We obtain the empirical cdf of the radiation at the solution of the CVaR and SigVaR approaches to compare the conservativeness of the different methods in enforcing the chance constraint. This verification is also used for the MM approach in order to confirm that the cdf follows the pre-defined
Figure 6: Top: Radiation cdf when flow follows a log-normal distribution, Bottom: Radiation cdf when flow follows an exponential distribution (moment matching, CVaR, and SigVaR results).

form and that any error in the pre-definition of the distribution does not incur an error in satisfying the chance constraint. Figure 5 compares the log-normal cdf against the empirical cdf (obtained with the MM approach). Here, the vertical dotted red line highlights the desired threshold value and the horizontal line the probability achieved at such threshold. We can see that the log-normal cdf and
the empirical cdf nearly overlap and that the probability $1 - \alpha$ is exactly 95% at the desired threshold value. This indicates that the radiation cdf is indeed log-normal when the flow is log-normal (top panel) and exponential when the flow is exponential (bottom panel). We can also see that the error in the cdf is slightly more marked for the exponential case, indicating that slight deviations from exponential are being observed due to the finite number of scenarios used but this does not seem to affect the enforcement of the CC in a perceptible manner.

Figure 6 compares the optimal radiation cdfs obtained with CVaR, SigVaR, and MM. We can see again that CVaR is very conservative, achieving probability levels for the chance constraint of 98-99% (when 95% is only required). Also, the cdfs for SigVaR and MM overlap for the log-normal case, reinforcing that SigVaR provides good quality approximations and that the distribution of the radiation is indeed log-normal and thus MM is a sound approach to solve the problem. For the exponential distribution we see that SigVar overestimates the probability (around 96%). This is because of numerical difficulties encountered in solving the SigVaR problem for large values of $\mu, \tau$. Nevertheless, SigVaR is significantly less conservative than the CVaR and AS approaches.

4 Conclusions and Future Work

We demonstrated how to use moment matching techniques to reformulate chance constraints when the shape of the underlying density function is known. We also demonstrated the use of conservative approximations, which can be applied to more general settings in which the shape of the density function is not known. A flare design study shows that a sigmoidal approximation overcomes extreme conservativeness of the popular conditional value-at-risk and scenario approaches while the moment matching approach (when applicable) is computationally more attractive. The proposed approaches enable the solution of large-scale NLPs with chance constraints. As part of future work, we are interested in developing solution strategies for joint chance constraints. Moreover, it is necessary to explore alternative conservative approximations that have more stable computational behavior for limiting parameter values.

Acknowledgements

VM Zavala was partially supported by the U.S. Department of Energy grant DE-SC0014114.

References


