13

Chance Constrained Programming and Its Applications to Energy Management

Wim van Ackooij[1], Riadh Zorgati[2], René Henrion[3] and Andris Möller[4]
[1,2]*EDF R&D, Department OSIRIS, 1 avenue du Général de Gaulle ; F-92141 Clamart Cedex*
[3,4]*Weierstrass Institute Berlin, Mohrenstraße 39, 10117 Berlin*
[1,2]*France*
[3,4]*Germany*

1. Introduction

Chance Constrained Programming belongs to the major approaches for dealing with random parameters in optimization problems. Typical areas of application are engineering and finance, where uncertainties like product demand, meteorological or demographic conditions, currency exchange rates etc. enter the inequalities describing the proper working of a system under consideration. The main difficulty of such models is due to (optimal) decisions that have to be taken prior to the observation of random parameters. In this situation, one can hardly find any decision which would definitely exclude later constraint violation caused by unexpected random effects. Sometimes, such constraint violation can be balanced afterwards by some compensating decisions taken in a second stage. For instance, making recourse to pumped storage plants or buying energy on the liberalized market is an option for power generating companies that are faced with unforeseen peaks of electrical load. As long as the costs of compensating decisions are known, these may be considered as a penalization for constraint violation. This idea leads to the important class of *twostage* or *multistage stochastic programs* Birge & Louveaux (1997); Kall & Wallace (1994); Ruszczyński & Shapiro (2003).
In many applications, however, compensations simply do not exist (e.g., for safety relevant restrictions like levels of a water reservoir) or cannot be modeled by costs in any reasonable way. In such circumstances, one would rather insist on decisions guaranteeing feasibility 'as much as possible'. This loose term refers once more to the fact that constraint violation can almost never be avoided because of unexpected extreme events. On the other hand, when knowing or approximating the distribution of the random parameter, it makes sense to call decisions feasible (in a stochastic meaning) whenever they are feasible with high probability, i.e., only a low percentage of realizations of the random parameter leads to constraint violation under this fixed decision. A generic way to express such a *probabilistic* or *chance constraint* as an inequality is

$$\mathbb{P}(h(x,\xi) \geq 0) \geq p. \tag{1}$$

Here, x and ξ are decision and random vectors, respectively, "$h(x,\xi) \geq 0$" refers to a finite system of inequalities and $\mathbb{P}$ is a probability measure. The value $p \in [0,1]$ is called the probability level, and it is chosen by the decision maker in order to model the safety

requirements. In the following we tacitly assume that (1) represents a constraint inside an optimization problem where some objective function $f(x)$ has to be minimized. Since the role of f is as in conventional optimization problems, we shall focus our attention to the special type of constraint as given by (1).
Sometimes, the probability level is strictly fixed from the very beginning (e.g., $p = 0.95, 0.99$ etc.). In other situations, the decision maker may only have a vague idea of a properly chosen level. Of course, he is aware that higher values of p lead to fewer feasible decisions x in (1), hence to optimal solutions at higher costs. Fortunately, it turns out that usually p can be increased over quite a wide range without affecting too much the optimal value of some problem, until it closely approaches 1 and then a strong increase of costs becomes evident. In this way, models with chance constraints can also give a hint to a good compromise between costs and safety.
Formally, the chance constraint (1) may be written as a usual inequality constraint:

$$\alpha(x) \geq p, \quad \text{where} \quad \alpha(x) := \mathbb{P}(h(x, \xi) \geq 0). \tag{2}$$

In contrast to conventional optimization problems, however, the challenge posed by chance constraints consists in the fact that the function α is not given explicitly. Therefore neither theoretical properties (continuity, differentiability, concavity) nor suitable algorithmic approaches are evident. Not surprisingly, there does not exist a general solution method for chance constrained programs. The choice strongly depends on how random and decision variables interact in the constraint model. Sometimes a linear programming solver will do the job. In other models, one has to have access to values and gradients of multidimensional distribution functions (e.g., the reservoir management model of Section 6). Of particular interest is the application of algorithms from convex optimization. Convexity of chance constraints, however, does not only depend on convexity properties of the constraint function h in (1) but also of the distribution of the random parameter ξ. The question of whether this distribution is continuous or discrete is another crucial aspect for algorithmic treatment. The biggest challenges from the algorithmic and theoretical points of view arise in chance constraints where random and decision variables cannot be decoupled.
All issues discussed up to now illustrate the close tie between algorithmic and structural properties. Some of these shall be briefly presented in the following sections. The chapter is organized as follows: Section 2 is dedicated to a discussion of structural properties of chance constraints. Section 3 will illustrate the importance of stochastic programming in general and chance constrained programming in particular for energy management problems. Moreover, we will present the generic look and feel of such problems. This will be further developed in Section 4. In Section 5 recent results on CCP for Energy management structured problems will be discussed. These results are illustrated on a typical example in Section 6, that also shows that CCP can be tractable/interesting for some problems in EM and with some research effort could become a very important tool for EM under uncertainty. Finally Section 7 sketches some perspectives.
Among the numerous applications of chance constrained programming one may find areas like water resource management, circuit manufacturing, chemical engineering, telecommunications, finance and Energy management. For basic monographs on this topic, we refer to Prékopa (2003) and relevant chapters in Ruszczyński & Shapiro (2003), Shapiro et al. (2009).

2. Models and structural properties

The properties of a concrete chance constrained optimization problem mainly hinge on the following items:

- Distribution of the random vector (e.g., continuous or discrete distribution, independent or correlated components)
- Type of constraint system (e.g., linear, separated random vector, coupled random and decision vectors)
- Type of chance constraints (individual or joint)

Different combinations of elements from these basic categories may lead to mathematical objects with drastically different theoretical properties and algorithmical requirements.

2.1 Models

The most important models in practical applications of chance constraints are linear in the random vector. This means that the constraint mapping h in (1) takes one of the forms

$$h(x,\xi) = g(x) - A\xi \quad \text{or} \quad h(x,\xi) = A(\xi)g(x) - b, \tag{3}$$

where A and $A(\xi)$ are determinstic or stochastic matrices, respectively, g is a mapping just depending on the decision vector x and b is a vector of appropriate size. The basic difference between both models is that in the first case the random vector appears separated from the decision vector, whereas both are coupled in the second model. Both models have numerous applications in engineering and, in particular, in energy management.

The chance constraint (1) can be written more explicitly as

$$\mathbb{P}(h_j(x,\xi) \geq 0 \quad (j = 1,\ldots,m)) \geq p. \tag{4}$$

Since here, the probability is taken over the whole stochastic inequality system, one also calls this a *joint* chance constraint. Alternatively, one could turn each component of the stochastic inequality system into several chance constraints individually, and thereby allowing individual probability levels for each chance constraint:

$$\mathbb{P}(h_j(x,\xi) \geq 0) \geq p_j \quad (j = 1,\ldots,m) \tag{5}$$

Such *individual* chance constraints, though formally yielding a larger system of m inequalities as compared to just one inequality in the joint case, may lead to much easier mathematical models in some special cases (see Section 2.2). Care has to be taken, however, with a correct interpretation of results for these two models. If one is interested in decisions guaranteeing satisfaction of the whole stochastic inequality system at the given probability level, then a formal solution via the individual model, though appealing for its simplicity, may result in completely unreliable optimal decisions (see, e.g., van Ackooij et al. (2010c)). On the other hand, individual chance constraints may be used to derive upper and lower bounds for the optimal value in an optimization problem with joint chance constraints. More precisely, if x is feasible for (4), then x is feasible for (5) too provided that $p \geq p_j$ for all j. Conversely, if x is feasible for (5), then x is feasible for (4) too provided that $\sum_{j=1}^{m} p_j \geq p + m - 1$.

Finally, it has to be mentioned that the chance constraint (1) is of static type. This means that, if decision and random vector represent discrete time processes, then the decision policy would

be designed in a way that it does not react on previously observed realizations of the random vector. Dynamic models for chance constraints lead to new challenges and complications which are outside the scope of this presentation. For a recently proposed approach in this direction, we refer to Henrion et al. (2010).

2.2 Random right-hand side

An important special case of the linear separated model (first case of (3)) arises if the linear transformation A reduces to the identity such that the chance constraint gets the form of random right-hand side. Then, the probability function of (2) can be written as a composition

$$\alpha(x) = \mathbb{P}(g(x) \geq \xi) = F_{\xi}(g(x)), \tag{6}$$

where F_{ξ} is the cumulative multivariate distribution function of the random vector ξ. This special structure has the advantage that the effort of verifying analytical properties or of implementing numerical algorithms for the solution of chance constrained problems can be focussed on distribution functions which are well-studied objects in stochastics. The composition formula $\alpha = F_{\xi} \circ g$ allows one to transfer properties like continuity, (local or global) Lipschitz continuity or differentiability from F_{ξ} and g to α. Since the mapping g is typically given in analytical form and thus its properties are well understood from the beginning, it remains to check or to rely on known analogous properties of F_{ξ}. For instance, F_{ξ} is always continuous if the random vector ξ has a density. Differentiability and convexity are a more involved issues but can be checked for important classes of distributions (see Sections 2.3 and 2.4).

Under random right-hand side the model of individual chance constraints (5) becomes

$$\alpha_j(x) = \mathbb{P}(g_j(x) \geq \xi_j) = F_{\xi_j}(g_j(x)) \geq p_j \quad (j = 1, \ldots, m),$$

where now F_{ξ_j} refers to the one-dimensional distribution function of the component ξ_j. As one-dimensional distribution functions can be inverted via the concept of quantile, the individual chance constraints can be rewritten as

$$\alpha_j(x) \geq p_j \Longleftrightarrow g_j(x) \geq q_{p_j}^{(j)} \quad (j = 1, \ldots, m),$$

where $q_{p_j}^{(j)} := \inf\{\tau | F_{\xi_j}(\tau) \geq p_j\}$ is the p_j- quantile of F_{ξ_j}. In other words: individual chance constraints with random right hand side inherit their structure from the underlying stochastic constraint. If the latter was linear then the induced individual chance constraints will be linear too.

Another important special case under random right-hand side arises if the random vector ξ has independent components, then the calculation of α breaks down to one dimensional distribution values again:

$$\alpha(x) = F_{\xi_1}(g_1(x)) \cdots F_{\xi_m}(g_m(x)).$$

Although the constraint $\alpha(x) \geq p$ cannot be further simplified to an explicit constraint involving just the g_j (as was possible for individual chance constraints), one may still benefit from the fact that one dimensional distribution functions are usually easy to calculate. On the other hand, the independence assumption is often not reasonable in practice.

2.3 Multivariate normal distribution

Perhaps the most important special case in practical applications arises from joint chance constraints with random right-hand side having a regular multivariate normal distribution. We shall use the standard notation $\xi \sim \mathcal{N}(\mu, \Sigma)$ to indicate that ξ has a multivariate normal distribution with mean vector μ and covariance matrix Σ. Such normal distribution is called regular if Σ is positive definite. According to (6) the constraint with random right-hand side takes the form $F_\xi(g(x)) \geq p$ then. As g is explicitly given by a formula, in general, the evaluation of such constraints by optimization algorithms requires the calculation of $F_\xi, \nabla F_\xi \ldots$, i.e., of values and (higher order) derivatives of a nondegenerate multivariate normal distribution function. Fortunately, gradients of such distribution functions can be reduced analytically to some lower dimensional multivariate normal distribution functions (see Prékopa (1995), p. 204). The precise formula can be found in Lemma 0.5 below. Thus, proceeding by induction for higher order derivatives (see also Section 5.2.4), the whole optimization issue hinges upon the evaluation of nondegenerate normal distribution functions in this situation. Much progress has been made in computing such distributions functions be it by using specially designed techniques of numerical integration (Genz & Bretz (2009)) or be it by developping efficient lower and upper bounds for their values combined with adapted simulation techniques (Bukszár & Szántai (2002); Szántai & Habib (1998)). Using those methods at hand, it is possible to deal with joint chance constraints under normally distributed random right-hand side with moderate precision in moderate dimension of ξ of say up to a few hundred.

It is important to observe that, given a tool for calculating multivariate normal distribution functions, it is not only possible to deal with the special case of random right-hand side but also with the more general linear models introduced in (3). If, for instance, ξ has a multivariate normal distribution with mean vector μ and covariance matrix Ξ, then the linearly transformed random vector $\eta := A\xi$ will have a multivariate normal distribution too with mean vector $A\mu$ and covariance matrix $A\Xi A^T$. Consequently, the first model in (3) can be written without loss of generality in the special form with random right-hand side η and one is back to the situation discussed before. A similar argument applies to the second model in (3). However, one must be aware of the fact that a linear transformation of the random vector may change the normal distribution from a nondegenerate one (i.e., with positive definite covariance matrix) to a singular one. This is necessarily the case, for instance, if the number of rows in A exceeds the dimension of the random vector as is typical for instance in network problems. Then, algorithms for calculating singular normal distribution functions Genz & Kwong (2000), for calculating normal probabilities of convex sets (in particular: polyhedra) Deák (1986) or for reducing singular normal distribution functions to regular ones via an efficient inclusion-exclusion formula Henrion & Römisch (2010) can be applied. At the same time, it is also possible to obtain gradients with respect to the decision variable x in the models (3) via reduction to the calculus of values of multivariate normal distribution functions pretty much the same way (though possibly more involved) as in the case of random right-hand side. As an instance of such models which are different from random right-hand side, we discuss two-sided chance constraints with multivariate normal distribution in Section 5.2. We note that beyond normal distributions and models of type (3) gradients of probability functions α in (2) may be very difficult to obtain. For a general, abstract gradient formula, we refer to Uryasev (1995).

2.4 Convexity

Convexity is a basic issue for theory (structure, stability) and algorithms (convergence towards global solutions) in any optimization problem. In chance constrained programming, the first question one could deal with is convexity of the feasible set defined say by a very simple probabilistic constraint of the type

$$\{x|\mathbb{P}(\xi \leq x) \geq p\} = \{x|F_{\xi}(x) \geq p\}. \tag{7}$$

It is well known that such a set is convex if F_{ξ} is a quasiconcave function. Although distribution functions can never be concave or convex (due to being bounded by zero and one) it turns out that many of them are quasiconcave. The left plot of Figure 1 shows the graph of the bivariate normal distribution function with independent components. It is neither concave nor convex, but all of its upper level sets are convex (the boundary of the upper level set corresponding to the level $p = 0.5$ is depicted by a curve on the graph). For algorithmic

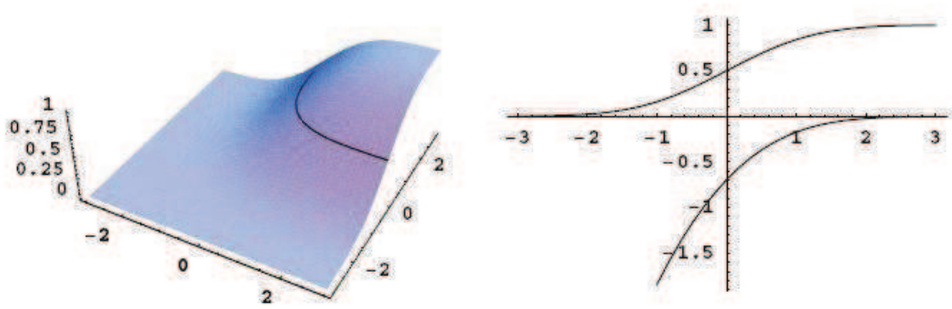


Fig. 1. Bivariate normal distribution function (left) and standard normal distribution and its logarithm (right).

purposes it is often desirable to know that the function defining an inequality constraint of type '$\geq$' is not just quasiconcave but actually concave. As mentioned above, this cannot hold for inequalities of type (7). However, a suitable transformation might do the job. Indeed, it turns out that most of the prominent multivariate distribution functions (e.g., multivariate normal, uniform distribution on convex compact sets, Dirichlet, Pareto, etc.) share the property of being log-concave, i.e., $\log F_{\xi}$ is concave (an illustration for the one-dimensional normal distribution and its log is given in the right plot of Figure 1). The key for verifying such a nontrivial property for the distribution function is to check the same property of log-concavity for the density of F_{ξ}, if it exists. The latter task is easy in general. For instance, a nondegenerate normal density is proportional to the exponential of a concave function, hence multivariate normal distributions are logarithmically concave. The mentioned result is a consequence of a celebrated theorem due to Prékopa (1995). Now, when F_{ξ} is log-concave, (7) may be equivalently rewritten as a concave inequality constraint $\{x|\log F_{\xi}(x) \geq \log p\}$ or equivalently as a convex inequality constraint $\{x|-\log F_{\xi}(x) \leq -\log p\}$. The same conclusions on convexity can be drawn for more general chance constraints of linear separated type

$$\{x|\mathbb{P}(Bx \geq A\xi) \geq p\},$$

i.e., the set of feasible decisions can be described by the convex inequality constraint

$$\{x| - \log F_{A\xi}(Bx) \leq -\log p\}$$

for the same family of distributions of ξ having log-concave densities.
Things become more involved in the feasible set

$$\{x|\mathbb{P}(A(\xi)g(x) - b) \geq p\} \tag{8}$$

of the coupled model (right case of (3)). A classical result by van de Panne & Popp (1963) and by Kataoka (1963) states that if the random matrix $A(\xi)$ reduces to just one line $A_1(\xi)$, the mapping g is the identity (i.e., $g(x) = x$), ξ has a regular multivariate normal distribution and $p \geq 0.5$, then the set (8) is convex. A first difference with the log-concavity properties stated above is that convexity of the feasible set does no longer hold true for arbitrary probability levels but only for sufficiently large ones. This, however, is not a severe restriction because in practice one is interested in large probability levels anyway (e.g., $p \geq 0.95$). This classical result has been generalized later on to other than normal distributions of ξ (e.g., elliptically symmetric or symmetric log-concave, Lagoa et al. (2005)) and to nonlinear mappings $g(x)$ (see Henrion (2007)).
Evidently, the previous results can be immediately applied to the feasible set of individual chance constraints:

$$\{x|\mathbb{P}(A_j(\xi)g(x) - b_j) \geq p_j \quad (j = 1, \ldots, m)\}.$$

Indeed, since the intersection of convex sets is convex again, it follows from the previously mentioned result for one single row $A_1(\xi)$ that this feasible set induced by a whole random matrix is convex provided that $p_j \geq 0.5$ for $j = 1, \ldots, m$. Not surprisingly, things are not that evident for the joint chance constraint (8) if $A(\xi)$ has more than just one line. Convexity results can then be found under the assumption of ξ having a normal distribution with specially structured covariance matrix (see Henrion & Strugarek (2008); Prékopa (1995)). Convexity in the general case is an open question.

2.5 Compactness

Compactness of the feasible domain is a very interesting property to check, since non-empty and compact feasible sets guarantee the existence of solutions and allow us to derive stability of results. It is interesting to observe that compactness of the coupled chance constraint (8) can be derived in case of a normal distribution without enforcing it by additional exterior deterministic constraints on the decision vector (e.g., box constraints). To be more precise, let the rows A_i of A in (8) be normally distributed according to $A_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ with positive definite covariance matrices Σ_i for $i = 1, ..., m$. Assuming that g is a homeomorphism (e.g., $g(x) = x$), then, (8) is compact provided that $p > \min_i \Phi_1(\sqrt{\mu_i^T \Sigma_i^{-1} \mu_i})$. Here, Φ_1 refers to the one-dimensional standard normal distribution function and, hence, the critical probability level beyond which compactness is guaranteed can be calculated explicitly from the distribution parameters of ξ. As a consequence, the Weierstrass Theorem ensures the existence of a solution to the optimization problem

$$\min\{f(x) \mid x \text{ satisfies (8)}\},$$

whenever the objective function f is continuous.

2.6 Discrete distributions

The setting of joint chance constraints with random right-hand side and nondegenerate multivariate normal distribution enjoys many desirable features such as differentiability or convexity (via log-concavity). Of course, other settings may have practical importance too. For instance, the distribution of the random right-hand side could be other than normal. The cases of multivariate Gamma or Dirichlet distributions are discussed in Prékopa (1995), Section 6.6. Here, log-concavity remains an important tool.

Things become different, however, when passing to discrete distributions. These are of interest for at least two reasons: first, the problem to be solved could have been directly modeled by discrete random variables (see, e.g., Beraldi & Ruszczyński (2002)). Second, there may be a need to approximate continuous distributions (e.g., multivariate normal) by discrete ones, for instance when treating probabilistic constraints in two stage models with scenario formulations Ruszczyński (2002). A key issue in discrete chance constrained programming is finding the so called p-efficient points (introduced in Prékopa (1990)) of the distribution function F_{ξ} of ξ. These are points z such that $F_{\xi}(z) \geq p$ and the relations $F_{\xi}(y) \geq p$, $y \leq z$ (partial order of vectors) imply that $y = z$. One easily observes that all the information about the p-level set of F_{ξ} is contained in these points because

$$\{y | F_{\xi}(y) \geq p\} = \bigcup_{z \in E} (z + \mathbb{R}^s_+),$$

where E is the set of p-efficient points and $\mathbb{R}^s_+$ is the positive orthant in the space of the random vector. In the case of ξ having integer-valued components and $p \in (0, 1)$, P is a finite set (see Theorem 1 in Dentcheva et al. (2000)). Algorithms for enumerating or generating p-efficient points are described, for instance, in Beraldi & Ruszczyński (2002); Dentcheva et al. (2000); Prékopa (2003); Prékopa et al. (1998). It is interesting to note that the log-concavity concept, even if not directly applicable, can be adapted with useful consequences to discrete distributions as well (see Dentcheva et al. (2000)).

Another powerful approach to solve chance constrained programs with discrete distributions via integer programming methods has been recently reported in Luedtke & Ahmed (2008).

3. Randomness and energy management optimization problems

In the electrical power industry, it is important to guarantee at each time step, the equilibrium between the offer and demand and hence avoid shortage supply. This is a major concern, whatever the time horizon. The traditional Unit Commitment Problem (UCP) consists of defining the minimal-cost power generation schedule for a given set of power plants satisfying at each time step the equilibrium between the production and the demand while respecting physical constraints. This problem, in a deterministic setting, is a challenging large-size, non-convex, non-linear optimization problem, due to many thermal and hydro power-plants constraints, which introduce discontinuous operation domains and give non-convex dynamic constraints. It has been solved satisfactory in an industrial way (Batut & Renaud (1992); Cohen & Zhu (1983); Lemaréchal & Sagastizábal (1994); Merlin & Sandrin (1983)).

Many uncertainties strongly impact the electrical power industry and should be taken into account in this problem. Uncertainty consists of the load charge curve and the hydraulic-inflows of each reservoir, both of which are climate sensitive (temperature and

cloud cover). Moreover, we have to consider the availability of the power plants, which are subject to random failure, the prices on both electricity and gas markets and wind generation. Extending, in the context of electricity markets, the traditional UCP leads to a modern Energy Management Problem (EMP). This modern version consists of optimizing the production planning, while keeping supply shortage risk under bounds, using both financial (interruption options, futures and markets) and physical (thermal and hydraulic production units) assets (Zorgati et al. (2009)). For economical reasons, one key point, which requires significant effort deals with the definition of an efficient Water Reservoir Management Problem. Such a problem can be considered as a sub problem of any EMP. Since the original deterministic UCP is already challenging, needless to say, adding uncertainty has not made things easier. Hence, in a logic of price decomposition (or optimization assets against market prices) we will typically consider subproblems of the huge EMP. It is important to note that the structures that occur in these subproblems are quite general and occur in many other Energy management problems. We refer to Section 4 for more on these structures.
When generally considering Optimization Problems encountered in Energy management, we can state that they are characterized by challenging key features such as:

- the stochastic nature of the problem, due to the uncertainty affecting the electrical system
- the stochastic nature of several physical constraints
- the nature of the decision variables of the problem (real, integer, binary/logical)
- the huge number of variables and constraints
- the non-linear (and non-convex) nature of many constraints
- bilateral constraints
- we are looking for closed loop strategies, i.e., decisions that adapt whenever the outcome of randomness is observed.

Considering the related Energy Management optimization Problems (EMOP) and a large class of other problems[1] such as long run marginal costs of energetic commodities or gas management, we clearly obtain the following generic structure of many EMOP:

$$\begin{array}{ll} \min & f[c(x,\xi)] \\ s.t. & b_l(\xi_\delta) \leq A(\xi_\alpha)x + \theta(\xi) \leq b_u(\xi_\delta) \\ & Px \leq h \\ & x \in X, \end{array} \tag{9}$$

where

- f is a risk measure on the cost function c,
- x are the controls of the problem,
- $A(\xi_\alpha)$ the matrix of the problem affected by random processes ξ_α and describing either
  - the offer. In this case α is the type of assets we consider, e.g., thermal, hydro, Futures, contracts, wind, etc...
  - a network. In this case α can be associated to coal mines, roads, gas compression stations, pipes and reservoirs, etc...

[1] with time horizons ranging from long to short term

- the (bilateral) stochastic inequality $b_l(\xi_\delta) \leq A(\xi_\alpha)x + \theta(\xi) \leq b_u(\xi_\delta)$ has to be given a meaning. For instance by using a probability constraints, i.e., $\mathbb{P}[b_l(\xi_\delta \leq A(\xi_\alpha)x + \theta(\xi) \leq b_u(\xi_\delta)] \geq p$ or by using Robust Optimization.
- the (possible void) deterministic constraints $Px \leq h$ models any polyhedral set of constraints on x
- $b_l(\xi_\delta)$ the demand affected by the random process ξ_δ. In that case b_u would be infinity. In the case of hydro management, b_l and b_u would be the lower and upper bounds on the reservoir capacity respectively.
- $\xi = [\xi_\alpha; \xi_\delta]$ is the concatenation of ξ_α and ξ_δ. Alternatively we can write $\xi = \Xi(\xi_\alpha, \xi_\delta)$ as some global random process, reflecting complex correlations and dependencies. We have expressed this feature through the use of the function Ξ.

All other specific constraints such as those appearing in water reservoir management or the nature of the controls are symbolically described by the set X. Such a set can contain all dynamic constraints on power plants for example (see Langrene et al. (2010) for the difficulties induced by such constraints).
We can distinguish three main classes of problems depending on the nature of randomness of the above stochastic inequalities:

- Only the right member b is random. We can think of coal, gas, hydro production or water reservoir management problems. In such problems, the matrix describes the topology of a system or a network and is considered fixed.
- Only the matrix A is random. This case occurs, in gas problems when considering investments on the network.
- Both A and b are random. This is the case in unit commitment and hedging problems.

4. Structure of energy management optimization problems

The general problem (9) can be declined in various subproblems. Each of these subproblems contains key features such as bilateral chance-constraints, random matrices with singularities and binary variables. The point of moving to subproblems is that these do not contain all problematic features of problem (9) at once. We can hence consider specific and adapted algorithms and methods. These models of the different subproblems often come with a robust counterpart or even an approximate chance constrained model. The results of the latter models can be compared with results obtained, using a chance-constraint formulation.
This section will detail the general structure derived from Energy management optimization problems. These structures are however far more general and can be found in many other problems. As such, the derived algorithms can be applied to problems from other contexts as well. The typical considered problems have the form

$$\begin{array}{ll} \min_x & c^\mathsf{T} x \\ s.t. & \mathbb{P}[b_l^k(\xi) \leq A^k(\xi)x + \theta^k(\xi) \leq b_u^k(\xi)] \geq p_k \ \forall k = 1, ..., K \\ & Qx \leq q \\ & x \in \mathbb{R}^{n_r} \times \{0,1\}^{n_b}, \end{array} \tag{10}$$

where the problem (10) can have unilateral (either b_l or b_u is $\pm\infty$) or bilateral constraints for any of the K (joint) chance constraints. Moreover, $n_r + n_b = n$, where n is the problem

dimension and either n_r or n_b can be zero. The matrix can be deterministic, as well as θ, b_l or b_u but never all together.
Assuming the law θ centered, and absence of uncertainty, problem (10) is basically an extension of the linear (mixed-integer) program:

$$\begin{array}{ll} \min_x & c^\top x \\ s.t. & b_l \leq Ax \leq b_u \\ & Qx \leq q \\ & x \in \mathbb{R}^{n_r} \times \{0,1\}^{n_b}, \end{array} \tag{11}$$

Since model (10) is quite general, we will give some specific versions of this model and point out the structure of the submodels.

4.1 Shortage supply hedging

In a simplified version of the stochastic unit-commitment problem we can only focus on shortage supply hedging under randomness on power plant generation and customer load. In such a setting, prices and randomness on hydro reservoirs would be considered absent (in order to simplify). This leads to a version of model (10), wherein the random matrix $A(\xi)$ has the following form :

$$A(\xi) = \begin{pmatrix} A^\theta(\xi_\theta) & A^\eta(\xi_\eta) & A^\mu(\xi_\mu) & A^\sigma(\xi_\sigma) & A^\varepsilon(\xi_\varepsilon), \end{pmatrix}$$

where $\theta, \eta, \mu, \sigma, \varepsilon$ stand for type of assets, respectively thermal, hydro, markets, contracts and renewable. The decisions x in this problem relate to production decisions on various assets. The lines of the matrix would typically correspond to different time steps in our problem and the entries of the matrix would correspond to random availability coefficients. The thermal coefficient matrix would typically have the following sparse random structure:

$$A^\alpha(\xi_\alpha) = \begin{pmatrix} a^\alpha_{11} & \dots & a^\alpha_{1N^\alpha} & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & \ddots & & & \\ 0 & \dots & 0 & \dots & a^\alpha_{i1} & \dots & a^\alpha_{iN^\alpha} & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & \ddots & & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & a^\alpha_{m1} & \dots & a^\alpha_{mN^\alpha} \end{pmatrix}.$$

A natural first idea is to use a unilateral probabilistic constraint for this model, i.e., we will assume that $b_l(\xi_\delta)$ is the random load. This would correspond to the idea that we are looking to produce a sufficient quantity (and avoid shortage supply) in most cases as randomness will affect our system after decision making. We can also argue that we would like to produce not too far from the load in a sufficient amount of cases. In that case $\theta(\xi)$ would be the negative customer load and b_l and b_u two bandwidth parameters (e.g., $\pm 500MW$). One can also imagine a series of such probabilistic constraints with increasing probability level and increasing margins. We refer to Zorgati et al. (2009) and Zorgati & van Ackooij (2010) for more on this model.
Variations of this model would consist of considering individual chance constraints rather than joint ones. The danger of such a model would be that we might avoid shortage supply with a sufficient level for each time step, but never on the global time horizon. Another

variation consists in considering that we decide on what asset to use and assume that it produces at its random maximal level. This greatly simplifies that problem as it ejects (or neglects) dynamic constraints on thermal plants. It also simplifies the hydro sub-problem as now, one only needs to know plausible hydro production trajectories, which can be pre-computed. This reduces problem (10) to a stochastic knapsack problem. Robust versions of which can be found in Klopfenstein (2007); Klopfenstein & Nace (2007; 2008).

4.2 Singularities in the random matrix?

In Section 4.1 we have seen that an important problem to consider, due to the random failure process that affects thermal units, is the following:

$$\begin{array}{ll} \min_x & c^\mathsf{T} x \\ s.c. & \mathbb{P}[A(\xi)x \geq b] \geq p \\ & x \geq 0 \end{array} \tag{12}$$

This problem generalizes the first one since it suffices to add the random b vector to the random matrix and introduce a single variable $x_{n+1} = -1$. Therefore the global setting wherein both the right member and the matrix are random can be reduced to problem (12).
However as seen in Section 4.1 the random matrix might have many (non-random) zeros and hence a priori has a distribution with many singularities. It can however be reformulated as a random-vector problem and do away with the singularities. This reformulation is very useful when we want to compute the probabilities for each x. To this end, let us define the following operator T, $T(x) = \text{diag}((x^\mathsf{T}, ..., x^\mathsf{T})^\mathsf{T})$. We can remark that T is actually a linear operator. We also define the following matrix operation $A \mapsto A^\odot$, which we shall call the vector transform, by $A^\odot = (A_{11}, ..., A_{1n}, A_{21}, ..., A_{nn})^\mathsf{T}$. Then the system in equation (12) can also be rewritten as follows

$$\mathbb{P}[T(x)A^\odot \geq b] \geq p. \tag{13}$$

What is very interesting about this transform is that if the original A matrix contained some non-random zero components due to a formulation issue, as is the case for the thermal production matrix A^θ then applying this transform we can actually place the zero components in the $T(x)$ decision matrix and obtain a random vector $A^\odot$ that does not contain any singularities. Moreover if we assume that $A^\odot$ is actually a normally distributed random vector with covariance matrix Σ, then computing the probability (13) comes down to computing a multivariate normal cumulative distribution function having covariance matrix $T(x)\Sigma T(x)^T$. One can therefore see that the number of columns of matrix A doesn't really matter here as the probability that has to be computed is normal of dimension the number of rows of A.

4.3 Hydro reservoir management

The hydro subproblem of problem (10) is of particular interest as it has a structure that is common to many other network flow problems with randomness. Indeed, in such problems, we typically have righthand side randomness. In particular matrix A describes the topology of the systems, i.e., the flow constraints. Randomness occurs as in each node of the network random quantities are withdrawn (customer load in a coal-mine investment model with random load) or added (random water inflows in a hydro reservoir model). The cost vector can describe investment and transportation costs (coal-mine model, Lepaul (2009)) or water turbining costs (where we assume that volume dependent water values are available). Further

deterministic constraints describe non-random parts of the model, such as reservoirs that are not impacted by random inflows, or nodes in the network not subject to random load (mines, roads). We refer to van Ackooij et al. (2010b;c) for more information on the hydro reservoir model. This subproblems also offers an alternative formulation as robust optimization (see Appariagliato et al. (2006)).

5. Chance constrained programming results for EM

When considering chance constrained optimization problems, such as the EMP (9) two important paths can be taken. We can either try to solve the problem exactly or we can try to find a good approximation of the problem. In the first setting it is important to dispose of a way to evaluate the probability constraint for any x quickly and dispose of a way to compute gradients (see van Ackooij et al. (2010c)), second derivatives (van Ackooij et al. (2010b)) and exploit information in the covariance matrices of the uncertainty factors (see van Ackooij et al. (2010a)) combined with Prékopa's LP method (Prékopa (1995)). In the second approach, the difficulty resides in finding a good approximation of the chance-constraint. This can be typically done by bounding the contraint. The advantage often resides in the fact that the approximation holds for all laws. Hence, we can obtain convex approximations of a CCP. In EM, for some problems with random matrices the decision vector contains binary variables. Such stochastic knapsack problems can be solved approximately by combining inner and outer bounds on the probability measure (see Zorgati & van Ackooij (2008; 2010)). Another approach is Robust Knapsack problems, such as those considered in Klopfenstein (2007); Klopfenstein & Nace (2007; 2008). These approaches can also be handsomely compared on the same problem. Such approximation schemes can also be used in a continuous setting, i.e., one wherein the decision vector x is real (see Zorgati et al. (2010)). The advantage of using such approximation techniques is that they transform the potentially non-convex chance constraint problem (if we take exotic laws) into a conic quadratic problem. The price of which is an approximation.
In this section we will discuss both paths.

5.1 Approximate chance constrained programming: Bounds

5.1.1 Minimal information about uncertainties

Two major questions have to be investigated in the aim of taking uncertainties into account in the optimization process. First, some knowledge about random processes has to be available. Secondly, provided that such knowledge is available, how can we integrate the associated information into the optimization process? These questions are key questions in stochastic optimization and are in practice very difficult.
Since laws are not precisely known or very complex, we aim to approximately solve the problem. We choose here a very simplistic solution based on minimal available information about uncertainties. We assume that, for any random parameter r, we know the average $r_{mean} = \mathbb{E}(r)$, the maximal value, r_{max} and its minimal value r_{min}, all derived from historically observed values.
No further hypothesis are made about the underlying random process. We will just suppose that all uncertain coefficients of the matrix A and vector b are bounded independent random variables. Boundedness is not a restrictive assumption as all borelian random variables are tight and can therefore be assumed to be almost bounded.

5.1.2 Conic approximations of individual chance constraint

Approximate solution of the probabilistic model can be obtained using the following result, the proof of which follows from an application of Hoeffding's Theorem (Hoeffding (1963)) and can be found in Zorgati et al. (2010) for each individual chance constraint, i.e., each line of the individual chance constrained stochastic matrix inequality system:

Lemma 0.1. *Let $a_j, b, j = 1, ..., n$ be almost surely bounded independent random variables and let A denote the random vector a. We will note these bounds by a_j^{min}, b^{min} and a_j^{max}, b^{max}. Furthermore we define the (semi positive definite) diagonal matrix Δ as $\Delta = \mathrm{diag}((a_1^{max} - a_1^{min}, ..., a_n^{max} - a_n^{min})^{\mathsf{T}})$. Any individual chance constraint*

$$\mathbb{P}[\langle A(\xi), x\rangle \geq b(\xi^{\delta})] \geq \alpha \tag{14}$$

can be bounded by the 2 following convex conic quadratic inequalities:

$$\begin{aligned} \langle \mathbb{E}[A(\xi)], x)\rangle - \sqrt{(1/2)|\ln(1-\alpha)|}\,\|\Delta x + \delta_b\| &\geq \mathbb{E}(b) \\ \langle \mathbb{E}[A(\xi)], x)\rangle &\geq \mathbb{E}(b), \end{aligned}$$

where bounded means that the feasible set of equation (14) contains the feasible set of the 2 convex conic inequalities.

As a consequence, the individualized and unilateralized version of the constraints in the general problem (10) related to time step i:

$$\begin{aligned} \mathbb{P}(\langle A_i, x\rangle \leq b_l + \theta_i(\xi)) &\geq \beta_i \\ \mathbb{P}(\langle A_i, x\rangle \geq b_u + \theta_i(\xi)) &\geq \beta_i \end{aligned}$$

can be approximated by

$$\begin{aligned} \langle A_i, x\rangle &\leq b_l + \mathbb{E}(\theta_i) + \sqrt{(1/2)\,|\ln 1 - \beta_i|}R_i \\ \langle A_i, x\rangle &\leq b_l + \mathbb{E}(\theta_i) \\ \langle A_i, x\rangle &\geq b_u + \mathbb{E}(\theta_i) + \sqrt{(1/2)\,|\ln 1 - \beta_i|}R_i \\ \langle A_i, x\rangle &\geq b_u + \mathbb{E}(\theta_i), \end{aligned}$$

where $R_i = [max(\theta_i) - min(\theta_i)]^2$
If the initial problem has m constraints and mn variables, the convex approximation using the result leads to a problem with $m(2n+5)$ constraints and mn variables.
This result implies that any individual chance-constrained optimization problem of the form (10) can be approximated by the convex conic quadratic problem :

$$\begin{aligned} \min_x \quad & c^t x \\ s.t. \quad & \left\|\tilde{A}_l x + \tilde{b}_l\right\|_2 \leq \tilde{f}_{l^t} x + \tilde{d}_l, l = 1, ..., L, \end{aligned}$$

since, by Lemma 0.1, any linear constraint corresponds to a particular case of conic quadratic constraint with null matrix A_i and null vector b_i and any positivity constraint can also be written in a conic quadratic form (with $f_i = 0$ (Alizadeh & Goldfarb (2001); Lobo et al. (1998))).

Then, by applying Schur's complement theorem, it is easy to give the Semi-Definite version of this conic quadratic approximation :

Corollary 0.2. *Any individual chance constraint:*

$$\mathbb{P}[\langle A_i(\xi), x\rangle \geq b_i(\xi_\delta)] \geq \alpha_i,\ \forall i \in I$$

can be bounded by the following semi-definite condition:

$$\begin{bmatrix} (\tilde{f}_l^t x + \tilde{d}_l) I & \tilde{A}_l x + \tilde{b}_l \\ (\tilde{A}_l x + \tilde{b}_l)^t & \tilde{f}_l^t x + \tilde{d}_l \end{bmatrix} \succeq 0,\ l \in (1, L),$$

where notations are as in Lemma 0.1

5.1.3 Approximations in the combinatorial case : Stochastic Knapsack problems

As indicated earlier, when problem (10) only contains binary decisions, we are facing a stochastic multi-knapsack problem. By considering individual chance constraints, using finite subadditivity of the probability measure and the inclusion-exclusion principle, we show that thanks to Hoeffding's inequality, any chance constraint can be approximated by an "outer" bound for m odd and by a "inner" bound for m even. This leads to a robust mixed inner-outer algorithm that allows us to approximately solve our binary chance-constrained program and, in general, any stochastic Multi-Knapsack Problem, i.e., canonical problems of the form

$$\begin{array}{ll} \min_{x\in\{0,1\}^n} & c^\top x \\ s.t. & \mathbb{P}[A(\xi)x \geq b] \geq 1 - p \end{array} \tag{15}$$

We refer to Zorgati & van Ackooij (2010) for the proofs of the theorems in this paragraph.

5.1.3.1 Method 1 : Mixed Inner Outer approximation (AMIO)

The following approximation is based on Hoeffding's inequality.

Lemma 0.3. *Let u be the all-one vector. Assuming $\langle \mathbb{E}(A_i), u\rangle \leq b_i$ and fixing $\tau_i \geq 0$ such that $b_i = \tau_i + \langle \mathbb{E}(A_i), u\rangle$, we obtain*

$$\mathbb{P}[\langle A_i, x\rangle \geq b_i] \leq \exp\left(-\frac{2\tau_i^2}{\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2}\right).$$

Whenever $\langle \mathbb{E}(A_i), u\rangle > b_i$, we obtain

$$\mathbb{P}[\langle A_i, x\rangle \geq b_i] \geq 1 - \exp\left(-\frac{2\tau^2}{\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2}\right),$$

where $\tau = \min_x \langle \mathbb{E}(A_i), x\rangle - b_i$.

Lemma 0.4. *Define $\tau_i(x) = b_i - \langle \mathbb{E}(A_i), x \rangle$. Let x be a feasible solution of the following problem*

$$\begin{aligned} \min_{(x,z)\in\{0,1\}^{n+m}} \quad & c^{\mathsf{T}} x \\ s.t. \quad & -\hat{C}_i z_i + (1-z_i)\ln(1-p) \leq \frac{-2(b_i - \langle \mathbb{E}(A_i), x\rangle)^2}{\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2} \\ & \frac{-2(b_i - \langle \mathbb{E}(A_i), x\rangle)^2}{\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2} \leq \ln(p) + C(1-z_i) \\ & -M_i z_i \leq \tau_i(x) \leq M_i(1-z_i) \end{aligned} \tag{16}$$

where C is such that $\exp(C) \geq \frac{1}{p}$, $\hat{C}_i = 2\frac{M_i^2}{\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2}$ and M_i some big constant such that $\tau_i(x) \in [-M_i, M_i]\ \forall x$, then x is feasible for the canonical problem on constraint i if $\tau_i(x) \leq 0$. If x is a feasible solution for constraint i of the canonical problem (15) and $\tau_i(x) \geq 0$ then x is feasible for constraint i for (16).

We will call the problem (16) the mixed-inner-outer approximation (MIO) of the canonical problem. It is a linear problem if we remark that z_i is binary, and the fact that the first constraint is active whenever $z_i = 0$ and the second when $z_i = 1$. Indeed the following problem is equivalent to MIO :

$$\begin{aligned} \min_{(x,z)\in\{0,1\}^{n+m}} \quad & c^{\mathsf{T}} x \\ s.t. \quad & -\tau_i(x) \geq \sqrt{-\frac{1}{2}\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2 \ln(p) z_i} \\ & \tau_i(x) \geq -M_i z_i \\ & -\tau_i(x) \geq -\sqrt{\frac{1}{2}\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2 \hat{C}_i z_i} + \sqrt{-\frac{1}{2}\sum_{j=1}^n (\overline{a}_{ij} - \underline{a}_{ij})^2 \ln(1-p)(1-z_i)} \\ & -\tau_i(x) \geq -M_i(1-z_i) \end{aligned} \tag{17}$$

The interpretation is noteworthy since the ratios inside the constraints are the expected difference between load and the production normalized by the total power available at time step k. Indeed on some constraints we will have enforced the original constraint, therefore obtaining a feasible, but potentially costly solution. However on some other constraints we will have relaxed the original constraint, therefore obtaining a potentially non-feasible but cheap solution. On some examples, this allows us to approximate rather accurately the optimal cost.
We will speak of the augmented MIO problem whenever the objective function is replaced by $\min_{x,z} c^T x - a^T z$, for some positive vector a. The point in adding the additional a vector is giving additional value to the event $\tau_i(x) \leq 0$, which is the average version of what we wish to achieve with our chance constraint! The more negative $\tau_i(x)$, the likelier the chance constraint is satisfied.

5.1.3.2 Method 2 : Robust Knapsack Formulation (RKP(Γ))

Following Klopfenstein & Nace (2008), we can build a Robust Knapsack version of our problem (10) $RKP(\Gamma)$. We thus obtain

$$\begin{array}{ll} \min & c^T x \\ s.t. & \sum_{j \in S} \underline{a}_{ij} x_j + \sum_{j \notin S} \overline{a}_{ij} x_j \geq b_i \forall i \forall S \subset \{1, ..., n\}, |S| = \Gamma. \end{array}$$

Here Γ is a hardness parameter. Taking for instance $\Gamma = n$ gives the full-robust solution, i.e., whatever the realization of uncertainty the chance-constraint is satisfied. If the above problem is infeasible for $\Gamma = 0$ there is no solution to problem (15) either. The problem $RKP(\Gamma)$ can be solved using a dynamic programming algorithm as indicated in Klopfenstein & Nace (2008). The main difficulty in these approximations is that for many Γ the $RKP(\Gamma)$ problem may be infeasible.

5.2 Gradients for two-sided chance constraints under multivariate normal distribution

In the hydro sub-problem that we consider (Section 4.3), probabilistic constraints are induced by two-sided stochastic inequalities. Indeed we have seen that it is of the following form:

$$\min\{c^{\mathsf{T}} x \mid \mathbb{P}(Ax + a \leq L\xi \leq Bx + b) \geq p\}, \tag{18}$$

where A, B, L and a, b, c are matrices and vectors, respectively, of appropriate orders. Assuming inflows normally distributed, these inequalities bound a normally distributed random vector by some decision-dependent functions. More precisely the probabilistic constraint may take the form

$$\mathbb{P}(\alpha(x) \leq \xi \leq \beta(x)) \geq p.$$

We refer to van Ackooij et al. (2010c) for more on these methods.

Here, ξ is a random vector having a regular multivariate normal distribution, $\mathbb{P}$ denotes the probability measure, $p \in (0, 1)$ is a probability level and x refers to a decision vector. In geometric terms, it is required that the probability of some x-dependent rectangle be not smaller than p. In order to determine an optimal decision x^* in the context of an optimization problem, one has to have access to values and derivatives of this probability function. As far as values are concerned, one may employ numerical algorithms designed for the calculus of normal distribution functions Szántai (2000), of normal probabilities of general convex sets Déak (1980) or directly of rectangles Genz (1992). However, none of these algorithms provides gradients of the probability function with respect to changes of the lower and upper limit of the rectangle. In case of one-sided constraints (i.e., $\alpha = -\infty$, so that one is dealing with distribution functions), there is no problem to reduce the computation of the gradient to that of a value of a distribution function (see Lemma 0.5 below). Formally, one could also do so with gradients of two-sided constraints by exploiting a representation of rectangle probabilities in terms of distribution functions (see (19)) and then taking derivatives of the latter ones term by term. We note that such representation allowing for reduction of derivatives to those of distribution functions is available even for general polyhedra Henrion & Römisch (2010). This approach, however, becomes impractical already in small dimension. For example in the case of an n-dimensional rectangle, the number of terms in the representation equals 2^n.

5.2.1 Constraints induced by rectangular sets and multivariate normal distributions

We present a simple formula for the derivative of the normal probability of rectangles with respect to their lower and upper limits. In particular, this formula allows to reduce the problem to the same calculus of probabilities of rectangles (in one dimension less). Consequently, the same algorithm in Genz (1992) can be used for computing values and derivatives of the probability function introduced above.

Let ξ be some n-dimensional random vector having a nondegenerate multivariate normal distribution with mean vector μ and covariance matrix Σ. We will write $\xi \sim \mathcal{N}(\mu, \Sigma)$ for short. Denote by

$$\Phi_{\xi}(z) := \mathbb{P}(\xi \leq z) \quad \forall z \in \mathbb{R}^n$$

its cumulative distribution function (with $\mathbb{P}$ referring to the underlying probability measure). We further introduce the rectangle probability function

$$F_{\xi}(a, b) := \mathbb{P}(a \leq \xi \leq b) \quad \forall a, b \in \mathbb{R}^n : a \leq b.$$

The following relation is well known to hold whenever $a \leq b$:

$$F_{\xi}(a, b) = \sum_{i_1, \ldots, i_n \in \{0,1\}} (-1)^{\left[n + \sum_{j=1}^{n} i_j\right]} \Phi_{\xi}(y_{i_1}, \ldots, y_{i_n}), \tag{19}$$

where

$$y_{i_j} := \begin{cases} a_j & \text{if } i_j = 0 \\ b_j & \text{if } i_j = 1 \end{cases}.$$

For instance, if $n = 2$, the probability of the rectangle $[a, b]$ calculates via the distribution function as

$$F_{\xi}(a, b) = \Phi_{\xi}(a_1, a_2) - \Phi_{\xi}(a_1, b_2) - \Phi_{\xi}(b_1, a_2) + \Phi_{\xi}(b_1, b_2).$$

The following lemma can be found (in its equivalent form for standard normal distributions) in Prékopa (1995). It shows how the derivative of a multivariate normal distribution can be reduced to values of a different multivariate normal distribution (in one dimension less):

Lemma 0.5. *Assume that $\xi \sim \mathcal{N}(\mu, \Sigma)$ with some positive definite covariance matrix $\Sigma = \left(\sigma_{ij}\right)$. Then, Φ_{ξ} is contiuously differentiable and*

$$\frac{\partial \Phi_{\xi}}{\partial z_i}(z) = f_{\xi_i}(z_i) \cdot \Phi_{\tilde{\xi}(z_i)}(z_1, \ldots, z_{i-1}, z_{i+1} \ldots, z_s) \quad (i = 1, \ldots, n).$$

Here, f_{ξ_i} denotes the one-dimensional probability density of the component ξ_i, $\tilde{\xi}(z_i)$ is an $n-1$-dimensional random vector distributed according to $\tilde{\xi}(z_i) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$, $\hat{\mu}$ results from the vector $\mu + \sigma_{ii}^{-1}(z_i - \mu_i)\sigma_i$ by deleting component i and $\hat{\Sigma}$ results from the matrix $\Sigma - \sigma_{ii}^{-1}\sigma_i\sigma_i^T$ by deleting row i and column i, where σ_i refers to column i of Σ.

In the next theorem, we generalize Lemma 0.5 to the case of probability functions F_{ξ} defined by rectangles. In particular, the presented formula allows to again reduce the derivative of F_{ξ} to the calculus of values of a similar function induced by a different normally distributed random vector. The proof of the theorem can be found in van Ackooij et al. (2010c)

Theorem 0.6. *Assume that $\xi \sim \mathcal{N}(\mu, \Sigma)$ with some positive definite covariance matrix Σ. Then, for $i = 1, \ldots, n$,*

$$\frac{\partial}{\partial b_i} F_\xi(a,b) = f_{\xi_i}(b_i)\, F_{\tilde{\xi}(b_i)}(\tilde{a},\tilde{b}) \tag{20}$$

$$\frac{\partial}{\partial a_i} F_\xi(a,b) = -f_{\xi_i}(a_i)\, F_{\tilde{\xi}(a_i)}(\tilde{a},\tilde{b}). \tag{21}$$

Here, f_{ξ_i} is as in Lemma 0.5, $\tilde{\xi}(b_i), \tilde{\xi}(a_i)$, are $n-1$-dimensional random vectors distributed according to $\tilde{\xi}(b_i), \tilde{\xi}(a_i) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$ such that $\hat{\mu}$ results from the vector $\mu + \sigma_{ii}^{-1}(b_i - \mu_i)\sigma_i$ (in case of b_i) or from the vector $\mu + \sigma_{ii}^{-1}(a_i - \mu_i)\sigma_i$ (in case of a_i) by deleting component i and $\hat{\Sigma}$ is defined as in Lemma 0.5. Moreover $\tilde{a}$ and $\tilde{b}$ result from a and b by deleting the respective component i.

In order to demonstrate the impact of the derived formula, we consider the optimization problem (18). Given that ξ (and so $L\xi$ too) has a multivariate normal distribution, we know from Prékopa (1995) that the function

$$x \mapsto \log \mathbb{P}(Ax \leq L\xi \leq Bx) \tag{22}$$

is concave. This allows to rewrite the optimization problem as a convex one:

$$\min\{c^T x \mid -\log \mathbb{P}(Ax + a \leq L\xi \leq Bx + b) \leq -\log p\}$$

Now one can apply, for instance, a supporting hyperplane type method as described in Prékopa (1995) in order to solve this problem. This requires, apart from functional values, also to calculate gradients of the function (22) which amounts to determine partial derivatives of the function $F_\xi(Ax, Bx)$ introduced above. The latter task can efficiently be realized with the aid of the formula given in Corollary 0.8. It resides in the fact that we rely on the same algorithm as used for determining values of F_ξ.

5.2.2 Convexity of rectangular constraint problems?

When looking at the definition of the function h in (4), we can see that we are dealing here with the special case

$$h(x, \xi) := Ax + B\xi - c, \tag{23}$$

of separated linear constraints. In (23), A and B may represent matrices which describe how releases x and inflows ξ accumulate over time and how reservoirs are interconnected. The vector c provides certain lower and upper levels in the reservoirs which have to be respected (possibly time-dependent).

Defining the linearly transformed random variable $\eta := -B\xi$, we may rewrite the probabilistic constraint associated with (23) as

$$\mathbb{P}(Ax + B\xi \geq c) \geq p \Longleftrightarrow \mathbb{P}(Ax - c \geq \eta) \geq p \Longleftrightarrow F_\eta(Ax - c) \geq p, \tag{24}$$

where F_η refers to the (multivariate) distribution function of η. This means, the probabilistic constraint is equivalent to a single inequality in the decision vector x which can be evaluated (e.g., in the framework of a nonlinear optimization code) if one is able to cope with multivariate distribution functions.

Clearly, problem (18) can be cast into the equivalent problem

$$\min\left\{\langle d,x\rangle \,|F_\eta\left(u\right)\geq p,\ u=Ax-c\right\}. \tag{25}$$

The key observation for a numerical treatment of (25) in the framework of convex optimization is that many prominent multivariate distribution functions (e.g., regular and singular normal, Dirichlet, Gamma, Wishart uniform etc.) share the property of *log-concavity*:

$$\log F_\eta\left(\lambda_1 u_1+\lambda_2 u_2\right)\geq \lambda_1 \log F_\eta\left(u_1\right)+\lambda_2 \log F_\eta\left(u_2\right)\quad \forall u_1,u_2\ \forall \lambda_1,\lambda_2\geq 0:\lambda_1+\lambda_2=1.$$

The verification of log-concavity for distribution functions is based on the celebrated Theorem by Prékopa (see Prékopa (1973)) stating that a distribution function is log-concave if and only if the density function has this property (see Henrion & Strugarek (2008)), for these distributions one may pass to the equivalent (by monotonicity of log) optimization problem

$$\min\left\{\langle d,x\rangle \,|\log F_\eta\left(u\right)\geq \log p,\ u=Ax-c\right\}. \tag{26}$$

5.2.3 A cutting planes algorithms for joint chance constrained programming

Being that $\log F_\eta$ is a concave function, (26) becomes a convex optimization problem[2]. This can be solved, for instance, by means of the cutting plane method. As it is well-known, the following ingredients are required for the application of the cutting plane method:

- a Slater point $(\hat{x},\hat{u})$ satisfying $F_\eta\left(\hat{u}\right)>p,\ \hat{u}=A\hat{x}-c$
- a procedure to calculate the distribution function F_η in order to determine in each iteration k a point $\tilde{u}_k$ on the line segment $[u_k,\hat{u}]$ satisfying $F_\eta\left(\tilde{u}_k\right)=p$. Here, u_k is part of the current iterate (x_k,u_k).
- a procedure to calculate the gradient ∇F_η in order to add in each iteration k a cut $\left\langle \nabla F_\eta\left(\tilde{u}_k\right),u-\tilde{u}_k\right\rangle\geq 0$.
- a linear programming solver for solving (26) but with the nonlinear constraint $\log F_\eta\left(u\right)$ replaced by the accumulated cuts (linear constraints) from the previous item.

The last requirement being standard, we adress the first three items in the following subsections. From now on we restrict our considerations to the - most important case - of normally distributed random vectors. For the calculation of other distributions like t-distribution, Gamma-distribution, Dirichlet-distribution or Exponential distribution, we refer to Genz (2002), Szántai (1996), Gouda & Szántai (2004) and Olieman & van Putten (2006).

5.2.3.1 Calculation of multi-variate normal distribution functions

As mentioned before, we assume from now on that η obeys a multi-variate normal distribution. We write $\eta\sim\mathcal{N}\left(\mu,\Sigma\right)$ to say that the expectation of η equals μ and the covariance matrix equals Σ. Codes for calculating the associated distribution function F_η typically assume that η be standardized, such that $\mu=0$ and $\Sigma_{ii}=1$ (i.e., Σ is actually a correlation matrix). This standardization is easily carried out by introducing the transformed random vector

$$\tilde{\eta}:=T\left(\eta-\mu\right),$$

[2] alternatively we may impose (u,x) to be in some general convex set and problem (26) remains a convex optimization problem

where T is a diagonal matrix with entries $\Sigma_{ii}^{-1/2}$. Then, $\tilde{\eta} \sim \mathcal{N}(0, R)$, where R is the correlation matrix associated with Σ. Then, the relation between the values of the original and the standardized distribution functions is given by

$$F_\eta(u) = \mathbb{P}(\eta \leq u) = \mathbb{P}(\tilde{\eta} \leq T(u - \mu)) = F_{\tilde{\eta}}(T(u - \mu)).$$

Therefore, it is sufficient to have access to algorithms calculating standardized distribution functions. For algorithms doing this job we refer as examples to Szántai (2000) or Genz (1992); Genz & Kwong (2000). The difference between the two approaches is that the first one relies on a combination of simulation and efficient probability bounds from modern graph theory, whereas the second one employs a clever scheme of numerical integration. There is one peculiarity to be respected in our model: the random vector η was already obtained from the original random vector ξ via a linear transformation: $\eta = -B\xi$ (see (24)). Of course, assuming that already ξ had a multi-variate distribution, say $\xi \sim \mathcal{N}(\mu', \Sigma')$ we know that so has η and we even know how the parameters of η's distribution are related to those of ξ:

$$\mu = -B\mu' \quad and \quad \Sigma = B\Sigma' B^T. \tag{27}$$

Many algorithms for calculating multi-variate normal distributions (such as Szántai (2000)) assume that this distribution is regular, i.e., the covariance matrix is positive definite. There is not much loss of generality to assume that original random vectors in practical applications, such as our ξ, follow indeed a regular normal distribution. However, in our optimization problem (26), we deal with the transformed random vector η rather than with ξ and it is clear that the transformation of covariance matrices in (27) destroys the regularity of the covariance matrix whenever B does not have full rank. But such is typically the case in network problems and it will turn out to be also the case in our application to water reservoirs due to considering lower and upper reservoir levels simultaneously. Then, one may benefit from the algorithm presented in Genz (1992) (see also Genz & Kwong (2000)). We mention that algorithms for calculating regular normal distributions can also be applied to problems with singular normal distributions (by using some efficient inclusion-exclusion formula presented in Bukszár et al. (2004)) and then turn out to be very fast but they require the determination of all vertices of a polyhedron which limits its use to small dimensions.

5.2.3.2 Calculation of gradients to multi-variate normal distribution functions

By combining the results of Theorem 0.6 with those from Corollary 0.8, computing the gradients of the chance constraint in problem (18) comes down to evaluation normal densities in dimension one and multi-variate normal density functions. The same remarks as those made in Section 5.2.3.1 apply however.

5.2.3.3 Determination of a Slater point

Given the probability level p one actually does not know in advance whether or not the optimization problem (25) has a feasible solution at all. Indeed, choosing a too large safety level p may lead to an empty feasible set. Much less one has direct access to a Slater point which strictly satisfies the probabilistic constraint. In order to get more information here, one may solve the following auxiliary problem which is also called 'max p'-problem:

$$\max \left\{ p | F_\eta(u) \geq p,\ u = Ax - c \right\}. \tag{28}$$

This looks pretty much the same as (25) but the difference is that the objective now is to maximize the safety level (rather than minimize some cost function) and that optimization takes place with respect to variables (x, u, p) (whereas in (25) p was fixed). Nevertheless, one may transform (28) again into a convex optimization problem. First, apply the same logarithmic transformation as above:

$$\max \left\{ p | \log F_\eta (u) \geq \log p,\ u = Ax - c \right\}. \tag{29}$$

Here, the mapping $\log F_\eta (u) - \log p$ defining the inequality constraint is not concave in both variables (u, p) simultaneously. However, (29) is easily seen to be equivalent with

$$\max \left\{ p' | \log F_\eta (u) \geq p',\ u = Ax - c \right\}. \tag{30}$$

Indeed, (x^*, u^*, p^*) is a solution of (29) if and only if $\left(x^*, u^*, e^{p^*}\right)$ is a solution of (30). On the other hand, (30) is a convex problem because the mapping $\log F_\eta (u) - p'$ defining the inequality constraint now is concave in both variables (u, p') simultaneously. Of course, now one is formally faced again with the four items required for a cutting plane method mentioned above. However, the last three items are covered by the same arguments as before (calculus of F_η, ∇F_η and linear optimization solver). Concerning the first item, the Slater point, this problem is solved very easily for (29) or (30), respectively, because the safety level is no longer fixed but becomes a variable. So it suffices to put in (30)

$$\left(\hat{x}, \hat{u}, \hat{p}'\right) := \left(0, -c, \log F_\eta (-c) - \varepsilon\right)$$

for some sufficiently small $\varepsilon > 0$ to see that

$$\log F_\eta (\hat{u}) > \hat{p}' \quad and \quad \hat{u} = A\hat{x} - c.$$

Once, (30) (and thus (29)) is solved, the optimal solution (x^*, u^*, p^*) of (29) can be used to derive a Slater point for the original optimization problem (25). Indeed, if it turns out that the maximum possible probability level p^* is smaller than the level p chosen by the decision maker in (25), then this latter program will not have any feasible solution at all and the decision maker will have to adjust (reduce) his safety level. Otherwise, if $p^* > p$, then (x^*, u^*) may obviously be used as a Slater point for the original problem (25). A part from the meaning of the 'max p'-problem for the determination of a Slater point in the original problem, its solution provides useful additional insight: indeed, the associated part x^* of its solution indicates the most robust decision possible. In the application to water reservoirs it will represent the most robust release control in order to keep the level constraints of the reservoir with maximum possible probability. Of course, this robust control will come at a significantly higher price (in terms of the cost function $c^\mathsf{T} x$ in (25)).

5.2.4 Second order derivatives

If one is interested in applying second order solution methods to increase the efficiency of the solution process, one has to work out second derivatives of the probability function φ (where notations are as in corollary 0.8) on the basis of the gradients obtained in theorem 1 of van Ackooij et al. (2010c). The results (van Ackooij et al. (2010b)), which follow from a straight-forward second application of theorem 1 in van Ackooij et al. (2010c) are collected in the following lemma.

Lemma 0.7. *Let ξ be a Gaussian random vector with mean μ and variance-covariance matrix Σ. We define the mapping $F_\xi(a,b) = \mathbb{P}[a \leq \xi \leq b]$ for any rectangle, i.e., $a \leq b$. Let D_n^i denote the n dimensional identity matrix from which the ith row has been deleted. Define $\mu^{c(i,z)} = D_n^i(\mu + \Sigma_{i,i}^{-1}(z-\mu_i)\Sigma_i)$ and $\Sigma^{c(i)} = D_n^i(\Sigma - \Sigma_{i,i}^{-1}\Sigma_i\Sigma_i^\mathsf{T})(D_n^i)^\mathsf{T}$, where Σ_i is the ith column of Σ. We define $\xi^{c(i,z)}$ as the Gaussian random variable with mean $\mu^{c(i,z)}$ and covariance matrix $\Sigma^{c(i)}$. The following holds:*

$$\begin{aligned}
\frac{\partial^2}{\partial a_j \partial a_i} F_\xi(a,b) &= f_{\mu^{c(i,a_i)},\Sigma_{j,j}^{c(i)}}(a_j) f_{\mu_i,\Sigma_{i,i}}(a_i) F_{(\xi^{c(i,a_i)})^{c(j,a_j)}}(D_{n-1}^j D_n^i a, D_{n-1}^j D_n^i b) \ \forall j \neq i \\
\frac{\partial^2}{\partial b_j \partial a_i} F_\xi(a,b) &= -f_{\mu^{c(i,a_i)},\Sigma_{j,j}^{c(i)}}(b_j) f_{\mu_i,\Sigma_{i,i}}(a_i) F_{(\xi^{c(i,a_i)})^{c(j,b_j)}}(D_{n-1}^j D_n^i a, D_{n-1}^j D_n^i b) \ \forall i,j \\
\frac{\partial^2}{\partial b_j \partial b_i} F_\xi(a,b) &= f_{\mu^{c(i,b_i)},\Sigma_{j,j}^{c(i)}}(b_j) f_{\mu_i,\Sigma_{i,i}}(b_i) F_{(\xi^{c(i,b_i)})^{c(j,b_j)}}(D_{n-1}^j D_n^i a, D_{n-1}^j D_n^i b) \ \forall j \neq i,
\end{aligned}$$

where $f_{\mu,\sigma}(x)$ is the standard gaussian density. Moreover, whenever $j = i$ and z is a or b we have:

$$\begin{aligned}
\frac{\partial^2}{\partial z_i^2} F_\xi(a,b) &= -\frac{z_i - \mu_i}{\Sigma_{i,i}^2} f_{\mu_i,\Sigma_{i,i}}(z_i) F_{\xi^{c(i,z_i)}}(D_n^i a, D_n^i b) \\
&+ f_{\mu_i,\Sigma_{i,i}}(z_i)(D_n^i \Sigma_{i,i}^{-1}\Sigma_i)^\mathsf{T}(\nabla_{D_n^i a} F_{\xi^{c(i,z_i)}}(D_n^i a, D_n^i b) + \nabla_{D_n^i b} F_{\xi^{c(i,z_i)}}(D_n^i a, D_n^i b))
\end{aligned}$$

The following corollary follows trivially from lemma 0.7 and theorem 1 of van Ackooij et al. (2010c).

Corollary 0.8. *Let ξ be a Gaussian Random variable of dimension n. Let x, A,B,a,b be vectors and matrices of appropriate dimension. Define furthermore, $\alpha = Ax + a$ and $\beta = Bx + b$. Now consider the mapping $\varphi : x \mapsto \mathbb{P}[a + Ax \leq \xi \leq Bx + b]$. We have:*

$$\begin{aligned}
\nabla\varphi &= \nabla_\alpha F_\xi(\alpha,\beta)^\mathsf{T} A + \nabla_\beta F_\xi(\alpha,\beta)^\mathsf{T} B \\
\triangle\varphi &= A^\mathsf{T} \triangle_{\alpha\alpha} F_\xi(\alpha,\beta) A + A^\mathsf{T} \triangle_{\alpha\beta} F_\xi(\alpha,\beta) B + B^\mathsf{T} \triangle_{\beta\alpha} F_\xi(\alpha,\beta) A + B^\mathsf{T} \triangle_{\beta\beta} F_\xi(\alpha,\beta) B.
\end{aligned}$$

6. Illustration : Feasibility of CCP for EMOP

In this section we will consider the hydro reservoir management example from van Ackooij et al. (2010b). We will consider a discretized time horizon. To this end let τ denote the set of (homogeneous) time steps. Let Δt be this time step size expressed in hours.

6.1 Topology

A hydro valley can be seen as a set of connected reservoirs. We can therefore represent this with a directed graph. Let $\mathcal{N}$ be the set of nodes and let A (of size $|\mathcal{N}| \times |\mathcal{N}|$) be the connection matrix, i.e., $A_{n,m} = 1$ whenever water released from reservoir n will flow into reservoir m. We will assume that D is the flow duration matrix, i.e., D_m is the amount of time (measured in time steps) it takes for water to flow from reservoir m to its child. Let $\mathcal{T} := \left\{ g^i \ , i = 1, ..., N_\mathcal{T} \right\}$ denote the set of turbines and $\mathcal{P} := \left\{ p^i, i = 1, ..., N_\mathcal{P} \right\}$ denote the set of pumping stations. We furthermore introduce the mapping $\sigma_\mathcal{T} : \{1, ..., N_\mathcal{T}\} \to \mathcal{N}$ ($\sigma_\mathcal{P} : \{1, ..., N_\mathcal{P}\} \to \mathcal{N}$) attributing to each turbine (pumping station) the reservoir number to which it belongs. We will also

introduce the sets $\mathcal{A}(n) = \{m \in \mathcal{N} : A_{m,n} = 1\}$ and $\mathcal{F}(n) = \{m \in \mathcal{N} : A_{n,m} = 1\}$. The set $\mathcal{A}(n)$ is empty for top reservoirs and the set $\mathcal{F}(n)$ for bottom reservoirs.

6.2 Controls

We will assume that each turbine (and pumping station) can be controlled for each time step. To this end we introduce the variables $x^i(t)$ for each $t \in \tau$ and $i = 1, ..., N_{\mathcal{T}}$. In a similar way we introduce the variables $y^i(t)$ for the pumping stations. The units are in m^3/h. Furthermore we assume that each of these variables are bounded from below by zero and from above by $\overline{x}^i$ ($\overline{y}^i$ respectively).

6.3 Water values

Let $\pi_n(V)$ be a given discretization of the water levels of reservoir n, i.e., $\pi_n(V) = \left\{V_0^n = V_{min}^n, ..., V_{K_n}^n = V_{max}^n\right\}$. We assume that a water value $W_i^n(t)$ (in €$/m^3$) is attributed to each interval $[V_{i-1}^n, V_i^n)$, $i = 1, ..., K_n$. We introduce two real variables $z_{x,i}^n(t)$ and $\gamma_{x,i}^n(t)$ for each time step $t \in \tau$, each $i = 1, ..., K_n$ and for each reservoir. We similarly introduce $z_{y,i}^n(t)$ and $\gamma_{y,i}^n(t)$ for turbining. In fact $z_i^n(t)$ represents the part of the water turbined ($z_{x,i}^n(t)$) / pumped ($z_{y,i}^n(t)$)that falls in the interval $[V_{i-1}^n, V_i^n)$. We impose the following constraints for each $n \in \mathcal{N}$ and $t \in \tau$:

$$\begin{aligned}
\sum_{i=1}^{K_n} z_{x,i}^n(t) = \Delta t \sum_{j \in \sigma_{\mathcal{T}}^{-1}[n]} x^j(t)\,,\ \sum_{i=1}^{K_n} z_{y,i}^n(t) &= \Delta t \sum_{j \in \sigma_{\mathcal{P}}^{-1}[n]} y^j(t) \\
(z_{x,i}^n(t) - \mathbb{E}\left(V^n(t)\right) - V_{i-1}^n + \gamma_{x,i}^n(t)) z_{x,i}^n(t) &\leq 0 \ \forall i = 1, ..., K_n \\
(z_{y,i}^n(t) - V_i^n + \mathbb{E}\left(V^n(t)\right) + \gamma_{y,i}^n(t)) z_{y,i}^n(t) &\leq 0 \ \forall i = 1, ..., K_n \\
0 \leq z_{u,i}^n(t) &\leq (V_i^n - V_{i-1}^n) \ \forall i = 1, ..., K_n \ u \in \{x, y\} \quad (31) \\
\gamma_{u,i}^n(t) &\geq 0 \ \forall \ \forall i = 1, ..., K_n \ u \in \{x, y\}
\end{aligned}$$

In fact $z_{x,i}^n(t)$ represents the part of the water turbined that falls in the interval $[V_{i-1}^n, V_i^n)$. A natural constraint is $z_{x,i}^n(t) \leq \max(V^n(t) - V_{i-1}^n, 0)$. However, in our example, $V^n(t)$ is random. Fortunately, when combining this with an objective function that we wish to optimize in expectation, the constraint becomes $z_{x,i}^n(t) \leq \max(\mathbb{E}\left(V^n(t)\right) - V_{i-1}^n, 0)$, hence erasing randomness from the objective function. The quadratic constraints arise as it is easily seen that the following problems are equivalent $\min_x \left\{f(x) : g(x) \leq [h(x)]^+\right\}$ and $\min_{x,\lambda \geq 0} \{f(x) : (g(x) - h(x) + \lambda) g(x) \leq 0\}$. In our numerical example (Section 6.7) we use a constant watervalue, removing the quadratic constraints.

6.4 Random inflows

We will assume that inflows (in m^3/h) in reservoirs are the result of some stochastic process. Let $A^n(t)$ denote this stochastic process for reservoir n. Not all reservoirs will have stochastic inflows, some of them will have deterministic inflows (typically zero). This can be explained by the fact that top reservoirs have random inflows due to the melting of snow in the high mountains, whereas rain can be neglected for lower reservoirs. Let $\mathcal{N}^r \subseteq \mathcal{N}$ denote the set of reservoirs receiving random inflows. We will assume that the stochastic inflow process is the sum of a deterministic trend s_t^n and a causal process (Shumway & Stoffer (2000)) generated by Gaussian innovations. To this end let $\zeta^n(t)$ be a gaussian white noise process, where

$(\zeta^{k_1}(t),...,\zeta^{k_l})$ is a Gaussian random vector of zero average and variance-covariance matrix $\Sigma(t)$ ($\{k_1,...,k_l\} = \mathcal{N}^r$). We will assume independence between time steps of the ζ vector. Since $A^n(t)$ is a causal process, we can write it as follows

$$A^n(t) = s_t^n + \sum_{j=0}^{\infty} \psi_j^n \zeta^n(t-j) = s_t^n + \sum_{j=t}^{\infty} \psi_j^n \zeta^n(t-j) + \sum_{j=0}^{t-1} \psi_j^n \zeta^n(t-j),$$

for some coefficient vector ψ^n. We will assume that randomness before $t = 0$ is known and as such we can assume WLOG that the random inflow process can be written as

$$A^n(t) = s_t^n + \sum_{j=0}^{t-1} \psi_j^n \zeta^n(t-j).$$

6.5 Flow constraints and volume bounds

Each reservoir is subject to flow constraints induced by pumping and turbining. The following equilibrium constraint applies

$$\begin{aligned} V_n(t) &= V_n(t-1) + \sum_{m\in\mathcal{A}(n)} \sum_{i\in\sigma_{\mathcal{T}}^{-1}[m]} x^i(t-D_m)\Delta t - \sum_{i\in\sigma_{\mathcal{T}}^{-1}[n]} x^i(t)\Delta t \\ &+ \sum_{m\in\mathcal{F}(n)} \sum_{i\in\sigma_{\mathcal{P}}^{-1}[m]} y^i(t)\Delta t - \sum_{i\in\sigma_{\mathcal{P}}^{-1}[n]} y^i(t)\Delta t + s_t^n\Delta t + \sum_{j=0}^{t-1} \psi_j^n \zeta^n(t-j)\Delta t. \end{aligned} \tag{32}$$

The above equation is entirely deterministic except for the reservoirs $n \in \mathcal{N}^r$. In order to deal with this randomness and reservoir bounds we will therefore add the following constraints

$$\mathbb{P}[V_{min}^n(t) \le V^n(t) \le V_{max}^n(t)\ \forall t \in \tau, n \in \mathcal{N}^r] \ge p \tag{33}$$

$$V_{min}^n(t) \le V^n(t) \le V_{max}^n(t)\ \forall t \in \tau, n \in \mathcal{N} \setminus \mathcal{N}^r, \tag{34}$$

this is a joint chance constraint.

6.6 Objective function

Often, in reality, each reservoir only has a single turbine. The power output of turbining x m^3/s is given by a function $\rho(x)$. This function is strictly increasing and concave, i.e., $\rho'(x) \ge 0$ and $\rho''(x) \le 0$. In our model we have split this range into several subsections (hence several turbines), each with efficiency $\rho_i = \rho'(s_i^*)/3600$ (MWh/m^3) for some s_i^* in each section. We can thus remark that for any two turbines i_1 and i_2 belonging to the same reservoir we either have $\rho_{i_1} \ge \rho_{i_2}$ or vice versa. This approximation comes down to approximating $\rho(x)$ by a piece-wise linear function.

We assume given a time dependent price signal $\lambda(t)$ (in €/MWh). The following objective function has to be minimized:

$$\sum_{t\in\tau} \sum_{n\in\mathcal{N}} \sum_{i=1}^{K_n} (W_i^n(t)(z_{x,i}^n(t) - z_{y,i}^n(t)) - \sum_{t\in\tau} \lambda(t)\Delta t(\sum_{i=1}^{N_{\mathcal{T}}} \rho_i(t)x^i(t) - \sum_{i=1}^{N_{\mathcal{P}}} \frac{1}{\theta_i(t)} y^i(t)),$$

where the first part corresponds to the cost of using water expressed by the water-values, and $\theta^i(t)$ is the efficiency of pumping.

6.7 Numerical example

Plugging some numerical values in the problem defined in this section 6. We can consider for example 24 time steps of 2 hours each, the valley 2 (Left) and $AR(3)$ uncertainty on inflows. More importantly that the actual numerical values (which can be found in van Ackooij et al. (2010b)), is a comparison of the individual chance constrained model (5) and the joint constrained model (4).

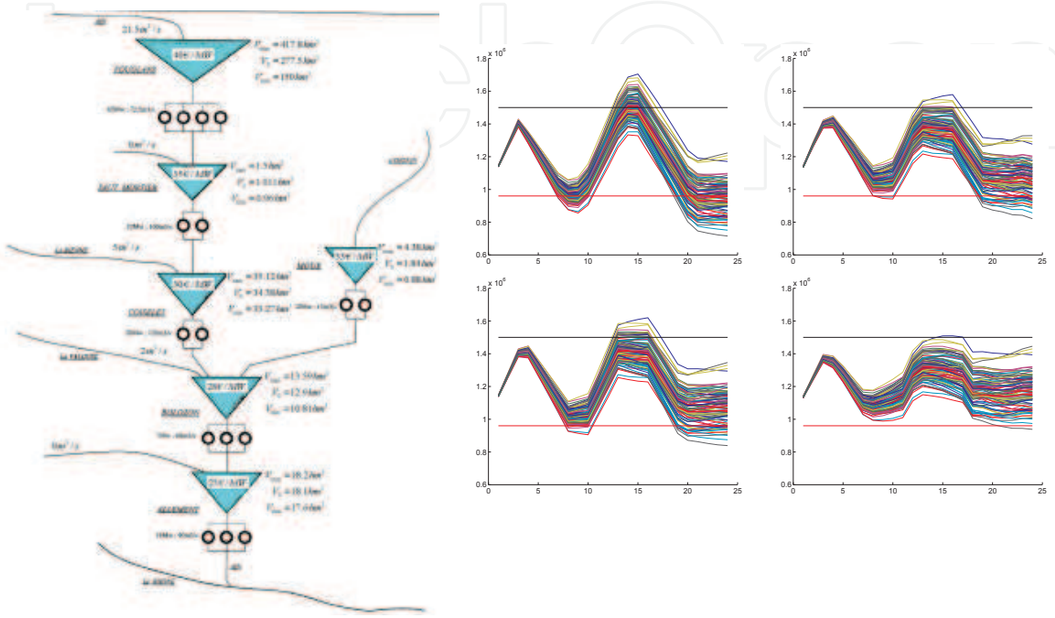


Fig. 2. (Left) The hydro Valley. (Right) : Water trajectories for reservoir "Saut Mortier". From top left to bottom right, solutions of problems (11), (4), (5) and (28).

Table 1 shows optimal costs and number of violations. Figure 2 shows simulations of water trajectories. Clearly we observe the advantage of using joint chance constrained programming. The additional cost with respect to the deterministic solution is only small, but robustness can be fine tuned. A full robust solution (max-p problem) turns out quite costly. Finally individual chance constrained programming can not be used to mimic joint chance constraints as we have no control over the number of violations over a period of time.

Item / Problem	(11)	(4)	(5)	(28)
nbViolation	100	20	35	2
Cost (€)	$-1.0478e^5$	$-1.0340e^5$	$-1.0422e^5$	$-9.9176e^4$

Table 1. Comparison of costs and number of violations

One can come up with a robust counterpart of problem (10), by defining an elipsoidal uncertainty set $\mathcal{E}$ for η. It can be easily seen that constraints (18) (derived from (33)) can be transformed in $Ax + a \leq \inf \mathcal{E}$ and $Bx + b \geq \sup \mathcal{E}$, where the latter has to be understood in the partial order of $\mathbb{R}^n$. Unfortunately, even when the uncertainty set $\mathcal{E}$ is very well calibrated, i.e., $\mathbb{P}(\mathcal{E}) = p$, the solution is often over-robust. Even worse, for larger values of p this often leads to an empty feasible set of the robust problem, even though solutions of (4) exist.
We can observe that the speed of Genz' code is not independent of the "nature" of a and b (see Lemma 0.7). The "shape" of the covariance matrix of ξ is pointed downwards. It seems that

whenever a and b mimic this shape, i.e., $a_1 \leq ... \leq a_n$, that evaluating $F_{\tilde{\xi}}(a, b)$ is about 20 times faster than having a uniform a and b.
Since this valley is a realistic example from Energy management exact joint-CCP can be tractable for problems. Moreover clearly the interest has been shown over an individual chance constraint formulation.

7. Perspectives / transgressing frontiers

Perspectives contain three main axis: The first axis is concerned with improved bounds for approximate chance constrained programming. Currently we have used Hoeffding's bound, but far better bounds exists. One could think of the bounds derived in Ben-Tal et al. (2009). By combining different bounding techniques and different levels of available information we can derive a whole class of approximate algorithms, much in the style of the MIO algorithm exposed here. A second important question to answer is that of classification of the solution. Is the approximate solution far from the optimal one?
A second axis is concerned with working on exact joint chance constraint programming for the separated linear setting. In particular efficient derivative formulae have to be derived for the case of a random matrix. Further clear extensions concern such questions for the case of other laws. Often laws in a problem are of a different nature and such special cases have to be considered. From an algorithmic perspective, instead of using a cutting planes idea, one could use a bundle method to hopefully improve computation times and stability. A second point that needs investigations is an improved use of Genz' code by using preconditioning and exploiting the observed shaping/computation time effect. Finally we could combine the use of Genz' code with Prekopa's LP estimation method for probability measures to increase the size of the model or improve the speed.
A third axis consists of considering the mixed integer formulation of (10). If the relaxed problem has good properties (convexity, etc..). We could, in theory apply a branch and bound technique combined with cuts, lift&projects, etc... But one could equally consider this a special case of discrete randomness.

8. References

Alizadeh, F. & Goldfarb, D. (2001). Second-order cone programming, *Technical Report RRR 51-2001, Available at http://rutcor.rutgers.edu/pub/rrr/reports2001/51.ps* .

Appariagliato, R., Vial, J. & Zorgati, R. (2006). Weekly management of a hydraulic valley by robust optimization., *19th international symposium on Mathematical programming, Rio de Janeiro* .

Batut, J. & Renaud, A. (1992). Daily scheduling with transmission constraints: A new class of algorithms, *IEEE Transactions on Power Systems* 7(3): 982–989.

Ben-Tal, A., Ghaoui, L. E. & Nemirovski, A. (2009). *Robust Optimization*, Princeton University Press.

Beraldi, P. & Ruszczyński, A. (2002). A branch and bound method for stochastic integer problems under probabilistic constraints., *Optimization Methods and Software* 17: 359–382.

Birge, J. & Louveaux, F. (1997). *Introduction to Stochastic Programming*, Springer, New York.

Bukszár, J., R. Henrion, M. H. & Szántai, T. (2004). Polyhedral inclusion-exclusion, *Preprint No. 913, Weierstrass Institute Berlin* .

Bukszár, J. & Szántai, T. (2002). Probability bounds given by hypercherry trees, *Optimization Methods and Software* 17: 409–422.

Cohen, G. & Zhu, D. (1983). Decomposition-coordination methods in large-scale optimization problems. the non-differentiable case and the use of augmented Lagrangians, *Large Scale Systems, Theory and Applications* 1.

Déak, I. (1980). Three digit accurate multiple normal probabilities, *Numerische Mathematik* 35: 369–380.

Deák, I. (1986). Computing probabilities of rectangles in case of multidimensional distribution., *Journal of Statistical Computation and Simulation* 26: 101–114.

Dentcheva, D., Prékopa, A. & Ruszczyński, A. (2000). Concavity and efficient points for discrete distributions in stochastic programming, *Mathematical Programming* 89: 55–79.

Genz, A. (1992). Numerical computation of multivariate normal probabilities, *J. Comp. Graph Stat.* 1: 141–149.

Genz, A. (2002). Methods for the computation of multivariate t-probabilities, *J. Comp. Graph. Stat.* 11: 950–971.

Genz, A. & Bretz, F. (2009). *Computation of multivariate normal and t probabilities.*, number 195 in *Lecture Notes in Statistics*, Springer, Dordrecht.

Genz, A. & Kwong, K.-S. (2000). Numerical evaluation of singular multivariate normal distributions, *J. Stat. Comp. Simul.* 68: 1–21.

Gouda, A. & Szántai, T. (2004). New sampling techniques for calculation of dirichlet probabilities, *Central European Journal of Operations Research* 12: 389–403.

Henrion, R. (2007). Structural properties of linear probabilistic constraints, *Optimization* 56(4): 425–440.

Henrion, R., Andrieu, L. & Römisch, W. (2010). A model for dynamic chance constraints in hydro power management, *European Journal of Operations Research* 207: 579–589.

Henrion, R. & Römisch, W. (2010). Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions, *Annals of Operations Research* 177: 115–125.

Henrion, R. & Strugarek, C. (2008). Convexity of chance constraints with independent random variables, *Computational Optimization and Applications* 41: 263–276.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical society* 58(301): 13–30.

Kall, P. & Wallace, S. (1994). *Stochastic Programming*, Wiley, New York.

Kataoka, S. (1963). A stochastic programming model, *Econometrica* 31: 181–196.

Klopfenstein, O. (2007). Tractable algorithms for chance-constrained combinatorial problems, *www.optimization-online.org* .

Klopfenstein, O. & Nace, D. (2007). Polyhedral aspects of a robust knapsack problem, *www.optimization-online.org* .

Klopfenstein, O. & Nace, D. (2008). A robust approach to the chance-constrained knapsack problem, *Operations Research Letters* 36: 628–632.

Lagoa, C., Li, X. & Sznaier, M. (2005). Probabilistically constrained linear programs and risk-adjusted controller design, *SIAM Journal on Control and Optimization* 15: 938–951.

Langrene, N., van Ackooij, W. & Bréant, F. (2010). Dynamic constraints for aggregated units: Formulation and application, *To Appear : IEEE transactions on Power Systems* p. 8.

Lemaréchal, C. & Sagastizábal, C. (1994). An approach to variable metric bundle methods, *Lecture Notes in Control and Information Science* 197: 144–162.

Lepaul, S. (2009). Stochastic approach for an endogenous capacity expansion long-term world coal market model, 32nd *IAEE International Conference, Energy, Economy, Environment: The Global View (June 21-24)* .

Lobo, M., Vandenberghe, L., Boyd, S. & Lebret, H. (1998). Applications of second-order cone programming, *Linear Algebra and its Applications* 284: 193–228.

Luedtke, J. & Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints, *SIAM Journal on Optimization* 19: 674–699.

Merlin, A. & Sandrin, P. (1983). A new method for unit commitment at electricité de france, *IEEE Transactions Power App. Syst.* PAS-102: 1218–1225.

Olieman, N. & van Putten, B. (2006). Estimation method of multivariate exponential probabilities based on a simplex coordinates transform, *Stochastic Programming E-Print Series (SPEPS)* 6.

Prékopa, A. (1973). On logarithmic concave measures and functions, *Acta Scientiarium Mathematicarum (Szeged)* 34: 335–343.

Prékopa, A. (1990). Sharp bound on probabilities using linear programming., *Operations Research* 38: 227–239.

Prékopa, A. (1995). *Stochastic Programming*, Kluwer, Dordrecht.

Prékopa, A. (2003). *Probabilistic programming. In Ruszczyński & Shapiro (2003) (Chapter 5)*.

Prékopa, A., Vizvári, B. & Badics, T. (1998). Programming under probabilistic constraint with discrete random variable., *In (F. Giannessi et al. eds.): New Trends in Mathematical Programming* pp. 235–255.

Ruszczyński, A. (2002). Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra., *Mathematical Programming* 93: 195–215.

Ruszczyński, A. & Shapiro, A. (2003). *Stochastic Programming*, Vol. 10 of *Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam.

Shapiro, A., Dentcheva, D. & Ruszczyński, A. (2009). *Lectures on Stochastic Programming. Modeling and Theory*, Vol. 9 of *MPS-SIAM series on optimization*, SIAM and MPS, Philadelphia.

Shumway, R. & Stoffer, D. (2000). *Time Series Analysis and Its Applications*, 1st edn, Springer.

Szántai, T. (1996). Evaluation of a special multivariate gamma distribution, *Mathematical Programming Study* 27: 1–16.

Szántai, T. (2000). Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function, *Annals of Operations Research* 100: 85–101.

Szántai, T. & Habib, A. (1998). On the *k*-out-of-*r*-from-*n*:probabilities., *In (F. Giannessi et al. eds.) New Trends in Mathematical Programming* 36: 289–303.

Uryasev, S. (1995). Derivatives of probability functions and some applications, *Annals of Operations Research* 56: 287–311.

van Ackooij, W., Henrion, R., Möller, A. & Zorgati, R. (2010a). Early evaluation of chance-constrained programming for energy management optimization problems, *Technical Report : H-R36-2010-00447-EN* p. 103.

van Ackooij, W., Henrion, R., Möller, A. & Zorgati, R. (2010b). Joint chance constrained programming for hydro reservoir management, *Submitted : EngOpt2010, 2nd International Conference on Engineering optimization* .

van Ackooij, W., Henrion, R., Möller, A. & Zorgati, R. (2010c). On probabilistic constraints induced by rectangular sets and multivariate normal distributions, *Mathematical Methods of Operations Research* 71(3): 535–549.

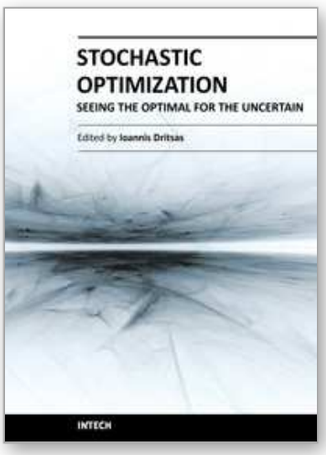
van de Panne, C. & Popp, W. (1963). Minimum-cost cattle feed under probabilistic protein constraints, *Managment Science* 9: 405–430.

Zorgati, R. & van Ackooij, W. (2008). Optimizing financial and physical assets with chance-constrained programming in the electrical industry, *EngOpt2008, International Conference on Engineering optimization* .

Zorgati, R. & van Ackooij, W. (2010). Optimizing financial and physical assets with chance-constrained programming in the electrical industry, *to Appear in : Optimization and Engineering* Accepted.

Zorgati, R., van Ackooij, W. & Apparigliato, R. (2009). Supply shortage hedging : estimating the electrical power margin for optimizing financial and physical assets with chance-constrained programming, *IEEE Transactions on Power Systems* 24(2): 533–540.

Zorgati, R., van Ackooij, W. & Gorge, A. (2010). Uncertainties on power systems. probabilistic approach and conic approximation, *PMAPS2010, 11th international Conference on Probabilistic Methods applied to Power systems (Prize Paper Award)* p. 8.



Stochastic Optimization - Seeing the Optimal for the Uncertain
Edited by Dr. Ioannis Dritsas

ISBN 978-953-307-829-8
Hard cover, 476 pages
Publisher InTech
Published online 28, February, 2011
Published in print edition February, 2011

Stochastic Optimization Algorithms have become essential tools in solving a wide range of difficult and critical optimization problems. Such methods are able to find the optimum solution of a problem with uncertain elements or to algorithmically incorporate uncertainty to solve a deterministic problem. They even succeed in â€œfighting uncertainty with uncertaintyâ€. This book discusses theoretical aspects of many such algorithms and covers their application in various scientific fields.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Wim van Ackooij, Riadh Zorgati, René Henrion and Andris Möller (2011). Chance Constrained Programming and Its Applications to Energy Management, Stochastic Optimization - Seeing the Optimal for the Uncertain, Dr. Ioannis Dritsas (Ed.), ISBN: 978-953-307-829-8, InTech, Available from: http://www.intechopen.com/books/stochastic-optimization-seeing-the-optimal-for-the-uncertain/chance-constrained-programming-and-its-applications-to-energy-management

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen