Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Probabilistic constraints via SQP solver: Application to a renewable energy management problem

Ingo Bremer[1], René Henrion[1], Andris Möller

submitted: September 11, 2014

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
E-Mail: ingo.bremer@wias-berlin.de
rene.henrion@wias-berlin.de

No. 2010
Berlin 2014

2010 *Mathematics Subject Classification.* Primary: 90C15; Secondary: 90B05.

Key words and phrases. probabilistic constraints, renewable energies, multivariate Gaussian probability, SQP with low precision data.

This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. The aim of this paper is to illustrate the efficient solution of nonlinear optimization problems with joint probabilistic constraints by means of an SQP method. Here, the random vector is assumed to obey some multivariate Gaussian distribution. The numerical solution approach is applied to a renewable energy management problem. We consider a coupled system of hydro and wind power production used in order to satisfy some local demand of energy and to sell/buy excessive or missing energy on a day-ahead and intraday market, respectively. A short term planning horizon of 2 days is considered and only wind power is assumed to be random. In the first part of the paper, we develop an appropriate optimization problem involving a probabilistic constraint reflecting demand satisfaction. Major attention will be payed to formulate this probabilistic constraint not directly in terms of random wind energy produced but rather in terms of random wind speed, in order to benefit from a large data base for identifying an appropriate distribution of the random parameter. The second part presents some details on integrating Genz' code for Gaussian probabilities of rectangles into the environment of the SQP solver SNOPT. The procedure is validated by means of a simplified optimization problem which by its convex structure allows to estimate the gap between the numerical and theoretical optimal values, respectively. In the last part, numerical results are presented and discussed for the original (non-convex) optimization problem.

1. INTRODUCTION

A probabilistic constraint is an inequality of the type

$$\mathbb{P}\left(g(x,\xi) \leq 0\right) \geq p, \tag{1}$$

where g is a mapping defining a (random) inequality system and ξ is an s- dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The constraint (1) expresses the fact that a decision vector x is feasible if and only if the random inequality system $g(x,\xi) \leq 0$ is satisfied at least with probability $p \in [0,1]$. Probabilistic constraints are important for engineering problems involving uncertain data. Applications can be found in water management, telecommunications, electricity network expansion, mineral blending, chemical engineering etc. For a comprehensive overview on the theory, numerics and applications of probabilistic constraints, we refer to, e.g., [23],[24], [26]. The analysis of probabilistic constraints has attracted much attention in recent years with a focus on algorithmic approaches. Without providing an exhaustive list, we refer here to models like robust optimization [8], penalty approach [12], p-level efficient points [11], scenario approximation [9], sample average approximation [21] or convex approximation [19].

In this paper, we want to pursue the classical idea that (1) is a nonlinear inequality constraint in the decision vector x and may be treated as such in the framework of nonlinear programming algorithms. Many of the recently proposed numerical approaches to probabilistic programming are based on a possibly large number of scenarios sampled according to the given distribution of ξ. These approaches are universal in the sense that they just require the possibility to draw samples which is no problem for most of the prominent multivariate distributions. On the other hand, the required sample size for guaranteeing a fairly good precision of optimal values and solutions to a problem under probabilistic constraints may become excessive with increasing dimension of the random vector. An alternative would consist in taking advantage of specific information about certain types of continuous distribution. This may yield a possibility not only to approximate values but also gradients with respect to x of (1). For instance, in the special case of separable constraints $g(x,\xi) = \xi - x$, and of ξ having a regular Gaussian distribution, one may employ an efficient code by Genz [13, 14] which is based on a numerical integration scheme combining separation and reordering of variables with randomized QMC. A similar technique has been proposed for the multivariate Student (or T-) distribution [14]. The numerical evaluation of other multivariate distribution functions such as Gamma or exponential distribution has been discussed, e.g., in [20, 27].

In particular the multivariate Gaussian distribution makes it possible to calculate in reasonable time at fairly high precision values of (1) in the general linear model $g(x,\xi) = A(x)\xi - b(x)$ and moreover analytically to reduce gradients of these probabilities to probability values of the same type as before (but with different parameters, [1, 2, 17]). A perspective for continuing this approach towards general nonlinear models and non-Gaussian (but Gaussian-like, such as log-normal or Student) multivariate distributions is offered by spherical-radial decomposition of Gaussian random vectors [4, 25]. So far, the nonlinear programming approach to

probabilistic programming has been primarily applied to settings yielding convex problems. This is the case, for instance, if the mapping g in (1) is linear and the distribution of ξ is log-concave (e.g. Gaussian, Student, uniform etc.). Among the applications considered, hydro reservoir problems played a central role [22, 3, 1, 2, 6]. The supporting hyper-plane method was the preferred one for the numerical solution of these problems. This method is quite robust, easy to implement and provides upper and lower bounds for the optimal value of the problem. On the other hand, it becomes less efficient with increasing dimension and it does not apply to non-convex problems. For this reason, we decided to embed the nonlinear programming approach into an SQP solver (SNOPT) and to check its efficiency for an example of renewable energies. The main challenge in this embedding consists in handling numerical imprecisions of function value and gradient evaluation for the probabilistic constraint.

The application we consider here is a coupled system of hydro and wind energy production which is supposed to meet a local demand of energy and to sell energy on a day ahead market. Unbalanced demand satisfaction is regulated on an intraday market (buy/sale in case of falling short of or exceeding demand). A short term future time horizon of 48 hours is considered and production of wind energy is the only process considered to be stochastic. A major step in modeling this problem will consist in providing a statistical model for wind energy production given a large historical data basis for wind speed observations. The paper is divided into three major sections: in the first one an appropriate optimization model involving a joint probabilistic constraint will be derived for the mentioned application. The next section is devoted to details of implementing probabilistic constraints in an SQP solver. Numerical tests including relative gaps for the optimal value and computing times will be reported for a simplified (convex) example. The final section reports and discusses the numerical results for the general (nonlinear, non-convex) optimization problem.

2. A MODEL FOR SHORT TERM PLANNING OF A COUPLED SYSTEM OF HYDRO/WIND ENERGY PRODUCTION

2.1. **A deterministic model.** We consider the following power management problem with renewable energy production units: a hydro plant is coupled with a wind farm in order to meet some local power demand and to sell any surplus electricity on the market. All time-dependent quantities of the problem are discretized by subdividing the time horizon into a specified number of T intervals (e.g., hours). Accordingly, given the time interval $[t-1, t)$, we denote by h_t the decision variable of hydro energy to be produced in this interval, by d_t the demand of electricity, by ξ_t the amount of electricity generated by the wind farm, by π_t the price for selling/buying one unit of electricity on a power market and by w_t the inflow of water into the hydro reservoir. Let us assume in the beginning that all data of the problem - i.e., d_t, ξ_t, w_t, π_t - are exactly known. Then, one may sell or buy energy on a day ahead market in order to balance the demand satisfaction. The total profit made over the whole time horizon equals

$$\sum\nolimits_{t=1}^{T} \pi_t \left(h_t + \xi_t - d_t\right), \tag{2}$$

where with respect to demand satisfaction surplus energy $[h_t + \xi_t - d_t]_+$ is sold and missing energy $[h_t + \xi_t - d_t]_-$ is bought at price π_t. Turbining of water in a hydro reservoir is subject to certain constraints: first, there are simple operational bounds:

$$0 \le h_t \le M \quad (t = 1, \ldots, T). \tag{3}$$

Second, the water level in the hydro reservoir has to stay between zero and some maximum capacity (these constraints make sure that water can be turbined only if present in the reservoir and stored only if there is enough capacity left). In practice, hard lower and upper level constraints are imposed which are more stringent for ecological (e.g., flood reserve), technological or economical reasons. One arrives at relations

$$l^{\min} \le l_0 + \sum\nolimits_{\tau=1}^{t} \left(w_\tau - \varkappa^{-1} h_\tau\right) \le l^{\max} \quad (t = 1, \ldots, T). \tag{4}$$

Here, l_0 denotes the initial filling level of the reservoir in the beginning of the time horizon and $\varkappa$ is a conversion constant between water turbined and hydro energy produced. Accordingly, the current filling level at time t equals the initial level plus the cumulative amount $\sum_{\tau=1}^{t} w_\tau$ of water inflow until time t minus the cumulative

amount $\varkappa^{-1}\sum_{\tau=1}^{t} h_\tau$ of water released from the reservoir and transformed into hydro energy $\sum_{\tau=1}^{t} h_\tau$ until time t.

Finally, in order to exclude production strategies which are optimal for the given time horizon but come at the expense of future ones (e.g., maximum turbining), a so-called end level constraint is imposed for the final water level in the hydro reservoir. In the simplest case, the end level l^* could be chosen equal to the initial level l_0 but one might also consider increasing it by a certain amount, for instance, if the initial level is low. We do not insist here on a more sophisticated water level evaluation strategy as described, for instance, in [3] because this issue is less relevant for the mathematical aspects discussed in this paper. Summarizing, we are led to the end-level constraint

$$l_0 + \sum_{\tau=1}^{T} \left(w_\tau - \varkappa^{-1} h_\tau\right) \geq l^*. \tag{5}$$

Observing that the only decision variable in the objective (2) is given by h_t, the resulting optimization problem becomes a conventional linear program which numerically to solve does not represent any challenge:

$$\max \sum_{t=1}^{T} \pi_t h_t \quad \text{subject to} \quad (3), (4), (5). \tag{6}$$

2.2. **A model with random wind energy and joint probabilistic constraint.** In reality, all data of (6) - prices, demand, inflow to reservoir and wind energy - are random with a degree of uncertainty increasing over time. In what follows, we shall consider a short time horizon of 2 days. In that case, prices, demand and water inflow may be assumed to be known with sufficient precision, so that we will consider them as deterministic data. Figure 1 shows some typical profiles of the price signal and the demand of electricity for a period of two days. For the inflow we will even assume that it comes at constant speed so that $w_t = w$ for all t. In contrast, the

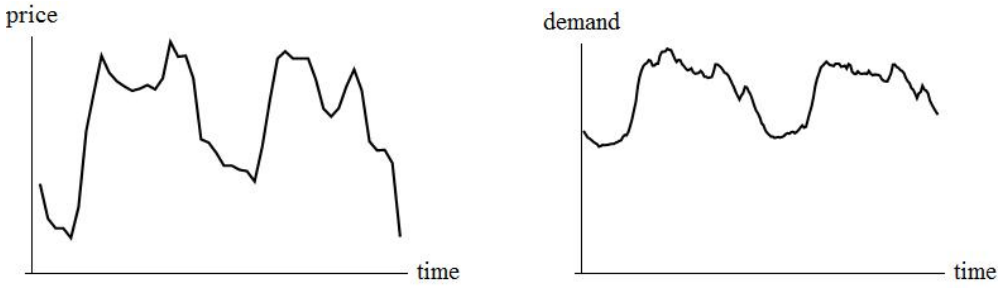

FIGURE 1. 2-days profiles for price signal (left) and demand (right) of power

more volatile wind energy production will be treated as a random vector $\xi = (\xi_1, \ldots, \xi_T)$. Since the (future) realizations of this random vector are not known at the time one has to decide on the amount of energy traded on a day ahead market, one has to take account of surplus or missing energy in the more flexible intraday market which allows for short term (e.g., 15 minutes ahead) contracts. These may come, of course, at quite different prices $\tilde{\pi}_t$ when compared with day ahead prices π_t. In order to distinguish between both kinds of transactions, we split the amount of hydro energy produced into two parts

$$h_t = x_t + y_t \quad (t = 1, \ldots, T), \tag{7}$$

where $x_t \geq 0$ denotes the part which is offered on the day ahead market and $y_t \geq 0$ refers to the remaining part used for demand satisfaction (together with random wind energy). Now, the overall profit will be

$$\sum_{t=1}^{T} \pi_t x_t + \sum_{t=1}^{T} \tilde{\pi}_t \left(y_t + \xi_t - d_t\right). \tag{8}$$

This profit is a random quantity not only due to the uncertain wind energy but also due to the presence of highly uncertain intraday prices $\tilde{\pi}_t$. Our aim consists in risk-averse maximization of the random profit. For instance, we could maximize some profit which can be guaranteed with some probability $p \in [0, 1]$ given the joint distribution of the random vector $(\xi, \tilde{\pi})$ which would correspond to value-at-risk maximization. However, information about the distribution of $\tilde{\pi}$ may be very difficult to obtain because intraday prices are strongly

influenced by rare events like outages of certain production units in the market pool. In contrast, modeling the distribution of wind energy production has a much better chance due to abundant historical data on wind speed, as will be described in Section 2.3. This leads us to the following partitioned risk-averse objective:

$$\max\left\{\eta|\mathbb{P}\left(\omega|\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)\geq\eta\quad\forall\tilde{\pi}\right)\geq p\right\}\quad(p\in[0,1]) \tag{9}$$

Its meaning is the following: we maximize some profit η which can be guaranteed with at least probability p with respect to uncertain wind energy production and **for all** possible intraday prices. This mixed probabilistic-worst case objective will be reformulated in the following section by using a joint probabilistic constraint.

Lemma 1. *Let $p>0$ and assume that intraday prices are non-negative ($\tilde{\pi}\geq 0$). Then, for arbitrarily fixed decisions x_t, y_t and level η in (9) one has the following equivalence:*

$$\mathbb{P}\left(\omega|\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)\geq\eta\quad\forall\tilde{\pi}\geq 0\right)\geq p \tag{10}$$

if and only if

$$\mathbb{P}(\omega|y_t+\xi_t\left(\omega\right)\geq d_t\quad\forall t=1,\ldots,T)\geq p,\quad\sum\nolimits_{t=1}^{T}\pi_t x_t\geq\eta \tag{11}$$

Proof. Evidently, by $\tilde{\pi}_t\geq 0$ for all t, (11) implies (10). Conversely, let (10) hold true. Assume that $\sum_{t=1}^{T}\pi_t x_t<\eta$. Then, we arrive at the following contradiction with (10):

$$\mathbb{P}\left(\omega|\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)\geq\eta\quad\forall\tilde{\pi}\geq 0\right)\leq\mathbb{P}\left(\omega|\sum\nolimits_{t=1}^{T}\pi_t x_t\geq\eta\right)=0<p.$$

This shows the second relation of (11). Next, let ω be such that

$$\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)\geq\eta\quad\forall\tilde{\pi}\geq 0. \tag{12}$$

Assume that $y_\tau+\xi_\tau\left(\omega\right)<d_\tau$ for some $\tau\in\{1,\ldots,T\}$. Then, defining $\tilde{\pi}$ by $\tilde{\pi}_t:=0$ if $t\in\{1,\ldots,T\}\setminus\{\tau\}$ and

$$\tilde{\pi}_\tau:=\frac{\eta-\sum_{t=1}^{T}\pi_t x_t}{y_\tau+\xi_\tau\left(\omega\right)-d_\tau}+1,$$

we observe that $\tilde{\pi}_\tau>0$ because of the already shown relation $\eta-\sum_{t=1}^{T}\pi_t x_t\leq 0$. Hence, $\tilde{\pi}\geq 0$. This establishes the contradiction

$$\begin{aligned}\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)&=\sum\nolimits_{t=1}^{T}\pi_t x_t+\tilde{\pi}_\tau\left(y_\tau+\xi_\tau\left(\omega\right)-d_\tau\right)\\&<\sum\nolimits_{t=1}^{T}\pi_t x_t+\eta-\sum\nolimits_{t=1}^{T}\pi_t x_t=\eta\end{aligned}$$

with (12). Consequently, any ω satisfying (12), satisfies $y_t+\xi_t\left(\omega\right)\geq d_t$ for all $t\in\{1,\ldots,T\}$ as well. Therefore, we have shown the first relation of (11):

$$\begin{aligned}\mathbb{P}(\omega|y_t+\xi_t\left(\omega\right)\geq d_t\quad\forall t=1,\ldots,T)&\geq\\\mathbb{P}\left(\omega|\sum\nolimits_{t=1}^{T}\pi_t x_t+\sum\nolimits_{t=1}^{T}\tilde{\pi}_t\left(y_t+\xi_t\left(\omega\right)-d_t\right)\geq\eta\quad\forall\tilde{\pi}\geq 0\right)&\geq p.\end{aligned}$$

■

In the light of Lemma 1, we may replace the maximization of the objective (9) by a maximization of η subject to the constraints (11). However, maximizing η subject to the second constraint of (11) is equivalent with simply

maximizing the day ahead profit $\sum_{t=1}^{T} \pi_t x_t$. Therefore, the resulting optimization problem reads as follows:

$$\max \sum_{t=1}^{T} \pi_t x_t \tag{13}$$

$$\text{subject to}$$

$$\mathbb{P}(\omega | y_t + \xi_t(\omega) \geq d_t \quad \forall t = 1, \ldots, T) \geq p \tag{14}$$

$$x_t \geq 0, \quad y_t \geq 0, \quad x_t + y_t \leq M \quad (t = 1, \ldots, T) \tag{15}$$

$$l^{\min} \leq l_0 + tw - \varkappa^{-1} \sum_{\tau=1}^{t} (x_\tau + y_\tau) \leq l^{\max} \quad (t = 1, \ldots, T) \tag{16}$$

$$l_0 + Tw - \varkappa^{-1} \sum_{\tau=1}^{T} (x_\tau + y_\tau) \geq l^* \tag{17}$$

Here, (14) follows from the first constraint in (11), (15) is a consequence of (3) and (7) and (16) and (17) result from (4) and (5) upon recalling that we assume $w_t = w$ for all t. Observe that, by Lemma 1 an optimal solution $(\bar{x}, \bar{y})$ satisfies the property that the random profit

$$\sum_{t=1}^{T} \pi_t \bar{x}_t + \sum_{t=1}^{T} \tilde{\pi}_t (\bar{y}_t + \xi_t(\omega) - d_t)$$

made on the day-ahead and intraday market will exceed the deterministic quantity $\sum_{t=1}^{T} \pi_t \bar{x}_t$ at least with probability p no matter what the prices $\tilde{\pi}_t$ on the intraday market will be. Moreover, $\sum_{t=1}^{T} \pi_t \bar{x}_t$ is the largest possible quantity with this property.

From the mathematical viewpoint, the most interesting ingredient of this optimization problem is the inequality (14) which acts as a constraint on the decision vector y. One refers to this as a *joint probabilistic* (or chance) *constraint* where the attribute 'joint' reminds of the fact that under the decision y the whole random inequality system

$$y_t + \xi_t(\omega) \geq d_t \quad (t = 1, \ldots, T)$$

has to be satisfied with probability at least p. The probabilistic constraint (14) is of static type, i.e., it is assumed that the whole vector y is decided on without reacting on possible past observations of the random process ξ_t. This approach is justified in our setting, because the sale x of hydro energy on the day-ahead market has to be decided on one day before the contribution of wind energy will be observed. But deciding on x amounts to deciding on y at the same time due to the constraint (15). In general, one might also consider dynamic (closed loop) decisions reacting on past realizations of the random variable leading to more complex dynamic models of probabilistic constraints (as discussed, e.g., in [6]).

2.3. **Modeling the distribution of wind energy.** In order to cope with the probabilistic constraint (14) one has to characterize the distribution of the random vector ξ representing the wind energy produced. Our approach consists in exploiting distribution information about hourly mean wind speed data and to transfer this information to the produced wind energy. More precisely, we assume that produced wind energy ξ and wind speed v are related by

$$\xi = \min\{cv^3, a\} \tag{18}$$

for certain coefficients $a, c > 0$. This relation reflects the fact that wind energy produced is roughly proportional to the third power of wind speed before reaching a maximum amount a which cannot be exceeded by further increasing wind speed. Figure 2 (left) shows that relation (18) provides a very good approximation to real data. For details on related issues, we refer, for instance, to [28]. A concrete wind speed scenario and associated power scenario with maximum production a being realized in the second half of the day is illustrated in Figure 2 (right).

As far as wind speed is concerned, we want to apply a statistical model which allows us to eventually establish a multivariate distribution for the discrete-time random vector which moreover is accessible to a numerical treatment of the probabilistic constraint (14). This seems to be difficult when assuming a usual Weibull distribution for wind speed data. Therefore, we shall follow here an approach presented in [5] where wind speed data v_t are raised to a certain power v_t^θ as to make them normal-like (with $\theta \approx 0.38$ for the data considered by the authors) and these transformed data are modeled as a stationary autoregressive process of order one

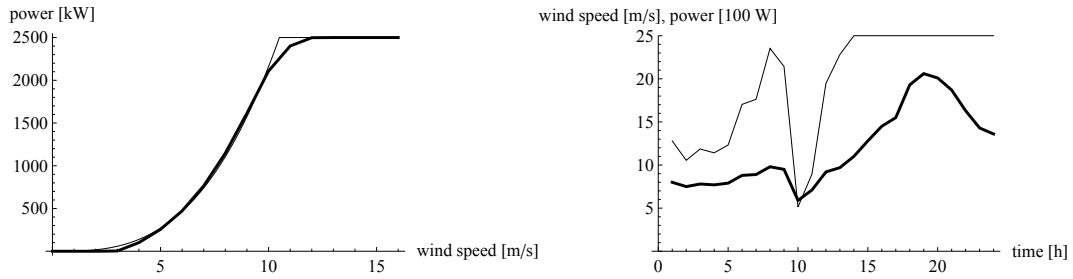

FIGURE 2. Left: Plot of generated power vs. wind speed for a real wind plant (thick curve) and approximation by (18) (thin curve). Right: historical scenario for hourly wind speed of one day (thick curve) and associated scenario of (ideal) power generation (thin curve).

with normally distributed innovations. The statistics of this autoregressive process (mean, standard deviation and correlation) may be used in order to derive some (truncated) multivariate Gaussian distribution for the transformed wind speed data. More precisely, putting

$$\eta_t := v_t^\theta, \tag{19}$$

one has that $(\eta_t)_{t\in\mathbb{Z}}$ is a truncation to nonnegative values of some Gaussian process $(\tilde{\eta})_{t\in\mathbb{Z}}$ obeying the relations

$$\tilde{\eta}_t = (1-\rho)\,\mu + \rho\tilde{\eta}_{t-1} + \sigma\varepsilon_t \quad \forall t \in \mathbb{Z}, \tag{20}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ and all ε_t are uncorrelated with all ε_τ ($\tau \in \mathbb{Z}, \tau \neq t$) and all $\tilde{\eta}_\tau$ ($\tau \in \mathbb{Z}, \tau < t$). Moreover, $\tilde{\eta}_t \sim \mathcal{N}(\mu, \sigma^2/(1-\rho^2))$ for all $t \in \mathbb{Z}$, and $cov(\tilde{\eta}_t, \tilde{\eta}_\tau) = \rho^{|t-\tau|}\sigma^2/(1-\rho^2)$ for all $t, \tau \in \mathbb{Z}$. Collecting this information for times $t \in \{1,\dots,T\}$, $\tilde{\eta}$ is a T-dimensional Gaussian vector with multivariate distribution

$$\tilde{\eta} \sim \mathcal{N}(\mu\mathbf{1}, \Sigma), \quad \Sigma_{ij} := \rho^{|i-j|}\sigma^2/\left(1-\rho^2\right) \quad (i,j \in \{1,\dots,T\}), \tag{21}$$

where $\mathbf{1} = (1,\dots,1)$. Accordingly, η is the truncated to $\mathbb{R}_+^T$ version of $\tilde{\eta}$ which means that the distribution of η is given by

$$\mathbb{P}(\eta \in A) = \frac{\mathbb{P}\left(\tilde{\eta} \in A \cap \mathbb{R}_+^T\right)}{\mathbb{P}(\tilde{\eta} \in \mathbb{R}_+^T)} \tag{22}$$

for all Borel measurable sets $A \subseteq \mathbb{R}_+^T$.

2.4. **A statistical model for the multivariate distribution of hourly wind speed data.** In order to determine the distribution parameters mentioned in the previous section, we used data from the data basis of the German Weather Service (Quelle: Deutscher Wetterdienst) for the station Kap Arkona (north east Germany). In order to disregard seasonal variations of wind speed data, we restricted the data base to hourly data for all days in month October from 1992 to 2013.

In a first step, we want to determine the appropriate exponent for our data in (19). To this aim, we fit an optimal normal distribution to the transformed wind speed data v_t^θ. The fit is carried out according to a minimum Kolmogorov distance between the empirical distribution of the transformed data and the fitted normal distribution. The use of the Kolmogorov distance is motivated by stability results for chance constrained optimization problems with separated random vector, a class which problem (13)-(17) belongs to. These confirm that the perturbation of solutions and optimal values to such problems can be controlled (in a Lipschitz or Hölder way) by the perturbation of the underlying distribution when measured by the Kolmogorov distance. Recall that the Kolmogorov distance between two univariate cumulative distribution functions F and G is defined as

$$d_K(F,G) := \sup_{z\in\mathbb{R}} |F(z) - G(z)|.$$

In the particular case of F being continuous and of $G(z) = N^{-1}\#\{i|x^{(i)} \leq z\}$ being the empirical distribution function of some random variable γ associated with a sample $x^{(1)}, \ldots, x^{(N)}$ of an i.i.d. sequence $\gamma_1, \ldots, \gamma_N$ of random variables having the same distribution as γ, the supremum in the definition of the Kolmogorov distance can be reduced to a finite maximum:

$$d_K(F,G) = \max_{i=1,\ldots,N} \max\{F(\tilde{x}^{(i)}) - N^{-1}(i-1), N^{-1}i - F(\tilde{x}^{(i)})\}. \tag{23}$$

Here, $\tilde{x}^{(1)} \leq \cdots \leq \tilde{x}^{(N)}$ is an ordered version of the sample $x^{(1)}, \ldots, x^{(N)}$. In order to get an i.i.d. sample for the wind speed, we used the total of 15.835 hourly October data from the data basis mentioned above and extracted a subsequence of values at distance of 50 hours in order to comply with the required independence of sampling. Hence, instead of using the total sample $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$, we considered a subsample $x^{(1)}, x^{(51)}, x^{(101)}, \ldots$ and fitted to its empirical distribution function G an optimal normal distribution $F^{\mu,\sigma}$ with mean value μ and standard deviation σ applied to an optimal power θ of the argument. Given (23), the best fitting parameters μ, σ, θ can be found by solving the following nonlinear optimization problem

$$\min\{d \,|\, F^{\mu,\sigma}(\left[\tilde{x}^{(i)}\right]^{\theta}) - m^{-1}(i-1) \leq d,\ m^{-1}i - F^{\mu,\sigma}(\left[\tilde{x}^{(i)}\right]^{\theta}) \leq d \quad i = 1, \ldots, m\}$$

in 4 variables (μ, σ, θ, d), where $\tilde{x}^{(1)} \leq \cdots \leq \tilde{x}^{(m)}$ is the ordered version of the subsample $x^{(1)}, x^{(51)}, x^{(101)}, \ldots$ of size $m = 316$. The optimal solution $\left(\bar{\mu}, \bar{\sigma}, \bar{\theta}, \bar{d}\right)$ of this problem identifies the best fitting normal distribution with parameters $\bar{\mu}, \bar{\sigma}$, the best exponent $\bar{\theta}$ at which the data have to be raised as well as the resulting minimum Kolmogorov distance $\bar{d}$ with the given empirical distribution. Figure 3 (left) shows the best fitting normal

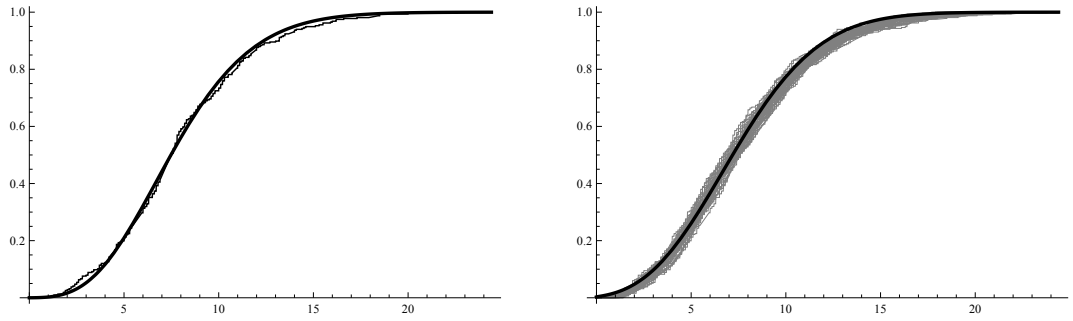

FIGURE 3. Left: Distribution function of the best fitting normal distribution applied to an optimal power transformation of arguments (thick curve) for a given empirical distribution function (thin curve) of a subsample of size 316. Right: Normal distribution function (thick curve) simultaneously best fitting to a set of 50 empirical distribution functions (cloud of thin curves) drawn from the total data basis.

distribution function with parameters $\bar{\mu} = 2.82, \bar{\sigma} = 0.65$ and applied to an optimal power $\bar{\theta} = 0.52$ of arguments as well as the given empirical distribution function. Of course one could also drive other independent subsamples of independent wind speed data, such as $x^{(2)}, x^{(52)}, x^{(102)}, \ldots$ or $x^{(3)}, x^{(53)}, x^{(103)}, \ldots$ yielding different optimal fitting parameters. In order to stabilize the estimation of these parameters on the basis of all given data, we looked for the best fitting normal distribution and optimal power transformation simultaneously with respect to all possible 50 subsamples. Figure 3 (left) shows the best fitting normal distribution function with parameters $\bar{\mu} = 4.23, \bar{\sigma} = 1.54$ and applied to an optimal power $\bar{\theta} = 0.73$ of arguments as well as the cloud of the total of 50 given empirical distribution function.

In the second step, we model the transformed wind speed v_t^{θ} with the determined optimal power $\theta = 0.73$ as a stationary autoregressive process of order one with normally distributed innovations according to (20) whose parameter values are estimated using the entire data basis of 15.835 October data for hourly wind speed. The correlation coefficient was found to be $\rho = 0.96$. Along with the previously determined values $\mu = 4.23, \sigma = 1.54$ for the stationary mean and standard deviation, respectively, this allows us to set up a multivariate Gaussian distribution according to (21) approximating the distribution of transformed wind

speed data v_t^θ for $t = 1, \ldots, T$. Since v_t^θ is nonnegative, we pass from the distribution (21) to a truncated multivariate Gaussian distribution according to (22).

2.5. **Reformulation of the probabilistic constraint.** The next Lemma allows us to transform the probabilistic constraint (14) involving the produced wind energy ξ with unknown distribution into a probabilistic constraint involving the multivariate distribution identified in the previous section:

Lemma 2. *Under the assumption that $y_t \leq d_t$ for $t = 1, \ldots, T$, and that $\theta > 0$ in (19), the probabilistic constraint (14) is equivalent with the constraints*

$$\mathbb{P}\left(\tilde{\eta}_t \geq \left(\frac{d_t - y_t}{c}\right)^{\theta/3} \quad (t = 1, \ldots, T)\right) \geq p\tilde{p} \quad \textit{and} \quad a \geq d_t - y_t \quad (t = 1, \ldots, T)\,, \tag{24}$$

where $\tilde{\eta}$ is the Gaussian random vector introduced in (21) and

$$\tilde{p} := \mathbb{P}\left(\tilde{\eta}_t \geq 0 \quad (t = 1, \ldots, T)\right). \tag{25}$$

Proof. Exploiting our assumption and using (18) and (19), we derive that for $t = 1, \ldots, T$,

$$\begin{aligned} \xi_t \geq d_t - y_t \quad &\Leftrightarrow \quad \left(\min\{cv_t^3, a\}\right)^\theta \geq (d_t - y_t)^\theta \\ &\Leftrightarrow \quad c^\theta \eta_t^3 \geq (d_t - y_t)^\theta \quad \text{and} \quad a \geq d_t - y_t \\ &\Leftrightarrow \quad \eta_t \geq \left(\frac{d_t - y_t}{c}\right)^{\theta/3} \quad \text{and} \quad a \geq d_t - y_t. \end{aligned}$$

Hence,

$$\begin{gathered} \mathbb{P}\left(\xi_t \geq d_t - y_t \quad (t = 1, \ldots, T)\right) \geq p \Leftrightarrow \\ \mathbb{P}\left(\eta_t \geq \left(\tfrac{d_t - y_t}{c}\right)^{\theta/3} \quad \text{and} \quad a \geq d_t - y_t \quad (t = 1, \ldots, T)\right) \geq p \Leftrightarrow \\ \mathbb{P}\left(\eta_t \geq \left(\tfrac{d_t - y_t}{c}\right)^{\theta/3} \quad (t = 1, \ldots, T)\right) \geq p \quad \text{and} \quad a \geq d_t - y_t \quad (t = 1, \ldots, T)\,. \end{gathered}$$

The last equivalence follows because we required $p > 0$ from the beginning (otherwise a probabilistic constraint is always automatically satisfied) and, hence, deterministic constraints can be isolated from probabilistic ones. Finally, for nonnegative arguments α_t the relation between the truncated (to non-negative values) Gaussian random vector η and its associated un-truncated Gaussian random vector $\tilde{\eta}$ is (see (22))

$$\mathbb{P}\left(\eta_t \geq \alpha_t \quad (t = 1, \ldots, T)\right) = \mathbb{P}\left(\tilde{\eta}_t \geq \alpha_t \quad (t = 1, \ldots, T)\right) / \mathbb{P}\left(\tilde{\eta}_t \geq 0 \quad (t = 1, \ldots, T)\right).$$

This entails the assertion of the Lemma. ■

We claim next that the assumption $y_t \leq d_t$ for $t = 1, \ldots, T$ of Lemma 2 is always satisfied for a solution of problem (13)-(17) under the reasonable assumption that all day-ahead prices π_t are strictly positive. Therefore, adding the relations $y_t \leq d_t$ as additional constraints to the problem would not change its solution. On the other hand, under these additional constraints, we may take for granted the conclusion of Lemma 2.

Lemma 3. *Let (x^*, y^*) be a solution to problem (13)-(17). Assume that $\pi_t > 0$ for all $t = 1, \ldots, T$. Then, $y_t^* \leq d_t$ for $t = 1, \ldots, T$.*

Proof. Assume to the contrary that $y_{t^*}^* > d_{t^*}$ for some $t^* \in \{1, \ldots, T\}$. Define a vector $(\hat{x}, \hat{y})$ by

$$\hat{x}_t := \begin{cases} x_t^* & \text{if } t \in \{1, \ldots, T\} \setminus \{t^*\} \\ x_{t^*}^* + y_{t^*}^* - d_{t^*} & \text{if } t = t^* \end{cases} \quad ; \quad \hat{y}_t := \begin{cases} y_t^* & \text{if } t \in \{1, \ldots, T\} \setminus \{t^*\} \\ d_{t^*} & \text{if } t = t^* \end{cases}.$$

Clearly, $\hat{x}, \hat{y} \geq 0$ because $x^*, y^* \geq 0$ (as a solution to problem (13)-(17)) and $d \geq 0$ (as a demand profile). From $\hat{x}_t + \hat{y}_t = x_t^* + y_t^*$ for all $t \in \{1, \ldots, T\}$ it follows that $(\hat{x}, \hat{y})$ satisfies all linear constraints (15)-(17)

because (x^*, y^*) does so. But the probabilistic constraint (14) is fulfilled too due to

$$\begin{aligned}\mathbb{P}\left(\hat{y}_t + \xi_t \geq d_t \quad (t = 1, \ldots, T)\right) &= \mathbb{P}\left(y_t^* + \xi_t \geq d_t \quad t \in \{1, \ldots, T\} \setminus \{t^*\} \quad \text{and} \quad \xi_{t^*} \geq 0\right) \\ &= \mathbb{P}\left(y_t^* + \xi_t \geq d_t \quad t \in \{1, \ldots, T\} \setminus \{t^*\}\right) \\ &\geq \mathbb{P}\left(y_t^* + \xi_t \geq d_t \quad (t = 1, \ldots, T)\right) \geq p.\end{aligned}$$

Here, the second equality follows from the fact that ξ_{t^*} as the produced wind energy is non-negative $\mathbb{P}-$almost surely and the last inequality relies on (x^*, y^*) satisfying the probabilistic constraint (14). Consequently, $(\hat{x}, \hat{y})$ is a feasible solution to problem (13)-(17). On the other hand,

$$\sum\nolimits_{t=1}^{T} \pi_t \hat{x}_t - \sum\nolimits_{t=1}^{T} \pi_t x_t^* = \pi_{t^*}\left(y_{t^*}^* - d_{t^*}\right) > 0.$$

This, however, means that the feasible solution $(\hat{x}, \hat{y})$ realizes a strictly larger objective value than the optimal solution (x^*, y^*) to problem(13)-(17), a contradiction proving our assertion. ■

2.6. **The resulting optimization problem.** We are now in a position to formulate the final optimization problem with identified distribution of the random vector. Lemma 2 and Lemma 3 allow us to replace the probabilistic constraint (14) by the relations (24) involving a Gaussian random vector and several linear deterministic constraints. Referring back to (13)-(17), we arrive at:

(26) $$\max \sum\nolimits_{t=1}^{T} \pi_t x_t$$

subject to

(27) $$\mathbb{P}\left(\tilde{\eta}_t \geq \left(\frac{d_t - y_t}{c}\right)^{\theta/3} \quad (t = 1, \ldots, T)\right) \geq p\tilde{p}$$

(28) $$0 \leq d_t - y_t \leq a \quad (t = 1, \ldots, T)$$

(29) $$x_t, y_t \geq 0; \quad x_t + y_t \leq M \quad (t = 1, \ldots, T)$$

(30) $$l^{\min} \leq l_0 + tw - \varkappa^{-1} \sum\nolimits_{\tau=1}^{t} (x_\tau + y_\tau) \leq l^{\max} \quad (t = 1, \ldots, T)$$

(31) $$l_0 + Tw - \varkappa^{-1} \sum\nolimits_{\tau=1}^{T} (x_\tau + y_\tau) \geq l^*.$$

3. Numerical solution via SQP method

As mentioned in the introduction, the traditional approach to probabilistic constraints in the context of nonlinear programming consisted in the verification of convexity and the application of first-order methods from convex numerical optimization such as supporting hyperplane, central cutting plane or reduced gradient methods. Two major reasons suggest rather to deal with probabilistic constraints in an SQP environment: first, the afore mentioned method converge rather slow with increasing dimension (both of the random and of the decision vector). Here, we have in mind problems where the dimension of the random vector amounts to a few hundred (talking, of course, about its joint distribution) while there is no a priori restriction to the dimension of the decision vector. Second, SQP methods provide the right framework also for potentially non-convex problems. Both aspects will be significant for the application discussed in this paper.

Formally, there is no problem to integrate a probabilistic constraint like (1) into the environment of an SQP solver: all one has to be able to provide is routines to compute values and gradients of the function

$$\varphi(x) := \mathbb{P}\left(g(x, \xi) \leq 0\right)$$

assigning to each decision vector the probability of satisfying the inequality system $g(x, \xi) \leq 0$. This being granted, one may treat (1) as a conventional constraint $\varphi(x) \geq p$ of nonlinear programming. The main challenge, however, arises from the fact that no explicit formula for evaluating φ or $\nabla\varphi$ is available and that

all one can hope for is a numerical approximation of these quantities whose accuracy is never comparable with that in case of analytic expressions. On the other hand, as far as the solution of an optimization problem is concerned, one usually does not insist on highly precise values, so one might have the idea of running an SQP code with less precise data but which are sufficient to provide solutions of reasonable accuracy. Our working horse for calculating φ and $\nabla\varphi$ will be Genz' code for probabilities of rectangles under multivariate Gaussian distribution [13, 14]. This code assigns to each multidimensional rectangle $[a, b]$ (with possible components $-\infty$ or ∞) the probability $\alpha_\xi(a, b) := \mathbb{P}(\xi \in [a, b])$, where ξ is a Gaussian random vector. In the context of problem (26)-(31), the probabilistic constraint (27) can be written as $\varphi(y) := \alpha_{\tilde{\eta}}(H(y), \widetilde{\infty})$, where $\widetilde{\infty} := (\infty, \ldots, \infty)$ and

$$H_t(y) := \left(\frac{d_t - y_t}{c}\right)^{\theta/3} \quad (t = 1, \ldots, T)\,.$$

Now, as H is given analytically, the evaluation of φ basically reduces to that of $\alpha_{\tilde{\eta}}$ and similarly, the evaluation of $\nabla\varphi$ basically reduces to that of $\nabla\alpha_{\tilde{\eta}}$. Although, similar to $\alpha_{\tilde{\eta}}$ itself, there is no explicit formula for $\nabla\alpha_{\tilde{\eta}}$, a fortunate consequence of the properties of Gaussian distributions leads to the possibility of analytically (!) expressing the partial derivatives $(\partial\alpha_{\tilde{\eta}}/\partial x_t)$ in terms of the values $\alpha_{\eta'}$ of some other Gaussian rectangle probability (see, e.g., [23, p. 204]). A remarkable consequence of this fact is that gradients (and even higher order derivatives by further differentiation) can be calculated by means of the same method as function values themselves. This underlines the importance of an efficient code for computing multivariate Gaussian distribution functions. Usually, this task is considered to be extremely time consuming in larger dimension, but it turns out that with a parallelized version of Genz' code - which will be discussed more in detail in the next section - one may obtain a distribution function value in dimension 100 at relative precision of 10^{-4} within a few seconds.

The preceding discussion motivates us to integrate Genz' code in an SQP code environment, taking care of the reduced precision in order to keep the code running smoothly. For our application, we have chosen the SQP solver SNOPT [15].

3.1. **Genz' code for Gaussian probabilities.** As mentioned in the previous section, Genz' code provides the probability $\alpha_\xi(a, b)$ of some s- dimensional Gaussian random vector ξ taking values in a rectangle $[a, b]$. Assume without loss of generality that ξ has mean zero and some covariance matrix Σ. The basic idea behind Genz' code is to transform the original multiple integral associated with this probability into the following iterated integral on the standard unit cube:

(32)
$$\alpha_\xi(a, b) = (e_1 - d_1)\int_0^1 (e_2(z_1) - d_2(z_1))\int_0^1 \cdots \int_0^1 (e_s\,(z_1, \ldots, z_{s-1}) - d_s\,(z_1, \ldots, z_{s-1}))\,dw,$$

where

(33)
$$d_i(\text{or } e_i)\,(z_1, \ldots, z_{i-1}) := \Phi\left(l_{ii}^{-1}\left(a_i(\text{or } b_i) - \sum_{j<i} l_{ij}\Phi^{-1}\,(z_j)\right)\right),$$

$L = (l_{ij})$ is the Cholesky factor of $\Sigma = LL^T$ and Φ refers to the one-dimensional standard Gaussian distribution function. Then, the integral (32) can be numerically approximated by generating random samples (preferably via randomized Quasi Monte Carlo) of the uniform distribution on $[0, 1]^s$ and passing them through the iterated integral which turns into a finite sum. The final result is obtained upon dividing by the sample size. In the practical application of the code, a certain accuracy for the desired probability is chosen by the user and the set of samples is increased step-wise (each step yielding an increase of accuracy by a factor of approximately 0.8) thereby averaging the final result over all steps. The final result is guaranteed to satisfy the chosen accuracy at 99% confidence. As observed in [14], the order of integration variables influences significantly the variance of the obtained estimator. For this purpose, a cheap preprocessing step taking into account the structure of the Cholesky factor L is carried out in the beginning.

The main computational effort in the approximation of (32) is spent by the frequent evaluation of Φ and Φ^{-1} in (33). As this computation is the same for each sample, it can be perfectly parallelized. Parallelization options for Genz' code have already been presented in [10]. In our implementation, we make some block-oriented rearrangements by means of a parallelizing compiler with OpenMP-support. On current hardware (Intel(R) Xeon(R) CPU E5-2680) using all 32 processors is a good choice with a good load-balancing.

3.2. **Embedding of Genz' code into SQP solver SNOPT.** As mentioned above, we chose the SNOPT code as an SQP environment for the solution of optimization problems involving joint nonlinear probabilistic constraints. In order to employ this code, the user has to provide routines for function values and gradients of the objective and constraints. Gradients are checked by an internal control and can be replaced by finite difference approximations. As described above, in our problem the evaluation of the probabilistic constraint can be basically reduced to Genz' code for calculating Gaussian probabilities of rectangles. As far as gradients of such (parameter dependent) probabilities are concerned, it has been shown in [1] that the partial derivatives can be reduced analytically to Gaussian probabilities of rectangles again albeit in different dimension and with different distribution parameters. This allows to employ the same code by Genz' to obtain gradients at the same time. Moreover, as shown in [16], the accuracy of this gradient can be explicitly controlled by that of function values. Inductive application of this reduction would allow, in principle, to use the given code in order to calculate derivatives of any order if desired. A generalization of this possibility from rectangle to arbitrary polyhedra is presented in [17].

Although Genz' method described above is able to provide fairly good approximations, the precision of the obtained values for the probability function φ is too low for a straight forward embedding into the SNOPT code. In particular, the internal gradient check via finite differences typically fails. Moreover, the accuracy levels for primal and dual feasibility imposed as stopping criteria in SNOPT cannot be chosen the same as for usual problems with high accuracy numerical function evaluations but have to be harmonized iteratively with the application of Genz' code. Therefore, we have taken several measures in order to stabilize the line search step and convergence of the iterates. As far as the line search is concerned, it is mainly affected by noisy function values mainly having the following three sources: first, when using a fixed precision for Genz' code, jumps in function values close to a given argument may occur due to a sudden increase of the sample size in the stepwise procedure mentioned in Section 3.1; second, the use of variable seeds for the random number generator will contribute to noise as well; third, a change of the order of integration variables (see Section 3.1) will lead to additional discontinuities. While this third source cannot and should not be eliminated due to its significance for computational efficiency, we were using a fixed seed for the random number generator and did not prescribe the accuracy for Genz' code but rather fix a sample size on the discrete scale available in Genz' code. This is what we will call the chosen accuracy level (see, e.g., Table 1) in what follows. It turns out that in contrast with considering this fixed level directly, the computation time can be significantly reduced when starting SNOPT with the smallest sample size (accuracy level 0) and then - using the obtained solution as a new starting point - stepwise increasing it until the chosen one has been reached. We could observe in numerical experiments a reduction of a factor up to around 5 depending on the dimension of the problem and on the chosen accuracy level. Running the SQP code with imprecise function values and gradients requires of course to adapt the appropriate stopping criteria for primal and dual feasibility in order to guarantee the termination of the algorithm for the corresponding accuracy level. As far as primal feasibility is concerned, which basically means satisfying the probabilistic constraint, we measure it by adding the achieved accuracy in Genz' code to the obtained probability. In this way it is guaranteed the criterion for primal feasibility being satisfied at 99% confidence. The dual feasibility criterion is set (and possibly modified) at the lowest accuracy level 0 in some heuristic way and then, passing to the next higher accuracy level, adapted by exploiting gradient information of the probabilistic constraint at the solution of the lower level. Here by gradient we mean the analytical reduction of the theoretical gradient to function values and its approximation by the latter ones mentioned in the beginning of this section. While this 'analytical' gradient is useful for adapting the dual feasibility criterion, it is less appropriate for defining the direction of line search or for updating the Jacobian in the SQP code. Rather, we employ here a handmade substitute of automatic differentiation for the code. This

turns out to fit better to the function values used in the line search step and to be less time consuming than the analytic counterpart in the Jacobian update.

3.3. **Validation of computing times and precision for a simplified convex model.** In this section we will illustrate the performance of the numerical solution approach by means of a simplified model. More precisely, we consider the optimization problem (13)-(17) under the assumption of wind energy ξ having some multivariate Gaussian distribution. Of course, this assumption is unrealistic and we will devote the final section of this paper to the solution of the optimization problem (26)-(31), which takes into account a statistically founded model on the basis of real life wind speed data. The simplified optimization problem, however has the advantage of being a (nonlinear) convex one which is more appropriate for discussing numerical issues of the solution of joint chance constraints. In particular, it allows us to validate solutions in terms of relative gaps for optimal values by (less efficient) algorithms from convex optimization such as the supporting hyperplane method. In our opinion, besides computing times, an estimation of the relative gap between the theoretical and the numerically obtained optimal value is essential in order to validate a specific method (as done, e.g., in [18]).

For modeling the wind energy ξ via a multivariate Gaussian distribution, we used appropriate adaptations of the parameters (mean values, standard deviations, correlations) obtained in the more realistic nonlinear model. The concrete values of these parameters is of not much interest here because we will not discuss the solutions themselves but rather check the performance of the algorithm. In this context, the advantage of the simplified model (13)-(17) is the following: the linear structure of the inequalities $y_t + \xi_t \geq d_t$ in the probabilistic constraint (14) along with the (log-concave) Gaussian distribution of the random vector ξ ensure that the feasible set (in the x, y space) defined by (14) is convex [23, Th. 10.2.1]. This allows us to derive upper and lower bounds for the optimal value by employing, for instance, the supporting hyperplane method. While this method suffers from slow convergence in larger dimension of the random vector, it provides us some upper bound $\hat{f}$ of the maximization problem (13)-(17). In order to make this upper bound as small as possible, we had this method run for a long time within these computational experiments. On the other hand, our numerical approach via SQP method always yields in the end some significantly feasible solution, by which we mean that this solution satisfies the probabilistic constraint at a significance level of 99%. Consequently, the optimal value f^{num} associated with this numerical solution is a lower bound for problem (13)-(17). Denoting by f^* the true optimal value, we consider the relative gap

$$\Delta := \frac{|f^{num} - f^*|}{f^*}$$

to be a measure of quality for the obtained numerical solution. Since f^* is unknown, we cannot directly determine Δ. However, taking into account that $\hat{f} \geq f^* \geq f^{num}$, we are able to calculate an upper estimate

$$\Delta \leq \frac{\hat{f} - f^{num}}{f^{num}}. \tag{34}$$

TABLE 1. Upper estimates for the relative gaps of optimal values in different dimensions of the random vector and for different accuracy levels

accuracy level	dim = 48	dim = 96	dim = 192
0	0.07% | 117s	0.32% | 679s	1.53% | 5077s
1	0.09% | 218s	0.32% | 2031s	1.36% | 4884s
3	0.08% | 545s	0.32% | 2600s	1.34% | 13100s
7	0.04% | 1716s	0.30% | 24260s	1.26% | 69346s
11	0.03% | 4278s	0.27% | 31090s	-
15	0.03% | 21087s	0.27% | 114815s	-

Table 1 compiles computing times and upper estimates for the relative gap Δ according to (34) for different dimensions of the random vector and different accuracy levels in Genz' code for computing values of the multivariate Gaussian distribution function. The dimensions correspond to different discretizations of our 2 days planning horizon into hours, half hours and quarter of hours as they will be used in the solution of the realistic model discussed in the following section. Not surprisingly, computing times increase with dimension and accuracy level, whereas the obtained upper estimates for Δ decrease with the accuracy level and increase with dimension. One may observe that very good precisions for the optimal value can be obtained in dimension 48 even for the lowest accuracy level resulting in a computation time of 2 minutes. It seems that significantly higher accuracy is bought by a considerably larger computing time and that the obtained upper estimates for Δ tend to a certain limit that cannot be improved. This is certainly due to the unavoidable gap $\hat{f} - f^*$ between the upper bound obtained by the supporting hyperplane method and the true optimal value. Hence, the true relative gaps Δ might improve much faster in reality than the values collected in the table.

4. Results for the problem with nonlinear probabilistic constraint

In this section we present the computational results for the solution of problem (26)-(31). Starting with an hourly discretization of a two-days horizon ($T = 48$) (which is extended later to an half hourly and quarter of an hour discretization), we used the data specified in Table 2 (in appropriate units and with MWh as basic unit for energy). The distribution data for the random vector $\tilde{\eta}$ were chosen according to (21) with parameters μ, ρ, σ as indicated above and corresponding to those determined in Section 2.4. The probability $\tilde{p}$ was computed according to (25). The data were designed in a way that hydro production alone cannot completely meet the demand, hence, wind power has to be added. As the latter is random, one cannot expect almost sure demand satisfaction ($p = 1$). It turns out that the maximum probability to meet the demand amounts to approximately $p = 0.85$ in this example. We illustrate several aspects of the obtained solutions for probability levels ranging from 0.1 to 0.85.

Table 2. Input data for problem (26)-(31).

Problem data

$$
\begin{aligned}
\pi &= (25.12, 15.59, 12.87, 12.86, 10.09, 18.9, 40.48, 51.36, 61.91, 57.13, 54.66, 53.2, 51.94,\\
&\quad 52.56, 53.59, 52.35, 55.51, 65.88, 61.65, 61.92, 55.45, 38.15, 37.21, 34.3, 30.63, 30.69,\\
&\quad 29.49, 29.1, 26.18, 36.07, 49.39, 61.15, 63.19, 61.13, 61.17, 61.12, 55.15, 46.87, 44.62,\\
&\quad 47.02, 53.25, 58.03, 52.08, 37.57, 34.99, 35.14, 31.38, 10.79)\\
d &= (8.25, 7.86, 7.65, 7.73, 7.85, 8.18, 9.12, 11.37, 12.61, 12.57, 13.06, 13.40, 12.88, 12.51,\\
&\quad 12.19, 12.04, 11.84, 12.04, 12.39, 11.88, 11.02, 10.78, 10.81, 9.69, 8.72, 8.21, 8.16,\\
&\quad 8.34, 8.46, 8.74, 9.41, 11.37, 12.52, 12.50, 12.53, 12.43, 12.13, 12.02, 12.02, 11.91,\\
&\quad 11.65, 11.95, 12.36, 11.85, 11.08, 10.86, 10.78, 9.85)\\
c &= 0.032; \quad a = 40; \quad w = 6 \cdot 10^5; \quad \varkappa = 1.8 \cdot 10^{-5}; \quad M = 16.2;\\
l^{\min} &= 2.4 \cdot 10^6; \quad l^{\max} = 4.8 \cdot 10^6; \quad l_0 = 3.2 \cdot 10^6; \quad l^* = 3.6 \cdot 10^6;\\
\mu &= 4.23; \quad \sigma = 1.54; \quad \rho = 0.96; \quad \theta = 0.73;
\end{aligned}
$$

Figure 4 illustrates the results of computations. The first diagram (first row, left) plots the given price signal π_t (thin curve) and the optimal profile x_t (thick curve) of hydro energy sold on the day ahead market for the probability level $p = 0.4$. It can be seen that the sale of hydro energy tries to follow as much as possible (under the given additional constraints) the price profile in order to maximize the profit on the day ahead market. The next diagram (first row, right) provides an analogous plot of the given demand signal d_t (thin curve) and the optimal profile y_t (thick curve) of hydro energy used for demand satisfaction for the same probability level $p = 0.4$. Evidently, the degree of freedom in the choice of decisions satisfying a probabilistic constraint at a given level is used according to the objective function: for an optimal splitting of the total amount of hydroenergy produced, the support of demand is maximum at time periods when intraday prices are low and minimum when those prices are high (see left diagram). The diagram also shows that the inequality $y_t \leq d_t$ is satisfied at any time (see (28)). The next diagram (second row, left) shows the filling level profiles resulting for probability levels $p = 0.3$ (black thick line) and $p = 0.7$ (black thin line). In both cases, the lower and upper filling limits $l^{\min}, l^{\max}$ (grey lines) are respected (see (30)) and also the specified end level l^* is realized (see (31)). While for the lower probability level, the limits of the reservoir are reached several times, the filling level stays strictly between these limits for the high probability requirement. The neighbouring diagram (second row, right), provides the corresponding plot of total hydro energy production $x_t + y_t$ for the same two probability levels. Again the upper production limit (thin line) is respected (see (29)), but the variation of the production profile is much lower in the case of high probability. The dependence of solutions on the probability level p is demonstrated in the next row of diagrams. These show the surfaces of hydroenergy x_t sold on the day ahead market (left) and hydroenergy y_t used for demand satisfaction (right). Both surfaces are interpolations based on computations for probability levels $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85$ (black curves on the surfaces). It can be clearly seen how the contribution to demand satisfaction increases while that to sale on the day ahead market decreases with increasing probability level. In particular, at probability level $p = 0.85$ no more sale takes place and the total amount hydro power production is used for demand satisfaction. Therefore, the right most profile on the right-hand side surface (at $p = 0.85$) can be interpreted as a solution which guarantees no losses (day-ahead+intraday) at a probability of 85%. Here we have to keep in mind, that arbitrary intraday prices are allowed - even in the sense of an adversary arranging a worst case situation -, hence the realistic situation is a much more optimistic one.

The first diagram in the fourth row superposes for $p = 0.5$ the optimal profiles x_t of hyrdo energy sold on the day ahead market for finer discretizations of the 2 day horizon: hourly discretization ($T = 48$, black thick line), half hourly discretization ($T = 96$, grey thick line) and discretization in quarters of hours ($T = 192$, black thin line). Similarly, the second diagram in the fourth row superposes the optimal profiles y_t of hydro energy used for demand satisfaction at the same levels of discretization as in the left diagram. On the one hand, the purpose of these computations was to demonstrate that our approach works well even for the nonlinear model in rather large dimension 192 of the random vector.

Second, the obtained solutions immediately lead to an interesting mathematical questions about their limit: the decrease of sale profiles and increase of the contribution for demand satisfaction with finer discretization is not surprising, because fulfilling the probabilistic constraint in each quarter of an hour is much harder than doing so in a cumulative way in each hour. Intuitively, one might guess that - in the limit - the discretized solutions would converge to the solutions of a continuous version of problem (26)-(31). Here, continuous means that not only all data profiles (demand, prices) and all control variables (profiles for hydro energy) are supposed to be elements of an appropriate function space but also the random process of wind speed (wind energy, respectively) is no longer discrete but continuous. This supposed convergence is somehow supported by those last two diagrams, where the transition from hours to half hours is much larger than that from half hours to quarters of hours.

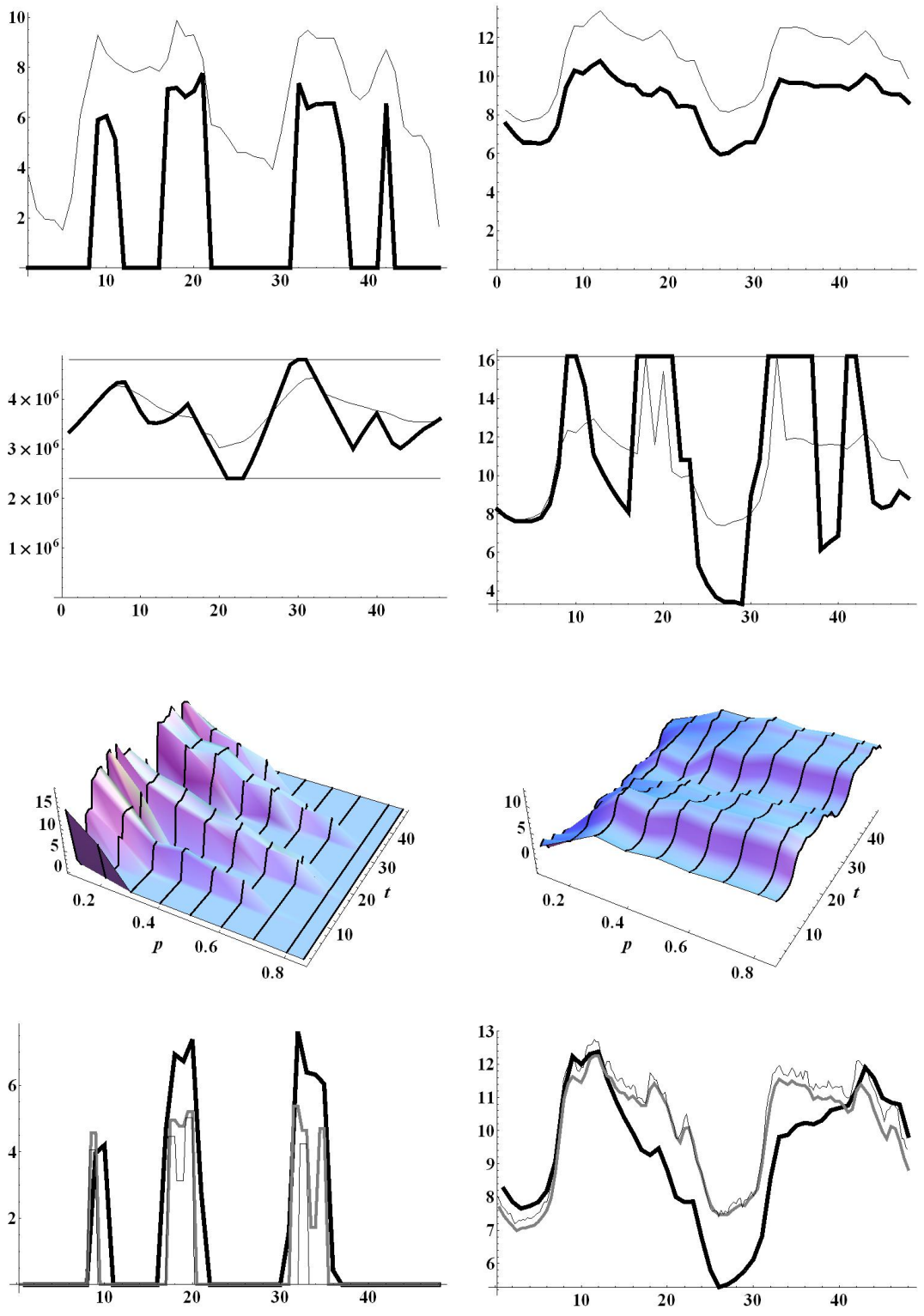

FIGURE 4. Illustration of results for the nonlinear problem (details see text).

In order to illustrate the feasibility of the obtained solutions with respect to the probabilistic demand satisfaction constraint, we make a posterior check using simulated and historical wind speed data. As an example, we consider the solution for the probability level $p = 0.7$. To the obtained profile y_t of hydro power devoted to demand satisfaction, we add 100 different wind energy profiles ξ_t and verify, whether the sum of both exceeds the demand profile. We have two ways of using wind energy data: the first way is to employ historical scenarios from the given data basis. The seconds way - which our formulation of optimization problem (26)-(31) relies on - consists in generating 100 scenarios for the random vector $\tilde{\eta}$ introduced in (21) according to the distribution parameters specified in Table 2. By omitting scenarios with possibly negative values, one ends up at scenarios η_t which via (19) and (18) finally yield simulated wind energy scenarios. Figure 5 illustrates the demand profile (black bold) and the 100 scenarios obtained by adding hydro power contribution and 100 simulated wind energy scenarios (left) as well as by adding hydro power contribution and 100 historical wind energy scenarios (right). It turns out that 76 out of the 100 simulated scenarios meet the demand throughout the whole time interval (i.e., 24 scenarios fall short of demand occasionally). This corresponds to an empirical probability of 76% for satisfying the probabilistic constraint. In the case of historical scenarios, the corresponding empirical probability equals 68%. Both values are fairly good approximations of the given level $p = 0.7$ given a data basis of only 100 scenarios.

Finally we want to illustrate the robustness of the solutions obtained with our model involving a probabilistic constraint. To this aim, we introduce first a risk-neglecting strategy obtained as a solution to the optimization problem (26)-(31) with the probabilistic constraint (27) removed. In this case, no hydro power is contributed to the demand satisfaction ($y_t = 0$) and a maximum possible hydro power generation x_t is sold on the day ahead market. This leads to a higher optimal value (= profit on day ahead market) of when compared to any solution involving the probabilistic constraint (27). Figure 6 shows the profiles for hydro energy sold on the day ahead market for the risk-neglecting solution and for a solution considering the probabilistic constraint at level $p = 0.7$. The optimal values (profit from day ahead market) of both solutions are 25698 and 868, respectively. Next we construct two different worst-case like situations for intraday prices. To this aim 1000 wind energy scenarios are simulated. In the first situation, for each scenario, we put the intraday price equal to zero whenever the demand is satisfied for the risk-neglecting solution and equal to the double of the day-ahead price whenever shortage of demand is observed for this solution. As a consequence, the intraday-market is never profitable for this solution and the final (scenario-dependent) profit is always smaller than the calculated day-ahead profit. The top left diagram of Figure 7 illustrates these scenario-dependent total profits for scenarios labeled from 1 to 1000 along the horizontal axis and opposes them to the day ahead profit (thick line). It can be seen that the loss on the intraday market due to shortage of demand can be that drastic as to yield strongly negative total profits in many cases, thus annihilating the apparently large profit on the day ahead market. The top right diagram illustrates the same situation for the probabilistic solution assuming the same set of wind energy scenarios and (scenario-dependent) intraday prices according to the construction given above. Here only very few values for the total profits fall below zero and very mildly only, if so. Most of the time the total profit is comparatively large and much larger, in particular, than the small day-ahead profit.

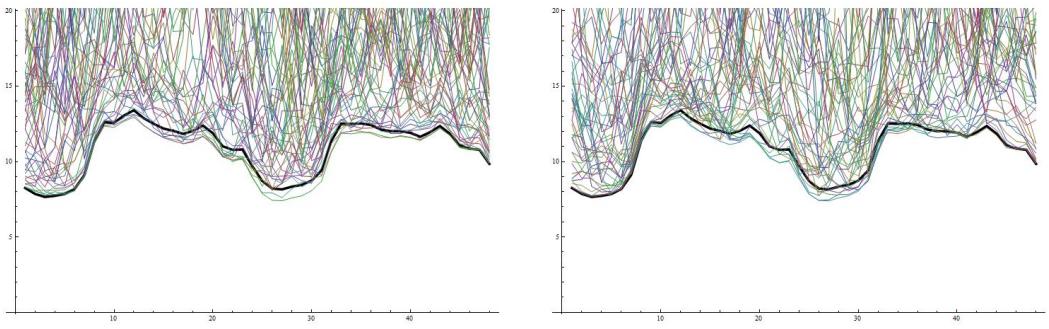

FIGURE 5. Illustration of demand satisfaction for 100 simulated (left) and 100 historical wind scenarios (right).

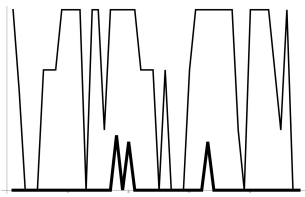

FIGURE 6. Hydro energy sold on day ahead market for a risk-neglecting (thin) and a probabilistic (thick) solution

In the second situation we relate the construction described above to the probabilistic rather than the risk-neglecting solution, i.e. for each scenario, the intraday price is put equal to zero in case that the demand is satisfied for the probabilistic solution and equal to the double of the day-ahead price in case of the probabilistic solution falling short of demand. The corresponding profits are plotted in the bottom diagrams of Figure 7. It can be seen that in this situation the risk-neglecting solutions may possibly yield a total profit exceeding the day-ahead profit. However, as before, there may occur rather frequent negative profits. In contrast, the probabilistic solution cannot yield total profits larger than the (small) day-ahead profit by construction. However, on the negative side it is still very robust so that only few total profits fall below zero and mildly only if so.

Of course, these constructed, worst-case like situations are not likely to occur in reality, but they demonstrate in the extreme, what might happen if intraday prices cannot be predicted well.

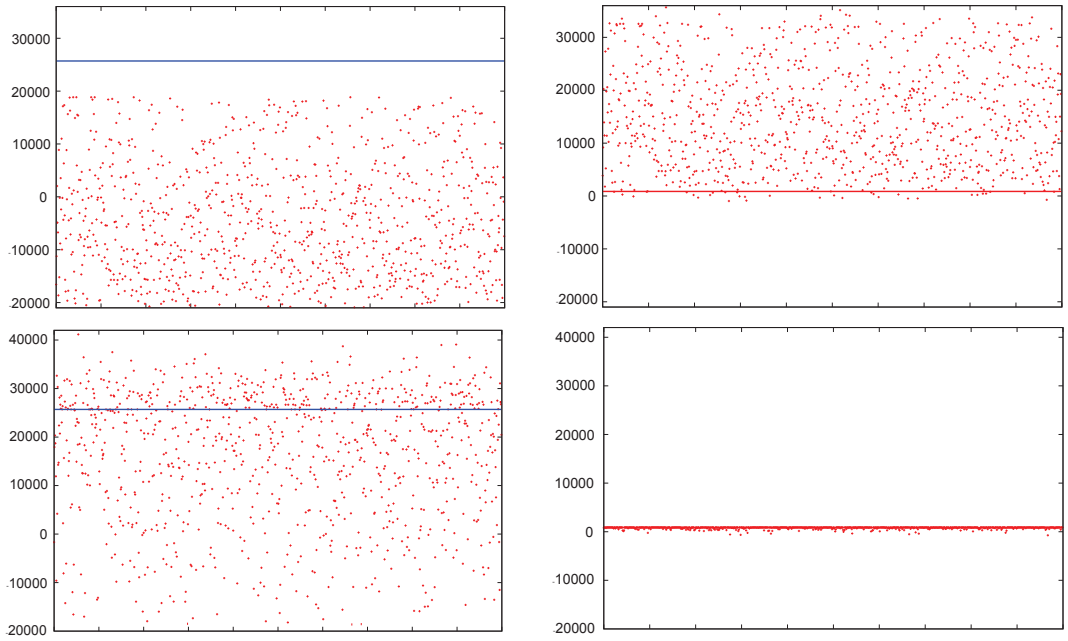

FIGURE 7. Comparison of random profits for a risk-neglecting (left) and a probabilistic (right) solution in nearly worst-case situations.

REFERENCES

[1] W. Van Ackooij, R. Henrion, A. Möller and R. Zorgati, On probabilistic constraints induced by rectangular sets and multivariate normal distributions, Mathematical Methods of Operations Research, 71 (2010), 535-549.

[2] W. Van Ackooij, R. Henrion, A. Möller and R. Zorgati, On joint probabilistic constraints with Gaussian coefficient matrix, Operations Research Letters, 39 (2011), 99-102.

[3] W. Van Ackooij, R. Zorgati, R. Henrion and A. Möller, Joint Chance Constrained Programming for Hydro Reservoir Management, Optimization and Engineering, 15 (2014), 509-531.

[4] W. Van Ackooij and R. Henrion, Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions, Weierstrass Institute Berlin, Preprint 1799 (2013), to appear in SIAM J. Optim.

[5] H. Aksoy, Z. Fuat Toprak, A. Aytek and N. Erdem Ünal, Stochastic generation of hourly mean wind speed data, Renewable Energy, 29 (2004), 2111-2131.

[6] L. Andrieu, R. Henrion and W. Römisch, A model for dynamic chance constraints in hydro power reservoir management, European Journal of Operations Research, 207 (2010) 579-589.

[7] T. Arnold, R. Henrion, A. Möller and S.Vigerske, A mixed-integer stochastic nonlinear optimization problem with joint probabilistic constraints, Pacific Journal of Optimization, 10 (2014) 5-20.

[8] A. Ben-Tal and A. Nemirovski, Robust solutions of Linear Programming problems contaminated with uncertain data, Mathematical Programming, 88 (2002), 411-424.

[9] G.C. Calafiore and M.C. Campi, The scenario approach to robust control design, IEEE Trans. Automat. Control, 51 (2006), 742-753.

[10] E. de Doncker, A. Genz, and M. Ciobanu, Parallel computation of multivariate normal probabilities, Computing Science and Statistics, 31 (1999), 89-93.

[11] D. Dentcheva and G. Martinez, Regularization methods for optimization problems with probabilistic constraints, Mathematical Programming, 138 (2013), 223-251.

[12] Y.M. Ermoliev, T.Y. Ermolieva, G.J. Macdonald and V.I. Norkin, Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risk, Annals of Operations Research 99 (2000), 207-225.

[13] A. Genz, Numerical Computation of Multivariate Normal Probabilities, Journal of Computational and Graphical Statistics 1 (1992), pp. 141-149.

[14] A. Genz and F. Bretz, Computation of multivariate normal and t probabilities., Lecture Notes in Statistics, Vol. 195, Springer, Dordrecht, 2009.

[15] P.E. Gill, W. Murray and M. A. Saunders, SNOPT: An SQP algorithm for large-scale constraint optimization, Numerical Analysis Report 97-1, Department of Mathmatics, University of California, San Diego, La Jolla, CA, 1997.

[16] R. Henrion, Gradient estimates for Gaussian distribution functions: Application to probabilistically constrained optimization problems, Numerical Algebra, Control and Optimization, 2 (2012) 655-668.

[17] R. Henrion and A. Möller, A gradient formula for linear chance constraints under Gaussian distribution, Mathematics of Operations Research, 37 (2012), 475-488.

[18] J. Luedtke and S. Ahmed, A sample approximation approach for optimization with probabilistic constraints, SIAM J. Optim. 19 (2008), 674-699.

[19] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, SIAM Journal of Optimization, 17 (2006), 969-996.

[20] N. Olieman and B. van Putten, Estimation method of multivariate exponential probabilities based on a simplex coordinates transform, Journal of Statistical Computation and Simulation 80 (2010), 355-361.

[21] B. Pagnoncelli, S. Ahmed and A. Shapiro, Sample average approximation method for chance constrained programming: Theory and applications, Journal of Optimization Theory and Applications, 142 (2009), 399-416.

[22] A. Prékopa and T. Szántai, Flood control reservoir system design using stochastic programming, Mathematical Programming Study 9 (1978), 138-151.

[23] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.

[24] A. Prékopa, *Probabilistic Programming*, Stochastic Programming (A. Ruszczyński and A. Shapiro, eds.), Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

[25] J.O. Royset and E. Polak, Extensions of stochastic optimization results to problems with system failure probability functions, Journal of Optimization Theory and Applications, 133 (2007), 1-18.

[26] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming*, MPS-SIAM series on optimization **9**, 2009.

[27] T. Szántai, Evaluation of a special multivariate Gamma distribution, Mathematical Programming Study 27 (1996), 1-16.

[28] J. Tamura, Calculation method of losses and efficiency of wind generators, in (M. Muyeen ed.): Wind energy conversion systems. Technology and trends, Springer, Berlin, 2012, pp. 25-51.