

# Optimization under rare chance constraints\*

Shanyin Tong<sup>†</sup>, Anirudh Subramanyam<sup>‡</sup>, and Vishwas Rao<sup>‡</sup>

**Abstract.** Chance constraints provide a principled framework to mitigate the risk of high-impact extreme events by modifying the controllable properties of a system. The low probability and rare occurrence of such events, however, impose severe sampling and computational requirements on classical solution methods that render them impractical. This work proposes a novel sampling-free method for solving rare chance constrained optimization problems affected by uncertainties that follow general Gaussian mixture distributions. By integrating modern developments in large deviation theory with tools from convex analysis and bilevel optimization, we propose tractable formulations that can be solved by off-the-shelf solvers. Our formulations enjoy several advantages compared to classical methods: their size and complexity is independent of event rarity, they do not require linearity or convexity assumptions on system constraints, and under easily verifiable conditions, serve as safe conservative approximations or asymptotically exact reformulations of the true problem. Computational experiments on linear, nonlinear and PDE-constrained problems from applications in portfolio management, structural engineering and fluid dynamics illustrate the broad applicability of our method and its advantages over classical sampling-based approaches in terms of both accuracy and efficiency.

**Key words.** chance constraints, extreme events, large deviation theory, bilevel optimization

**AMS subject classifications.** 90C15, 60F10, 65K05

**1. Introduction.** Rare high-impact events are phenomena that occur with low probability, but can have widespread—and sometimes even catastrophic—impacts on the economy, society and environment. Examples of such phenomena include cascading blackouts in the power grid, stress fractures and material failures in bridges and other civil structures, extreme oceanic turbulence leading to natural disasters like hurricanes and tsunamis, financial stock market crashes, and large-scale pandemics like the ongoing COVID-19 outbreak. It is of paramount importance to *quantify the risk* of such rare high-impact events in a given system, by estimating their probability of occurrence, and to *actively reduce this risk* by tuning the controllable properties of the system.

Chance constraints provide an attractive modeling framework to mitigate the risk of such rare high-impact phenomena, since it is often difficult—or even impossible—to estimate the costs associated with the underlying events. This motivates our study of chance constrained optimization problems of the following form.

$$(1.1) \quad \begin{aligned} & \underset{u \in \mathcal{U}}{\text{minimize}} && J(u) \\ & \text{subject to} && \mathbb{P}(F(u, \xi) \geq z) \leq \alpha, \quad \text{where } \alpha \ll 1. \end{aligned}$$

---

\*Submitted November 2020.

**Funding:** This material was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347. We acknowledge partial support by the US National Science Foundation (NSF) through grants DMS #1723211 to Shanyin Tong.

<sup>†</sup>Courant Institute of Mathematical Sciences, New York University, New York, NY ([shanyin.tong@nyu.edu](mailto:shanyin.tong@nyu.edu)).

<sup>‡</sup>Argonne National Laboratory, Lemont, IL ([asubramanyam@anl.gov](mailto:asubramanyam@anl.gov), [vhebbur@anl.gov](mailto:vhebbur@anl.gov)).

Here,  $u$  represents the set of control or decision variables with domain  $\mathcal{U}$ ,  $J : \mathcal{U} \rightarrow \mathbb{R}$  represents the cost associated with decisions  $u$ , and  $\xi$  denotes uncertain parameters that can influence the decisions via the function  $F$ . Thus,  $\mathcal{U}$  represents deterministic constraints on our decisions that are unaffected by uncertainties, and we consider  $\mathcal{U} \subseteq \mathbb{R}^m$  or  $\mathcal{U} \subseteq L^2(\mathcal{D})$  for some  $\mathcal{D} \subseteq \mathbb{R}^m$ . We shall assume that the uncertain parameters  $\xi$  are modeled as random vectors with probability distribution  $\mathcal{P}$  supported on  $\Xi \subseteq \mathbb{R}^n$ , and use  $\mathbb{P}(A)$  to denote the probability of an event  $A$  under the measure induced by  $\mathcal{P}$ . Therefore,  $F : \mathcal{U} \times \Xi \rightarrow \mathbb{R}$  represents the uncertainty-affected system metric (also variously referred to as the *parameter-to-observation map* or *limit-state function*), and  $z$  is a specified upper bound for the metric  $F$  such that, for any fixed decision vector  $u$  and fixed realization  $\xi$  of the random vector,  $F(u, \xi) \geq z$  leads to an “unsafe condition” or “bad event” that we seek to mitigate.

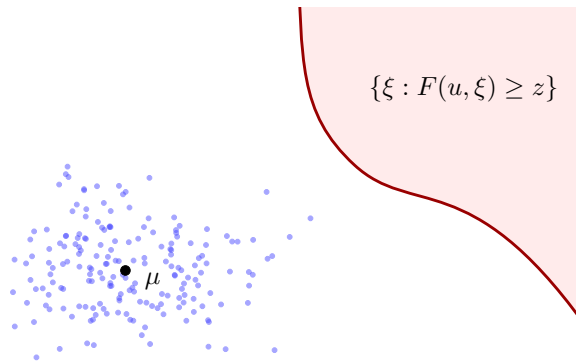
Problem (1.1) thus seeks to determine decisions  $u$  such that the probability of the system being unsafe,  $\mathbb{P}(F(u, \xi) \geq z)$ , remains below a certain specified risk threshold  $\alpha \in (0, 1)$ . We are particularly interested in the **rare event regime** where, for reasonable values of  $u \in \mathcal{U}$ , the latter quantities—the probability of being unsafe and the risk threshold  $\alpha$ —are very close to 0. We defer further assumptions and concrete examples to the subsequent sections.

One of the main challenges in solving (1.1) is to tackle the chance constraint for extreme values of the risk threshold  $\alpha \ll 1$  or the upper bound  $z \rightarrow \infty$ , since the resulting probability  $\mathbb{P}(F(u, \xi) \geq z)$  can be very small in these cases. For general nonlinear  $F$ , classical methods approximate this probability by generating samples from the original distribution of (or some other ‘trial’ distribution closely related to)  $\xi$ . However, the rare nature of the event coupled with the potential nonlinearity and complexity of  $F$  render sampling-based methods generally intractable and potentially inaccurate. As Figure 1 illustrates, such samples concentrate around regions of high likelihood, and relatively few samples cover the unsafe region. Indeed, assuming that  $F$  satisfies certain convexity properties, the number of samples required to obtain a feasible solution to (1.1) is  $\mathcal{O}(\alpha^{-1})$  (e.g., see [10]), which becomes impractical when  $\alpha$  is  $10^{-6}$  or  $10^{-10}$  (say). When taking into account problem-dependent constants as well, the actual required sample size (even for modest  $\alpha$  and Gaussian  $\xi$ ) can become arbitrarily large [24, Example 1]. This is made doubly worse when these samples are included in the optimization problem, since  $F$  must be evaluated at each generated sample and this can lead to computational difficulties if  $F$  itself is complex. This motivates the development of an alternate *sampling-free* method to solve (1.1), whose computational complexity does not scale with extreme values of  $z$  or  $\alpha$ .

**1.1. Overview of methodology and contributions.** We propose a sampling-free method to solve (1.1) by exploiting—and extending—recently developed results based on large deviation theory (LDT). Under certain assumptions on the distribution of  $\xi$  and the function  $F$ , it was shown in [16, 47] that, for fixed  $u$  and  $z$ , the rare event probability  $\mathbb{P}(F(u, \xi) \geq z)$  can be estimated using only local derivative information of  $F$  evaluated at the *LDT minimizer*:

$$(1.2) \quad \xi^* \in \operatorname{argmin}_{\xi \in \Xi} \{I(\xi) : F(u, \xi) \geq z\},$$

where  $I$  is the so-called *rate function* that depends only on the distribution of  $\xi$ . Although the method is more generally applicable, we focus our attention on the case where  $\xi$  follows a general Gaussian mixture distribution. Depending on the extent of local derivative information



**Figure 1.** Drawbacks of sampling-based methods for estimating rare events. The blue dots represent samples from the original distribution  $\mathcal{P}$  with mean  $\mu$ , whereas the shaded region represents the rare event set  $\{\xi : F(u, \xi) \geq z\}$  for some fixed decision vector  $u$ . Observe how very few samples fall inside the rare event set.

of  $F$  that one can exploit or access, we then propose first- and second-order approximations of the rare event probability by quantifying the measure of the corresponding set bounded by the first- and second-order Taylor expansions of  $F$  at  $\xi^*$ , and provide easily verifiable conditions under which the proposed estimates are exact or constitute upper bounds on the true probability. However, using these estimates in place of the true probability in (1.1) amounts to a bilevel optimization problem, since the LDT minimizer (1.2) is a function of the decision variables  $u$ ; to tackle this, we replace (1.2) with its optimality conditions, and elucidate when the latter are sufficient to characterize the minimizer. To demonstrate the effectiveness of our method, and its potential benefits over classical sampling-based methods, we consider a variety of applications from different domains. Specifically, via problems in financial portfolio management, structural column design and optimal control of a system of partial differential equations (PDE), we illustrate the accuracy of our probability estimates, feasibility of the obtained optimal solutions, improvement in computation time, and ability to handle complex nonlinear models.

The main contributions of our work may be summarized as follows.

1. We provide explicit formulas for the first- and second-order LDT approximations of rare event probabilities when the constraint function  $F$  is a continuously differentiable function of random parameters that follow a general Gaussian mixture distribution. This generalizes existing results [47] for Gaussians, and can be used to tackle general continuous distributions by approximating them with a Gaussian mixture. We also provide an alternate characterization of the second-order LDT approximation for Gaussians that is more numerically efficient for our purposes.
2. We apply the LDT approximations to optimization problems affected by rare chance constraints, and provide reformulations that can be solved by off-the-shelf solvers. Notably, our formulations are sampling-free and scalable: their size and complexity does not scale with the rarity of the chance constraints. We also contribute easily verifiable conditions under which their optimal values are exact or provide conservative upper bounds to the true optimal values. Finally, we do not limit ourselves to linear

or convex  $F$  and only require that it is twice differentiable, thus enabling fast local solutions to nonlinear and nonconvex problems.

3. We conduct computational experiments on linear, nonlinear and PDE-constrained optimization problems from several application areas to demonstrate the viability of our method and compare its performance with traditional sampling-based methods.

This paper is structured as follows. After introducing some notation and basic assumptions in [subsection 1.2](#), we provide a literature review of solution methods for chance constrained problems in [Section 2](#). In [Section 3](#), we provide an overview of our techniques, the large deviation results we use, as well as easily verifiable sufficient conditions under which we can establish optimality and feasibility of our formulations; notably, this section does not make specific assumptions on the nature of the distributions. In [Section 4](#) and [Section 5](#), we then specialize our results to Gaussian and Gaussian mixture distributions, respectively, and provide closed-form expressions of the probability estimates and optimization formulations. In [Section 6](#), we illustrate the computational performance of our method on several applications, and finally, in [Section 7](#), we provide concluding remarks and directions for future work.

**1.2. Notation and setup.** We use  $\mathbb{R}_+$  to denote the set of non-negative reals, and  $I_n$  to denote the  $n \times n$  identity matrix. We define  $[x]_+ := \max\{x, 0\}$ , and  $\mathbb{1}_X$  to be the indicator function of set  $X$ :  $\mathbb{1}_X(x)$  is equal to 1 if  $x \in X$  and 0 otherwise. Given a symmetric matrix  $Q \in \mathbb{R}^{n \times n}$ , we use  $Q \succ 0$  ( $\succeq 0$ ) to denote that it is positive definite (positive semidefinite), and in such cases, we denote the corresponding weighted inner product between  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^n$  by  $\langle x, y \rangle_Q := \langle x, Qy \rangle$ , and the induced norm by  $\|x\|_Q$ . We denote the standard Euclidean inner product interchangeably using both  $\langle x, y \rangle$  and  $x^\top y$ .

With a slight abuse of notation, we use the same symbol  $\xi$  to denote the random parameter vector as well as its realizations; the meaning should be clear from the context. We use  $\mathcal{P}$  to denote the distribution of  $\xi$ ,  $\Xi$  to denote its support, and  $\mathbb{P}(A)$  to denote the probability of an event  $A$  under the measure  $\mathcal{P}$ . Given a measurable function  $f(\xi)$ , we let  $\mathbb{E}(f(\xi))$  denote the expectation of  $f(\xi)$ . Finally, we use  $S : \mathbb{R}^n \rightarrow \mathbb{R}$  to denote the cumulant generating function:

$$(1.3) \quad S(\eta) := \log \mathbb{E} e^{\langle \eta, \xi \rangle} = \log \left\{ \int_{\Xi} e^{\langle \eta, \xi \rangle} d\mathcal{P}(\xi) \right\},$$

and define the rate function  $I : \Xi \rightarrow \mathbb{R}$  as the convex (Fenchel) conjugate of  $S$ :

$$(1.4) \quad I(\xi) := \max_{\eta \in \mathbb{R}^n} \{ \langle \eta, \xi \rangle - S(\eta) \}.$$

Throughout the paper, we assume that  $\mu := \mathbb{E}(\xi)$  exists and that  $S$  is continuously differentiable on  $\mathbb{R}^n$ . These are easily verified to be true whenever  $\mathcal{P}$  is a Gaussian (or a mixture of Gaussians). In particular, when working with multivariate Gaussian parameters in  $\mathbb{R}^n$ , we use  $\xi \sim \mathcal{N}(\mu, \Sigma)$  to denote that  $\xi$  is a multivariate Gaussian with mean  $\mu \in \mathbb{R}^n$  and covariance matrix  $\Sigma \succ 0$ , and  $\xi \sim \sum_{i=1}^M w_i \mathcal{N}(\mu_i, \Sigma_i)$  to indicate that  $\xi$  follows a Gaussian mixture distribution with  $M$  components, where component  $i \in \{1, \dots, M\}$  is a Gaussian with mean  $\mu_i$ , covariance  $\Sigma_i \succ 0$  and weight  $w_i > 0$  such that  $\sum_{i=1}^M w_i = 1$ . Finally, we use  $\Phi$  to denote the cumulative distribution function (CDF) of the standard normal distribution  $\mathcal{N}(0, 1)$ .

Regarding the structure of  $F$ , we shall assume that  $F(\cdot, \xi)$  is twice continuously differentiable on  $\mathcal{U}$  for every  $\xi \in \Xi$  and  $F(u, \cdot)$  is twice continuously differentiable on closure of  $\Xi$  for every  $u \in \mathcal{U}$ . In addition, we assume that there exist  $\emptyset \neq \mathcal{U}_0 \subset \mathcal{U}$ ,  $z_0 < \infty$  and  $K_0 > 0$  such that, for all  $u \in \mathcal{U}_0$  and all  $z \in [z_0, \infty)$ , we have  $F(u, \mu) < z_0$ ,  $\Xi(u, z) := \{\xi \in \Xi : F(u, \xi) = z\} \neq \emptyset$  and  $\|\nabla_{\xi} F(u, \xi)\| \geq K_0 > 0$  for all  $\xi \in \Xi(u, z)$ .

We note that the existence of  $u \in \mathcal{U}$  and  $x$  in a neighborhood of  $z$  for which  $\Xi(u, x) \neq \emptyset$  is not restrictive; indeed, if one has  $F(u, \xi) < z$  for all  $\xi \in \Xi$  and  $u \in \mathcal{U}$ , then (1.1) is trivially feasible for all  $u \in \mathcal{U}$  (*i.e.*, the chance constraint can be ignored), and similarly it is trivially infeasible if  $F(u, \xi) > z$  for all  $\xi \in \Xi$  and all  $u \in \mathcal{U}$ . The assumption  $\|\nabla_{\xi} F(u, \xi)\| > 0$  ensures that the hypersurface  $\Xi(u, x) \subset \Xi$  does not degenerate and remains smooth and simply connected, and that the minimization problem defined in (3.6) is always feasible for any  $u \in \mathcal{U}_0$ . We defer further assumptions to the sections where they are used.

Finally, note that we only consider the case of a single *individual chance constraint*. The extension to multiple individual chance constraints is straightforward and we do not present it for ease of exposition. On the other hand, the extension to *joint chance constraints* where  $F : \mathcal{U} \times \Xi \rightarrow \mathbb{R}^p$  and  $z \in \mathbb{R}^p$ , with  $p > 1$  is not straightforward. In such cases, one can either resort to a Bonferroni approximation (*i.e.*, union bound) to convert the joint chance constraint into  $p$  individual chance constraints, or apply our method to a smooth approximation of the function  $\phi(x, u) = \min\{F_1(u, \xi) - z_1, \dots, F_p(u, \xi) - z_p\}$  (*e.g.*, see [40]).

**2. Literature review.** Chance constrained problems have been intensely studied for over five decades; see [42, 46] for background, references and a detailed treatment. Unfortunately, they are generically NP-hard since the set of feasible decisions is usually nonconvex, and admit exact or tractable reformulations only in very simple or restrictive settings, such as when  $F$  is bi-affine and  $\xi$  follows a normal distribution [42]. Therefore, the majority of existing research takes one of two approaches: exploit or assume some special structure in  $F$  or the distribution of  $\xi$ , or resort to some form of generic sampling to approximate the chance constraint. Examples of the former include the study of linear or convex  $F$  [31, 37], finitely supported distributions of  $\xi$  [34], and making use of specialized numerical integration techniques to compute Gaussian-like probabilities and their gradients [50]. Sampling-based methods are reviewed in subsection 2.1. In subsection 2.2, we discuss connections of our proposed method to so-called reliability-based methods that are prevalent in engineering design.

**2.1. Sampling-based methods.** Importance sampling is by far the most commonly used sampling method to estimate rare event probabilities [29]. These methods belong to the class of variance reduction techniques, and sample from a (typically problem-specific) biasing distribution that is constructed so that probability mass is concentrated in the region of interest (*e.g.*, see [3, 8, 21]). Existing methods for rare chance constrained optimization are also based on variants of importance sampling [4, 9, 44]. Given  $N$  independent samples or realizations,  $\xi^1, \dots, \xi^N$ , that have been obtained via importance sampling (or some other method), the probability,  $\mathbb{P}(F(u, \xi) \geq z)$  in (1.1), can then be approximated via one of two broad classes of approaches which we discuss in the following subsections. Examples of sampling-based solution methods that do not fall into these classes can be found in [13, 40].

Philosophically, however, we note that none of these approaches can be expected to work well in general for rare chance constraints, since the required number of samples can be

$O(\alpha^{-1})$  in the worst case. Indeed, even if we manage to reduce the required number of samples (by importance sampling for example), it still entails modeling additional (at least  $\mathcal{O}(N)$ ) variables and constraints in the optimization problem, which can make its solution computationally difficult.

**2.1.1. Indicator formulations.** The methods in this class are all based on reformulating the probability in (1.1) as the expectation of an indicator function and then replacing the expectation with a *sample average approximation* (SAA) [33, 39]:

$$\mathbb{P}(F(u, \xi) \geq z) = \mathbb{E} [\mathbf{1}_{[z, \infty)}(F(u, \xi))] \approx \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{[z, \infty)}(F(u, \xi^i)).$$

The discontinuous indicator function  $\mathbf{1}_{[z, \infty)}$  is then either exactly reformulated using  $\mathcal{O}(N)$  binary variables (e.g., see [34]) resulting in a mixed-integer nonlinear optimization problem, or approximated using continuous functions to avoid introducing binary variables. Examples of the latter include difference-of-convex approximations [26], sigmoidal or other smooth approximations [1, 11, 23] as well as convex outer approximations [37]. An example of a smooth approximation (that we also compare in our numerical experiments) is the sigmoidal approximation proposed in [11]. It is based on the observation that, for any  $\nu, \tau > 0$ , one can construct an outer-approximation of the indicator function that is asymptotically optimal as  $\nu, \tau \rightarrow \infty$ .

$$\mathbf{1}_{[z, \infty)}(x) \leq \left[ \frac{2(\nu + 1)}{\nu + \exp(-\tau(x - z))} - 1 \right]_+, \quad \lim_{\nu, \tau \rightarrow \infty} \left[ \frac{2(\nu + 1)}{\nu + \exp(-\tau(x - z))} - 1 \right]_+ = \mathbf{1}_{[z, \infty)}(x).$$

This allows us to obtain the following approximation to (1.1):

$$(2.1) \quad \begin{aligned} & \underset{u \in \mathcal{U}, p \in \mathbb{R}_+^N}{\text{minimize}} && J(u) \\ & \text{subject to} && p_i \geq \frac{2(\nu + 1)}{\nu + \exp(-\tau(F(u, \xi^i) - z))} - 1, \quad i \in \{1, \dots, N\}, \quad (\text{SAA}) \\ & && \frac{1}{N} \sum_{i=1}^N p_i \leq \alpha. \end{aligned}$$

Another class of general outer approximations was proposed in [37] by using any non-decreasing convex function  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  satisfying  $\psi(0) = 1$  and  $\psi(x) > 1$  for all  $x > 0$ . In particular, by noting that  $\mathbf{1}_{[z, \infty)}(x) \leq \psi(t^{-1}(x - z))$  is satisfied for any such function and scalar  $t > 0$ , it can be shown that satisfaction of the inequality,

$$\inf_{t > 0} \{t \mathbb{E}(\psi(t^{-1}(x - z))) - t\alpha\} \leq 0,$$

implies satisfaction of the probability constraint in (1.1). Under the choice  $\psi(x) = [1 + x]_+$ , the above inequality is equivalent to  $\text{CVaR}_{1-\alpha}(F(u, \xi) - z) \leq 0$ , where, for any measurable function  $f(\xi)$  and any  $\alpha \in (0, 1)$ ,  $\text{CVaR}_{1-\alpha}(f(\xi)) := \inf_{t \in \mathbb{R}} \{t + \alpha^{-1} \mathbb{E}[f(\xi) - t]_+\}$  is the conditional-value-at-risk (CVaR) of the random variable  $f(\xi)$ . By approximating the expectation in the expression for CVaR with a sample average, we obtain the following conservative

approximation to (1.1):

$$(2.2) \quad \begin{aligned} & \underset{u \in \mathcal{U}, p \in \mathbb{R}_+^N, t \in \mathbb{R}}{\text{minimize}} && J(u) \\ & \text{subject to} && p_i \geq F(u, \xi^i) - z + t, \quad i \in \{1, \dots, N\}, \quad (\text{CVaR}) \\ & && \frac{1}{N} \sum_{k=1}^N p_k \leq \alpha t. \end{aligned}$$

The CVaR approximation has the advantage of preserving any convexity structure in  $F$  (with respect to  $u$ ), and so we use it as a benchmark in our numerical experiments.

**2.1.2. Scenario approaches.** An alternative approach is to enforce the constraints deterministically for each of the  $N$  realizations of  $\xi$ :  $F(u, \xi^i) \leq z$ , for all  $i \in \{1, \dots, N\}$ . As in the case of the CVaR approximation, this preserves any convexity structure in  $F$  and the hope is that for sufficiently large  $N$ , optimal solutions of the scenario approximation will be feasible in (1.1) with high probability. Under the assumption that  $F(\cdot, \xi)$  is convex for all  $\xi \in \Xi$ , the work [10] shows that choosing  $N = \mathcal{O}(m\alpha^{-1} \log(\alpha^{-1}))$  suffices, where  $m$  is the number of decision variables  $u$ . This size can be prohibitively large for small values of  $\alpha$ . Therefore, [5, 36] consider extensions of this approach to reduce the required sample size by making further assumptions on  $F$  and the distribution of  $\xi$ .

**2.2. Reliability-based methods.** In engineering, the structural reliability of buildings or bridges is assessed by solving a probability estimation problem similar to the one we consider, and reliability is enforced as part of the structural design by solving a so-called inverse reliability problem [18] or reliability-based design optimization problem [49] that is similar in spirit to the chance constrained problem (1.1). These methods approximate the distribution of  $\xi$  using a (standard) multivariate normal distribution, and then compute the so-called most probable point (*i.e.*, the point with maximum probability density) over the failure region  $\{\xi \in \Xi : F(u, \xi) \geq z\}$  (*e.g.*, see [20]). The distance of the most probable point from the origin is referred to as the reliability index, which is not only used for estimating probabilities, but also explicitly constrained as part of the optimal design problem. In particular, the probability estimates are obtained by employing first- or second-order Taylor expansions of the limit-state function around the most probable point and then integrating the Gaussian densities over the set bounded by the Taylor approximations, resulting in the so-called first- and second-order reliability methods (FORM and SORM), respectively [20, 45]. In contrast to a chance constraint on the limit-state function  $F$ , the reliability-based design problem explicitly constrains the reliability index and is solved using specialized procedures; *e.g.*, see [32, 49].

These methods share similarities with our LDT approach, especially for Gaussian distributions. In particular, both methods are sampling-free, and connect the probability estimations to an optimization problem. A crucial difference is that whereas reliability-based methods focus on finding the point with maximal likelihood, our LDT approach uses the point which characterizes the asymptotic behavior of the rare event probability, and thus includes information beyond just local probability densities. Another caveat is that reliability-based methods typically require transformation of the distribution to a standard normal, which can be difficult for a general high-dimensional distribution, whereas our methods are generally applicable

for distributions with well-defined rate functions, including Gaussian mixtures that can, in limit, approximate arbitrary continuous distributions. Finally, an important conceptual difference lies in the formulation of the optimization problem: whereas reliability-based methods use the reliability index as a surrogate for the probability to enforce constraints, we treat the more traditional chance constrained formulation (1.1), and use the LDT minimizer along with local derivative information directly to approximate the probability in the chance constraint.

**3. Large deviation theory for rare chance constraints.** Large deviation theory (LDT) [17, 48, 51] is concerned with the asymptotic behavior of tails of probability distributions, and specifically the rates of exponential decay of probability measures of extreme events. Classical LDT enables characterizing the asymptotic behavior of rare probabilities, such as the one in (1.1), with respect to some small (or large) parameter associated with the random variable  $\xi$ , while keeping  $z$  fixed. It states that the behavior of the probability with respect to that small (or large) parameter is dominated by a decaying exponential term that can be characterized by the rate function of the random variable. Recently, [16] adapted classical theory and sharp asymptotics based on the dominating point of large deviations [6, 7, 28, 38] to study the asymptotic behavior with respect to  $z \rightarrow \infty$ , while keeping the distribution of  $\xi$  fixed. The key idea is to find a dominating point in the rare event set by solving an appropriately defined optimization problem and estimating probabilities solely using this dominating point. In [47], it was shown that improved estimates can be obtained by exploiting Gaussianity as well as local approximations of  $F$  around this dominating point. The approach has been successfully applied to quantify rare probabilities of extreme oceanic phenomena [14, 15, 47].

In the following subsections, we present an overview of this approach and show how it can be used for rare chance constrained optimization. Specifically, in subsection 3.1, we first reduce the problem of estimating the measure of the set bounded by (the potentially nonlinear)  $F$  to estimating the measure of the set bounded by its first- and second-order Taylor expansions around a dominating point. In subsection 3.2, we provide a specific choice of the dominating point based on large deviation arguments; and finally, in Section subsection 3.3 we incorporate these probability estimates in the chance constrained problem.

**3.1. Taylor approximation of nonlinear limit state function.** For a fixed choice of  $u \in \mathcal{U}$  and  $z \in \mathbb{R}$ , a key challenge in computing the probability in (1.1),  $\mathbb{P}(F(u, \xi) \geq z)$  is the potential nonlinearity of  $F$  with respect to  $\xi$ . A simple way to tackle this is to construct first- and second-order Taylor approximations of  $F$  around some point  $\xi^* \in \Xi$ ,

$$(3.1a) \quad F_1(u, \xi; \xi^*) := F(u, \xi^*) + \langle \nabla_{\xi} F(u, \xi^*), \xi - \xi^* \rangle,$$

$$(3.1b) \quad F_2(u, \xi; \xi^*) := F(u, \xi^*) + \langle \nabla_{\xi} F(u, \xi^*), \xi - \xi^* \rangle + \frac{1}{2} \langle \xi - \xi^*, \nabla_{\xi}^2 F(u, \xi^*) (\xi - \xi^*) \rangle.$$

We can then compute first- and second-order probability estimates,  $P_1$  and  $P_2$  of  $\mathbb{P}(F(u, \xi) \geq z)$ , by computing the measure of the sets bounded by the corresponding Taylor approximations:

$$(3.2) \quad P_k(u, z, \xi^*) = \mathcal{P}(\{\xi \in \Xi : F_k(u, \xi; \xi^*) \geq z\}), \quad k \in \{1, 2\}.$$

It is important to note that this approach approximates the nonlinear (in  $\xi$ ) chance constrained problem (1.1) with a linear and quadratic (in  $\xi$ ) chance constrained problem, respectively. In



particular, it allows us to exploit special structure (if any) in the distribution of  $\xi$  to efficiently compute the first- and second-order approximations in (3.2). We provide examples of this computation for standard Gaussian and mixture distributions in the subsequent sections. First, we make the immediate observation that  $P_1$  provides a conservative approximation to the true probability whenever  $F$  is concave in the uncertainty  $\xi$ .

**Proposition 3.1 (First-order probability estimate for concave functions).** *Fix  $u \in \mathcal{U}$ . Suppose that  $\Xi$  is convex and the function  $F(u, \cdot)$  is concave on  $\Xi$ . Then,  $\mathbb{P}(F(u, \xi) \geq z) \leq P_1(u, z, \xi^*)$  for any  $z \in \mathbb{R}$  and  $\xi^* \in \Xi$ .*

*Proof.* Since  $F(u, \cdot)$  is continuously differentiable (refer to subsection 1.2), it is concave on  $\Xi$  if and only if

$$F(u, \xi) \leq F(u, \xi^*) + \langle \nabla_{\xi} F(u, \xi^*), \xi - \xi^* \rangle = F_1(u, \xi; \xi^*), \quad \forall \xi \in \Xi.$$

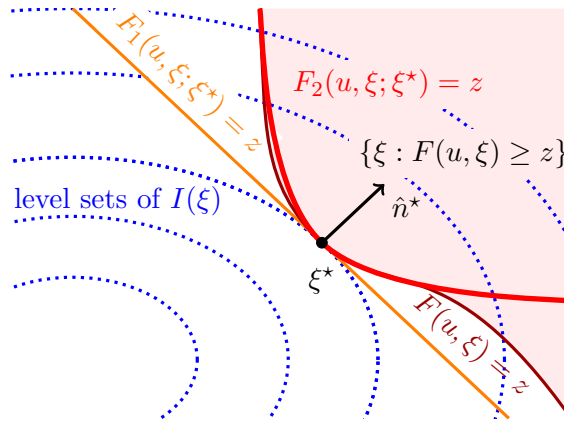
Therefore, we have the relationship

$$\{\xi \in \Xi : F(u, \xi) \geq z\} \subseteq \{\xi \in \Xi : F_1(u, \xi; \xi^*) \geq z\},$$

i.e., the rare event set is included in the half-space bounded by  $F_1(u, \xi; \xi^*) = z$ . Therefore,

$$\mathbb{P}(F(u, \xi) \geq z) = \mathcal{P}(\{\xi \in \Xi : F(u, \xi) \geq z\}) \leq \mathcal{P}(\{\xi \in \Xi : F_1(u, \xi; \xi^*) \geq z\}) = P_1(u, z, \xi^*),$$

and the claim follows. ■



**Figure 2.** *Illustration of first and second-order approximations of probability using Taylor expansions at the LDT minimizer  $\xi^*$ , which is the solution to (3.3). The blue dotted lines are level sets of the rate function  $I(\cdot)$ , and the red shaded region is the extreme event set  $\{\xi : F(u, \xi) \geq z\}$ . The normal  $\hat{n}^*$  aligns with the gradients  $\nabla_{\xi} F(u, \xi^*)$  and  $\nabla I(\xi^*)$ . The figure is adapted from [47].*

**3.2. Large deviation theory.** Although the Taylor approximations of  $F$  provide a tractable means to estimate the true probability, their accuracy depends critically on the choice of  $\xi^*$ . In the following, we invoke large deviation results from [16] to justify a choice of  $\xi^*$  that

is asymptotically optimal as  $z \rightarrow \infty$ . The formal statement of the result needs additional assumptions on the distribution of  $\xi$  and the structure of  $F$ . The procedure for choosing  $\xi^*$  based on this result, and the first- and second-order approximations of  $F$  at this point are shown in [Figure 2](#).

**Theorem 3.2** (Theorem 2.1 in [16]). *Fix  $u \in \mathcal{U}_0$ . Suppose the following assumptions hold:*

(A1) *The optimization problem  $\min_{\xi \in \Xi} \{I(\xi) : F(u, \xi) \geq z\}$  has a unique global minimizer for all  $z \in [z_0, \infty)$ . Therefore, we can define:*

$$(3.3) \quad \xi^* : \mathcal{U}_0 \times [z_0, \infty) \rightarrow \Xi, \quad \xi^*(u, z) := \operatorname{argmin}_{\xi \in \Xi} \{I(\xi) : F(u, \xi) \geq z\}.$$

(A2) *The function  $\xi^*$  in (3.3) is continuously differentiable, and there exists  $K'_0 > 0$  such that the rate function  $I(\xi^*(u, \cdot))$  in (1.4) is strictly increasing with*

$$(3.4) \quad I(\xi^*(u, z)) \rightarrow \infty \quad \text{and} \quad \|\nabla I(\xi^*(u, z))\| \geq K'_0 > 0 \quad \text{as} \quad z \rightarrow \infty.$$

(A3) *The following relation holds for all  $z \in [z_0, \infty)$ :*

$$\{\xi \in \Xi : F(u, \xi) \geq z\} \subseteq \{\xi \in \Xi : \langle \nabla_\xi F(u, \xi^*(u, z)), \xi - \xi^*(u, z) \rangle \geq 0\},$$

(A4) *There exists  $K_1 > 0$  such that, for all  $u \in \mathcal{U}_0$  we have*

$$\lim_{z \rightarrow \infty} \frac{\log \left( \int_{\Lambda(u, z, K_1)} \exp(I(\xi^*(u, z)) + \langle \nabla I(\xi^*(u, z)), \xi - \xi^*(u, z) \rangle) d\mathcal{P}(\xi) \right)}{I(\xi^*(u, z))} = 0$$

$$\text{where } \Lambda(u, z, s) := \left\{ \xi \in \Xi : F(u, \xi) \geq z, \frac{\langle \nabla_\xi F(u, \xi^*(u, z)), \xi - \xi^*(u, z) \rangle}{\|\nabla_\xi F(u, \xi^*(u, z))\|} < s \right\}.$$

Then, the following result holds.

$$(3.5) \quad \lim_{x \rightarrow \infty} \frac{\log \mathbb{P}(F(u, \xi) \geq z)}{I(\xi^*(u, z))} = -1.$$

[Theorem 3.2](#) implies that under its stated assumptions, the rare event probability in (1.1) satisfies the asymptotic equality

$$(3.6) \quad \mathbb{P}(F(u, \xi) \geq z) \asymp \exp(-I(\xi^*(u, z))).$$

where  $\asymp$  indicates that the ratio of logarithms of both sides tends to 1 as  $z \rightarrow \infty$ . For large but finite values of  $z$ , the right-hand of (3.6) can therefore be expected to serve as a good approximation of the true probability. This justifies the LDT minimizer  $\xi^*(u, z)$  defined in (3.3) as the ideal point around which one should Taylor approximate the nonlinear  $F$ .

We now review the assumptions leading to the result of [Theorem 3.2](#) in detail, with the goal of simplifying them. Although not explicitly stated in [16], assumption (A1) is necessary to ensure that the LDT minimizer in (3.3) is well defined. Assumption (A2) formally states that the event  $\{F(u, \xi) \geq z\}$  becomes rare as  $z \rightarrow \infty$ , whereas assumption (A3) requires that  $\xi^*$  is a so-called dominating point [38]: that is, the rare event set  $\{\xi \in \Xi : F(u, \xi) \geq z\}$  is

completely enclosed in the halfspace that is tangent to the hypersurface  $\Xi(u, z)$  at  $\xi^*$ . Finally, Assumption (A4) is an equivalent rephrasing of Assumption 5 in [16]. Specifically, it ensures that the true rare event probability does not decay much faster than  $\exp(-I(\xi^*(u, z)))$  as  $z \rightarrow \infty$ ; *i.e.*, that the right-hand side of (3.5) is a lower bound on its left-hand side expression.

The assumptions leading to the result of Theorem 3.2 may not always be straightforward to verify. With a goal of obtaining easily verifiable conditions, the following theorem provides alternative sufficient conditions that lead to the same large deviation result as Theorem 3.2.

**Theorem 3.3 (Sufficient conditions for Theorem 3.2).** *Fix  $u \in \mathcal{U}_0$ . Suppose that the following condition is satisfied:*

(C1)  $\Xi = \mathbb{R}^n$  and the function  $F(u, \cdot)$  is concave on  $\Xi$ .

Then, the following inequality is valid:

$$(3.7) \quad \lim_{x \rightarrow \infty} \frac{\log \mathbb{P}(F(u, \xi) \geq z)}{I(\xi^*(u, z))} \leq -1.$$

Additionally, suppose that the following conditions also hold:

(C2)  $\nabla_{\xi} F(u, \cdot)$  is Lipschitz continuous on  $\Xi$ .

(C3) For any finite  $r > 0$ , the measure  $\mathcal{P}$  satisfies

$$\log \left( \int_{\mathcal{B}(r, \hat{\xi})} \exp \left( I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle \right) d\mathcal{P}(\xi) \right) = o(I(\hat{\xi})) \quad \text{as } \|\hat{\xi}\| \rightarrow \infty,$$

where  $\mathcal{B}(r, \hat{\xi}) \subset \Xi$  denotes a closed ball of radius  $r$  that contains  $\hat{\xi} \in \Xi$ .

Then, Equation (3.5) holds.

*Proof.* To show the first part, we will show that condition (C1) implies assumptions (A1), (A2) and (A3) stated in Theorem 3.2. The proof for the validity of (3.7) then follows from the proof of Theorem 2.1 in [16]. By a similar argument, we will show that the conditions taken together also imply assumption (A4). The proof that the left-hand side of (3.7) is  $\geq$  to its right-hand side then follows from similar arguments as Theorem 2.1 in [16].

*Verification of (A1).* Since the cumulant generating function  $S$  is finite on  $\mathbb{R}^n$  (see subsection 1.2), [41, Theorem 2] implies that it is strictly convex and twice continuously differentiable. Therefore, the rate function  $I$ , which by definition (1.4) is the convex conjugate of  $S$ , satisfies the following properties: (i)  $I$  is strictly convex and twice continuously differentiable on  $\mathbb{R}^n$  [25, Corollary 4.2.10], and (ii)  $I$  is coercive [43, Theorem 11.8], *i.e.*,  $I(\xi) \rightarrow +\infty$  as  $\|\xi\| \rightarrow \infty$ . Together with the concavity of  $F$  from condition (C1), these properties imply that  $\min_{\xi \in \Xi} \{I(\xi) : F(u, \xi) \geq z\}$  is a convex optimization problem (which is always feasible for  $z \in [z_0, \infty)$ , see subsection 1.2) with a finite optimal value that is attained by a unique global minimizer  $\xi^*(u, z)$  defined in (3.3). Notably, this implies assumption (A1).

*Verification of (A2).* We show the implication of assumption (A2) in two parts.

- First, note that continuity of  $F(u, \cdot)$  implies that  $\sup_{\xi \in \Xi} \{F(u, \xi) : \|\xi\| \leq r_0\} < \infty$  for all finite  $r_0 > 0$ . This, in turn, implies that any point in the set  $\{\xi \in \Xi : F(u, \xi) \geq z\}$  must satisfy  $\|\xi\| \rightarrow \infty$  as  $z \rightarrow \infty$ . In particular, we have  $\|\xi^*(u, z)\| \rightarrow \infty$  and from the coercivity of  $I$  established in (ii) above, this implies that  $I(\xi^*(u, z)) \rightarrow \infty$ .

- Second, for any  $z \in [z_0, \infty)$ , there always exists a point  $\xi \in \Xi = \mathbb{R}^n$  satisfying  $F(u, \xi) > z$ , see [subsection 1.2](#). This implies that the convex problem  $\min_{\xi \in \Xi} \{I(\xi) : F(u, \xi) \geq z\}$  satisfies Slater's constraint qualification, and strong Lagrangian duality holds:

$$I(\xi^*(u, z)) = \max_{\lambda \in \mathbb{R}_+} \left\{ \lambda z + \min_{\xi \in \Xi} \{I(\xi) - \lambda F(u, \xi)\} \right\}$$

For sufficiently large values of  $z$ , the maximizer in the right-hand side of the above expression is always attained at some  $\lambda \geq \lambda_0 > 0$ ; indeed, if it is attained at  $\lambda = 0$ , then  $I(\xi^*(u, z)) = \min_{\xi \in \Xi} I(\xi) < +\infty$  which would contradict  $I(\xi^*(u, z)) \rightarrow \infty$ . Therefore, as a function of  $z$ , the right-hand side of the above expression, and hence  $I(\xi^*(u, z))$ , is convex and strictly increasing. Moreover,  $\|\nabla I(\xi^*(u, z))\| = \lambda \|\nabla_\xi F(u, \xi)\| \geq \lambda_0 K_0 > 0$ , (Note that, by the same reasoning, we also have  $F(u, \xi^*(u, z)) = z$  which will be used in the next paragraph).

*Verification of (A3).* This follows from the definition of concavity in condition (C1).

*Verification of (A4).* Let  $L > 0$  denote the Lipschitz constant in condition (C2). Consider now the closed ball  $B(r, \xi^*(u, z))$  of radius  $r = K_0 L^{-1} > 0$  centered at  $\xi^*(u, z) + r \hat{n}(u, z)$ , where  $K_0$  was defined in [subsection 1.2](#), and  $\hat{n}(u, z)$  is the unit vector along  $\nabla_\xi F(u, \xi^*(u, z))$ . We claim that  $B(r, \xi^*(u, z)) \subseteq \Lambda(u, z, 2r)$ , where  $\Lambda$  is defined in assumption (A4). If true, then observe that, the fact that  $\|\xi^*(u, z)\| \rightarrow \infty$  as  $z \rightarrow \infty$  (established above), along with condition (C3) evaluated at  $\hat{\xi} = \xi^*(u, z)$  implies (A4) is satisfied with  $K_1 = 2K_0 L^{-1} > 0$ . To show  $B(r, \xi^*(u, z)) \subseteq \Lambda(u, z, K_1)$ , we proceed as follows, dropping the dependence on  $u$  and  $z$  for simplicity of exposition.

- First, we show  $\xi \in B(r, \xi^*)$  implies that  $\langle \hat{n}, \xi - \xi^* \rangle \leq K_1$ .

$$\xi \in B(r, \xi^*) \iff \|\xi - \xi^* - r \hat{n}\| \leq r \iff \|\xi - \xi^*\| \leq 2r \implies \langle \hat{n}, \xi - \xi^* \rangle \leq 2r = K_1.$$

- Next, we show that  $\xi \in B(r, \xi^*)$  implies that  $F(\xi) \geq z$ . Lipschitz continuity of  $\nabla_\xi F(u, \cdot)$  along with its twice differentiability implies that  $\nabla_\xi^2 F(u, \xi) \succeq -L I_n$  holds for all  $\xi \in \Xi$ . Along with Taylor's theorem, this implies

$$\begin{aligned} F(u, \xi) &\geq F(u, \xi^*) + \langle \nabla_\xi F(u, \xi^*), \xi - \xi^* \rangle - \frac{L}{2} \|\xi - \xi^*\|^2 \\ &= z + \|\nabla_\xi F(u, \xi^*)\| \left[ \langle \hat{n}, \xi - \xi^* \rangle - \frac{L}{2 \|\nabla_\xi F(u, \xi^*)\|} \|\xi - \xi^*\|^2 \right] \\ &\geq z + \|\nabla_\xi F(u, \xi^*)\| \left[ \langle \hat{n}, \xi - \xi^* \rangle - \frac{1}{2r} \|\xi - \xi^*\|^2 \right] \\ &= z \quad (\text{since } \xi \in B(r, \xi^*)) \end{aligned}$$

where we used the fact that  $F(u, \xi^*) = z$ ,  $\|\nabla_\xi F(u, \xi^*)\| \geq K_0$  and the definition of  $r = K_0 L^{-1}$ . Thus, we have also verified assumption (A4). ■

Unlike [Theorem 3.2](#), it is crucial to note that the conditions in [Theorem 3.3](#) either depend only on the structure of  $F$  or on the measure  $\mathcal{P}$ , and hence, are relatively easier to verify in general. For example, [Theorem 5.1](#) in [Section 5](#) shows that when  $\mathcal{P}$  is a Gaussian or mixture of Gaussian distributions, then condition (C3) is automatically satisfied.

**3.3. Incorporating large deviation estimates in the chance constrained problem.** We can now incorporate in the chance constrained problem (1.1), the probability estimates obtained using the first- and second-order Taylor expansions of  $F$  that we introduced in subsection 3.1, around the dominating point  $\xi^*$  from subsection 3.2. Depending on whether we use the first- ( $k = 1$ ) or second-order ( $k = 2$ ) approximation, we obtain the following conceptual problem:

$$(3.8) \quad \begin{aligned} & \underset{u \in \mathcal{U}}{\text{minimize}} && J(u) \\ & \text{subject to} && P_k(u, z, \xi^*) \leq \alpha \\ & && \xi^* \in \underset{\xi \in \Xi}{\text{argmin}} \{I(\xi) : F(u, \xi) \geq z\} \end{aligned}$$

Observe that, since the dominating point depends on the choice of the decision variables  $u$ , the resulting problem is a generic nonlinear bilevel optimization problem. Moreover, even though we use Taylor approximations of  $F$  in constructing  $P_k$ , see (3.2), the lower-level problem that characterizes the dominating point  $\xi^*$  utilizes the original nonlinear function. This is crucial for the validity of the asymptotic results in Theorem 3.2 and Theorem 3.3.

To tackle the bilevel structure, we replace the lower-level problem with its first-order optimality conditions. Although it is straightforward to consider the extension where the support  $\Xi = \{\xi \in \mathbb{R}^n : G_j(\xi) \leq 0, j \in \{1, \dots, J\}\}$  has an explicit algebraic description in terms of inequalities, we assume that  $\Xi = \mathbb{R}^n$  for simplicity of exposition. We have the following single-level approximation of the chance-constrained problem (1.1):

$$(3.9) \quad \begin{aligned} & \underset{u, \xi^*, \lambda}{\text{minimize}} && J(u) \\ & \text{subject to} && u \in \mathcal{U}, \xi^* \in \Xi, \lambda \in \mathbb{R}_+ \\ & && P_k(u, z, \xi^*) \leq \alpha, \\ & && F(u, \xi^*) = z, \\ & && \nabla I(\xi^*) = \lambda \nabla_{\xi} F(u, \xi^*). \end{aligned}$$

Note that we have replaced  $F(u, \xi^*) \geq z$  with equality. Doing so relies on an implicit (yet mild) assumption that  $F(u, \tilde{\xi}) < z$ , where  $\tilde{\xi} := \underset{\xi \in \Xi}{\text{argmin}} I(\xi)$ , and we provide a formal argument in Theorem 3.4 below. In any case, note that  $\tilde{\xi}$  is unique and always exists (see proof of Theorem 3.3), and continuity of  $F$  implies that this assumption is expected to hold for sufficiently large values of  $z$ . It is important to note that this argument does not rely on any convexity assumptions on  $F$ .

Finally, we note that the rate function  $I$  may not always be explicitly available depending on the distribution of  $\xi$ . In such cases, we can instead rely on the fact that  $I$  is the conjugate of the cumulant generating function  $S$ , for which we typically have efficient statistical codes that can compute its values and gradients for several distributions. Indeed, from the definition of the rate function (1.4) and convexity and differentiability of  $S$  (see subsection 1.2), we have

$$(3.10) \quad \begin{aligned} & \xi^* \in \underset{\xi \in \Xi}{\text{argmin}} \{I(\xi) : F(u, \xi) \geq z\} \\ \iff & \exists \eta^* \in \mathbb{R}^n : (\xi^*, \eta^*) \in \left\{ \begin{array}{l} \underset{\xi \in \Xi, \eta \in \mathbb{R}^n}{\text{argmin}} \langle \eta, \xi \rangle - S(\eta) \\ \text{subject to } F(u, \xi) \geq z, \xi = \nabla_{\eta} S(\eta) \end{array} \right\}. \end{aligned}$$

It is straightforward to verify that we then have  $\nabla I(\xi^*) = \eta^*$ . We then obtain the following single-level approximation of the chance-constrained problem (1.1):

$$\begin{aligned}
(3.11) \quad & \underset{u, \xi^*, \eta^*, \lambda}{\text{minimize}} && J(u) \\
& \text{subject to} && u \in \mathcal{U}, \xi^* \in \Xi, \eta^* \in \mathbb{R}^n, \lambda \in \mathbb{R}_+ \\
& && P_k(u, z, \xi^*) \leq \alpha, \\
& && F(u, \xi^*) = z, \\
& && \eta^* = \lambda \nabla_{\xi} F(u, \xi^*), \\
& && \xi^* = \nabla S(\eta^*).
\end{aligned}$$

We establish some properties of the single-level approximation (3.9) or equivalently (3.11).

**Theorem 3.4 (Properties of the LDT chance-constrained optimization).** *Suppose that  $\Xi = \mathbb{R}^n$  and  $F(u, \bar{\xi}) < z$  for all feasible solutions  $u \in \mathcal{U}$  of (1.1) and (3.8), where  $\bar{\xi} := \operatorname{argmin}_{\xi \in \Xi} I(\xi)$ . Then, the following statements are true.*

1. *If  $(u, \xi^*)$  is feasible in the bilevel problem (3.8), then there exists  $\lambda \in \mathbb{R}_+$  such that  $(u, \xi^*, \lambda)$  is feasible in the single-level problem (3.9).*
2. *Under condition (C1), the converse is also true: if  $(u, \xi^*, \lambda)$  is feasible in (3.9), then  $(u, \xi^*)$  is feasible in (3.8). Therefore, (3.9) is an exact single-level reformulation of (3.8).*
3. *Under condition (C1) and  $k = 1$ , the first-order LDT problem (3.9) is a conservative approximation of the chance constrained problem (1.1) for any  $z \in \mathbb{R}$ ; that is, if  $(u, \xi^*, \lambda)$  is feasible in (3.9), then  $u$  is feasible in (1.1).*
4. *Under conditions (C1), (C2) and (C3), the feasible solutions of both the first- ( $k = 1$ ) and second-order ( $k = 2$ ) LDT problems (3.9) are in one-one correspondence with those of the chance constrained problem (1.1) and therefore, the former constitutes an asymptotic reformulation of the latter, as  $z \rightarrow \infty$ .*

*Proof.* We consider each of the claims separately.

*Proof of claim 1.* From subsection 1.2, we know that  $\|\nabla_{\xi} F(u, \xi)\| > 0$  for all  $u \in \mathcal{U}_0$ . Therefore, all local minimizers of the lower-level problem in (3.8) satisfy the linearly independent constraint qualification. Therefore, if  $(u, \xi^*)$  is feasible in (3.8), then  $\xi^*$  must satisfy the first-order optimality conditions of the lower-level problem for some  $\lambda \in \mathbb{R}_+$ . If  $\lambda = 0$ , then the stationarity condition  $\nabla I(\xi^*) = \lambda \nabla_{\xi} F(u, \xi^*)$  would imply that  $\xi^* = \bar{\xi}$ ; the primal feasibility of the lower-level problem then implies that  $F(u, \xi^*) = F(u, \bar{\xi}) \geq z$  which leads to a contradiction since  $F(u, \bar{\xi}) < z$  by construction. Therefore, we must have  $\lambda > 0$  and complementarity slackness then requires that  $F(u, \xi^*) = z$ .

*Proof of claim 2.* Under condition (C1), the lower-level problem in (3.8) satisfies Slater's constraint qualification, see also the proof of Theorem 3.3. Therefore, its first-order optimality conditions are necessary and sufficient.

*Proof of claim 3.* This is a direct consequence of Proposition 3.1 and claim 2.

*Proof of claim 4.* Suppose that  $(u, \xi^*, \lambda)$  is feasible in (3.9). Then, under conditions (C1), (C2) and (C3), Theorem 3.3 ensures that this  $\xi^*$  is precisely the unique global minimizer appearing in (3.3). Moreover, for this  $u$  and  $\xi^*$ , the measure of the rare event set  $\mathcal{P}(\{\xi \in \Xi : F(u, \xi) \geq z\}) \asymp \exp(-I(\xi^*))$  as  $z \rightarrow \infty$ . An application of the same theorem to

$F_k(u, \xi; \xi^*)$  (see (3.1a) and (3.1b)) instead of  $F$ , must require that  $\xi^*$  again be the unique global minimizer appearing in (3.3); this is because of  $\nabla_\xi F(u, \xi^*) = \nabla_\xi F_k(u, \xi^*; \xi^*)$  (by construction) along with the strict convexity of the rate function  $I$ . Therefore, again we have  $\mathcal{P}(\{\xi \in \Xi : F_k(u, \xi; \xi^*) \geq z\}) \asymp \exp(-I(\xi^*))$  as  $z \rightarrow \infty$ . The claim now follows from claim 2. ■

We remark that the single-level reformulations (3.9) and (3.11) are *conceptual* in the sense that they require an instantiation of the probability estimate  $P_k$  as well as gradients of the rate function  $I$  and/or the cumulant generating function  $S$ . In Section 4 and Section 5, we provide explicit single-level formulations in the case where  $\xi$  is a Gaussian and a mixture of Gaussians, respectively.

**4. Explicit formulations for Gaussian random parameter.** Throughout this section, we shall assume that  $\xi \sim \mathcal{N}(\mu, \Sigma)$  follows a Gaussian distribution with mean  $\mu$  and covariance  $\Sigma \succ 0$ . Our goal will be to provide explicit formulations of the first- and second-order LDT chance constrained optimization problems (3.9).

**4.1. Rate function.** The moment generating function of  $\xi$  is  $\exp(\eta^\top \mu + \frac{1}{2} \eta^\top \Sigma \eta)$ , and therefore, its cumulant generating function,  $S(\eta) = \eta^\top \mu + \frac{1}{2} \eta^\top \Sigma \eta$ , is quadratic. Its rate function can be computed explicitly through the Legendre transform of  $S$ , and is also quadratic:

$$(4.1) \quad I(\xi) = \frac{1}{2} \|\xi - \mu\|_{\Sigma^{-1}}^2.$$

Observe that  $I(\xi)$  is, up to a normalization constant, the negative log-probability density of  $\mathcal{N}(\mu, \Sigma)$ . Hence, for a Gaussian distribution, minimizing the rate function over some domain, such as in (1.2), is equivalent to maximizing the probability density function over that domain. We note, however, this is a special property of Gaussians, and it does not hold in general; indeed, the next section shows that this property no longer holds for non-Gaussian parameters.

**4.2. First- and second-order probability estimates.** For  $\xi \sim \mathcal{N}(\mu, \Sigma)$ , the probability estimates in (3.2) are Gaussian integrals over a half-space ( $k = 1$ ), or a set bounded by a quadric ( $k = 2$ ). The explicit formulas for  $P_k$  can be computed by first splitting the respective domains into a normal and orthogonal component (with respect to  $\nabla_\xi F(u, \xi^*)$ ), and then by computing the integral along each direction directly [47]. In the remainder of this discussion, we fix  $u$  and  $\xi^*$  and assume that they satisfy  $F(u, \xi^*) = z$ .

The first-order approximation is a Gaussian integral over a half-space bounded by a plane perpendicular to  $\hat{n}^* := \nabla_\xi F(u, \xi^*) / \|\nabla_\xi F(u, \xi^*)\|$ , and passing through  $\xi^*$ . It is straightforward to verify that this integral degenerates to a one-dimensional Gaussian integral and can be computed using the dominating point  $\xi^*$  as follows:

$$(4.2) \quad P_1(u, z, \xi^*) = \Phi(-\sqrt{2I(\xi^*)}) = \Phi(-\|\xi^* - \mu\|_{\Sigma^{-1}}).$$

The second-order probability estimate  $P_2$  is slightly more complicated, but can still be approximated using an analytical expression. Following [47], we first approximate the surface  $F_2(u, \xi; \xi^*) = z$  by a paraboloid with axis of symmetry  $\hat{n}^*$  at  $\xi^*$ , and curvatures from  $\nabla_\xi^2 F(u, \xi^*)$  in the directions on the orthogonal plane.  $P_2(u, z, \xi^*)$  can then be approximated as the measure of the set bounded by that paraboloid, which turns out to be equal to  $P_1(u, z, \xi^*)$  multiplied by a correction term. This correction term can be computed either through

principle curvatures of  $F$  at  $\xi^*$  [20], or by using the eigenvalues of the rotated covariance-preconditioned Hessian of  $F$  at  $\xi^*$ ; see [47] for details. In this paper, we introduce another way to compute this correction term, which turns out to be more numerically tractable for optimization purposes.

**Theorem 4.1 (Second-order probability estimate for Gaussians).** *Suppose that  $\xi \sim \mathcal{N}(\mu, \Sigma)$ ,  $u \in \mathcal{U}$  and  $\xi^* \in \Xi$  satisfy  $F(u, \xi^*) = z$ , and that the following matrix is invertible:*

$$H := \mathbf{I}_n - \frac{\|\xi^* - \mu\|_{\Sigma^{-1}}}{\left\| \Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*) \right\|} \Sigma^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma^{\frac{1}{2}}$$

Then the second-order approximation of the probability  $\mathbb{P}(F(u, \xi) \geq z)$  is given by:

$$(4.3) \quad P_2(u, z, \xi^*) \approx \Phi(-\|\xi^* - \mu\|_{\Sigma^{-1}}) \det_{\perp \hat{n}}(H)^{-\frac{1}{2}},$$

where the covariance-preconditioned normal  $\hat{n} = \frac{\Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*)}{\left\| \Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*) \right\|}$  and orthogonal determinant

$$(4.4) \quad \det_{\perp \hat{n}} H = \left( \hat{n}^{\top} H^{-1} \hat{n} \right) \det H.$$

*Proof.* Following the proof of Theorem 4.4 in [47], the second-order probability approximation can be shown to be equal to:

$$P_2(u, z, \xi^*) = P_1(u, z, \xi^*) C(u, \xi^*),$$

where  $C(u, \xi^*)$  is a correction term computed through the curvatures of  $F$  at  $\xi^*$ . To define it, we first denote  $S(\vartheta)$  to be the Lebesgue measure on the orthogonal component  $\vartheta \in \{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}$  of  $\hat{n}$  at  $\xi^*$ . It can then be shown that

$$\begin{aligned} C(u, \xi^*) &\approx (2\pi)^{-\frac{n-1}{2}} \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} \exp \left( -\frac{1}{2} \left\langle \vartheta, \mathbf{I}_n - \frac{\|\Sigma^{\frac{1}{2}} \nabla I(\xi^*)\|}{\left\| \Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*) \right\|} \Sigma^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma^{\frac{1}{2}} \vartheta \right\rangle \right) dS(\vartheta) \\ &= (2\pi)^{-\frac{n-1}{2}} \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} \exp \left( -\frac{1}{2} \langle \vartheta, H \vartheta \rangle \right) dS(\vartheta) \end{aligned}$$

where we have used  $\nabla I(\xi^*) = \Sigma^{-1}(\xi^* - \mu)$  along with the definition of  $H$ . The above expression can be further simplified by defining the unit vector:

$$(4.5) \quad \hat{t} := \frac{H^{-1} \hat{n}}{\|H^{-1} \hat{n}\|}.$$

Then, for all  $\xi \in \mathbb{R}^n$ , there exist  $s \in \mathbb{R}$  and  $\vartheta \in \{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}$  such that  $\xi = \vartheta + s\hat{t}$ . From Fubini's theorem and the orthogonality between  $\hat{n}$  and  $\xi$ , we have

$$(4.6) \quad \begin{aligned} \int_{\mathbb{R}^n} e^{-\frac{1}{2} \langle \xi, H \xi \rangle} d\xi &= \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} \int_{\mathbb{R}} e^{-\frac{1}{2} \langle \vartheta + s\hat{t}, H(\vartheta + s\hat{t}) \rangle} \langle \hat{n}, \hat{t} \rangle ds dS(\vartheta) \\ &= \langle \hat{n}, \hat{t} \rangle \int_{\mathbb{R}} e^{-\frac{1}{2} \langle \hat{t}, H \hat{t} \rangle s^2} ds \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} e^{-\frac{1}{2} \langle \vartheta, H \vartheta \rangle} dS(\vartheta). \end{aligned}$$



Applying the Gauss integral formula to both sides of (4.6), we obtain

$$(\det H)^{-\frac{1}{2}} = (2\pi)^{-\frac{n-1}{2}} \langle \hat{n}, \hat{t} \rangle \langle \hat{t}, H\hat{t} \rangle^{-\frac{1}{2}} \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} e^{-\frac{1}{2} \langle \vartheta, H\vartheta \rangle} dS(\vartheta).$$

Plugging in the definition of  $\hat{t}$  in (4.5), we obtain

$$\begin{aligned} (2\pi)^{-\frac{n-1}{2}} \int_{\{\vartheta \in \mathbb{R}^n : \langle \hat{n}, \vartheta \rangle = 0\}} e^{-\frac{1}{2} \langle \vartheta, H\vartheta \rangle} dS(\vartheta) &= \langle \hat{n}, \hat{t} \rangle^{-1} \langle \hat{t}, H\hat{t} \rangle^{\frac{1}{2}} (\det H)^{-\frac{1}{2}} \\ &= \left( \frac{\langle \hat{n}, H^{-1}\hat{n} \rangle^2}{\|H^{-1}\hat{n}\|^2} \frac{\|H^{-1}\hat{n}\|^2}{\langle H^{-1}\hat{n}, HH^{-1}\hat{n} \rangle} \det H \right)^{-\frac{1}{2}} \\ &= \left( (\hat{n}^\top H^{-1}\hat{n}) \det H \right)^{-\frac{1}{2}} \\ &= \det_{\perp \hat{n}}(H)^{-\frac{1}{2}}. \end{aligned}$$

This establishes the statement of the theorem. ■

Note that the expression (4.3) in Theorem 4.1 is not the same as  $P_2$  in (3.2), and only provides an approximation of the latter. Nevertheless, under the assumption that  $\xi^*$  is the LDT minimizer in the lower-level problem of (3.8), it is asymptotically equal to (3.2) as  $z \rightarrow \infty$ . Crucially, it has the added advantage of being computable in closed-form.

**4.3. LDT chance-constrained problems.** An application of the approximations in (4.2) and (4.3) to (3.9) yield the following first- ( $k = 1$ ) and second-order ( $k = 2$ ) LDT approximations of the chance-constrained problem (1.1), respectively:

$$(4.7) \quad \begin{aligned} &\underset{u, \xi^*, \lambda}{\text{minimize}} && J(u) \\ &\text{subject to} && u \in \mathcal{U}, \xi^* \in \Xi, \lambda \in \mathbb{R}_+ \\ &&& P_k(u, z, \xi^*) \leq \alpha, \\ &&& F(u, \xi^*) = z, \\ &&& \Sigma^{-1}(\xi^* - \mu) = \lambda \nabla_{\xi} F(u, \xi^*). \end{aligned}$$

Specifically, we replace the first-order probability constraint  $P_1(u, z, \xi^*) \leq \alpha$  as follows:

$$(4.8) \quad -\sqrt{2I(\xi^*)} \leq \Phi^{-1}(\alpha).$$

Similarly, we replace the second-order probability constraint  $P_2(u, z, \xi^*) \leq \alpha$  as follows:

$$(4.9) \quad \Phi(-\sqrt{2I(\xi^*)}) \det_{\perp \hat{n}} \left( \mathbf{I}_n - \lambda \Sigma^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma^{\frac{1}{2}} \right)^{-\frac{1}{2}} \leq \alpha,$$

where we have used the fact that

$$(4.10) \quad \frac{\|\xi^* - \mu\|_{\Sigma^{-1}}}{\left\| \Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*) \right\|} = \frac{\left\| \Sigma^{\frac{1}{2}} \lambda \nabla_{\xi} F(u, \xi^*) \right\|}{\left\| \Sigma^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*) \right\|} = \lambda.$$

We close this section with a few important remarks. First, note that since all of the functions involved in the closed-form expressions (4.8) and (4.9) are fairly standard, we can use standard automatic differentiation routines to compute their function values and gradients. Notably, this allows us to use off-the-shelf solvers for the solution of (4.7).

Second, the matrix  $H = I_n - \lambda \Sigma^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma^{\frac{1}{2}}$  appearing in (4.9) is guaranteed to be positive definite (and hence, invertible) whenever the local solutions of the lower-level problem in (3.8) satisfy the sufficient second-order optimality conditions. Although the solutions of (4.7) cannot guarantee this in general, we have nonetheless empirically found in our numerical experiments that the linear system  $Hx = \hat{n}$  was almost always solvable; in such cases, we can equivalently reformulate (4.4) as  $\hat{n}^\top x \det(H)$  and evaluate its function values and gradients.

Third, and finally, although the above derivations are for Gaussian random parameters, the probability estimates (4.2) and (4.3) as well as the resulting LDT chance-constrained problem (4.7) can be also used for random parameters that can be transformed to a Gaussian distribution. Specifically, if for a random parameter  $\zeta$ , there exists an invertible function  $f$ , such that  $f(\zeta)$  is Gaussian, then we can use  $\xi = f(\zeta)$  instead as the random parameter and apply the above results to  $\xi$  while involving  $f^{-1}$  in  $F$ . For example, when  $\zeta$  follows a log-normal distribution, we can use  $\xi = \log(\zeta)$  as the random parameter, while replacing  $\zeta$  with  $\exp(\xi)$  in  $F$ . We discuss these techniques further in subsections 6.1 and 6.2.

**5. Explicit formulations for Gaussian mixture random parameter.** We now generalize the approach from the last section for non-Gaussian random parameters to provide explicit formulations of the first- and second-order LDT chance constrained optimization problems (3.9). Specifically, we assume that  $\xi \sim \sum_{i=1}^M w_i \mathcal{N}(\mu_i, \Sigma_i)$  follows a Gaussian mixture distribution with  $M$  components, where  $w_i > 0$  and  $\sum_{i=1}^M w_i = 1$ . Here, we use  $\mathcal{P}_i = \mathcal{N}(\mu_i, \Sigma_i)$  to denote the distribution of the  $i^{\text{th}}$  component with mean  $\mu_i$  and covariance  $\Sigma_i \succ 0$ . Note that  $\mathcal{P}$  is a convex combination of  $\mathcal{P}_i$ ,  $i \in \{1, \dots, M\}$ :

$$(5.1) \quad \mathcal{P} = \sum_{i=1}^M w_i \mathcal{P}_i.$$

**5.1. Rate function.** Unlike in the case of Gaussians, the rate function is no longer equal to the negative log-density of  $\xi$ , and is available only implicitly. We first compute the cumulant generating function using its definition as follows:

$$(5.2) \quad \begin{aligned} S(\eta) &= \log \left\{ \int_{\mathbb{R}^n} e^{\eta^\top \xi} d\mathcal{P}(\xi) \right\} = \log \left\{ \int_{\mathbb{R}^n} e^{\eta^\top \xi} \sum_{i=1}^M w_i d\mathcal{P}_i(\xi) \right\} \\ &= \log \left\{ \int_{\mathbb{R}^n} e^{\eta^\top \xi} \sum_{i=1}^M \left[ w_i (2\pi)^{-\frac{n}{2}} (\det \Sigma_i)^{-\frac{1}{2}} e^{-\frac{1}{2}(\xi - \mu_i)^\top \Sigma_i^{-1}(\xi - \mu_i)} \right] d\xi \right\} \\ &= \log \left\{ \sum_{i=1}^M w_i e^{\eta^\top \mu_i + \frac{1}{2} \eta^\top \Sigma_i \eta} \right\}. \end{aligned}$$

Since the rate function  $I$  is the Legendre transform of  $S$ , it is non-trivial to compute it explicitly in this case. Nevertheless, as noted in Section 3, we can still compute the LDT

minimizer  $\xi^*$  by solving (3.10) instead:

$$(5.3) \quad \begin{aligned} (\xi^*, \eta^*) \in \operatorname{argmin}_{\xi, \eta \in \mathbb{R}^n} & \quad \eta^\top \xi - \log \left\{ \sum_{i=1}^M w_i e^{\eta^\top \mu_i + \frac{1}{2} \eta^\top \Sigma_i \eta} \right\} \\ \text{subject to } & \quad F(u, \xi) = z, \quad \xi = \frac{\sum_{i=1}^M w_i e^{\eta^\top \mu_i + \frac{1}{2} \eta^\top \Sigma_i \eta} (\mu_i + \Sigma_i \eta)}{\sum_{i=1}^M w_i e^{\eta^\top \mu_i + \frac{1}{2} \eta^\top \Sigma_i \eta}}. \end{aligned}$$

Moreover, even though the rate function is not available in closed form, the following theorem establishes that Gaussian mixture distributions (and hence, Gaussian distributions) always satisfy condition (C3) in Theorem 3.3. Therefore, the conditions of Theorem 3.3 depend only on the behavior of the constraint function  $F$  with respect to the uncertain parameters  $\xi$ , and not on the form of the mixture distribution.

**Theorem 5.1 (Sufficient conditions for Gaussian mixture distributions).** *Suppose that  $\xi \sim \sum_{j=1}^M w_j \mathcal{N}(\mu_j, \Sigma_j)$ . Then condition (C3) in Theorem 3.3 is satisfied.*

*Proof.* We shall show that

- (a) the cumulant generating function  $S$  is strongly convex, and
- (b) the rate function  $I$  satisfies  $I(\xi) \geq \min_{j \in \{1, \dots, M\}} I_j(\xi)$  for all  $\xi \in \mathbb{R}^n$ , where  $I_j(\xi) = \frac{1}{2} \|\xi - \mu_j\|_{\Sigma_j^{-1}}^2$  denotes the rate function of the  $j^{\text{th}}$  Gaussian  $\mathcal{N}(\mu_j, \Sigma_j)$ , see (4.1).

Assume for the moment that the above statements are true, and let  $\sigma > 0$  denote the strong convexity parameter of  $S$  in (a). Since the rate function (1.4) is the convex conjugate of  $S$ , strong convexity of  $S$  implies that the gradient of the rate function  $\nabla I(\xi)$  is Lipschitz continuous with constant  $\sigma$ ; that is,

$$(5.4) \quad I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle \geq I(\xi) - \frac{\sigma}{2} \|\xi - \hat{\xi}\|^2, \quad \forall \xi, \hat{\xi} \in \mathbb{R}^n.$$

Now, fix  $r > 0$  and  $\hat{\xi} \in \mathbb{R}^n$ , and let  $\tilde{\xi}$  be the (unique) minimizer of the integrand in (C3) over  $B(r, \hat{\xi})$ . Since for the given Gaussian mixture distribution  $\mathcal{P}$ , (see Section 5)

$$d\mathcal{P}(\xi) = \sum_{j=1}^M \left[ w_j (2\pi)^{-\frac{n}{2}} (\det \Sigma_j)^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \|\xi - \mu_j\|_{\Sigma_j^{-1}}^2 \right) \right] d\xi,$$

we have that

$$\tilde{\xi} = \operatorname{argmin}_{\xi \in B(r, \hat{\xi})} \left\{ \exp \left( I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle \right) \sum_{j=1}^M \left[ w_j (2\pi)^{-\frac{n}{2}} (\det \Sigma_j)^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \|\xi - \mu_j\|_{\Sigma_j^{-1}}^2 \right) \right] \right\}.$$

Combining the above with (5.4), we obtain

$$\int_{B(r, \hat{\xi})} \exp \left( I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle \right) d\mathcal{P}(\xi) \geq \sum_{j=1}^M c_j \exp \left( I(\tilde{\xi}) - I_j(\tilde{\xi}) - \frac{\sigma}{2} \|\tilde{\xi} - \hat{\xi}\|^2 \right)$$

$$\geq \sum_{j=1}^M c_j e^{-\frac{\sigma}{2}r^2} \exp\left(I(\tilde{\xi}) - I_j(\tilde{\xi})\right),$$

$$\left(\text{with } c_j := w_j(2\pi)^{-\frac{n}{2}}(\det \Sigma_j)^{-\frac{1}{2}} \int_{B(r, \hat{\xi})} d\xi > 0\right),$$

where we have used the definition of the  $j^{\text{th}}$  component rate function  $I_j(\xi) = \frac{1}{2}\|\xi - \mu_j\|_{\Sigma_j^{-1}}^2$ , along with the fact that  $\tilde{\xi}, \hat{\xi} \in B(r, \hat{\xi}) \implies \|\tilde{\xi} - \hat{\xi}\| \leq r$ . Claim (b) implies that there exists some  $k \in \{1, \dots, M\}$  such that  $I(\hat{\xi}) \geq I_k(\hat{\xi})$ . Therefore, we have shown that

$$\int_{B(r, \hat{\xi})} \exp\left(I(\hat{\xi}) + \langle \nabla_{\xi} I(\hat{\xi}), \xi - \hat{\xi} \rangle\right) d\mathcal{P}(\xi) \geq \min_{j \in \{1, \dots, M\}} c_j e^{-\frac{\sigma}{2}r^2} > 0, \quad \forall \hat{\xi} \in \mathbb{R}^n.$$

Also, we have

$$\int_{B(r, \hat{\xi})} \exp\left(I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle\right) d\mathcal{P}(\xi) \leq \int_{\mathbb{R}^n} \exp\left(I(\hat{\xi}) + \langle \nabla I(\hat{\xi}), \xi - \hat{\xi} \rangle\right) d\mathcal{P}(\xi) = 1.$$

where the equality can be verified easily (*e.g.*, see the proof of Theorem 2.1 in [16]). Therefore, the left-hand side expression in condition (C3) is bounded away from 0 by a constant that is independent of  $\hat{\xi}$ , whereas its right-hand side  $\rightarrow \infty$  as  $\|\hat{\xi}\| \rightarrow \infty$  (see proof of Theorem 3.3). This proves the statement of this theorem.

It remains to prove claims (a) and (b). To show (a), let  $S_j(\eta) = \eta^\top \mu_j + \frac{1}{2}\eta^\top \Sigma_j \eta$  denote the cumulant generating function of the  $j^{\text{th}}$  Gaussian  $\mathcal{N}(\mu_j, \Sigma_j)$ . Then, the cumulant generating function of the Gaussian mixture  $S(\eta) = \log\left(\sum_{j=1}^M w_j \exp(S_j(\eta))\right)$ , see (5.2). A straightforward calculation shows that its Hessian satisfies

$$\begin{aligned} \nabla^2 S(\eta) &= \begin{pmatrix} \left[ \sum_{j=1}^M e^{S_j(\eta)} \right] \left[ \sum_{j=1}^M e^{S_j(\eta)} \left( \nabla S_j(\eta) (\nabla S_j(\eta))^\top + \nabla^2 S_j(\eta) \right) \right] \\ - \left[ \sum_{j=1}^M e^{S_j(\eta)} \nabla S_j(\eta) \right] \left[ \sum_{j=1}^M e^{S_j(\eta)} (\nabla S_j(\eta))^\top \right] \end{pmatrix} \cdot \left[ \sum_{j=1}^M e^{S_j(\eta)} \right]^{-2} \\ &= \begin{pmatrix} \sum_{j=1}^M e^{2S_j(\eta)} \nabla^2 S_j(\eta) + \sum_{j=1}^M \sum_{i=j+1}^M e^{S_i(\eta) + S_j(\eta)} \left( \nabla^2 S_i(\eta) + \nabla^2 S_j(\eta) \right) \\ + [\nabla S_i(\eta) - \nabla S_j(\eta)] [\nabla S_i(\eta) - \nabla S_j(\eta)]^\top \end{pmatrix} \cdot \left[ \sum_{j=1}^M e^{S_j(\eta)} \right]^{-2} \\ &\succeq \varepsilon \mathbf{I}_n \cdot \left( \sum_{j=1}^M e^{2S_j(\eta)} + \sum_{j=1}^M \sum_{i=j+1}^M 2e^{S_i(\eta) + S_j(\eta)} \right) \cdot \left[ \sum_{j=1}^M e^{S_j(\eta)} \right]^{-2} \\ &= \varepsilon \mathbf{I}_n, \end{aligned}$$

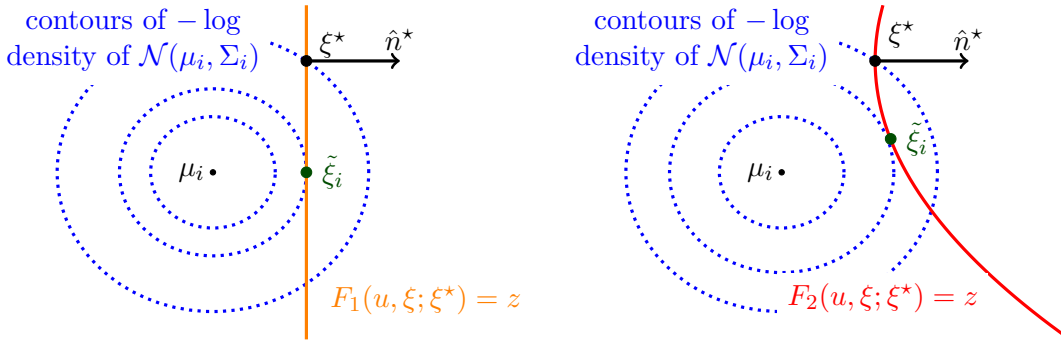
where  $\varepsilon = \min_{j \in \{1, \dots, M\}} \lambda_{\min}(\nabla^2 S_j) = \min_{j \in \{1, \dots, M\}} \lambda_{\min}(\Sigma_j) > 0$ , and  $\lambda_{\min}(A)$  denotes the minimum eigenvalue of a matrix  $A$ .

To show (b), let  $f^*$  denote the conjugate of a function  $f$ . Convexity and continuity of the cumulant generating function implies  $I^* = (S^*)^* = I$ , and hence, for any  $\eta \in \mathbb{R}^n$ , we have

$$\begin{aligned} I^*(\eta) = S(\eta) &= \log \left( \sum_{j=1}^M w_j \exp(S_j(\eta)) \right) \leq \log \left( \max_{j \in \{1, \dots, M\}} \exp(S_j(\eta)) \right) \\ &= \max_{j \in \{1, \dots, M\}} S_j(\eta) \\ &= \max_{j \in \{1, \dots, M\}} \max_{\xi \in \mathbb{R}^n} \left\{ \eta^\top \xi - I_j(\xi) \right\} \\ &= \max_{\xi \in \mathbb{R}^n} \left\{ \eta^\top \xi - \min_{j \in \{1, \dots, M\}} I_j(\xi) \right\} \\ &= \left( \min_{j \in \{1, \dots, M\}} I_j \right)^* (\eta) \end{aligned}$$

where the first inequality follows from the fact that  $w_j \geq 0$ ,  $\sum_{j=1}^M w_j = 1$ , the first equality follows from the monotonicity of  $\exp$ , the second and final equalities follow from the definition of the convex conjugate, and the third equality follows because we can interchange the order of the max over  $j$  with the max over  $\xi$ . Finally, the order reversing property of the conjugate  $f^* \leq g^* \implies f \geq g$ , implies that  $I(\xi) \geq \min_{j \in \{1, \dots, M\}} I_j(\xi)$  which is claim (b).  $\blacksquare$

**5.2. First- and second-order probability estimates.** Recall from subsection 3.1 that the first-order probability estimate  $P_1$  in (3.2) is the measure of the half-space bounded by  $F_1(u, \xi; \xi^*) = z$ , where  $\xi^*$  solves (5.3). Since a Gaussian mixture is just a convex combination of Gaussians, we can exploit the analogous result for the latter (4.2) to compute the first-order estimate analytically.



**Figure 3.** Illustration of finding  $\tilde{\xi}_i$  for each Gaussian component  $\mathcal{N}(\mu_i, \Sigma_i)$ . The left figure shows the procedure for the first-order approximation, namely of finding a point on the hyperplane  $F_1(u, \xi; \xi^*) = z$  where the contour of negative log-density of  $\mathcal{N}(\mu_i, \Sigma_i)$  is tangential to the plane, by solving (5.7). Similarly, the right figure shows the procedure for the second-order approximation, by finding  $\tilde{\xi}_i$  on  $F_2(u, \xi; \xi^*) = z$  by solving (5.9).

**Theorem 5.2 (First-order probability estimate for Gaussian mixtures).** Suppose that  $\xi \sim \sum_{i=1}^M w_i \mathcal{N}(\mu_i, \Sigma_i)$  and for some fixed  $u \in \mathcal{U}$ ,  $\xi^*$  is the solution to (5.3). Then the first-order

approximation of the probability  $\mathbb{P}(F(u, \xi) \geq z)$  is given by

$$(5.5) \quad P_1(u, z, \xi^*) = \sum_{i=1}^M w_i \Phi \left( -\frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\|\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*)\|} \right).$$

*Proof.* Substituting (5.1) in (3.2) for  $k = 1$ , we obtain:

$$(5.6) \quad P_1(u, z, \xi^*) = \sum_{i=1}^M w_i \mathcal{P}_i(\{\xi : F_1(u, \xi; \xi^*) \geq z\}).$$

Therefore, we can consider each  $\mathcal{P}_i(\{\xi : F_1(u, \xi; \xi^*) \geq z\})$ ,  $i \in \{1, \dots, M\}$ , separately. Similar to (4.2) for Gaussians, this quantity is a Gaussian integral over a half-space. However, unlike the latter, the contours of negative log-density of  $\mathcal{N}(\mu_i, \Sigma_i)$ , i.e., of  $\frac{1}{2}\|\xi - \mu_i\|_{\Sigma_i}^2$ , need not necessarily be tangential to the half-space  $F_1(u, \xi; \xi^*) = z$  at  $\xi^*$ . Therefore, to compute  $\mathcal{P}_i(\{\xi : F_1(u, \xi; \xi^*) \geq z\})$ , we need to find the point  $\tilde{\xi}_i$  on the hyperplane  $F_1(u, \xi; \xi^*) = z$  where the contour of  $\frac{1}{2}\|\xi - \mu_i\|_{\Sigma_i}^2$  is tangential to the plane (see Figure 3):

$$(5.7) \quad \begin{aligned} \tilde{\xi}_i &= \operatorname{argmin}_{\xi \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\xi - \mu_i\|_{\Sigma_i}^2 : F_1(u, \xi; \xi^*) = z \right\} \\ &= \operatorname{argmin}_{\xi \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\xi - \mu_i\|_{\Sigma_i}^2 : \langle \nabla_{\xi} F(u, \xi^*), \xi - \xi^* \rangle = 0 \right\}, \end{aligned}$$

where we used the definition of the first-order Taylor expansion (3.1a) and the fact that  $\xi^*$  satisfies  $F(u, \xi^*) = z$  since it solves (5.3).

The optimality conditions of (5.7) imply that there exists  $\lambda_i \geq 0$  such that:

$$\Sigma_i^{-1}(\tilde{\xi}_i - \mu_i) = \lambda_i \nabla_{\xi} F(u, \xi^*).$$

Therefore,  $\tilde{\xi}_i = \mu_i + \lambda_i \Sigma_i \nabla_{\xi} F(u, \xi^*)$ . Substituting this in  $\langle \nabla_{\xi} F(u, \xi^*), \tilde{\xi}_i - \xi^* \rangle = 0$ , we obtain

$$\lambda_i = \frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\langle \nabla_{\xi} F(u, \xi^*), \Sigma_i \nabla_{\xi} F(u, \xi^*) \rangle}.$$

Therefore, we have that

$$\tilde{\xi}_i = \mu_i + \frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\langle \nabla_{\xi} F(u, \xi^*), \Sigma_i \nabla_{\xi} F(u, \xi^*) \rangle} \Sigma_i \nabla_{\xi} F(u, \xi^*).$$

Finally, replacing  $\xi^*$  in the formula of the first-order probability estimate for Gaussians (4.2)

with  $\tilde{\xi}_i$  yields

$$\begin{aligned}
\mathcal{P}_i(\{\xi : F_1(u, \xi; \xi^*) \geq z\}) &= \Phi\left(-\|\tilde{\xi}_i - \mu_i\|_{\Sigma_i^{-1}}\right) \\
&= \Phi\left(-\left\|\mu_i + \frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\langle \nabla_{\xi} F(u, \xi^*), \Sigma_i \nabla_{\xi} F(u, \xi^*) \rangle} \Sigma_i \nabla_{\xi} F(u, \xi^*) - \mu_i\right\|_{\Sigma_i^{-1}}\right) \\
&= \Phi\left(-\frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\langle \nabla_{\xi} F(u, \xi^*), \Sigma_i \nabla_{\xi} F(u, \xi^*) \rangle} \|\Sigma_i \nabla_{\xi} F(u, \xi^*)\|_{\Sigma_i^{-1}}\right) \\
&= \Phi\left(-\frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\|\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*)\|}\right).
\end{aligned}$$

Combining this with (5.6) gives us the claimed formula (5.5). ■

The second-order probability estimate  $P_2$  satisfies a linear relationship similar to the first-order estimate  $P_1$ ; that is,  $P_2(u, z, \xi^*) = \sum_{i=1}^M w_i \mathcal{P}_i(\{\xi : F_2(u, \xi; \xi^*) \geq z\})$ . For each  $i \in \{1, \dots, M\}$ , we can then use formula (4.3) for Gaussians to compute  $\mathcal{P}_i(\{\xi : F_2(u, \xi; \xi^*) \geq z\})$ . As in the case of the first-order estimate, this requires us to first find the point on the surface  $F_2(u, \xi; \xi^*) = z$  where the contours of  $\frac{1}{2}\|\xi - \mu_i\|_{\Sigma_i^{-1}}^2$  are tangential to the surface, see Figure 3. The final closed-form expression is presented in the following theorem, whose proof is omitted for brevity.

**Theorem 5.3 (Second-order approximation of probability for Gaussian mixtures).** *Suppose that  $\xi \sim \sum_{i=1}^M w_i \mathcal{N}(\mu_i, \Sigma_i)$  and for some fixed  $u \in \mathcal{U}$ ,  $\xi^*$  is the solution to (5.3). Then the second-order approximation of the probability  $\mathbb{P}(F(u, \xi) \geq z)$  is given by*

$$(5.8) \quad P_2(u, z, \xi^*) = \sum_{i=1}^M w_i \Phi\left(-\|\tilde{\xi}_i - \mu_i\|_{\Sigma_i^{-1}}\right) \det_{\perp \tilde{n}_i} \left( \mathbf{I}_n - \frac{\|\tilde{\xi}_i - \mu_i\|_{\Sigma_i^{-1}}}{\|\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F_2(u, \tilde{\xi}_i; \xi^*)\|} \Sigma_i^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma_i^{\frac{1}{2}} \right)^{-\frac{1}{2}},$$

where

$$(5.9) \quad \tilde{\xi}_i = \underset{\xi \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\xi - \mu_i\|_{\Sigma_i^{-1}}^2 : F_2(u, \xi; \xi^*) = z \right\}, \quad \tilde{n}_i = \frac{\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F_2(u, \tilde{\xi}_i; \xi^*)}{\|\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F_2(u, \tilde{\xi}_i; \xi^*)\|}.$$

**5.3. LDT chance-constrained problems.** Substituting (5.3) and (5.5) in (3.11) ( $k = 1$ ) yields the following first-order LDT approximation of the chance-constrained problem (1.1):

$$\begin{aligned}
(5.10) \quad & \underset{u, \xi^*, \eta^*, \lambda}{\text{minimize}} && J(u) \\
& \text{subject to} && u \in \mathcal{U}, \xi^* \in \mathbb{R}^n, \eta^* \in \mathbb{R}^n, \lambda \in \mathbb{R}_+ \\
& && \sum_{i=1}^M w_i \Phi \left( -\frac{\langle \nabla_{\xi} F(u, \xi^*), \xi^* - \mu_i \rangle}{\|\Sigma_i^{\frac{1}{2}} \nabla_{\xi} F(u, \xi^*)\|} \right) \leq \alpha, \\
& && F(u, \xi^*) = z, \\
& && \eta^* = \lambda \nabla_{\xi} F(u, \xi^*), \\
& && \xi^* = \frac{\sum_{i=1}^M w_i \exp \left( \eta^{*\top} \mu_i + \frac{1}{2} \eta^{*\top} \Sigma_i \eta^* \right) (\mu_i + \Sigma_i \eta^*)}{\sum_{i=1}^M w_i \exp \left( \eta^{*\top} \mu_i + \frac{1}{2} \eta^{*\top} \Sigma_i \eta^* \right)}.
\end{aligned}$$

To obtain the second-order LDT approximation, we note that we first have to reformulate the nonconvex quadratic optimization problem (5.9) that appears in the closed-form expression of the second-order probability estimate (5.8). The following lemma provides a characterization of  $\tilde{\xi}_i$  that we shall use subsequently.

**Lemma 5.4.** *Fix  $u \in \mathcal{U}_0$  and suppose that  $\xi^*$  is the solution to (5.3). Then, for each  $i \in \{1, \dots, M\}$ ,  $\tilde{\xi}_i$  is a global minimizer of  $\min_{\xi \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\xi - \mu_i\|_{\Sigma_i^{-1}}^2 : F_2(u, \xi; \xi^*) = z \right\}$  if and only if there exists  $\tilde{\lambda}_i \in \mathbb{R}_+$  such that*

$$(5.11a) \quad F_2(u, \tilde{\xi}_i; \xi^*) = z,$$

$$(5.11b) \quad \Sigma_i^{-1}(\tilde{\xi}_i - \mu_i) = \tilde{\lambda}_i \left( \nabla_{\xi} F(u, \xi^*) + \nabla_{\xi}^2 F(u, \xi^*)(\tilde{\xi}_i - \xi^*) \right),$$

$$(5.11c) \quad \mathbf{I}_n - \tilde{\lambda}_i \Sigma_i^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma_i^{\frac{1}{2}} \succeq 0.$$

*Proof.* First, note that  $\nabla_{\xi} F_2(u, \xi^*; \xi^*) = \nabla_{\xi} F(u, \xi^*) \neq 0$  (see subsection 1.2). Therefore, [35, Theorem 3.2] shows that  $\tilde{\xi}_i$  is a global minimizer if and only if (5.11a) and (5.11b) are satisfied, along with  $\Sigma_i^{-1} - \tilde{\lambda}_i \nabla_{\xi}^2 F_2(u, \tilde{\xi}_i; \xi^*) \succeq 0$ . Since  $\nabla_{\xi}^2 F_2(u, \tilde{\xi}_i; \xi^*) = \nabla_{\xi}^2 F(u, \xi^*)$  by construction, this condition is equivalent to  $\Sigma_i^{-\frac{1}{2}} \left( \mathbf{I}_n - \tilde{\lambda}_i \Sigma_i^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma_i^{\frac{1}{2}} \right) \Sigma_i^{-\frac{1}{2}} \succeq 0$ , which is seen to be equivalent to (5.11c).  $\blacksquare$

To obtain the second-order LDT chance-constrained approximation, we relax the positive definiteness condition (5.11c), and combine (5.3), (5.8), (5.11a) and (5.11b) along with (3.11)



( $k = 2$ ) to obtain the following formulation:

(5.12)

$$\begin{aligned}
& \underset{u, \xi^*, \eta^*, \lambda, \tilde{\xi}, \tilde{\lambda}}{\text{minimize}} && J(u) \\
& \text{subject to} && u \in \mathcal{U}, \xi^* \in \mathbb{R}^n, \eta^* \in \mathbb{R}^n, \lambda \in \mathbb{R}_+, \\
& && \tilde{\xi}_i \in \mathbb{R}^n, \tilde{\lambda}_i \in \mathbb{R}_+, \quad i \in \{1, \dots, M\}, \\
& && \left\langle \nabla_{\xi} F(u, \xi^*), \tilde{\xi}_i - \xi^* \right\rangle + \frac{1}{2} \left\langle \tilde{\xi}_i - \xi^*, \nabla_{\xi}^2 F(u, \xi^*)(\tilde{\xi}_i - \xi^*) \right\rangle = 0, \quad i \in \{1, \dots, M\}, \\
& && \Sigma_i^{-1}(\tilde{\xi}_i - \mu_i) = \tilde{\lambda}_i \left( \nabla_{\xi} F(u, \xi^*) + \nabla_{\xi}^2 F(u, \xi^*)(\tilde{\xi}_i - \xi^*) \right), \quad i \in \{1, \dots, M\}, \\
& && \sum_{i=1}^M w_i \Phi \left( -\|\tilde{\xi}_i - \mu_i\|_{\Sigma_i^{-1}} \right) \det_{\perp \tilde{n}_i} \left( \mathbf{I}_n - \tilde{\lambda}_i \Sigma_i^{\frac{1}{2}} \nabla_{\xi}^2 F(u, \xi^*) \Sigma_i^{\frac{1}{2}} \right)^{-\frac{1}{2}} \leq \alpha, \\
& && F(u, \xi^*) = z, \\
& && \eta^* = \lambda \nabla_{\xi} F(u, \xi^*), \\
& && \xi^* = \frac{\sum_{i=1}^M w_i \exp \left( \eta^{*\top} \mu_i + \frac{1}{2} \eta^{*\top} \Sigma_i \eta^* \right) (\mu_i + \Sigma_i \eta^*)}{\sum_{i=1}^M w_i \exp \left( \eta^{*\top} \mu_i + \frac{1}{2} \eta^{*\top} \Sigma_i \eta^* \right)}.
\end{aligned}$$

As in the case of the LDT chance-constrained formulations for Gaussians, the function values and gradients of all functions involved in (5.10) and (5.12) can be computed using standard automatic differentiation codes, enabling their solution using off-the-shelf solvers.

**6. Applications.** We illustrate the computational performance of our first- and second-order LDT probability estimates and corresponding chance-constrained optimization problems using three examples from different applications. In subsection 6.1, we consider a portfolio selection problem with random stock prices; in subsection 6.2, we consider a structural short column design with uncertain material properties and loads; and finally in subsection 6.3, we consider optimal control of a steady-state advection-diffusion PDE. We implemented all of our formulations in Julia using JuMP [22], and solved all optimization problems using Ipopt [52] with default solver options. We used the forward mode automatic differentiation provided by JuMP for gradient computations. Since JuMP does not support Hessian computations of user-defined nonlinear constraint functions, which is needed for the second-order LDT chance-constrained formulations (for example, to encode the orthogonal determinant (4.4)), we used Ipopt's in-built quasi-newton method for the solution of these problems. It is important to note that our goal is to demonstrate the advantage and competitive performance of our method using a purely off-the-shelf implementation. Reported solution times correspond to wall clock times on a single thread of an Intel Xeon Gold CPU at 2.30GHz with 512 GB RAM.

**6.1. Portfolio management.** Consider an optimal asset allocation problem, where we wish to allocate our wealth among  $n$  stocks which will be held over  $T$  time periods. Let  $u_i \in \mathbb{R}^+$  denote the fraction of wealth invested in the  $i^{\text{th}}$  stock. Our goal is to determine the optimal allocation  $u = [u_1, \dots, u_n] \in \mathbb{R}_+^n$  of our wealth for each stock. Let  $v_i$  denote the value

of the  $i^{\text{th}}$  stock at the end of  $T$  time periods. We model the uncertainty in stock price  $v_i$  using a traditional log-normal model [27, 30],

$$(6.1) \quad v_i(\xi_i) = \exp\left(\left(\theta_i - \frac{\sigma_i^2}{2}\right)T + \sqrt{T}\xi_i\right), \quad i \in \{1, \dots, n\},$$

where  $\theta_i \in \mathbb{R}$  and  $\sigma_i \in \mathbb{R}$  represent the drift and infinitesimal standard deviation of the Lévy process for the  $i^{\text{th}}$  stock, respectively, and  $\xi_i \in \mathbb{R}$  is a zero-mean continuously compounded rate of return which represents the true uncertainty driver. Since choosing the right distribution for  $\xi$  is an unresolved question (e.g., see [30]), we model  $\xi$  using Gaussian (as in traditional models) as well as Gaussian mixture distributions. Our goal is to ensure, with very high probability, that the total return after  $T$  time periods,  $\sum_{i=1}^n v_i(\xi_i)u_i$ , is higher than some value  $z$ ; note that this is equivalent to constraining the probability  $\mathbb{P}(\sum_{i=1}^n v_i(\xi_i)u_i \leq z)$ .

**6.1.1. Probability estimation.** Before we consider the chance-constrained formulation, we empirically assess the accuracy of the first- and second-order LDT estimates of the rare event probability  $\mathbb{P}(\sum_{i=1}^n v_i(\xi_i)u_i \leq z)$  for a fixed stock allocation  $u$  and fixed value of  $z$ . Defining

$$(6.2) \quad F(u, \xi) = -\sum_{i=1}^n v_i(\xi_i)u_i,$$

this is equivalent to estimating  $\mathbb{P}(F(u, \xi) \geq -z)$ . Note that the function  $F$  is concave in  $\xi$  and linear in  $u$ , and in particular, it satisfies condition (C1) in Theorem 3.3.

We downloaded 1,000 days' worth of daily NASDAQ stock price data for 50 stocks to estimate the drift parameter  $\theta_i$ , standard deviation  $\sigma_i$  and the distribution for the random parameter  $\xi$  (Gaussian or Gaussian mixture). Specifically, we first compute the compounded rates of  $v_i$ ; *i.e.*, the logarithm of the ratio of stock prices in two consecutive days. For a Gaussian distribution, we then compute the empirical mean and covariance corresponding to these rates to obtain  $\theta_i - \sigma_i^2/2$  and covariance for  $\xi_i$  respectively. For a Gaussian mixture, we use the `GaussianMixtures.jl` package to estimate the parameters of a 2- and 3-component Gaussian mixture using the expectation-maximization algorithm. The allocation  $u$  is fixed and chosen randomly from a uniform distribution in  $[0, 1]^n$  subject to the constraint  $\sum_{i=1}^n u_i = 1$ . We fix  $T = 10$ , and compare the accuracy of estimating  $\mathbb{P}(F(u, \xi) \geq -z)$  for different fixed values of  $z$  using: (i) our first- ( $P_1$ ) and second-order ( $P_2$ ) LDT probability approximations, with (ii) a Monte Carlo sample average approximation  $P_N^{MC}(u, z)$  using  $N$  independent samples. We use  $N = 10^7$  samples to estimate the true probability as  $P^{true} = P_{10^7}^{MC}(u, z)$ , and define errors

$$\begin{aligned} \epsilon_k^{LDT}(u, z) &:= \log_{10} P_k(u, z, \xi^*) - \log_{10} P^{true}, \quad k \in \{1, 2\}, \\ \epsilon_N^{MC}(u, z) &:= \log_{10} P_N^{MC}(u, z) - \log_{10} P^{true}, \quad N \in \{10^4, 10^5\}. \end{aligned}$$

We note that positive errors denote an over-estimation of the true probability, whereas negative values imply under-estimation.

Table 1 shows that, as  $z$  becomes smaller, the event  $F(u, \xi) \geq -z$  becomes rarer and a sample average approximation with  $N = 10^4$  or  $10^5$  samples is insufficient to accurately estimate the true probability. For example we observe that for  $z = 0.80$ , the true probability

Table 1

The comparison of the first- ( $P_1$ ) and second-order ( $P_2$ ) LDT probability estimates with sample average approximations ( $P_N^{MC}$ ) for estimating  $\mathbb{P}(F(u, \xi) \geq -z)$  in the portfolio management problem. We drop the dependence on  $u$  and  $z$  for simplicity.

$z$	$P^{true}$	$P_1$	$P_2$	$P_{10^4}^{MC}$	$P_{10^5}^{MC}$	$\epsilon_1^{LDT}$	$\epsilon_2^{LDT}$	$\epsilon_{10^4}^{MC}$	$\epsilon_{10^5}^{MC}$
<i>Gaussian</i>									
0.80	$5.0 \cdot 10^{-7}$	$1.9 \cdot 10^{-6}$	$4.8 \cdot 10^{-7}$	0.0	0.0	0.57	-0.01	$-\infty$	$-\infty$
0.82	$5.4 \cdot 10^{-6}$	$2.2 \cdot 10^{-5}$	$6.6 \cdot 10^{-6}$	0.0	$1.0 \cdot 10^{-5}$	0.61	0.08	$-\infty$	0.27
0.84	$5.8 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$	$6.3 \cdot 10^{-5}$	0.0	$4.0 \cdot 10^{-5}$	0.51	0.04	$-\infty$	-0.16
0.86	$3.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	$4.4 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	$2.9 \cdot 10^{-4}$	0.46	0.05	0.10	-0.13
<i>Gaussian mixture (<math>M = 2</math>)</i>									
0.80	$6.6 \cdot 10^{-5}$	$2.2 \cdot 10^{-4}$	$7.5 \cdot 10^{-5}$	0.0	$6.0 \cdot 10^{-5}$	0.53	0.06	$-\infty$	-0.04
0.82	$1.8 \cdot 10^{-4}$	$5.7 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$3.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	0.49	0.06	0.22	0.04
0.84	$4.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$	$5.1 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$	0.45	0.04	-0.37	-0.13
0.86	$1.1 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	0.42	0.00	0.02	0.05
<i>Gaussian mixture (<math>M = 3</math>)</i>									
0.80	$5.7 \cdot 10^{-5}$	$2.5 \cdot 10^{-4}$	$5.8 \cdot 10^{-5}$	0.0	$4.0 \cdot 10^{-5}$	0.65	0.01	$-\infty$	-0.15
0.82	$2.1 \cdot 10^{-4}$	$7.9 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$	0.57	-0.02	-0.32	0.02
0.84	$6.5 \cdot 10^{-4}$	$2.1 \cdot 10^{-3}$	$6.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$	$7.9 \cdot 10^{-4}$	0.51	-0.03	0.19	0.09
0.86	$1.7 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	0.47	-0.03	0.13	0.04

is much smaller than  $10^{-4}$ ; in such cases, none of the  $N = 10^4$  samples record the rare event, on average, resulting in a probability estimate of 0. Stated differently,  $P_{10^4}^{MC}$  and  $P_{10^5}^{MC}$  fail to approximate probabilities lower than  $10^{-4}$  and  $10^{-5}$ , respectively, illustrating their high sample size requirement for accurate rare event estimation.

In contrast, our proposed LDT approximations are relatively less sensitive to the extremeness of the event (*i.e.*, to values of  $z$ ). In particular, since  $F$  is concave in  $\xi$ , [Proposition 3.1](#) ensures that the first-order approximation  $P_1(u, z, \xi^*)$  always overestimates the probability, and this is verified in [Table 1](#), since the errors  $\epsilon_1^{LDT}$  are positive in all cases. Nevertheless, the first-order estimates can be about three times the magnitude of the true probability, with log-errors of roughly 0.5, across different  $z$ . In contrast, the second-order estimates are highly accurate, with log-errors consistently less than  $10^{-1}$ , regardless of values of  $z$ , and they typically outperform the Monte Carlo estimates for both Gaussians and mixture distributions. We highlight that, for rare events, it is much more important to correctly estimate the order of magnitude of the probability rather than its exact value; for example, it matters less that we estimate an event probability of  $10^{-6}$  as  $2 \cdot 10^{-6}$ , as compared to estimating it as  $10^{-3}$ .

**6.1.2. Optimal stock allocation.** We now turn to determining the optimal stock allocation. We first define the value-at-risk (VaR) of a fixed allocation vector  $u$ :

$$(6.3) \quad \text{VaR}_{1-\alpha}(u) := \underset{z}{\operatorname{argmax}} \left\{ \mathbb{P} \left( \sum_{i=1}^n v_i(\xi_i) u_i \geq z \right) \geq 1 - \alpha \right\}.$$

Our goal is to maximize the VaR of the total return at level  $1 - \alpha$ , and consider  $\alpha \ll 1$  reflecting a conservative investment approach. This problem can be reformulated as the following variant

of the chance-constrained optimization problem (1.1) using the definition of  $F$  in (6.2):

$$(6.4) \quad \begin{aligned} & \underset{u \in \mathbb{R}_+^n, z \in \mathbb{R}}{\text{maximize}} && z \\ & \text{subject to} && \sum_{i=1}^n u_i = 1 \\ & && \mathbb{P}(F(u, \xi) \geq -z) \leq \alpha. \end{aligned}$$

For  $T = 10$  and  $\alpha = 10^{-4}$ , Table 2 compares the computational performance of our LDT chance-constrained formulations with the smooth sigmoidal sample average formulation (2.1) (with  $\nu = \tau = 1$ ) and the CVaR sample average formulation (2.2), each with  $N = 10^5$  samples. Here,  $u^*$  denotes the optimal allocation determined by each method,  $z^*$  denotes the corresponding objective value of (6.4), and  $\text{VaR}_{1-\alpha}(u^*)$  denotes the Value-at-Risk (6.3) of the corresponding allocation  $u^*$  which we would like to be maximized. As before, we estimate the true probability  $P^{\text{true}}(u^*, z^*) = P_{10^7}^{MC}(u^*, z^*)$  using  $N = 10^7$  Monte Carlo samples. Therefore, the column  $\log_{10} P^{\text{true}}(u^*, z^*) - \log_{10} \alpha$  shows the extent of feasibility of the chance constraint in each case, where negative values indicate satisfaction of the chance constraint whereas positive values indicate that the corresponding  $u^*$  and  $z^*$  fail to satisfy the chance constraint in (6.4). We also report solution times of the different methods.

**Table 2**

*Comparison of solution methods for the optimal stock allocation problem (6.4) with  $T = 10$  and  $\alpha = 10^{-4}$ .*

Method	$z^*$	$\text{VaR}_{1-\alpha}(u^*)$	$P^{\text{true}}(u^*, z^*)$	$\log_{10} P^{\text{true}}(u^*, z^*) - \log_{10} \alpha$	Time [sec]
<i>Gaussian</i>					
LDT ( $k = 1$ )	$9.53 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$7.7 \cdot 10^{-5}$	-0.11	0.5
LDT ( $k = 2$ )	$9.54 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$9.5 \cdot 10^{-5}$	-0.02	197.0
SAA ( $N = 10^5$ )	$9.02 \cdot 10^{-1}$	$9.53 \cdot 10^{-1}$	0.0	$-\infty$	638.7
CVaR ( $N = 10^5$ )	$9.53 \cdot 10^{-1}$	$9.52 \cdot 10^{-1}$	$1.5 \cdot 10^{-4}$	0.17	527.6
<i>Gaussian mixture (<math>M = 2</math>)</i>					
LDT ( $k = 1$ )	$9.15 \cdot 10^{-1}$	$9.20 \cdot 10^{-1}$	$5.4 \cdot 10^{-5}$	-0.26	11.3
LDT ( $k = 2$ )	$9.21 \cdot 10^{-1}$	$9.23 \cdot 10^{-1}$	$8.3 \cdot 10^{-5}$	-0.08	85.7
SAA ( $N = 10^5$ )	$9.20 \cdot 10^{-1}$	$9.18 \cdot 10^{-1}$	$1.2 \cdot 10^{-4}$	0.09	1,147.4
CVaR ( $N = 10^5$ )	$9.19 \cdot 10^{-1}$	$9.17 \cdot 10^{-1}$	$1.2 \cdot 10^{-4}$	0.07	420.4

From Table 2, we observe that our proposed second-order LDT formulation consistently attains the maximum VaR. Moreover, its optimal objective value  $z^*$  is quite close to the true VaR, indicating that the second-order probability estimate is also accurate. Notably, these improved objective values also come at a much lower solution time compared to the sampling-based methods. The first-order LDT formulation attains a slightly smaller VaR compared to the second-order formulation, but it is still better than the sampling-based methods. Moreover, the solutions  $(u^*, z^*)$  obtained by the sampling-based methods are not feasible in the original problem (6.4), since the corresponding  $\log_{10} P^{\text{true}}(u^*, z^*) - \log_{10} \alpha > 0$ , implying that we need to increase the number of samples  $N$  for these methods to obtain a feasible solution. However, given that their solution times are already quite excessive with  $N = 10^5$  samples;

further increasing  $N$  will only amplify these times even more. In contrast, the first- and second-order LDT formulations solve in  $\sim 10$  and  $\sim 100$  seconds respectively, with their solution times being largely unaffected by the choice of  $\alpha$ . In conclusion, the LDT formulations outperform the sampling-based methods both in terms of objective value and solution time. The choice between the first- and second-order formulations can be made on a case-by-case basis; whereas the former is more computationally efficient, the latter is more accurate, in general.

Figure 4 shows the optimal stocks allocations obtained by different methods in the case of the Gaussian mixture. Interestingly, all methods indicate that we should invest majority of the initial wealth in the stock “AGZD”, and never invest in some of the other stocks. However, the first-order LDT formulation determines a slightly different allocation for this stock compared with other methods, implying that we should use the second-order LDT allocation to obtain higher returns. Curiously, a closer look at Table 2 reveals that the VaR of the total return is less than the initial wealth ( $= 1$ ) in all cases, indicating that it may not be wise to invest at all over the short time period  $T = 10$  with an overly conservative risk threshold  $\alpha = 10^{-4}$ .

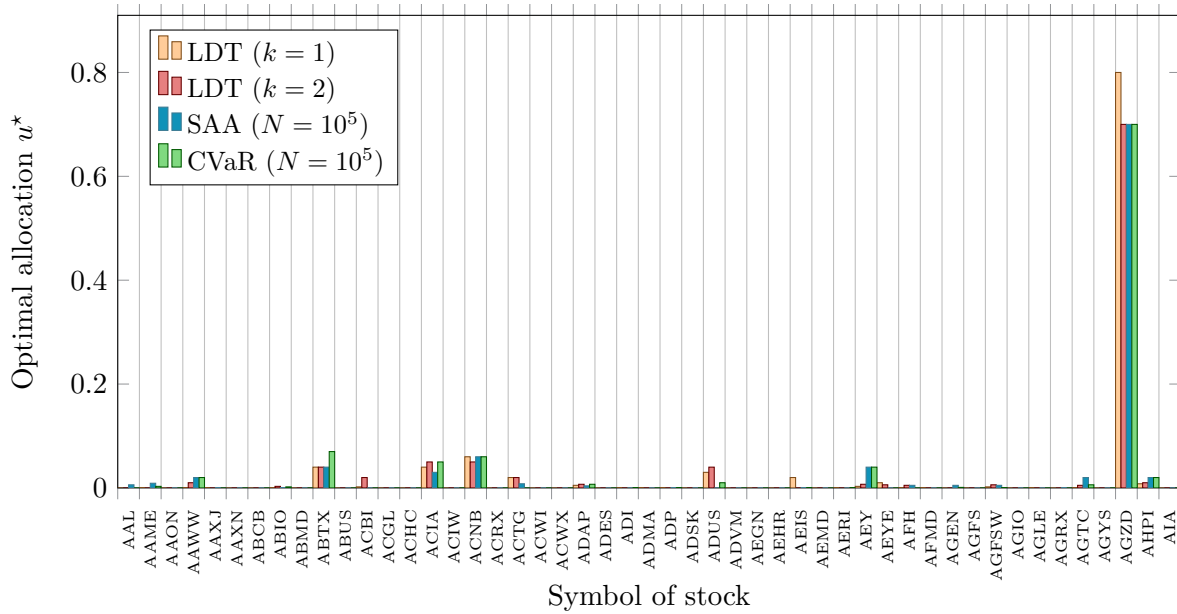


Figure 4. Optimal stock allocation determined by different methods with  $T = 10$  and  $\alpha = 10^{-4}$ , for the case where  $\xi$  follows a Gaussian mixture distribution with two components.

**6.2. Short column design.** In the last application, the function  $F$  was linear in the decisions  $u$ . We now consider a nonlinear model for the design of a short column [2, 12] with uncertain material properties, and subject to a chance constraint on material failure.

Consider a short column with a rectangular cross-section of width  $w$  and height  $h$ , that is subject to an uncertain axial load  $\xi_F$ , bending moment  $\xi_M$  and an uncertain yield stress  $\xi_Y$ . We model these as a three-dimensional random vector  $\xi = [\xi_F, \xi_M, \xi_Y]$ . Our goal is to determine  $w$  and  $h$  so that the area of the cross-section  $wh$  is minimized, and the design

should not experience material failure with very high probability of  $1 - \alpha$ , where  $\alpha \in (0, 1)$  is some given threshold. The parameter-to-observation map  $F$  can be derived from the elastic-plastic constitutive law, and it has a general nonlinear dependence on both the decisions  $u = (w, h) \in \mathbb{R}^2$  as well as the uncertainty  $\xi \in \mathbb{R}^3$ :

$$(6.5) \quad F(u, \xi) := \frac{4\xi_M}{wh^2 \exp(\xi_Y)} + \frac{\xi_F^2}{w^2h^2 \exp(2\xi_Y)}.$$

The optimization problem is

$$(6.6) \quad \begin{aligned} & \underset{u=(w,h) \in \mathbb{R}^2}{\text{minimize}} && wh \\ & \text{subject to} && L_w \leq w \leq U_w, \quad L_h \leq h \leq U_h, \\ & && \mathbb{P}(F(u, \xi) \geq 1) \leq \alpha, \end{aligned}$$

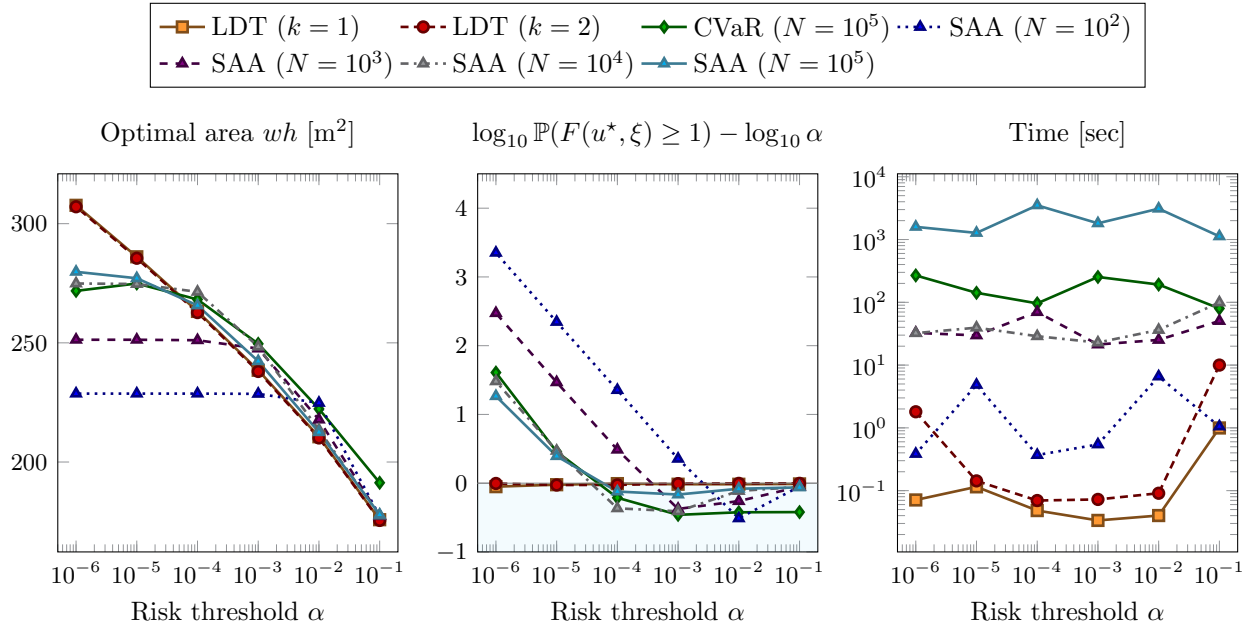
where  $L_w, L_h$  and  $U_w, U_h$  are given lower and upper bounds on  $w$  and  $h$ , respectively, and we set  $[L_w, U_w, L_h, U_h] = [5, 15, 15, 25]$ .

**6.2.1. Gaussian random parameter.** We first consider the random vector  $\xi \sim \mathcal{N}(\mu, \Sigma)$  with mean and covariance given as follows, adapted from the example in [12]:

$$(6.7) \quad \mu = [500, 2000, 1.604], \quad \Sigma = \begin{bmatrix} 10000 & 20000 & 0 \\ 20000 & 160000 & 0 \\ 0 & 0 & 0.00995 \end{bmatrix}.$$

The covariance matrix indicates that the correlation coefficient between the axial force  $\xi_F$  and bending moment  $\xi_M$  is 0.5, whereas the yield stress  $\exp(\xi_Y)$  is uncorrelated to them and follows a log-normal distribution with mean 5 and standard deviation 0.5.

In [Figure 5](#), we compare the performance of different methods for solving the short column design problem with Gaussian random parameter (6.7), as a function of the risk threshold  $\alpha$  varying from  $10^{-1}$  to  $10^{-6}$ . The left-most figure shows that our LDT chance constrained formulations provide similar optimal solutions, indicating that  $F$  is near-linear. Moreover, when  $\alpha$  becomes smaller, although the optimal areas determined by the sampling-based methods appear to be smaller than the ones determined by the LDT approaches (near-horizontal parts in the left-most figure), the middle figure shows that these solutions are, in fact, not feasible because they fail to satisfy the probability constraint (since their curves fall outside the negative part highlighted in blue). In fact, as far as the sampling-based methods go, the solutions determined using  $N$  samples can give reliable solutions only for  $\alpha \geq 10^{-(N-1)}$ . In contrast, the middle figure shows that the solutions determined by the LDT formulations always satisfy the chance constraint for all values of  $\alpha$ , and their corresponding optimal areas coincide with those of the sampling-based methods, whenever the latter are also feasible. In fact, they are able to find feasible designs even for extreme cases where sampling-based methods with  $N \leq 10^5$  samples fail. Finally, the right-most figure shows that the computational time increases severely for sampling-based methods as the number of samples  $N$  increases, since the problem size in (2.1) and (2.2) scales as  $O(N)$ . The middle figure suggests that, for very high reliability  $\alpha = 10^{-6}$ , we need to increase  $N$  to obtain a feasible solution, and thus the corresponding



**Figure 5.** Comparison of different methods for solving the short column design problem with Gaussian random parameter. The parameters for SAA are  $\nu = 200$  and  $\tau = 200$ . The left figure shows the optimal area (objective value) for each methods. The middle one shows the feasibility of the solutions, we expect the values  $\log_{10} \mathbb{P}(F(u^*, \xi) \geq 1) - \log_{10} \alpha$  to be negative (fall in the cyan area), meaning the solution  $u^*$  satisfies the chance constraint. The right figure shows the time costs for different methods.

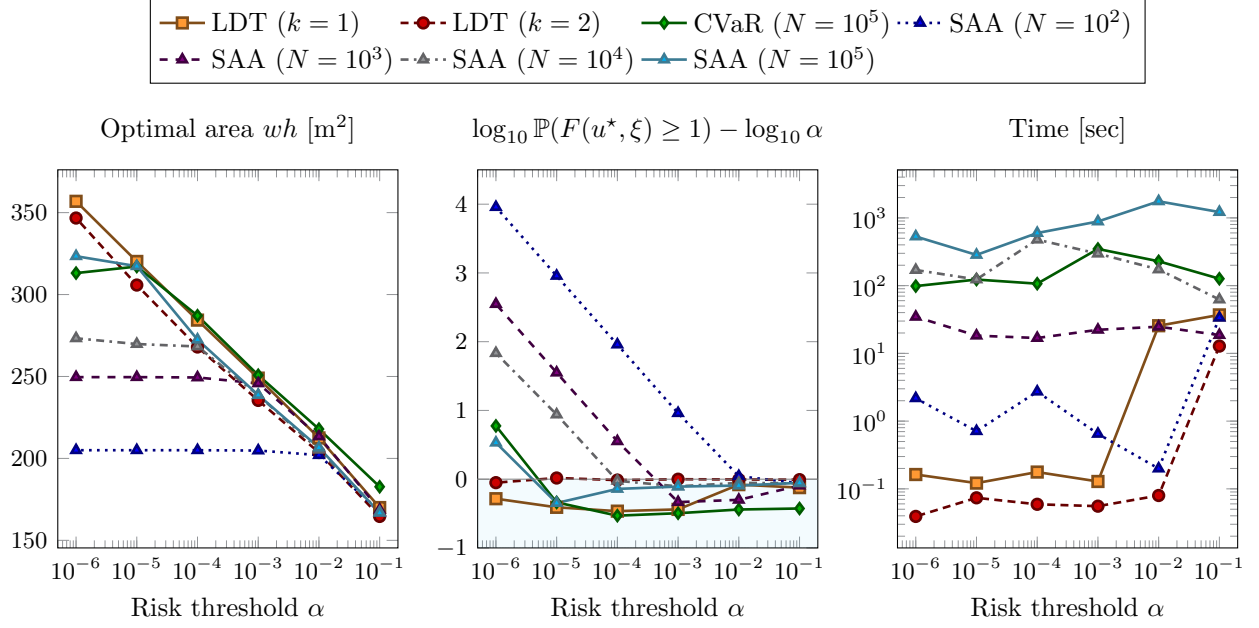
time is also expected to increase. The LDT approaches are much more efficient, with both the first- and second-order formulations solving in roughly 10% of the time compared to SAA with  $N = 10^2$  samples.

**6.2.2. Gaussian mixture random parameter.** We now consider  $\xi \sim 0.5\mathcal{N}(\mu_1, \Sigma_1) + 0.5\mathcal{N}(\mu_2, \Sigma_2)$  to follow a Gaussian mixture distribution with two components, where the mean  $\mu_1$  and covariance  $\Sigma_1$  of the first Gaussian component have the same values as in (6.7), whereas the mean and covariance of the second Gaussian component is given as follows:

$$(6.8) \quad \mu_2 = [100, 1000, 1.0849], \quad \Sigma_2 = \begin{bmatrix} 10000 & 20000 & 0 \\ 20000 & 160000 & 0 \\ 0 & 0 & 0.0274 \end{bmatrix}.$$

The results are summarized in Figure 6, and are largely similar to the case of Gaussians. The main differences are the different optimal areas determined by the two LDT approaches, with the second-order formulation yielding a slightly better objective value while still obtaining a feasible design. As before, both LDT approaches are significantly more efficient compared with sampling-based methods in terms of computational time and sensitivity to event extremeness.

**6.3. Optimal boundary control for steady-state advection-diffusion.** In the previous sections, we showed the performance of our proposed formulations approach for problems



**Figure 6.** Comparison of different methods for solving the short column design problem with Gaussian mixture random parameter with two components. The parameters for SAA are  $\nu = 200$  and  $\tau = 200$ .

that are nonlinear in the uncertainty, but still admit closed-form expressions of the parameter-to-observation map  $F$ . In this section, we consider a much more complicated PDE system, where we do not have closed-form expressions. In particular, these represent examples where it is infeasible to use sampling-based methods, since they would need to incorporate as many PDE solves as the number of samples. In contrast, we demonstrate that our LDT approaches continue to provide satisfactory results.

Consider a 2D steady-state advection-diffusion equation in the unit square  $\Omega = [0, 1] \times [0, 1]$ , and denote the boundary  $\Gamma_c = 0 \times [0, 1]$  and  $\Gamma_n = \partial\Omega \setminus \Gamma_c$ .

$$(6.9a) \quad -\nabla \cdot (\kappa(x, \xi) \nabla y(x)) + w(x) \cdot \nabla y(x) = f(x, \xi), \quad x \in \Omega,$$

$$(6.9b) \quad (\kappa(x, \xi) \nabla y(x)) \cdot n(x) = \frac{1}{\epsilon_0} (u(x) - y(x)), \text{ on } \Gamma_c,$$

$$(6.9c) \quad (\kappa(x, \xi) \nabla y(x)) \cdot n(x) = 0, \quad \text{on } \Gamma_n.$$

Here,  $y$  is the state representing the temperature at each location  $x$  and  $u$  is the Dirichlet boundary control on the boundary  $\Gamma_c$ . We enforce  $y(x) = u(x)$  on  $\Gamma_c$  through a Robin-type condition (6.9b) to avoid regularity limitations of Dirichlet boundary control. Also,  $\epsilon_0 = 10^{-4}$  is a penalty parameter,  $w = [1, 0]^\top$  is a fixed velocity field, and  $n(x)$  represents the outward-pointing normal at the boundary. The diffusion coefficient  $\kappa(x, \xi)$  and force  $f(x, \xi)$  are subject



to random uncertainty  $\xi$  with the following nonlinear dependence:

$$(6.10) \quad \begin{aligned} \kappa(x, \xi) &= \begin{cases} 0.8, & x \in [0, 1] \times [0.6, 1], \\ \exp(\xi_1), & x \in [0, 1] \times [0, 0.6] \end{cases} \\ f(x, \xi) &= 20 \exp\left(-\frac{(x_1 - \xi_2)^2}{0.1}\right) \exp\left(-\frac{(x_2 - 0.5)^2}{0.1}\right). \end{aligned}$$

Observe that, since the temperature  $y$  is a solution of (6.9), it is uncertain due to the randomness of  $\xi$ . In applications like hyperthermia treatment for cancer therapy [19], it is important to ensure that the (uncertain) temperature in some areas (*e.g.*, essential organs) do not exceed some critical threshold to protect organs from thermal damage. Motivated by these applications, our goal is to ensure that the average temperature in the sub-domain  $\Omega_0 = [0.4, 0.6] \times [0.4, 0.6]$  remains less than  $z$  with high probability  $1 - \alpha$ . This leads to the following definition of the parameter-to-observation  $F(u, \xi)$  as the average temperature in  $\Omega_0$  for a given realization of the random parameter  $\xi$  and control vector  $u$ :

$$(6.11) \quad F(u, \xi) := \frac{1}{|\Omega_0|} \int_{\Omega_0} y(x; u, \xi) dx,$$

where  $y(x; u, \xi)$  is the solution of PDE (6.9) at location  $x$  given  $\xi$  and  $u$ . Our goal is to control the temperature at the boundary  $\Gamma_c$  as close to 0 as possible, while ensuring that the average temperature in  $\Omega_0$  does not exceed  $z$  with high probability:

$$(6.12) \quad \begin{aligned} \underset{u}{\text{minimize}} \quad & \frac{1}{2} \int_{\Gamma_c} u^2(x) dx, \\ \text{subject to} \quad & \mathbb{P}(F(u, \xi) \geq z) \leq \alpha, \end{aligned}$$

where  $z$  is a given upper bound on the average temperature in  $\Omega_0$ , and  $\alpha \ll 1$ .

Unlike the other applications, note that we do not have a closed-form expression for  $F(u, \xi)$  with respect to  $u$  or  $\xi$ , and we can only access it through the solution of the PDE in (6.9). Nevertheless, it can still be characterized implicitly as a linear system upon discretization of the PDE. If we denote the dimension of the design  $u$  to be  $m$  and use a finite difference method to discretize the PDE, then the resulting linear system has size  $m \times m$ . Since the complexity of solving one linear problem is  $O(m^2)$ , resorting to a sampling-based methods will require us to model  $N$  realizations of  $F$  as constraints in the optimization problem, each involving a PDE-based linear solve. This would result in a very large-scale problem even for modest numbers of samples and control inputs, making sampling-based methods practically inefficient. This motivates alternate LDT approaches.

We use a finite-difference method to discretize the PDE (6.9) with  $m = 30$  elements. Therefore, the mesh is  $30 \times 30$  and the state  $y$  is 900-dimensional. To avoid issues of infeasibility and convergence to non-optimal stationary points, we successively warm-start the solution of the  $\alpha = 10^{-(k+1)}$  problem using the optimal values determined for the  $\alpha = 10^{-k}$  problem. In case of the first-order LDT approximation, we use a simple finite difference method to compute gradients, since the random parameter is low-dimensional  $n = 2$  (this also explains the faster solution times in Table 3). In case of the second-order LDT approximation, we use forward-mode automatic differentiation to compute gradients, since we need accurate gradients of the correction term (4.9) to ensure convergence.

Table 3

Optimal objective values and constraint feasibilities of the boundary control problem (6.12) with  $z = 1$  for different  $\alpha$ 's using our first- and second order LDT chance-constrained formulations.

$\alpha$	$\frac{1}{2} \int_{\Omega_0} (u^*)^2 dx$	$\mathbb{P}(F(u^*, \xi) \geq z)$	Time [sec]
<i>LDT (k = 1)</i>			
$10^{-1}$	3.71	$4.9 \cdot 10^{-2}$	14.8
$10^{-2}$	4.68	$5.4 \cdot 10^{-3}$	15.1
$10^{-3}$	5.49	$3.9 \cdot 10^{-4}$	14.4
$10^{-4}$	6.12	$4.5 \cdot 10^{-5}$	17.5
$10^{-5}$	6.63	$2.8 \cdot 10^{-6}$	21.8
$10^{-6}$	6.97	$3.0 \cdot 10^{-7}$	31.7
<i>LDT (k = 2)</i>			
$10^{-1}$	3.52	$8.9 \cdot 10^{-2}$	$3.5 \cdot 10^5$
$10^{-2}$	4.43	$9.5 \cdot 10^{-3}$	3,430.0
$10^{-3}$	5.22	$9.2 \cdot 10^{-4}$	3,899.0
$10^{-4}$	5.88	$9.5 \cdot 10^{-5}$	3,913.0
$10^{-5}$	6.41	$9.6 \cdot 10^{-6}$	4,137.5
$10^{-6}$	6.80	$9.4 \cdot 10^{-7}$	5,069.5

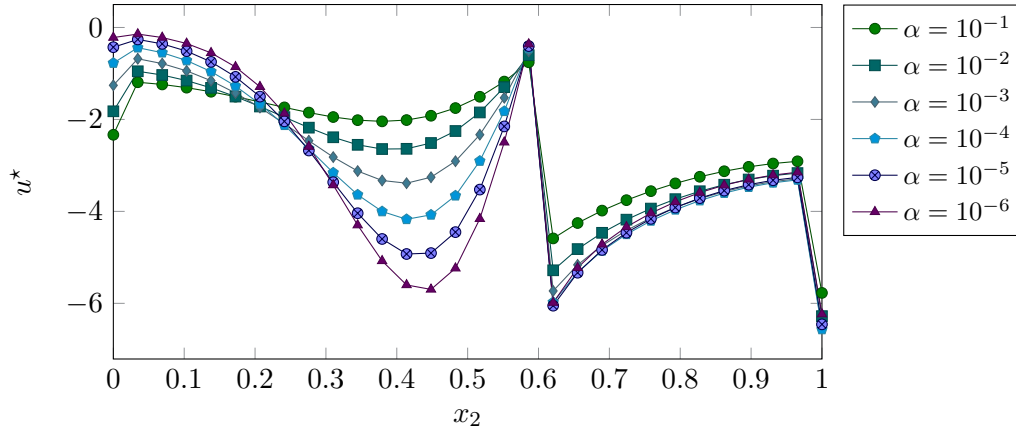


Figure 7. The optimal boundary condition  $u^*$  from problem (6.12) with  $z = 1$  for different  $\alpha$ 's using the second-order LDT approximation.

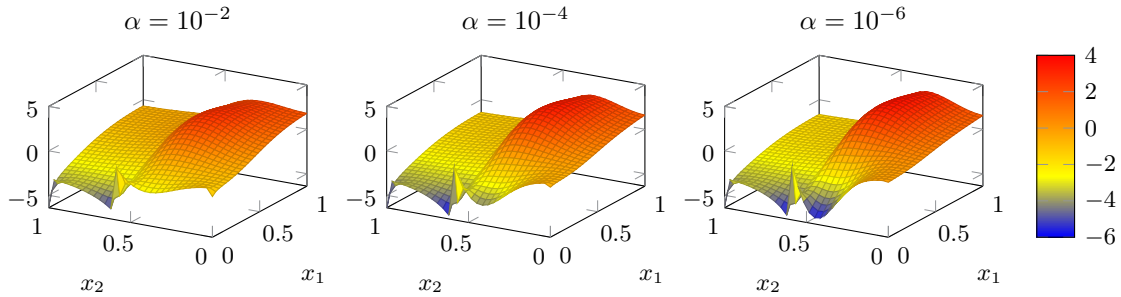


Figure 8. The temperature profile  $y$  under the optimal boundary condition  $u^*$  in Figure 7 and  $\xi^*$  for the minimization problem (6.12) with  $z = 1$  for different  $\alpha$ 's using the second-order LDT approximation.

Table 3 summarizes the computational performance. We observe that the optimal objective values become larger for more extreme events; *i.e.*, the optimal boundary control  $u^*$  has larger variations. The terms  $\mathbb{P}(F(u^*, \xi) \geq z)$  show chance constraint feasibilities, and the fact that they are all less than  $\alpha$  indicates that all obtained solutions are feasible. Comparing the two approaches, we observe that the optimal objective values of the second-order approximations are better compared with those of first-order approximations; however, this improvement comes at a significantly higher computational cost. We note that a large portion of the computational time is due to the inability of forward-mode automatic differentiation in JuMP to differentiate a sparse solve that required us to modify the PDE solve to a dense solve. The latter can be reduced by using a more efficient automatic differentiation implementation (or perhaps by using tailored codes for evaluating adjoints and gradients).

Finally, Figures 7 and 8 show the optimal boundary control  $u^*$  and corresponding temperature profile  $y$  under the optimal  $u$  and  $\xi^*$  obtained through (4.7). The optimal control inputs  $u$  have similar shapes but scale with  $\log \alpha$ ; moreover, they all have a jump at  $x_2 = 0.6$ , which comes from the discontinuity of the diffusion coefficient  $\kappa$  at  $x_2 = 0.6$ . The temperature profile  $y$  under the optimal boundary condition  $u^*$  and the LDT realization  $\xi^*$ , corresponds to the most representative profile among all feasible profiles in (6.12). Similar to the boundary conditions in Figure 7, the temperatures for smaller risk thresholds have larger variations.

**7. Discussions and conclusions.** Despite their rare occurrence, factors such as climate change and human population growth are causing a steady increase in the frequency of extreme high-impact events. The paucity of available data and inability to accurately model socioeconomic costs during these events motivates chance constraints as a viable modeling framework to mitigate the risk of extreme events.

In this paper, we combined ideas from large deviation theory and convex and bilevel optimization to propose a novel solution method for decision-making problems constrained by probabilities of rare events. Unlike classical approaches, our formulations are sampling-free, independent of event extremeness, and applicable to a broad class of nonlinear problems. Under certain regularity assumptions on the system constraints, they constitute safe conservative approximations or even asymptotically equivalent reformulations of the true problem. Furthermore, by utilizing local Taylor approximations of the system functions, we provide refined approximations over classical large deviation estimates that turn out to be empirically accurate even in non-asymptotic regimes. Although we only considered problems affected by Gaussian mixture uncertainties, it should be noted that our methods are also directly applicable to more general settings where the uncertainties can be transformed to a Gaussian or approximated well using a mixture of Gaussians, for which there are efficient off-the-shelf codes. Our proposed formulations complement this well since they can also be solved by off-the-shelf optimization solvers. Computational experiments on applications from diverse domains confirm the broad applicability, and improved efficiency and accuracy of our method over classical sampling-based methods in the rare event regime: the portfolio application illustrates its applicability to nonlinear high-dimensional uncertainties, the structural design application showcases its potential advantages over sampling-based methods in systems that are nonlinear in the decisions as well as uncertainties, whereas the optimal control application shows its ability to address complicated PDE-constrained systems.

We believe that our approach is but a small step towards tackling rare chance constraints, and much more work needs to be done. Various regularity assumptions on the structure of the constraint function (*e.g.*, smoothness, concavity etc.) as well as the uncertainty (*e.g.*, sufficiently smooth rate functions) need to be lifted and generalized. For example, our findings suggest that we cannot expect our method, and perhaps even large deviation theory, to work well when the system constraints are convex functions of the uncertainty. We would also like to extend our methodology to nonsmooth constraint functions, where the current Taylor-based probability estimates would be inapplicable, as well as to joint chance constrained problems. From a computational viewpoint, although our method has the advantage of using off-the-shelf codes, this also means that it can be inefficient at tackling more structured large-scale problems. For example, we need more efficient methods to compute derivatives of the probability estimates, and particularly of the second-order correction term. We believe this may also spur the development of new techniques in automatic differentiation.

**Acknowledgments.** We thank Eric Vanden-Eijnden and Georg Stadler from Courant Institute at New York University for helpful suggestions at various stages of this work.

#### REFERENCES

- [1] L. ADAM, M. BRANDA, H. HEITSCH, AND R. HENRION, *Solving joint chance constrained problems using regularization and benders' decomposition*, Annals of Operations Research, (2018), pp. 1–27.
- [2] Y. AOUES AND A. CHATEAUNEUF, *Benchmark study of numerical methods for reliability-based design optimization*, Structural and multidisciplinary optimization, 41 (2010), pp. 277–294.
- [3] S. ASMUSSEN AND P. W. GLYNN, *Stochastic simulation: algorithms and analysis*, vol. 57, Springer Science & Business Media, 2007.
- [4] J. BARRERA, T. HOMEM-DE MELLO, E. MORENO, B. K. PAGNONCELLI, AND G. CANESSA, *Chance-constrained problems and rare events: an importance sampling approach*, Mathematical Programming, 157 (2016), pp. 153–189.
- [5] J. BLANCHET, F. ZHANG, AND B. ZWART, *Optimal scenario generation for heavy-tailed chance constrained optimization*, arXiv preprint arXiv:2002.02149, (2020).
- [6] A. BOROVKOV AND B. ROGOZIN, *On the multi-dimensional central limit theorem*, Theory of Probability & Its Applications, 10 (1965), pp. 55–62.
- [7] M. BRONIATOWSKI AND A. FUCHS, *Tauberian theorems, Chernoff inequality, and the tail behavior of finite convolutions of distribution functions*, Advances in Mathematics, 116 (1995), pp. 12–33.
- [8] J. BUCKLEW, *Introduction to rare event simulation*, Springer Science & Business Media, 2013.
- [9] A. BUDHIRAJA, S. LU, Y. YU, AND Q. TRAN-DINH, *Minimization of a class of rare event probabilities and buffer probabilities of exceedance*, arXiv preprint arXiv:1902.07829, (2019).
- [10] G. CALAFIORE AND M. C. CAMPI, *Uncertain convex programs: randomized solutions and confidence levels*, Mathematical Programming, 102 (2005), pp. 25–46.
- [11] Y. CAO AND V. M. ZAVALA, *A sigmoidal approximation for chance-constrained nonlinear programs*, arXiv preprint arXiv:2004.02402, (2020).
- [12] A. CHAUDHURI, M. NORTON, AND B. KRAMER, *Risk-based design optimization via probability of failure, conditional Value-at-Risk, and buffered probability of failure*, in AIAA Scitech 2020 Forum, 2020, p. 2130.
- [13] F. E. CURTIS, A. WÄCHTER, AND V. M. ZAVALA, *A sequential algorithm for solving nonlinear optimization problems with chance constraints*, SIAM Journal on Optimization, 28 (2018), pp. 930–958.
- [14] G. DEMATTEIS, T. GRAFKE, M. ONORATO, AND E. VANDEN-EIJNDEN, *Experimental evidence of hydrodynamic instantons: The universal route to rogue waves*, Phys. Rev. X, 9 (2019), p. 041057, <https://doi.org/10.1103/PhysRevX.9.041057>.
- [15] G. DEMATTEIS, T. GRAFKE, AND E. VANDEN-EIJNDEN, *Rogue waves and large deviations in deep sea*,

- Proceedings of the National Academy of Sciences, 115 (2018), pp. 855–860.
- [16] G. DEMATTEIS, T. GRAFKE, AND E. VANDEN-ELIJNDEN, *Extreme event quantification in dynamical systems with random components*, SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 1029–1059.
  - [17] A. DEMBO AND O. ZEITOUNI, *Large Deviations Techniques and Applications*, Applications of mathematics, Springer, 1998.
  - [18] A. DER KIUREGHIAN, Y. ZHANG, AND C.-C. LI, *Inverse reliability problem*, Journal of engineering mechanics, 120 (1994), pp. 1154–1159.
  - [19] P. DEUFLHARD, A. SCHIELA, AND M. WEISER, *Mathematical cancer therapy planning in deep regional hyperthermia*, Acta Numerica, 21 (2012), pp. 307–378.
  - [20] X. DU AND W. CHEN, *A most probable point-based method for efficient uncertainty analysis*, Journal of Design and Manufacturing automation, 4 (2001), pp. 47–66.
  - [21] W. L. DUNN AND J. K. SHULTIS, *Exploring Monte Carlo methods*, Elsevier, 2011.
  - [22] I. DUNNING, J. HUCHETTE, AND M. LUBIN, *Jump: A modeling language for mathematical optimization*, SIAM Review, 59 (2017), pp. 295–320.
  - [23] A. GELETU, A. HOFFMANN, M. KLOPPEL, AND P. LI, *An inner-outer approximation approach to chance constrained optimization*, SIAM Journal on Optimization, 27 (2017), pp. 1834–1857.
  - [24] R. HENRION, *A Critical Note on Empirical (Sample Average, Monte Carlo) Approximation of Solutions to Chance Constrained Programs*, in System Modeling and Optimization, D. Hömberg and F. Tröltzsch, eds., Springer Berlin Heidelberg, 2013, pp. 25–37.
  - [25] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, Springer Berlin Heidelberg, 1993.
  - [26] 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.
  - [27] J. HULL ET AL., *Options, futures and other derivatives*, Upper Saddle River, NJ: Prentice Hall, 2009.
  - [28] M. ILTIS, *Sharp asymptotics of large deviations for general state-space markov-additive chains in  $\mathbb{R}^d$* , Statistics & probability letters, 47 (2000), pp. 365–380.
  - [29] H. KAHN AND A. W. MARSHALL, *Methods of reducing sample size in Monte Carlo computations*, Journal of the Operations Research Society of America, 1 (1953), pp. 263–278.
  - [30] B. KAWAS AND A. THIELE, *A log-robust optimization approach to portfolio management*, OR Spectrum, 33 (2011), pp. 207–233.
  - [31] C. M. LAGOA, X. LI, AND M. SZNAIER, *Probabilistically constrained linear programs and risk-adjusted controller design*, SIAM Journal on Optimization, 15 (2005), pp. 938–951.
  - [32] J. LIM, B. LEE, AND I. LEE, *Second-order reliability method-based inverse reliability analysis using hessian update for accurate and efficient reliability-based design optimization*, International Journal for Numerical Methods in Engineering, 100 (2014), pp. 773–792.
  - [33] J. LUEDTKE AND S. AHMED, *A sample approximation approach for optimization with probabilistic constraints*, SIAM Journal on Optimization, 19 (2008), pp. 674–699.
  - [34] 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.
  - [35] J. J. MORÉ, *Generalizations of the trust region problem*, Optimization methods and Software, 2 (1993), pp. 189–209.
  - [36] A. NEMIROVSKI AND A. SHAPIRO, *Scenario approximations of chance constraints*, in Probabilistic and randomized methods for design under uncertainty, G. Calafiore and F. Dabbene, eds., Springer, 2006, pp. 3–47.
  - [37] A. NEMIROVSKI AND A. SHAPIRO, *Convex approximations of chance constrained programs*, SIAM Journal on Optimization, 17 (2007), pp. 969–996.
  - [38] P. NEY, *Dominating points and the asymptotics of large deviations for random walk on  $\mathbb{R}^d$* , The Annals of Probability, 11 (1983), pp. 158–167.
  - [39] B. K. PAGNONCELLI, S. AHMED, AND A. SHAPIRO, *Sample average approximation method for chance constrained programming: theory and applications*, Journal of optimization theory and applications, 142 (2009), pp. 399–416.
  - [40] A. PEÑA-ORDIERES, J. R. LUEDTKE, AND A. WÄCHTER, *Solving chance-constrained problems via a smooth sample-based nonlinear approximation*, SIAM Journal on Optimization, 30 (2020), pp. 2221–

- 2250.
- [41] G. PISTONE AND H. P. WYNN, *Finitely generated cumulants*, *Statistica Sinica*, 9 (1999), pp. 1029–1052.
  - [42] A. PRÉKOPA, *Probabilistic programming*, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., vol. 10 of *Handbooks in Operations Research and Management Science*, Elsevier, 2003, pp. 267–351.
  - [43] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer Berlin Heidelberg, 2009.
  - [44] R. Y. RUBINSTEIN, *Optimization of computer simulation models with rare events*, *European Journal of Operational Research*, 99 (1997), pp. 89–112.
  - [45] G. I. SCHUËLLER AND R. STIX, *A critical appraisal of methods to determine failure probabilities*, *Structural Safety*, 4 (1987), pp. 293–309.
  - [46] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on stochastic programming: modeling and theory*, SIAM, 2014.
  - [47] S. TONG, E. VANDEN-EIJNDEN, AND G. STADLER, *Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis*, arXiv preprint arXiv:2007.13930, (2020).
  - [48] H. TOUCHETTE, *A basic introduction to large deviations: Theory, applications, simulations*, arXiv preprint arXiv:1106.4146, (2011).
  - [49] M. A. VALDEBENITO AND G. I. SCHUËLLER, *A survey on approaches for reliability-based optimization*, *Structural and Multidisciplinary Optimization*, 42 (2010), pp. 645–663.
  - [50] 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.
  - [51] S. S. VARADHAN, *Large deviations and applications*, vol. 46, SIAM, 1984.
  - [52] A. WÄCHTER AND L. T. BIEGLER, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, *Mathematical Programming*, 106 (2006), pp. 25–57.