

Single and joint chance-constrained optimization with continuous distributions

Daniela Bernhard, Frauke Liers, Michael Stingl

Abstract The input parameters of an optimization problem are often affected by uncertainties. Chance constraints are a common way to model stochastic uncertainties in the constraints. Typically, algorithms for solving chance-constrained problems require convex functions or discrete probability distributions. In this work, we go one step further and allow non-convexities as well as continuous distributions. We propose a gradient-based approach to approximately solve joint chance-constrained models. We approximate the original problem by smoothing indicator functions. Then, the smoothed chance constraints are relaxed by penalizing their violation in the objective function. The approximation problem is solved with the Continuous Stochastic Gradient method that is an enhanced version of the stochastic gradient descent and has recently been introduced in the literature. We present a convergence theory for the smoothing and penalty approximations. Under very mild assumptions, our approach is applicable to a wide range of chance-constrained optimization problems. As an example, we illustrate its computational efficiency on difficult practical problems arising in the operation of gas networks. The numerical experiments demonstrate that the approach quickly finds nearly feasible solutions for joint chance-constrained problems with non-convex constraint functions and continuous distributions, even for realistically-sized instances.

1 Introduction

In this work, we consider joint chance-constrained optimization problems of the form

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & \mathbb{P}_{\boldsymbol{\delta} \sim \zeta}(g_j(\mathbf{x}, \boldsymbol{\delta}) \geq 0 \ \forall j = 1, \dots, \mathcal{J}) \geq p \\ & \mathbf{x} \in \mathcal{X} \end{aligned}$$

with $g_0 : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_j : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ for $j = 1, \dots, \mathcal{J}$ and $\mathcal{X} \subseteq \mathbb{R}^n$. The functions g_j depend on the optimization variables \mathbf{x} and stochastic parameters $\boldsymbol{\delta}$ that follow a probability distribution ζ over a set $\Delta \subseteq \mathbb{R}^d$. The research on chance-constrained problems dates back to Charnes et al. [1958], where single chance constraints are introduced. Joint chance constraints were first analyzed in Miller and Wagner [1965]. For a general overview on chance-constrained optimization we refer to Prékopa [2003], Shapiro et al. [2021], Nemirovski [2012], Bai et al. [2012] and references therein.

In this work, we consider joint chance-constrained problems with continuous distributions. Our contribution consists in the derivation of a practically very efficient solution approach for models with non-convex constraints g_j . In addition, it is not necessary to perform Monte-Carlo sampling. In our method, we approximate the chance-constrained problem by smoothing the indicator functions in the probability distribution. Violation of the smoothed constraint is penalized in the objective function. The approximation problem is solved with the Continuous Stochastic Gradient method introduced in Pflug et al. [2020]. On the theoretical side, we derive a convergence theory for the smoothing and penalization

approximations, respectively. On the practical side, we illustrate that our algorithm solves non-convex joint chance-constrained problems with continuous distributions very quickly.

In general, chance constraints can be modeled with discrete and continuous distributions. For discrete ones, the model can be reformulated as a standard mixed-integer non-linear optimization problem (MINLP) by introducing additional binary variables (Ruszczynski [2002]). Nevertheless, these reformulations are often difficult to solve already for small instances since they typically include non-linear functions together with discrete variables. Furthermore, the problem size increases with an increasing number of scenarios in the distribution. Contributions Adam and Branda [2016] and Adam et al. [2020] present a regularized version of the linear relaxation for the MINLP reformulation by enlarging the feasible set based on methods for complementarity constraints. Furthermore, the first paper proposes an iterative algorithm that uses necessary optimality conditions of the regularized problem. Under some assumptions, the solution converges to a stationary point of the original problem. In the second paper, the regularized problem is solved by adding Benders cutting planes. In Beraldi and Ruszczyński [2002], the authors introduce a tailored Branch-and-Bound approach to solve the original MINLP reformulation. In Branda [2013], the author provides an asymptotically equivalent penalty model for the original chance-constrained problem. All these approaches are applicable to general functions g_j . For specific chance constraints, for example, linear functions g_j , see exemplarily Luedtke et al. [2010], Luedtke [2014], Ahmed and Xie [2018] and Bai et al. [2021].

For continuous distributions tractable reformulations for chance constraints are in general not available. Even the check for feasibility of a given vector is often challenging. To avoid evaluating the chance constraint exactly, one often approximates the continuous distribution with a discrete one. For example, one can use scenario approximations (e.g. Calafiore and Campi [2005], Nemirovski and Shapiro [2006], Yang and Sutanto [2019]) or sample approximation techniques (e.g. Ahmed and Shapiro [2008], Luedtke and Ahmed [2008], Pagnoncelli et al. [2009], Porras et al. [2023]). Additionally, Branda [2012b] studies sample approximation techniques for several chance constraints and Peña-Ordieres et al. [2020] use the approach for quantile reformulations. In Sun et al. [2014] sample approximation is applied to a conditional Value-at-Risk approximation to the chance-constrained problem. Although these approaches allow using MINLP techniques to approximate chance-constrained problems with continuous distributions, the problem size increases with an increasing number of scenarios in the approximating discrete distribution. Naturally, the question arises how many scenarios are necessary to obtain a nearly feasible solution for the original problem. More specifically, one can ask how large the sample size has to be chosen to guarantee feasibility for the original chance constraint with a high probability. Most of the aforementioned papers provide upper bounds for the sample size to guarantee this property. All these bounds grow at least linearly in the dimension of \mathcal{X} . Furthermore, the necessary number of samples increases with increasing probability level p . As a result, discrete approximations of continuous distributions can lead to very large and complex problems in practice if feasibility for the original chance constraint is required. Therefore, there also exist approaches that deal directly with the continuous distributions. We mention some of these approaches in the following paragraphs.

Another challenge arising for chance constraints with continuous distributions is non-convexity. Even if all constraint functions g_j are convex, the chance constraint is in general non-convex. Therefore, a typical solution approach are convex approximations like the Bernstein approximation (Nemirovski and Shapiro [2007]), the conditional Value-at-

Risk approximation (Rockafellar et al. [2000]) or robust safe approximations (e.g. Ben-Tal and Nemirovski [1998], Ben-Tal and Nemirovski [2000], Bertsimas and Sim [2004], Yuan et al. [2017]). Concerning robust approximations, Li and Li [2015] provide a study about the optimal uncertainty set to approximate chance-constrained problems and Chen et al. [2010] discuss different robust approximations and connections to conditional Value-at-Risk bounds. Furthermore, a combination of robust optimization and a scenario approach is presented in Margellos et al. [2014]. The method is enhanced for robust approximation of joint chance constraints with decision-dependent uncertainties with an application in electricity networks in Aigner et al. [2022] and Aigner et al. [2023].

Hong et al. [2011] propose an approximation to the joint chance constraint with DC (difference of convex) functions and show convergence to KKT-points under some additional assumptions. Furthermore, the authors describe a sequential convex approximation method to solve the occurring problems. Since the obtained approximation contains difficult non-smooth functions, Shan et al. [2014] and Shan et al. [2016] expand the analysis to smooth functions and show convergence to KKT-points for certain asymptotic regimes. Furthermore, Hu et al. [2013] study smooth DC approximations with logarithm-sum-exponential smoothing functions and solve the approximations with a Monte-Carlo method. Ren et al. [2022] use an approximation of the indicator function with sigmoid functions that also provides convergence to KKT-points. In Cao and Zavala [2020], the authors approximate the Value-at-Risk with sigmoid functions to provide an approximation for chance-constrained problems and solve the resulting problem with a sample average approximation. Geletu et al. [2017] propose a smooth approximation consisting of an inner and outer approximation for single chance constraints that asymptotically converges to the optimal solution of the original problem. The last two methods work for non-convex constraint functions g_j , whereas the other methods described in this paragraph are restricted to convex functions.

Another possibility to handle chance-constrained problems are penalty formulations like in Branda and Dupačová [2012] or Branda [2012c]. Kannan and Luedtke [2021] combine smoothing approaches and penalization to approximate the efficient frontier between optimal solutions and violations of the chance constraint for different probability levels. To solve the approximation problems they apply a projected stochastic subgradient method.

In the following, we consider a very general type of joint chance-constrained problems without convexity assumptions on the constraint functions g_j . We derive a solution method that combines smoothing and penalization techniques. Furthermore, we deal directly with the continuous distributions and do not approximate them with discrete ones. First, we reformulate the joint chance constraint to a single chance constraint in the spirit of Branda [2012a]. After smoothing the indicator function to obtain continuously differentiable expressions, we penalize the violation of the smoothed constraint in the objective function. To solve the approximation problem we apply an advanced version of the stochastic gradient descent algorithm, called Continuous Stochastic Gradient method (CSG) that is introduced in Pflug et al. [2020], Grieshammer et al. [2023a] and Grieshammer et al. [2023b]. If the gradients of the constraint functions and the objective function are locally Lipschitz continuous, the CSG method converges to stationary points of the approximation problem for suitably chosen step sizes. Furthermore, we present a convergence analysis for the smoothing and the penalty approximations, respectively. To illustrate the efficiency of our approach, we apply it to non-convex chance-constrained gas transport problems. Our numerical results show that we can compute feasible solutions to the original problem by solving the approximation problem with the CSG method. Our approach is applicable to

very general problems since we deal directly with continuous distributions of joint chance constraints and do neither require convexity for the functions g_j nor special conditions for the continuous distributions.

We start this paper by presenting the approximations of the joint chance-constrained problem that allow the application of the CSG method in Chapter 2. Chapter 3 provides convergence results with increasing smoothing and penalty parameters. In Chapter 4 we briefly summarize the CSG method and in Chapter 5 we discuss some algorithmic improvements. The applicability of our method is illustrated in Chapter 6 with a small analytical example and detailed computational results for an application to chance-constrained Nomination Validation problems in gas networks. The paper closes with a short summary of the results and an outlook to future research questions.

2 Reformulations and approximations for joint chance-constrained problems

In this section, we reformulate and approximate the joint chance-constrained problem presented in Chapter 1 in order to apply the CSG method. First, we formulate the joint chance constraint as a single chance constraint. A similar reformulation is presented in Branda [2012a]. The joint constraint $\mathbb{P}_{\delta \sim \zeta}(g_j(\mathbf{x}, \boldsymbol{\delta}) \geq 0 \forall j = 1, \dots, \mathcal{J}) \geq p$ is equivalent to

$$\mathbb{P}_{\delta \sim \zeta} \left(\underbrace{\sum_{j=1}^{\mathcal{J}} -\min\{0, g_j(\mathbf{x}, \boldsymbol{\delta})\}^2}_{=:g(\mathbf{x}, \boldsymbol{\delta})} \geq 0 \right) \geq p.$$

So, the joint chance-constrained problem can be equivalently written as

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & \mathbb{P}_{\delta \sim \zeta}(g(\mathbf{x}, \boldsymbol{\delta}) \geq 0) \geq p \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (P)$$

We see that g is always non-positive and continuously differentiable if g_j is continuously differentiable for each j . Obviously, transforming the joint chance constraint to a single one complicates the expressions. But we will see in Chapter 4 that the CSG method only requires continuously differentiable functions with Lipschitz continuous gradients. This property is transferred from the functions g_j to g , and therefore the reformulation works for our approach. Additionally, the non-convexity of the minimum, even for linear functions g_j , is not a problem for our method since it is especially designed to handle non-convex functions.

Next, we rewrite the probability with indicator functions χ , where $\chi_{[0, \infty)}(y) = 1$ for $y \geq 0$ and $\chi_{[0, \infty)}(y) = 0$ for $y < 0$, and obtain

$$\mathbb{P}_{\delta \sim \zeta}(g(\mathbf{x}, \boldsymbol{\delta}) \geq 0) = \int_{\Delta} \chi_{[0, \infty)}(g(\mathbf{x}, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}).$$

Here, μ denotes the probability measure associated with the distribution ζ . The following example demonstrates that even for simple functions g_j the probability function is not necessarily continuous, and therefore also not differentiable.

Example 1. Let $g(x, \delta) = x \cdot \delta$, $\mathcal{X} = [0, 1]$ and $\Delta = [-2, -1]$. Then, for every continuous probability distribution ζ the probability function \mathbb{P} is not continuous in $x = 0$ because we have $\mathbb{P}_{\delta \sim \zeta}(g(\epsilon, \delta) \geq 0) = 0$ for all $\epsilon > 0$ but for $x = 0$ we obtain $\mathbb{P}_{\delta \sim \zeta}(g(x, \delta) \geq 0) = 1$.

To obtain differentiability we approximate the indicator function with a continuously differentiable function, called smoothing function, with the following properties:

Definition 1 (Smoothing function). A function $s : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ is called a smoothing function for the indicator function $\chi_{[0, \infty)}$ if

1. $s(\cdot, \beta)$ is continuously differentiable and monotonically increasing for all $\beta > 0$,
2. $\lim_{y \rightarrow -\infty} s(y, \beta) = 0$ for all $\beta > 0$,
3. $s(0, \beta) = 1$ for all $\beta > 0$ and
4. $\lim_{\beta \rightarrow +\infty} s(y, \beta) = 0$ for all $y < 0$.

We call β the smoothing parameter. We note that there is no condition for s for positive inputs y except the first one. Since the argument of the indicator function is always non-positive for our modeling, the behavior of the smoothing function for positive y is not relevant. Nevertheless, we define the function on \mathbb{R} since positive arguments occur in the algorithmic improvements in Chapter 5. The first and second property guarantee that the smoothing function approaches the indicator function from above if the first argument becomes small. Owing to the third property, s always equals the indicator function if the condition in the chance constraint is fulfilled, this means $g(\mathbf{x}, \boldsymbol{\delta}) \geq 0$. Property 4 ensures that the smoothing function approximates the indicator function better and better for increasing smoothing parameters β . Replacing the indicator function by a smoothing function is obviously not exact, but there is a connection between feasibility for the original chance constraint and the smoothed version.

Theorem 1. Let \mathbf{x}^* fulfil the smoothed chance constraint

$$\int_{\Delta} s(g(\mathbf{x}^*, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) \geq p$$

for a smoothing function s and a smoothing parameter $\beta > 0$. Then, there exists for every $\tilde{\epsilon} > 0$ an $\epsilon > 0$ with $\mathbb{P}_{\delta \sim \zeta}(g(\mathbf{x}^*, \boldsymbol{\delta}) \geq -\epsilon) \geq p - \tilde{\epsilon}$.

Proof. Since $\lim_{g(\mathbf{x}, \boldsymbol{\delta}) \rightarrow -\infty} s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) = 0$, there exists for every $\tilde{\epsilon} > 0$ an $\epsilon > 0$ with $s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \leq \tilde{\epsilon}$ for $g(\mathbf{x}, \boldsymbol{\delta}) < -\epsilon$. Now, let $\Delta^- := \{\boldsymbol{\delta} \in \Delta : g(\mathbf{x}^*, \boldsymbol{\delta}) < -\epsilon\}$ and $\Delta^+ := \Delta \setminus \Delta^-$. Then, we obtain with the properties of the smoothing function

$$\begin{aligned} p &\leq \int_{\Delta} s(g(\mathbf{x}^*, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) \leq \int_{\Delta^-} \tilde{\epsilon} \mu(d\boldsymbol{\delta}) + \int_{\Delta^+} 1 \mu(d\boldsymbol{\delta}) \\ &= \tilde{\epsilon} \cdot \mu(\Delta^-) + \int_{\Delta} \chi_{[-\epsilon, \infty)}(g(\mathbf{x}^*, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}) \leq \tilde{\epsilon} + \int_{\Delta} \chi_{[-\epsilon, \infty)}(g(\mathbf{x}^*, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}) \\ &= \tilde{\epsilon} + \mathbb{P}_{\delta \sim \zeta}(g(\mathbf{x}^*, \boldsymbol{\delta}) \geq -\epsilon), \end{aligned}$$

which implies the desired inequality. □

The Theorem shows that feasible solutions for the smoothed chance constraint fulfil a relaxed feasibility for the original one, where the probability level might be smaller and the original condition $g(\mathbf{x}, \boldsymbol{\delta}) \geq 0$ might be violated by a small quantity $\epsilon > 0$. We can avoid the inexactness in the probability level by solving the smoothed problem for a higher probability level $p + \tilde{\epsilon}$. In practical computations, it is relevant not to increase the new probability level more than necessary, also taking into account that ϵ increases for decreasing $\tilde{\epsilon}$ and vice versa. In the computational results in Chapter 6.2 we show that it is not difficult to determine appropriate values for a real energy-management application and large instances.

In total, the original joint chance-constrained problem (P) can be approximated with

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & \int_{\Delta} s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) \geq p \\ & \mathbf{x} \in \mathcal{X} \end{aligned} \quad (P_s)$$

for a smoothing parameter $\beta > 0$. To use the CSG method we have to project vectors onto the feasible set. To obtain an easier feasible set, we penalize the violation of the smoothed chance constraint in the objective function. For tractability, we impose the mild assumption that the projection onto \mathcal{X} needs to be algorithmically tractable:

Assumption 1. *The projection onto \mathcal{X} is algorithmically tractable.*

The resulting penalized and smoothed problem (P_{app}) reads for a penalty parameter $\lambda > 0$ and a smoothing parameter $\beta > 0$

$$\min_{\mathbf{x} \in \mathcal{X}} g_0(\mathbf{x}) + \frac{\lambda}{2} \max \left\{ 0, p - \int_{\Delta} s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) \right\}^2.$$

Assuming that the objective function g_0 and the constraint functions g_j for $j = 1, \dots, \mathcal{J}$ are continuously differentiable in \mathbf{x} , the objective function of (P_{app}) is also continuously differentiable.

3 Theory for smoothing and penalization

In this section, we present a convergence theory for the smoothing and penalty approximations presented in Chapter 2. This theory demonstrates that the approximation problem (P_{app}) provides a good approximation to the original one. We consider sequences of increasing smoothing and penalty parameters and show that the solutions of the smoothed and penalized problems, respectively, converge to solutions of the original one under certain conditions. In the following, we denote subsequences that converge to an accumulation point of the whole sequence always with the indices of the whole sequence to unburden notation.

3.1 Smoothing

We start our convergence analysis with the smoothing approximation. We recall that the original problem (P) can be written as

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & \int_{\Delta} \chi_{[0,\infty)}(g(\mathbf{x}, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}) \geq p \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (P_o)$$

The smoothed version of (P_o) for a smoothing parameter $\beta > 0$ and a smoothing function s is given by (P_s) . We define the sets $\Delta_1(\mathbf{x}) = \{\boldsymbol{\delta} \in \Delta : g(\mathbf{x}, \boldsymbol{\delta}) < 0\}$ and $\Delta_2(\mathbf{x}) = \{\boldsymbol{\delta} \in \Delta : g(\mathbf{x}, \boldsymbol{\delta}) = 0\}$. We note that Δ is the disjoint union of $\Delta_1(\mathbf{x})$ and $\Delta_2(\mathbf{x})$ for each \mathbf{x} . With these definitions we show that a sequence of feasible points for the smoothed problems with $\beta^k \rightarrow \infty$ converges to a feasible point for the original one.

Theorem 2. *Let $\{\mathbf{x}^k\}$ be a sequence of feasible points for the smoothed problems (P_s^k) with smoothing parameters $\beta^k > 0$ and $\beta^k \rightarrow \infty$. Furthermore, let g be continuous in \mathbf{x} and \mathcal{X} be compact. Then, every accumulation point $\bar{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ is feasible for the original problem (P_o) .*

Proof. Let $\bar{\mathbf{x}}$ be an accumulation point of $\{\mathbf{x}^k\}$. Owing to the feasibility of \mathbf{x}^k for (P_s^k) we have $\mathbf{x}^k \in \mathcal{X}$ for all $k \in \mathbb{N}$. Since \mathcal{X} is compact, this implies $\bar{\mathbf{x}} \in \mathcal{X}$. It remains to be shown that the chance constraint is fulfilled. Because \mathbf{x}^k is feasible for (P_s^k) we have for all $k \in \mathbb{N}$

$$\int_{\Delta} s(g(\mathbf{x}^k, \boldsymbol{\delta}), \beta^k) \mu(d\boldsymbol{\delta}) \geq p.$$

Since the integral term on the left-hand side is bounded, there exists a convergent subsequence $\{k_l\}$ and we obtain

$$\lim_{k_l \rightarrow \infty} \int_{\Delta} s(g(\mathbf{x}^{k_l}, \boldsymbol{\delta}), \beta^{k_l}) \mu(d\boldsymbol{\delta}) \geq p.$$

Now, we consider $\boldsymbol{\delta} \in \Delta_1(\bar{\mathbf{x}})$. Then, there exists an $\epsilon^\delta > 0$ with $g(\bar{\mathbf{x}}, \boldsymbol{\delta}) + \epsilon^\delta < 0$. Since $\mathbf{x}^{k_l} \rightarrow \bar{\mathbf{x}}$ and g is continuous there also exists a $k^\delta \in \mathbb{N}$ with $g(\mathbf{x}^{k_l}, \boldsymbol{\delta}) \leq g(\bar{\mathbf{x}}, \boldsymbol{\delta}) + \epsilon^\delta$ for all $k_l > k^\delta$. Together with the monotonicity and non-negativity of s this implies for all $k_l > k^\delta$

$$0 \leq s(g(\mathbf{x}^{k_l}, \boldsymbol{\delta}), \beta^{k_l}) \leq s(g(\bar{\mathbf{x}}, \boldsymbol{\delta}) + \epsilon^\delta, \beta^{k_l}).$$

Since $g(\bar{\mathbf{x}}, \boldsymbol{\delta}) + \epsilon^\delta < 0$, the left and right expression in this inequality converge to zero, and therefore the sandwich lemma implies that $f_{k_l}(\boldsymbol{\delta}) := s(g(\mathbf{x}^{k_l}, \boldsymbol{\delta}), \beta^{k_l})$ converges pointwise to zero on $\Delta_1(\bar{\mathbf{x}})$. Furthermore, we have $|f_{k_l}(\boldsymbol{\delta})| \leq 1$ for all $\boldsymbol{\delta} \in \Delta_1(\bar{\mathbf{x}})$ and all $k_l \in \mathbb{N}$, which implies that f_{k_l} is dominated by the constant and integrable function 1. So, Lebesgue's dominated convergence theorem gives

$$\lim_{k_l \rightarrow \infty} \int_{\Delta_1(\bar{\mathbf{x}})} s(g(\mathbf{x}^{k_l}, \boldsymbol{\delta}), \beta^{k_l}) \mu(d\boldsymbol{\delta}) = \int_{\Delta_1(\bar{\mathbf{x}})} 0 \mu(d\boldsymbol{\delta}).$$

In total, we obtain

$$\begin{aligned} p &\leq \lim_{k_l \rightarrow \infty} \int_{\Delta} s(g(\mathbf{x}^{k_l}, \boldsymbol{\delta}), \beta^{k_l}) \mu(d\boldsymbol{\delta}) \leq \int_{\Delta_1(\bar{\mathbf{x}})} 0 \mu(d\boldsymbol{\delta}) + \int_{\Delta_2(\bar{\mathbf{x}})} 1 \mu(d\boldsymbol{\delta}) \\ &= \int_{\Delta} \chi_{[0,\infty)}(g(\bar{\mathbf{x}}, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}). \end{aligned}$$

So, $\bar{\mathbf{x}}$ is feasible for (P_o) . □

Furthermore, the feasible set of (P_o) is a subset of the feasible set of each smoothed problem.

Lemma 1. *Let \mathbf{x} be feasible for the original problem (P_o) . Then, \mathbf{x} is also feasible for the smoothed problem (P_s) with an arbitrary smoothing parameter $\beta > 0$.*

Proof. Since \mathbf{x} is feasible for the original problem, we have $\mathbf{x} \in \mathcal{X}$ and

$$\int_{\Delta} \chi_{[0,\infty)}(g(\mathbf{x}, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}) \geq p.$$

So, we obtain for all $\beta > 0$

$$\int_{\Delta} s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) \geq \int_{\Delta_1(\mathbf{x})} 0 \mu(d\boldsymbol{\delta}) + \int_{\Delta_2(\mathbf{x})} 1 \mu(d\boldsymbol{\delta}) = \int_{\Delta} \chi_{[0,\infty)}(g(\mathbf{x}, \boldsymbol{\delta})) \mu(d\boldsymbol{\delta}) \geq p.$$

Therefore \mathbf{x} is feasible for all smoothed problems. \square

Using these two results we obtain the convergence of globally optimal solutions for the smoothed problems to globally optimal solutions for the original problem.

Theorem 3. *Let $\{\mathbf{x}^k\}$ be a sequence of globally optimal solutions to the smoothed problems (P_s^k) with positive smoothing parameters β^k that converge to infinity. Let furthermore g_0 and g be continuous and \mathcal{X} be compact. Then, every accumulation point $\bar{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ is a globally optimal solution to the original problem (P_o) .*

Proof. Let $\bar{\mathbf{x}}$ be an accumulation point of $\{\mathbf{x}^k\}$. Since \mathbf{x}^k is feasible for (P_s^k) for all $k \in \mathbb{N}$, Theorem 2 gives the feasibility of $\bar{\mathbf{x}}$ for the original problem. Now, assume for a contradiction that $\bar{\mathbf{x}}$ is not optimal for (P_o) . Then, there exists a vector \mathbf{x}^* that is feasible for the original problem with $g_0(\mathbf{x}^*) < g_0(\bar{\mathbf{x}})$. Since g_0 is continuous this implies $g_0(\mathbf{x}^*) < \lim_{k \rightarrow \infty} g_0(\mathbf{x}^k)$. So, there exists a $\tilde{k} \in \mathbb{N}$ such that for all $k > \tilde{k}$ we have $g_0(\mathbf{x}^*) < g_0(\mathbf{x}^k)$. Now, let $l > \tilde{k}$. Owing to Lemma 1, \mathbf{x}^* is feasible for the smoothed problem with parameter β^l . This is a contradiction to the global optimality of \mathbf{x}^l for (P_s^l) . So, $\bar{\mathbf{x}}$ is globally optimal for the original problem. \square

Since a smoothed chance constraint is in general not convex, globally optimal solutions to the smoothed problems are difficult to compute, which means that the above Theorem is in particular of theoretical interest. An alternative are KKT-points. For a sequence of KKT-points, we typically cannot expect convergence to a globally optimal solution of (P_o) . Since the original chance constraint is in general neither differentiable, nor convex, nor locally Lipschitz continuous, we can also not expect that a typical version of a sub-differential, like Clarke's subdifferential (Clarke [1990]), exists for the chance constraint, and therefore it is in general not possible to define generalized KKT-conditions for (P_o) . In the literature, it is sometimes assumed that the probability function is locally Lipschitz continuous (see e.g. Shan et al. [2014], Shan et al. [2016]) or even continuously differentiable (see e.g. Hong et al. [2011]). In this case, generalized KKT-conditions for (P_o) can be defined and the authors in the aforementioned publications could show for their special smoothing functions that a series of KKT-points for the smoothed problems converges to a generalized KKT-point of the original problem under some additional conditions. There also exist assumptions on the distributions and the constraint functions g_j that guarantee

the existence of gradients or at least subgradients for the original chance constraint. For example, the authors in Hantoute et al. [2019] discuss the existence for Gaussian distributions and constraint functions g_j that are convex in the uncertain parameters. Since we want to avoid restrictive conditions on the distributions and constraint functions in this work, we do not consider generalized KKT-points for (P_o) in our theoretical analysis.

We note that all results in this section also apply for problems with multiple joint chance constraints. We use this result later for practical adjustments of our approach. In the following, we do not consider sequences of increasing smoothing parameters β , but restrict to a fixed value. In Chapter 6.3 we illustrate that choosing β large already gives a good approximation to the chance constraint, and therefore the restriction to fixed values is acceptable in practice.

3.2 Penalization

It remains to discuss the penalty approximation. We consider the general optimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & f(\mathbf{x}) \leq 0 \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (P_p)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ denotes an arbitrary function, for example the smoothed chance constraint for a fixed smoothing parameter β , and \mathcal{X} fulfils Assumption 1. We penalize the violation of the constraint $f(\mathbf{x}) \leq 0$ in the objective function with a quadratic penalty term. For a penalty parameter $\lambda > 0$ we obtain the penalty problem (P_λ)

$$\min_{\mathbf{x} \in \mathcal{X}} g_0(\mathbf{x}) + \frac{\lambda}{2} \max\{0, f(\mathbf{x})\}^2 =: \min_{\mathbf{x} \in \mathcal{X}} G(\mathbf{x}, \lambda).$$

We note that only one constraint is penalized. This approach is based on the necessity to calculate projections onto the feasible set in the CSG method. Owing to Assumption 1 the projection onto \mathcal{X} is computationally tractable, and therefore it is not necessary to penalize the constraints in \mathcal{X} . Unfortunately, this partial penalization prevents using classical convergence theory for penalty problems directly. For completeness, we adapt the well-known results in this field to our special case.

We show that a sequence of optimal solutions for the penalty problems (P_λ) with increasing penalty parameters $\lambda^k \rightarrow \infty$ converges to an optimal solution for (P_p) .

Theorem 4. *Let g_0 and f be continuous, \mathcal{X} be closed and let $\lambda^k > 0$ be a sequence of strictly monotonically increasing penalty parameters that converges to infinity. Let furthermore $\{\mathbf{x}^k\}$ be a sequence of globally optimal solutions for the penalty problems with penalty parameters λ^k . Then, every accumulation point $\bar{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ is globally optimal for the original problem (P_p) .*

Proof. Let $\bar{\mathbf{x}}$ be an accumulation point of the sequence $\{\mathbf{x}^k\}$. Owing to the feasibility of \mathbf{x}^k for the corresponding penalty problem, we have $\mathbf{x}^k \in \mathcal{X}$ for all $k \in \mathbb{N}$. Since \mathcal{X} is closed this implies $\bar{\mathbf{x}} \in \mathcal{X}$. The optimality follows directly from Theorem 2.1 in Freund [2004] since the penalty function $p(\mathbf{x}) = \frac{1}{2} \max\{0, f(\mathbf{x})\}^2$ is continuous. \square

This result also holds for problems with multiple penalized constraints (see Freund [2004]). Unfortunately, the Theorem is often not applicable in practice, since computing globally optimal solutions to the penalty problems is not possible in general. In Chapter 4 we discuss that the CSG method is able to compute stationary points for the penalty problems under some mild assumptions. A point \mathbf{x} is called stationary for (P_λ) if and only if there exists a scalar $t > 0$ with

$$\mathcal{P}_{\mathcal{X}}(\mathbf{x} - t\nabla_1 G(\mathbf{x}, \lambda)) = \mathbf{x},$$

where $\mathcal{P}_{\mathcal{X}}$ denotes the Euclidean projection onto \mathcal{X} . In the following, we assume that the set \mathcal{X} is given by an n -dimensional box.

Assumption 2. *There exist vectors $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$ with $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in [\mathbf{l}, \mathbf{u}]\}$ and $l_i < u_i$ for all $i = 1, \dots, n$.*

We note that the projection onto boxes can be calculated separately for each coordinate. The next Theorem shows under which conditions a series of stationary points for the penalty problems converges to a KKT-point of the original one.

Theorem 5. *Let $\{\mathbf{x}^k\}$ be a sequence of stationary points for (P_{λ^k}) with $\lambda^k \rightarrow \infty$, let g_0 and f be continuously differentiable and let $\bar{\mathbf{x}}$ be an accumulation point of $\{\mathbf{x}^k\}$. Furthermore, let Assumption 2 be true. If $\bar{\mathbf{x}}$ is infeasible for the original problem (P_p) , it is a KKT-point of*

$$\min_{\mathbf{x} \in [\mathbf{l}, \mathbf{u}]} f(\mathbf{x}). \quad (P_f)$$

Otherwise, $\bar{\mathbf{x}}$ is feasible for (P_p) and is either a KKT-point of (P_p) or fulfils at least one of the following conditions:

- A) $\nabla f(\bar{\mathbf{x}}) = 0$.
- B) *There exists an index $i \in \{1, \dots, n\}$ with $\bar{x}_i \in (l_i, u_i)$, $\nabla_i f(\bar{\mathbf{x}}) = 0$ and $\nabla_i g_0(\bar{\mathbf{x}}) \neq 0$.*
- C) *There exists an index $i \in \{1, \dots, n\}$ with $\nabla_i f(\bar{\mathbf{x}}) = 0$ and either $\bar{x}_i = l_i$ and $\nabla_i g_0(\bar{\mathbf{x}}) < 0$ or $\bar{x}_i = u_i$ and $\nabla_i g_0(\bar{\mathbf{x}}) > 0$.*

Proof. If there exist infinitely many $k \in \mathbb{N}$, where the updated point $\mathbf{x}^k - t^k \nabla_1 G(\mathbf{x}^k, \lambda^k)$ is contained in $[\mathbf{l}, \mathbf{u}]$, the proof of Theorem 17.2 in Nocedal and Wright [1999], adapted to inequality constraints and $\tau_k = 0$, gives the desired result. We do not treat this case explicitly, as it is covered by the case distinctions below.

Define $\mathcal{I} := \{1, \dots, n\}$ and let $\bar{\mathbf{x}}$ be an accumulation point of $\{\mathbf{x}^k\}$. Since $[\mathbf{l}, \mathbf{u}]$ is compact and \mathbf{x}^k is feasible for (P_{λ^k}) we have $\bar{\mathbf{x}} \in [\mathbf{l}, \mathbf{u}]$. So, the feasibility of $\bar{\mathbf{x}}$ for the original problem is determined by the constraint $f(\mathbf{x}) \leq 0$.

The stationarity of \mathbf{x}^k gives the existence of scalars $t^k > 0$ with

$$\mathcal{P}_{[\mathbf{l}, \mathbf{u}]}(\mathbf{x}^k - t^k \nabla_1 G(\mathbf{x}^k, \lambda^k)) = \mathbf{x}^k,$$

where $(\nabla_1 G(\mathbf{x}^k, \lambda^k))_i = \nabla_i g_0(\mathbf{x}^k) + \lambda^k \max\{0, f(\mathbf{x}^k)\} \nabla_i f(\mathbf{x}^k)$. Together with the convergence of \mathbf{x}^k to $\bar{\mathbf{x}}$ this implies that there exists a $\tilde{k} \in \mathbb{N}$ such that for all $k \geq \tilde{k}$

$$\left(\nabla_1 G(\mathbf{x}^k, \lambda^k) \right)_i \begin{cases} \geq 0, & \bar{x}_i = l_i \\ = 0, & \bar{x}_i \in (l_i, u_i) \\ \leq 0, & \bar{x}_i = u_i. \end{cases} \quad (1)$$

For simplicity, let $c := \lim_{k \rightarrow \infty} \lambda^k \max\{0, f(\mathbf{x}^k)\} \in \mathbb{R}_{\geq 0} \cup \{\infty\}$. If the limit does not exist in $\mathbb{R}_{\geq 0} \cup \{\infty\}$ we can always restrict to a subsequence, for which it exists. Now, we distinguish between feasible and infeasible vectors $\bar{\mathbf{x}}$:

1. $f(\bar{\mathbf{x}}) > 0$, i. e. $\bar{\mathbf{x}}$ is not feasible for (P_p) .

We have to show that $\bar{\mathbf{x}}$ is a KKT-point of (P_f) , this means we have to find Lagrange multipliers $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}_{\geq 0}^n$ with

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) - \boldsymbol{\mu} + \boldsymbol{\nu} &= 0 \\ \bar{\mathbf{x}} &\in [\mathbf{l}, \mathbf{u}] \\ \mu_i(l_i - \bar{x}_i) &= 0, \quad \nu_i(\bar{x}_i - u_i) = 0 \quad \forall i \in \mathcal{I}. \end{aligned}$$

We have already shown that the second KKT-condition holds. Furthermore, we have $c = \infty$ since $\lambda^k \rightarrow \infty$ and $\max\{0, f(\mathbf{x}^k)\} \rightarrow f(\bar{\mathbf{x}}) > 0$. So, there exists an $l \geq \tilde{k}$ such that the sign of (1) is given by the sign of $\nabla_i f(\mathbf{x}^k)$ for all $k \geq l$ if $\nabla_i f(\mathbf{x}^k) \neq 0$. We distinguish between the three possible cases for \bar{x}_i :

- i) $\bar{x}_i \in (l_i, u_i)$.

Owing to (1) we have for all $k \geq l$

$$\nabla_i g_0(\mathbf{x}^k) + \lambda^k \max\{0, f(\mathbf{x}^k)\} \nabla_i f(\mathbf{x}^k) = 0.$$

Therefore, we obtain $\nabla_i f(\bar{\mathbf{x}}) = 0$. Choosing $\mu_i = \nu_i = 0$, the corresponding KKT-conditions of (P_f) are satisfied.

- ii) $\bar{x}_i = l_i$.

If $\nabla_i f(\bar{\mathbf{x}}) = 0$, we can argue like in case i). Otherwise, equation (1) implies $\nabla_i f(\bar{\mathbf{x}}) > 0$. Choosing $\mu_i = \nabla_i f(\bar{\mathbf{x}}) > 0$ and $\nu_i = 0$, the respective KKT-conditions for (P_f) are fulfilled.

- iii) $\bar{x}_i = u_i$.

For $\nabla_i f(\bar{\mathbf{x}}) = 0$ see again case i). Otherwise, (1) implies $\nabla_i f(\bar{\mathbf{x}}) < 0$, and therefore choosing $\mu_i = 0$ and $\nu_i = -\nabla_i f(\bar{\mathbf{x}}) > 0$ gives the respective Lagrange multipliers.

In total, we have shown that $\bar{\mathbf{x}}$ is a KKT-point of (P_f) .

2. $f(\bar{\mathbf{x}}) \leq 0$, i. e. $\bar{\mathbf{x}}$ is feasible for (P_p) .

First, we state the KKT-conditions for the original problem in $\bar{\mathbf{x}}$:

$$\begin{aligned} \nabla g_0(\bar{\mathbf{x}}) + \gamma \nabla f(\bar{\mathbf{x}}) - \boldsymbol{\mu} + \boldsymbol{\nu} &= 0 \\ f(\bar{\mathbf{x}}) &\leq 0, \quad \bar{\mathbf{x}} \in [\mathbf{l}, \mathbf{u}] \\ \gamma f(\bar{\mathbf{x}}) &= 0, \quad \mu_i(l_i - \bar{x}_i) = 0, \quad \nu_i(\bar{x}_i - u_i) = 0 \quad \forall i \in \mathcal{I} \\ \gamma, \boldsymbol{\mu}, \boldsymbol{\nu} &\geq 0. \end{aligned}$$

For $c < \infty$ we choose the Lagrange multipliers $\gamma = c$,

$$\begin{aligned} \mu_i &= \begin{cases} \nabla_i g_0(\bar{\mathbf{x}}) + c \nabla_i f(\bar{\mathbf{x}}), & \bar{x}_i = l_i \\ 0, & \text{otherwise,} \end{cases} \\ \nu_i &= \begin{cases} -\nabla_i g_0(\bar{\mathbf{x}}) - c \nabla_i f(\bar{\mathbf{x}}), & \bar{x}_i = u_i \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Owing to the limit of (1) all KKT-conditions except $\gamma f(\bar{\mathbf{x}}) = 0$ are obviously true. If $f(\bar{\mathbf{x}}) = 0$, we are done and $\bar{\mathbf{x}}$ is a KKT-point of the original problem. Otherwise, we have $f(\bar{\mathbf{x}}) < 0$ which implies for k large enough $f(\mathbf{x}^k) < 0$. So, $\lambda^k \max\{0, f(\mathbf{x}^k)\} = 0$ for large enough k and this implies $\gamma = c = 0$. So, $\bar{\mathbf{x}}$ is a KKT-point of (P_p) .

It remains to consider the case $c = \infty$. This implies immediately $f(\bar{\mathbf{x}}) = 0$ since we need $\max\{0, f(\mathbf{x}^k)\} > 0$ for large k . Furthermore, there exists an $l \geq \tilde{k}$ such that the sign of (1) is given by the sign of $\nabla_i f(\mathbf{x}^k)$ for all $k \geq l$ if $\nabla_i f(\mathbf{x}^k) \neq 0$, and therefore we know

$$\nabla_i f(\bar{\mathbf{x}}) \begin{cases} \geq 0, & \bar{x}_i = l_i \\ = 0, & \bar{x}_i \in (l_i, u_i) \\ \leq 0, & \bar{x}_i = u_i. \end{cases}$$

If case A) in the Theorem is not true, i. e. $\nabla f(\bar{\mathbf{x}}) \neq 0$, we choose the Lagrange multipliers

$$\begin{aligned} \gamma &= \max_{i \in \mathcal{I}: \nabla_i f(\bar{\mathbf{x}}) \neq 0} \left| \frac{\nabla_i g_0(\bar{\mathbf{x}})}{\nabla_i f(\bar{\mathbf{x}})} \right|, \\ \mu_i &= \begin{cases} \nabla_i g_0(\bar{\mathbf{x}}) + \gamma \nabla_i f(\bar{\mathbf{x}}), & \bar{x}_i = l_i \\ 0, & \text{otherwise,} \end{cases} \\ \nu_i &= \begin{cases} -\nabla_i g_0(\bar{\mathbf{x}}) - \gamma \nabla_i f(\bar{\mathbf{x}}), & \bar{x}_i = u_i \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Owing to this choice, the complementarity condition (third KKT-condition) is fulfilled. For $\bar{x}_i \in \{l_i, u_i\}$ the first KKT-condition holds by the definition of γ , $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Now, we consider $\bar{x}_i \in (l_i, u_i)$. Then, we know $\nabla_i f(\bar{\mathbf{x}}) = 0$. So, either case B) in the Theorem holds or $\nabla_i g_0(\bar{\mathbf{x}}) = 0$ for all $\bar{x}_i \in (l_i, u_i)$, which implies the first KKT-condition. It remains to be shown that all Lagrange multipliers are non-negative. This is immediately true for γ . For $\bar{x}_i = l_i$ and $\nabla_i f(\bar{\mathbf{x}}) \neq 0$ we obtain

$$\begin{aligned} \mu_i &= \nabla_i g_0(\bar{\mathbf{x}}) + \gamma \underbrace{\nabla_i f(\bar{\mathbf{x}})}_{>0} \geq \nabla_i g_0(\bar{\mathbf{x}}) + \left| \frac{\nabla_i g_0(\bar{\mathbf{x}})}{\nabla_i f(\bar{\mathbf{x}})} \right| \nabla_i f(\bar{\mathbf{x}}) \\ &= \nabla_i g_0(\bar{\mathbf{x}}) + |\nabla_i g_0(\bar{\mathbf{x}})| \geq 0. \end{aligned}$$

For $\bar{x}_i = u_i$ and $\nabla_i f(\bar{\mathbf{x}}) \neq 0$ we obtain analogously

$$\begin{aligned} \nu_i &= -\nabla_i g_0(\bar{\mathbf{x}}) - \gamma \underbrace{\nabla_i f(\bar{\mathbf{x}})}_{<0} \geq -\nabla_i g_0(\bar{\mathbf{x}}) - \left| \frac{\nabla_i g_0(\bar{\mathbf{x}})}{\nabla_i f(\bar{\mathbf{x}})} \right| \nabla_i f(\bar{\mathbf{x}}) \\ &= -\nabla_i g_0(\bar{\mathbf{x}}) + |\nabla_i g_0(\bar{\mathbf{x}})| \geq 0. \end{aligned}$$

So, we only have to consider the cases with $\bar{x}_i \in \{l_i, u_i\}$ and $\nabla_i f(\bar{\mathbf{x}}) = 0$. If C) in the Theorem is not true, the respective Lagrange multipliers are non-negative. So, we have shown that either $\bar{\mathbf{x}}$ is a KKT-point or at least one of the conditions A), B) or C) in the Theorem is true. □

The Theorem shows that a convergent sequence of stationary points for the penalty problems either converges to an infeasible point that is a KKT-point of (P_f) or to a feasible point that is under some additional conditions a KKT-point of the original problem.

Remark 1. *In contrast to the previous results, the proof cannot be adapted easily to problems with multiple penalized constraints. But for $c < \infty$ the result remains true. This case also includes the well-known result for multiple penalized constraints in Nocedal and Wright [1999].*

Next, we present some examples demonstrating that infeasible accumulation points and feasible ones that are not KKT-points can even occur for quadratic functions.

Example 2 (Example for infeasibility). *We consider the optimization problem*

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) := 0 \\ \text{s.t.} \quad & f(\mathbf{x}) := -x_1^2 + x_2 + 1 \leq 0 \\ & \mathbf{x} \in [0, 1]^2 =: \mathcal{X}. \end{aligned}$$

Let $\mathbf{x}^k = (0, 0)^T$ for all $k \in \mathbb{N}$. Then, we have for all $t > 0$

$$\mathcal{P}_{\mathcal{X}}(\mathbf{x}^k - t\nabla_1 G(\mathbf{x}^k, \lambda^k)) = \mathcal{P}_{\mathcal{X}}((0, -t\lambda^k)^T) = (0, 0)^T.$$

So, \mathbf{x}^k is a stationary point for the respective penalty problem for each $k \in \mathbb{N}$, but the limit $\bar{\mathbf{x}} = (0, 0)^T$ is obviously not feasible for the original problem. Choosing the Lagrange multipliers $\boldsymbol{\mu} = (0, 1)^T$ and $\boldsymbol{\nu} = (0, 0)^T$, $\bar{\mathbf{x}}$ is a KKT-point of (P_f) .

Example 3 (Example for A)). *We consider the problem*

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) := -2x_1 + x_2 \\ \text{s.t.} \quad & f(\mathbf{x}) := x_1^2 - x_2^2 \leq 0 \\ & \mathbf{x} \in [0, 1]^2 =: \mathcal{X}. \end{aligned}$$

We choose $\mathbf{x}^k = (1/k, 0)^T$ and $\lambda^k = k^3$. The sequence converges to $\bar{\mathbf{x}} = (0, 0)^T$ that is feasible for the original problem. We have for all $k \in \mathbb{N}$ and all $t > 0$

$$\mathcal{P}_{\mathcal{X}}(\mathbf{x}^k - t\nabla_1 G(\mathbf{x}^k, \lambda^k)) = \mathcal{P}_{\mathcal{X}}((1/k, -t)^T) = (1/k, 0)^T.$$

So, \mathbf{x}^k is a stationary point for all $k \in \mathbb{N}$. It remains to be shown that $\bar{\mathbf{x}}$ is not a KKT-point of the original problem. Owing to the complementary slackness we need $\nu_1 = 0$. Therefore, the first line of the multiplier rule reads

$$-2 + \gamma \cdot 0 - \mu_1 + 0 = 0 \quad \Leftrightarrow \quad \mu_1 = -2 < 0,$$

which is a contradiction to the non-negativity of the Lagrange multipliers. So, $\bar{\mathbf{x}}$ is not a KKT-point of the original problem. But we see that $\nabla f(\bar{\mathbf{x}}) = 0$. So, case A) in Theorem 5 is fulfilled.

Example 4 (Example for B)). *We consider the problem*

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) := -2x_1 \\ \text{s.t.} \quad & f(\mathbf{x}) := x_1^2 + x_2 \leq 0 \\ & \mathbf{x} \in [-1, 1] \times [0, 1] =: \mathcal{X}. \end{aligned}$$

We choose $\mathbf{x}^k = (1/k, 0)^T$ and $\lambda^k = k^3$. The sequence converges to $\bar{\mathbf{x}} = (0, 0)^T$ that is feasible for the original problem. We have for all $k \in \mathbb{N}$ and all $t > 0$

$$\mathcal{P}_{\mathcal{X}}(\mathbf{x}^k - t\nabla_1 G(\mathbf{x}^k, \lambda^k)) = \mathcal{P}_{\mathcal{X}}((1/k, -tk)^T) = (1/k, 0)^T.$$

So, \mathbf{x}^k is a stationary point for all $k \in \mathbb{N}$. It remains to be shown that $\bar{\mathbf{x}}$ is not a KKT-point for the original problem. Owing to the complementary slackness we need $\mu_1 = \nu_1 = 0$. Therefore, the first line of the multiplier rule is given by

$$-2 + \gamma \cdot 0 = -2 \neq 0,$$

which is a contradiction. So, $\bar{\mathbf{x}}$ is not a KKT-point of the original problem. But we have $\bar{x}_1 \in (l_1, u_1)$, $\nabla_1 f(\bar{\mathbf{x}}) = 0$ and $\nabla_1 g_0(\bar{\mathbf{x}}) \neq 0$ and case B) in Theorem 5 is true.

Example 5 (Example for C)). We consider the problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) := -2x_1 + x_2 \\ \text{s.t.} \quad & f(\mathbf{x}) := x_1^2 + x_2 \leq 0 \\ & \mathbf{x} \in [0, 1]^2 =: \mathcal{X}. \end{aligned}$$

We choose $\mathbf{x}^k = (1/k, 0)^T$ and $\lambda^k = k^3$. The sequence converges to $\bar{\mathbf{x}} = (0, 0)^T$ that is feasible for the original problem. We have for all $k \in \mathbb{N}$ and all $t > 0$

$$\mathcal{P}_{\mathcal{X}}(\mathbf{x}^k - t\nabla_1 G(\mathbf{x}^k, \lambda^k)) = \mathcal{P}_{\mathcal{X}}((1/k, -t(1+k))^T) = (1/k, 0)^T.$$

So, \mathbf{x}^k is a stationary point for all $k \in \mathbb{N}$. It remains to be shown that $\bar{\mathbf{x}}$ is not a KKT-point of the original problem. Owing to the complementary slackness we need $\nu_1 = 0$. Therefore, the first line of the multiplier rule is given by

$$-2 + \gamma \cdot 0 - \mu_1 + 0 = 0 \quad \Leftrightarrow \quad \mu_1 = -2 < 0,$$

which is a contradiction to the non-negativity of the Lagrange multipliers. So, $\bar{\mathbf{x}}$ is not a KKT-point of the original problem. But case C) in Theorem 5 is fulfilled since we have $\nabla_1 f(\bar{\mathbf{x}}) = 0$, $\bar{x}_1 = 0 = l_1$ and $\nabla_1 g_0(\bar{\mathbf{x}}) = -2 < 0$.

In general, KKT-conditions are necessary optimality conditions if a constraint qualification holds (Nocedal and Wright [1999]). In the following, we use the linear independence constraint qualification (LICQ) as defined in Definition 12.4 in Nocedal and Wright [1999]. If the LICQ holds in a feasible accumulation point $\bar{\mathbf{x}}$ we show that the point is always a KKT-point of the original problem. But first, we need the following Lemma:

Lemma 2. Let $\{\mathbf{x}^k\}$ be a sequence of stationary points for (P_{λ^k}) with $\lambda^k \rightarrow \infty$ and let g_0 and f be continuously differentiable. Let furthermore $\bar{\mathbf{x}}$ be a feasible accumulation point that fulfils the LICQ for the original problem and let Assumption 2 be true. Then, the sequence $\{\lambda^k \max\{0, f(\mathbf{x}^k)\}\}$ is bounded from above.

Proof. Let $\mathcal{I} := \{1, \dots, n\}$. For a contradiction, assume that there exists a subsequence $\{k_l\}$ with $\lambda^{k_l} \max\{0, f(\mathbf{x}^{k_l})\} \rightarrow \infty$. This implies the existence of an $m \in \mathbb{N}$ with $f(\mathbf{x}^{k_l}) > 0$ for all $k_l > m$, and therefore we have $f(\bar{\mathbf{x}}) = 0$ since $\bar{\mathbf{x}}$ is feasible. Owing to the stationarity of \mathbf{x}^{k_l} for $(P_{\lambda^{k_l}})$, there exists a $\tilde{k} > m$ such that (1) holds for all $k_l > \tilde{k}$, i. e. for $\bar{x}_i \in (l_i, u_i)$ we have

$$\nabla_i g_0(\mathbf{x}^{k_l}) + \lambda^{k_l} \max\{0, f(\mathbf{x}^{k_l})\} \nabla_i f(\mathbf{x}^{k_l}) = 0.$$

Because of $\lambda^{k_l} \max\{0, f(\mathbf{x}^{k_l})\} \rightarrow \infty$ this implies $\nabla_i f(\bar{\mathbf{x}}) = 0$ for $\bar{x}_i \in (l_i, u_i)$.

Since $f(\bar{\mathbf{x}}) = 0$, the constraint $f(\bar{\mathbf{x}}) \leq 0$ is active. Therefore, the LICQ is fulfilled if

$$a\nabla f(\bar{\mathbf{x}}) + \sum_{i \in \mathcal{I}: \bar{x}_i \in \{l_i, u_i\}} b_i \mathbf{e}_i = 0$$

holds if and only if $a = b_i = 0$ for all i in the sum. We note that the vector \mathbf{e}_i denotes the i -th unit vector. Choosing $a = 1$ and $b_i = -\nabla_i f(\bar{\mathbf{x}})$, we have a non-trivial linear combination of the gradients that is equal to zero since $\nabla_i f(\bar{\mathbf{x}}) = 0$ for all indices i that are not contained in the sum. So, we have a contradiction to the LICQ and our assumption is false. This implies that the sequence $\lambda^k \max\{0, f(\mathbf{x}^k)\}$ is bounded from above. \square

Now, we can state the following Corollary for accumulation points $\bar{\mathbf{x}}$:

Corollary 1. *Let $\{\mathbf{x}^k\}$ be a sequence of stationary points for the penalty problems (P_{λ^k}) with $\lambda^k \rightarrow \infty$ and let g_0 and f be continuously differentiable. Furthermore, let Assumption 2 be true. If an accumulation point $\bar{\mathbf{x}}$ is infeasible for the original problem, it is a KKT-point of (P_f) . If an accumulation point $\bar{\mathbf{x}}$ is feasible and fulfils the LICQ for the original problem, then it is a KKT-point of (P_p) .*

Proof. The infeasible case follows directly from the proof of Theorem 5. So, let $\bar{\mathbf{x}}$ be feasible and let the LICQ be fulfilled in $\bar{\mathbf{x}}$. The proof of Theorem 5 tells us that $\bar{\mathbf{x}}$ is always a KKT-point of the original problem, if $\lambda^k \max\{0, f(\mathbf{x}^k)\}$ is bounded. Owing to Lemma 2 this always holds if the LICQ is fulfilled. So, feasible accumulation points that fulfil the LICQ are always KKT-points of the original problem. \square

Remark 2. *Like in Theorem 5 the proof cannot be adapted directly to multiple penalized constraints for infeasible accumulation points, but the statement in the Corollary for feasible points that fulfil the LICQ also holds for this case.*

Similar to the smoothing parameters β we do not consider sequences of increasing penalty parameters in the following, but deal with a fixed value for λ . The numerical results in Chapter 6.3 indicate that the fixed λ can be chosen large enough to obtain nearly feasible solutions for the original unpenalized problem.

4 Continuous Stochastic Gradient method

In this section, we present the basics of the Continuous Stochastic Gradient method introduced in Pflug et al. [2020] and extended in Grieshammer et al. [2023a] and Grieshammer et al. [2023b]. For further details we refer to these references. Furthermore, we show that the convergence theory for the CSG method is applicable to the approximation problem (P_{app}) . The CSG method was introduced to solve optimization problems with an integral in the objective function. In contrast to many other stochastic optimization schemes, like the Stochastic Gradient method (Robbins and Monro [1951]) or the Stochastic Average Gradient method (Schmidt et al. [2017b]), the CSG method can also handle optimization problems with objective functions that consist of nested expectations or the composition of measure integrals with general non-linear functions (Grieshammer et al. [2023a]). The

approximation problem (P_{app}) belongs to the second class since the integral is composed with the maximum function. In this section, we consider the general optimization problem

$$\min_{\mathbf{x} \in \tilde{\mathcal{X}}} J(\mathbf{x}) := \int_{\tilde{\Delta}} j(\mathbf{x}, \boldsymbol{\delta}) \tilde{\mu}(d\boldsymbol{\delta}) \quad (P_{CSG})$$

with $j : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$. $\tilde{\Delta} \subseteq \mathbb{R}^d$ is the parameter set and $\tilde{\mathcal{X}} \subseteq \mathbb{R}^n$ the feasible set of the optimization problem. The measure $\tilde{\mu}$ corresponds to the distribution of the uncertain parameters $\boldsymbol{\delta}$. As in Grieshammer et al. [2023a] we need some additional assumptions:

Assumption 3. *The problem (P_{CSG}) fulfils:*

- *The sequence of samples $\{\boldsymbol{\delta}^n\}_{n \in \mathbb{N}}$ is independent and identically distributed with distribution $\tilde{\mu}$.*
- *$\tilde{\mathcal{X}}$ is compact and convex.*
- *$\tilde{\Delta}$ is bounded with $\text{supp}(\tilde{\mu}) \subseteq \tilde{\Delta}$.*
- *j is continuously differentiable in \mathbf{x} and its gradient is bounded and Lipschitz continuous in both arguments.*

The general idea of the CSG method is to approximate the gradient of the objective function by only drawing one random sample in each iteration. The approximation quality is improved over time by reusing the information of previous iterations. Therefore, in each iteration we draw a random sample $\boldsymbol{\delta}^n$ and approximate the original gradient with a weighted linear combination of the gradients with respect to the sample points of all previous iterations. So, we approximate the gradient in iteration n with

$$\hat{\mathbf{G}}^n = \sum_{k=1}^n \alpha_k \nabla_{\mathbf{x}} j(\mathbf{x}^k, \boldsymbol{\delta}^k), \quad (2)$$

where α_k are given integration weights. In Grieshammer et al. [2023a] the authors present four different ways how the integration weights can be calculated. In this paper, we restrict to the easiest one, namely the empirical integration weights, since they can be computed efficiently even for high dimensions of $\tilde{\Delta}$. The computation of all weights is based on a nearest neighbour heuristic and Voronoi diagrams. For more details see Grieshammer et al. [2023a]. We define in each iteration $n \in \mathbb{N}$ the empirical weight of the point $(\mathbf{x}^k, \boldsymbol{\delta}^k)$ by

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{M_k}(\boldsymbol{\delta}^i), \quad (3)$$

where

$$\mathbb{1}_{M_k}(y) = \begin{cases} 1, & y \in M_k \\ 0, & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} M_k &:= \{ \boldsymbol{\delta} \in \tilde{\Delta} : \|\mathbf{x}^n - \mathbf{x}^k\| + \|\boldsymbol{\delta} - \boldsymbol{\delta}^k\| \\ &< \|\mathbf{x}^n - \mathbf{x}^j\| + \|\boldsymbol{\delta} - \boldsymbol{\delta}^j\| \text{ for all } j \in \{1, \dots, n\} \setminus \{k\} \}. \end{aligned}$$

So, M_k contains all points $\boldsymbol{\delta} \in \tilde{\Delta}$ such that $(\mathbf{x}^n, \boldsymbol{\delta})$ is closer to $(\mathbf{x}^k, \boldsymbol{\delta}^k)$ than to any other point $(\mathbf{x}^j, \boldsymbol{\delta}^j)$ obtained from a previous iteration $j \neq k$. In the following, we restrict to Euclidean norms although all results also hold for general ones.

Algorithm 1: CSG method

```
while Termination condition not met do
  Sample gradient  $\mathbf{g}^n := \nabla_1 j(\mathbf{x}^n, \boldsymbol{\delta}^n)$ 
  Calculate empirical integration weights  $\boldsymbol{\alpha}$  via (3)
  Calculate search direction  $\hat{\mathbf{G}}^n$  via (2)
  Choose step size  $\tau_n$ 
  Projected gradient step  $\mathbf{x}^{n+1} := \mathcal{P}_{\tilde{\mathcal{X}}}(\mathbf{x}^n - \tau_n \hat{\mathbf{G}}^n)$ 
  Update index  $n \leftarrow n + 1$ 
end
```

The whole procedure of the CSG method is presented in Algorithm 1. Starting from an initial choice \mathbf{x}^0 for the variables, the algorithm computes in each iteration one new sample $\boldsymbol{\delta}^n$ and the corresponding gradient \mathbf{g}^n . After deriving the empirical integration weights $\boldsymbol{\alpha}$, the approximation $\hat{\mathbf{G}}^n$ to the gradient is given as a weighted linear combination of the sampled gradients of the previous iterations. Then, the algorithm selects a step size τ_n and the new iterate \mathbf{x}^{n+1} is given by the Euclidean projection of the gradient step onto the set $\tilde{\mathcal{X}}$.

In Grieshammer et al. [2023a] the authors present the following convergence result for the CSG method with constant step sizes:

Theorem 6. *If Assumption 3 is fulfilled and if the objective function J of problem (P_{CSG}) has only finitely many stationary points on $\tilde{\mathcal{X}}$, the CSG method with constant step size $\tau_n = \tau < 2/L$, where L is the Lipschitz constant of ∇J , converges to a stationary point of J .*

We note that stationary points for (P_{CSG}) are defined equivalently as for the penalty problems in Chapter 3.2.

Our goal is to solve the approximation problem (P_{app}) with the CSG method. To obtain the convergence result in Theorem 6 we have to ensure that Assumption 3 holds. First, we note that (P_{app}) can be written in the form of (P_{CSG}) with $\tilde{\mathcal{X}} = \mathcal{X}$, $\tilde{\Delta} = \Delta$, $\tilde{\mu} = \mu$ and

$$j(\mathbf{x}, \boldsymbol{\delta}) = g_0(\mathbf{x}) + s(g(\mathbf{x}, \boldsymbol{\delta}), \beta).$$

For these definitions and $\text{supp}(\mu) \subseteq \Delta$ the objective function of (P_{CSG}) reads

$$J(\mathbf{x}) = \int_{\Delta} g_0(\mathbf{x}) + s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}) = g_0(\mathbf{x}) + \int_{\Delta} s(g(\mathbf{x}, \boldsymbol{\delta}), \beta) \mu(d\boldsymbol{\delta}),$$

which is the objective function of (P_{app}) . If g_0 and g_j for $j = 1, \dots, \mathcal{J}$ are continuously differentiable in \mathbf{x} and their gradients are bounded and Lipschitz continuous on their domain, the objective function J is also continuously differentiable and its gradient is bounded and Lipschitz continuous. In total, Theorem 6 is applicable to the approximation problem (P_{app}) , if

- \mathcal{X} is compact and convex,
- Δ is bounded with $\text{supp}(\mu) \subseteq \Delta$ and

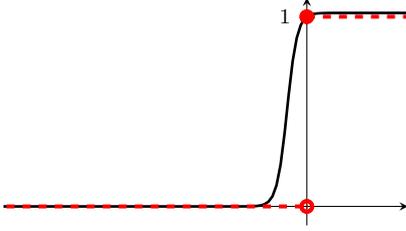


Figure 1: Approximation of indicator function (red) by smoothing function h^ν for $\nu = 0.51$ and $\beta = 20$ (black).

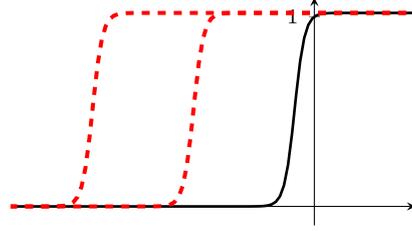


Figure 2: Smoothing function h^ν with $\nu = 0.51$ and $\beta = 20$ for $r = 0$ (black) and $r < 0$ (red).

- g_0 and g_j for $j = 1, \dots, \mathcal{J}$ are continuously differentiable in \mathbf{x} and their gradients are bounded and Lipschitz continuous on their domain. For the constraint functions g_j the Lipschitz condition is required in both arguments, i. e. for each $j = 1, \dots, \mathcal{J}$ there exists a constant $L_j > 0$ with

$$\|\nabla_1 g_j(\mathbf{x}_1, \boldsymbol{\delta}_1) - \nabla_1 g_j(\mathbf{x}_2, \boldsymbol{\delta}_2)\| \leq L_j (\|\mathbf{x}_1 - \mathbf{x}_2\| + \|\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2\|)$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and all $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \in \Delta$.

5 Algorithmic improvements for CSG

In the last section, we have shown that the CSG method computes stationary points of (P_{app}) under some additional conditions. Now, we present two extensions to improve its performance for our problems. For the numerical experiments in Chapter 6 we restrict to a special form of a smoothing function following the conditions in Definition 1, namely a scaled and shifted version of the hyperbolic tangent given by

$$h^\nu(g(\mathbf{x}, \boldsymbol{\delta}), \beta) = \nu \cdot (\tanh(\beta g(\mathbf{x}, \boldsymbol{\delta}) + \gamma) + 1)$$

for a smoothing parameter $\beta > 0$ and a scalar $\nu \geq 0.5$. Owing to the third requirement in Definition 1 we need $h^\nu(0, \beta) = 1$. So, the parameter γ is uniquely determined by ν and β and given by $\gamma = \text{artanh}(1/\nu - 1)$. It can be verified easily that h^ν is a smoothing function in the sense of Definition 1. The function h^ν is illustrated in Figure 1.

5.1 "Steepening of Gradients"

In theory, the smoothing function h^ν has non-zero gradient on its domain, but for a fixed smoothing parameter β we have $\nabla_1 h^\nu(y, \beta) \rightarrow 0$ for $y \rightarrow \pm\infty$. Especially, for fixed β the gradient approaches zero if $g(\mathbf{x}, \boldsymbol{\delta}) \ll 0$, in other words if the condition in the chance constraint is strongly violated. This fact enables the gradient of the whole penalty term to approach zero although the penalized constraint is violated. Therefore, it is possible that the CSG method converges numerically to the optimal solution of the "unconstrained" problem $\min_{\mathbf{x} \in \mathcal{X}} g_0(\mathbf{x})$.

In general, this solution can be arbitrarily far from the feasible set of the original problem and also from the feasible set for the smoothed chance constraint. We conclude that large violations of the condition in the chance constraint can numerically produce gradient

steps that do not create an improving move with respect to feasibility. In our numerical experiments, we counteract this problem by the following trick. Instead of solving the original chance-constrained problem (P) we replace it by

$$\begin{aligned} \min_{\mathbf{x}} \quad & g_0(\mathbf{x}) \\ \text{s.t.} \quad & \mathbb{P}_{\delta \sim \zeta}(g(\mathbf{x}, \delta) - r \geq 0) \geq p \quad \forall r \in \mathcal{R} \\ & \mathbf{x} \in \mathcal{X} \end{aligned} \quad (P_r)$$

for a finite set $\mathcal{R} \subseteq \mathbb{R}_{\leq 0}$ with $0 \in \mathcal{R}$. It can be easily checked that (P) and (P_r) are equivalent. Reformulating and approximating problem (P_r) as discussed in Chapter 2 we obtain the penalty problem

$$\min_{\mathbf{x} \in \mathcal{X}} g_0(\mathbf{x}) + \frac{\lambda}{2} \sum_{r \in \mathcal{R}} \max \left\{ 0, p - \int_{\Delta} s(g(\mathbf{x}, \delta) - r, \beta) \mu(d\delta) \right\}^2.$$

Most of the optimality results for smoothing and penalization presented in Chapter 3 also hold for this formulation. For further details we refer to Remarks 1 and 2 in Chapter 3. For a visualization of the procedure see Figure 2.

In general, for each $\mathbf{x} \in \mathcal{X}$ and each $\delta \in \Delta$ there should exist an $r \in \mathcal{R}$ with $g(\mathbf{x}, \delta) - r$ close to zero. This choice of \mathcal{R} ensures that the gradient of the penalty term does not vanish if the smoothed chance constraint is violated. In practice, this choice is often not possible since large sets \mathcal{R} extend the computation time drastically. Instead of approximating all possible values of g , one often has to restrict to approximate the values of g that are attained during the optimization procedure. In total, one has to ensure that the minimal value in \mathcal{R} is small enough and that the whole set is "dense" enough to derive a good approximation for the values of g . For our numerical results in Chapter 6 preliminary computations suggest choosing $\mathcal{R} = \{\dots, -0.02, -0.01, 0\}$ up to a suitably chosen bound for each problem. In general, the set \mathcal{R} has to be "denser" for larger smoothing parameters β . As mentioned before, large sets \mathcal{R} require more computation time since the smoothing function and its gradient have to be evaluated in each iteration for each element in \mathcal{R} . For large-sized problems the number of evaluations is often not tractable. Therefore, we need a possibility to decrease it. We note that the smoothing function is monotonically increasing and only attains values between 0 and 2ν . So, when reaching a scalar $r \in \mathcal{R}$, where the function value is numerically close to 0, this also holds for all smaller r and we can stop the evaluations and set the smoothing function for all smaller scalars to 0. Contrarily, if we find some r with a numerical function value larger or equal to 2ν this also holds for all larger scalars r and we can again stop and set all function values for larger scalars r to 2ν . To minimize the number of evaluations we start at the scalar $r \in \mathcal{R}$ that is closest to the actual value of g . For the gradients we can proceed in a similar way. Since both limits for $\pm\infty$ are 0, we can stop the procedure in both directions if the gradient value is numerically equal to 0.

5.2 Step size rule

The convergence result for the CSG method (Grieshammer et al. [2023a]), mentioned in Theorem 6, requires a constant step size. In contrast, we choose a more advanced step size strategy that allows larger penalty parameters. We initialize each iteration with the constant size τ , but decrease it if the gradient step would become too large. This

mainly happens for points with positive penalty term if the penalty parameter λ is large. Concretely, we define a scalar $t > 0$ and reduce the constant step size in iteration n to

$$\tilde{\tau} = \tau \cdot t \cdot \frac{\|\nabla g_0(\mathbf{x})\|}{\|\hat{\mathbf{G}}^n\|}, \quad \text{if } \|\hat{\mathbf{G}}^n\| > t \cdot \|\nabla g_0(\mathbf{x})\|.$$

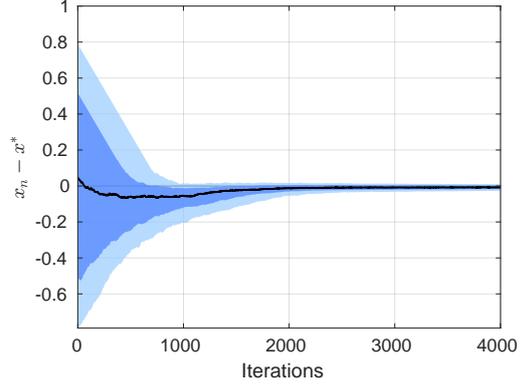
In practice, the step size in each iteration should be chosen as large as possible to still obtain convergence. Our strategy for the numerical experiments is to choose the constant step size τ large enough to ensure enough descent if we are in the feasible set for the smoothed constraint. The scalar t has to be chosen large enough to prohibit too much descent in the infeasible region. So, t has to be chosen large, if the penalty parameter λ is large. This strategy ensures that at the transfer from the infeasible to the feasible region large penalty parameters do not result in jumping wide into the interior of the feasible set, but staying close to the boundary.

6 Numerical results

In this section, we illustrate the effectiveness of the CSG method for the approximation problem (P_{app}) with some numerical examples. First, we test the method on a simple analytical example. Then, the applicability to practical problems is demonstrated with some instances for the Nomination Validation problem in gas networks. Additionally, we study the method's performance for different smoothing and penalty parameters in Chapter 6.3 to support our strategy of fixed values for β and λ . In Chapter 6.4 we discuss two different methods to verify feasibility for the original and the smoothed chance constraint. Our method is implemented in Matlab R2022a. All computations are executed on a machine with an Intel Xeon E3-1240 v6 CPU (3.7 gigahertz base frequency), 4 cores and 32 gigabyte RAM.

To apply the CSG method to (P_{app}) we have to specify several parameters. In general, the choice of parameter values is problem-specific, but some general rules for appropriate choices are given next. Parameter ν determines 0.5 times the limit of the smoothing function h^ν if the argument tends to infinity. It should be chosen slightly larger than 0.5 to approximate the original indicator function well. We note that a good approximation for positive arguments is only necessary due to the "Steepening of Gradients". For the smoothing and penalty parameters β and λ , the convergence theory in Chapter 3 implies large values. However, choosing them too large can impact the stability and number of iterations in CSG negatively. Therefore, one has to find a compromise between good approximations of the original problem and practical requirements for the CSG method. For our instances, good values for β lie in the range of 10^3 to 10^4 and for λ in the range of 10^5 . For the CSG stopping criterion, we restrict the maximal number of iterations. Here, we have to ensure that we choose this number large enough to obtain convergence, but it also has to be small enough to keep computational tractability of computing the empirical weights. For our experiments we perform 4000 iterations. Typically, a larger number of iterations increases the running time, but does not change the results considerably for the tested instances. For the step size and the parameter set \mathcal{R} we follow the suggestions in Chapter 5. The concrete values for the parameters for all tested problems are specified by preliminary computational experiments. Furthermore, we perform 500 random runs of the CSG method for each instance.

Figure 3: Error $x_n - x^*$ over iterations for the median (black line). Shaded areas for quantiles $P_{0.1,0.9}$ (light) and $P_{0.25,0.75}$ (dark).



Since we solve an approximation of the original joint chance-constrained problem, a natural question is whether the obtained solutions are feasible for the original constraints. Because the measure integral in the chance constraint cannot be evaluated exactly, we check feasibility with a Monte-Carlo approximation. We use Monte-Carlo to approximate the left-hand sides of the original chance constraint and the smoothed version evaluated at the median and several quantile solutions of the 500 returned solutions \mathbf{x} .

6.1 Analytical example

In this section, we consider the one-dimensional joint chance-constrained problem

$$\begin{aligned} \min_x \quad & x \\ \text{s.t.} \quad & \mathbb{P}_{\delta \sim \zeta}(x + \delta \geq 0, -x \cdot \delta + 0.5 \geq 0) \geq 0.5 \\ & x \in [-1, 1] =: \mathcal{X}, \end{aligned}$$

where $\Delta = [-1, 1]$ and ζ is the uniform distribution over Δ . The optimal solution is given by $x^* = 0$. We evaluate our approach on this convex example to illustrate its ability to solve joint chance-constrained problems. In Chapter 6.2 we solve large gas network instances with non-convex constraint functions.

With $g(x, \delta) = -\min\{x + \delta, 0\}^2 - \min\{-x \cdot \delta + 0.5, 0\}^2$ the approximation problem reads

$$\min_{x \in \mathcal{X}} x + \frac{\lambda}{2} \max \left\{ 0, 0.5 - \int_{\Delta} h^{\nu}(g(x, \delta), \beta) \mu(d\delta) \right\}^2.$$

Owing to the preliminary computational tests we choose $\nu = 0.51$, $\beta = 2 \cdot 10^4$ and $\lambda = 2 \cdot 10^5$. The initial constant step size is given by $\tau = 10^{-3}$ and is decreased with parameter $t = 2$. The set \mathcal{R} is chosen as $\mathcal{R} := \{-5, -4.99, -4.98, \dots, 0\}$. The results for the 500 runs with randomly chosen starting points in $[-1, 1]$ are presented in Figure 3. The graphic shows that the sequences approximate the true solution x^* better over the iterations. Instead of returning the solution found in the last iteration, the CSG method outputs the iteration with the smallest penalized objective function value over the last 50 iterations as solution for the approximation problem. The median of these returned solutions is -0.0096 . So, the absolute error to the optimal solution of the original problem is about 0.01. Furthermore, one random run of CSG only takes about 40 – 45 seconds for this problem.

With respect to feasibility we observe that about 75% of the returned solutions are negative, and therefore not feasible, however the violation is small. For concrete values, we

CC	90%	75%	median	25%	10%
smoothed	0.51	0.50	0.50	0.49	0.49
original	0.50	0.50	0.49	0.49	0.48

Table 1: Monte-Carlo approximation of left-hand side of smoothed CC and original CC evaluated at the median and the respective quantile values of the returned solutions x over 500 runs of CSG.

check feasibility with a Monte-Carlo approximation using at most 10^5 random samples. We stop the approximation if either the maximal number of iterations is reached or if the absolute error between the actual approximation and the average over the last 50 iterations is smaller than 10^{-10} . The left-hand side of the smoothed constraint for the median value is given by 0.50 and for the original problem we obtain 0.49. The approximations for the quantiles can be found in Table 1. The values indicate that the CSG method finds nearly feasible solutions for the approximation problem. Nevertheless, the returned solutions are not in all runs feasible for the smoothed constraint. We observe that in about 20% of the runs the CSG approximation to the violation of the smoothed constraint is in the range of 10^{-4} . The maximal violation in one run is given by 0.02. So, the infeasibility in some runs is caused by the inexact penalty approach. Since the smoothed chance constraint is nearly fulfilled with probability 0.5 and the penalty term approaches zero, we can conclude that our approach converges to the optimal solution of the smoothed optimization problem. This implies that the absolute error of about 0.01 between the median and the optimal solution x^* is caused by the smoothing approximation.

6.2 Nomination Validation for gas networks

In contrast to the analytical example above, these problems contain non-convex constraint functions. Furthermore, we consider real-world sized problems. We briefly introduce a reformulated, nominal version of a Nomination Validation problem, explain how chance constraints are included and present the numerical results of CSG for slightly modified realistic gas network instances from the library by Schmidt et al. [2017a].

6.2.1 Chance-constrained Nomination Validation problem for gas networks.

For a general overview on the mathematical treatment of gas networks see for example Koch et al. [2015]. In general, a gas network can be modeled as a directed graph $\mathcal{G} = (V, E)$. We consider a stationary gas model. Our goal is to minimize the cost $w(\mathbf{x})$ for modifying active elements, like compressors or control valves, in the network. We assume that w has a unique and strict minimum for the zero-vector, which means that no modification of the active elements induces no cost and every modification induces a strictly positive cost. The gas network's physics is determined by the gas flow on each edge and a pressure value for each node. A feasible gas flow fulfils flow conservation with respect to a given balanced demand vector \mathbf{d} , where balanced means $\sum_{v \in V} d_v = 0$. Furthermore, a nonlinear pressure loss equation that depends on pressure loss coefficients $\phi_a \geq 0$ for each edge $a \in E$, is required and we have bounds $\underline{\pi}$ and $\overline{\pi}$ for the pressure vector. We note that we use a linear compressor model. For a more detailed overview on the nominal problem we refer to Gotzes et al. [2016] and Aßmann et al. [2019]. In Gotzes et al. [2016] and Aßmann et al. [2019], the authors introduce a reformulation of the nominal problem by eliminating the

flow and pressure variables. Afmann et al. [2019] show that the reformulation works if there are no active elements on cycles, whereas Stangl [2014] proves the statement if every cycle contains at least one pipe with strictly positive pressure loss coefficient ϕ_a . The latter result is for example also used in Kuchlbauer et al. [2022a] and Kuchlbauer et al. [2022b]. The reformulation fixes the pressure at a chosen root node r and provides functions π_v for the pressures at all nodes, except r , that only depend on the control variables \mathbf{x} , the demands \mathbf{d} and the pressure loss coefficients ϕ . Then, the whole problem reads

$$\begin{aligned} \min_{\mathbf{x}} \quad & w(\mathbf{x}) \\ \text{s.t.} \quad & \pi_v(\mathbf{x}, \mathbf{d}, \phi) \in [\underline{\pi}_v, \bar{\pi}_v] \quad \forall v \in V \setminus \{r\} \\ & \mathbf{x} \in [\underline{\mathbf{x}}, \bar{\mathbf{x}}] =: \mathcal{X}. \end{aligned}$$

After introducing the nominal problem, we additionally consider uncertainties in the demands \mathbf{d} and assume that the demands follow a probability distribution ζ over the set

$$D = \left\{ \mathbf{d} \in \mathbb{R}^{|V|} : \sum_{v \in V} d_v = 0, d_v \in [l_v, u_v] \forall v \in V \right\},$$

where $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{|V|}$ are fixed vectors with $l_v \leq u_v$ for all $v \in V$. Now, we require that the pressure bounds $\underline{\pi}$ and $\bar{\pi}$ have to be satisfied with a given probability p which induces the joint chance constraint

$$\mathbb{P}_{\mathbf{d} \sim \zeta} (\pi_v(\mathbf{x}, \mathbf{d}, \phi) - \underline{\pi}_v \geq 0, \bar{\pi}_v - \pi_v(\mathbf{x}, \mathbf{d}, \phi) \geq 0 \forall v \in V \setminus \{r\}) \geq p.$$

So, the function g is defined as

$$g(\mathbf{x}, \mathbf{d}, \phi) = \sum_{v \in V \setminus \{r\}} (-\min\{0, \pi_v(\mathbf{x}, \mathbf{d}, \phi) - \underline{\pi}_v\}^2 - \min\{0, \bar{\pi}_v - \pi_v(\mathbf{x}, \mathbf{d}, \phi)\}^2)$$

and the corresponding approximation problem reads

$$\min_{\mathbf{x} \in \mathcal{X}} w(\mathbf{x}) + \frac{\lambda}{2} \max \left\{ 0, p - \int_D h^\nu(g(\mathbf{x}, \mathbf{d}, \phi), \beta) \mu(d\mathbf{d}) \right\}^2.$$

CSG requires the gradient of g with respect to \mathbf{x} . Kuchlbauer et al. [2022a] and Kuchlbauer [2023] present a gradient formula. It can be easily verified that it is Lipschitz continuous and bounded. If furthermore w is continuously differentiable and its gradient is locally Lipschitz continuous and bounded, the whole objective function has a bounded and Lipschitz-continuous gradient. Together with the definition of D and the feasible set for \mathbf{x} Assumption 3 is fulfilled, and therefore Theorem 6 guarantees the convergence of the CSG method for suitably chosen step sizes.

6.2.2 Numerical Results.

For our numerical studies in the gas context we use slightly modified instances from a library of realistic gas instances (Schmidt et al. [2017a]), namely GasLib 24, GasLib 40 and GasLib 134. General information with respect to the size of these instances is collected in Table 2. We note that the instances are of the size of real-world gas networks. To create the uncertainty set D we use the given nominal demands n_v for each node v in the instance

instance	number nodes	number edges	number active elements	dimension uncertainty	number constraints in CC
GasLib 24	24	25	4	8	46
GasLib 40	40	45	6	32	78
GasLib 134	134	133	2	48	266

Table 2: Information for GasLib instances.

files and perturb them by 5% in each direction. So, the feasible set for the demands is given by

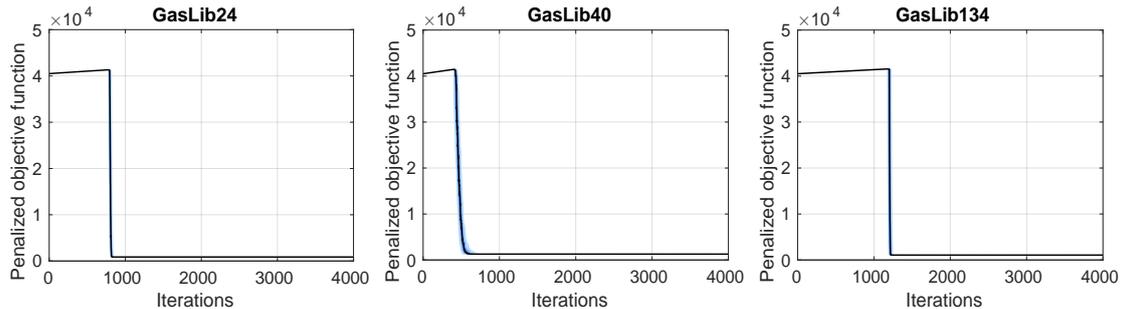
$$D = \left\{ \mathbf{d} \in \mathbb{R}^{|V|} : \sum_{v \in V} d_v = 0, d_v \in [l_v, u_v] \forall v \in V \right\}$$

with $[l_v, u_v] = [0.95 \cdot n_v, 1.05 \cdot n_v]$ for $n_v \geq 0$ and $[l_v, u_v] = [1.05 \cdot n_v, 0.95 \cdot n_v]$ for $n_v < 0$. We note that the size of the uncertainty does not impact the general performance of the CSG method. For the chance constraint we consider the uniform distribution over D . CSG requires independent, uniformly distributed samples on D . Since D is a box intersected with a hyperplane, uniform sampling on D cannot be done by simply using built-in Matlab functions. First, we reduce the dimension of D by 1 to obtain a set with non-empty interior. Choosing an arbitrary node $u \in V$ and defining

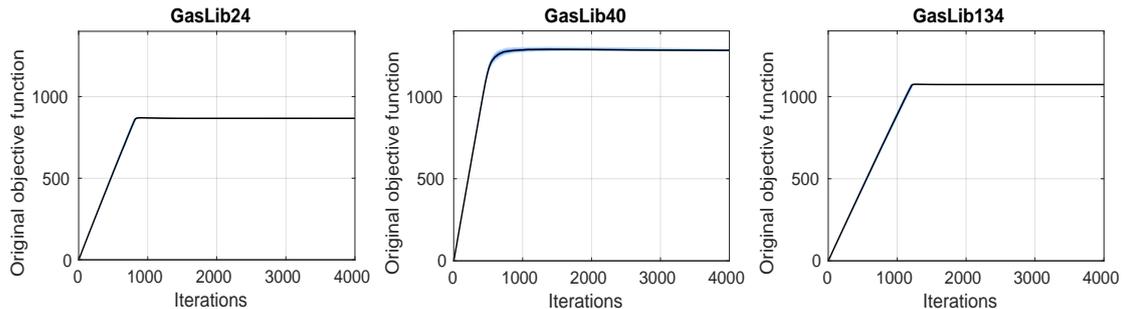
$$\tilde{D} := \left\{ \tilde{\mathbf{d}} \in \mathbb{R}^{|V|-1} : \tilde{d}_v \in [l_v, u_v] \forall v \in V \setminus \{u\}, - \sum_{v \in V \setminus \{u\}} \tilde{d}_v \in [l_u, u_u] \right\},$$

there is a one-to-one correspondence between points in D and \tilde{D} following the bijection $\psi : \tilde{D} \rightarrow D$ with $\psi(\tilde{\mathbf{d}}) = (\tilde{\mathbf{d}}^T, -\sum_{v \in V \setminus \{u\}} \tilde{d}_v)^T$. We thus sample uniformly on D by sampling uniformly on \tilde{D} . The uniform sample on \tilde{D} is constructed with rejection sampling, this means we sample uniformly at random on the box $[l_v, u_v]_{v \in V \setminus \{u\}}$ until the sampled vector is contained in \tilde{D} . The accepted vectors build the sample on \tilde{D} . We note that the rejection method works well for \tilde{D} , whereas it is not applicable in practice to D since the equality constraint is hardly ever satisfied for a random sample in the box.

For the numerical experiments we choose the linear objective function $w(\mathbf{x}) := \mathbf{1}^T \mathbf{x}$, where $\mathbf{1}$ is the vector that contains only 1-entries, and a probability level $p = 0.9$. We again demonstrate the results of the CSG method for 500 random runs. Since there do not exist strict bounds on the variables \mathbf{x} we choose the feasible set $[\underline{\mathbf{x}}, \bar{\mathbf{x}}]$ with $\mathbf{0} \in [\underline{\mathbf{x}}, \bar{\mathbf{x}}]$ very large to never reach the bounds in an optimum. Owing to the large feasible set we do not start the CSG method at random points, but initialize the algorithm with $\mathbf{x} = \mathbf{0}$ since this vector gives the lowest cost and would be preferred if no constraint is violated. To choose the parameters we perform again preliminary computations which lead to $\nu = 0.51$, $\beta = 5 \cdot 10^3$ and $\lambda = 10^5$. The initial constant step size is given by $\tau = 10^{-2}$ and is decreased with parameter $t = 100$. The set \mathcal{R} is chosen as $\mathcal{R} := \{-2000, -1999.99, -1999.98, \dots, 0\}$. Since the true optimum of the Nomination Validation problems is not known, we cannot quantify the error of the returned solutions. Therefore, we illustrate in Figures 4a and 4b the sequences of function values for the objective function in the approximation problem including the penalty term and the original objective function w . Obviously, the initial vector $\mathbf{x} = \mathbf{0}$ is infeasible, and therefore the penalized objective function is very large in the beginning. By driving the solution towards the feasible set, the penalized objective function increases until the solution becomes feasible. At this point, there is a large jump downwards for the penalized function. The original function also increases until this jump



(a) Penalized objective function.



(b) Original objective function.

Figure 4: Objective functions over iterations for the median (black line). Shaded areas for quantiles $P_{0.1,0.9}$ (light) and $P_{0.25,0.75}$ (dark).

is reached. In total, we observe that after the jump the values for both objective functions stay nearly constant, which means that the CSG method converges after about 500 – 1200 iterations for the tested networks. Additionally, we note that the quantiles are very small, which emphasizes the efficiency of our method. One random run needs at most 150 seconds for the three instances.

Next, we discuss the feasibility. Again, for the CSG iteration with the smallest penalized objective function value over the last 50 iterations, we approximate the left-hand side of the original chance constraint and the smoothed version with Monte-Carlo using 10^4 samples. The results for the median and the quantiles can be found in Table 3. Indeed, CSG finds nearly feasible solutions for the smoothed constraints. Nevertheless, the desired probability level is not reached in all runs and also not for the median. Table 4 tells us that the violation of the smoothed constraint in the CSG approximation is often not exactly equal to zero. For GasLib 24 we obtain in 61% of the runs a violation, for GasLib 40 in 36% and for GasLib 134 in 48%. In average, the violations are in the range of 10^{-4} . For GasLib 134 the maximal violation is in the range of 10^{-3} , whereas for the other two instances the maximal one also lies in the range of 10^{-4} . So, slightly infeasible solutions for the smoothed chance constraint occur due to the penalization technique. Since the values for the smoothed chance constraints are nearly 0.9, this also indicates that the found solutions are not far from the boundary of the smoothed feasible region. So, the returned solutions yield high-quality objective function values with respect to minimization of $w(\mathbf{x})$.

instance	CC	90%	75%	median	25%	10%
GasLib 24	smoothed	0.88	0.89	0.89	0.90	0.91
	original	0.84	0.85	0.86	0.86	0.87
GasLib 40	smoothed	0.91	0.90	0.89	0.89	0.89
	original	0.91	0.89	0.89	0.89	0.89
GasLib 134	smoothed	0.87	0.88	0.88	0.89	0.91
	original	0.79	0.80	0.81	0.82	0.83

Table 3: Monte-Carlo approximation of left-hand side of smoothed CC and original CC evaluated at the median and the respective quantile values of the returned solutions \mathbf{x} over 500 runs of CSG.

instance	number of runs with violation	average violation	maximal violation
GasLib 24	61%	$1.7 \cdot 10^{-4}$	$4.5 \cdot 10^{-4}$
GasLib 40	36%	$2.3 \cdot 10^{-4}$	$9.7 \cdot 10^{-4}$
GasLib 134	48%	$1.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$

Table 4: Violation in CSG method for returned solution over all 500 runs.

For GasLib 40 the solutions are also nearly feasible for the original chance constraint, which is a very good result. But for the other two instances the original constraint is violated. Since the approximation to the smoothed constraint is good for all instances, the violation of the original one results from the smoothing technique. To heal this, we use the trick presented in Chapter 2 and increase the probability level to $0.9 + \tilde{\epsilon}$. In general, the value for $\tilde{\epsilon}$ can be chosen small if the violation of the original chance constraint is also small. Like for all other parameters we perform preliminary computational experiments. If the violation of the original chance constraint is small, like for GasLib 24, we start testing the performance for $\tilde{\epsilon} = 0.01$ and perform a bisection method to obtain a reasonable value for $\tilde{\epsilon}$. If the initial violation is large, like for GasLib 134, we start with a larger initial value for $\tilde{\epsilon}$ and proceed in the same way. For GasLib 24, we obtain $\tilde{\epsilon} = 0.03$ as a good choice, whereas for GasLib 134 we use $\tilde{\epsilon} = 0.06$. Then, Theorem 1 ensures that the chance constraint

$$\mathbb{P}_{\mathbf{d} \sim \zeta}(g(\mathbf{x}, \mathbf{d}, \boldsymbol{\phi}) \geq -\epsilon) \geq p + \tilde{\epsilon} - \tilde{\epsilon} = p$$

is fulfilled, whenever the smoothed chance constraint is fulfilled with probability level $p + \tilde{\epsilon}$. Here, ϵ is a scalar such that $h^\nu(g(\mathbf{x}, \mathbf{d}, \boldsymbol{\phi}), \beta) \leq \tilde{\epsilon}$ for $g(\mathbf{x}, \mathbf{d}, \boldsymbol{\phi}) < -\epsilon$. Inserting the special form of the smoothing function h^ν we obtain

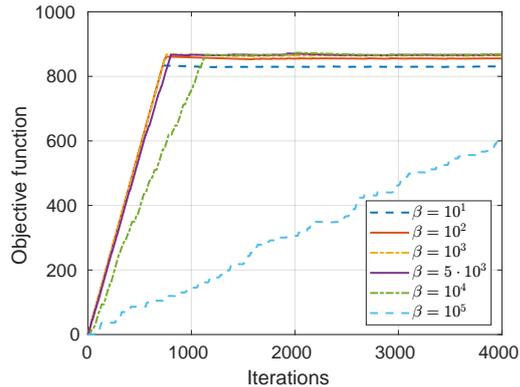
$$\epsilon = -\frac{\operatorname{artanh}\left(\frac{\tilde{\epsilon}}{\nu} - 1\right) - \gamma}{\beta}.$$

For the chosen $\tilde{\epsilon}$ for GasLib 24 and GasLib 134 the values for ϵ are in the range of 10^{-4} . Results for solving the approximation problem with $0.9 + \tilde{\epsilon}$ can be seen in Table 5. We notice that the values for the smoothed versions are nearly equal to the new probability levels $0.9 + \tilde{\epsilon}$, which implies that the CSG method also works for these approximation problems. Concerning the original chance constraint, the returned solutions are nearly feasible. So, our correction procedure works for appropriately chosen parameters $\tilde{\epsilon}$. Studying the results for the perturbed chance constraint, one notes that the probability level is sometimes much higher than requested. This illustrates that the result in Theorem 1 gives an upper bound for the probability that the perturbed chance constraint is satisfied, but in practice the probability is often much higher. In conclusion, our approach consistently produces

instance	CC	90%	75%	median	25%	10%
GasLib 24	smoothed	0.91	0.92	0.92	0.93	0.94
	original	0.88	0.88	0.89	0.90	0.91
	perturbed	0.92	0.93	0.94	0.94	0.95
GasLib 134	smoothed	0.94	0.95	0.95	0.96	0.96
	original	0.88	0.89	0.89	0.91	0.91
	perturbed	0.96	0.96	0.97	0.97	0.97

Table 5: Monte-Carlo approximation of left-hand side of smoothed CC, original CC and perturbed CC evaluated at the median and the respective quantile values of the returned solutions \mathbf{x} over 500 runs of CSG with probability level $0.9 + \tilde{\epsilon}$.

Figure 5: Sequence of original objective function values for different smoothing parameters in one fixed run.



solutions that are remarkably close to feasibility, even when dealing with real-world sized gas network instances featuring non-convex constraint functions and stochastic parameters spanning up to 50 dimensions.

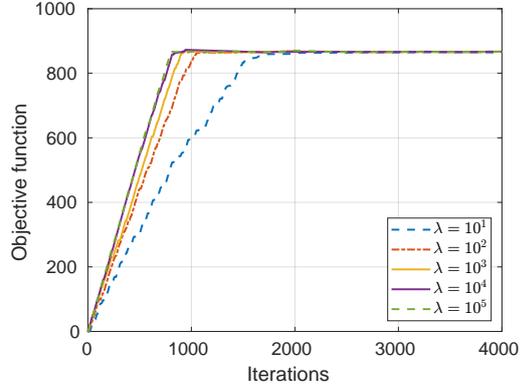
6.3 Different smoothing and penalty parameters

The convergence results in Chapter 3 suggest to solve sequences of approximation problems for increasing smoothing and penalty parameters. In contrast, we solve the approximation problem for one fixed value of β and λ . In this section, we analyze why this procedure is appropriate. First, we consider the smoothing approximation for a fixed penalty parameter $\lambda = 10^5$ and smoothing parameters $\beta \in \{10^1, 10^2, 10^3, 5 \cdot 10^3, 10^4, 10^5\}$. We present the results exemplarily for the instance GasLib 24, but similar results can also be observed for other problems. The results for one run of CSG with the same fixed samples for all values of β can be seen in Figure 5. We observe that the CSG method needs more iterations to converge for larger values of β . Especially, for the largest value CSG does not converge within 4000 iterations. Concerning the feasibility, the Monte-Carlo approximations are shown in Table 6. We note that the smoothed constraint is satisfied with the required probability for convergent runs, whereas the original one is only nearly satisfied for $\beta = 5 \cdot 10^3$ and $\beta = 10^4$. The violation for all other values is large. As a consequence, small parameters β are not meaningful in practice, but using constant values is nevertheless acceptable if they are chosen large enough. Our choice of $\beta = 5 \cdot 10^3$ is reasonable since it gives the best trade-off between fast convergence and feasibility for all tested values. In total, for our purposes it is enough to solve one approximation problem for a fixed and large smoothing parameter instead of a sequence of problems to obtain nearly feasible solutions.

CC	$\beta = 10^1$	$\beta = 10^2$	$\beta = 10^3$	$\beta = 5 \cdot 10^3$	$\beta = 10^4$	$\beta = 10^5$
smoothed	0.90	0.90	0.90	0.91	0.90	0
original	0.01	0.53	0.81	0.87	0.88	0

Table 6: Monte-Carlo approximation of left-hand side of smoothed CC and original CC of returned solution \mathbf{x} for different smoothing parameters β in one fixed run.

Figure 6: Sequence of original objective function values for different penalty parameters in one fixed run.



Now, we consider different penalty parameters $\lambda \in \{10^1, 10^2, 10^3, 10^4, 10^5\}$ and fix the smoothing parameter to $\beta = 5 \cdot 10^3$. The results for one fixed random run are presented in Figure 6. We observe faster convergence for larger penalty parameters. Thus, choosing a large parameter that guarantees a small penalty term is a very good choice. Table 7 shows the Monte-Carlo approximation of the left-hand side of the smoothed chance constraint for the returned solution \mathbf{x} for the different penalty parameters. We observe that the smoothed constraint is violated for the smallest value for λ , whereas the remaining tests give feasible solutions for the smoothed constraint. Therefore, solving the approximation problems for a fixed value of λ is also reasonable and our choice $\lambda = 10^5$ is good since it gives no violation and a small convergence time.

6.4 Checking feasibility for the chance constraint

As mentioned before, checking feasibility of a given vector for a chance constraint can be difficult for continuous distributions. To verify the solution quality, we approximate the measure integral by Monte-Carlo simulation in the last subsections. Obviously, this procedure results in additional computational effort, in particular if sampling is complicated. An alternative idea to check feasibility is using the information gained during the CSG method. Since CSG is proven to produce approximations to the integrand that converge to the true value, this also holds for the left-hand side of the smoothed chance constraint. Since the original chance constraint is neither differentiable nor continuous, the theory for the CSG method gives no result for the original constraint. Nevertheless, it is natural to assume that the approximation also works for the indicator functions forming the chance constraint since CSG approximates function values with a piecewise constant model. So, the jumps in the indicator functions can be approximated better over the iterations. We illustrate this assumption for the tested GasLib instances. Therefore, we perform for each instance one random run of CSG and approximate in each iteration the left-hand side of the smoothed chance constraint and the original one. The approximation to the smoothed constraint is already computed for the penalized objective function values, therefore we

$\lambda = 10^1$	$\lambda = 10^2$	$\lambda = 10^3$	$\lambda = 10^4$	$\lambda = 10^5$
0.85	0.90	0.90	0.90	0.90

Table 7: Monte-Carlo approximation of left-hand side of smoothed CC of returned solution \mathbf{x} for different penalty parameters λ in one fixed run.

instance	CC	MC	CSG
GasLib 24	smoothed	0.90	0.90
	original	0.86	0.87
GasLib 40	smoothed	0.90	0.90
	original	0.89	0.90
GasLib 134	smoothed	0.88	0.90
	original	0.80	0.83

Table 8: Monte-Carlo and CSG approximation of left-hand side of smoothed CC and original CC of returned solution \mathbf{x} for one fixed run.

do not need additional computation time here. For the original chance constraint we compute the value of the indicator function in each iteration, this means we check, whether $g(\mathbf{x}^k, \boldsymbol{\delta}^k) \geq 0$. Then, we approximate the left-hand side with a weighted linear combination of the indicator function values, where the coefficients are given by the empirical weights $\boldsymbol{\alpha}$. In the end, the algorithm outputs the approximation obtained in the iteration, where the returned solution appears. The obtained results and the Monte-Carlo approximations can be found in Table 8. We see that the values for the Monte-Carlo and CSG approximation are nearly equal for GasLib 24 and GasLib 40, whereas we observe a difference of up to 3% for GasLib 134. But since the difference also occurs for the smoothed version and the CSG method is proven to converge to the true value for the smoothed constraint, the difference for the largest instance seems to result from the number of iterations that is too small to guarantee the convergence of the function values for the high-dimensional uncertainty. Nevertheless, although the approximation of the left-hand sides of the chance constraint is not exact after 4000 iterations the results of the Monte-Carlo approximation imply that the obtained solution vector \mathbf{x} is exact enough. So, the number of iterations suffices to compute solutions to the approximation problem (P_{app}). In total, replacing the Monte-Carlo simulation by the CSG approximation provides a good indication for the feasibility of the returned solutions if the stochastic uncertainty is not extraordinarily large. This enables us to evaluate our approach without using Monte-Carlo techniques.

7 Conclusion

In this paper, we introduce an approximation for joint chance-constrained optimization problems using smoothing and penalization that enables the application of the CSG method. The procedure allows non-convexities in the constraint functions and does not rely on Monte-Carlo sampling. The numerical experiments show the efficiency of our approach, even for high-dimensional uncertainties in the gas context. Furthermore, we demonstrate with a convergence theory for the smoothing and penalty approach that our approximation problem provides a good approximation for the original problem for large smoothing and penalty parameters. For the penalty approach we show convergence results for sequences of optimal solutions and sequences of KKT-points if the LICQ holds. For

the smoothing approximation we only present a convergence result with respect to optimal solutions. A question for further research is whether a similar result for KKT-points can be obtained, if the cumulative distribution function is locally Lipschitz continuous and whether there exist meaningful conditions on the distribution and the constraint functions that guarantee the Lipschitz continuity of the distribution function in our setting. Since our approach does not rely on any assumptions on the distributions or the constraint functions, except of Lipschitz continuous gradients, the approach is applicable in a wide context. Another interesting question for further research is, whether the presented approach can be used for solving distributionally robust optimization problems with chance constraints, where the distribution is not known exactly.

8 Acknowledgments

The authors thank the DFG for their support within Project B06 in the Sonderforschungsbereich/Transregio 154 Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks with Project-ID 239904186. Furthermore, they thank Andrian Uihlein for providing the basic implementation of the CSG method and Martina Kuchlbauer for implementations in the gas context.

References

- L. Adam and M. Branda. Nonlinear Chance Constrained Problems: Optimality Conditions, Regularization and Solvers. *Journal of Optimization Theory and Applications*, 170:419–436, 2016. doi:10.1007/s10957-016-0943-9.
- L. Adam, M. Branda, H. Heitsch, and R. Henrion. Solving joint chance constrained problems using regularization and Benders’ decomposition. *Annals of Operations Research*, 292:683–709, 2020. doi:10.1007/s10479-018-3091-9.
- S. Ahmed and A. Shapiro. Solving Chance-Constrained Stochastic Programs via Sampling and Integer Programming. In *State-of-the-art decision-making tools in the information-intensive age*, pages 261–269. Informs, 2008. doi:10.1287/educ.1080.0048.
- S. Ahmed and W. Xie. Relaxations and approximations of chance constraints under finite distributions. *Mathematical Programming*, 170:43–65, 2018. doi:10.1007/s10107-018-1295-z.
- K. Aigner, J. Clarner, F. Liers, and A. Martin. Robust approximation of chance constrained DC optimal power flow under decision-dependent uncertainty. *Eur. J. Oper. Res.*, 301(1):318–333, 2022. doi:10.1016/J.EJOR.2021.10.051.
- K. Aigner, P. Schaumann, F. v. Loeper, A. Martin, V. Schmidt, and F. Liers. Robust DC optimal power flow with modeling of solar power supply uncertainty via R-vine copulas. *Optimization and Engineering*, 24(3):1573–2924, 2023. doi:10.1007/s11081-022-09761-0.
- D. Afmann, F. Liers, and M. Stingl. Decomposable robust two-stage optimization: An application to gas network operations under uncertainty. *Networks*, 74(1):40–61, 2019. doi:10.1002/net.21871.

- X. Bai, X. Zheng, and X. Sun. A survey on probabilistically constrained optimization problems. *Numerical Algebra, Control and Optimization*, 2(4):767–778, 2012. doi:10.3934/naco.2012.2.767.
- X. Bai, J. Sun, and X. Zheng. An Augmented Lagrangian Decomposition Method for Chance-Constrained Optimization Problems. *INFORMS Journal on Computing*, 33(3):1056–1069, 2021. doi:10.1287/ijoc.2020.1001.
- A. Ben-Tal and A. Nemirovski. Robust Convex Optimization. *Mathematics of operations research*, 23(4):769–805, 1998. doi:10.1287/moor.23.4.769.
- A. Ben-Tal and A. Nemirovski. Robust solutions of Linear Programming problems contaminated with uncertain data. *Mathematical programming*, 88:411–424, 2000. doi:10.1007/PL00011380.
- P. Beraldi and A. Ruszczyński. A branch and bound method for stochastic integer problems under probabilistic constraints. *Optimization Methods and Software*, 17(3):359–382, 2002. doi:10.1080/1055678021000033937.
- D. Bertsimas and M. Sim. The Price of Robustness. *Operations research*, 52(1):35–53, 2004. doi:10.1287/opre.1030.0065.
- M. Branda. Chance constrained problems: penalty reformulation and performance of sample approximation technique. *Kybernetika*, 48(1):105–122, 2012a. URL <http://eudml.org/doc/246573>.
- M. Branda. Sample approximation technique for mixed-integer stochastic programming problems with several chance constraints. *Operations Research Letters*, 40(3):207–211, 2012b. doi:10.1016/j.orl.2012.01.002.
- M. Branda. Stochastic programming problems with generalized integrated chance constraints. *Optimization*, 61(8):949–968, 2012c. doi:10.1080/02331934.2011.587007.
- M. Branda. On relations between chance constrained and penalty function problems under discrete distributions. *Mathematical Methods of Operations Research*, 77:265–277, 2013. doi:10.1007/s00186-013-0428-7.
- M. Branda and J. Dupačová. Approximation and contamination bounds for probabilistic programs. *Annals of Operations Research*, 193:3–19, 2012. doi:10.1007/s10479-010-0811-1.
- G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005. doi:10.1007/s10107-003-0499-y.
- Y. Cao and V. M. Zavala. A Sigmoidal Approximation for Chance-Constrained Nonlinear Programs. *arXiv preprint arXiv:2004.02402*, 2020. doi:10.48550/arXiv.2004.02402.
- A. Charnes, W. W. Cooper, and G. H. Symonds. Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil. *Management science*, 4(3):235–263, 1958. doi:10.1287/mnsc.4.3.235.

- W. Chen, M. Sim, J. Sun, and C.-P. Teo. From CVaR to Uncertainty Set: Implications in Joint Chance-Constrained Optimization. *Operations research*, 58(2):470–485, 2010. doi:10.1287/opre.1090.0712.
- F. H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM, 1990.
- R. M. Freund. Penalty and Barrier Methods for Constrained Optimization. *Lecture Notes, Massachusetts Institute of Technology*, 2004.
- A. Geletu, A. Hoffmann, M. Kloppel, and P. Li. An Inner-Outer Approximation Approach to Chance Constrained Optimization. *SIAM Journal on Optimization*, 27(3):1834–1857, 2017. doi:10.1137/15M1049750.
- C. Gotzes, H. Heitsch, R. Henrion, and R. Schultz. On the quantification of nomination feasibility in stationary gas networks with random load. *Mathematical Methods of Operations Research*, 84:427–457, 2016. doi:10.1007/s00186-016-0564-y.
- M. Grieshammer, L. Pflug, M. Stingl, and A. Uihlein. The continuous stochastic gradient method: part I—convergence theory. *Computational Optimization and Applications*, 2023a. doi:10.1007/s10589-023-00542-8.
- M. Grieshammer, L. Pflug, M. Stingl, and A. Uihlein. The continuous stochastic gradient method: part II—application and numerics. *Computational Optimization and Applications*, 2023b. doi:10.1007/s10589-023-00540-w.
- A. Hantoute, R. Henrion, and P. Pérez-Aros. Subdifferential characterization of probability functions under Gaussian distribution. *Mathematical Programming*, 174:167–194, 2019. doi:10.1007/s10107-018-1237-9.
- L. J. Hong, Y. Yang, and L. Zhang. Sequential Convex Approximations to Joint Chance Constrained Programs: A Monte Carlo Approach. *Operations Research*, 59(3):617–630, 2011. doi:10.1287/opre.1100.0910.
- Z. Hu, L. J. Hong, and L. Zhang. A smooth Monte Carlo approach to joint chance-constrained programs. *IIE Transactions*, 45(7):716–735, 2013. doi:10.1080/0740817X.2012.745205.
- R. Kannan and J. R. Luedtke. A stochastic approximation method for approximating the efficient frontier of chance-constrained nonlinear programs. *Mathematical Programming Computation*, 13(4):705–751, 2021. doi:10.1007/s12532-020-00199-y.
- T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. *Evaluating Gas Network Capacities*. SIAM, 2015.
- M. Kuchlbauer. *Mixed-integer nonlinear robust optimization via nonsmooth methods*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2023. URL <https://nbn-resolving.org/urn:nbn:de:bvb:29-opus4-219102>.
- M. Kuchlbauer, F. Liers, and M. Stingl. Adaptive Bundle Methods for Nonlinear Robust Optimization. *Informatics Journal on Computing*, 34(4):2106–2124, 2022a. doi:10.1287/ijoc.2021.1122.

- M. Kuchlbauer, F. Liers, and M. Stingl. Outer Approximation for Mixed-Integer Nonlinear Robust Optimization. *Journal of Optimization Theory and Applications*, 195(3):1056–1086, 2022b. doi:10.1007/s10957-022-02114-y.
- Z. Li and Z. Li. Optimal robust optimization approximation for chance constrained optimization problem. *Computers & Chemical Engineering*, 74:89–99, 2015. doi:10.1016/j.compchemeng.2015.01.003.
- J. Luedtke. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming*, 146(1-2):219–244, 2014. doi:https://doi.org/10.1007/s10107-013-0684-6.
- J. Luedtke and S. Ahmed. A Sample Approximation Approach for Optimization with Probabilistic Constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008. doi:10.1137/070702928.
- J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming*, 122(2):247–272, 2010. doi:10.1007/s10107-008-0247-4.
- K. Margellos, P. Goulart, and J. Lygeros. On the Road Between Robust Optimization and the Scenario Approach for Chance Constrained Optimization Problems. *IEEE Transactions on Automatic Control*, 59(8):2258–2263, 2014. doi:10.1109/TAC.2014.2303232.
- B. L. Miller and H. M. Wagner. Chance Constrained Programming with Joint Constraints. *Operations Research*, 13(6):930–945, 1965. doi:10.1287/opre.13.6.930.
- A. Nemirovski. On safe tractable approximations of chance constraints. *European Journal of Operational Research*, 219(3):707–718, 2012. doi:10.1016/j.ejor.2011.11.006.
- A. Nemirovski and A. Shapiro. Scenario Approximations of Chance Constraints. *Probabilistic and randomized methods for design under uncertainty*, pages 3–47, 2006. doi:10.1007/1-84628-095-8_1.
- A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007. doi:10.1137/050622328.
- J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.
- B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications. *Journal of optimization theory and applications*, 142(2):399–416, 2009. doi:10.1007/s10957-009-9523-6.
- A. Peña-Ordieres, J. R. Luedtke, and A. Wächter. Solving Chance-Constrained Problems via a Smooth Sample-Based Nonlinear Approximation. *SIAM Journal on Optimization*, 30(3):2221–2250, 2020. doi:10.1137/19M1261985.
- L. Pflug, N. Bernhardt, M. Grieshammer, and M. Stingl. CSG: A new stochastic gradient method for the efficient solution of structural optimization problems with infinitely many states. *Structural and Multidisciplinary Optimization*, 61:2595–2611, 2020. doi:10.1007/s00158-020-02571-x.

- Á. Porras, C. Domínguez, J. M. Morales, and S. Pineda. Tight and Compact Sample Average Approximation for Joint Chance-Constrained Problems with Applications to Optimal Power Flow. *INFORMS Journal on Computing*, 2023. doi:10.1287/ijoc.2022.0302.
- A. Prékopa. Probabilistic Programming. *Handbooks in operations research and management science*, 10:267–351, 2003. doi:10.1016/S0927-0507(03)10005-9.
- Y. H. Ren, Y. Xiong, Y. H. Yan, and J. Gu. A smooth approximation approach for optimization with probabilistic constraints based on sigmoid function. *Journal of Inequalities and Applications*, 2022(1):38, 2022. doi:10.1186/s13660-022-02774-4.
- H. Robbins and S. Monro. A Stochastic Approximation Method. *The annals of mathematical statistics*, pages 400–407, 1951. URL <http://www.jstor.org/stable/2236626>.
- R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000. doi:10.21314/JOR.2000.038.
- A. Ruszczyński. Probabilistic programming with discrete distributions and prece-
dence constrained knapsack polyhedra. *Mathematical Programming*, 93:195–215, 2002. doi:10.1007/s10107-002-0337-7.
- M. Schmidt, D. Aßmann, R. Burlacu, J. Humpola, I. Joormann, N. Kanelakis, T. Koch, D. Oucherif, M. E. Pfetsch, L. Schewe, R. Schwarz, and M. Sirvent. GasLib – A Library of Gas Network Instances. *Data*, 2(4):article 40, 2017a. doi:10.3390/data2040040.
- M. Schmidt, N. Le Roux, and F. Bach. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162:83–112, 2017b. doi:10.1007/s10107-016-1030-6.
- F. Shan, L. Zhang, and X. Xiao. A Smoothing Function Approach to Joint Chance-Constrained Programs. *Journal of Optimization Theory and Applications*, 163:181–199, 2014. doi:10.1007/s10957-013-0513-3.
- F. Shan, X. Xiao, and L. Zhang. Convergence analysis on a smoothing approach to joint chance constrained programs. *Optimization*, 65(12):2171–2193, 2016. doi:10.1080/02331934.2016.1233551.
- A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.
- C. Stangl. *Modelle, Strukturen und Algorithmen für stationäre Flüsse in Gasnetzen*. PhD thesis, Universitätsbibliothek Duisburg-Essen, 2014. URL <https://nbn-resolving.org/urn:nbn:de:hbz:464-20140813-133151-1>.
- H. Sun, H. Xu, and Y. Wang. Asymptotic Analysis of Sample Average Approximation for Stochastic Optimization Problems with Joint Chance Constraints via Conditional Value at Risk and Difference of Convex Functions. *Journal of Optimization Theory and Applications*, 161:257–284, 2014. doi:10.1007/s10957-012-0127-1.
- Y. Yang and C. Sutanto. Chance-constrained optimization for nonconvex programs using scenario-based methods. *ISA transactions*, 90:157–168, 2019. doi:10.1016/j.isatra.2019.01.013.

Y. Yuan, Z. Li, and B. Huang. Robust optimization approximation for joint chance constrained optimization problem. *Journal of Global Optimization*, 67:805–827, 2017. doi:10.1007/s10898-016-0438-0.