A tractable approximation of non-convex chance constrained optimization with non-Gaussian uncertainties

Abebe Geletua\textsuperscript{a}, Michael Klöppel\textsuperscript{a}, Armin Hoffmann\textsuperscript{b} and Pu Li\textsuperscript{a}

\textsuperscript{a}Department of Simulation and Optimal Processes, Institute for Automation and Systems Engineering, Technische Universität Ilmenau, P.O. Box 10 05 65, 98684 Ilmenau, Germany;\textsuperscript{b}Department of Operations Research and Stochastic, Institute of Mathematics, Technische Universität Ilmenau, P.O. Box 10 05 65, 98684 Ilmenau, Germany

(Received 7 March 2013; accepted 24 February 2014)

Chance constrained optimization problems in engineering applications possess highly nonlinear process models and non-convex structures. As a result, solving a nonlinear non-convex chance constrained optimization (CCOPT) problem remains as a challenging task. The major difficulty lies in the evaluation of probability values and gradients of inequality constraints which are nonlinear functions of stochastic variables. This article proposes a novel analytic approximation to improve the tractability of smooth non-convex chance constraints. The approximation uses a smooth parametric function to define a sequence of smooth nonlinear programs (NLPs). The sequence of optimal solutions of these NLPs remains always feasible and converges to the solution set of the CCOPT problem. Furthermore, Karush–Kuhn–Tucker (KKT) points of the approximating problems converge to a subset of KKT points of the CCOPT problem. Another feature of this approach is that it can handle uncertainties with both Gaussian and/or non-Gaussian distributions.

Keywords: chance constraints; non-convex optimization; analytic approximation; non-Gaussian distribution

1. Introduction

In recent years, there has been growing interest in chance constrained optimization (CCOPT) methods in engineering and financial applications. Applied optimization problems subject to constraints involving uncertain variables may become infeasible if they are required to be satisfied deterministically. Therefore, chance constrained optimization provides decision and planning strategies that guarantee optimal system performance as well as reliability in product (output) quality. In addition, CCOPT methods have become indispensable tools of risk management and risk-metrics in financial applications in the context of value-at-risk (VaR) and conditional value-at-risk (cVaR).

Chance constrained optimization models were first introduced by Charnes, Cooper, and Symmonds (1958) and Charnes and Cooper (1959) in relation to financial planning problems. Some recent areas of application are: water reservoir management (Andrieu, Henrion, and Römisch 2010; Dhar and Datta 2006; Ouarda and Labadie 2001); chemical process engineering (Arellano-Garcia and Wozny 2009; Li et al. 2002; Flemming, Bartl, and Li 2007; Li,
nonlinear non-convex chance constrained optimization problem of the form

\[
\text{(CCOPT)} \quad \min_u E \left[ f(x, u, \xi) \right]
\]

subject to

\[
G(x, u, \xi) = 0,
\]

\[
\Pr \left\{ g_j(x, u, \xi) \leq 0 \right\} \geq \alpha_j, \quad j = 1, \ldots, q,
\]

\[
u \in \mathcal{U},
\]

where \( x \in \mathbb{R}^n \) is a vector of state variables, \( u \in \mathcal{U} \subset \mathbb{R}^m \) is a vector of inputs, \( \mathcal{U} \) is a compact set and \( \xi \) is a vector of Borel-measurable random (uncertain) input variables from a probability space into \( \mathbb{R}^p \). The random vector \( \xi \) is assumed to have a known joint probability density function \( \phi(\xi) \) with support on the Borel-measurable set \( \Omega \subset \mathbb{R}^p \). The density \( \phi(\xi) \) can be either a Gaussian or non-Gaussian probability density function. The functions \( f, g_j, G : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R} \) are at least one-time continuously differentiable with respect to \( x, u \) and \( \xi \). The operators \( E(\cdot) \) and \( \Pr(\cdot) \) represent expected value and probability, respectively. The expression \( \Pr\{g_j(x, u, \xi) \leq 0\} \geq \alpha_j \) is a chance constraint. It specifies the probability of holding the constraint \( g_j(x, u, \xi) \leq 0 \) with a reliability level \( \alpha_j \), where \( 1/2 \leq \alpha_j \leq 1 \). The nonlinear equality constraints \( G(x, u, \xi) = 0 \) are required to be satisfied for all possible realizations of the uncertain variables \( \xi \).

Nonlinear equality constraints of the type \( G(x, u, \xi) = 0 \) usually arise in practical engineering applications. In general, the decision (control) variables \( u \) are deterministic, while the state variables \( x \) of the processes are influenced by both controls \( u \) and uncertainties \( \xi \) as indicated by \( x(u, \xi) \). Hence, as a result of the nonlinearity of \( G(x, u, \xi) \) and the nature of the distribution of the uncertainties \( \xi \), the functions \( p_j(u) := \Pr\{g_j(x, u, \xi) \leq 0\}, j = 1, \ldots, q, \) may not be convex. Furthermore, random input variables \( \xi \) are traditionally assumed to have Gaussian distributions. Nevertheless, this is not always the case in many practical applications. For instance, the wind speed in wind power generation (Celik, Makkawi, and Muneer 2010) and the clearness index in solar energy (Ibanez, Rosell, and Beckman 2003) are found to be Beta and Gamma distributed, respectively. Therefore, there is a strong practical need for the consideration of nonlinear non-convex chance constrained optimization problems in the presence of non-Gaussian distributed random variables. Despite their prevalence in engineering applications, there is no work known to the best knowledge of the authors that studies CCOPT with non-convex \( g(u, x, \xi) \) and non-Gaussian uncertainties. This work proposes a smooth analytic approximation strategy in order to improve the tractability of nonlinear non-convex CCOPT models with non-Gaussian uncertainties. Nevertheless, if the uncertain input variables cannot be modelled as random variables with known probability distributions, it is appropriate to use the robust optimization approach instead of CCOPT, in which the input uncertainties are commonly assumed to belong to a bounded set. Since the robust optimization approach (Ben-Tal and Nemirovski 1998, 1999) attempts to satisfy constraints under almost all possible realizations of the uncertainties, generally it may lead to a semi-infinite
optimization problem (López and Still 2007). Some classes of chance constrained or robust optimization problems can also be modelled as semi-definite optimization problems, e.g. Bertsimas, David, and Caramanis (2011), Feng, Dabbene, and Lagoa (2011), Jasour and Lagoa (2012), Zymler, Kuhn, and Rustem (2013), etc. In particular, Özmen et al. (2011, 2013) Özmen, Weber, and Karimov (2013), Weber, Çavuşoğlu, and Özmen (2012), Werner (2008) and Schöttle and Werner (2013) consider robust optimization approaches in financial risk analysis and portfolio optimization problems that lead to conic or semi-definite optimization problems.

The remainder of the article is organized as follows. Section 2 presents a brief review on latest developments in methods of chance constrained optimization. Section 3 puts to-date-available analytic approximation schemes into a general context, suggests a novel smooth and monotonic function for analytic approximation and studies its properties. Section 4 presents a sequence of nonlinear programs (NLPs) as an approximation to CCOPT. This section investigates Hausdorff convergence of the sequence of feasible sets of these NLPs to the feasible set of CCOPT problems, approximability of optimal solutions of CCOPT problems by optimal solutions of the NLPs and convergence of the set of Karush–Kuhn–Tucker (KKT) points of the NLPs to a subset of KKT-points of CCOPT problems. Numerical computation results and a case study in Section 5 demonstrate the viability and effectiveness of the proposed approximation scheme. The article concludes in Section 6 with a summary and some future research directions. Furthermore, details of mathematical proofs and additional references are given in an online supplement to this article (see the section ‘Supplemental data’ before the references list).

2. State-of-the-art solution of nonlinear chance constrained optimization problems

For the sake of brevity and clarity of presentation, the discussion here considers CCOPT with one chance constraint \( \Pr\{g(x, u, \xi) \leq 0\} \geq \alpha \). However, the results presented easily carry over to the case of multiple chance constraints as in inequality (3) (see, for instance, Section 4.3). Since the state (output) variables \( x \) depend on the input control variables \( u \) and the random variables \( \xi \), the CCOPT problem can be projected onto the solution set of the equation \( G(x, u, \xi) = 0 \) for given \( u \) and a realization of \( \xi \). Subsequently, CCOPT can be written equivalently as

\[
\text{(CCOPT)} \quad \min_u F(u) \tag{5}
\]

subject to

\[
p(u) \geq \alpha, \tag{6}
\]

\[
u \in U, \tag{7}
\]

where \( F(u) := E[f(x(u, \xi), u, \xi)] \) and \( p(u) := \Pr\{g(x(u, \xi), u, \xi) \leq 0\} \). To make the ensuing presentation more transparent, in the following the exclusive dependence on \( x \) can be dropped so that \( f(u, \xi) = f(x(u, \xi), u, \xi) \) and \( p(u) := \Pr\{g(u, \xi) \leq 0\} = \Pr\{g(x(u, \xi), u, \xi) \leq 0\} \).

Remark 1 Assuming the applicability of the implicit function theorem for Equation (2), the smoothness properties of \( G \) can be translated to corresponding smoothness properties of the solution function \( x(u, \xi) \). Hence, \( f(u, \xi) = f(x(u, \xi), u, \xi) \) and \( g(u, \xi) = g(x(u, \xi), u, \xi) \) are smooth, given the smoothness of \( G(x, u, \xi), f(x, u, \xi) \) and \( g(x, u, \xi) \).

In CCOPT problems, the main object of investigation is the probability function \( p(u) := \Pr\{g(u, \xi) \leq 0\} \). Currently, the structural properties of \( p(u) \) such as continuity (Theorem 10.1.1., p. 302 in Prékopa 1995), convexity (Prékopa, Yoda, and Subasi 2011; Henrion and Strugarek...
The potential difficulty associated with SAA is that relative-frequency approximation of a probability requires a very large sample size. Thus, a solution obtained through the SAA approach may fail to be feasible to CCOPPT.

The work by Garnier, Omrane, and Rouchdy (2009) considers a linearization of the form

\[ g(u, \xi) \cong g(u, \mu) + \sum_{k=1}^{p} \frac{\partial g}{\partial \xi_k}(u, \mu)\xi_k, \quad \text{where} \quad \xi \sim N(\mu, \sigma^2). \]
This provides a decoupling of the random variables $\xi$ from the (deterministic) problem variables $u$. Nevertheless, such a linearization scheme is valid only when the random variables $\xi$ have small variance. In fact, recently, Nemirovski (2012) considered tractability issues of chance constraints of the form $\Pr\{w(u) + \sum_{k=1}^{p} \xi_k w_k(u) \geq \alpha\}$.

Possible infeasibility of solutions obtained from scenario generation and SAA methods are major drawbacks. In fact, these methods require enormous computational effort to achieve only approximate feasibility of solutions. Therefore, analytic methods replace chance constraints with some approximations and attempt to provide an a priori guarantee of feasibility along with tractability. Pinter (1989) was the first to propose such analytic approximation schemes. Nemirovski and Shapiro (2006a) and Nemirovski (2012) have suggested analytic approximations for convex chance constraints that are affine with respect to the uncertain variables. Recently, Feng, Dabbene, and Lagoa (2011) proposed an analytic approximation approach through the concept of a kinship function that is applicable when the chance constrained function is convex w.r.t. the decision variable $u$ and polynomial w.r.t. to the uncertain variables $\xi$. This approach benefits from enormous computational advantages by providing efficient evaluation of chance constraints. Furthermore, based on the strategy suggested by Rockafellar and Uryasev (2000), Hong, Yang, and Zhang (2011) (see also Xiao et al. 2012, Wozabal 2012 and Wozabal, Hochreiter, and Pflug 2010) have successfully shown that approximate solutions of convex chance constrained optimization problems can be obtained by solving a sequence of difference-convex (DC) deterministic optimization problems. This is a breakthrough in the numerical solution of nonlinear convex chance constrained optimization problems. Nevertheless, difference-convex algorithms are inherently limited to problems of smaller dimensions. At the same time, the DC approach benefits from little computational advantage for general non-convex CCOPT. Furthermore, the work by Norkin (1993) suggests a method similar to non-parametric density estimation (for $Z = g(u, \xi)$) to setup an analytic approximation, but it needs a modification (as indicated in Section 3.3) in order to make it applicable for problems with non-symmetric and non-Gaussian distributions. The works of Feng, Dabbene, and Lagoa (2011) and Jasour and Lagoa (2012) also suggest relaxation strategies.

This work proposes an analytic approximation strategy for non-convex chance constraints that guarantees a priori feasibility of solutions to the chance constraints, a smooth approximation, simplicity of implementation and applicability to large-scale non-convex engineering optimization problems.

3. A general analytic approximation

3.1. Further properties of CCOPT

The probability function $p(u) = \Pr\{g(u, \xi) \leq 0\}$ can alternatively be expressed as

$$1 - p(u) = \Pr\{g(u, \xi) > 0\} = E[h(u, \xi)],$$

where $h(u, \xi) := \begin{cases} 0, & \text{if } g(u, \xi) \leq 0; \\ 1, & \text{if } g(u, \xi) > 0. \end{cases}$

This immediately yields the equivalence

$$\Pr\{g(u, \xi) \leq 0\} \geq \alpha \iff [1 - p(u)] \leq 1 - \alpha \iff E[h(u, \xi)] \leq 1 - \alpha. \quad (8)$$

It is assumed that $p(u)$ is a continuous function of $u$ (see Prékopa 1995, p. 302). By Remark 1, $g(u, \xi)$ is differentiable with respect to $u$ and $E[h(\cdot, \xi)]$ is a lower semi-continuous function of $u$ (according to Lebesgue’s Integration Theorem). Similarly, by Remark 1, $u \mapsto f(u, \xi)$ is continuous and $E[f(u, \xi)] < \infty$ for all $u \in U$. Then the function $F : U \rightarrow \mathbb{R}$ such that $F(u) = E[f(u, \xi)]$ is continuous on $U$. 

The feasible set of CCOPT $\mathcal{K} := \{u \in \mathcal{U} \mid p(u) \geq \alpha\}$ and the set $\mathcal{M} := \{u \in \mathcal{U} \mid E[h(u, \xi)] \leq 1 - \alpha\}$ are compact, by the compactness of $\mathcal{U}$. Furthermore, from Equation (8), it follows that

$$\mathcal{K} = \{u \in \mathcal{U} \mid p(u) \geq \alpha\} = \{u \in \mathcal{U} \mid E[h(u, \xi)] \leq 1 - \alpha\} = \mathcal{M}.$$ 

Therefore, both problems $\min_{u \in \mathcal{K}} F(u)$ and

$$\begin{equation}
(NLP2) \quad \min_u F(u) \tag{9}
\end{equation}$$

subject to

$$E[h(u, \xi)] \leq 1 - \alpha, \quad u \in \mathcal{U}, \tag{10}$$

possess (the same) optimal solution set. Despite the fact that the constraint $E[h(u, \xi)] \leq 1 - \alpha$ provides an exact representation of the chance constraint, the function $h$ causes difficulties in numerical computations. Even so, the function $E[h(u, \xi)]$ provides some insight for the construction of a smooth, tight and tractable approximation of CCOPT.

### 3.2. A generalized analytic approximation of chance constraints

A general setup for the analytic approximation of chance constraints with an a priori feasibility guarantee is to determine a continuous function $\psi : \mathbb{R}_{++} \times \mathbb{R}^m \to \mathbb{R}_+$ with the following properties:

- **P1:** $E[h(u, \xi)] \leq \psi(\tau, u)$, for each $\tau > 0$ and $u \in \mathcal{U}$,
- **P2:** $\inf_{\tau > 0} \psi(\tau, u) = E[h(u, \xi)]$, for each $u \in \mathcal{U}$,
- **P3:** $\psi(\tau, u)$ is non-decreasing w.r.t. $\tau$,

where $\mathbb{R}_+ := [0, +\infty)$ and $\tau \in \mathbb{R}_{++} := (0, +\infty)$ is a parameter. From properties P2, P3 and the continuity of $\psi(\tau, u)$ w.r.t. $\tau$ it follows that

$$\inf_{\tau > 0} \psi(\tau, u) = \lim_{\tau \searrow 0^+} \psi(\tau, u).$$

Once such a function is available, the constraint $E[h(u, \xi)] \leq 1 - \alpha$ in (NLP2) can be replaced by the constraint $\inf_{\tau > 0} \psi(\tau, u) \leq 1 - \alpha$. As a result of property P1, for each fixed $\tau > 0$, the optimal solution of the problem

$$\begin{equation}
(NLP_{\tau}) \quad \min_u F(u) \tag{12}
\end{equation}$$

subject to

$$\psi(\tau, u) \leq 1 - \alpha, \quad u \in \mathcal{U}, \tag{13}$$

is feasible to (NLP2); consequently, feasible to CCOPT. Moreover, the feasible set $M(\tau) = \{u \in \mathcal{U} \mid \psi(\tau, u) \leq 1 - \alpha\}$ of NLP$_\tau$ is always a subset of the feasible set of NLP2, so of CCOPT. Most importantly, properties P2 and P3 imply that

$$\limsup_{\tau \searrow 0^+} M(\tau) = \mathcal{M} = \mathcal{K}, \tag{15}$$

in the sense of convergence of sets (see the online supplement ‘Supplemental data’ before the references list). In fact, in Equation (5) the equality $\limsup_{\tau \searrow 0^+} M(\tau) = \lim_{\tau \searrow 0^+} M(\tau) = \mathcal{M} = \mathcal{K}$.
\( \mathcal{K} \) holds true, due to the compactness of each \( M(\tau) \) and the monotony property \( M(\tau_1) \subset M(\tau_2) \) for \( 0 < \tau_2 < \tau_1 \).

Basically, the so-called concentration of measure inequalities from probability theory can be used as a background to construct a function \( \psi \) with properties P1–P3 (Pinter 1989; Nemirovski and Shapiro 2006a). For example, the function \( \psi(\tau, u) := E[\Theta(\tau, g(u, \xi))] \) for some given real-valued function \( \Theta : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+ \). In the literature, there are some suggestions for \( \Theta \) where the associated function \( \psi \) satisfies some of the properties P1–P3:

- \( \Theta_1(\tau, s) = \exp(\tau^{-1}s), \tau > 0 \) (Pinter 1989; Nemirovski and Shapiro 2006a);
- \( \Theta_2(\tau, s) = \tau^{-1} \int_{-\infty}^{0} k((z + s)/\tau) \, dz, \tau > 0 \) (Norkin 1993), where \( k(s) \) is a non-negative integrable and symmetric (i.e. \( k(s) = k(-s) \)) function satisfying \( \int_{-\infty}^{\infty} k(s) \, ds = 1 \);
- \( \Theta_3(\tau, s) = \tau + (1 - \alpha)^{-1}[s - \tau]_+, \tau > 0 \) (Rockafellar and Uryasev 2000), where \( [s - \tau]_+ = \max\{s - \tau, 0\} \);
- \( \Theta_4(\tau, s) = p_d(s) \) (Feng, Dabbene, and Lagoa 2011) with \( \tau = 1/d \), where \( p_d(s) \) is a non-negative and non-decreasing polynomial function on \([-1, +\infty)\) with degree \( d, d \in \mathbb{N} \), and with properties that \( p_d(0) = 1, p_d(-1) = 0 \) and \( p''_d(s) \geq 0, s \in [-1, +\infty) \).

The strategy of using the function \( \Theta_1(\tau, s) \) provides a convex approximation of \( p(u) \) when \( g(u, \xi) \) is an affine function of \( \xi \). Nevertheless, this approximation scheme is not directly applicable to the computation of general non-convex chance constraints. The approximation based on \( \Theta_2(\tau, s) \) can be generalized by lifting the symmetric assumption on \( k(s) \). While concrete non-symmetric function \( k(s) \) is easy to construct, the work of Norkin (1993) attempts to solve CCOPT problems through a sequence of non-smooth optimization problems which limits its wider applicability. The function \( \Theta_4(\tau, s) \) preserves convexity and polynomial structures in the chance constrained function. However, when problems have complex nonlinearities, these structures can be hard to verify unless they are readily available (see Section 5 for numerical examples). Recent investigations related to value-at-risk (VaR) and conditional VaR (cVaR) make use of the continuous piecewise-differentiable function \( \Theta_3(\tau, s) \). This function can help approximate CCOPT through DC optimization problems (Hong, Yang, and Zhang 2011; Wozabal 2012; Wozabal, Hochreiter, and Pflug 2010; Xiao et al. 2012). The following section introduces a novel smooth function \( \Theta(\tau, s) \) satisfying the properties P1–P3 which can be used for the approximation of general smooth non-convex chance constraints.

### 3.3. A new analytic approximation and its properties

The method proposed here uses an analytic approximation based on the parametric function

\[
\Theta(\tau, u, s) = \frac{1 + \tau m_1(u)}{1 + \tau m_2(u) \exp\left(-\frac{1}{\tau}\right)}, \text{ such that } \psi(\tau, u) := E\left[\Theta(\tau, u, g(u, \xi))\right], \tag{16}
\]

where \( m_1(u) \) and \( m_2(u) \) are given positive functions (or constants), \( m_2(u) \leq m_1(u) \) and \( 0 < \tau < 1 \). The function \( \Theta(\tau, u, s) \) possesses several salient properties that make it amenable for the intended approximation.

**Proposition 3.1** Suppose \( 0 < m_2(u) \leq m_1(u) \) and \( \tau > 0 \). Then the function \( \Theta \) has the following properties:

1. \( \Theta(\tau, u, s) > 0 \), for any values of \( s \),
2. \( \Theta(\tau, u, s) \geq 1 \), for \( s \geq 0 \).
Proof Part (1) is trivial. To verify (2), observe that
\[ m_2(u) \tau \leq m_1(u) \tau \Rightarrow m_2(u) \tau e^{-s/\tau} \leq m_1(u) \tau e^{-s/\tau}, \text{ since } e^{-s/\tau} \leq 1 \text{ for } s \geq 0. \]
This implies
\[ 1 + m_2(u) \tau e^{-s/\tau} \leq 1 + m_1(u) \tau \Rightarrow \frac{1 + m_1(u) \tau}{1 + m_2(u) \tau e^{-s/\tau}} \geq 1. \]
\[
\square
\]

Assumption 3.2

(A) The functions \( m_1(u) \) and \( m_2(u) \) are chosen in such a way that \( 0 < m_2(u) \leq \frac{m_1(u)}{1 + m_1(u)} \).

(B) There are constants \( C_1, C_2 \) and \( C_3 \) such that
\[
\sup_{u \in U} m_1(u) = C_1 < \infty, \quad \sup_{u \in U} m_2(u) = C_2 < \infty \quad \text{and} \quad \inf_{u \in U} m_2(u) = C_3 > 0.
\]

Proposition 3.3 Suppose Assumption 3.2 holds true. Then

(1) \( \Theta(\tau, u, \cdot) \) is a strictly increasing function w.r.t. \( s \in \mathbb{R} \);

(2) \( \Theta(\cdot, u, s) \) is a non-decreasing function w.r.t. \( \tau \) for \( 0 < \tau < 1 \);

(3) and
\[
\lim_{\tau \searrow 0^+} \Theta(\tau, u, s) = \begin{cases} 
1, & \text{if } s \geq 0, \\
0, & \text{if } s < 0,
\end{cases}
\]
uniformly for \( u \in U \) and uniformly for \( s \in (-\infty, -\varepsilon] \cup [0, +\infty) \) and arbitrary \( \varepsilon > 0 \).

Proof For the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list. \[
\square
\]

Assumption 3.4 For each fixed \( u \in U \), the random set \( \Gamma_0(u) := \{ \xi \in \Omega \mid g(u, \xi) = 0 \} \) is of probability zero; i.e. \( \Pr(\Gamma_0(u)) = 0 \).

Corollary 3.5 If in Proposition 3.3 the variable \( s \) is replaced by \( g(u, \xi) \), and if Assumptions 3.2 and 3.4 hold true, then
\[
\lim_{\tau \searrow 0^+} \psi(\tau, u) = \lim_{\tau \searrow 0^+} E \left[ \Theta(\tau, u, g(u, \xi)) \right] = E \left[ h(u, \xi) \right] = 1 - p(u). \tag{17}
\]

Proposition 3.3 remains valid if \( m_1 \) and \( m_2 \) are chosen as positive constants according to Assumption 3.2(A).

Figure 1 depicts a comparison of the functions \( \Theta_1(\tau, s) \) of Nemirovski and Shapiro (2006a) and \( \Theta(\tau, u, s) \) of the new approximation (16) for \( m_1 = 1 \) and \( m_2 = 0.5 \). It can be seen that the approximation profile of \( \Theta(\tau, u, s) \) becomes much tighter as \( \tau \) decreases towards \( 0^+ \). In addition, instead of designing a tighter approximation for \( E[h(x, \xi)] \) by taking infimum with respect to \( \tau \) (cf. Nemirovski and Shapiro 2006a), a monotonic reduction of \( \Theta(\tau, u, s) \) is computationally much more preferable.
4. A sequential approximation of CCOPT and convergence properties

For a fixed parameter value $\tau_k \in (0, 1)$, consider the nonlinear optimization problem

$$(\text{NLP}_k) \quad \min_u E[f(u, \xi)]$$

subject to

$$u \in M(\tau_k),$$

where the feasible set $M(\tau_k) = \{u \in \mathcal{U} | \psi(\tau_k, u) \leq 1 - \alpha\}$ and the objective function $E[f(u, \xi)]$ are the same as in CCOPT. Let $\{\tau_k\}_{k=1}^{+\infty} \subset (0, 1)$ be a decreasing sequence of parameters (i.e. $\tau_{k+1} < \tau_k$) and

$$\tau_k \to 0^+. \quad (19)$$

Then this leads to a sequence of nonlinear optimization problems $\text{NLP}_k$, $k = 1, 2, \ldots$ Since $f(\cdot, \xi)$ and $\psi(\tau_k, \cdot)$ are continuous functions on the compact set $u \in \mathcal{U}$, then $M(\tau_k)$ is a compact set and $\text{NLP}_k$ has an optimal solution $u_k \in M(\tau_k)$. Furthermore, assuming that the functions $f(\cdot, \xi)$ and $\psi(\tau_k, \cdot)$ are at least once continuously differentiable (see Section 4.2), these optimal solutions of $\text{NLP}_k$ can be determined using a gradient-based nonlinear optimization algorithm. The main interest now lies in verifying how the sequence of solutions $\{u_k\}_{k=0}^{+\infty}$ of $\text{NLP}_k$ are related to the original CCOPT problem.

4.1. Approximability of optimal solutions of CCOPT

4.1.1. Convergence of feasible sets

This section makes use of definitions and properties for the convergence of a sequence of sets. Details and relevant references are given in the online supplement.
Now, based on property P2 (Section 3), for any \( \tau \in (0, 1) \), the set \( M(\tau) \subset \mathcal{K} \). Specially, the sequence of sets \( \{M(\tau_k)\}_{k \in \mathbb{N}} \) has the convergence property that \( \limsup_{k \to \infty} M(\tau_k) \subset \mathcal{K} \), for any decreasing to zero sequence \( \{\tau_k\}_{k \in \mathbb{N}} \). Furthermore, the equality \( \lim_{k \to \infty} M(\tau_k) = \mathcal{K} \) can be guaranteed under the following assumption.

**Assumption 4.1 – Regularity.** Let \( \alpha \in [1/2, 1) \). The chance constraint \( \Pr\{g(u, \xi) \leq 0\} \geq \alpha \) is said to be regular if for each \( u \in \mathcal{U} \) with \( \Pr\{g(u, \xi) \leq 0\} = \alpha \) there is a sequence \( \{u_k\}_{k \in \mathbb{N}} \) such that \( \lim_{k \to +\infty} u_k = u \) and \( \Pr\{g(u_k, \xi) \leq 0\} > \alpha \); i.e. \( p(u_k) > \alpha \).

**Theorem 4.2** The compactness of \( \mathcal{U} \), the monotonicity and continuity of \( \psi(\cdot, u) \) with respect to \( \tau \) as well as properties P1–P3 imply that, for each decreasing to zero sequence \( \{\tau_k\}_{k \in \mathbb{N}} \) and for a regular chance constraint, the relation \( \lim_{k \to \infty} M(\tau_k) = \mathcal{K} \) holds true.

**Proof** For the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list. \( \blacksquare \)

The following statement strengthens the result of Theorem 4.2 guaranteeing the uniform convergence of the feasible sets \( M(\tau_k) \) of NLP\(_k\) to the feasible set \( \mathcal{K} \) of CC OPT; i.e. convergence w.r.t. the Hausdorff metric. The Hausdorff distance between two closed and bounded sets \( A \subset \mathbb{R}^n \) and \( B \subset \mathbb{R}^m \) is given by \( H(A, B) = \max\{\sup_{x \in A} \text{dist}(x, B), \sup_{x \in B} \text{dist}(x, A)\} \).

**Corollary 4.3** Let \( \{\tau_k\}_{k \in \mathbb{N}} \subset (0, 1) \) be any sequence that decreases to zero. Then Assumption 4.1 implies that
\[
\lim_{k \to \infty} H(M(\tau_k), \mathcal{K}) = 0.
\]

**Proof** For the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list. \( \blacksquare \)

**Remark 2** Observe that the results of Theorem 4.2 and Corollary 4.3 hold true for any sequence \( \{\tau_k\} \) such that \( \tau_{k+1} < \tau_k \) and \( \tau_k \searrow 0^+ \). These results can be further generalized as
\[
\inf_{0 < \tau < 1} H(M(\tau), \mathcal{K}) = \limsup_{\tau \searrow 0^+} H(M(\tau), \mathcal{K}) = 0,
\]
where \( M(\tau) \) is the feasible set of NLP\(_\tau\).

The discussions in this section reveal that \( M(\tau) \) can be made as close as possible to \( \mathcal{K} \) asymptotically, becoming equal to \( \mathcal{K} \) in the limit. This equality holds true irrespective of how the parameter \( \tau \) is driven to 0, from above. Consequently, instead of working with the computationally difficult set \( \mathcal{K} \) for the numerical solution of the CC OPT problem, it is convenient to work with \( M(\tau_k) \) for an arbitrary monotonically decreasing sequence \( \tau_k \searrow 0^+ \).

### 4.1.2. Convergence of approximate solutions

According to the definitions of NLP\(_k\) and the properties of \( \psi(\tau_k, u) \), it is obvious that each solution \( u_k \) of NLP\(_k\) is a feasible point of NLP2, hence, also of CC OPT. Furthermore, for each \( \tau_k \), the optimal value of NLP\(_k\) is an upper bound for the optimal value of CC OPT, i.e. \( E[f(u_k, \xi)] \geq \min_{u \in \mathcal{K}} E[f(u, \xi)] \). In the following, the compactness of the set \( \mathcal{U} \) will be used to show that any limit point of a sequence \( \{u_k\}_{k \in \mathbb{N}} \) of local optimal solutions of NLP\(_k\) is not only a feasible point, but also a local optimal solution of the original CC OPT problem.
Theorem 4.4 – The approximability of optimal solutions

(1) Let \( \{u_k\}_{k\in\mathbb{N}} \) be a sequence such that \( u_k \) is a local optimal solution of the NLP \( k \), \( k = 1, 2, \ldots \). Then there is a subsequence \( \{u_{k_l}\}_{l\in\mathbb{N}} \) of \( \{u_k\}_{k\in\mathbb{N}} \) such that \( u_{k_l} \to u^* \), there is an open ball \( B(u^*) \) around \( u^* \), \( u^* \in K \cap B(u^*) \), and

\[
E\left[f(u^*, \xi)\right] = \min_{u \in K \cap B(u^*)} E\left[f(u, \xi)\right];
\]

i.e. \( u^* \) is a local optimal solution of (CCOPT).

(2) Conversely, if \( u_0 \) is a strict local minimizer of (CCOPT), then there is a sequence of local minimizers \( u_k \) of NLP \( k \) which converges to \( u_0 \).

Proof For the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list.

The most important result in Theorem 4.4 is that any sequence \( \{u_k\}_{k\in\mathbb{N}} \) of local (or global) optimal solutions of the problems NLP \( k \) has a convergent subsequence whose limit is a local optimal solution of the original CCOPT problem. Therefore, the solution of a sequence of smooth nonlinear (non-convex) optimization problems NLP \( k \), for any zero-sequence \( \{\tau_k\}_{k\in\mathbb{N}} \), provides a working approximation for the optimal solution of the CCOPT problem where at the same time all the obtained solutions of NLP \( k \) remain feasible to CCOPT.

Since the CCOPT problem and NLP \( \tau \) are non-convex, their respective set of KKT points may not be singletons. Hence, it is necessary to show that the set of KKT points of NLP \( k \) converge to a subset of the KKT set of the CCOPT problem. To obtain this result, there should be a relation between the derivatives \( \nabla p(u) \) and \( \nabla_u \psi(\tau, u) \).

4.2. Convergence of derivatives

Given \( u_0 \in \mathcal{U} \), this section verifies the convergence

\[
\lim_{k \to \infty} \nabla \psi(\tau_k, u) = -\nabla p(u),
\]

for any sequence \( \{\tau_k\}_{k\in\mathbb{N}} \) with \( \tau_k \downarrow 0^+ \) uniformly for \( u \) in some ball around \( u_0 \).

Assumption 4.5 For each \( u \in \mathcal{U} \), \( \nabla_\xi g(u, \xi) \neq 0 \) for each \( \xi \) on the boundary of the random set \( \Gamma(u) := \{\xi \in \Omega \mid g(u, \xi) \leq 0\} \).

Obviously the set \( \Gamma_0(u) := \{\xi \in \Omega \mid g(u, \xi) = 0\} \) is a subset of the boundary \( \partial \Gamma(u) \) of the random set \( \Gamma(u) = \{\xi \in \Omega \mid g(u, \xi) \leq 0\} \). The differentiability of the probability function \( p(u) = \text{Pr}\{g(u, \xi) \leq 0\} \) has been studied based on Assumption 4.5 by Uryasev (1995) and Marti (1996). The following is a direct consequence of Theorem 2.1, p. 200 of Uryasev (1995).

Proposition 4.6 Suppose Assumption 4.5 be satisfied. If the function \( g(u, \xi) \) has continuous partial derivatives w.r.t. \( u \) and the density \( \phi \) is continuous in some small neighbourhood of the union \( \bigcup_{\eta \in B(u)} \Gamma_0(\tilde{u}) \) with some small ball \( B(u) \) around \( u \), then \( p(u) \), respectively \( 1 - p(u) \), is continuously differentiable.

In the following, the notation \( \Theta_k(u, \xi) := \Theta(\tau_k, u, \xi) \) and \( \psi_k(u) = \psi(\tau_k, u) \) are used for the sake of brevity.
Lemma 4.7 Let \( \{\tau_k\}_{k \in \mathbb{N}} \subset (0, 1) \) be a decreasing sequence with \( \lim_{k \to \infty} \tau_k = 0^+ \). Then the sequence \( \{\psi_k(u)\}_{k \in \mathbb{N}} \) converges uniformly to the probability function \( 1 - p(u) \).

Proof Since \( \mathcal{U} \) is a compact set, the sequence of functions \( \{\psi_k(u)\}_{k \in \mathbb{N}} \) is monotonic (i.e. \( \psi_k(u) \geq \psi_{k+1}(u) \)) and \( \psi_k(u) = E[\Theta_k(u, \xi)] \to 1 - p(u) \) (see Corollary 3.5), the well-known Theorem of Dini guarantees the uniform convergence. 

Proposition 4.8 Suppose \( \{\tau_k\}_{k \in \mathbb{N}} \subset (0, 1) \) be a decreasing sequence with \( \lim_{k \to \infty} \tau_k = 0^+ \). If \( E[\sup_{u \in B(\tau_k, \xi)} \|\nabla \Theta_k(u, \xi)\|] < \infty \) for each \( k \), then, for each \( k \in \mathbb{N} \), the function \( \psi_k(u) = E[\Theta_k(u, \xi)] \) is continuously differentiable and \( \nabla \psi_k(u) = \nabla E[\Theta_k(u, \xi)] = E[\nabla \Theta_k(u, \xi)] \).

Proof The proof of this proposition can be found in the standard literature. 

In order to obtain the required convergence in Equation (21), referring to the differentiability of \( p(u) \) (Proposition 4.6) and the uniform convergence in Lemma 4.7, it suffices to show that the sequence \( \{\nabla \psi_k(u)\}_{k \in \mathbb{N}} \) is uniformly convergent in some bounded neighbourhood \( W \) of \( u_0 \), where \( W \subset \mathcal{U} \). Accordingly, Proposition 4.8 implies that

\[
\nabla \psi(\tau_k, u) = \int_{\Omega} \nabla \Theta(\tau_k, u, \xi) \phi(\xi) \, d\xi, \quad k = 1, 2, \ldots
\]

Note also that

\[
\nabla \Theta(\tau_k, u, \xi) = \left[ \frac{\partial \Theta(\tau_k, u, s)}{\partial m_1} \nabla_{m_1} m_1(u) + \frac{\partial \Theta(\tau_k, u, s)}{\partial m_2} \nabla_{m_2} m_2(u) + \frac{\partial \Theta(\tau_k, u, s)}{\partial s} \nabla u \phi(u, \xi) \right]_{s=g(u, \xi)}.
\]

If \( m_1 \) and \( m_2 \) are chosen as constants, then the first two terms vanish and the gradient \( \nabla \psi(\tau_k, u) \) takes a simpler form.

Theorem 4.9 Let \( W \) be an open bounded subset of \( \mathcal{U} \) and \( \Omega \) be an open set with \( \Omega \subset \hat{\Omega} \). Define \( B(u, \varepsilon) := \{ \xi \in \hat{\Omega} | |g(u, \xi)| < \varepsilon \} \). Then for each \( u \in W \) and each \( \varepsilon > 0 \) it follows that

(i)

\[
\limsup_{k \to \infty} \sup_{u \in W} \left| \int_{\Omega \setminus B(u, \varepsilon)} \left[ \frac{\partial \Theta(\tau_k, u, s)}{\partial m_1} \nabla_{m_1} m_1(u) + \frac{\partial \Theta(\tau_k, u, s)}{\partial m_2} \nabla_{m_2} m_2(u) \right]_{s=g(u, \xi)} \nabla u \phi(u, \xi) \, d\xi \right| = 0;
\]

(ii)

\[
\limsup_{k \to \infty} \sup_{u \in W} \left| \int_{B(u, \varepsilon)} \left[ \frac{\partial \Theta(\tau_k, u, s)}{\partial m_1} \nabla_{m_1} m_1(u) + \frac{\partial \Theta(\tau_k, u, s)}{\partial m_2} \nabla_{m_2} m_2(u) \right]_{s=g(u, \xi)} \nabla u \phi(u, \xi) \, d\xi \right| = 0;
\]

(iii)

\[
\lim_{k \to \infty} \nabla \psi(\tau_k, u) = \lim_{k \to \infty} \int_{B(u, \varepsilon)} \left[ \frac{\partial \Theta(\tau_k, u, s)}{\partial s} \nabla u \phi(u, \xi) \right]_{s=g(u, \xi)} \, d\xi
\]

uniformly on \( W \).
Proof  For a sketch of the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list.

Theorem 4.9 can be interpreted according to the result

$$\lim_{\tau \downarrow 0^+} \Theta(\tau, u, s) = \begin{cases} 1, & \text{if } s \geq 0 \\ 0, & \text{if } s < 0 \end{cases}$$

of Proposition 3.3 and the definition of $\Theta$ in Equation (16). Since the set $\tilde{\Omega} \setminus B(u, \varepsilon)$ corresponds to $s > 0$, for which $\lim_{\tau \downarrow 0^+} \Theta(\tau, u, s) = 1$, Theorem 4.9(i) implies that all partial derivatives

$$\frac{\partial \Theta(\tau_k, u, s)}{\partial m_1}, \quad \frac{\partial \Theta(\tau_k, u, s)}{\partial m_2} \quad \text{and} \quad \frac{\partial \Theta(\tau_k, u, s)}{\partial s}$$

vanish on $\tilde{\Omega} \setminus B(u, \varepsilon)$ for $s = g(u, \xi)$ in the limit $\tau_k \downarrow 0^+$. On the other hand the set $B(u, \varepsilon)$ corresponds to $s < 0$ and at the same time it contains the set $\Gamma_0(u) := \{ \xi \in \Omega \mid g(u, \xi) = 0 \}$. Thus, Theorem 4.9(ii) indicates again that the partial derivatives

$$\frac{\partial \Theta(\tau_k, u, s)}{\partial m_1} \quad \text{and} \quad \frac{\partial \Theta(\tau_k, u, s)}{\partial m_2}$$

vanish on $B(u, \varepsilon)$ in the limit, while the only contribution to the gradient $\lim_{k \to \infty} \nabla \psi(\tau_k, u)$ is coming from the partial derivative

$$\left. \frac{\partial \Theta(\tau_k, u, s)}{\partial s} \right|_{s = g(u, \xi)} \nabla g(u, \xi)$$

as indicated in Theorem 4.9(iii).

**Theorem 4.10**  Under the assumptions of Propositions 4.6 and 4.8 and Theorem 4.9, it follows that

$$\lim_{\tau \to 0^+} \sup_{u \in \tilde{U}} \| \nabla u \psi(\tau, u) + \nabla p(u) \| = 0$$

in a small open neighbourhood $\tilde{U}$ of $U$ and $\nabla p$ is continuous in $u \in \tilde{U}$. Furthermore, for each $u \in \tilde{U}$ and each $\varepsilon > 0$ it follows that

$$\lim_{k \to \infty} \nabla \psi(\tau_k, u) = -\nabla p(u) = \lim_{k \to \infty} \int_{B(u, \varepsilon)} \left[ \left. \frac{\partial \Theta(\tau_k, u, s)}{\partial s} \right|_{s = g(u, \xi)} \nabla g(u, \xi) \right] \phi(\xi) \, d\xi.$$

In summary, for a differentiable probability function $p(u)$, the function $\psi(\tau, u)$ not only approximates the function $(1 - p(u))$, but the gradients $\nabla \psi(\tau_k, u)$ also approximate the gradient $-\nabla p(u)$ for any sequence of parameters $\{\tau_k\}$ with $\tau_k \downarrow 0^+$. This justifies the use of $\nabla \psi(\tau_k, u)$ in optimization algorithms instead of $-\nabla p(u)$, since the latter can be harder to compute.

### 4.3. Convergence properties of the KKT sets

The results of the previous sections are easy to translate to a CCOPT problem with multiple single chance constraints $p_j(u) = \Pr\{g_j(u, \xi) \leq 0\} \geq \alpha_j, j = 1, \ldots, q$. Let $\{\tau_k\}_{k \in \mathbb{N}} \subset (0, 1)$ with $\tau_k \downarrow 0^+$. 
ASSUMPTION 4.11 – MFCQ. At each stationary point of $u \in K$ of CCOPT the Mangasarian–Fromowitz Constraint Qualification holds; i.e. there is a vector $\gamma \in \mathbb{R}^m$ such that
\[ \gamma^T \nabla p_j(u) > 0, \quad j \in \{1, 2, \ldots, q\} \mid p_j(u) = \alpha \].

ASSUMPTION 4.12 – LICQ. At each local optimal solution $u_0 \in K$ of CCOPT the Linear Independence Constraint Qualification holds; i.e. the set of vectors \{\nabla p_j(u_0) \mid j \in A(u_k)\} has $q$ elements that are linearly independent, where $A(u_k) := \{j \in \{1, 2, \ldots, q\} \mid p_j(u_0) = 1 - \alpha\}$.

Define the sets
\[
\Pi := \{(u, \eta) \in U \times \mathbb{R}^q \mid \nabla F(u) - \sum_{j=1}^q \eta_j \nabla p_j(u) = 0; \quad \eta_j p_j(u) = 0, \quad \eta_j \geq 0, \quad j = 1, \ldots, q\},
\]
\[
\Pi_0 := \{(u, \eta) \in \Pi \mid (u, \eta) \text{ is a local optimal pair for CCOPT}\},
\]
\[
\Pi_1 := \{(u, \eta) \in \Pi \mid (u, \eta) \text{ is a strict local optimal pair for CCOPT}\},
\]
\[
\Lambda_k := \{(u, \eta) \in U \times \mathbb{R}^q \mid \nabla F(u) + \sum_{j=1}^q \eta_j \nabla \psi_j(\tau_k, u) = 0, \eta_j \psi_j(\tau_k, u) = 0, \quad \eta_j \geq 0, \quad j = 1, \ldots, q\}, \quad k = 1, 2, 3, \ldots,
\]
\[
\Lambda_k^0 := \{(u, \eta) \in \Lambda_k \mid (u, \eta) \text{ is a local optimal pair for NLP}_k\}.
\]
Since $U$ is a compact set, the sets $\Pi_0$ and $\Lambda_k^0$, for each $k$, are non-empty sets.

THEOREM 4.13 Let $\{\tau_k\}_{k \in \mathbb{N}} \subset (0, 1)$ be any sequence with $\tau_k \downarrow 0^+$ and let Assumptions 4.11 and 4.12 be satisfied. Then the following hold true:

(1) if Assumption 4.11 (MFCQ) holds, then
   (a) $\limsup_{k \to \infty} \Lambda_k \subset \Pi$, and
   (b) $\limsup_{k \to \infty} \Lambda_k^0 \subset \Pi_0$.

(2) if Assumption 4.12 (LICQ) holds, then $\Pi_1 \subset \limsup_{k \to \infty} \Lambda_k^0$.

Proof For the proof, see the online supplement available via the link given in the section ‘Supplemental data’ before the references list.

In contrast to the result given in Theorem 2.9 of Xiao et al. (2012), the result provided in Theorem 4.13 is much stronger. In any case, the MFCQ Assumption 4.11 is indispensable and it also implies Assumption 4.1 ‘Regularity’. From a computational point of view, Theorems 4.4 and 4.13 guarantee that CCOPT problems can be solved approximately by using any state-of-the-art gradient-based optimization solver.

5. Numerical experiments and a case-study
5.1. Computation of probability integrals

The values and gradients of the objective function and constraints are obtained by computing multidimensional probability integrals of the form
\[
E[f(u_k, \xi)] = \int_{\Omega \subset \mathbb{R}^p} \cdots \int_{\Omega \subset \mathbb{R}^p} f(u_k, \xi) \phi(\xi) \, d\xi \quad \text{and} \quad \nabla E[f(u_k, \xi)] = \int_{\Omega \subset \mathbb{R}^p} \cdots \int_{\Omega \subset \mathbb{R}^p} \nabla f(u_k, \xi) \phi(\xi) \, d\xi
\]
at each iterate $u_k$ of the optimization algorithm. The numerical evaluation of these integrals is computationally expensive. Generally, there are two major approaches for the evaluation of
multidimensional integrals: Monte Carlo (MC)/quasi-Monte Carlo(QMC) methods and cubature rules.

Traditionally, multidimensional probability integrals have been computed using MC (see Deák 2011) or QMC methods (Caflisch 1998; Niederreiter 1992; Sloan 1994). The QMC methods rely on the generation of low-discrepancy integration points that are uniformly distributed in a bounded domain of integration. Despite the fact that QMC methods require a very large number of integration nodes, which result in a very large number of function evaluations in higher dimensions (Heiss and Winschel 2006; Gerstner and Griebel 1998), they provide quite accurate approximation, even for non-smooth integrands.

Recently, on the other hand, sparse-grid integration techniques (Smolyak 1963) are being widely used to reduce computational expense in the evaluation of high dimensional integrals. Hence, the work Klöppel et al. (2011) demonstrates the efficiency of sparse grid integration methods for the evaluation of chance constraints (see also Geletu et al. 2011). Integration nodes of the sparse grid integration technique (Bungartz and Griebel 2004; Gerstner and Griebel 1998) are constructed through a skillful combination of one-dimensional quadrature nodes. As a result, sparse grids methods can be constructed so as to compute integrals of polynomial functions exactly and efficiently. However, for general functions, the possible existence of negative weights in the weighted-sum approximation of integrals is a major drawback of sparse grid methods. In fact, negative integration weights are known to cause numerical cancellations and inaccuracies in the approximation. Furthermore, as \( \tau \) decreases, the smoothness of the integrand reduces. Since sparse grid methods depend on the smoothness of the integrand, the quasi-Monte Carlo integration method is found to provide more stable results than the sparse grid method. Hence, in this work,

- integrals are computed through the quasi-Monte Carlo method;
- at each step of the optimization algorithm, a set of multidimensional integrals need to be computed corresponding to the values and gradients of the objective function and constraints; since these integrations are highly parallelizable, code parallelization techniques are used in order to ameliorate computation burdens;
- the experiments reported here show that overall CPU-time decreases linearly with the number of processors used.

In order to demonstrate the viability and efficiency of the proposed approach, two numerical examples are presented in Section 5.2 and a case-study is discussed in Section 5.3. The optimization problems in Section 5.2 are solved by using MATLAB™’s fmincon solver; while, for the case-study in Section 5.3, optimization problems are solved through the interior point optimization solver IpOpt (Wächter and Biegler 2006). IpOpt is capable of working with first-order information and it is highly efficient for the solution of NLPs with smaller values of the parameter \( \tau_k \). Furthermore, the theory developed above holds true when \( m_1 \) and \( m_2 \) are chosen as positive constants and \( m_2 \leq m_1 \). All numerical computations are done on a desktop PC with an Intel 980X hexacore processor with 3.33 GHz per core and 6 GB RAM.

### 5.2. Numerical experiments

#### 5.2.1. Model predictive control of a batch reactor

This section presents some numerical experiments in order to demonstrate the efficiency of the proposed approximation strategy. For this purpose, the chance constrained model predictive control of a batch reactor is studied here (see also Geletu et al. 2012). The dynamic optimization
problem can be formulated as

$$\min f(F_J, x_1, x_2, T, T_M, T_J, \hat{t}, h)$$

subject to

$$x_1(t) = x_1(t-1) - \Delta t(k_1(t)x_1^2(t-1)),$$

$$x_2(t) = x_2(t-1) + \Delta t(k_1(t)x_1(t-1) - k_2(t)x_2(t-1)),$$

$$T(t) = T(t-1) + \Delta t \left[ (\Delta Hr_1k_1(t)x_1(t-1) + \Delta Hr_2k_2(t)x_2(t-1))V_p \right.$$

$$- hA_T[T(t-1) - T_M(t-1)] \left( V_p\rho_p c_p \right)^{-1},$$

$$T_M(t) = T_M(t-1) + \Delta t \left[ hA_T[T(t-1) - 2T_M(t-1) - Z_J(t-1)] \right] \left( V_M\rho_M c_M \right)^{-1},$$

$$\Pr \{T(t) \leq 328 \text{ K}\} \geq 0.8,$$

$$t \in \left\{ \hat{t}, \ldots, \hat{t} + h - 1 \right\},$$

$$x_1(0) = 1, \quad x_2(0) = 0, \quad T(0) = T_M(0) = T_J(0) = 320 \text{ K},$$

where $k_1(t) = k_{10} \exp(-E_1/RT(t-1))$, $k_2(t) = k_{20} \exp(-E_2/RT(t-1))$, $x_1$ and $x_2$ are the concentrations of species A and B, and $T$, $T_M$ and $T_J$ are the temperatures of the reaction mass, wall and jacket, respectively. The quantities $k_1$, $k_2$ describe the reaction rates, $E_1$, $E_2$ the activation energy, $k_{10}$, $k_{20}$ the pre-exponential Arrhenius constants and $F_J$ is the amount of cooling water per hour. The variables $E_1$, $E_2$, $k_{10}$ and $k_{20}$ are assumed to be random with a joint normal distribution (see Table 1). Physical parameters of the batch reactor are shown in Table 2.
The objective function is given by

\[
    f(F_J, x_1, x_2, T, T_M, T_J, \hat{t}, h) = \omega_1 \left[ \sum_{t=\hat{t}}^{\hat{t}+h-1} \left\{ -E(x_2(t)) + \text{Var}(x_2(t)) \right\} \right]
    + \omega_2 \left[ F_J^2 \Delta F_J + \sum_{t=\hat{t}}^{\hat{t}+h-1} F_J^2(t) \right],
\]

where \( \Delta F_J := (F(\hat{t}) - F(\hat{t} - 1), \ldots, F(\hat{t} - h - 1) - F(\hat{t} - h - 2)) \) and \( \omega_1, \omega_2 \in \mathbb{R} \) are weighting factors. The objective is to maximize the expected concentration of product B (i.e. \( E[x_2(t)] \)) while decreasing its variations (\( \text{Var}[x_2(t)] \)) and at the same time to minimize the fluctuation in the cooling water stream. Further details on the solution of the problem can be found in Geletu et al. (2012).

Here, this example is used to study the influence of the number of chance constraints on the computation time. On each prediction horizon, the problem represents a chance constrained nonlinear optimization problem. The number of chance constraints in Equation (27) equals the length of the prediction horizon \( h \). For the sake of comparison, the same problem is solved with the previously proposed back-mapping approach of Wendt, Li, and Wozny (2002) and the analytic approximation strategy, for various lengths of the prediction horizon \( h \). The back-mapping approach uses a sparse grid integration with 441 grid points. The analytical approximation approach uses \( m_1 = 1.1, m_2 = 1, \tau = 0.001 \) and quasi-Monte Carlo integration with 500 grid points. The computation here considers a one hour operation of the process on a total of 30 prediction horizons. Considering the dynamic nature of the problem, computation time per iteration is examined and the result of the experiment is shown in Figure 2.

It is apparent that the back-mapping approach needs significantly more computation time per iteration. Moreover, the computation time per iteration in the analytical approximation approach appears to depend only linearly on the length of the prediction horizon, whereas with the back-mapping approach it appears to increase quadratically with increasing prediction horizon. This
can be explained by the fact that the analytical approximation approach requires the solution of the model equations only once at every grid point of the underlying integration routine of the chance constraints. In contrast, the back-mapping approach requires the solution of \( h \) different modified versions of the original model equations at every grid point, resulting in a quadratic dependence of \( O(h^2) \). In general, if \( N_c \) is the number of chance constraints, \( T_{\text{sol}} \) is the computation time necessary to solve the model equations, and \( N_i \) is the number of grid points in the integration routine, then one would expect a computation time of order \( O(N_c N_i T_{\text{sol}}) \) for the back-mapping approach, respectively. These are rough estimations based on problem data and customized implementations. A theoretical analysis needs to be further studied.

In general, the computational effort increases with decreasing value of \( \tau \).

It should be remarked that analytic approximation methods based on convex polynomial kinship functions (Feng, Dabbene, and Lagoa 2011) would require comparatively fewer integration nodes \( N_i \), since these methods allow the efficient usage of sparse grid integration for the evaluation of chance constraints (see also Klöppel et al. 2011). Hence, computational complexities are of the order of \( O(N_i T_{\text{sol}}) \). This indicates that a kinship-function approach seems to be the best choice when the chance constrained function \( g(x, \xi) \) is convex w.r.t. \( x \) and has a polynomial structure w.r.t. \( \xi \). Even then, the next example indicates that for specific problems the analytical approximation approach results in tighter bounds of the chance constraints and, therefore, improved optimal values.

### 5.2.2. A robust LP

This section studies the robust operation of a plant under uncertainty. The problem formulation as presented in Feng, Dabbene, and Lagoa (2011) is

\[
\begin{align*}
\min_{x \in \mathbb{R}^3} & \quad x_2 \\
\text{subject to} & \quad A(\xi) = \begin{pmatrix} A(\xi) \\ -A(\xi) \end{pmatrix}, \\
& \quad B(\xi) = \begin{pmatrix} d(\xi) - (14 \ 45.25 \ 77 \ 54.25)^T \\ -d(\xi) + (6 \ 31.25 \ 57 \ 38.25)^T \end{pmatrix}, \\
& \quad \Pr \{ A_i x + B_i \leq 0, \ i = 1, \ldots, 8 \} \geq 0.99, \\
& \quad x^T x \leq 1000,
\end{align*}
\]

where \( x \in \mathbb{R}^3 \) are the control variables, \( A_i \) and \( B_i \) are the rows of the matrices \( A(\xi) \) and \( B(\xi) \), respectively; \( \xi \in \{ \xi \in \mathbb{R}^4 | \|\xi_1\| \leq 0.05, \|\xi_2\| \leq 0.05, \|\xi_3\| \leq 0.1, \|\xi_4\| \leq 0.05 \} \) are the uncertain disturbances, which are uniformly distributed, and the remaining quantities are defined as follows:

\[
A(\xi) = \begin{pmatrix} b_0(\xi) & 0 & a_0(\xi) \\ b_1(\xi) & b_0(\xi) & a_1(\xi) \\ b_2(\xi) & b_1(\xi) & a_2(\xi) \\ 0 & b_2(\xi) & 1 \end{pmatrix}, \quad d(\xi) = \begin{pmatrix} 0 \\ a_0(\xi) \\ a_1(\xi) \\ a_2(\xi) \end{pmatrix}
\]

and

\[
a_0(\xi) = -0.116\xi_3 - 4.72\xi_3\xi_4 - 2.36\xi_3^2\xi_4 - 0.472, \\
a_1(\xi) = -1.236\xi_3 - 17.64\xi_3\xi_4 - 10\xi_3^2\xi_4 - 2.236,
\]
\[ a_2(\xi) = 10\xi_3\xi_4 - \xi_3 - 0.764, \]
\[ b_0(\xi) = 2 + 40\xi_1\xi_2 + 20\xi_1^2, \]
\[ b_1(\xi) = 1.5(2 + 40\xi_1\xi_2 + 20\xi_1^2)(1 + \xi_2), \]
\[ b_2(\xi) = 2 + 40\xi_1\xi_2 + 20\xi_1^2. \]

The optimization problem (30)–(34) has a joint chance constraint. In order to apply the proposed analytic approximation method, the relation

\[ \Pr \{ Ax + B_i \leq 0, i = 1, \ldots, 8 \} = \Pr \left\{ \max_{i \in \{1, \ldots, 8\}} (Ax + B_i) \leq 0 \right\} \]
\[ \geq \Pr \left\{ \ln \left( \sum_{i=1}^{8} \exp(A_i x + B_i) \right) \leq 0 \right\} \]

will be helpful. In fact, this follows from the generic inequality that \( \max\{x, y\} \leq \ln(e^x + e^y) \). Consequently,

\[ \Pr \left\{ \ln \left( \sum_{i=1}^{8} \exp(A_i x + B_i) \right) \leq 0 \right\} \geq \alpha \]

implies that

\[ \Pr \{ Ax + B_i \leq 0, i = 1, \ldots, 8 \} \geq \alpha. \]

As a result, the joint chance constraint (33) can be treated through the analytic approximation method. The solution obtained by Feng, Dabbene, and Lagoa (2011), using the kinship-function approach, is \( x = (6.2922 \ 20.4695 \ 9.7037)^T \). Using the analytical approximation of the constraint \( \Pr[\ln(\sum_{i=1}^{8} \exp(A_i x + B_i))] \leq 0 \geq \alpha \) with \( \alpha = 0.99 \), the solution found is \( x = (8.4598 \ 18.1449 \ 9.923)^T \). This indicates a significant decrease in the objective function value in comparison to that given by Feng, Dabbene, and Lagoa (2011). Moreover, a Monte Carlo simulation, using 1,000,000 samples from the uncertainty set, reveals that violations of the constraints \( Ax + B_i \leq 0, i = 1, \ldots, 8 \), actually occur, but \( \Pr\{Ax + B_i \leq 0, i = 1, \ldots, 8\} \) is still greater than the desired 0.99. This indicates that, at least for this example, the analytical approximation method provides a better approximation of the probability values and results in a significantly lower objective function value. A detailed study of the tractable analytic approximation of joint chance constraints remains for future research.

Here, optimization problems are solved by using the MATLAB™’s fmincon solver coupled with a MATLAB™ implementation of a QMC integration routine. The overall required computation time is less than 4.2 minutes. In fact, this CPU-time can be reduced 100- to 500-fold if implementations are done based in C/C++ along with parallelization strategies.

### 5.3. A case-study

The applications of CCOPT in the area of power systems has received increased interest in the last years. The first application was by Ozturk, Mazumdar, and Norman (2004) for the solution of the unit commitment problem (i.e. the problem of the optimal scheduling of the operation of generators in an electrical network). Recent applications include transmission network expansion planning (Yu et al. 2009) and optimal power flow under Gaussian distributed uncertainty (Zhang and Li 2011). In this case-study, a problem comparable with that presented in Zhang and Li (2011) is solved. In contrast to the solution presented there, there is no need here for monotonic relations.
According to Fabbri et al. (2005), the generation of a wind generator can be described by a network. The random input variables are the so-called power flow equations which describe the distribution of power throughout the network. It is assumed there is a base power of 10 MVA and that a total of 15 wind generators (randomly allocated) are installed in the system, where each generator has a maximum generation capacity of \( P_{r,i,\text{max}} = 0.15 \). All units are given in the per unit system, unless otherwise specified. According to Fabbri et al. (2005), the generation of a wind generator can be described by a Beta distribution and the distribution parameters can be calculated from weather forecasts. In the following optimization, results for different weather scenarios are discussed.

The aforementioned operation task can be formulated as a chance constrained optimization problem for one time horizon (e.g. 1 hour).

\[
\min_{\beta} \sum_{i \in I_g} (1 - \beta_i) E[P_{r,i}] \tag{37}
\]

subject to:

\[
P_i - V_i \sum_{j=1}^{41} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) = 0, \quad i \in \{1, \ldots, 41\}, \tag{38}
\]

\[
Q_i - V_i \sum_{j=1}^{41} V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) = 0, \quad i \in \{1, \ldots, 41\}, \tag{39}
\]

\[
P_i = P_S + \beta_i P_{r,i} - P_{d,i}, \quad i \in \{1, \ldots, 41\}, \tag{40}
\]

\[
Q_i = Q_S - Q_{d,i}, \quad i \in \{1, \ldots, 41\}, \tag{41}
\]

\[
\Pr \{0 \leq P_s \leq 1\} \geq \alpha, \tag{42}
\]

\[
\Pr \{-2 \leq Q_s \leq 2\} \geq \alpha, \tag{43}
\]

\[
\Pr \{0.95 \leq V_k \leq 1.05\} \geq \alpha, \quad k \in \{2, \ldots, 41\}, \tag{44}
\]

\[
0 \leq \beta_i \leq 1, \quad \beta_i \in I_g, \tag{45}
\]

where \( \beta_i, i \in I_g \), is a curtailment factor for the active power generation \( P_{r,i} \) of a renewable energy source at bus \( i \), \( I_g = \{17, 18, 19, 22, 26, 27, 28, 33, 34, 36, 37, 39, 40, 41\} \) is the set of indices of buses with installed wind generators, \( V_i \) is the voltage magnitude at bus \( i \), \( \theta_{ij} \) is the difference of the phase angles between buses \( i \) and \( j \), \( i, j \in \{1, \ldots, 41\} \), the matrices \( G \) and \( B \) are the bus conductance and susceptance matrix of the system, \( P_S \) and \( Q_S \) are active and reactive injections at the slack bus (in this study the slack bus is bus 1), \( P_{d,i} \) and \( Q_{d,i} \) are active and reactive loads at bus \( i \) and, \( P_i \) and \( Q_i \) are the active and reactive power injections into bus \( i \). Equations (38)–(41) are the so-called power flow equations which describe the distribution of power throughout the network. The random input variables \( P_{r,i}, i \in I_g \), depend on Beta-distributed random variables; i.e., \( P_{r,i} = P_{r,i,\text{max}} \times \text{Beta}(a_i, b_i), i \in I_g \), where \( \text{Beta}(a, b) \) describes a Beta-distributed uncertain variable with the parameters \( a \) and \( b \). Then the expected values for the generation at bus \( i \) can be calculated by \( E[P_{r,i}] = [a_i/(a_i + b_i)] \times P_{r,i,\text{max}}, i \in I_g \). The curtailment factors \( \beta_i, i \in I_g \), are
control variables, where a value of one means that no curtailment occurs and a value of zero that no power is injected into the network. The chance constraints (42) and (43) guarantee a reliable operation of the network which connects the network to a higher voltage grid, whereas the chance constraints (44) guarantee a certain voltage magnitude at the PQ-buses (buses 2–41).

The optimization task was carried out using IpOpt (Wächter and Biegler 2006) and the chance constraints were evaluated using the method introduced in this article along with quasi-Monte Carlo integration. The two-sided constraints are approximated in the following way:

\[
\Pr\{0.95 \leq V_j \leq 1.05\} = 1 - \Pr\{V_j \leq 0.95\} - \Pr\{1.05 \leq V_j\} \\
\approx 1 - E\left[\frac{1 + 1.002\tau}{1 + \tau \exp\left\{-\frac{1}{\tau}(0.95 - V_j)\right\}}\right] - E\left[\frac{1 + 1.002\tau}{1 + \tau \exp\left\{-\frac{1}{\tau}(V_j - 1.05)\right\}}\right],
\]

where \(0 < \tau < 0.0015, m_1 = 1.002\) and \(m_2 = 1\). Results for different wind scenarios (i.e. different parameters of the underlying Beta distribution) and for different probability levels \(\alpha\) are presented in Table 3. Solutions corresponding to a probability level of \(\alpha = 0.99\) are marked with a subscript H. All other solutions correspond to \(\alpha = 0.95\). The first scenario is a low wind speed scenario; i.e. the expected generation is near the minimum possible value. It can be seen that under such conditions nearly all of the available power can be used. In addition, the solutions for \(\alpha = 0.95\) and \(\alpha = 0.99\) are similar, although, in general, a higher desired reliability leads to increased costs. This can be attributed to the NLP solver which converges toward a local optimum inside the interior of the feasible set. The second scenario is a medium wind speed scenario; i.e. the expected generation is near 50% of the maximum possible generation. In comparison to scenario 1 there is more curtailment and more than 50% of the available wind power is lost. In this case, a higher reliability level leads to increased costs; i.e. more curtailment. The last scenario describes a high wind scenario. Here, only 33% of the available wind energy is utilized for \(\alpha = 0.95\) and only 20% for \(\alpha = 0.99\). In general, the given setting leads to a loss of potentially available energy in
Table 3. Optimization results for different scenarios with $\tau = 0.0009$.

| Solution no. | $v$ (m s$^{-1}$) | Available wind power | Curtailed wind power | $\min_{2 \leq k \leq 41} \Pr[0.95 \leq V_k \leq 1.05]$ | $\Pr[0 \leq P_s \leq 1]$ | $\Pr[|Q_s| \leq 2]$ |
|--------------|------------------|----------------------|---------------------|---------------------------------|----------------------------|-------------------|
| 1            | 5                | 0.22                 | 0.00255             | 0.9965                          | 1                          | 1                 |
| 1$_H$        |                  |                      |                     |                                 |                            |                   |
| 2            | 9                | 1.1                  | 0.663               | 0.95                            | 1                          | 1                 |
| 2$_H$        |                  |                      |                     |                                 |                            |                   |
| 3            | 13               | 1.98                 | 1.540               | 0.95                            | 1                          | 1                 |
| 3$_H$        |                  |                      |                     |                                 |                            |                   |

Figure 4. Objective function values for scenario 2 and different values of $\tau$.

most scenarios due to the constraints on the network. The computation time here is less than 10 minutes.

To give further insight into the proposed method, the problem in scenario 2 was solved for different levels of $\tau$. As can be seen in Figure 4, a decrease of $\tau$ leads to a tighter approximation of the chance constraints and thereby a decrease in the objective function value, since the feasible set becomes larger monotonically with decreasing $\tau$. Furthermore, the problem could also be solved for $\tau < 0.0009$, but this requires a more accurate cubature rule, which in return would increase the computational burden due to a higher number of grid points in the computation of integrals.

6. Conclusions

Optimization under uncertainty becomes a necessity in many engineering applications, such as chemical process design and operation, energy distribution networks, water resources management, and unmanned vehicle control, to name but a few. Chance constrained optimization provides an appropriate way to tackle such stochastic optimization tasks due to the fact that the solution achieves maximum yield while ensuring the desired operational reliability. However, the high complexity of chance constrained optimization problems confronted in engineering applications
(nonlinear, non-convex, high dimensional and with uncertain parameters being non-Gaussian distributed) makes them extremely difficult to solve. The approach developed in this article paves a solid step toward the goal of eventually solving complex engineering optimization problems under uncertainty.

Nonlinear non-convex chance constrained optimization problems can be classified as hard problems in computational optimization. In general, the numerical solution of these problems can only be done through approximation strategies. To date, available analytic approximation strategies either fail to provide tight approximations or lead to non-smooth optimization problems. This work presents a general analytic approximation of nonlinear chance constraints and indicates how to set up conservative but tighter approximations with a priori guaranteed feasibility, where solutions of the approximating problem remain always feasible to the chance constraints, converging to a solution of the chance constrained optimization problem in the limit. The proposed approach can be used irrespective of whether the uncertainties are Gaussian or non-Gaussian continuous density functions. In addition, this approach can address CCOPT problems, in which a monotonic relation between the constrained output and the uncertain input does not exist. Furthermore, a concrete function has been presented which satisfies the desired properties and facilitates the solution of general CCOPT problems.

The case study demonstrates the viability of the proposed approach for practical problems that are large scale with complex model equations.

The consideration here is restricted only to single (separate) chance constraints. An investigation with joint chance constraints remains for future research. Furthermore, in the computations, the choice of the parameter $\tau$ is a compromise between the accuracy of the approximation scheme and the efficiency of numerical integrations. A smaller $\tau$ provides a good approximation of chance constraints, while requiring more integration nodes. On the other hand, a larger $\tau$ leads to smoother integrands at the expense of the tightness of the approximation. In the case studies, convenient values of $\tau$ are selected based on experimentation. Consequently, the choice of an ‘optimal’ $\tau$, balancing tighter analytic approximation and the computational accuracy of integrals, is also an important issue for future research.

Acknowledgements

Many thanks are due to Mr A. Gabash for his support in the case study.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) [Grant No. LI 806/13-1] (all the authors); the Carl-Zeiss Foundation (M. Klöppel).

Supplemental data

Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/0305215X.2014.905550.

Notes

1. An optimal system performance should guarantee at least a 50% ($\alpha = 0.5$) reliability in product quality. On the other hand, guaranteeing a near 100% reliability can deteriorate optimal performance.
2. This can be achieved using the Implicit Function Theorem based on appropriate assumptions. The solution for $x$ from $G(x, u, \xi) = 0$, for a given $u$ and realizations of $\xi$, can be done algorithmically. This effectively eliminates the equality constraint $G(x, u, \xi) = 0$ and the state variables $x$ from CCOPT.
3. The regularity assumption can be guaranteed, for instance, (i) if \( \text{int}K \neq \emptyset \) and \( K = \text{cl}(\text{int}K) \), where cl is closure and int is interior of K, or (ii) if the probability function \( P(u) := \Pr\{g(u, \xi) \leq 0\} \) is differentiable – see Uryasev (1995) and Marti (1996) – and for each \( u \in \mathcal{U} \) there is a (Mangasarian–Fromowitz) vector \( \eta_{u} \) such that \( \eta_{u}^{T} \nabla P(u) > 0 \).

4. It should be remarked (without proof) that, under suitable regularity assumptions, this limit set consists of the so-called strongly stationary points.

References

Andrieu, L., R. Henrion, and W. Römisch. 2010. “A Model for Dynamic Chance Constraints in Hydropower Reservoir Management.” *European Journal of Operations Research* 207 (2): 579–589.

Arellano-Garcia, H., and G. Wozny. 2009. “Chance Constrained Optimization of Process Systems Under Uncertainty: I. Strict Monotonicity.” *Computers and Chemical Engineering* 33 (10): 1568–1583.

Ben-Tal, A., and A. Nemirovski. 1998. “Robust Convex Optimization.” *Mathematics of Operations Research* 23 (4): 769–805.

Ben-Tal, A., and A. Nemirovski. 1999. “Robust Solutions of Uncertain Linear Programs.” *Operations Research Letters* 25 (1): 1–13.

Bertsimas, D., B. B. K. David, and C. Caramanis. 2011. “Theory and Applications of Robust Optimization.” *SIAM Review* 53 (3): 464–501.

Birge, John R., and François Louveaux. 1997. *Introduction to Stochastic Programming*. New-York: Springer-Verlag.

Blackmore, L., and M. Ono. 2009. “Convex Chance Constrained Predictive Control Without Sampling.” In *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Chicago, IL, August 10–13. doi: 10.2514/6.2009-5876

Blackmore, L., M. Ono, A. Bektassov, and B. C. Williams. 2010. “A Probabilistic Particle-Control Approximation of Chance-Constrained Predictive Control.” *IEEE Transactions on Robotics* 26 (3): 502–517.

Bungartz, H.-J., and M. Griebel. 2004. “Sparse Grids.” *Acta Numerica* 13: 147–269.

Caflisch, Russel E. 1998. “Monte Carlo and Quasi-Monte Carlo Methods.” *Acta Numerica* 7: 1–49.

Califore, G., and M. C. Campi. 2005. “Uncertain Convex Programs: Randomized Solutions and Confidence Levels.” *Mathematical Programming* 102 (1): 25–46.

Califore, G., and M. C. Campi. 2010. “Random Convex Programs.” *SIAM Journal on Optimization* 20 (6): 3427–3464.

Campi, M.C., and S. Garetti. 2011. “Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality.” *Journal of Optimization Theory and Applications* 148 (2): 257–280.

Cannon, M., B. Kouvaritakis, and X. Wu. 2009. “Model Predictive Control for Systems with Stochastic Multiplicative Uncertainty and Probabilistic Constraints.” *Automatica* 45 (1): 167–172.

Celik, A. N., A. Makkawi, and T. Muneer. 2010. “Critical Evaluation of Wind Speed Frequency Distribution Functions.” *Renewable Energy* 35 (3): 365–370.

Charnes, A., and W. W. Cooper. 1958. “Cost Horizons and Certainty Equivalence: An Approach to Stochastic Programming Heating Oil.” *Management Science* 6 (1): 73–79.

Charnes, A., W. W. Cooper, and G. H. Symmonds. 1958. “Cost Horizons and Certainty Equivalence: An Approach to Stochastic Programming.” *Management Science* 4 (3): 235–263.

Deák, I. 2011. “Efficiency in Monte Carlo Computation in Very High Dimensional Spaces.” *Central European Journal of Operations Research* 19 (2): 177–189.

Dhar, A., and B. Datta. 2006. “Chance Constrained Water Quality Management Model for Reservoir Systems.” *ISH Journal of Hydraulic Engineering* 12 (3): 39–48.

Fabbri, A., T. G. San Román, J. R. Abbad, and V. H. M. Quezada. 2005. “Assessment of the Cost Associated with Wind Generation Prediction Errors in a Liberalized Electricity Market.” *IEEE Transactions on Power Systems* 20 (3): 1440–1446.

Feng, C., F. Dabbene, and C. M. Lagoa. 2011. “A Kinship Function Approach for Robust and Probabilistic Optimization Under Polynomial Uncertainty.” *IEEE Transactions on Automatic Control* 56 (7): 1509–1523.

Flemming, T., M. Bartl, and P. Li. 2007. “Set-Point Optimization for Closed-Loop Control Systems Under Uncertainty.” *Automatica* 43 (1): 167–172.

Flemming, T., M. Bartl, and P. Li. 2007. “Set-Point Optimization for Closed-Loop Control Systems Under Uncertainty.” *Automatica* 43 (1): 167–172.

Gabash, A., and P. Li. 2012. “Active–Reactive Optimal Power Flow in Distribution Networks with Embedded Generation and Battery Storage.” *IEEE Transactions on Power Systems* 27 (4): 2026–2035.

Garnier, J., A. Omrane, and Y. Rouchdy. 2009. “Asymptotic Formulas for the Derivatives of Probability Functions and Their Monte Carlo Estimations.” *European Journal of Operations Research* 198 (3): 848–858.

Geletu, A., A. Hoffmann, M. Klöppel, and P. Li. 2011. “Monotony Analysis and Sparse-Grid Integration for Nonlinear Chance Constrained Process Optimization.” *Engineering Optimization* 43 (10): 1019–1041.

Geletu, A., M. Klöppel, H. Zhang, and P. Li. 2012. “Advances and Applications of Chance-Constrained Approaches to Systems Optimization Under Uncertainty.” *International Journal of Systems Science* 44 (7): 1209–1232.

Gerstner, T., and M. Griebel. 1998. “Numerical Integration on Sparse Grids.” *Numerical Algorithms* 18 (3–4): 209–232.

Heiss, Florian, and Viktor Winschel. 2006. “Estimation with Numerical Integration on Sparse Grids.” *Münchner Wirtschaftswissenschaftliche Beiträge (VWL)* 2006-15. http://epub.ub.uni-muenchen.de/916/1/HeissWinschel2006.pdf

Henrion, R. 2012. “Gradient Estimates for Gaussian Distribution Functions: Applications to Probabilistically Constrained Optimization Problems.” *Numerical Algebra, Control and Optimization* 2 (4): 655–668.

Henrion, R., C. Küchler, and W. Römisch. 2009. “Scenario Reduction in Stochastic Programming With Respect to Discrepancy Distances.” *Computational Optimization and Applications* 43 (1): 67–93.
van Ackooij, W., R. Henrion, A. Möller, and R. Zorgati. 2010. “On Probabilistic Constraints Induced by Rectangular Sets and Multivariate Normal Distributions.” *Mathematical Methods of Operations Research* 71 (3): 535–549.

Yu, H., C. Y. Chung, K. P. Wong, and J. H. Zhang. 2009. “A Chance Constrained Transmission Network Expansion Planning Method with Consideration of Load and Wind Farm Uncertainties.” *IEEE Transactions on Power Systems* 24 (3): 1568–1576.

Wächter, A., and L. T. Biegler. 2006. “On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming.” *Mathematical Programming* 106 (1): 25–57.

Wang, W., and S. Ahmed. 2008. “Sample Average Approximation of Expected Value Constrained Stochastic Programs.” *Operations Research Letters* 36 (5): 515–519.

Weber, G.-W., Z. Çavuşoğlu, and A. Özmen. 2012. “Predicting Default Probabilities in Emerging markets by New Conic Generalized Partial Linear Models and Their Optimization.” *Optimization* 61 (4): 443–457.

Wendt, M., P. Li, and G. Wozny. 2002. “Nonlinear Chance-Constrained Process Optimization Under Uncertainty.” *Industrial & Engineering Chemistry Research* 41 (15): 3621–3629.

Werner, R. 2008. “Cascading: An Adjusted Exchange Method for Robust Conic Programming.” *Central European Journal of Operations Research* 16 (2): 179–189.

Wozabal, D. 2012. “Value-At-Risk Optimization Using the Difference of Convex Algorithm.” *OR Spectrum* 34 (4): 861–883.

Wozabal, D., R. Hochreiter, and G. Ch. Pflug. 2010. “A Difference of Convex Formulation of Value-At-Risk Constrained Optimization.” *Mathematical Methods of Operations Research* 59 (3): 377–400.

Xiao, Xiantao, Jian Gu, Liwei Zhang, and Shaowu Zhang. 2012. “A Sequential Convex Program Method to DC Program with Joint Chance Constraints.” *Journal of Industrial and Management Optimization* 8 (4): 733–747.

Zhang, H., and P. Li. 2010. “Probabilistic Analysis for Optimal Power Flow Under Uncertainty.” *IET Generation Transmission and Distribution* 4 (5): 553–561.

Zhang, H., and P. Li. 2011. “Chance Constrained Programming for Optimal Power Flow Under Uncertainty.” *IEEE Transactions on Power Systems* 26 (4): 2417–2424.

Zymler, S., D. Kuhn, and B. Rustem. 2013. “Distributionally Robust Joint Chance Constraints with Second-Order Moment Information.” *Mathematical Programming* 137 (1-2): 167–198.