# On Solving Chance-Constrained Models with Gaussian Mixture Distribution

Shibshankar Dey
Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL
shibshankardey2025@u.northwestern.edu

Sanjay Mehrotra
Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL
mehrotra@northwestern.edu

Anirudh Subramanyam
Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, PA
subramanyam@psu.edu

**Abstract.** We study linear chance-constrained problems where the coefficients follow a Gaussian mixture distribution. We provide mixed-binary quadratic programs that give inner and outer approximations of the chance constraint based on piecewise linear approximations of the standard normal cumulative density function. We show that $O\left(\sqrt{\ln(1/\tau)/\tau}\right)$ pieces are sufficient to attain $\tau$-accuracy in the chance constraint. We also show that any desired optimality gap can be achieved under a constraint qualification condition by controlling the approximation accuracy. Extensive computations using a commercial solver show that problems with up to one thousand random coefficients specified with up to fifteen Gaussian mixture components, generated under diverse settings, can be solved to near optimality within 18 hours, while satisfying chance constraint satisfaction probabilities of up to $0.999$. The solution times are significantly lower for problems with fewer random coefficients and mixture terms. For example, problems with one hundred random coefficients, ten mixture terms, and a constraint satisfaction probability of $0.999$ can be solved in a minute or less. Sample average approximations fail to provide meaningful solutions even for the smaller problems.

**Key words:** Gaussian mixture model (GMM), piecewise linear (PWL), approximation accuracy, continuity, differentiability, finite breakpoints, standard normal distribution.

## 1. Introduction

Chance-constrained models [8, 33] ensure that constraints involving random parameters are satisfied with a desired probability. These models are known to be hard to solve, even in the presence

of a single chance constraint, due to their non-convex feasible set [28,35]. We study linear chance-constrained optimization problems of the form:

$$Z^\star(\theta) := \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x} \tag{1a}$$

$$\text{s.t. } \mathbb{P}\left[\boldsymbol{\xi}^\top \boldsymbol{x} \le b\right] \ge \theta. \tag{1b}$$

Here, $\boldsymbol{x} \in \mathbb{R}^n$ is the decision vector and $\mathcal{X}$ represents its feasible region specified by deterministic constraints. We assume that $\mathcal{X}$ is compact. We suppose that $\boldsymbol{\xi}$ is a random vector taking realizations in $\mathbb{R}^n$. The constant $\theta \in (0,1)$ specifies the desired probability of constraint satisfaction.

We focus on the solvability of (1) when the probability distribution of $\boldsymbol{\xi}$ is described by a $K$-component Gaussian mixture model (GMM). Specifically, the density function of $\boldsymbol{\xi}$ is $\sum_{k=1}^K w_k \mathcal{N}(\cdot \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, where $\mathcal{N}(\cdot \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the density function of a multivariate normal vector with mean $\boldsymbol{\mu}_k \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{R}^{n\times n}$. The vector $\boldsymbol{w} \in \mathbb{R}^K_+$ specifies the mixture weights, satisfying $\mathbf{1}^\top \boldsymbol{w} = 1$, where $\mathbf{1}$ denotes the vector of ones. We assume that $\boldsymbol{\Sigma}_k$ is positive definite for all $k \in [K]$, where we define $[K] \coloneqq \{1, 2, \ldots, K\}$. Our motivation for specifying the chance constraint using a GMM stems from the fact that any probability density function can be approximated using a Gaussian mixture distribution to arbitrary accuracy [47, 51]. This allows chance constraint models of problems where the uncertain data is multi-modal. We list several recent applications of Gaussian mixture models to optimization problems arising in various domains in Section 2.

**Remark.** *The chance constraint* (1b) *is an example of left-hand side uncertainty. Although we assume the right-hand side coefficient* $b$ *to be a given constant, it may also follow a (univariate) Gaussian mixture distribution. This case can be reduced to* (1) *by augmenting the decision vector* $\boldsymbol{x}$ *to* $(\boldsymbol{x}, x')$*, adding the constraint* $x' = 1$ *to* $\mathcal{X}$*, replacing the objective coefficients* $\mathbf{c}$ *with* $(\mathbf{c}, 0)$*, and updating the chance constraint to* $\mathbb{P}[\boldsymbol{\xi}^\top \boldsymbol{x} - bx' \le 0] \ge \theta$*. Similarly, the form of problem* (1) *is also general enough to accommodate constraints of the form,* $\boldsymbol{\xi}^\top (\boldsymbol{H}\boldsymbol{x} + \boldsymbol{h}) + \boldsymbol{a}^\top \boldsymbol{x} \le b$*, where* $\boldsymbol{H} \in \mathbb{R}^{m\times m'}$*,* $\boldsymbol{h} \in \mathbb{R}^m$*, and* $\boldsymbol{a} \in \mathbb{R}^{m'}$*. This can be captured by setting* $n = m + m'$*, augmenting the decision vector* $\boldsymbol{x}$ *to* $(\boldsymbol{x}, \boldsymbol{x}')$*, adding the constraints* $\boldsymbol{x}' = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{h}$ *to* $\mathcal{X}$*, and replacing the objective coefficients* $\mathbf{c}$ *with* $(\mathbf{c}, \mathbf{0})$*.*

**Assumption 1.** *The covariance matrix* $\Sigma_k$ *is positive definite for all* $k \in [K]$*; specifically,* $\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x} > 0$ *for all* $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$*.*

Define $p(\boldsymbol{x}) \coloneqq \mathbb{P}\left[\boldsymbol{\xi}^\top \boldsymbol{x} \le b\right]$ to be the probability function. Then, Assumption 1 implies that we can equivalently write $p(\boldsymbol{x})$ as

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathbb{P}\left[\boldsymbol{\xi}_k^\top \boldsymbol{x} \le b \,\middle|\, \boldsymbol{\xi}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right] \tag{2a}$$

$$= \sum_{k=1}^{K} w_k \mathbb{P}\left[\boldsymbol{\xi}_k^\top \boldsymbol{x} \le b \,\middle|\, \boldsymbol{\xi}_k^\top \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_k^\top \boldsymbol{x}, \boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x})\right] \tag{2b}$$

$$= \sum_{k=1}^{K} w_k p_k(\boldsymbol{x}), \tag{2c}$$

$$\text{where } p_k(\boldsymbol{x}) := \begin{cases} \mathbb{1}_{\ge 0}(b), & \text{if } \boldsymbol{x} = \boldsymbol{0}, \\ \Phi\left(\dfrac{b - \boldsymbol{\mu}_k^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}}\right) & \text{otherwise.} \end{cases} \tag{2d}$$

Here, $\Phi$ denotes the standard normal cumulative distribution function (CDF) and $\mathbb{1}_{\ge 0}$ denotes the indicator function of the nonnegative reals. We will see in Proposition 1 and Lemma 2 that problem (1) is infeasible if $b < 0$. The chance constraint (1b) can now be equivalently expressed as $p(\boldsymbol{x}) \ge \theta$. We denote the feasible region of problem (1) as:

$$P(\theta) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{H}\boldsymbol{x} = \boldsymbol{h}, \boldsymbol{A}\boldsymbol{x} \ge \boldsymbol{d},\ p(\boldsymbol{x}) \ge \theta\}, \tag{3}$$

whenever $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n |\ \boldsymbol{H}\boldsymbol{x} = \boldsymbol{h}, \boldsymbol{A}\boldsymbol{x} \ge \boldsymbol{d}\}$ is specified explicitly using linear equality and inequality constraints.

**Remark.** *Assumption 1 and the resulting reformulation* (2d) *of the chance constraint* (1b) *is without loss of generality. Indeed, suppose that* $\boldsymbol{\Sigma}_k$ *has rank* $r_k < n$*. Then, it admits an eigenvalue decomposition; specifically, there exists* $\boldsymbol{Q}_k \in \mathbb{R}^{n \times n}$ *with columns formed from the eigenvectors of* $\boldsymbol{\Sigma}_k$ *such that* $\boldsymbol{D}_k = \boldsymbol{Q}_k^\top \boldsymbol{\Sigma}_k \boldsymbol{Q}_k = \begin{pmatrix} \boldsymbol{D}_k' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$*, where* $\boldsymbol{D}_k' \in \mathbb{R}^{r_k \times r_k}$ *is a diagonal matrix. Now, augment the decision vector* $\boldsymbol{x}$ *to* $(\boldsymbol{x}, \boldsymbol{y}_k)$ *and add the constraints,* $\boldsymbol{y}_k = \boldsymbol{Q}_k \boldsymbol{x}$*, to* $\mathcal{X}$*. Also, partition* $\boldsymbol{y}_k = (\boldsymbol{y}_k', \boldsymbol{y}_k'')$ *such that* $\boldsymbol{y}_k' \in \mathbb{R}^{r_k}$*. Then, it can be readily verified that the associated* $p_k(\boldsymbol{x})$ *in* (2d) *can be written as* $\Phi\left(\frac{b - {\mu_k}^\top \boldsymbol{x}}{\sqrt{{\boldsymbol{y}_k'}^\top \boldsymbol{D}_k' \boldsymbol{y}_k'}}\right)$ *with* ${\boldsymbol{y}_k'}^\top \boldsymbol{D}_k' \boldsymbol{y}_k' > 0$ *for all* $\boldsymbol{y}_k' \ne \boldsymbol{0}$*.*

The chance constraint, $p(\boldsymbol{x}) \ge \theta$, can be nonconvex even in simple cases. To visualize this, consider a two-dimensional problem with $\mathcal{X} = [-15, 15]^2$, $K = 2$, eigenvalues of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are $\boldsymbol{D}_1 = \mathrm{diag}(1.15, 0.65)$ and $\boldsymbol{D}_2 = \mathrm{diag}(1.47, 0.33)$, respectively, with eigenvectors $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ shown below:

$$w_1 = w_2 = 0.5,\ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \begin{bmatrix} 0.875 \\ 1.784 \end{bmatrix},\ \boldsymbol{Q}_1 = \begin{bmatrix} 1 & -0.08 \\ 0.08 & 1 \end{bmatrix},\ \boldsymbol{Q}_2 = \begin{bmatrix} 1 & -0.02 \\ 0.02 & 1 \end{bmatrix}.$$

The chance-constraint right-hand side coefficient $b = 6.7$. The left-hand side of Figure 1 plots $p_1(\boldsymbol{x})$ and $p_2(\boldsymbol{x})$ whereas the right-hand side plots $p(\boldsymbol{x})$. Observe that $p(\boldsymbol{x})$ is nonconvex over $\mathcal{X}$.

The non-convexity can also be seen from Figure 2, which shows the contour plots of $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$ on the left and of $p(\boldsymbol{x})$ on the right.

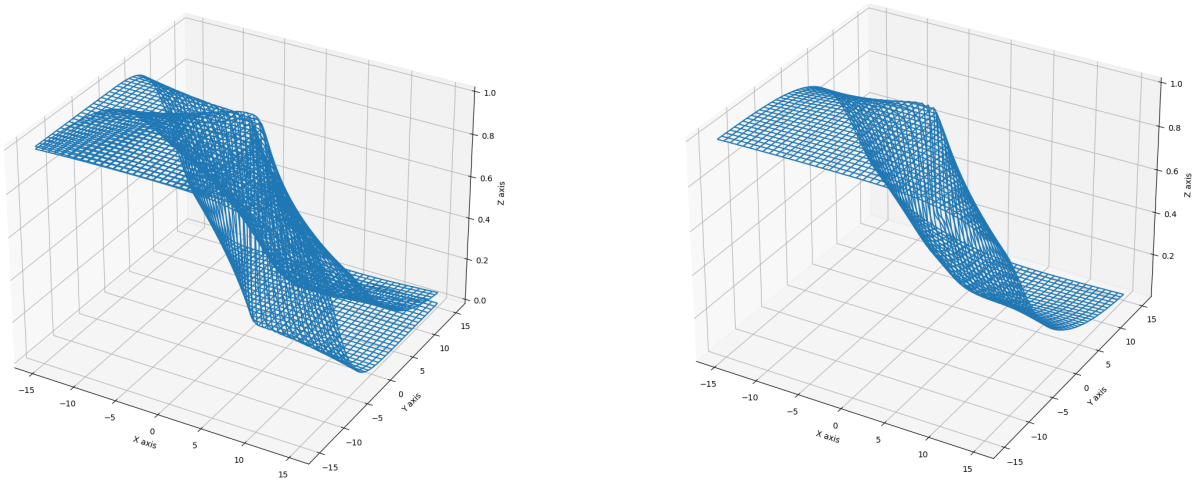


**Figure 1** **Plots of $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$ (left) and $p(\boldsymbol{x})$ (right) over $\boldsymbol{x} \in \mathcal{X}$.**

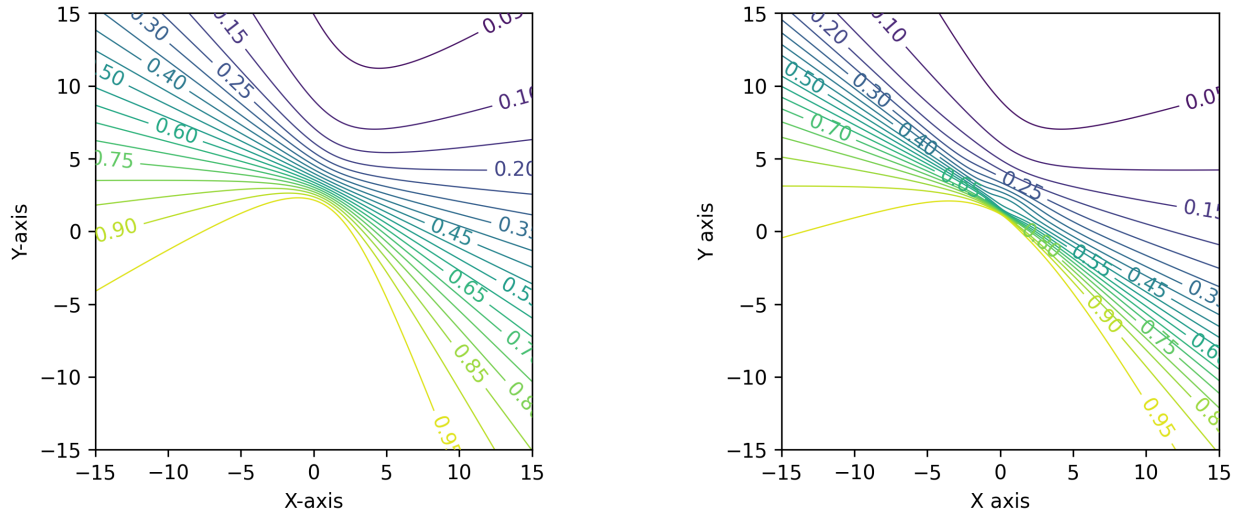


**Figure 2** **Contour plots of $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$ (left) and $p(\boldsymbol{x})$ (right) over $\boldsymbol{x} \in \mathcal{X}$.**

The shape of the distribution of $\boldsymbol{\xi}^\top \boldsymbol{x}$ can also change significantly as a function of $\boldsymbol{x}$. This is illustrated in Figure 3, where each plot corresponds to different solutions sampled from the feasible region of a problem in which the random vector follows a GMM with $K = 5$ components [1]. In

[1] The data for this experiment is generated as follows. We set $\theta = 0.9$, $n = 100$, $\mathcal{X} = [-20, 20]^{100}$, and each entry of $\boldsymbol{c} \in \mathbb{R}^{100}$ is randomly sampled from $[-1, 1]$. The chance constraint right-hand side coefficient $b = 1303.223$ is set to be the average of $\boldsymbol{x}^\top \boldsymbol{\mu}_k + 4\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}$ across all $k$ and one thousand randomly sampled decisions $\boldsymbol{x} \in \mathcal{X}$. We set $K = 5$, $\boldsymbol{w} = (0.30, 0.10, 0.20, 0.05, 0.35)$, and each entry of $\boldsymbol{\mu}_k$ is sampled from $[\underline{\mu}_k, \bar{\mu}_k]$, which are $[10^2, 10^3]$, $[5 \cdot 10^2, 1.2 \cdot 10^3]$, $[-20 \cdot 10^3, -8 \cdot 10^3]$, $[1.4 \cdot 10^3, 2.2 \cdot 10^3]$, $[-3 \cdot 10^3, -2 \cdot 10^3]$, for $k = 1, 2, \ldots, 5$, respectively. The eigenvalues; i.e., diagonal entries of $\boldsymbol{D}_k$, are uniformly sampled from $[0, \bar{\nu}_k]$, where $\bar{\nu}_1, \bar{\nu}_2, \ldots, \bar{\nu}_5$ are evenly spaced points in $(0, 5]$. The eigenvector matrices $\boldsymbol{Q}_k$ are randomly generated orthogonal matrices to ensure that $\boldsymbol{\Sigma}_k = \boldsymbol{Q}_k^\top \boldsymbol{D}_k \boldsymbol{Q}_k$ is positive definite.

each plot, we normalize the $x$-axis and show the individual component-specific normal densities of $\boldsymbol{\xi}_k^\top \boldsymbol{x}$, see equation (2b), for $k = 1, 2, \ldots, 5$ using blue, green, cyan, magenta, and orange colored curves, respectively. The density of $\boldsymbol{\xi}^\top \boldsymbol{x}$ is shown in black color.

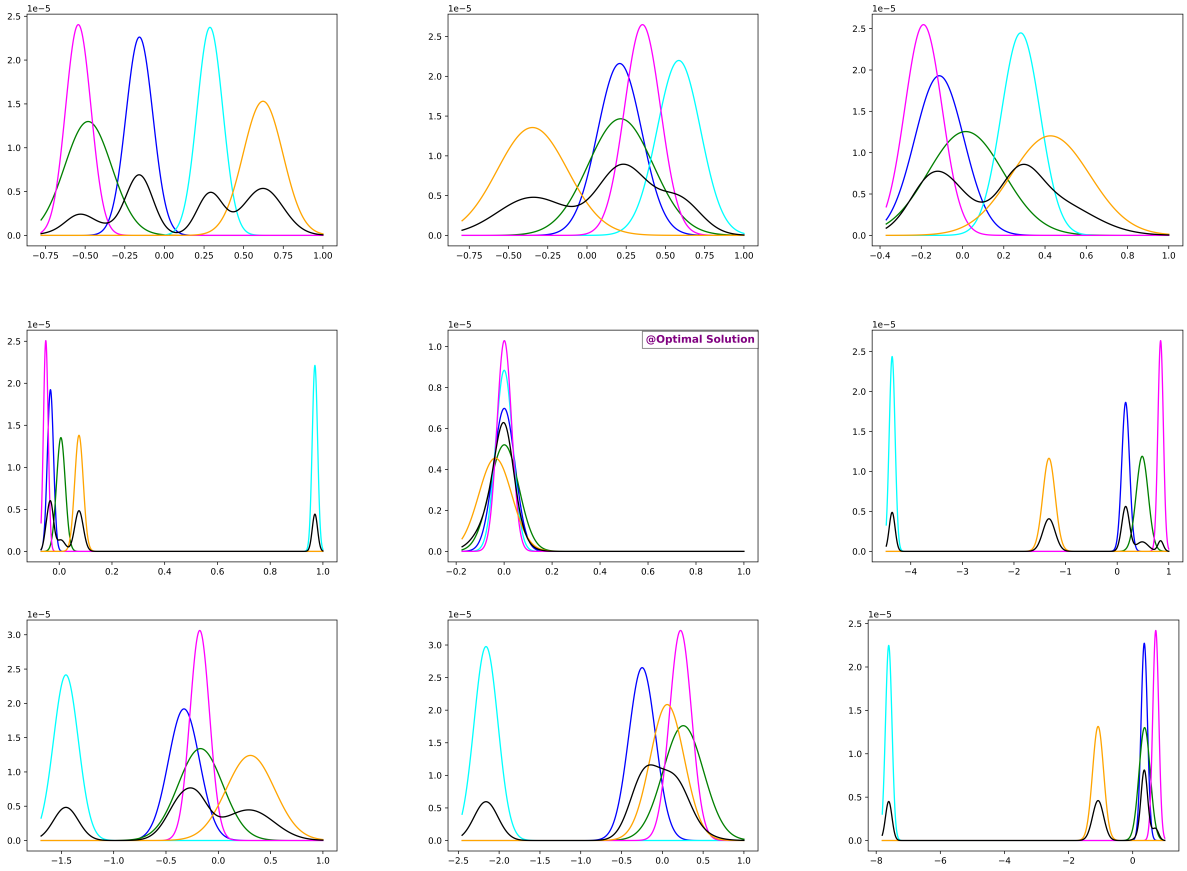


Figure 3: **Probability density of $\boldsymbol{\xi}_k^\top \boldsymbol{x}$ for $k = 1, 2, \ldots, 5$ in blue, green, cyan, magenta, and orange colors, respectively, and of $\boldsymbol{\xi}^\top \boldsymbol{x}$ in black color. Each plot uses a distinct $\boldsymbol{x}$ sampled from the feasible region.**

We observe from Figure 3 that for certain decisions (e.g., middle row and bottom right), the means of most component-specific distributions tend to cluster around a particular value. In contrast, for certain others (e.g., top left and right), the means are spread further apart. Additionally, certain component-specific distributions can completely shift from the negative end of the $x$-axis (e.g., blue, magenta, green in the middle left plot) to the positive end (e.g., see bottom and middle right plots). Moreover, the mass of $\boldsymbol{\xi}^\top \boldsymbol{x}$ can vary significantly (e.g., middle and bottom right) with the decision. The plots highlight the difficulty in estimating or predicting the location and shape of the overall mixture distribution, let alone the individual component-specific distributions, at the optimal solution $\boldsymbol{x}^*(\theta)$.

### 1.1. Contributions

The key technical challenge lies in handling the weighted sum of the cumulative distribution functions in reformulation (2d), each of which is composed with a nonlinear fractional function of $\boldsymbol{x}$. When $\boldsymbol{\xi}$ is multivariate normal (i.e., when $K = 1$), it is well known that the chance constraint can be reformulated as a tractable second-order conic constraint [46]. Unfortunately, this is no longer the case when $\boldsymbol{\xi}$ follows a GMM with $K > 1$ components.

To address this gap, we develop and analyze approximations of problem (1). This is achieved by exploiting the special convex-concave structure of the univariate standard normal CDF $\Phi$ and replacing it with an efficient piecewise linear (PWL) approximation. We provide mixed-integer optimization formulations that can be iteratively refined to either inner- or outer-approximate the feasible set, namely $P(\theta)$, to any arbitrary user-specified accuracy. We refer to these formulations as 'PWL-I' and 'PWL-O', respectively. Unlike previous studies (reviewed in the next section), we do not impose any restrictive structural assumptions on the GMM, such as dependence structures among the component covariance matrices, or bounds on the total number of components. Under suitable regularity assumptions, we also show that our formulations provide solutions up to any desired optimality tolerance. All of the proposed formulations can be solved using standard mixed-integer quadratic programming solvers.

We perform an extensive computational study on the PWL-I, and PWL-O formulations, across five thousand synthetically generated problem instances with 100, 500, and 1,000 variables and up to fifteen mixture components for chance constraint satisfaction levels $\theta$ set at $0.95$, $0.99$, and $0.999$. If the complement of the chance constraint in (1b) models an unsafe system condition that causes failure, then $\theta = 0.999$ can be interpreted as a minimum reliability requirement of $99.9\%$, or equivalently, as a maximum allowed failure probability of $0.001$, thus modeling a decision-dependent rare event [3, 49]. Our experimental findings show that the majority of PWL-I and PWL-O approximation models are solved to optimality within a time limit of 18 hours, while also attaining the desired probabilities of constraint satisfaction. Most instances consisting of ten and five mixture components are solved within 12 and 4 hours, respectively. Finally, we also compare the computational performance of the proposed formulations with classical sample average approximations (SAA) [24, 28, 36]. We find that the SAA models take very large computational times and are unable to produce solutions with the desired optimality or feasibility tolerances.

### 1.2. Notation

All vectors and matrices as well as vector- and matrix-valued functions are typeset in boldface, whereas all scalars and scalar-valued functions are typeset in normal font. The cumulative distribution and the probability density functions of the standard normal random variable are denoted by $\Phi$ and $\phi$, respectively. The derivative of $\phi$ is denoted by $\phi'$. We use $\mathbb{R}_+$ to denote the set of non-negative real numbers. For any non-negative integer $N$, we use $[N]$ to denote the set $\{1, 2, \ldots, N\}$ and $[N]_0$ to denote $\{0, 1, 2, \ldots, N\}$. Unless stated otherwise, $\|\cdot\|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices; we sometimes also use $\|\cdot\|_F$ to denote the latter. We use $\overset{\mathcal{U}}{\sim}$ to indicate sampling from a uniform distribution.

## 2. Related Literature

For an overview of the theory and algorithms for chance-constrained optimization problems, we refer the reader to [12, 43]. Approaches based on distributionally robust optimization are reviewed in [26, 37, 57]. For the models we study in this paper, the decision vector $\boldsymbol{x}$ and the random parameters $\boldsymbol{\xi}$ are not separable. In the case of separable chance-constrained models, one can move the uncertainty entirely to the right-hand side. In this case, existing approaches are based on integer programming ideas [25, 29, 48] and so-called $p$-efficient points to approximate the feasible set [13, 42].

In the general non-separable case, existing methodologies for solving chance-constrained optimization problems can be categorized into sample-based and sample-free analytical approaches. The sample average approximation (SAA) method [4, 7, 9, 28, 34, 36], belonging to the first category, is a general strategy to approximate the probability in (1b). The method is based on drawing several, say $N$, i.i.d. Monte Carlo samples, $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots, \boldsymbol{\xi}^N$, of the random parameters and then approximating the probabilistic constraint function using the empirical estimate based on these samples,

$$p(\boldsymbol{x}) = \mathbb{E}[\mathbb{1}_{\geq 0}\left(b - \boldsymbol{x}^\top \boldsymbol{\xi}\right)] \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\geq 0}\left(b - \boldsymbol{x}^\top \boldsymbol{\xi}^i\right), \tag{4}$$

where $\mathbb{1}_{\geq 0}(z)$ is the indicator function of the non-negative reals that equals 1 if $z \geq 0$ and 0 otherwise. Alternative sample-based approaches include quasi-Monte Carlo-based methods tailored for elliptically symmetric distributions [53, 54]. The quality of the resulting SAA solutions depends critically on the number $N$ of generated samples. A key issue is their deviation from the true (unknown) optimal solution. The sampled-based solution might be unstable when $N$ is small, meaning that small changes in the samples may have a significant impact on the obtained solutions [1, 11, 20]. Larger sample sizes are thus required to better approximate the optimal solution,

but they come at the cost of increased computational difficulty [19, 55]. This is exacerbated in high-dimensional regimes where incorporating large sample sizes renders the solution of the SAA models impractical (see Section 4.4).

Indeed, the challenge in SAA is the discontinuity of the indicator function $\mathbb{1}_{\geq 0}$ which needs to be further reformulated. This reformulation is typically done using additional binary variables to yield a mixed-integer optimization problem [10, 29], or approximated using continuous nonlinear functions, such as difference-of-convex functions [21], smooth differentiable functions [16, 39], or monotone convex functions [35] that also subsume the conditional-value-at-risk (CVaR) approximation. Each of these strategies requires introducing at least $N$ new variables or constraints, thus increasing the computational complexity of solving the chance-constrained problem.

In contrast to sample-based methods, sample-free approaches typically exploit the structure of the distribution of $\boldsymbol{\xi}$ to obtain analytically computable approximations or reformulations of the chance constraints. Early pioneering results are based on so-called $\alpha$-concave measures and functions [6, 40, 41]. In these cases, one can show that feasible region is convex, so that classical results in convex optimization can be adapted to design algorithms and optimality conditions. Convexity of the feasible set can also be achieved when $\boldsymbol{\xi}$ has symmetric logarithmically concave density function [27].

Another line of work aims to build (convex) outer approximations of the chance constraints. The popular CVaR approximation and other approaches based on Bernstein-type large deviation inequalities are examples of this approach [35, 44]. More recent sample-free approaches based on large deviation theory develop analytical closed-form (albeit nonlinear) expressions of the probabilistic constraint function for GMM distributed uncertainties [52].

The use of GMMs to model uncertainties in stochastic optimization models has gained significant interest in recent literature. Applications of GMM to quantify uncertainties can be found in portfolio optimization [31], chemical engineering [59, 60], and power system operations [14, 23, 58, 61, 62]. Joint chance-constrained optimization problems involving Gaussian mixture distributed randomness have been studied in [5, 59], although such problems typically rely on using Boole's inequality to obtain conservative individual chance-constrained problems.

The most closely relevant studies to our work are [14, 22, 37, 56]. All of these consider GMM-distributed uncertainties affecting either individual or two-sided chance constraints. In [22, 37], methods based on nonlinear optimization techniques, specifically gradient-based and spatial

branch-and-bound algorithms, are proposed to solve chance-constrained problems to global optimality. The work of [56] improves the performance of this spatial branch-and-bound algorithm by proposing an enhanced pruning strategy. In [14], a second-order cone programming (SOCP) model is proposed as a candidate approximation of chance-constrained problems under GMM-distributed uncertainties. Similar to our work, this reformulation also uses a piecewise linear approximation of the standard normal CDF by selecting an optimal number of breakpoints to obtain an SOCP approximation. However, the component-specific covariance matrices in their GMM are assumed to be proportional to the same general covariance matrix: $\boldsymbol{\Sigma}_k = \eta_k \boldsymbol{\Sigma}$ for some fixed $\eta_k > 0$. Also, their resulting SOCP model is optimal only when the risk threshold $\theta$ is sufficiently large; namely, when $\theta \geq 1 - \frac{1}{2}\min\{w_1, w_2, \ldots, w_K\}$. In contrast, our study considers a general form of the GMM without additional constraints on the component-specific weights, means, or covariances, and focuses on finding provably optimal solutions of linear chance-constrained models without restrictive assumptions on its problem parameters.

## 3. Mixed-Integer Optimization Formulations

We first present a mixed-integer quadratically constrained reformulation of the chance-constrained optimization problem (1) that motivates our subsequent developments. Our key insight is to isolate and separate the nonconvexities in the probability function $p(\boldsymbol{x})$, shown in (2), stemming from the fractional term, $(b - \boldsymbol{\mu}_k^\top \boldsymbol{x})/\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}$, and from the standard normal CDF, $\Phi$.

**Proposition 1.** *Problem* (1) *has the following equivalent reformulation.*

$$\begin{aligned} \min\ & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{s.t.}\ & \sum_{k=1}^{K} w_k \zeta_k \geq \theta, \quad \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{z} \in \mathbb{R}^K, \quad \boldsymbol{\zeta} \in [0,1]^K, \quad \boldsymbol{\lambda} \in \mathbb{R}^K_+, \\ & \left.\begin{array}{l} \Phi(z_k) \geq \zeta_k, \quad b - \boldsymbol{x}^\top \boldsymbol{\mu}_k \geq z_k \lambda_k, \quad \boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x} = \lambda_k^2 \end{array}\right\}\ k \in [K]. \end{aligned} \tag{5}$$

*Proof.* See Appendix A. □

Observe that the original non-convexities are captured in three separate sets of constraints for each $k \in [K]$; namely, *(i)* the constraint, $\Phi(z_k) \geq \zeta_k$, specified using the standard normal CDF, *(ii)* the bilinear constraint with the term $z_k \lambda_k$, and *(iii)* the quadratic equality constraint. Note that each of these three sets of constraints is non-convex. Recent developments in quadratic optimization solvers allow efficient handling of the nonconvex bilinear and quadratic constraints. Therefore, we only focus on approximating the constraint, $\Phi(z_k) \geq \zeta_k$, using suitable piecewise linear (PWL) approximations. In particular, we exploit the convex-concave structure of $\Phi(z)$, which is convex for $z \in (-\infty, 0]$ and concave for $z \in [0, \infty)$.

### 3.1. Piecewise Linear Outer Approximation

We now provide an outer approximation model called 'PWL-O' whose optimal objective value provides a lower bound on the optimal objective value of the original chance-constrained problem (1) and which "under-satisfies" the desired chance constraint probability. For notational simplicity, we temporarily drop subscript $k$ and focus on approximating $\Phi(z) \geq \zeta$, where $z, \zeta \in \mathbb{R}$.

Let $\check{\boldsymbol{z}} = (\check{z}_{-L}, \check{z}_{-L+1}, \ldots, \check{z}_{-1}, \check{z}_0, \check{z}_1, \ldots, \check{z}_{R-1}, \check{z}_R) \in \mathbb{R}^{L+R+1}$ denote a valid array of breakpoints, parameterized by integers $L > 0$ and $R > 0$ with

$$-\infty < \check{z}_{-L} < \check{z}_{-L+1} < \ldots < \check{z}_{-1} < \check{z}_0 = 0 < \check{z}_1 < \ldots < \check{z}_{R-1} < \check{z}_R < \infty. \tag{6}$$

The proposed piecewise linear outer approximation uses $L$ and $R+1$ linear pieces on the non-positive (left) and non-negative (right) sides of the domain of $\Phi$, respectively. We allow $L \neq R$; as we shall later discuss, the approximation of $\Phi$ on the convex and concave sides of $\Phi$ allows identification of a different number of breakpoints to achieve the same accuracy. We define the following quantities:

$$\begin{aligned} &g_i \coloneqq \phi(\check{z}_i), && g_i^0 = \Phi(\check{z}_i) - \phi(\check{z}_i)\check{z}_i, && i \in [R]_0, \\ &h_i \coloneqq \frac{\Phi\left(\check{z}_{-i+1}\right) - \Phi\left(\check{z}_{-i}\right)}{\check{z}_{-i+1} - \check{z}_{-i}}, && h_i^0 = \Phi(\check{z}_{-i}) - \frac{\Phi\left(\check{z}_{-i+1}\right) - \Phi\left(\check{z}_{-i}\right)}{\check{z}_{-i+1} - \check{z}_{-i}}\check{z}_{-i}, && i \in [L], \end{aligned}$$

where $g$ represents approximation of $\Phi$ on the non-negative reals and $h$ represents its approximation on the negative reals. Given the array of breakpoints $\check{\boldsymbol{z}}$, we define the PWL outer approximation of $\Phi$ as follows:

$$\bar{\Phi}(z; \check{\boldsymbol{z}}) \coloneqq \begin{cases} \min\left\{1, \min\limits_{i \in [R]_0}\left\{g_i z + g_i^0\right\}\right\}, & \text{if } z \geq 0, \quad \text{(Tangent)} \\ \max\left\{\Phi(\check{z}_{-L}), \max\limits_{i \in [L]}\left\{h_i z + h_i^0\right\}\right\}, & \text{otherwise.} \quad \text{(Secant)} \end{cases} \tag{7}$$

**Remark.** *Note that* (7) *ensures correctness of the outer approximation of* $\bar{\Phi}(z, \check{z})$ *for* $z \geq \check{z}_R$ *and* $z \leq \check{z}_{-L}$*, respectively, by using* $\min\left\{1, \min_{i \in [R]_0}\left\{g_i z + g_i^0\right\}\right\}$ *and* $\max\left\{\Phi(\check{z}_{-L}), \max_{i \in [L]}\left\{h_i z + h_i^0\right\}\right\}$ *to approximate the tails of the Gaussian distribution. In particular,* $\bar{\Phi}(z, \check{z})$ *equals* $\Phi(\check{z}_{-L})$ *for any* $z \leq \check{z}_{-L}$*, and* $\bar{\Phi}(z, \check{z})$ *equals* $1$ *for* $z \geq \check{z}_R$*. As we shall show, this allows us to bound the outer approximation error even when* $z$ *is outside* $[\check{z}_{-L}, \check{z}_R]$.

By construction, the graph of $\bar{\Phi}(\cdot; \check{\boldsymbol{z}})$, is above the graph of $\Phi$, as shown in Figure 4. This is because on the non-negative real line $(z \geq 0)$, it uses a tangent-based PWL approximation, whereas on the negative real line $(z < 0)$, it uses a secant-based PWL approximation.

Similar to $\Phi(z)$, the PWL approximation in (7) is also convex on the domain $z \leq 0$, and concave when $z \geq 0$. On the concave part of the domain, constraint $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ can be represented using

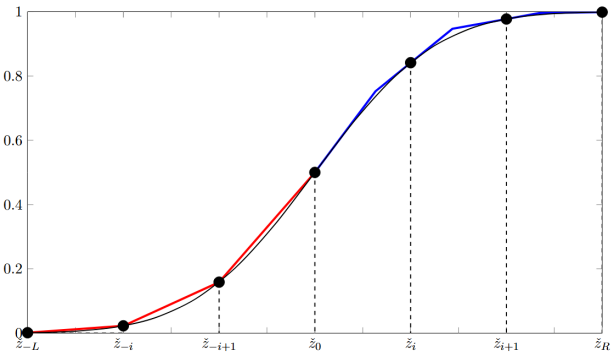


**Figure 4** **Piecewise linear outer approximation of $\Phi$.**

linear constraints, whereas on the convex part it is reformulated using Type-2 Special Ordered Set (SOS2) constraints. This is formalized in the following proposition.

**Proposition 2.** *If $\check{\boldsymbol{z}}$ is any valid array of breakpoints as in* (6) *and $\underline{z}, \bar{z} \in \mathbb{R}$ are given numbers satisfying $\underline{z} \leq \check{z}_{-L} < \check{z}_R \leq \bar{z}$, then there exist $(z, \zeta) \in [\underline{z}, \bar{z}] \times \mathbb{R}$ satisfying $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ if and only if $\mathcal{H}(\check{\boldsymbol{z}}) \neq \emptyset$, where*

$$\mathcal{H}(\check{\boldsymbol{z}}) := \left\{ \begin{array}{l|l} & \Phi(\check{z}_{-L})t_1 + t_2 + g_i y_3 + g_i^0 t_3 \geq \zeta \quad \forall i \in [R]_0, \\ \boldsymbol{\alpha} \in \mathbb{R}^L_+, & \Phi(\check{z}_{-L})t_1 + \sum_{i=0}^{L} \alpha_i \Phi(\check{z}_{-i}) + t_3 \geq \zeta, \\ \boldsymbol{t} \in \{0,1\}^3, & t_1 + t_2 + t_3 = 1,\ z = y_1 + y_2 + y_3, \\ \boldsymbol{y} \in \mathbb{R}^3, & -t_1 \underline{z} \leq y_1 \leq t_1 \check{z}_{-L}, \quad 0 \leq y_3 \leq t_3 \bar{z}, \\ z, \zeta \in \mathbb{R} & y_2 = \sum_{i=0}^{L} \alpha_i \check{z}_{-i},\ t_2 = \sum_{i=0}^{L} \alpha_i,\ \boldsymbol{\alpha} \in \mathrm{SOS2} \end{array} \right\}. \tag{8}$$

*Proof.* See AppendixA. □

In formulation (8), the binary variables, $t_1$, $t_2$, and $t_3$, are used to indicate if $z \in [\underline{z}, \check{z}_{-L}]$, $z \in (\check{z}_{-L}, 0)$, and $z \in [0, \bar{z}]$, respectively. The nonnegative variables $\boldsymbol{\alpha}$ can be nonzero only when $z \in (\check{z}_{-L}, 0)$. The SOS2 constraint ensures that at most two components of $\boldsymbol{\alpha}$ are positive, and that if two components are indeed positive, then these component indices are consecutive integers in the set $[L]_0$. Several modern mixed-integer optimization solvers enforce SOS2 constraints algorithmically during branching, as opposed to reformulating them as explicit constraints.

An application of Proposition 2 to each of the $K$ constraints, $\bar{\Phi}(z_k) \geq \zeta_k$, yields the following MIQP approximation of the reformulated problem (5), and hence, of the original chance con-

strained problem (1):

$$\begin{aligned}
&\min\ \boldsymbol{c}^\top \boldsymbol{x} \\
&\text{s.t.}\ \sum_{k=1}^{K} w_k \zeta_k \geq \theta, \quad \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{z} \in \mathbb{R}^K, \quad \boldsymbol{\zeta} \in [0,1]^K, \quad \boldsymbol{\lambda} \in \mathbb{R}^K_+ \\
&\qquad (\boldsymbol{\alpha}_k, \boldsymbol{t}_k, \boldsymbol{y}_k, z_k, \zeta_k) \in \mathcal{H}(\check{\boldsymbol{z}}), \quad b - \boldsymbol{x}^\top \boldsymbol{\mu}_k \geq z_k \lambda_k, \quad \boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x} = \lambda_k^2 \Big\}\ k \in [K].
\end{aligned} \tag{9}$$

The quality of the above PWL-O model depends on the accuracy of $\bar{\Phi}$, which in turn depends on the vector of breakpoints $\check{\boldsymbol{z}}$. The following lemma, inspired from [30, Appendix C.1], shows that $\bar{\Phi}$ is a $\tau$-accurate outer-approximation of $\Phi$ whenever the discretization of the breakpoint array, $\check{\boldsymbol{z}}$, is small in regions where the magnitude of the second derivative of $\Phi$ is large and when $\check{z}_{-L}$ and $\check{z}_R$ are sufficiently far from the origin.

**Lemma 1.** *For any $z_1, z_2 \in \mathbb{R}$, let $C(z_1, z_2) \coloneqq \max_{z \in [z_1, z_2]} |\Phi''(z)|$, $\check{\boldsymbol{z}}$ be any valid array of breakpoints as in* (6)*, and $\tau > 0$. Then the following holds:*

*a) If $\check{z}_i - \check{z}_{i-1} \leq \sqrt{\frac{2\tau}{C(\check{z}_{i-1}, \check{z}_i)}}$ for all $i \in [R]$, then $0 \leq \bar{\Phi}(z; \check{\boldsymbol{z}}) - \Phi(z) \leq \tau$ for all $z \geq 0$.*

*b) If $\check{z}_{-i+1} - \check{z}_{-i} \leq 2\sqrt{\frac{2\tau}{C(\check{z}_{-i}, \check{z}_{-i+1})}}$ for all $i \in [L]$, then $0 \leq \bar{\Phi}(z; \check{\boldsymbol{z}}) - \Phi(z) \leq \tau$ for all $z < 0$.*

*Proof.* See AppendixA. □

Lemma 1 generalizes the approximation conditions established in [14,30]. It shows that we only need the discretization interval to be small in regions where $C(\check{z}_{i-1}, \check{z}_i)$ is large. This is particularly important to control the number of breakpoints since $|\Phi''(z)|$ is small for large values of $z$ (see Figure 5). The number of breakpoints differs in (a) and (b) by a factor of 2 because, in comparison to the secant-based approximation, the tangent-based approximation has lower approximation error estimates.

We now discuss the calculation of the constant $C(\check{z}_{i-1}, \check{z}_i)$ appearing in Lemma 1. First note that $\Phi''(z) = -\phi(z)\, z$, and the maximum and minimum of $\Phi''(z)$ is attained at $\pm 1$ with the value $\pm\frac{e^{-0.5}}{\sqrt{2\pi}}$. Thus, $|\Phi''(z)| \leq \frac{e^{-0.5}}{\sqrt{2\pi}}, \forall z \in \mathbb{R}$. A worst-case bound on the number of breakpoints required to achieve a $\tau$-approximation in Lemma 1 can be obtained based on this absolute constant. However, as we note from Figure 5, $\Phi''(z)$ varies significantly on its domain, and it exhibits monotonicity on parts of its domain. A significant reduction in the number of breakpoints is possible when we take advantage of the monotonically changing values of $\Phi''(\cdot)$. These breakpoints are computed *a priori* based on the value of $\tau$ and varying $\Phi''(z)$. We now show how to achieve this.

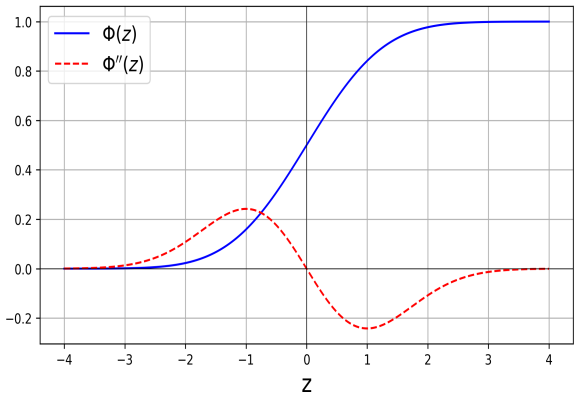


**Figure 5 CDF of standard Normal distribution and its second derivative**

**Algorithm 1 Breakpoints for approximating $\Phi(z)$**

(a) Tangent approximation of $\Phi(z), z \geq 0$

**Require:** $\check{z}_R, \tau$
**Ensure:** $\mathcal{A}^R$
1: **Part 1:** $0 \leq z \leq 1$
2: $i \leftarrow 1, \check{z}_i \leftarrow 1, \mathcal{A}^R \leftarrow \mathcal{A}^R \cup \{\check{z}_i\}$
3: $\check{z}_{i+1} = \check{z}_i - \sqrt{\frac{2\tau}{\frac{e^{-0.5}}{\sqrt{2\pi}}}}$
4: **while** $\check{z}_{i+1} > 0$ **do**
5: $\quad \mathcal{A}^R \leftarrow \mathcal{A}^R \cup \{\check{z}_{i+1}\}, i \leftarrow i+1$
6: $\quad \check{z}_{i+1} = \check{z}_i - \sqrt{\frac{2\tau}{\phi(\check{z}_i)\check{z}_i}}$
7: **end while**
8: $\mathcal{A}^L \leftarrow \mathcal{A}^R \cup \{0\}, i \leftarrow i+1$
9: Reverse the order of $\mathcal{A}^R$
10: $\mathcal{A}^R \leftarrow \mathcal{A}^R \backslash \{1\}$
11: **Part 2:** $1 \leq z \leq \check{z}_R$
12: $i \leftarrow i+1, \check{z}_{i-1} \leftarrow 1, \mathcal{A}^R \leftarrow \mathcal{A}^R \cup \{\check{z}_{i-1}\}$
13: $\check{z}_i = \check{z}_{i-1} + \sqrt{\frac{2\tau}{\phi(\check{z}_{i-1})\check{z}_{i-1}}}$
14: **while** $\check{z}_i < \check{z}_R$ **do**
15: $\quad \mathcal{A}^R \leftarrow \mathcal{A}^R \cup \{\check{z}_i\}, i \leftarrow i+1$
16: $\quad \check{z}_i = \check{z}_{i-1} + \sqrt{\frac{2\tau}{\phi(\check{z}_{i-1})\check{z}_{i-1}}}$
17: **end while**
18: $\mathcal{A}^R \leftarrow \mathcal{A}^R \cup \{\check{z}_R\}$

(b) Secant approximation of $\Phi(z)$, $z < 0$

**Require:** $\check{z}_{-L}, \tau$
**Ensure:** $\mathcal{A}^L$
1: **Part 1:** $-1 \leq z \leq 0$
2: $i \leftarrow 1, \check{z}_{-i} \leftarrow -1, \mathcal{A}^L \leftarrow \mathcal{A}^L \cup \{\check{z}_{-i}\}$
3: $\check{z}_{-i-1} = \check{z}_{-i} - 2\sqrt{\frac{2\tau}{\phi(\check{z}_{-i})\check{z}_{-i}}}$
4: **while** $\check{z}_{-i-1} < 0$ **do**
5: $\quad \mathcal{A}^L \leftarrow \mathcal{A}^L \cup \{\check{z}_{-i-1}\}, i \leftarrow -i-1$
6: $\quad \check{z}_{-i-1} = \check{z}_{-i} - 2\sqrt{\frac{2\tau}{\phi(\check{z}_{-i})\check{z}_{-i}}}$
7: **end while**
8: Reverse the order of $\mathcal{A}^L$
9: $\mathcal{A}^L \leftarrow \mathcal{A}^L \backslash -1$
10: **Part 2:** $\check{z}_{-L} \leq z \leq -1$
11: $i \leftarrow i+1, \check{z}_{-i+1} \leftarrow 1,$
$\quad \mathcal{A}^L \leftarrow \mathcal{A}^L \cup \{\check{z}_{-i+1}\}$
12: $\check{z}_{-i} = \check{z}_{-i+1} + 2\sqrt{\frac{2\tau}{\phi(\check{z}_{-i+1})\check{z}_{-i+1}}}$
13: **while** $\check{z}_{-i} > \check{z}_{-L}$ **do**
14: $\quad \mathcal{A}^L \leftarrow \mathcal{A}^L \cup \{\check{z}_{-i}\}, i \leftarrow i+1$
15: $\quad \check{z}_{-i} = \check{z}_{-i+1} + 2\sqrt{\frac{2\tau}{\phi(\check{z}_{-i+1})\check{z}_{-i+1}}}$
16: **end while**
17: $\mathcal{A}^L \leftarrow \mathcal{A}^L \cup \{\check{z}_{-L}\}$

Observe that $\Phi''(\cdot)$ is monotonically increasing in the intervals $[\check{z}_L, -1]$ and $[1, \check{z}_R]$, and it is monotonically decreasing in the interval $[-1, 1]$. For $z \geq 0$, we first divide the values of $z$ in intervals $[0, 1]$ (Part 1), $[1, \check{z}_R]$ (Part 2) (see lines 1-10 and 11-18 of Algorithm 1a respectively for Part 1 and Part 2). For the interval $[0, 1]$, we determine the breakpoints as a decreasing sequence starting at $z = 1$. We use the value of $\Phi''$ at the current point $\check{z}_i$ for $C(\check{z}_{i+1}, \check{z}_i)$ that determines the next point $\check{z}_{i+1}$. At the end of Part 1, we reverse the order of the breakpoints calculations (note that array

$\mathcal{A}^R$ maintains the breakpoints in increasing order (line 9)). $z = 1$ serves as the starting point for Part 2. Starting from $z = 1$, for the interval $[1, \check{z}_R]$ we determine the breakpoints as an increasing sequence. Once again, we use the value of $\Phi''$ at the current point $\check{z}_{i-1}$ for $C(\check{z}_{i-1}, \check{z}_i)$ to determine the next point $\check{z}_i$. Similar to a tangent-based approximation, Algorithm 1a outlines a secant-based approximation for $z < 0$ in two parts. Here, both parts start from $z = -1$ and add breakpoints at twice the interval used for tangent-based approximation and constitute an ordered sequence of breakpoints $\mathcal{A}^L$. Combining these two sequences, we achieve $\mathcal{A} \leftarrow \mathcal{A}^L \cup \mathcal{A}^R$.

Theorem 1 bounds the worst-case number of breakpoints required by Algorithms 1a-1b to achieve a $\tau$ error in the resulting PWL approximation.

**Theorem 1.** *Let* $\check{z}_{-L}, \check{z}_R$ *be such that* $\max\{1 - \Phi(\check{z}_R), \Phi(\check{z}_{-L})\} \leq \tau$, *where* $0 < \tau \ll \min\{|\check{z}_{-L}|, \check{z}_R\}$. *Then* $O(\sqrt{\frac{1}{\tau}\log(\frac{1}{\tau})})$ *breakpoints are sufficient (*$|\mathcal{A}|$ *in Algorithms 1a-1b) to ensure* $0 \leq \bar{\Phi}(z; \check{\boldsymbol{z}}) - \Phi(z) \leq \tau$, $\forall z \in \mathbb{R}$.

*Proof.* See AppendixA. □

### 3.2. Piecewise Linear Inner Approximation

An inner approximation of $\Phi(z)$ can be constructed similar to the outer approximation by taking secant approximation for $z \geq 0$ and tangent approximation for $z < 0$. This ensures that the inner approximation $\underline{\Phi}(\cdot; \check{\boldsymbol{z}})$ is always below the original curve $\Phi(z)$. More formally,

$$\underline{\Phi}(z; \check{\boldsymbol{z}}) = \begin{cases} \min\left\{\Phi(\check{z}_R), \min\limits_{i\in[R]}\left\{g_i z + g_i^0\right\}\right\}, & \text{if } z \geq 0, \quad \text{(Secant)} \\ \max\left\{0, \max\limits_{i\in[L]_0}\left\{h_i z + h_i^0\right\}\right\}, & \text{otherwise, (Tangent)} \end{cases} \tag{10}$$

where

$$g_i := \frac{\Phi(\check{z}_i) - \Phi(\check{z}_{i-1})}{\check{z}_i - \check{z}_{i-1}}, g_i^0 := \Phi(\check{z}_{i-1}) - \frac{\Phi(\check{z}_i) - \Phi(\check{z}_{i-1})}{\check{z}_i - \check{z}_{i-1}}\check{z}_{i-1}, i \in [R]$$

$$h_i := \phi(\check{z}_i), \qquad h_i^0 = \Phi(\check{z}_i) - \phi(\check{z}_i)\check{z}_i, \qquad i \in [L]_0.$$

Note that we have followed the convention that $g$ represents approximation of $\Phi(\cdot)$ on the non-negative reals and $h$ represents its approximation on the negative reals. Similar to $\Phi(z)$ and its outer approximation, its PWL inner approximation, $\underline{\Phi}(z; \check{\boldsymbol{z}})$, is also concave over the domain $z \geq 0$. Therefore, the constraint, $\underline{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$, can be reformulated using linear constraints over this domain. For $z \leq 0$, this function is convex, and it is reformulated using binary variables. The following proposition provides this formulation.

**Proposition 3.** *If $\check{\boldsymbol{z}}$ is any valid array of breakpoints and $\underline{z}, \bar{z} \in \mathbb{R}$ are given numbers satisfying $\underline{z} \leq \check{z}_{-L} < \check{z}_R \leq \bar{z}$, then there exist $(z, \zeta) \in [\underline{z}, \bar{z}] \times \mathbb{R}$ satisfying $\underline{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ if and only if $\mathcal{G}(\check{\boldsymbol{z}}) \neq \emptyset$, where*

$$\mathcal{G}(\check{\boldsymbol{z}}) := \left\{ \begin{array}{l|l} \boldsymbol{\alpha} \in \{0,1\}^L, & t_1 + t_2 + g_i y_3 + g_i^0 t_3 + \Phi(\check{z}_R) t_4 \geq \zeta \quad \forall i \in [R], \\ \boldsymbol{t} \in \{0,1\}^4, & h_L y_1 + h_L^0 t_1 + \sum_{i=0}^{L-1} (h_i \mathfrak{z}_i + h_i^0 \alpha_i) + t_3 + \Phi(\check{z}_R) t_4 \geq \zeta, \\ \boldsymbol{y} \in \mathbb{R}^3, & h_L y_1 + h_L^0 t_1 \geq 0, \quad t_1 + t_2 + t_3 + t_4 = 1, \\ z, \zeta \in \mathbb{R} & z = y_1 + y_2 + y_3 + y_4, \quad -\underline{z}\, t_1 \leq y_1 \leq \check{z}_{-L} t_1, \\ \mathfrak{z} \in \mathbb{R}^L & t_2 = \sum_{i=0}^{L-1} \alpha_i, \;\; \check{z}_{-i-1} \alpha_i \leq \mathfrak{z}_i \leq 0 \;\; \forall i \in [L-1]_0, \;\; y_2 = \sum_{i=0}^{L-1} \mathfrak{z}_i, \\ & 0 \leq y_3 \leq \check{z}_R t_3, \quad \bar{z}\, t_4 \geq y_4 \geq \check{z}_R t_4 \end{array} \right\}. \tag{11}$$

*Proof.* See Appendix A. □

In the above proposition, the binary variables, $t_1, t_2, t_3$, and $t_4$, are used to indicate if $z \in (\underline{z}, \check{z}_{-L}]$, $z \in [\check{z}_{-L}, 0]$, $z \in [0, \check{z}_R]$, and $z \in [\check{z}_R, \bar{z})$, respectively. Applying Proposition 3 to each of the $K$ constraints, $\hat{\Phi}(z_k) \geq \zeta_k$, we obtain the following MIQP approximation of the reformulated problem (5).

$$\begin{aligned} \min \;\; & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \;\; & \sum_{k=1}^{K} w_k \zeta_k \geq \theta, \quad \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{z} \in \mathbb{R}^K, \quad \boldsymbol{\zeta} \in [0,1]^K, \quad \boldsymbol{\lambda} \in \mathbb{R}^K_+, \\ & \left. \begin{array}{l} (\boldsymbol{\alpha}_k, \boldsymbol{t}_k, \boldsymbol{y}_k, z_k, \zeta_k) \in \mathcal{G}(\check{\boldsymbol{z}}), \quad b - \boldsymbol{x}^\top \boldsymbol{\mu}_k \geq z_k \lambda_k, \quad \boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x} = \lambda_k^2 \end{array} \right\} \; k \in [K]. \end{aligned} \tag{12}$$

We state the counterparts of Lemma 1 and Theorem 1 for the inner approximation in the Appendix B as Lemma 5 and Theorem 3, respectively.

**Remark.** *If the original problem* (1) *is feasible, the PWL-O model* (9) *is always feasible for any valid array of breakpoints. In contrast, the PWL-I model* (12) *may not always be feasible despite the feasibility of the original problem. However, note that a feasible inner approximation always over-satisfies the chance constraint.*

### 3.3. Optimality Guarantees to a Given Tolerance

We now show that the optimal objective values of the PWL-O and PWL-I models can be arbitrarily close to the optimal objective of the original chance-constrained model by controlling the accuracy threshold $\tau$ in the breakpoint calculation algorithms. To that end, we require the following additional assumptions and definition.

**Assumption 2.** *There exists $\rho > 0$ such that the feasible set, $P(\theta')$, is nonempty for all $\theta' \in [\theta - \rho, \theta + \rho]$.*

**Assumption 3.** *If $b = 0$, then there exists $\hat{\delta} > 0$ such that $P(\theta) \cap \{\boldsymbol{x} \in \mathbb{R}^n | \|\boldsymbol{x}\| \leq \hat{\delta}\} = \emptyset$. In other words, if $b = 0$, then $\boldsymbol{x} = \mathbf{0}$ is not a feasible solution of the chance-constrained problem* (1).

**Assumption 4.** *The linear independence constraint qualification is satisfied at all optimal solutions of the chance-constrained problem* (1).

**Definition 1** (Uniform compactness)**.** *[15, Definition 1.3]. The set $P(\theta)$ is uniformly compact near $\theta$ if there is a neighborhood $[\theta - \rho, \theta + \rho]$ of $\theta$ for some $\rho > 0$ such that the closure of the set, $\bigcup_{\theta' \in [\theta-\rho, \theta+\rho]} P(\theta')$, is compact.*

**Remark.** *Assumption 3 rules out $\mathbf{0}$ as an optimal solution when $b = 0$. This is a mild assumption even when $b \neq 0$ since this can be explicitly checked separately, and it might not even be of practical interest.*

Now, let $G : \mathbb{R}^n \to \mathbb{R}$ be defined such that

$$G(\boldsymbol{x}) = \begin{cases} \mathbb{1}_{\geq 0}(b), & \text{if } \boldsymbol{x} = \mathbf{0}, \\ \Phi\left(\dfrac{b - \boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}\right), & \text{otherwise.} \end{cases} \tag{13}$$

We show some graphs of $G(\boldsymbol{x})$ and its derivative in a one-dimensional setting below. We find that $G(\boldsymbol{x})$ and its derivative are well-behaved near $\boldsymbol{x} = 0$. We will formally prove this in the general case.

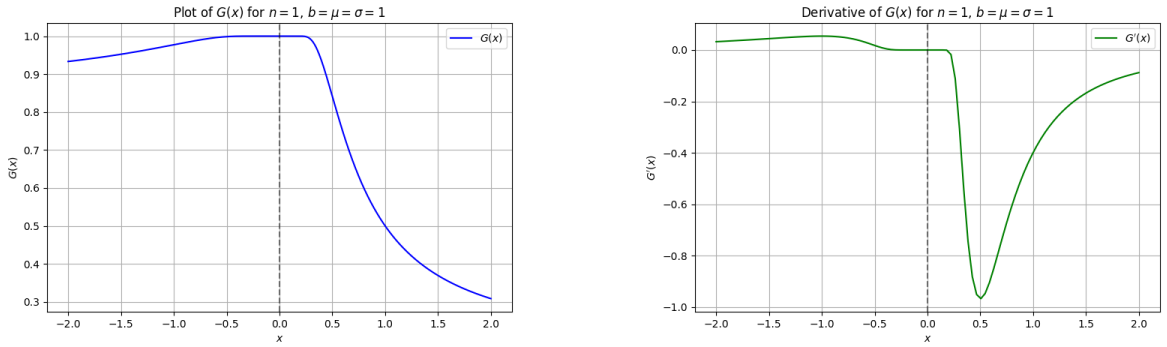


**Figure 6 The graph of $G(\boldsymbol{x})$ (left) and $G'(\boldsymbol{x})$ (right) when $\boldsymbol{x}$ is one-dimensional.**

**Lemma 2.** *Let $G$ be defined as in* (13)*. Then, the following holds:*

*(i) If $b < 0$, then $\mathbf{0} \notin P(\theta)$.*

*(ii) Under Assumptions 1-3, $G(\boldsymbol{x})$ is continuous on $P(\theta)$.*

*(iii) Under Assumptions 1-3, $P(\theta)$ is uniformly compact near $\theta$.*

*Proof.* We prove part *(i)* using (2a) which clearly shows that for any $\theta > 0$, $\sum_{k=1}^{K} w_k \mathbb{P}\left[0 \le b \,\middle|\, \boldsymbol{\xi}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right] \ge \theta$ does not hold with $b < 0$.

We prove part *(ii)* by establishing continuity of $G$ at an arbitrary point $\bar{\boldsymbol{x}} \in P(\theta)$. We consider two cases: (case-1) $\bar{\boldsymbol{x}} \neq \mathbf{0}$ and (case-2) $\bar{\boldsymbol{x}} = \mathbf{0}$.

To show (case-1), observe that $G(\boldsymbol{x}) = \Phi(g(\boldsymbol{x}))$ in this case, where $g(\boldsymbol{x}) = \frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}$. Now, let $\{\boldsymbol{x}_m\}_{m=1}^{\infty}$ be any sequence in $\mathbb{R}^n \setminus \{\mathbf{0}\}$ converging to $\bar{\boldsymbol{x}}$. Then, since Assumption 1 implies $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} > 0$ for any $\boldsymbol{x} \neq \mathbf{0}$, and $\bar{\boldsymbol{x}} \neq \mathbf{0}$, we have:

$$\lim_{m\to\infty} g(\boldsymbol{x}_m) = \lim_{m\to\infty} \frac{b - \boldsymbol{\mu}^\top \boldsymbol{x}_m}{\sqrt{\boldsymbol{x}_m^\top \boldsymbol{\Sigma} \boldsymbol{x}_m}} = \frac{\lim_{m\to\infty}(b - \boldsymbol{\mu}^\top \boldsymbol{x}_m)}{\lim_{m\to\infty} \sqrt{\boldsymbol{x}_m^\top \boldsymbol{\Sigma} \boldsymbol{x}_m}} = \frac{b - \boldsymbol{\mu}^\top \bar{\boldsymbol{x}}}{\sqrt{\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma} \bar{\boldsymbol{x}}}} = g(\bar{\boldsymbol{x}}).$$

The second equality holds by [45, Theorem 4.4] and the second-to-last equality follows due to the continuity of the functions, $b - \boldsymbol{\mu}^\top \boldsymbol{x}$ and $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}$. Thus $g$ is continuous at $\boldsymbol{x} = \bar{\boldsymbol{x}} \neq \mathbf{0}$. Additionally, since $\Phi$ is continuous on $\mathbb{R}$, $\Phi\big(g(\boldsymbol{x})\big)$ is also continuous at all $\boldsymbol{x} \neq \mathbf{0}$ by [45, Theorem 4.7].

To prove (case-2), note that part *(i)* implies that we must have $b \ge 0$. Moreover, Assumption 3 ensures that it suffices to consider only $b > 0$. Observe that the definition of $G$ implies that it is continuous at $\mathbf{0}$, if for any $\varepsilon > 0$, there exists some $\delta > 0$ such that all $\boldsymbol{x} \neq \mathbf{0}$ with $\|\boldsymbol{x}\| < \delta$ satisfy $\left|1 - \Phi\big(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}\big)\right| < \varepsilon$, which is equivalent to $1 - \Phi\big(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}\big) \le \varepsilon$ since $\Phi$ is always less than 1. To that end, define $\delta = \frac{b}{\Phi^{-1}(1-\varepsilon)\|\boldsymbol{\Sigma}^{\frac{1}{2}}\|\kappa} > 0$ with $\kappa = 1 + \frac{\|\boldsymbol{\mu}\|}{\Phi^{-1}(1-\varepsilon)\|\boldsymbol{\Sigma}^{\frac{1}{2}}\|}$. Then, observe that

$$\begin{aligned}
&\|\boldsymbol{x}\| \le \frac{b}{\Phi^{-1}(1-\varepsilon)\|\boldsymbol{\Sigma}^{1/2}\|\kappa} \implies \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} \le \|\boldsymbol{\Sigma}^{1/2}\|\|\boldsymbol{x}\| \le \frac{b}{\Phi^{-1}(1-\varepsilon)\kappa} \\
\implies &\Phi^{-1}(1-\varepsilon)\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} \le \frac{b}{\kappa} \implies \boldsymbol{\mu}^\top \boldsymbol{x} + \Phi^{-1}(1-\varepsilon)\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} \le \frac{b}{\kappa} + \boldsymbol{\mu}^\top \boldsymbol{x} \\
&\le \frac{b}{\kappa} + \|\boldsymbol{\mu}\| \frac{b}{\Phi^{-1}(1-\varepsilon)\|\boldsymbol{\Sigma}^{1/2}\|\kappa} \le \frac{b}{\kappa}\big(1 + \frac{\|\boldsymbol{\mu}\|}{\Phi^{-1}(1-\varepsilon)\|\boldsymbol{\Sigma}^{1/2}\|}\big) = b \\
\implies &\Phi^{-1}(1-\varepsilon)\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} \le b - \boldsymbol{\mu}^\top \boldsymbol{x} \implies 1 - \varepsilon \le \Phi\big(\frac{b - \boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}\big) \implies 1 - \Phi\big(\frac{b - \boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}\big) \le \varepsilon.
\end{aligned}$$

To prove part *(iii)*, we use the continuity property from part *(ii)*. By definition 1, uniform compactness of $P(\theta)$ near $\theta$ requires the closure of the set $\bigcup_{\theta' \in [\theta-\rho, \theta+\rho]} P(\theta')$ to be closed and bounded. Note that by Assumption 2, $\bigcup_{\theta' \in [\theta-\rho, \theta+\rho]} P(\theta')$ is not empty. Additionally, part *(ii)* implies that $p_k(\boldsymbol{x})$ and hence, $p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k p_k(\boldsymbol{x})$, is continuous everywhere as $\Phi(g_k(\boldsymbol{x}))$ is continuous. Then, given any $\theta' \in [\theta - \rho, \theta + \rho]$, for any convergent sequence $\{\boldsymbol{x}_m\}$ with $\boldsymbol{x}_m \in P(\theta')\ \forall m$, i.e.

$$\boldsymbol{H}\boldsymbol{x}_m = \boldsymbol{h}, \boldsymbol{A}\boldsymbol{x}_m \ge \boldsymbol{d}, p(\boldsymbol{x}_m) \ge \theta',\ \forall m, \text{ we have}$$

$$\lim_{m\to\infty} \boldsymbol{H}\boldsymbol{x}_m = \boldsymbol{H}\bar{x} = \boldsymbol{h},\ \lim_{m\to\infty} \boldsymbol{A}\boldsymbol{x}_m = \boldsymbol{A}\bar{\boldsymbol{x}} \ge \boldsymbol{d},\ \lim_{m\to\infty} p(\boldsymbol{x}_m) = p(\bar{\boldsymbol{x}}) \ge \theta'.$$

Thus, $P(\theta')$ also includes the limit point $\bar{\boldsymbol{x}}$. Therefore, $\bigcup_{\theta' \in [\theta-\rho,\theta+\rho]} P(\theta')$ is closed. Additionally, since $\mathcal{X}$ is a bounded set, $\mathcal{X} \cap \{\boldsymbol{x} \mid p(\boldsymbol{x}) \geq \theta'\} \subseteq \mathcal{X}$ is also bounded for every $\theta' \in [\theta-\rho, \theta+\rho]$. Therefore, $\bigcup_{\theta' \in [\theta-\rho,\theta+\rho]} P(\theta')$ is compact. □

**Lemma 3.** *$G(\boldsymbol{x})$ is continuously differentiable on $\mathbb{R}^n$ under Assumption 3. Specifically, for all $\boldsymbol{x} \in \mathbb{R}^n$,*

$$\nabla G(\boldsymbol{x}) = \Phi'(g(\boldsymbol{x}))\nabla g(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}} e^{-g(\boldsymbol{x})^2/2} \left\{ \frac{-\boldsymbol{\mu}}{(\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x})^{1/2}} - \frac{(b - \boldsymbol{\mu}^\top \boldsymbol{x})\boldsymbol{\Sigma}\boldsymbol{x}}{(\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x})^{3/2}} \right\} \tag{14}$$

*Proof.* Let $g(\boldsymbol{x}) \coloneqq \frac{b - \boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}}}$. We consider two cases: $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{x} = \mathbf{0}$.

Following the steps in the proof of Lemma 2, it can be shown that $g'(\boldsymbol{x})$ exists for all $\boldsymbol{x} \neq \mathbf{0}$, and since $\Phi(\cdot)$ is differentiable everywhere in $\mathbb{R}$, $\nabla G(\boldsymbol{x})$ is well-defined and continuous at all $\boldsymbol{x} \neq \mathbf{0}$. Hence, for all $\boldsymbol{x} \neq \mathbf{0}$, we can apply the chain rule [45] and obtain (14).

To prove differentiability at $\boldsymbol{x} = \mathbf{0}$ it suffices to prove it for $b > 0$ (Assumption 3 and Lemma 2(i)). Recall that a function $G(\boldsymbol{x})$ is continuously differentiable at $\mathbf{0}$ if and only if its partial derivatives exist and are continuous at $\mathbf{0}$ [32]. The partial derivatives of $G$ exist at $\mathbf{0}$ if following limits exist:

$$\frac{\partial G}{\partial \boldsymbol{x}_j}\Big|_{\boldsymbol{x}=0} = \underbrace{\lim_{h \to 0^+} \frac{\Phi(g(\mathbf{0} + h\boldsymbol{e}_j)) - G(\mathbf{0})}{h}}_{\text{A}} = \underbrace{\lim_{h \to 0^+} \frac{\Phi(g(\mathbf{0} - h\boldsymbol{e}_j)) - G(\mathbf{0})}{h}}_{\text{B}} \quad \forall j \tag{15}$$

Let $\|\boldsymbol{\Sigma}_j\|$ be the norm of the $j^{th}$ column vector of $\boldsymbol{\Sigma}$. Note that $g(\mathbf{0} + h\boldsymbol{e}_j) = (b - \mu_j h)/h\|\boldsymbol{\Sigma}_j\| = b/h\|\boldsymbol{\Sigma}_j\| - \mu_j/\|\boldsymbol{\Sigma}_j\| = q_j/h - r_j$, where $q_j = b/\|\boldsymbol{\Sigma}_j\| > 0$ since $b > 0$, and $r_j \coloneqq \mu_j/\|\boldsymbol{\Sigma}_j\|$. Since $G(\mathbf{0}) = 1$, using the definition of $\Phi$ in term A of (15), we have:

$$A = \lim_{h \to 0^+} \frac{\int_{-\infty}^{\frac{q_j}{h} - r_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - 1}{h} = \lim_{h \to 0^+} \frac{-\frac{1}{\sqrt{2\pi}} \int_{\frac{q_j}{h} - r_j}^{\infty} e^{-\frac{z^2}{2}} dz}{h} \ \forall j.$$

According to [50, Proposition 3], for $t > -1$, $\frac{2}{t + \sqrt{t^2+4}} \leq e^{t^2/2} \int_t^\infty e^{-z^2/2} dz \leq \frac{4}{3t + \sqrt{t^2+8}}$. Let $t_j = q_j/h - r$. Then as $h \to 0^+$, $t_j \to \infty$ since $q_j > 0$. Thus, the condition $t_j > -1$ is satisfied for all $j$. Therefore,

$$\frac{2e^{-(\frac{q_j}{h} - r)^2/2}}{h\Big((\frac{q_j}{h} - r) + \sqrt{(\frac{q_j}{h} - r)^2 + 4}\Big)} \leq \frac{\int_{\frac{q_j}{h} - r}^{\infty} e^{-z^2/2} dz}{h} \leq \frac{4e^{-(\frac{q_j}{h} - r)^2/2}}{h\Big(3(\frac{q_j}{h} - r) + \sqrt{(\frac{q_j}{h} - r)^2 + 8}\Big)}.$$

Note that the lower bound $\frac{2e^{(\frac{q_j}{h} - r)^2/2}}{h\left((\frac{q_j}{h} - r) + \sqrt{\left(\frac{q_j}{h} - r\right)^2 + 2}\right)} = \frac{2e^{-\left(\frac{q_j}{h} - r\right)^2/2}}{(q_j - rh) + \sqrt{(q_j - rh)^2 + 2h^2}}$ goes to 0 as $h \to 0^+$. Similarly, the upper bound also goes to 0 as $h \to 0$.

Now for the term $B$ in (15), $g(\mathbf{0} - h\boldsymbol{e}_j) = (b + \mu_j h)/h\|\boldsymbol{\Sigma}_j\| = b/h\|\boldsymbol{\Sigma}_j\| + \mu_j/\|\boldsymbol{\Sigma}_j\| = q_j/h + r_j = t_j$. Thus, as $h \to 0^+$, $t_j \to \infty$ (since $q_j > 0$). Hence, the condition $t_j > -1$ is satisfied. Therefore,

$$\frac{2e^{-(\frac{q_j}{h}+r)^2/2}}{h\big((\frac{q_j}{h}+r)+\sqrt{(\frac{q_j}{h}+r)^2+4}\big)} \le \frac{\int_{\frac{q_j}{h}+r}^{\infty} e^{-z^2/2}dz}{h} \le \frac{4e^{-(\frac{q_j}{h}+r)^2/2}}{h\big(3(\frac{q_j}{h}+r)+\sqrt{(\frac{q_j}{h}+r)^2+8}\big)}.$$

Similar to the proof for the part A in (15), we can show that both lower and upper bounds in the above inequality go to 0 as $h \to 0^+$. Thus, $\nabla G(\boldsymbol{x}) = \mathbf{0}$ at $\boldsymbol{x} = \mathbf{0}$ for $b > 0$. It shows that $G(\boldsymbol{x})$ is differentiable at all $\boldsymbol{x}$.

We now show that (14) provides an expression for $\nabla G(\boldsymbol{x})$ at all $\boldsymbol{x}$, including $\boldsymbol{x} = \mathbf{0}$, and that $\nabla G(\boldsymbol{x})$ is continuous. From (14),

$$\frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j} = \frac{1}{\sqrt{2\pi}} e^{-g(\boldsymbol{x})^2/2}\left\{\frac{-\boldsymbol{\mu}_j}{(\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x})^{1/2}} - \frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})(\boldsymbol{\Sigma}\boldsymbol{x})_j}{(\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x})^{3/2}}\right\}. \tag{16}$$

To prove continuity of $\nabla G(\boldsymbol{x})$ at $\boldsymbol{x} = \mathbf{0}$, when $b > 0$, we show that

$$\lim_{\boldsymbol{x}\to 0} \frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j} = \frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j}|_{\boldsymbol{x}=\mathbf{0}} = 0,\ \forall j$$

by showing that $\limsup_{\boldsymbol{x}\to\mathbf{0}} \frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j} = \liminf_{\boldsymbol{x}\to\mathbf{0}} \frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j} = 0$. Note that since $-\|\boldsymbol{\Sigma}^{1/2}\|\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\| \le (\boldsymbol{\Sigma}\boldsymbol{x})_j = (\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{x})_j \le \|\boldsymbol{\Sigma}^{1/2}\|\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|$,

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\exp\{-(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x}}})^2\}\left\{-\frac{|b-\boldsymbol{\mu}^\top \boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^3} - \frac{\|\mu\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}\right\} \le \frac{\partial G(\boldsymbol{x})}{\partial \boldsymbol{x}_j} \\
&\le \frac{1}{\sqrt{2\pi}}\exp\{-(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x}}})^2\}\left\{\frac{|b-\boldsymbol{\mu}^\top \boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^3} + \frac{\|\mu\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}\right\}\ \forall j.
\end{aligned} \tag{17}$$

Let the left and right hand side of (17) be $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$, respectively. It suffices to show that $\limsup_{\boldsymbol{x}\to\mathbf{0}} L(\boldsymbol{x}) = \liminf_{\boldsymbol{x}\to\mathbf{0}} U(\boldsymbol{x}) = 0$.

We first show that $\liminf_{\boldsymbol{x}\to\mathbf{0}} U(\boldsymbol{x}) = 0$. Since $e^y = \sum_{\ell=0}^{\infty} \frac{(y)^\ell}{\ell!} = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \cdots$, for any $j \in [n]$, $\ell \in \mathbb{N}$ and for $\ell \ge 2$,

$$\begin{aligned}
U(\boldsymbol{x}) &= \frac{1}{\sqrt{2\pi}}\left\{\left\|\frac{\frac{|b-\boldsymbol{\mu}^\top \boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^2}}{1+(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^2+\frac{1}{2!}(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^4+\frac{1}{3!}(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^6+\cdots}\right\|\right. \\
&\left.+\left\|\frac{\frac{\|\mu\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}}{1+(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^2+\frac{1}{2!}(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^4+\frac{1}{3!}(\frac{b-\boldsymbol{\mu}^\top \boldsymbol{x}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|})^6+\cdots}\right\|\right\} \\
&= \frac{1}{\sqrt{2\pi}}\left\{\left\|\frac{|b-\boldsymbol{\mu}^\top \boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^2+(b-\boldsymbol{\mu}^\top \boldsymbol{x})^2+\frac{1}{2!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^4}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^2}+\frac{1}{3!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^6}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^4}+\cdots}\right\|\right. \\
&\left.+\left\|\frac{\|\mu\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|+\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^2}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}+\frac{1}{2!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^4}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^3}+\frac{1}{3!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^6}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^5}+\cdots}\right\|\right\} \\
&\le \frac{1}{\sqrt{2\pi}}\left\{\left\|\frac{|b-\boldsymbol{\mu}^\top \boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^2+(b-\boldsymbol{\mu}^\top \boldsymbol{x})^2+\frac{1}{2!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^4}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^2}+\cdots+\frac{1}{\ell!}\frac{(b-\boldsymbol{\mu}^\top \boldsymbol{x})^{2\ell}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-2}}}\right\|\right.
\end{aligned}$$

$$+\left\|\frac{\|\mu\|}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|+\frac{(b-\boldsymbol{\mu}^\top\boldsymbol{x})^2}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|}+\frac{1}{2!}\frac{(b-\boldsymbol{\mu}^\top\boldsymbol{x})^4}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^3}+\cdots+\frac{1}{\ell!}\frac{(b-\boldsymbol{\mu}^\top\boldsymbol{x})^{2\ell}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-1}}}\right\|\Bigg\}$$
$$=\frac{1}{\sqrt{2\pi}}\Bigg\{\left\|\frac{\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-2}|b-\boldsymbol{\mu}^\top\boldsymbol{x}|\|\boldsymbol{\Sigma}^{1/2}\|}{\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell}+(b-\boldsymbol{\mu}^\top\boldsymbol{x})^2\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-2}+\cdots+(b-\boldsymbol{\mu}^\top\boldsymbol{x})^{2\ell}}\right\|$$
$$+\left\|\frac{\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-1}\|\mu\|}{\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell}+(b-\boldsymbol{\mu}^\top\boldsymbol{x})^2\ell!\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{x}\|^{2\ell-2}+\cdots+(b-\boldsymbol{\mu}^\top\boldsymbol{x})^{2\ell}}\right\|\Bigg\},$$

where the first inequality in the above follows since summation is truncated and all terms are positive. Now, since $b>0$, taking limit on both sides of the above expression, we obtain $\liminf_{\boldsymbol{x}\to\mathbf{0}} U(\boldsymbol{x})=0$. A similar argument shows that $\limsup_{\boldsymbol{x}\to\mathbf{0}} L(\boldsymbol{x})=0$. Thus, $\nabla G(\boldsymbol{x})$ in (14) is continuous at $\boldsymbol{x}=\mathbf{0}$ and $G(\boldsymbol{x})$ is continuously differentiable on $\mathbb{R}^n$. □

We are now in a position to present the main result of this section in Theorem 2 below. In particular, we show that using $O(\sqrt{\frac{1}{\tau}\log(\frac{1}{\tau})})$ breakpoints can provide a solution that achieves a user-specified tolerance $\hat{\tau}$ in the optimal objective value. We first establish the continuity of the optimal objective function of (1) by using a result from Gauvin and Tolle [15].

**Lemma 4.** *Let Assumptions 1-4 hold. Then, the optimal value function $Z^\star(\theta)$ is continuous at $\theta$.*

*Proof.* By Lemma 2, $P(\theta)$ is uniformly compact near $\theta$. Additionally, Lemma 3 ensures that $p(\boldsymbol{x})$ is continuously differentiable on $\mathbb{R}^n$. The KKT condition then implies that for dual variables $\boldsymbol{u}(\theta),\rho(\theta)\geq 0$, and $\boldsymbol{v}(\theta)$,

$$\boldsymbol{c}-\boldsymbol{H}^\top\boldsymbol{v}(\theta)-\boldsymbol{A}^\top\boldsymbol{u}(\theta)-\rho(\theta)\nabla p(\boldsymbol{x}^*(\theta))=\mathbf{0},\ \ \boldsymbol{u}_i(\theta)(\boldsymbol{A}_i\boldsymbol{x}^*(\theta)-\boldsymbol{d}_i)=0.$$

Let $\boldsymbol{B}\boldsymbol{x}^*(\theta)=\boldsymbol{d}_{\boldsymbol{B}}$ be the set of binding linear inequality constraints at an optimal solution, $\boldsymbol{x}^*(\theta)$. Since by Assumption 4, linear independence constraint qualification (LICQ) holds at $\boldsymbol{x}^*(\theta)$, there do not exist non-zero dual variables $\boldsymbol{u}(\theta),\rho(\theta)\geq 0$ and $\boldsymbol{v}(\theta)$ such that

$$\boldsymbol{H}^\top\boldsymbol{v}(\theta)+\boldsymbol{B}^\top\boldsymbol{u}(\theta)+\rho(\theta)\nabla p(\boldsymbol{x}^*(\theta))=\mathbf{0}.$$

Gordan's theorem [2, Theorem 2.4.9] then implies the existence of $\mathfrak{z}$ such that

$$\boldsymbol{B}\mathfrak{z}>\mathbf{0},\qquad \nabla p(\boldsymbol{x}^*(\theta))^\top\mathfrak{z}>0.$$

Thus, the Mangasarian-Fromovitz constraint qualification conditions hold at $\boldsymbol{x}^*(\theta)$. Then, [15, Theorem 2.6], implies that $Z^\star(\theta)$ is continuous at $\theta$. □

The following result is a direct consequence of the above lemma.

**Definition 2.** *A solution $\hat{\boldsymbol{x}}$ is called $\hat{\tau}-$feasible to problem* (1) *if $\boldsymbol{H}\hat{\boldsymbol{x}}=\boldsymbol{h}, \boldsymbol{A}\hat{\boldsymbol{x}}\geq\boldsymbol{d}$ and $p(\hat{\boldsymbol{x}})\geq\theta-\hat{\tau}$. A solution $\hat{\boldsymbol{x}}$ is called $\hat{\tau}-$optimal if it is $\hat{\tau}-$feasible and $|\boldsymbol{c}^\top\boldsymbol{x}^*(\theta)-\boldsymbol{c}^\top\hat{\boldsymbol{x}}|\leq\hat{\tau}$.*

**Theorem 2.** *Let $\boldsymbol{x}^*(\theta)$ be an optimal solution of* (1)*. Suppose Assumptions 1-4 hold, and we are given a user-specified tolerance $\hat{\tau}$. Then, there exists $\tau$ satisfying $\hat{\tau} \geq \tau > 0$, such that the outer and inner approximation problems* (9) *and* (12) *generated using $O(\sqrt{\frac{1}{\tau}\log(\frac{1}{\tau})})$ breakpoints have optimal solutions $\boldsymbol{x}^{\mathrm{OA}}$ and $\boldsymbol{x}^{\mathrm{IA}}$ that are $\hat{\tau}$-optimal. Specifically,*

$$|\boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{IA}} - \boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{OA}}| \leq |\boldsymbol{c}^\top \boldsymbol{x}^*(\theta+\tau) - \boldsymbol{c}^\top \boldsymbol{x}^*(\theta-\tau)| \leq \hat{\tau}.$$

*Proof.* To show $\hat{\tau}-$ feasibility, let us denote the feasible set corresponding to the PWL-O formulation (9) with $\tau-$accuracy as $\bar{P}(\theta,\tau)$. Then from Definition (3) and Theorem 1 we have $P(\theta) \subseteq \bar{P}(\theta,\tau) \subseteq P(\theta-\tau)$. Thus $\hat{\tau}-$feasibility follows since $\hat{\tau} \geq \tau$.

To prove $\hat{\tau}-$optimality, using the continuity of $Z^\star(\theta)$ at $\theta$ from Lemma 4, for every $\varepsilon > 0$, $\exists\, \tau^+(\varepsilon) > 0$ such that $|\theta - \theta'| \leq \tau^+(\varepsilon)$ implies $|Z^\star(\theta) - Z^\star(\theta')| \leq \varepsilon$. Let $\theta' \in [\theta-\tau, \theta+\tau]$ for some $\tau > 0$. Then, given the user-specified $\hat{\tau}$ we have a $\tau^+(\hat{\tau})$ for which $|Z^\star(\theta) - Z^\star(\theta')| \leq \hat{\tau}/2$ is satisfied for all sufficiently small $\hat{\tau} > 0$, with $\theta'$ satisfying $|\theta - \theta'| \leq \tau^+(\hat{\tau}/2) = \tau$. Now $|Z^\star(\theta+\tau) - Z^\star(\theta-\tau)| \leq |Z^\star(\theta+\tau) - Z^\star(\theta)| + |Z^\star(\theta) - Z^\star(\theta-\tau)| \leq \hat{\tau}$. The claim follows because $|\boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{IA}} - \boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{OA}}| \leq |Z^\star(\theta+\tau) - Z^\star(\theta-\tau)|$. □

An immediate consequence of Theorem 2 is that for any $\hat{\tau}$, there exists a $\tau > 0$ satisfying $|\boldsymbol{c}^\top \boldsymbol{x}^*(\theta) - \boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{OA}}| \leq \hat{\tau}$ and $|\boldsymbol{c}^\top \boldsymbol{x}^{\mathrm{IA}} - \boldsymbol{c}^\top \boldsymbol{x}^*(\theta)| \leq \hat{\tau}$. Although this is only an existence result, we note that we in practice can follow an iterative procedure to achieve the desired user-specified tolerance in the optimal objective value. In this procedure, we start with an initial value of $\tau$, and solve the outer and inner approximation problems. If the optimality gap determined from their solutions do not satisfy the desired tolerance $\hat{\tau}$, we reduce $\tau$ by a suitable factor (e.g., by a factor of 2) and update the previously generated break points. We can stop this procedure once the desired tolerance is satisfied.

## 4. Numerical Experience with a Commercial Solver

This section studies the computational performance of the PWL-I and PWL-O approximations, and the sample average approximation (SAA). Our discussion is based on extensive computational experiments conducted across a range of data generation parameters for the problems of different dimensions. We use Gurobi [18] to solve the instances. We describe problem generation in Section 4.1 and computational setup in Section 4.2. The break point selection in the PWL approximations is described in Section 4.3.

### 4.1. Data Generation

We randomly generated problems of 100, 500, and 1000 dimensions, i.e., $n \in \{100, 500, 1000\}$. We set $\theta \in \{0.95, 0.99, 0.999\}$ for chance constraint satisfaction probability. The value of $\theta = 0.999$ indicates a solution with a very high satisfaction probability. The number of Gaussian mixture components used was set at $K \in \{5, 10, 15\}$. For the results reported in the main body of this section, the mixture weights were $\boldsymbol{w} = (0.05, 0.1, 0.2, 0.3, 0.35)$ for $K = 5$, $\boldsymbol{w} = (0.001, 0.009, 0.02, 0.05, 0.08, 0.09, 0.1, 0.15, 0.2, 0.3)$ for $K = 10$ and $\boldsymbol{w} = (0.001, 0.005, 0.009, 0.01, 0.01, 0.015, 0.02, 0.05, 0.08, 0.09, 0.1, 0.12, 0.13, 0.17, 0.19)$ for $K = 15$. These weights are chosen to have both small and large values while ensuring that they sum to 1.0. We also conducted experiments using equal mixture weights. These results are reported in Appendix C.

Mixture component mean vectors and covariance matrices were generated using component-mean and variance control hyper-parameters, denoted by $\varrho$ and $\zeta$, respectively. We used six different values of $\varrho \in \{2, 5, 10, 15, 20, 25\}$. For each value of $\varrho > 0$, we randomly generated interval bounds $[\underline{\boldsymbol{\mu}}_k, \overline{\boldsymbol{\mu}}_k] \subset [0, \varrho\sqrt{n}\log n]^n$ for every $k \in [K]$. Given these bounds, each component $\mu_{ki}$ of $\boldsymbol{\mu}_k = (\mu_{k1}, \ldots, \mu_{kn})^\top$ is sampled as $\mu_{ki} \sim \text{Unif}([\underline{\mu}_{ki}, \overline{\mu}_{ki}])$ for all $i \in [n]$. Similarly, for covariance matrices, five different values of variance-control hyperparameter $\varsigma \in \{2, 5, 10, 15, 20\}$ were considered. Eigenvalues $\boldsymbol{\nu}_k, k \in [K]$ are uniformly sampled from $(0, \varsigma\ell_k]$, where $\ell_k \in [0, 1]$ was chosen through quasi-Monte-Carlo (QMC) sampling. QMC sampling ensures that $\ell_k$ is not clustered. Given a $\ell_k$, we finally constructed $\boldsymbol{\Sigma}_k \in \mathbb{S}^N_{++}$ for all $k \in [K]$ from an $n$-dimensional diagonal matrix of eigenvalues $\boldsymbol{D}_k$, where each entry was uniformly sampled from $[0, \varsigma\ell_k]$. The eigenvector matrices $\boldsymbol{Q}_k$ are randomly generated orthogonal matrices. They were generated by taking $2n$ products of randomly generated $2 \times 2$ Givens rotation matrices [17]. The generated covariance matrix $\boldsymbol{\Sigma}_k = \boldsymbol{Q}_k^\top \boldsymbol{D}_k \boldsymbol{Q}_k$ is positive definite.

We generated samples for the SAA approach as follows. We used the scikit-learn (`sklearn`) package in Python [38]. In particular, we used the `GaussianMixture` class from `sklearn.mixture`, initializing it with our Gaussian mixture parameters and then calling its `sample` method. Generated samples are then used in the general SAA model from [28]:

$$\min_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{y}\in\{0,1\}^S} \quad \boldsymbol{c}^\top \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{\xi}^s \boldsymbol{x} - b \le M y^s \quad \forall s \in [S] \qquad \sum_{s=1}^{S} y^s \le (1-\theta)\, S, \qquad \text{(GMM-SAA)}$$

where the binary vector $\boldsymbol{y}$ is of the same dimension as the sample size $S$. Big-$M$ value was set to $1e^6$. Sample size $S$ depends on $\theta$: for $\theta = 0.95$ and $\theta = 0.99$, we set $S = 100/(1-\theta)$. For

$\theta = 0.999$, we set $S = 20/(1-\theta)$, since generating problems with a large number of scenarios was not practical for $\theta = 0.999$.

To generate the right-hand side $b$ in constraint (1b), we sampled 1000 decision vectors $\boldsymbol{x}$ from its domain. We then computed $b$ as the average of $\boldsymbol{x}^\top \boldsymbol{\mu}_k + \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}$ over all $k$ and sampled decisions $\boldsymbol{x}$. One standard deviation of $\boldsymbol{x}^\top \boldsymbol{\mu}_k$ (i.e., $\boldsymbol{x}^\top \boldsymbol{\mu}_k + \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}$) was used to prevent an excessive proportion of infeasible instances. We did not use linear equality constraints and generated inequality constraints as follows: $\boldsymbol{A}$ and $\boldsymbol{d}$ are randomly generated as $\boldsymbol{A} \overset{\mathcal{U}}{\sim} [0,1]$ and $\boldsymbol{d} \overset{\mathcal{U}}{\sim} [n/4, 3N/4]$. For $n = \{100, 500\}$ we specify the dimension of vector $\boldsymbol{d}$ as $n/10$; for $n = 1000$ we reduce this dimension to $n/20$ to avoid too many infeasible instances.

### 4.2. Experimental Setup

All computations were performed on a server with Intel Xeon 2.80 GHz CPUs. We used Python 3.9 and Gurobi 11.0.3. We set $(1-\theta)/10$ as the `optimality (MIP)` gap in Gurobi. We allow Gurobi to use up to 2 cores per instance with other parameters set at their default values. A 18-hour (64,800s)-time limit is set. Let the solution from PWL-I, PWI-O, and SAA models within the time-limit be $\check{\boldsymbol{x}}^I, \check{\boldsymbol{x}}^O$, and $\check{\boldsymbol{x}}^S$. We substitute these solutions into (2d) to compute chance constraint probability satisfied by them. These probabilities are denoted by $\check{\theta}^I, \check{\theta}^O$ and $\check{\theta}^S$, respectively.

### 4.3. Break point selection and number of pieces in Piecewise Linear Approximation

We set $\tau = (1-\theta)/10$ in specifying piecewise linear approximation accuracy. For this value of $\tau$, we use Algorithm 1a-1b to adaptively construct the breakpoint-array $\check{\boldsymbol{z}}$. We fix breakpoint-array endpoints for all $\tau$ by setting $\check{z}_L = -6.466$ and $\check{z}_R = 6.466$. It meets the requirement of Theorems 1 and 3: $|\bar{\Phi}(\check{z}_L) - \Phi(\check{z}_L)| = |\bar{\Phi}(\check{z}_R) - \Phi(\check{z}_R)| < 1 \times e^{-10}$ (for PWL-O), and $|\underline{\Phi}(\check{z}_L) - \Phi(\check{z}_L)| = |\underline{\Phi}(\check{z}_R) - \Phi(\check{z}_R)| < 1 \times e^{-10}$ (for PWL-I).

### 4.4. Computational Results and Discussion

We present detailed numerical results for $\varrho = 2, \varsigma = 2$ and $\varrho = 5, \varsigma = 10$ with unequal mixture weight setting in Tables 1-2, respectively. Results for all the remaining parameter combinations of $(\varrho, \varsigma)$ with unequal weights are summarized in Table 3. Similar results with equal mixture weights are presented in Appendix C.

Table 1: Comparison of PWL-I, PWL-O, and SAA approaches for $\varrho = 2, \varsigma = 2$ with different mixture weights and 64,800s (18-hour) time limit

| | PWL-I | PWL-O | SAA |
|---|---|---|---|

| $K$ | $\theta$ | Time (sec) | Obj* | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj* | $\check{\theta}^S$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $n = 100$ | | | | | | |
| | 0.95 | 0.7 | -735 | 0.9500 | 0.4 | 0.0000 | 0.9499 | 0.011 | 0.5609 | -844 | 0.9565 |
| 5 | 0.99 | 1.2 | -735 | 0.9900 | 0.9 | 0.0002 | 0.9899 | 0.010 | 215.58 | -844 | 0.9452 |
| | 0.999 | 5.0 | -734 | 0.9990 | 5.3 | 0.0005 | 0.9990 | 0.000 | 338.77 | -844 | 0.9555 |
| | 0.95 | 2.2 | -727 | 0.9500 | 2.0 | 0.0002 | 0.9498 | 0.021 | 64,800 | -759 | 0.8636 |
| 10 | 0.99 | 3.4 | -726 | 0.9900 | 2.7 | 0.0001 | 0.9900 | 0.000 | 64,800 | -743 | 0.9186 |
| | 0.999 | 90 | -725 | 0.9990 | 57.1 | 0.0001 | 0.9990 | 0.000 | 23,893 | -730 | 0.9799 |
| | 0.95 | 3.5 | -704 | 0.9504 | 2.9 | 0.3555 | 0.9499 | 0.06 | 23,089 | -730 | 0.8790 |
| 15 | 0.99 | 2.6 | -706 | 0.9900 | 3.1 | 0.0006 | 0.9900 | 0.000 | 64,800 | -707 | 0.9573 |
| | 0.999 | 54 | -705 | 0.9990 | 55.2 | 0.0000 | 0.9990 | 0.000 | 64,805 | -706 | 0.9799 |
| | | | | | $n = 500$ | | | | | | |
| | 0.95 | 111.9 | -2,454 | 0.9758 | 109.7 | 0.0000 | 0.9995† | - | 5,238.8 | -2,474 | 0.5791 |
| 5 | 0.99 | 157.6 | -2,454 | 0.9901 | 199.2 | 0.0002 | 0.9900 | 0.010 | 64,800 | -2,473 | 0.5307 |
| | 0.999 | 201.4 | -2,453 | 0.9990 | 227.7 | 0.0001 | 0.9990 | 0.000 | 64,801 | -2,473 | 0.6910 |
| | 0.95 | 391.1 | -1,812 | 0.9500 | 329.0 | 0.0003 | 0.9498 | 0.021 | 64,800 | -1,857 | 0.6029 |
| 10 | 0.99 | 252.0 | -1,810 | 0.9900 | 251.6 | 0.0002 | 0.9900 | 0.000 | 64,801 | -1,812 | 0.7697 |
| | 0.999 | 498.7 | -1,809 | 0.9990 | 2,225.8 | 0.0001 | 0.9990 | 0.000 | 64,801 | -1,812 | 0.8628 |
| | 0.95 | 578.4 | -2,338 | 0.9501 | 658.4 | 0.0000 | 0.9496 | 0.053 | 64,800 | -2,665 | 0.6546 |
| 15 | 0.99 | 748.4 | -2,338 | 0.9900 | 686.6 | 0.0004 | 0.9899 | 0.010 | 64,800 | -2,411 | 0.8926 |
| | 0.999 | 1,556 | -2,337 | 0.9990 | 6,116 | 0.0000 | 0.9990 | 0.000 | 64,801 | -2,338 | 0.9018 |
| | | | | | $n = 1000$ | | | | | | |
| | 0.95 | 839.8 | -7,794 | 0.9510 | 823.2 | 0.0001 | 0.9499 | 0.116 | 64,800 | -7,794 | 0.3706 |
| 5 | 0.99 | 958.9 | -7,794 | 0.9901 | 760.5 | 0.0000 | 0.9900 | 0.010 | 64,800 | -7,794 | 0.7984 |
| | 0.999 | 798.4 | -7,793 | 0.9990 | 796.1 | 0.0000 | 0.9990 | 0.000 | 64,809 | -7,794 | 0.8662 |
| | 0.95 | 3,343 | -5,810 | 0.9500 | 2,947 | 0.0000 | 0.9496 | 0.053 | 64,800 | -5,824 | 0.4744 |
| 10 | 0.99 | 2,851 | -5,810 | 0.9900 | 2,363 | 0.0001 | 0.9899 | 0.010 | 64,800 | -5,819 | 0.5889 |
| | 0.999 | 8,008 | -5,809 | 0.9990 | 6,312 | 0.0000 | 0.9990 | 0.000 | 64,805 | -5,810 | 0.8211 |
| | 0.95 | 3,069 | -4,349 | 0.9506 | 4,275 | 0.0000 | 0.9498 | 0.105 | 64,800 | -4,592 | 0.6923 |
| 15 | 0.99 | 2,774 | -4,349 | 0.9900 | 4,587 | 0.0000 | 0.9899 | 0.010 | 64,801 | -4,445 | 0.8392 |
| | 0.999 | 2,836 | -4,348 | 0.9990 | 3,797 | 0.0000 | 0.9990 | 0.000 | 64,838 | -4,349 | 0.9049 |

Table 2: PWL-I, PWL-O vs SAA for $\varrho = 2, \varsigma = 5$ with different mixture weights

| $K$ | $\theta$ | Time (sec) | Obj* | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj* | $\check{\theta}^S$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $n = 100$ | | | | | | |
| | 0.95 | 0.53 | -988 | 0.9500 | 0.32 | 0.0000 | 0.9447 | 0.560 | 0.341 | -894 | 1.000 |
| 5 | 0.99 | 0.36 | -988 | 0.9900 | 0.476 | 0.0000 | 0.9899 | 0.010 | 3.504 | -894 | 1.000 |
| | 0.999 | 0.68 | -988 | 0.9990 | 0.76 | 0.0001 | 0.9990 | 0.000 | 11.25 | -894 | 1.000 |
| | 0.95 | 2.18 | -891 | 0.9500 | 1.37 | 0.0000 | 0.9499 | 0.011 | 1,978.9 | -894 | 0.9458 |
| 10 | 0.99 | 1.82 | -890 | 0.9900 | 4.02 | 0.0000 | 0.9889 | 0.111 | 64,853 | -892 | 0.9882 |
| | 0.999 | 14.3 | -890 | 0.9995 | 60.25 | 0.0190 | 0.9989 | 0.060 | 19,829 | -890 | 0.9970 |
| | 0.95 | 3.39 | -928 | 0.9500 | 3.65 | 0.0003 | 0.9499 | 0.011 | 30.99 | -981 | 0.9434 |
| 15 | 0.99 | 11.8 | -926 | 0.9900 | 25.1 | 0.0000 | 0.9899 | 0.010 | 36,789 | -976 | 0.9857 |
| | 0.999 | 58.91 | -922 | 0.9990 | 69.24 | 0.0001 | 0.9989 | 0.010 | 27,442 | -924 | 0.9922 |
| | | | | | $n = 500$ | | | | | | |
| | 0.95 | 115 | -2,344 | 0.9673 | 117.4 | 0.0010 | 0.9664† | - | 928.68 | -2,334 | 0.7964 |
| 5 | 0.99 | 166.2 | -2,343 | 0.9900 | 213.8 | 0.0005 | 0.9899 | 0.010 | 64,803 | -2,332 | 0.8411 |
| | 0.999 | 292.6 | -2,342 | 0.9990 | 268.3 | 0.0002 | 0.9989 | 0.010 | 64,801 | -2,330 | 0.8527 |
| | 0.95 | 410.4 | -2,500 | 0.9502 | 235.6 | 0.0002 | 0.9498 | 0.042 | 64,800 | -2,642 | 0.9194 |
| 10 | 0.99 | 267.7 | -2,499 | 0.9900 | 257.5 | 0.0003 | 0.9899 | 0.010 | 64,801 | -2,639 | 0.9826 |
| | 0.999 | 14,400 | -2,498 | 0.9990 | 2,460 | 0.0002 | 0.9989 | 0.010 | 25,398 | -2,593 | 0.9817 |
| | 0.95 | 737.7 | -2,188 | 0.9501 | 779.5 | 0.0001 | 0.9499 | 0.021 | 64,800 | -2,421 | 0.6018 |
| 15 | 0.99 | 633.6 | -2,187 | 0.9900 | 818.6 | 0.0006 | 0.9899 | 0.010 | 64,800 | -2,207 | 0.9324 |
| | 0.999 | 4,694 | -2,185 | 0.9990 | 15,318 | 0.0001 | 0.9989 | 0.010 | 64,804 | -2,186 | 0.9427 |
| | | | | | $n = 1000$ | | | | | | |

*(continued on next page)*

* Objective values are reported after truncating decimal digits

† For these instances, chance constraint appears inactive at target satisfaction probability

| $K$ | $\theta$ | Time (sec) | Obj | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj | $\check{\theta}^S$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.95 | 858 | -4,556 | 0.9509 | 930 | 0.0004 | 0.9498 | 0.116 | 64,800 | -4,558 | 0.4818 |
| | 0.99 | 1,176 | -4,554 | 0.9900 | 968.5 | 0.0001 | 0.9899 | 0.010 | 64,801 | -4,556 | 0.7318 |
| | 0.999 | 948 | -4,553 | 0.9990 | 1,007 | 0.0001 | 0.9989 | 0.010 | 64,806 | -4,554 | 0.9052 |
| 10 | 0.95 | 2,580 | -7,286 | 0.9500 | 2,596 | 0.0000 | 0.9495 | 0.053 | 64,800 | -7,599 | 0.5847 |
| | 0.99 | 2,540 | -7,285 | 0.9900 | 2,126 | 0.0001 | 0.9899 | 0.010 | 64,800 | -7,557 | 0.8751 |
| | 0.999 | 6,267 | -7,285 | 0.9990 | 4,117 | 0.0000 | 0.9989 | 0.010 | 64,801 | -7,285 | 0.9507 |
| 15 | 0.95 | 3,652 | -4,046 | 0.9500 | 3,747 | 0.0001 | 0.9497 | 0.032 | 64,800 | -4,365 | 0.7277 |
| | 0.99 | 564.5 | -4,046 | 0.9900 | 2,769 | 0.0000 | 0.9899 | 0.010 | 64,800 | -4,183 | 0.8842 |
| | 0.999 | 25,548 | -4,045 | 0.9990 | 2,933 | 0.0000 | 0.9989 | 0.010 | 64,801 | -4,046 | 0.9584 |

Table 3: Computational performance summary of PWL-I, PWL-O and SAA approaches with different mixture weights across hyperparameters

| | | Average Time (sec) | | | $\check{\theta}^S$ | % of Instances not Solved to Optimality Gap | | | Avg. $\check{\theta}^I, \check{\theta}^O, \check{\theta}^S$ for Instances not Solved to Optimality Gap | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $\theta$ | PWL-I | PWL-O | SAA | SAA | PWL-I | PWL-O | SAA | PWL-I | PWL-O | SAA |
| $n = 100$ | | | | | | | | | | | |
| 5 | 0.95 | 0.6 | 0.5 | 0.4 | 0.9927 | 0 | 0 | 0 | - | - | - |
| | 0.99 | 0.9 | 0.6 | 52 | 0.9908 | 0 | 0 | 0 | - | - | - |
| | 0.999 | 2.2 | 2.1 | 72 | 0.9925 | 0 | 0 | 0 | - | - | - |
| 10 | 0.95 | 2.6 | 1.56 | 15,108 | 0.9313 | 0 | 0 | 16.7 | - | - | 0.86 |
| | 0.99 | 14.0 | 14.4 | 64,809 | 0.9698 | 0 | 0 | 100 | - | - | 0.96 |
| | 0.999 | 52.5 | 52.0 | 50,488 | 0.9895 | 0 | 0 | 66.7 | - | - | 0.99 |
| 15 | 0.95 | 51.3 | 2.6 | 26,046 | 0.9054 | 0 | 0 | 28.6 | - | - | 0.89 |
| | 0.99 | 46.3 | 40.5 | 50,059 | 0.9604 | 0 | 0 | 66.7 | - | - | 0.94 |
| | 0.999 | 178 | 198.8 | 55,462 | 0.9793 | 0 | 0 | 75.0 | - | - | 0.97 |
| $n = 500$ | | | | | | | | | | | |
| 5 | 0.95 | 123 | 133 | 38,760 | 0.4934 | 0 | 0 | 50.0 | - | - | 0.52 |
| | 0.99 | 144 | 183 | 64,802 | 0.4253 | 0 | 0 | 100 | - | - | 0.42 |
| | 0.999 | 732 | 655 | 64,801 | 0.5539 | 0 | 0 | 100 | - | - | 0.55 |
| 10 | 0.95 | 501 | 250 | 64,800 | 0.8916 | 0 | 0 | 100 | - | - | 0.89 |
| | 0.99 | 260 | 366 | 64,813 | 0.8410 | 0 | 0 | 100 | - | - | 0.84 |
| | 0.999 | 2,504 | 2,384 | 58,235 | 0.9267 | 0 | 0 | 83.3 | - | - | 0.92 |
| 15 | 0.95 | 6,298 | 5,327 | 58,910 | 0.7653 | 0 | 0 | 66.7 | - | - | 0.76 |
| | 0.99 | 6,835 | 10,636 | 64,800 | 0.9090 | 0 | 0 | 100 | - | - | 0.91 |
| | 0.999 | 11,958 | 21,322 | 64,802 | 0.9063 | 0 | 13.3 | 100 | - | 0.9989 | 0.91 |
| $n = 1000$ | | | | | | | | | | | |
| 5 | 0.95 | 679 | 742 | 40,204 | 0.6654 | 0 | 0 | 50.0 | - | - | 0.53 |
| | 0.99 | 996 | 830 | 64,807 | 0.8376 | 0 | 0 | 100 | - | - | 0.84 |
| | 0.999 | 2,112 | 2,360 | 64,804 | 0.8836 | 0 | 0 | 100 | - | - | 0.88 |
| 10 | 0.95 | 2,400 | 3,294 | 54,106 | 0.6501 | 0 | 0 | 83.3 | - | - | 0.65 |
| | 0.99 | 2,400 | 1,976 | 64,812 | 0.8077 | 0 | 0 | 100 | - | - | 0.81 |
| | 0.999 | 2,537 | 8,525 | 64,802 | 0.8911 | 0 | 3.3 | 100 | - | 0.9999 | 0.89 |
| 15 | 0.95 | 4,929 | 13,426 | 64,800 | 0.7053 | 20 | 16.7 | 100 | 0.9978 | 0.9989 | 0.70 |
| | 0.99 | 17,313 | 22,778 | 64,817 | 0.8828 | 30 | 26.7 | 100 | 0.9989 | 0.9991 | 0.88 |
| | 0.999 | 21,022 | 31,354 | 64,863 | 0.9310 | 33.3 | 43.3 | 100 | 0.9997 | 0.9998 | 0.93 |

Tables 1–2 show the time required to solve the instances, objective value at termination, and the values of $\check{\theta}^I, \check{\theta}^O, \check{\theta}^S$. For PWL-O, we also report $\%$-Obj[§] and $\%\check{\theta}^{\ddagger}$ respectively defined as: $\max\left(\boldsymbol{c}^\top(\check{\boldsymbol{x}}^I - \check{\boldsymbol{x}}^O), 0\right)/\boldsymbol{c}^\top\check{\boldsymbol{x}}^I \times 100$ and $(\check{\theta}^I - \check{\theta}^O)/\theta \times 100$. The results in these tables show that SAA reaches the 18-hour limit for nearly all instances (even some with $K = 5$ components), whereas PWL-I and PWL-O are able to close the optimality gap much faster. In particular, PWL-I and PWL-O instances take at most 60 seconds for $n = 100$, and less than 6 and 9 hours for $n = 500$,

and $1000$, respectively. Runtimes of PWL-I and PWL-O grow with $n$ and $K$. Both approximations take a similar amount of time.

Both the reported values of $\check{\theta}^O$ and $\check{\theta}^I$ are within the approximation accuracy ($\tau$) of the target value of $\theta$. The few exceptions highlighted by symbol$^\dagger$ indicate that the solution we obtain for them lies in the feasible interior of the chance constraint. Except for those few instances, chance constraint satisfaction $\check{\theta}^I$ from PWL-I, in general, is at least as large as the target satisfaction probability $\theta$ and satisfaction $\check{\theta}^O$ obtained from PWL-O. Such an observation is consistent with the statements of Lemmas 1 and 5 made for the outer and inner approximation models, respectively. In contrast, chance constraint satisfaction probability $\check{\theta}^S$ at the SAA solutions is below the target value frequently—sometimes markedly below—suggesting inadequacy of the number of samples in the sample average approximation. Finally, PWL-I and PWL-O achieve almost the same optimal objective, with PWL-O having an objective value that is slightly smaller (as expected since the feasible set from inner approximation is contained in that by outer approximation). SAA objective values, however, are typically much smaller than those of PWL-O. Such smaller values come at the expense of their poor chance constraint satisfaction probability.

**4.4.1. Discussion on the summary Table 3** The results for PWL-I, PWL-O, and SAA summarized in Tables 3 report average computation time across instances, percentage of instances failing to achieve the desired optimality gap within 18 hours, and $\check{\theta}^I, \check{\theta}^O, \check{\theta}^S$ for the unsolved instances. This summary also highlights the superior performance of the PWL-I and PWL-O approximations over SAA in terms of both computational efficiency and probability satisfied by their solutions. The aggregate performance statistics are highlighted below:

- **Average computation time-based comparison:** For SAA, approximately 60% of instances reach the 18-hour time limit, and in 74% of the cases the computation time exceeds 14 hours. On the other hand, in any of the cases, the average time reported for PWL-I and PWL-O never reaches the time limit. Additionally, no significant difference is observed in the average time required to solve PWL-I and PWL-O.
- **Comparison of failure rates in achieving desired optimality gap:** In $55\%$ of the test settings, SAA fails to achieve the prescribed gap ($\frac{1-\theta}{10}$) within the time limit. In the remaining $45\%$ cases, many SAA instances with $K = 10$ or $K = 15$ also fail to achieve the desired gap. In contrast, PWL-I and PWL-O approximations almost always achieve the gap, except for $n = 1000$ and $K = 15$ instances. Even in these cases, the percentage of instances that fail to meet the prescribed

gap stays below $50\%$. Additionally, for $n = 1000, K = 10$ and $n = 500, K = 15$, fewer than $10\%$ of the PWL-O instances fail to meet the required gap, on average.

- **Comparison based on the satisfied chance constraint probability:** For PWL-I and PWL-O approximations instances that failed to meet the optimality gap, the reported average $\check{\theta}^I, \check{\theta}^O$ values are mostly observed to be greater than the target probability $\theta$. This indicates that even when the desired optimality gap is not attained, solutions we obtain at termination satisfy the required chance constraint probability on average. This however may come at the cost of increased objective value.

  On the other hand, SAA-generated solutions yield $\check{\theta}^S$ significantly below the target $\theta$, even when averaged over those instances that meet the required optimality gap. Understandably, average $\check{\theta}^S$ values become worse when instances that fail to satisfy the optimality requirement are also incorporated (see last column of Table 6). Furthermore, in comparison to PWL models, worsening performance trends of SAA are more noticeable with increasing $n$. Even with $n = 500$ and $K = 5$, some SAA runs yield $\check{\theta}^S$ values 2.5 times below the target, and with $n = 1000$, these values can be 3-fold smaller. Interestingly, the $\check{\theta}^S$ values for SAA appear to improve with $K$.

**4.4.2. Effect on solution time with number of breakpoints** We conducted additional experiments to examine the effect of the number of breakpoints and linear pieces on computation time and optimality gap at termination for both PWL-I and PWL-O approximations. For $n = 500$ and $n = 1000$, we randomly selected about 10% instances from those used in our previous experimentation settings. We, however, now set the approximation accuracy $\tau$ to $(1-\theta)/2$ and $(1-\theta)/50$, which require respectively fewer and more breakpoints (and linear pieces) than the baseline value $(1-\theta)/10$. Tables 7-8 and Tables 9-10 report the findings respectively for $n = 500$ and $n = 1000$. The first four columns of these tables show the values of mean and variance control parameters $\varrho$ and $\varsigma$, the number of components $K$, and the probability threshold $\theta$. As expected, increasing accuracy requirement increases solution time for both inner and outer approximations, with the effect being more pronounced for $n = 1000$ than for $n = 500$. For $n = 500$, every instance is solved within the 18-hour limit for both error tolerances and all values of $K$. In contrast, for $n = 1000$ cases, many instances fail to reach the desired optimality gap that was set in Gurobi. About 40% of instances with $n = 1000$ yield no feasible solution within 18 hours for $\tau = (1-\theta)/50$. However, in several instances, the optimality gap at termination is almost identical for $(1-\theta)/2$ and $(1-\theta)/50$ cases, indicating that Gurobi cannot tighten the optimality gap further regardless of the

approximation accuracy. Such observation is more noticeable for PWL-I than PWL-O approximation, indicating that outer approximation is easier to solve, especially when a finer approximation is used. The reasons are unclear. However, we hypothesize that it is because the feasible set of PWL-I approximation is more restricted, thus making it difficult for the MIP solver to identify feasible solutions.

**4.4.3. Solution time and mixture distribution** To examine the effect of the number of Gaussian mixture components and their parameters on solution time, we perform linear regression with the following predictors:

(i) $n$ – problem dimension

(ii) $K$ – number of mixture components

(iii) $\varrho$ – mean-control hyperparameter

(iv) $\varsigma$ – variance-control hyperparameter

(v) $\bar{d}_\mu$ – average of the distances between any two Gaussian component means

(vi) $d_\mu^{\max}$ – maximum of the distances between any two Gaussian component means

(vii) average value of eigenvalue ratio $\nu^{\max}/\nu^{\min}$ across mixture components

(viii) maximum value of eigenvalue ratio $\nu^{\max}/\nu^{\min}$ across mixture components

Analyses are performed separately for different values of $\theta$. Regression results are reported in Appendix D for both PWL-I and PWL-O approximations. Predictably, solution time increases with both $n$ and $K$, with $K$ showing a greater effect. Other observations are as follows:

- For both PWL-I and PWL-O approximations, predictors $\varsigma, \bar{d}_\mu$ and $d_\mu^{\max}$ exhibit statistically significant influence on solution time, particularly with larger $\theta$.
- For both PWL-I and PWL-O approximations, solution time and optimality gap tend to increase with greater variance in the component distribution. Such effects become more pronounced at higher $\theta$ values. The influence of variance, as captured by the regression coefficient and statistical significance of $\varsigma$, is slightly stronger for PWL-O than PWL-I approximation.
- For PWL-I approximation with $\theta = 0.99, 0.999$, it is observed that the smaller the $\bar{d}_\mu$ value (*average* distance across all pairs), the faster it is to reach the desired optimality gap. However, the effect of $d_\mu^{\max}$ value (maximum distance across all pairs) appears to be opposite.
- Similar effect of $\bar{d}_\mu$ and $d_\mu^{\max}$ are observed for PWL-O approximation but they are statistically significant only when $\theta = 0.999$. Additionally, corresponding regression coefficient values are observed to be smaller than those for PWL-I approximation.

## 5. Conclusion

We studied a linear chance-constrained problem with continuous decision variables, assuming that the data follows a GMM distribution. We related the approximation accuracy of $\Phi$ to a perturbation in chance constraint satisfaction probability $\theta$ and showed that the desired optimality gap can be ensured through controlling the approximation accuracy. An extensive computational study with thousands of instances gives a clear indication of the superior performance of piecewise-linear inner and outer approximation over the popular SAA approach, in terms of both computation time and the chance constraint probability satisfied by the solution of these problems. We believe that insights from our numerical findings could facilitate customized algorithm development for solving problems with multiple or joint GMM-based chance constraints. Extending this study to discrete decision variables and studying distributionally robust chance constraints for mixture models, with unknown component parameters, weights, and counts, are some promising future research directions.

## References

[1] Shabbir Ahmed and Alexander Shapiro. Solving chance-constrained stochastic programs via sampling and integer programming. *INFORMS TutORials in Operations Research: State-of-the-Art Decision-Making tools in the Information-Intensive Age*, pages 261–269, 2008.

[2] Mokhtar S Bazaraa, Hanif D Sherali, and Chitharanjan M Shetty. *Nonlinear programming: theory and algorithms*. John wiley & sons, 2006.

[3] Jose Blanchet, Joost Jorritsma, and Bert Zwart. Optimization under rare events: scaling laws for linear chance-constrained programs. *arXiv preprint arXiv:2407.11825*, 2024.

[4] Jose Blanchet, Fan Zhang, and Bert Zwart. Efficient scenario generation for heavy-tailed chance constrained optimization. *Stochastic Systems*, 14(1):22–46, 2024.

[5] Spencer Boone and Jay McMahon. Spacecraft maneuver design with non-gaussian chance constraints using gaussian mixtures. In *2022 AAS/AIAA Astrodynamics Specialist Conference*, 2022.

[6] Christer Borell. Convex measures on locally convex spaces. *Arkiv för matematik*, 12(1):239–252, 1974.

[7] Giuseppe Calafiore and Marco C Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005.

[8] Abraham Charnes and William W Cooper. Chance-constrained programming. *Management science*, 6(1):73–79, 1959.

[9] Jaeseok Choi, Anand Deo, Constantino Lagoa, and Anirudh Subramanyam. Reduced sample complexity in scenario-based control system design via constraint scaling. *IEEE Control Systems Letters*, 8:2793–2798, 2024.

[10] Frank E Curtis, Andreas Wächter, and Victor M Zavala. A sequential algorithm for solving nonlinear optimization problems with chance constraints. *SIAM Journal on Optimization*, 28(1):930–958, 2018.

[11] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, 55(5):798–812, 2009.

[12] Darinka Dentcheva. Optimization Models with Probabilistic Constraints. In Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, editors, *Lectures on Stochastic Programming: Modeling and Theory*, pages 87–153. Society for Industrial and Applied Mathematics, 2009.

[13] Darinka Dentcheva and Gabriela Martinez. Regularization methods for optimization problems with probabilistic constraints. *Mathematical Programming*, 138(1):223–251, 2013.

[14] Abolhassan Mohammadi Fathabad, Jianqiang Cheng, Kai Pan, and Boshi Yang. Asymptotically tight conic approximations for chance-constrained ac optimal power flow. *European Journal of Operational Research*, 305(2):738–753, 2023.

[15] Jacques Gauvin and Jon W Tolle. Differential stability in nonlinear programming. *SIAM Journal on Control and Optimization*, 15(2):294–311, 1977.

[16] Abebe Geletu, Armin Hoffmann, Michael Kloppel, and Pu Li. An inner-outer approximation approach to chance constrained optimization. *SIAM Journal on Optimization*, 27(3):1834–1857, 2017.

[17] Wallace Givens. Computation of plain unitary rotations transforming a general matrix to triangular form. *Journal of the Society for Industrial and Applied Mathematics*, 6(1):26–50, 1958.

[18] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024.

[19] René Henrion. A critical note on empirical (sample average, Monte Carlo) approximation of solutions to chance constrained programs. In Dietmar Hömberg and Fredi Tröltzsch, editors, *System Modeling and Optimization*, pages 25–37. Springer Berlin Heidelberg, 2013.

[20] René Henrion and Werner Römisch. Hölder and lipschitz stability of solution sets in programs with probabilistic constraints. *Mathematical Programming*, 100:589–611, 2004.

[21] L Jeff Hong, Yi Yang, and Liwei Zhang. Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Operations Research*, 59(3):617–630, 2011.

[22] Zhaolin Hu, Wenjie Sun, and Shushang Zhu. Chance constrained programs with gaussian mixture models. *IISE Transactions*, 54(12):1117–1130, 2022.

[23] Deping Ke, CY Chung, and Yuanzhang Sun. A novel probabilistic optimal power flow model with uncertain wind power generation described by customized gaussian mixture model. *IEEE Transactions on Sustainable Energy*, 7(1):200–212, 2015.

[24] Sujin Kim, Raghu Pasupathy, and Shane G Henderson. A guide to sample average approximation. *Handbook of simulation optimization*, pages 207–243, 2015.

[25] Simge Küçükyavuz. On mixing sets arising in chance-constrained programming. *Mathematical programming*, 132(1):31–56, 2012.

[26] Simge Küçükyavuz and Ruiwei Jiang. Chance-constrained optimization under limited distributional information: A review of reformulations based on sampling and distributional robustness. *EURO Journal on Computational Optimization*, 10:100030, 2022.

[27] Constantino M Lagoa, Xiang Li, and Mario Sznaier. Probabilistically constrained linear programs and risk-adjusted controller design. *SIAM Journal on Optimization*, 15(3):938–951, 2005.

[28] James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[29] James Luedtke, Shabbir Ahmed, and George L Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming*, 122(2):247–272, 2010.

[30] Fengqiao Luo and Sanjay Mehrotra. Decomposition algorithm for distributionally robust optimization using wasserstein metric with an application to a class of regression models. *European Journal of Operational Research*, 291(2):450–470, 2021.

[31] Eric Luxenberg and Stephen Boyd. Portfolio construction with gaussian mixture returns and exponential utility via convex optimization. *Optimization and Engineering*, 25(1):555–574, 2024.

[32] Jerrold E Marsden and Anthony Tromba. *Vector calculus*. Macmillan, 2003.

[33] Bruce L Miller and Harvey M Wagner. Chance constrained programming with joint constraints. *Operations research*, 13(6):930–945, 1965.

[34] Arkadi Nemirovski and Alexander Shapiro. Scenario approximations of chance constraints. In Giuseppe Calafiore and Fabrizio Dabbene, editors, *Probabilistic and randomized methods for design under uncertainty*, pages 3–47. Springer, 2006.

[35] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.

[36] Bernardo K Pagnoncelli, Shabbir Ahmed, and Alexander Shapiro. Sample average approximation method for chance constrained programming: theory and applications. *Journal of optimization theory and applications*, 142(2):399–416, 2009.

[37] Xiaochuan Pang, Shushang Zhu, and Zhaolin Hu. Chance constrained program with quadratic randomness: A unified approach based on gaussian mixture distribution. *arXiv preprint arXiv:2303.00555*, 2023.

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[39] Alejandra Peña-Ordieres, James R Luedtke, and Andreas Wächter. Solving chance-constrained problems via a smooth sample-based nonlinear approximation. *SIAM Journal on Optimization*, 30(3):2221–2250, 2020.

[40] András Prékopa. Logarithmic concave measures with applications to stochastic programming. 1971.

[41] András Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34:335–343, 1973.

[42] András Prékopa. Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probability distribution. *Zeitschrift für Operations Research*, 34:441–461, 1990.

[43] András Prékopa. Probabilistic programming. In Andrzej Ruszczyński and Alexander Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267 – 351. Elsevier, 2003.

[44] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[45] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[46] Andrzej Ruszczyński and Alexander Shapiro. Stochastic Programming Models. In Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, editors, *Lectures on Stochastic Programming: Modeling and Theory*, pages 1–25. Society for Industrial and Applied Mathematics, 2009.

[47] D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. Wiley, 2015.

[48] Suvrajeet Sen. Relaxations for probabilistically constrained programs with discrete random variables. *Operations Research Letters*, 11(2):81–86, 1992.

[49] Anirudh Subramanyam. Chance-constrained programming: Rare events. In *Encyclopedia of Optimization*, pages 1–6. Springer, 2022.

[50] Stanislaw J Szarek and Elisabeth Werner. A nonsymmetric correlation inequality for gaussian measure. *Journal of multivariate analysis*, 68(2):193–211, 1999.

[51] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley, 1985.

[52] Shanyin Tong, Anirudh Subramanyam, and Vishwas Rao. Optimization under rare chance constraints. *SIAM Journal on Optimization*, 32(2):930–958, 2022.

[53] W Van Ackooij, I Aleksovska, and M Munoz-Zuniga. (Sub-) differentiability of probability functions with elliptical distributions. *Set-Valued and Variational Analysis*, 26(4):887–910, 2018.

[54] Wim Van Ackooij and René Henrion. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. *SIAM Journal on Optimization*, 24(4):1864–1889, 2014.

[55] Shanshan Wang, Jinlin Li, and Sanjay Mehrotra. Chance-constrained multiple bin packing problem with an application to operating room planning. *INFORMS Journal on Computing*, 33(4):1661–1677, 2021.

[56] Jinxiang Wei, Zhaolin Hu, Jun Luo, and Shushang Zhu. Enhanced branch-and-bound algorithm for chance constrained programs with gaussian mixture models. *Annals of Operations Research*, pages 1–33, 2024.

[57] Weijun Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, 186(1):115–155, 2021.

[58] Shuwei Xu and Wenchuan Wu. Tractable reformulation of two-side chance-constrained economic dispatch. *IEEE Transactions on Power Systems*, 37(1):796–799, 2021.

[59] Yu Yang. Optimal blending under general uncertainties: A chance-constrained programming approach. *Computers & Chemical Engineering*, 171:108170, 2023.

[60] Yu Yang. Two-stage chance-constrained programming based on gaussian mixture model and piecewise linear decision rule for refinery optimization. *Computers & Chemical Engineering*, 184:108632, 2024.

[61] Yue Yang, Wenchuan Wu, Bin Wang, and Mingjie Li. Analytical reformulation for stochastic unit commitment considering wind power uncertainty with gaussian mixture model. *IEEE Transactions on Power Systems*, 35(4):2769–2782, 2019.

[62] Tianyang Yi, Shibshankar Dey, D Adrian Maldonado, Sanjay Mehrotra, and Anirudh Subramanyam. Discrete-continuous gaussian mixture models for wind power generation. In *2024 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2024.

## Appendix A: Proofs

***Proof of Proposition 1:*** Let $\boldsymbol{x}$ be any feasible solution of (1). First, suppose $\boldsymbol{x} = \mathbf{0}$. Then $p(\mathbf{0}) = \mathbb{P}[0 \leq b]$, which is equal to 1 if $b \geq 0$. If $b < 0$, then $p(\boldsymbol{x}) = 0 \not\geq \theta$, contradicting feasibility of $\boldsymbol{x}$. Thus, it must be true that $b \geq 0$ for $\boldsymbol{x} = \mathbf{0}$. Now, observe that one constructs a feasible solution in (2a) with the same objective value by setting $\boldsymbol{x} = \mathbf{0}$, $\lambda_k = 0$, $\zeta_k = \theta$, and $z_k = \Phi^{-1}(\theta)$ for all $k \in [K]$. Suppose now that $\boldsymbol{x} = \bar{\boldsymbol{x}} \neq \mathbf{0}$. Again, we construct a feasible solution in (2a) with the same objective value by setting $\boldsymbol{x} = \bar{\boldsymbol{x}}$, $\lambda_k = \sqrt{\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_k \bar{\boldsymbol{x}}}$, $z_k = \frac{b - \bar{\boldsymbol{x}}^\top \boldsymbol{\mu}_k}{\sqrt{\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_k \bar{\boldsymbol{x}}}}$, and $\zeta_k = \Phi(z_k)$ for all $k \in [K]$.

Now, let $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\zeta}, \boldsymbol{\lambda})$ be any feasible solution for (1). We claim that $\boldsymbol{x}$ is also feasible to (1) with the same objective value. First, suppose $\boldsymbol{x} = \mathbf{0}$. Then for all $k \in [K]$, $\lambda_k = 0$ implies that $b \geq 0$ for any arbitrary value of $z_k$. Hence, $\Phi(z_k) \geq \zeta_k$ is always satisfied for all $k \in [K]$ and therefore $\sum_{k \in [K]} w_k \zeta_k \geq \theta$. Suppose now that $\boldsymbol{x} \neq \mathbf{0}$. Then

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \Phi\left(\frac{b - \boldsymbol{x}^\top \boldsymbol{\mu}_k}{\sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma}_k \boldsymbol{x}}}\right) = \sum_{k=1}^{K} w_k \Phi\left(\frac{b - \boldsymbol{x}^\top \boldsymbol{\mu}_k}{\lambda_k}\right) \geq \sum_{k=1}^{K} w_k \Phi\left(z_k\right) \geq \sum_{k=1}^{K} w_k \zeta_k \geq \theta.$$

The equality follows from (2d) because of Assumption 1 and that $\boldsymbol{x} \neq \mathbf{0}$, whereas the inequalities follow from feasibility of $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\zeta}, \boldsymbol{\lambda})$ and monotonicity of $\Phi$. Hence, $p(\boldsymbol{x}) \geq \theta$ of (1) is satisfied having same objective value as that of (1). □

***Proof of Proposition 2:*** First, consider the forward direction of the claim starting from $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$. Let us take some $z \geq 0$, $\zeta$ such that $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$. Then, $z = y_3 \geq 0$ with $t_3 = 1$ implies that $t_1, t_2 = 0$, $g_i z + g_i^0 \geq \zeta$, $i \in [R]_0$ and $1 \geq \zeta$. Thus, for a given solution $z \geq 0$ satisfying $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$, one can construct a feasible solution to the set $\mathcal{H}(\check{\boldsymbol{z}})$ using $t_1, t_2 = 0$, $t_3 = 1$, $\boldsymbol{\alpha} = \mathbf{0}$, $y_1, y_2 = 0, z = y_3 \geq 0$.

Next we consider some $z \in [\check{z}_{-L}, 0)$. Then, $z$ will be in $[\check{z}_{-i}, \check{z}_{-i+1}]$ for some $i \in [L]$ which indicates $t_2 = 1$, $t_1 = t_3 = 0$, and therefore, $y_1 = y_3 = 0$. Then there exists some weights $\hat{\alpha}_i, \hat{\alpha}_{i-1} \geq 0$ such that $\hat{\alpha}_i + \hat{\alpha}_{i-1} = 1 = t_2$. Furthermore, from constraint $y_2 = \sum_{i=0}^{L} \alpha_i \check{z}_{-i}$ we obtain $y_2 = \hat{\alpha}_i \check{z}_{-i} + \hat{\alpha}_{i-1} \check{z}_{-i+1}$. Hence, $\alpha_i \Phi(\check{z}_{-i}) + \alpha_{i-1} \Phi(\check{z}_{-i+1}) \geq \zeta$. Thus for any given $z \in [\check{z}_{-L}, 0)$, one can find an index $i \in [L]$ such that $\alpha_i = \hat{\alpha}_i$, $\alpha_{i-1} = \hat{\alpha}_{i-1}$ with $y_2 = \hat{\alpha}_i \check{z}_{-i} + \hat{\alpha}_{i-1} \check{z}_{-i+1}$, $y_1 = y_3 = 0$, and $\alpha_j = 0 \; \forall j \in [L] \backslash \{i-1, i\}$, which satisfy all the constraints of $\mathcal{H}(\check{\boldsymbol{z}})$. Similarly, if we take some $z < \check{\boldsymbol{z}}_{-L}$, then $t_1 = 1$, and $y_1 = z$. This implies $t_2, t_3, y_2, y_3 = 0, \Phi(\check{z}_{-L}) \geq \zeta$. Therefore, $\mathcal{H}(\check{\boldsymbol{z}}) \neq \emptyset$.

Now we prove that any point in the set $\mathcal{H}(\check{\boldsymbol{z}})$ also satisfies $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ based on the definition of $\bar{\Phi}(z; \check{\boldsymbol{z}})$ in (7). It is formally shown below using three separate cases:

i) Let us choose $y_3 = z \geq 0$ with $t_3 = 1$ and $t_1, t_2, y_1, y_2 = 0$ such that $g_i y_3 + g_i^0 \geq \zeta$ for all $i \in [R]_0$. Then $\bar{\Phi}(z; \check{\boldsymbol{z}}) = \min_{i \in [R]_0} \{g_i y_3 + g_i^0\} \geq \zeta$.

ii) Let $\check{z}_{-L} \leq y_2 = z < 0$ with $t_2 = 1$ and $t_1, t_3, y_1, y_3 = 0$ such that $y_2 = \alpha_{\check{i}} \check{z}_{-\check{i}} + \alpha_{\check{i}-1} \check{z}_{-\check{i}+1}$, $\alpha_{\check{i}} + \alpha_{\check{i}-1} = 1 = t_2$, and $\alpha_{\check{i}} \Phi(\check{z}_{-\check{i}}) + \alpha_{\check{i}-1} \Phi(\check{z}_{-\check{i}+1}) \geq \zeta$ for some $\check{i} \in [L]$ and $\alpha_j = 0$ for all $j \in [L] \backslash \{\check{i} - 1, \check{i}\}$. Thus there exists an index $\check{i} \in [L]$ that satisfies $\bar{\Phi}(z; \check{\boldsymbol{z}}) = h_{\check{i}} y_2 + h_{\check{i}}^0 \geq \zeta$. Hence, $\bar{\Phi}(z; \check{\boldsymbol{z}}) = \max_{i \in [L]} h_i y_2 + h_i^0 \geq \zeta$ is satisfied.

iii) Finally, $y_1 = z < \check{z}_{-L}$ with $t_1 = 1$, $t_2, t_3, y_2, y_3 = 0$ indicates that $\Phi(\check{z}_{-L}) = \bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$.

Additionally, since all of these cases allow $\zeta$ to be at most 1, using the definition of $\bar{\Phi}(z; \check{\boldsymbol{z}})$ in (7), $\bar{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ is satisfied. □

***Proof of Lemma 1:*** We split the proof into two cases depending on the convexity/concavity of $\Phi$. Recall that for any $\hat{z}$ between $z_1$ and $z_2$, and for some $\mathfrak{z} \in (z_1, z_2)$, the Taylor expansion with Lagrange remainder around $z_1$ gives

$$\Phi(\hat{z}) = \Phi(z_1) + \Phi'(z_1)(\hat{z} - z_1) + \tfrac{1}{2} \Phi''(\mathfrak{z}) (\hat{z} - z_1)^2. \tag{18}$$

**Case 1 (Secant approximation when $z < 0$).** Let $z_1 < \hat{z} < z_2 < 0$. For notational convenience let the secant line be $S(z) = \bar{\Phi}(z; \check{z})$ for $z < 0$:

$$S(z) := \Phi(z_1) + \frac{\Phi(z_2) - \Phi(z_1)}{z_2 - z_1}(z - z_1). \tag{19}$$

From Taylor's theorem at $z_1$, we have

$$\Phi(\hat{z}) = \Phi(z_1) + \Phi'(z_1)(\hat{z} - z_1) + \tfrac{1}{2}\Phi''(\mathfrak{z})\,(\hat{z} - z_1)^2,$$

for some $\mathfrak{z} \in (z_1, \hat{z})$. Similarly, expanding at $z_1$ and plugging $z = z_2$ gives

$$\Phi(z_2) = \Phi(z_1) + \Phi'(z_1)(z_2 - z_1) + \tfrac{1}{2}\Phi''(\eta)\,(z_2 - z_1)^2,$$

for some $\eta \in (z_1, z_2)$. Hence, the slope of the secant is

$$\frac{\Phi(z_2) - \Phi(z_1)}{z_2 - z_1} = \Phi'(z_1) + \tfrac{1}{2}\Phi''(\eta)(z_2 - z_1).$$

Substituting this slope expression into (19) and using $z = \hat{z}$, we obtain

$$S(\hat{z}) = \Phi(z_1) + \big(\Phi'(z_1) + \tfrac{1}{2}\Phi''(\eta)(z_2 - z_1)\big)\,(\hat{z} - z_1).$$

Next, subtracting (18) from $S(\hat{z})$, we find

$$S(\hat{z}) - \Phi(\hat{z}) = \tfrac{1}{2}\Phi''(\eta)(z_2 - z_1)(\hat{z} - z_1) - \tfrac{1}{2}\Phi''(\mathfrak{z})(\hat{z} - z_1)^2.$$

Factoring out $(\hat{z} - z_1)$,

$$S(\hat{z}) - \Phi(\hat{z}) = \tfrac{1}{2}(\hat{z} - z_1)\Big(\Phi''(\eta)(z_2 - z_1) - \Phi''(\mathfrak{z})(\hat{z} - z_1)\Big). \tag{20}$$

Next, we show that for some $\tilde{\mathfrak{z}} \in (\hat{z}, z_2)$,

$$\Phi''(\eta)(z_2 - z_1) - \Phi''(\mathfrak{z})(\hat{z} - z_1) = \Phi''(\tilde{\mathfrak{z}})(z_2 - \hat{z}), \tag{21}$$

using the fact that $\Phi$ is twice continuously differentiable on an interval containing $[z_1, z_2]$ with $z_1 < \hat{z} < z_2$ and applying the Mean Value Theorem to $\Phi'$ respectively on the intervals $[z_1, \hat{z}]$, $[\hat{z}, z_2]$ and $[z_1, z_2]$.

By the Mean Value Theorem on $[\hat{z}, z_2]$, there exists $\tilde{\mathfrak{z}} \in (\hat{z}, z_2)$ such that

$$\Phi'(z_2) - \Phi'(\hat{z}) = \Phi''(\tilde{\mathfrak{z}})\,(z_2 - \hat{z}), \text{ and}$$

By the Mean Value Theorem on $[z_1, \hat{z}]$, there exists $\mathfrak{z} \in (z_1, \hat{z})$ such that

$$\Phi'(\hat{z}) - \Phi'(z_1) = \Phi''(\mathfrak{z})\,(\hat{z} - z_1).$$

By adding the above two equalities, we get: $\Phi'(z_2) - \Phi'(z_1) = \Phi''(\tilde{\mathfrak{z}})\,(z_2 - \hat{z}) + \Phi''(\mathfrak{z})\,(\hat{z} - z_1)$. Now, by the Mean Value Theorem on $[z_1, z_2]$, there exists $\eta \in (z_1, z_2)$ such that

$$\Phi'(z_2) - \Phi'(z_1) = \Phi''(\eta)\,(z_2 - z_1).$$

Combining the last two equality expressions, we obtain (21). Thus, plugging (21) into (20) gives us:

$$S(\hat{z}) - \Phi(\hat{z}) = \tfrac{1}{2}\,\Phi''(\tilde{\mathfrak{z}})(z_2 - \hat{z})(\hat{z} - z_1), \quad \text{for some } \tilde{\mathfrak{z}} \in (\hat{z}, z_2).$$

Now, since $\Phi$ is convex on $(-\infty, 0)$, we have $\Phi''(\tilde{\mathfrak{z}}) \geq 0$. Hence,

$$0 \;\leq\; S(\hat{z}) - \Phi(\hat{z}) \;\leq\; \tfrac{1}{2} C(z_1, z_2)\,(z_2 - \hat{z})(\hat{z} - z_1),$$

where $C(z_1, z_2) := \max_{z \in [z_1, z_2]} |\Phi''(z)|$. Furthermore, by geometric inequality

$$(z_2 - \hat{z})(\hat{z} - z_1) \;\leq\; \tfrac{1}{4}(z_2 - z_1)^2.$$

This inequality holds as equality at $\hat{z} = \frac{z_1 + z_2}{2}$. Therefore, the secant-based approximation error is bounded by

$$S(\hat{z}) - \Phi(\hat{z}) \;\leq\; \tfrac{1}{8} C(z_1, z_2)\,(z_2 - z_1)^2.$$

**Case 2 (Tangent approximation when $z \geq 0$).** Let $z_1 < \hat{z} < z_2$ with $z_1 \geq 0$. Consider the tangent line:

$$T(z) := \Phi(z_1) + \Phi'(z_1)(z - z_1).$$

Subtracting the Taylor expansion (18) from $T(z)$ gives

$$T(\hat{z}) - \Phi(\hat{z}) = -\tfrac{1}{2}\Phi''(\mathfrak{z})\,(\hat{z} - z_1)^2$$

for some $\mathfrak{z} \in (z_1, \hat{z})$. Since $\Phi(z)$ is concave in $z$ for $z \geq 0$, we have $\Phi''(\mathfrak{z}) \leq 0$. Hence, the error is nonnegative:

$$0 \;\leq\; T(\hat{z}) - \Phi(\hat{z}) \;\leq\; \tfrac{1}{2} C(z_1, z_2)\,(\hat{z} - z_1)^2 \;\leq\; \tfrac{1}{2} C(z_1, z_2)\,(z_2 - z_1)^2.$$

Similarly, expanding around $z_2$ yields

$$0 \;\leq\; T(\hat{z}) - \Phi(\hat{z}) \;\leq\; \tfrac{1}{2} C(z_1, z_2)\,(z_2 - \hat{z})^2 \;\leq\; \tfrac{1}{2} C(z_1, z_2)\,(z_2 - z_1)^2.$$

□

**Proposition 4.** *Let $z \in \mathbb{R}$. Then*

$$\Phi''(z) \quad \textit{is} \quad \begin{cases} \textit{strictly increasing for } z \in (-\infty, -1] \\ \textit{strictly decreasing for } z \in [-1, 1], \\ \textit{strictly increasing for } z \in [1, \infty), \end{cases}$$

*with* $\max_{z \le 0} \Phi''(z) = \Phi''(-1) = \frac{1}{\sqrt{2\pi}} e^{-1/2}$ *and* $\min_{z \ge 0} \Phi''(z) = \Phi''(1) = -\frac{1}{\sqrt{2\pi}} e^{-1/2}$*, respectively.*

*Proof.* Note that $\Phi''(z) = \frac{d}{dz}\Big(\phi(z)\Big) = \frac{d}{dz}\left(\frac{1}{\sqrt{2\pi}} e^{-z^2/2}\right) = -z\,\phi(z)$. Set $\kappa(z) := -z\,\phi(z)$. Then

$$\kappa'(z) = -\phi(z) - z\,\phi'(z).$$

Using $\phi'(z) = -z\,\phi(z)$, we get

$$\kappa'(z) = -\phi(z) - z(-z\,\phi(z)) = \phi(z)\,(z^2 - 1).$$

Since $\phi(z) > 0$ for all $z$, the sign of $\kappa'(z)$ is the sign of $z^2 - 1$ and we obtain:

$\kappa'(z) > 0$ for $z \in (-\infty, -1)$, $\quad \kappa'(-1) = 0$ with $\kappa(-1) = \frac{1}{\sqrt{2\pi}} e^{-1/2}$,
$\kappa'(z) < 0$ for $z \in (-1, 1)$, $\quad \kappa'(1) = 0$ with $\kappa(1) = -\frac{1}{\sqrt{2\pi}} e^{-1/2}$, $\quad$ and
$\kappa'(z) > 0$ for $z \in (1, \infty)$,

which gives us the desired result. □

***Proof of Theorem 1**:* We prove for the cases: $z \ge 0$ and $z < 0$.

First, let the number of linear pieces required for the outer approximating standard normal CDF curve when $0 \le z \le 1$ be $N_1$. Observe that corresponding iterative rule in Part 1 of Algorithm 1a is such that $\sum_{i=1}^{N_1} (\check{z}_i - \check{z}_{i-1}) = 1$ where $\check{z}_0 = 0, \check{z}_{N_1} = 1$ (line 1 of Algorithm 1a). However, by Lemma 1 and since $\max_{z \in \mathbb{R}} |\Phi''(z)| = \frac{1}{\sqrt{2\pi}} e^{-1/2}$,

$$1 = \sum_{i=1}^{N_1} (\check{z}_i - \check{z}_{i-1}) \ge N_1 \sqrt{\frac{2\tau}{\frac{1}{\sqrt{2\pi}} e^{-1/2}}} = (N_1) \sqrt{2\tau\sqrt{2\pi} e^{0.5}}.$$

Hence, we have $N_1 \le \sqrt{\frac{1}{2\tau\sqrt{2\pi} e^{0.5}}}$.

For Part 2 of Algorithm 1a ($z \ge 1$), let the number of pieces be $N_2 + 1$ starting from $z = 1$ and the corresponding constructed breakpoints be

$$\check{z}_i - \check{z}_{i-1} = \sqrt{\frac{2\tau}{\phi(\check{z}_{i-1})\check{z}_{i-1}}} \qquad \forall i \in [N_2].$$

Let $I_i = \check{z}_i - \check{z}_{i-1},\ \forall i \in [N_2]$. By construction and curvature property of $\Phi(\cdot)$ from Proposition 4, clearly $I_i \geq I_{i-1}\, \forall i \in [N_2]$. Then,

$$\hat{I} := \sum_{i=1}^{N_2} I_i \geq I_{N_2} = \sqrt{\frac{2\tau\sqrt{2\pi}e^{0.5(\check{z}_{N_2-1})^2}}{\check{z}_{(N_2-1)}}} \geq \sqrt{\frac{2\tau\sqrt{2\pi}e^{0.5(\check{z}_{N_2-1})^2}}{\check{z}_R}}$$
$$\implies (\check{z}_{N_2-1})^2 \leq 2\log\frac{\hat{I}^2\check{z}_R}{2\tau\sqrt{2\pi}} \tag{22}$$

However, since Part 2 of Algorithm 1a starts from $z = 1$,

$$\check{z}_{N_2-1} = 1 + \sum_{i=1}^{N_2-1} I_i \geq 1 + (N_2 - 1)I_1 = 1 + (N_2 - 1)\sqrt{2\tau\sqrt{2\pi}e^{0.5}}.$$

Plugging this $\check{z}_{N_2-1}$ value into (22) gives us:

$$(1 + (N_2 - 1)\sqrt{2\tau\sqrt{2\pi}e^{0.5}})^2 \leq 2\log\frac{\hat{I}^2\check{z}_R}{2\tau\sqrt{2\pi}}.$$

Note that since $\hat{I} \leq \check{z}_R,\ \hat{I}^2\check{z}_R \leq \check{z}_R^3$. Hence, it follows that

$$N_2 \leq \frac{1}{\sqrt{2\tau\sqrt{2\pi}e^{0.5}}}\left(\sqrt{2\log\frac{\check{z}_R^3}{2\tau\sqrt{2\pi}}} - 1\right) + 1.$$

Thus, the total number of breakpoints we obtain using Algorithm 1a is

$$N_1 + N_2 + 1 \leq \sqrt{\frac{1}{2\tau\sqrt{2\pi}e^{0.5}}} + \frac{1}{\sqrt{2\tau\sqrt{2\pi}e^{0.5}}}\left(\sqrt{2\log\frac{\check{z}_R^3}{2\tau\sqrt{2\pi}}} - 1\right) + 2,$$

i.e., $|\mathcal{A}^R| = O\left(\max\{\frac{1}{\sqrt{\tau}}, \frac{1}{\sqrt{\tau}}\sqrt{log(\frac{\check{z}_R^3}{\tau})}\}\right)$. Since the allowed approximation error is at most $\tau$ and $\check{z}_R^3 \gg \tau$, $\sqrt{log(\frac{\check{z}_R^3}{\tau})} > 1$ and hence $\frac{1}{\sqrt{\tau}}\sqrt{log(\frac{\check{z}_R^3}{\tau})} > \frac{1}{\sqrt{\tau}}$. Therefore, $|\mathcal{A}^R| = O\left(\frac{1}{\sqrt{\tau}}\sqrt{log(\frac{\check{z}_R^3}{\tau})}\right)$.

Because of the symmetric nature of the curvature of $\Phi(\cdot)$ with respect to $z = 0$, one can derive the same complexity order for Algorithm 1b. Hence, if the choice of $\check{z}_R < \infty$ is such that $1 - \Phi(\check{z}_R) \leq \tau$ and $\check{z}_L > -\infty$ is such that $\Phi(\check{z}_L) \leq \tau$, we obtain the worst-case complexity $O\left(\frac{1}{\sqrt{\tau}}\sqrt{log(\frac{1}{\tau})}\right)$. □

***Proof of Proposition 3:*** First, consider the forward direction of the claim starting from $\underline{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$. Let $0 \leq z < \check{z}_R$. Then $z = y_3$ and $t_3 = 1$ must hold in (11). Hence, $t_1, t_2, t_4, y_1, y_2, y_4 = 0$, which leads to $g_i y_3 + g_i^0 t_3 \geq \zeta, i \in [R]$ and $1 \geq \zeta$. Thus, $\mathcal{G}(\check{\boldsymbol{z}}) \neq \emptyset$. Similarly, if $\check{z}_R \leq z$, then $z = y_4$, and $t_4 = 1$ enforcing that $t_1, t_2, t_3, y_1, y_2, y_3 = 0$. Therefore, $\Phi(\check{z}_R) \geq \zeta$. Hence, feasibility is satisfied.

Next, for $\check{z}_{-L} \leq z < 0$, there must exist an $\alpha_{\hat{i}} = 1$ for some $\hat{i} \in [L-1]_0$ implying that $t_2 = 1, y_2 = \mathfrak{z}_{\hat{i}} = z$. Therefore, $y_1, y_3, y_4, t_1, t_3, t_4 = 0$ and feasibility holds since $1 \geq \zeta$ and $h_{\hat{i}} y_2 + h^0_{\hat{i}} \geq \zeta$.

Finally for $z < \check{z}_{-L}$, $t_1 = 1$, $z = y_1$ and $t_2, t_3, t_4, y_2, y_3, y_4 = 0$ is a feasible choice that satisfies $h_L y_1 + h^0_L \geq \zeta$, $1 \geq \zeta$, $h_L y_1 + h^0_L \geq 0$. Therefore, for any $z \leq 0, \zeta \in \mathbb{R}$ such that $\underline{\Phi}(z, \check{\boldsymbol{z}}) = \max\{0, \max_{i \in [L]_0} h_i z + h^0_i\} \geq \zeta$ also satisfy the constraints defining the set $\mathcal{G}(\check{\boldsymbol{z}})$ in (11).

We now prove the other direction. We claim that any point in the set $\mathcal{G}(\check{\boldsymbol{z}})$ also satisfies $\underline{\Phi}\,(z; \check{\boldsymbol{z}}) \geq \zeta$, which follows from the definition of $\underline{\Phi}\,(z; \check{\boldsymbol{z}})$ in (10) as follows:

i) $y_4 = z \geq \check{z}_R$ with $t_4 = 1$ and $t_1, t_2, t_3, y_1, y_2, y_3 = 0$ indicates $\underline{\Phi}\,(z; \check{\boldsymbol{z}}) = \Phi(\check{z}_R) \geq \zeta$,

ii) $\check{z}_R > z = y_3 \geq 0$ with $t_3 = 1$ and $t_1, t_2, t_4, y_1, y_2, y_4 = 0$ satisfies $g_i y_3 + g^0_i \geq \zeta$ for all $i \in [R]$,

iii) $\check{z}_{-L} \leq y_2 = z < 0$ with $t_2 = 1$ and $t_1, t_3, t_4, y_1, y_3, y_4 = 0$ indicates that there exists some $\check{i} \in [L-1]$ such that $y_2 = \mathfrak{z}_{\check{i}}, \alpha_{\check{i}} = 1$ satisfying $\underline{\Phi}\,(z; \check{\boldsymbol{z}}) = h_{\check{i}} \mathfrak{z}_{\check{i}} + h^0_{\check{i}} \geq \zeta$, and

iv) finally, $y_1 = z < \check{z}_{-L}$ with $t_1 = 1$ and $t_2, t_3, y_2, y_3 = 0$ leads to $h_L y_1 + h^0_L \geq \zeta \geq 0$.

Therefore, $\zeta$ satisfy both the secant and tangent cases of definition (10) and $\underline{\Phi}(z; \check{\boldsymbol{z}}) \geq \zeta$ holds. □

## Appendix B: Inner Approximation (PWL-I)

**Lemma 5.** *Let, for any $z_1, z_2 \in \mathbb{R}$, $C(z_1, z_2) \coloneqq \max_{z \in [z_1, z_2]} |\Phi''(z)|$, $\check{\boldsymbol{z}}$ be any valid array of breakpoints as defined in* (6)*, and $\tau > 0$. Then the following holds:*

*a) If $\check{z}_i - \check{z}_{i-1} \leq 2\sqrt{\frac{2\tau}{C(\check{z}_{i-1}, \check{z}_i)}}$ for all $i \in [R]$, then $0 \leq \Phi(z) - \underline{\Phi}(z; \check{\boldsymbol{z}}) \leq \tau$ for all $z \geq 0$.*

*b) If $\check{z}_{-i+1} - \check{z}_{-i} \leq \sqrt{\frac{2\tau}{C(\check{z}_{-i}, \check{z}_{-i+1})}}$ for all $i \in [L]$, $0 \leq \Phi(z) - \underline{\Phi}(z; \check{\boldsymbol{z}}) \leq \tau$ for all $z < 0$.*

*Proof.* Proof of this Lemma is similar to that of Lemma 1. □

The next result depends on an algorithm that is the counterpart to Algorithms 1a and 1b for inner approximation.

**Theorem 3.** *Let $\check{z}_{-L}, \check{z}_R$ be such that $\max\{1 - \Phi(\check{z}_R), \Phi(\check{z}_{-L})\} \leq \tau$, where $0 < \tau \ll \min\{\check{z}_R, |\check{z}_{-L}|\}$. Then $O(\frac{1}{\sqrt{\tau}}\sqrt{log(\frac{1}{\tau})})$ breakpoints are sufficient to ensure $0 \leq \Phi(z) - \underline{\Phi}(z; \check{\boldsymbol{z}}) \leq \tau$, $\forall z \in \mathbb{R}$.*

*Proof.* Omitted because of the similarity to the proof of Theorem 1. □

## Appendix C: Additional Computational Results

Table 4: Comparison of PWL-I, PWL-O, and SAA approximations for $\varrho = 2, \varsigma = 2$ with equal mixture weight and 64,800s (18-hour) time limit

| | | PWL-I | | | PWL-O | | | | SAA | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $\theta$ | Time (sec) | Obj | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj* | $\check{\theta}^S$ |
| $n = 100$ | | | | | | | | | | | |
| 5 | 0.95 | 0 | -735 | 0.9500 | 0 | 0.0003 | 0.9496 | 0.042 | 0.7 | -844 | 0.9129 |
| | 0.99 | 1 | -735 | 0.9900 | 1 | 0.0002 | 0.9900 | 0.000 | 528 | -844 | 0.9032 |
| | 0.999 | 4 | -734 | 0.9990 | 4 | 0.0005 | 0.9990 | 0.000 | 109 | -844 | 0.9036 |
| 10 | 0.95 | 1 | -726 | 0.9500 | 2 | 0.0003 | 0.9496 | 0.042 | 64,800 | -727 | 0.8551 |
| | 0.99 | 16 | -726 | 0.9900 | 15 | 0.0002 | 0.9899 | 0.010 | 64,800 | -726 | 0.9092 |
| | 0.999 | 71 | -725 | 0.9990 | 31 | 0.0002 | 0.9990 | 0.000 | 64,800 | -726 | 0.9474 |
| 15 | 0.95 | 2 | -706 | 0.9501 | 1 | 0.0005 | 0.9499 | 0.021 | 64,800 | -707 | 0.9099 |
| | 0.99 | 10 | -705 | 0.9900 | 7 | 0.0005 | 0.9900 | 0.000 | 64,800 | -706 | 0.9416 |
| | 0.999 | 15 | -705 | 0.9990 | 66 | 0.0002 | 0.9990 | 0.000 | 64,800 | -706 | 0.9648 |
| $n = 500$ | | | | | | | | | | | |
| 5 | 0.95 | 120 | -2,455 | 0.9506 | 107 | 0.0004 | 0.9497 | 0.095 | 1,198 | -2,474 | 0.5709 |
| | 0.99 | 109 | -2,454 | 0.9900 | 141 | 0.0001 | 0.9900 | 0.000 | 64,801 | -2,473 | 0.4687 |
| | 0.999 | 147 | -2,453 | 0.9990 | 147 | 0.0002 | 0.9990 | 0.000 | 64,804 | -2,473 | 0.4958 |
| 10 | 0.95 | 293 | -1,811 | 0.9504 | 321 | 0.0007 | 0.9497 | 0.074 | 64,800 | -1,812 | 0.7463 |
| | 0.99 | 192 | -1,810 | 0.9900 | 276 | 0.0004 | 0.9900 | 0.000 | 64,810 | -1,812 | 0.8386 |
| | 0.999 | 2,021 | -1,809 | 0.9990 | 1,159 | 0.0001 | 0.9990 | 0.000 | 64,801 | -1,811 | 0.8799 |
| 15 | 0.95 | 493 | -2,338 | 0.9502 | 600 | 0.0002 | 0.9497 | 0.053 | 64,802 | -2,339 | 0.7268 |
| | 0.99 | 575 | -2,337 | 0.9900 | 714 | 0.0001 | 0.9900 | 0.000 | 64,801 | -2,338 | 0.8804 |
| $n = 1000$ | | | | | | | | | | | |
| 5 | 0.95 | 904 | -7,794 | 0.9506 | 858 | 0.0000 | 0.9496 | 0.105 | 63,161 | -7,794 | 0.3805 |
| | 0.99 | 993 | -7,793 | 0.9901 | 798 | 0.0000 | 0.9900 | 0.010 | 64,801 | -7,794 | 0.6332 |
| | 0.999 | 923 | -7,793 | 0.9990 | 876 | 0.0000 | 0.9990 | 0.000 | 64,800 | -7,794 | 0.8626 |
| 10 | 0.95 | 3,069 | -5,810 | 0.9501 | 2,651 | 0.0001 | 0.9496 | 0.053 | 64,800 | -5,811 | 0.5688 |
| | 0.99 | 2,847 | -5,810 | 0.9900 | 2,147 | 0.0001 | 0.9899 | 0.010 | 64,801 | -5,810 | 0.6370 |
| | 0.999 | 6,530 | -5,809 | 0.9990 | 3,722 | 0.0000 | 0.9990 | 0.000 | 64,801 | -5,810 | 0.7392 |
| 15 | 0.95 | 3,306 | -4,345 | 0.9748 | 3,691 | 0.0820 | 0.9801† | - | 64,800 | -4,349 | 0.7403 |
| | 0.99 | 3,427 | -4,348 | 0.9901 | 4,593 | 0.0000 | 0.9900 | 0.010 | 64,800 | -4,349 | 0.7599 |
| | 0.999 | 2,984 | -4,348 | 0.9990 | 3,098 | 0.0000 | 0.9990 | 0.000 | 64,800 | -4,349 | 0.8167 |

Table 5: PWL-I, PWL-O vs SAA for $\varrho = 2, \varsigma = 5$ with equal mixture weight

| $K$ | $\theta$ | Time (sec) | Obj | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj | $\check{\theta}^S$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 100$ | | | | | | | | | | | |
| 5 | 0.95 | 0.44 | -988 | 0.9500 | 0.337 | 0.0000 | 0.9498 | 0.021 | 0.244 | -894 | 1.0000 |
| | 0.99 | 0.45 | -988 | 0.9900 | 0.364 | 0.0000 | 0.9900 | 0.000 | 4.01 | -894 | 1.0000 |
| | 0.999 | 0.67 | -988 | 0.9990 | 0.592 | 0.0001 | 0.9989 | 0.010 | 11.56 | -894 | 1.0000 |
| 10 | 0.95 | 1.34 | -891 | 0.9500 | 3.578 | 0.0000 | 0.9499 | 0.011 | 797.5 | -891 | 0.9438 |
| | 0.99 | 2.10 | -890 | 0.9900 | 1.921 | 0.0001 | 0.9899 | 0.010 | 64,819 | -891 | 0.9758 |
| | 0.999 | 16.1 | -890 | 0.9990 | 24.87 | 0.0003 | 0.9989 | 0.010 | 16,043 | -890 | 0.9973 |
| 15 | 0.95 | 1.73 | -927 | 0.9500 | 0.937 | 0.0003 | 0.9499 | 0.011 | 5,404 | -928 | 0.9388 |
| | 0.99 | 20.1 | -924 | 0.9900 | 13.51 | 0.0001 | 0.9899 | 0.010 | 64,800 | -925 | 0.9815 |
| | 0.999 | 10.1 | -922 | 0.9990 | 25.11 | 0.0009 | 0.9989 | 0.010 | 64,801 | -924 | 0.9940 |
| $n = 500$ | | | | | | | | | | | |
| 5 | 0.95 | 117 | -2,344 | 0.9508 | 112.3 | 0.0007 | 0.9496 | 0.021 | 1,009 | -2,332 | 0.7247 |
| | 0.99 | 118 | -2,343 | 0.9902 | 154.6 | 0.0007 | 0.9899 | 0.030 | 33,944 | -2,332 | 0.7345 |
| | 0.999 | 169 | -2,341 | 0.9990 | 175.2 | 0.0003 | 0.9989 | 0.010 | 64,804 | -2,330 | 0.7553 |
| 10 | 0.95 | 284 | -2,500 | 0.9500 | 234.2 | 0.0002 | 0.9497 | 0.032 | 64,801 | -2,501 | 0.9152 |
| | 0.99 | 185 | -2,499 | 0.9900 | 251.9 | 0.0003 | 0.9899 | 0.010 | 64,804 | -2,500 | 0.9239 |
| | 0.999 | 685 | -2,498 | 0.9990 | 668 | 0.0002 | 0.9989 | 0.010 | 64,800 | -2,499 | 0.9697 |
| 15 | 0.95 | 586 | -2,187 | 0.9503 | 709.4 | 0.0008 | 0.9497 | 0.063 | 64,800 | -2,188 | 0.8253 |
| | 0.99 | 691 | -2,185 | 0.9900 | 814.1 | 0.0004 | 0.9899 | 0.010 | 64,801 | -2,187 | 0.8708 |
| | 0.999 | 19,812 | -2,184 | 0.9990 | 7,811 | 0.0001 | 0.9989 | 0.010 | 64,806 | -2,186 | 0.9031 |

*(continued on next page)*

† For these instances, chance constraint appears inactive at target satisfaction probability

| $K$ | $\theta$ | Time (sec) | Obj | $\check{\theta}^I$ | Time (sec) | %- Obj | $\check{\theta}^O$ | $\%\check{\theta}$ | Time (sec) | Obj | $\check{\theta}^S$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 1000$ | | | | | | | | | | | |
| 5 | 0.95 | 1,033 | -4,556 | 0.9507 | 1,007 | 0.0003 | 0.9498 | 0.095 | 32,337 | -4,557 | 0.5991 |
| | 0.99 | 1,197 | -4,554 | 0.9900 | 942.3 | 0.0001 | 0.9899 | 0.010 | 64,800 | -4,556 | 0.8160 |
| | 0.999 | 1,087 | -4,553 | 0.9990 | 1,008 | 0.0001 | 0.9989 | 0.010 | 57,555 | -4,555 | 0.9204 |
| 10 | 0.95 | 2,324 | -7,286 | 0.9503 | 2,414 | 0.0001 | 0.9495 | 0.084 | 64,800 | -7,286 | 0.6534 |
| | 0.99 | 1,440 | -7,285 | 0.9900 | 2,099 | 0.0000 | 0.9899 | 0.010 | 64,800 | -7,286 | 0.8194 |
| | 0.999 | 5,683 | -7,285 | 0.9990 | 2,975 | 0.0000 | 0.9989 | 0.010 | 64,802 | -7,285 | 0.8988 |
| 15 | 0.95 | 3,960 | -4,046 | 0.9501 | 3,492 | 0.0000 | 0.9500 | 0.011 | 64,800 | -4,047 | 0.8121 |
| | 0.99 | 557.2 | -4,046 | 0.9901 | 6,801 | 0.0000 | 0.9900 | 0.010 | 64,800 | -4,047 | 0.8997 |
| | 0.999 | 19,580 | -4,044 | 0.9990 | 8,935 | 0.0000 | 0.9989 | 0.010 | 64,800 | -4,046 | 0.9431 |

Table 6: Computational performance summary of PWL-I, PWL-O and SAA based approaches with equal mixture weights and 18-hour (64,800s) time limit across different hyperparameters

| | | Average Time (sec) | | | $\check{\theta}^S$ | % of Instances not solved to Optimality Gap | | | Avg. $\check{\theta}^I, \check{\theta}^O, \check{\theta}^S$ for Instances not Solved to Optimality Gap | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $K$ | $\theta$ | PWL-I | PWL-O | SAA | SAA | PWL-I | PWL-O | SAA | PWL-I | PWL-O | SAA |
| $n = 100$ | | | | | | | | | | | |
| 5 | 0.95 | 0.65 | 0.47 | 0.34 | 0.9854 | 0 | 0 | 0 | - | - | - |
| | 0.99 | 0.80 | 0.68 | 108 | 0.9838 | 0 | 0 | 0 | - | - | - |
| | 0.999 | 2.20 | 2.10 | 39 | 0.9839 | 0 | 0 | 0 | - | - | - |
| 10 | 0.95 | 2.30 | 1.70 | 34,648 | 0.9182 | 0 | 0 | 50.0 | - | - | 0.89 |
| | 0.99 | 4.50 | 3.40 | 64,804 | 0.9625 | 0 | 0 | 100 | - | - | 0.96 |
| | 0.999 | 27.4 | 33.0 | 56,674 | 0.9817 | 0 | 0 | 83.3 | - | - | 0.97 |
| 15 | 0.95 | 15.4 | 2.50 | 35,150 | 0.9201 | 0 | 0 | 40.0 | - | - | 0.89 |
| | 0.99 | 65.8 | 16.7 | 64,801 | 0.9552 | 0 | 0 | 100 | - | - | 0.95 |
| | 0.999 | 143 | 161 | 64,800 | 0.9749 | 0 | 0 | 100 | - | - | 0.97 |
| $n = 500$ | | | | | | | | | | | |
| 5 | 0.95 | 100 | 104 | 39,315 | 0.4511 | 0 | 0 | 50.0 | - | - | 0.47 |
| | 0.99 | 104 | 143 | 59,658 | 0.4745 | 0 | 0 | 83.3 | - | - | 0.42 |
| | 0.999 | 696 | 555 | 63,038 | 0.5061 | 0 | 0 | 83.3 | - | - | 0.49 |
| 10 | 0.95 | 954 | 901 | 64,800 | 0.7256 | 0 | 0 | 100 | - | - | 0.72 |
| | 0.99 | 364 | 696 | 64,806 | 0.8744 | 0 | 0 | 100 | - | - | 0.87 |
| | 0.999 | 2,227 | 1,923 | 64,801 | 0.8918 | 0 | 0 | 100 | - | - | 0.89 |
| 15 | 0.95 | 10,555 | 9502 | 64801 | 0.7393 | 0 | 0 | 100 | - | - | 0.73 |
| | 0.99 | 5,960 | 4,290 | 64,801 | 0.8490 | 0 | 0 | 100 | - | - | 0.84 |
| | 0.999 | 6,585 | 21,120 | 64,813 | 0.8924 | 0 | 16.7 | 100 | - | 0.999 | 0.89 |
| $n = 1000$ | | | | | | | | | | | |
| 5 | 0.95 | 670 | 1,136 | 53,418 | 0.4996 | 0 | 0 | 50.0 | - | - | 0.36 |
| | 0.99 | 778 | 1,355 | 64,805 | 0.7633 | 0 | 0 | 100 | - | - | 0.76 |
| | 0.999 | 952 | 1,764 | 63,593 | 0.9059 | 0 | 0 | 83.3 | - | - | 0.90 |
| 10 | 0.95 | 8,414 | 10,823 | 64,800 | 0.6878 | 0 | 0 | 100 | - | - | 0.68 |
| | 0.99 | 5,789 | 9,306 | 64,801 | 0.8001 | 0 | 3.4 | 100 | - | 0.9999 | 0.80 |
| | 0.999 | 5,926 | 7,439 | 64,803 | 0.8724 | 0 | 0 | 100 | - | - | 0.87 |
| 15 | 0.95 | 18,411 | 6,833 | 64,800 | 0.7293 | 13.3 | 6.7 | 100 | 0.9999 | 0.9939 | 0.72 |
| | 0.99 | 19,892 | 21,198 | 64,800 | 0.8548 | 27.5 | 26.7 | 100 | 0.9998 | 0.9998 | 0.85 |
| | 0.999 | 22,210 | 28,666 | 64,801 | 0.8915 | 13.3 | 36.7 | 100 | 0.9998 | 0.9996 | 0.89 |

Table 7: Effect of number of breakpoints on solution time and optimality gap for PWL-I ($n = 500$). Solution time is reported in *hours* (hr) from now onwards for ease of reading the table. Gap % represents the optimality gap (in percentage) obtained at termination from Gurobi.

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 10 | 5 | 10 | 0.99 | 0.0[5] | 0 | 0.1 | 0 | 3.3 | 0 | 4.6 | 0 |
| 10 | 5 | 10 | 0.95 | 3.8 | 0 | 4.7 | 0 | 3.8 | 0 | 4.8 | 0 |
| 10 | 5 | 15 | 0.99 | 4.7 | 0 | 6.1 | 0 | 5.0 | 0 | 6.4 | 0 |
| 10 | 10 | 15 | 0.99 | 5.4 | 0 | 6.2 | 0 | 5.3 | 0 | 6.2 | 0 |
| 10 | 10 | 15 | 0.95 | 2.1 | 0 | 6.5 | 0 | 5.0 | 0 | 6.8 | 0 |
| 10 | 15 | 10 | 0.95 | 3.1 | 0 | 3.8 | 0 | 1.1 | 0 | 1.4 | 0 |
| 10 | 20 | 15 | 0.99 | 5.1 | 0 | 7.0 | 0 | 3.1 | 0 | 3.4 | 0 |
| 10 | 20 | 15 | 0.95 | 2.1 | 0 | 3.2 | 0 | 3.5 | 0 | 5.1 | 0 |
| 15 | 5 | 10 | 0.95 | 2.5 | 0 | 3.0 | 0 | 1.8 | 0 | 2.2 | 0 |
| 15 | 5 | 15 | 0.999 | 0.1 | 0 | 0.0 | 0 | 10.7 | 0 | 13.5 | 0 |
| 15 | 10 | 10 | 0.99 | 1.1 | 0 | 2.3 | 0 | 0.8 | 0 | 1.1 | 0 |
| 15 | 15 | 10 | 0.99 | 0.0 | 0 | 3.0 | 0 | 3.4 | 0 | 5.0 | 0 |
| 15 | 20 | 10 | 0.95 | 0.0 | 0 | 0.0 | 0 | 0.1 | 0 | 0.1 | 0 |
| 20 | 2 | 15 | 0.95 | 2.1 | 0 | 3.5 | 0 | 3.2 | 0 | 3.5 | 0 |
| 20 | 5 | 15 | 0.999 | 2.1 | 0 | 18.0 | 0 | 5.3 | 0 | 14.2 | 0 |
| 20 | 10 | 10 | 0.99 | 2.1 | 0 | 4.0 | 0 | 5.6 | 0 | 6.0 | 0 |
| 20 | 10 | 15 | 0.999 | 8.8 | 0 | 11.4 | 0 | 5.2 | 0 | 11.4 | 0 |
| 20 | 15 | 10 | 0.99 | 2.2 | 0 | 3.1 | 0 | 8.3 | 0 | 10.3 | 0 |
| 20 | 15 | 15 | 0.95 | 3.1 | 0 | 17.8 | 0 | 5.0 | 0 | 12.7 | 0 |
| 20 | 20 | 10 | 0.99 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 |
| 20 | 20 | 15 | 0.95 | 5.8 | 0 | 17.9 | 0 | 3.7 | 0 | 17.0 | 0 |
| 25 | 2 | 15 | 0.95 | 3.7 | 0 | 4.0 | 0 | 4.5 | 0 | 5.9 | 0 |
| 25 | 10 | 15 | 0.999 | 5.8 | 0 | 6.8 | 0 | 5.8 | 0 | 6.8 | 0 |
| 25 | 15 | 10 | 0.99 | 1.0 | 0 | 1.5 | 0 | 9.6 | 0 | 10.8 | 0 |

Table 8: Effect of number of breakpoints on solution time and optimality gap for PWL-O ($n = 500$). Solution time is reported in *hours* (hr). Gap % represents the optimality gap (in percentage) obtained at termination from Gurobi.

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 5 | 5 | 10 | 0.95 | 0.1 | 7e-4 | 0.1 | 7e-4 | 0.1 | 4e-4 | 0.1 | 4e-4 |
| 5 | 5 | 15 | 0.999 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.2 | 2e-4 | 0.4 | 2e-5 |
| 10 | 5 | 10 | 0.95 | 0.1 | 8e-5 | 0.1 | 8e-5 | 0.1 | 1e-4 | 0.1 | 1e-4 |
| 10 | 5 | 10 | 0.99 | 0.1 | 3e-4 | 0.1 | 3e-4 | 0.1 | 3e-4 | 0.2 | 5e-5 |
| 10 | 5 | 15 | 0.99 | 0.1 | 2e-4 | 0.2 | 2e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 |
| 10 | 10 | 15 | 0.95 | 0.1 | 1e-2 | 0.1 | 1e-2 | 0.1 | 2e-3 | 0.2 | 2e-3 |
| 10 | 10 | 15 | 0.99 | 0.2 | 8e-4 | 1.8 | 2e-4 | 3.6 | 4e-4 | 4.5 | 2e-4 |
| 10 | 15 | 10 | 0.95 | 1.6 | 2e-4 | 1.9 | 2e-4 | 0.1 | 1e-4 | 0.1 | 1e-4 |
| 10 | 20 | 15 | 0.95 | 2.2 | 2e-4 | 2.7 | 2e-4 | 2.5 | 1e-4 | 3.2 | 1e-4 |
| 10 | 20 | 15 | 0.99 | 2.0 | 5e-4 | 2.0 | 2e-4 | 5.7 | 4e-4 | 5.8 | 2e-4 |
| 15 | 5 | 10 | 0.95 | 0.1 | 1e-4 | 0.1 | 1e-4 | 0.1 | 3e-5 | 0.1 | 3e-5 |
| 15 | 5 | 15 | 0.999 | 2.6 | 1e-4 | 18.0 | 2e-5 | 3.3 | 1e-4 | 3.4 | 1e-4 |
| 15 | 10 | 10 | 0.99 | 1.0 | 3e-4 | 1.7 | 3e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 |
| 15 | 15 | 10 | 0.99 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 1e-3 | 0.1 | 1e-3 |
| 15 | 20 | 10 | 0.95 | 0.1 | 1e-4 | 0.1 | 1e-4 | 0.1 | 6e-5 | 0.1 | 6e-5 |
| 20 | 2 | 15 | 0.95 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 1e-4 | 0.1 | 1e-4 |

*(continued on next page)*

[5] Since time is reported in hours, we report 0.0 when consumed time is less than a minute

Table 8 – *continued from previous page*

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 20 | 5 | 15 | 0.999 | 5.1 | 1e-4 | 18.0 | 6e-5 | 5.3 | 3e-5 | 6.3 | 1e-5 |
| 20 | 10 | 10 | 0.99 | 0.1 | 1e-4 | 0.1 | 1e-4 | 0.1 | 1e-4 | 0.1 | 1e-4 |
| 20 | 10 | 15 | 0.999 | 2.5 | 6e-5 | 12.3 | 2e-5 | 8.3 | 1e-4 | 11.4 | 2e-5 |
| 20 | 15 | 10 | 0.99 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 |
| 20 | 20 | 10 | 0.99 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 | 0.1 | 2e-4 |
| 20 | 20 | 15 | 0.95 | 4.1 | 5e-5 | 18.0 | 1e-3 | 3.1 | 3e-5 | 3.4 | 3e-5 |
| 25 | 2 | 15 | 0.95 | 0.1 | 4e-5 | 0.1 | 4e-5 | 0.1 | 1e-5 | 0.1 | 1e-5 |
| 25 | 10 | 15 | 0.999 | 18.0 | 3e-5 | 18.0 | 3e-5 | 18.0 | 3e-5 | 18.0 | 3e-5 |
| 25 | 15 | 10 | 0.99 | 1.8 | 4e-4 | 1.9 | 4e-4 | 0.1 | 3e-3 | 0.1 | 3e-3 |

Table 9: Effect of the number of breakpoints on solution time and optimality gap for PWL-I ($n = 1000$). Solution time is reported in *hours* (hr). Gap % represents the optimality gap (in percentage) obtained at termination from Gurobi.

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 2 | 10 | 15 | 0.999 | 0.3 | 0 | 0.3 | 0 | 0.3 | 0 | 0.3 | 0 |
| 2 | 15 | 15 | 0.999 | 0.2 | 0 | 0.2 | 0 | 0.3 | 0 | 0.5 | 0 |
| 2 | 20 | 15 | 0.999 | 0.2 | 0 | 0.3 | 0 | 1.0 | 0 | 1.2 | 0 |
| 10 | 2 | 15 | 0.999 | 0.3 | 0 | 0.3 | 0 | – | – | – | – |
| 10 | 10 | 15 | 0.95 | 0.2 | 0 | 0.2 | 0 | 0.3 | 0 | 0.3 | 1.1e-5 |
| 10 | 20 | 15 | 0.999 | 0.4 | 0 | 0.7 | 0 | – | – | – | – |
| 15 | 5 | 10 | 0.999 | 0.3 | 0 | 0.3 | 0 | – | – | – | – |
| 20 | 5 | 10 | 0.999 | 0.1 | 0 | 0.2 | 0 | – | – | – | – |
| 25 | 5 | 15 | 0.95 | 18.0 | 0.19 | 18.0 | 0.19 | – | – | – | – |
| 25 | 20 | 15 | 0.999 | – | – | – | – | 18.0 | 0.008 | 18 | 0.008 |

Table 10: Effect of the number of breakpoints on solution time and optimality gap for PWL-O ($n = 1000$). Solution time is reported in *hours* (hr). Gap % represents the optimality gap (in percentage) obtained at termination from Gurobi.

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 2 | 10 | 15 | 0.999 | 1.1 | 2.6e-5 | 18.0 | 4.1e-4 | 0.8 | 0.0009 | 18.0 | 2.6e-4 |
| 2 | 15 | 15 | 0.999 | 0.1 | 0.0011 | 18.0 | 0.0011 | 18.0 | 0.0075 | 18.0 | 0.0075 |
| 2 | 20 | 15 | 0.999 | 18.0 | 0.0009 | 18.0 | 0.0009 | 1.0 | 0.0009 | 18.0 | 0.0009 |
| 5 | 20 | 15 | 0.99 | 18.0 | 0.0083 | 18.0 | 0.0087 | – | – | – | – |
| 10 | 2 | 15 | 0.999 | 18.0 | 0.3056 | 18.0 | 0.3056 | – | – | – | – |
| 10 | 5 | 15 | 0.999 | 1.1 | 5.0e-5 | 1.2 | 1.7e-5 | 10.9 | 0.0002 | 18.0 | 0.0002 |
| 10 | 10 | 15 | 0.95 | 1.1 | 3.9e-5 | 8.1 | 0.0048 | 18.0 | 2.0e-5 | 0.8 | – |
| 10 | 20 | 15 | 0.999 | 0.8 | 0.0002 | 3.7 | 1.9e-5 | 18.0 | 0.0059 | 18.0 | 0.0059 |
| 15 | 5 | 10 | 0.999 | 13.3 | 1.1e-5 | 18.0 | 2.6e-5 | – | – | – | – |
| 15 | 15 | 15 | 0.999 | 1.0 | 0.0002 | 18.0 | 9.7e-5 | 18.0 | 0.0404 | 18.0 | 0.0404 |
| 20 | 5 | 10 | 0.999 | 0.5 | 2.4e-5 | 0.8 | 2.0e-5 | 12.9 | 2.1e-6 | 13.0 | 2.4e-5 |
| 20 | 10 | 15 | 0.999 | 0.8 | 5.3e-5 | 18.0 | 0.0008 | – | – | – | – |
| 20 | 20 | 15 | 0.95 | 0.9 | 0.0056 | 11.0 | 0.0056 | – | – | – | – |
| 25 | 5 | 10 | 0.999 | 0.5 | 3.0e-1 | 0.7 | 1.3e-5 | – | – | – | – |
| 25 | 5 | 15 | 0.95 | 0.2 | 0.0043 | 0.3 | 0.0043 | 18.0 | 7.4e-6 | 18.0 | 0.0043 |

*(continued on next page)*

Table 10 – *continued from previous page*

| $\varrho$ | $\varsigma$ | $K$ | $\theta$ | *Equal weight* | | | | *Different weight* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | | $\tau=(1-\theta)/2$ | | $\tau=(1-\theta)/50$ | |
| | | | | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % | Time (hr) | Gap % |
| 25 | 15 | 15 | 0.95 | 0.6 | 7.4e-6 | 18.0 | 0.0047 | 15.9 | 1.7e-6 | 18.0 | 1.7e-6 |
| 25 | 20 | 15 | 0.999 | 18.0 | 0.0009 | 18.0 | 0.0009 | 18.0 | 0.0368 | 18.0 | 0.0368 |

## Appendix D: Regression Tables

**Table 11 Regression Results for PWL-I Approximations**

| | $\theta=0.95$ | | $\theta=0.99$ | | $\theta=0.999$ | |
|---|---|---|---|---|---|---|
| | Estimate | Pr($>\|t\|$) | Estimate | Pr($>\|t\|$) | Estimate | Pr($>\|t\|$) |
| Intercept | -8.12e+3 | 2.37e-05 *** | -6.60e+3 | 0.00135 ** | -1.06e+4 | 1.23e-06 *** |
| $n$ | 4.29e+1 | 0.0154 * | 6.09e+1 | 0.00150 ** | 1.23e+2 | 2.00e-09 *** |
| $K$ | 6.27e+2 | 1.99e-08 *** | 5.79e+2 | 1.20e-06 *** | 8.46e+2 | 2.86e-11 *** |
| $\varrho$ | -4.44e+1 | 0.5046 | -5.70e+1 | 0.42722 | -8.40e+1 | 0.26805 |
| $\varsigma$ | 1.26e+2 | 0.0459 * | 6.83e+1 | 0.31215 | 2.32e+2 | 0.00124 ** |
| MIP-gap | 2.63e+2 | <2e-16 *** | 2.50e+2 | <2e-16 *** | 3.61e+2 | 0.00882 ** |
| avg $\frac{\nu^{\max}}{\nu^{\min}}$ | -1.09e+2 | 0.7093 | 5.12e+1 | 0.87082 | -1.48e+2 | 0.65528 |
| $\bar{d}_\mu$ | 2.06e+1 | 0.7439 | -1.19e+2 | 0.08258 . | -2.81e+2 | 9.85e-05 *** |
| max $\frac{\nu^{\max}}{\nu^{\min}}$ | -2.89e+0 | 0.9063 | -1.49e+1 | 0.57307 | 1.51e+1 | 0.58955 |
| $d_\mu^{\max}$ | 9.98e+0 | 0.7206 | 6.76e+1 | 0.02670 * | 1.28e+2 | 7.25e-05 *** |
| weight | 1.43e+3 | 0.0782 . | -2.11e+2 | 0.80926 | -1.20e+3 | 0.19495 |
| Signif. codes: <0.001 '***' 0.001 '**' 0.01 '*' 0.05 '.' | | | | | | |

**Table 12 Regression Results for PWL-O Approximations**

| | $\theta=0.95$ | | $\theta=0.99$ | | $\theta=0.999$ | |
|---|---|---|---|---|---|---|
| | Estimate | Pr($>\|t\|$) | Estimate | Pr($>\|t\|$) | Estimate | Pr($>\|t\|$) |
| Intercept | -6.76e+03 | 3.12e-05 *** | -1.00e+04 | 1.26e-06 *** | -1.54e+04 | 4.03e-08 *** |
| $n$ | 5.32e+01 | 3.97e-04 *** | 9.19e+01 | 1.86e-06 *** | 1.33e+02 | 3.51e-07 *** |
| $K$ | 4.72e+02 | 5.49e-07 *** | 7.89e+02 | 4.75e-11 *** | 1.41e+03 | <2e-16 *** |
| $\varrho$ | -1.57e+01 | 7.80e-01 | -4.65e+01 | 5.17e-01 | -1.65e+02 | 9.02e-02 . |
| $\varsigma$ | 1.39e+02 | 9.02e-03 ** | 1.75e+02 | 9.44e-03 ** | 3.20e+02 | 5.07e-04 *** |
| MIP-gap | 2.77e+02 | <2e-16 *** | 2.40e+02 | <2e-16 *** | 2.13e+02 | 1.34e-10 *** |
| avg $\frac{\nu^{\max}}{\nu^{\min}}$ | 1.05e+00 | 9.97e-01 | -9.94e+01 | 7.53e-01 | -1.58e+02 | 7.13e-01 |
| $\bar{d}_\mu$ | -3.72e+01 | 4.83e-01 | 2.74e+00 | 9.68e-01 | -2.10e+02 | 2.39e-02 * |
| max $\frac{\nu^{\max}}{\nu^{\min}}$ | -5.46e+00 | 7.93e-01 | -1.86e+00 | 9.44e-01 | 1.26e+01 | 7.27e-01 |
| $d_\mu^{\max}$ | 1.99e+01 | 3.98e-01 | -2.95e+00 | 9.23e-01 | 1.02e+02 | 1.31e-02 * |
| weight | 9.32e+02 | 1.74e-01 | -7.68e+02 | 3.80e-01 | -5.68e+02 | 6.32e-01 |
| Signif. codes: <0.001 '***' 0.001 '**' 0.01 '*' 0.05 '.' | | | | | | |