Mathematical Programming manuscript No.
(will be inserted by the editor)

Josselin Garnier · Abdennebi Omrane · Youssef Rouchdy

Efficient Monte Carlo simulations for stochastic programming

Received: date / Accepted: date

Abstract This paper is concerned with chance constrained programming to deal with nonlinear optimization problems with random parameters. Specific Monte Carlo methods to evaluate the gradient and Hessian of probabilistic constraints are proposed and discussed. These methods are implemented in penalization optimization routines adapted to stochastic optimization. They are shown to reduce the computational cost of chance constrained programming substantially.

Keywords stochastic programming · optimization with constraints · random constraints · Monte Carlo methods

Mathematics Subject Classification (2000) 90C15 · 65C05

1 Introduction

The presence of uncertainty in optimization problems creates a need for new tools dealing with stochastic programming. Stochastic programming is based on advanced mathematical tools such as non smooth calculus, abstract optimization, probability theory and statistical techniques, including Monte Carlo (MC) simulations [12, 7]. It addresses a variety of models where random data are present in optimization problems encountered in engineering

J. Garnier
Laboratoire de Probabilités et Modèles Aléatoires & Laboratoire Jacques-Louis Lions, Université Paris VII, 2 Place Jussieu 75251 Paris Cedex 5, France E-mail: garnier@math.jussieu.fr

A. Omrane
Université des Antilles et de la Guyane, UFR Sciences Exactes et Naturelles, DMI, 97159 Pointe à Pitre, Guadeloupe, France E-mail: aomrane@univ-ag.fr

Y. Rouchdy
Institut Camille Jordan, INSA Lyon, 20 avenue A. Einstein, 69621 Villeurbanne Cedex, France E-mail: youssef.rouchdy@insa-lyon.fr

or finance: chance-constrained models, transportation and logistics models, economics, or models involving risk measures (see the books by Prékopa [15], Ruszczyński & Shapiro [17], and Kall & Wallace [13] and the survey articles by Birge [4] and Sen & Higle [18]).

It is possible to distinguish two types of optimization problems in presence of randomness [19]. In the first type, the cost function is random, or it is the expectation of a random cost function, and the set of feasible solutions is deterministic [1]. In this paper we address a second type of random optimization problems. Here, a deterministic cost function $J(x)$ has to be minimized under a set of N_c random constraints

$$g_p(x, \Lambda) \leq c_p\,, \quad p = 1, ..., N_c, \tag{1}$$

where $g_p : \mathbb{R}^{N_x} \times \mathbb{R}^{N_\Lambda} \to \mathbb{R}$ is a real-valued constraints mapping and $c_p \in \mathbb{R}$ is the constraint level, for each $p = 1, ..., N_c$,

$$x = (x_1, \ldots, x_{N_x}) \quad \in \quad \mathbb{R}^{N_x} \tag{2}$$

is the vector of physical parameters, and

$$\Lambda = (\Lambda_1, \ldots, \Lambda_{N_\Lambda}) \quad \in \quad \mathbb{R}^{N_\Lambda} \tag{3}$$

is a continuous random vector modeling the uncertainty of the model. In many situations, the set of points x such that the N_c constraints (1) are satisfied for (almost) all realizations of the random vector Λ is empty or quasi-empty. Indeed, even if constraint violation at x happens only for a very unlikely subset of realizations of Λ, the point x has to be excluded. It then makes sense to look for an admissible set of points x that satisfy the constraints with high probability $1 - \alpha$, $0 < \alpha \ll 1$, in the sense that only a small proportion α of realizations of the random vector Λ leads to constraint violation at x. The set of constraints (1) is then substituted by the probabilistic constraint

$$P(x) \geq 1 - \alpha, \tag{4}$$

with

$$P(x) = \mathbb{P}\Big(g_p(x, \Lambda) \leq c_p\,, \quad p = 1, ..., N_c \Big). \tag{5}$$

The probabilistic constraint prescribes a lower bound for the probability of simultaneous satisfaction of the N_c random constraints. Situations where probabilistic constraints of the form (4-5) are encountered are listed in the book by Prékopa [15] and in the papers by Henrion & Romisch [10], Henrion & Strugarek [11], and Prékopa [14].

The convexity of the admissible set, i.e. of the set of points x satisfying (4), is a crucial point for the optimization problem [6]. We shall review this question in this paper. In particular we shall present the recent contribution [11], where r-concavity properties of the functions g_p together with a r-decreasing assumption of the density of Λ are used to investigate the convexity of the admissible set. The second crucial issue is that optimization problems under probabilistic constraints of the form (4-5) are particularly

difficult to solve for continuous random vectors, such as multivariate normal, because the probability (5) has the form of a high-dimensional integral that cannot be evaluated directly, but has to be approximated by numerical quadrature or by MC simulations. Because MC simulation appears to offer the best possibilities for higher dimensions, it seems to be the natural choice for use in stochastic programs. In [5] Birge and Louveaux describe some basic approaches built on sampling methods. They also consider some extentions of the MC method to include analytical evaluations exploiting problem structure in probabilistic constraints estimation. However, the evaluation of the gradient and/or Hessian of (5) is required in nonlinear optimization routines, and this is not compatible with the random fluctuations of MC simulations. We propose a specific MC method to evaluate the gradient and Hessian of the probabilistic constraint and we implement this method in penalization optimization routines adapted to stochastic optimization. The proposed MC estimators are based on probabilistic representations of the partial derivatives of the probabilistic constraint, and they are shown to reduce the computational cost of chance constrained programming.

The paper is organized as follows. In Section 2 we apply standard convex analysis results to study the convexity of the admissible set. Section 3 is devoted to numerical illustrations of convex and non convex examples. In Sections 4-6 we show how to construct MC estimators of the gradient and of the Hessian of the probabilistic constraint. We finally implement these methods in two simple examples.

2 Structural properties of the constraints set

In this section, we investigate the convexity of the admissible set

$$\mathcal{A}_{1-\alpha} = \Big\{ x \in \mathbb{R}^{N_x} \ : \ P(x) \geq 1-\alpha \Big\}, \tag{6}$$

where $\alpha \in (0,1)$. Let us first recall some definitions and results.

Definition 1 A non-negative real-valued function f is said to be log-concave if $\ln(f)$ is a concave function. Since the logarithm is a strictly concave function, any concave function is also log-concave.

Equivalently, a function $f : \mathbb{R}^{N_x} \to [0,\infty)$ is log-concave if

$$f(s\,x + (1-s)y) \geq (f(x))^s (f(y))^{1-s}, \qquad \text{for any } \ x, y \in \mathbb{R}^{N_x},\ s \in [0,1].$$

A random variable is said to be log-concave if its density function is log-concave. The uniform, normal, beta, exponential, and extreme value distributions have this property [2].
The log-concavity of the density of a real-valued random variable implies the log-concavity of the distribution function (Theorem 4.2.4 [15]). As an example, the standard normal distribution $\mathcal{N}(0,1)$ with density $\phi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ is log-concave, since $(\ln\phi(x))'' = -1 < 0$. Therefore, the normal distribution has a log-concave distribution function.

If $F(x) = \mathbb{P}(X \leq x)$ is log-concave and if $p_0 \in (0,1)$ is fixed, then the set of x such that $\mathbb{P}(X \leq x) \geq p_o$ is convex. Indeed, if x, y are such that $\mathbb{P}(X \leq x) \geq p_o$ and $\mathbb{P}(X \leq y) \geq p_o$, then for any $s \in [0,1]$,

$$\mathbb{P}(X \leq sx + (1-s)y) \geq \mathbb{P}(X \leq y)^s \mathbb{P}(X \leq x)^{1-s} \geq p_o^s p_o^{1-s} = p_o.$$

Definition 2 A function $f : \mathbb{R}^{N_x} \to \mathbb{R}$ is said to be quasi-convex, resp. quasi-concave, if for any $x, y \in \mathbb{R}^{N_x}$ and $s \in [0,1]$

$$f(sx + (1-s)y) \leq \max(f(x), f(y)) \quad \text{resp. } \geq \min(f(x), f(y)).$$

Note that convexity implies quasi-convexity. We now give an important result which follows directly from Theorem 10.2.1 by Prékopa [15]:

Proposition 1 *Let $g_1(x, \Lambda), g_2(x, \Lambda), ..., g_{N_c}(x, \Lambda)$ be quasi-convex functions in (x, Λ). If the random vector Λ has a log-concave density, then P is log-concave and the admissible set $\mathcal{A}_{1-\alpha}$ defined by (6) is convex for any $\alpha \in (0,1)$.*

In this section, we shall focus our attention to the case where $N_c = N_\Lambda$ and the constraint function $g = (g_p)_{p=1,...,N_c}$ is of the form

$$g_p(x, \Lambda) = f_p(x) - \Lambda_p.$$

If, additionnally, we suppose that the Λ_p, $p = 1, ..., N_c$, are independent random Gaussian variables with mean μ_p and variance σ_p^2, then the distribution function of Λ_p is:

$$F_p(x) = \Phi\left(\frac{x - \mu_p}{\sigma_p}\right), \tag{7}$$

where Φ is the distribution function of the $\mathcal{N}(0,1)$ distribution. As the Λ_p are independent, we get the following :

$$P(x) = \mathbb{P}\Big(f_p(x) \leq \Lambda_p \,,\, p = 1, ..., N_c\Big) = \prod_{p=1}^{N_c} \Phi\left(\frac{\mu_p - f_p(x)}{\sigma_p}\right). \tag{8}$$

By Proposition 1 we obtain the following result : If $f_1(x), f_2(x), ..., f_{N_c}(x)$ are quasi-convex functions then the admissible set $\mathcal{A}_{1-\alpha}$ defined by (6) is convex.

In the linear case the functions f_p are of the form

$$f_p(x) = \sum_{j=1}^{N_x} A_{pj} x_j \,, \qquad p = 1, ..., N_c, \tag{9}$$

where A is a $N_x \times N_c$ matrix. Therefore, the hypotheses of Proposition 1 are fulfilled and $\mathcal{A}_{1-\alpha}$ is convex.

In the general case where the functions f_p are nonlinear in the x-variable, the general theory is based on the Karush-Kuhn-Tucker Theorem:

Theorem 1 (KKT Theorem) *Suppose that J and $G_k(x)$, $k = 1, ..., K$, are all convex functions. Then under very mild conditions, $\bar{x}$ solves*

$$\min_{x \in \mathbb{R}^{N_x}} J(x) \text{ s. t. } G_k(x) \leq \ell_k\,,\ k = 1, ..., K \tag{10}$$

if and only if there exists $\mu \in [0, \infty)^K$ such that
(i) $\nabla J(\bar{x}) + \mu \cdot \nabla G(\bar{x}) = 0$,
(ii) $G_k(\bar{x}) \leq \ell_k, \quad k = 1, ..., K$.
(i) and (ii) are known as the KKT conditions.

The convexity of the admissible set (6) in the nonlinear case corresponds to $K = 1$ and $G_1(x) = -\log(P(x))$, where $P(x)$ is given by (8). If the functions f_p are quasi-convex, then the admissible set $\mathcal{A}_{1-\alpha}$ is convex by Proposition 1, the function $P(x)$ is logconcave, and $G_1(x)$ is convex. Theorem 1 can therefore be applied.

When the problem is non convex (i.e. the cost function J and/or the constraint function G_1 are not convex) we can at least assert that any optimal solution must satisfy the KKT conditions, i.e. under very mild conditions, if $\bar{x}$ solves (10), there exists μ_1 that the KKT conditions (i) and (ii) are fulfilled.

As noted in [11], it is not necessary for the admissible domain $\mathcal{A}_{1-\alpha}$ to be convex for all values of $\alpha \in (0, 1)$. In practical situations the convex property is sufficient if it is satisfied only for small values of α. It turns out that an extension of the previous results based on this remark is possible. We will see in Section 7 configurations where randomness leads to the convexification of the admissible set for small α. We give in the final part of this section the basis of this theory developed by Henrion and Strugareck [11], where the convexity of the admissible set is obtained under more general hypotheses satisfied by r-concave and r-decreasing functions.

Definition 3 A function $f : \mathbb{R}^{N_x} \to [0, +\infty)$ is said to be r-concave for $r \in [-\infty, +\infty]$ if

$$f(sx + (1-s)y) \geq \left[sf(x)^r + (1-s)f(y)^r\right]^{\frac{1}{r}} \tag{11}$$

for all $(x, y) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_x}$, and for any $s \in [0, 1]$.

For $r = 1$, a function f is 1-concave if and only if (iff) f is concave. For $r = -\infty$, a function f is $-\infty$-concave iff f is quasi-concave, i.e. for any $x, y \in \mathbb{R}^{N_x}$ and $s \in [0, 1]$

$$f(sx + (1-s)y) \geq \min(f(x), f(y)).$$

Definition 4 A function $f : \mathbb{R} \to \mathbb{R}$ is said to be r-decreasing for some $r \in \mathbb{R}$ if it is continuous on $(0, +\infty)$ and if there exists some $t^* > 0$ such that the function $t^r f(t)$ is strictly decreasing for $t > t^*$.

For $r = 0$, a function f is 0-decreasing iff it is strictly decreasing. This property can be seen as a generalization of the log-concavity, as shown by the following proposition, which follows from Borell's theorem [6] and is proved in [11].

Proposition 2 *Let $F : \mathbb{R} \to [0,1]$ be a log-concave distribution function with differentiable density ϕ and such that $F(t) < 1$ for all $t \in \mathbb{R}$. Then, ϕ is r-decreasing for all $r > 0$.*

Many classical density functions (including Gaussian, log-normal densities) are r-decreasing. This notion is important, as shown by the following theorem [11].

Theorem 2 *Suppose that there exist $r_p > 0$ such that $-f_p(x)$ are $(-r_p)$-concave functions, and suppose that Λ_p are independent random variables with (r_p+1)-decreasing densities ϕ_p for $p = 1, ..., N_c$. Then*

$$\mathcal{A}_{1-\alpha} = \Big\{ x \in \mathbb{R}^{N_x} \, : \, \mathbb{P}\Big(f_p(x) \le \Lambda_p \, , \, p = 1, ..., N_c \Big) \ge 1 - \alpha \Big\} \tag{12}$$

is a convex subset of $\mathbb{R}^{N_x}$, for all $\alpha < \alpha^* := \max\{F_p(t_p^*) \, , \, p = 1, ..., N_c\}$, where F_p denotes the distribution function of Λ_p and the t_p^* refer to Definition 4 in the context of ϕ_p being (r_p+1)-decreasing.

3 Applications

In this section we plot admissible domains $\mathcal{A}_{1-\alpha}$ of the form (6) in the case where

$$g_p(x, \Lambda) = f_p(x) - \Lambda_p \ , \ \ p = 1, ..., N_c,$$

and $\Lambda = (\Lambda_1, ..., \Lambda_{N_c})$ is a Gaussian vector in $\mathbb{R}^{N_c}$ with covariance matrix Γ and mean zero. In this case it is possible to express analytically the probabilistic constraint. In this section we address cases where the f_p's are linear and nonlinear functions such that, for each realization of the random vector Λ, the set of constraints $f_p(x) - \Lambda_p \le c_p$ defines a bounded, convex or non-convex, domain in $\mathbb{R}^{N_x}$. In these situations, the convexity of the admissible set $\mathcal{A}_{1-\alpha}$ is investigated.

Example 1 (Convex case) Let us consider the set G1 of random constraints

$$f_p(x) - \Lambda_p \le 1 \, , \ \ p = 1, ..., 4, \tag{13}$$

where $f : \mathbb{R}^2 \to \mathbb{R}^4$ is defined by

$$\begin{aligned} f_1(x) &= -x_1 + x_2, \\ f_2(x) &= x_1 + x_2, \\ f_3(x) &= x_1 - x_2, \\ f_4(x) &= -x_1 - x_2, \end{aligned} \tag{14}$$

and Λ_p are independent zero-mean Gaussian random variables with standard deviations σ_p. The set of points x such that $f_p(x) \le 1, p = 1, ..., 4$ is a simple convex polygon (namely, a rectangle). Figure 1 plots the $1-\alpha$ level sets of the probabilistic constraint

$$P(x) = \mathbb{P}\Big(f_p(x) - \Lambda_p \le 1, p = 1, ..., 4 \Big)$$

for different values of α between 0 and 1. The admissible sets in this case are convex, as predicted by the theory.

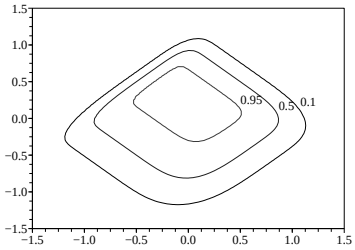

Fig. 1 Level sets of the probabilistic constraint $P(x)$ with the set G1 of random constraints. The random variables Λ_p are independent with standard deviations $\sigma_p = 0.1p, 1 \leq p \leq 4$.

Example 2 (Non-convex cases) We revisit the previous example and replace one edge by a circular curve so that the obtained domain (for a realization of Λ) is not convex. We consider the set of random constraints G2:

$$f_p(x) - \Lambda_p \leq 1\,, \;\; p = 1, ..., 5, \tag{15}$$

where $f : \mathbb{R}^2 \to \mathbb{R}^5$ is defined by

$$\begin{aligned} f_1(x) &= -(x_1+1)^2 - (x_2+1)^2 + 2, \\ f_2(x) &= -x_1 - x_2 - 1, \\ f_3(x) &= -x_1 + x_2, \\ f_4(x) &= x_1 + x_2, \\ f_5(x) &= x_1 - x_2, \end{aligned} \tag{16}$$

and Λ_p are independent zero-mean Gaussian random variables with standard deviations σ_p. As shown in Figure 2 a,c, the level sets of the probabilistic constraint $P(x) = \mathbb{P}\big(f_p(x) - \Lambda_p \leq 1\,,\, p = 1, ..., 5\big)$ are convex for high values of the level $1 - \alpha$. This is in qualitative agreement with the conclusion of Theorem 2: The admissible domain is convex for high levels $1 - \alpha$. Note, however, that the constraint G2 does not fulfill the hypotheses of Therorem 2, which means that it should be possible to extend the validity of this result.

Finally, we consider the set G2 of random constraints and assume that the coordinates of the Gaussian random vector Λ are not independent. More precisely, we consider the case where $\Lambda_p \equiv \Lambda$, $p = 1, ..., 5$, where Λ is a real Gaussian random variable with mean zero and variance σ^2. Even in this case, for high values of the probability level $1 - \alpha$, the level set is convex as seen in Figure 2 b,d, although the convexification is less efficient in the correlated case.

To summarize. The hypotheses of $(-r)$-concavity and independence of the entries of the Gaussian random vector are important in the proof of the convexity result stated in Theorem 2. However, as we have seen in our simulations, the convexification of the admissible domain for large (i.e. close to one) probability levels $1-\alpha$ seems to be valid for a large class of problems.

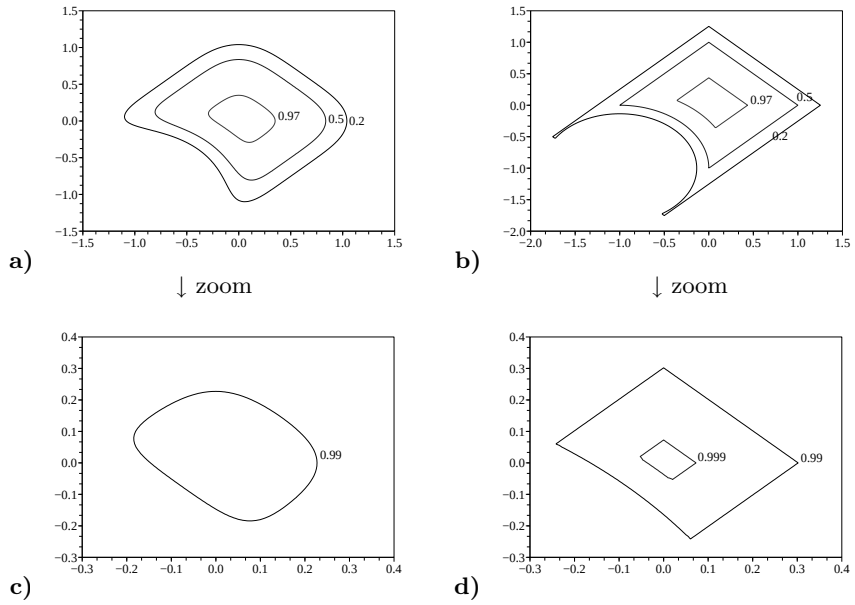

Fig. 2 Level sets of the probabilistic constraint $P(x)$ with the set G2 of random constraints. In pictures (a) and (c) the random variables Λ_p are independent zero-mean Gaussian variables with standard deviations 0.3. In pictures (b) and (d) the random variables Λ_p are all equal to a single real random variable with Gaussian statistics, mean zero and standard deviations $\sigma = 0.3$.

When the functions $f_p(x)$, $p = 1, ..., N_c$, are upper semi-continuous, then the probabilistic constraint $P(x)$ is upper semi-continuous and the admissible domain $\mathcal{A}_{1-\alpha}$ is closed. Consequently, the minimization problem of a strictly convex cost function is well posed (i.e. the problem has a unique solution). According to the previous section, it is important to use nonlinear and non-convex minimization algorithms. The computation of the gradient (and the Hessian) of the probabilistic constraint $P(x)$ is required in the nonlinear optimization. When simple analytical expressions of $P(x)$ (using the erf function for instance) are available, finite differences can be used to evaluate the partial derivatives of $P(x)$. However, in many situations, the expression of $P(x)$ involves an integral in high dimension, whose evaluation by deterministic quadrature or Monte Carlo methods is not easy. In this case, the computation of the the partial derivatives of $P(x)$ with finite differences can lead to significant errors. In Sections 5-6, specific techniques to compute gradient and Hessian of probabilistic constraints are proposed and discussed.

4 The probabilistic constraint

In this section, we consider the general framework introduced in Section 1 and consider the probabilistic constraint (4-5). Here we assume that Λ

is a Gaussian random vector with mean 0 and covariance matrix D_2. In many applications the Λ_i's are independent and have Gaussian distribution with mean 0 and variance σ_i^2, so that D_2 is the diagonal $N_\Lambda \times N_\Lambda$ matrix $(\sigma_i^2)_{i=1,...,N_\Lambda}$, but we shall carry out our analysis with a general covariance matrix.

The constraints are of the form:

$$g_p(x, \Lambda) \leq c_p\,, \quad p = 1, ..., N_c\,,$$

where $g : \mathbb{R}^{N_x} \times \mathbb{R}^{N_\Lambda} \to \mathbb{R}^{N_c}$ is a complicated function. We assume in the following that the function g is of class $\mathcal{C}^2$. The goal is to solve an optimization problem with such constraints. Due to the complexity of the function g in typical industrial problems, the computational cost of an evaluation method of the probabilistic constraint, its gradient and Hessian, is proportional to the number of calls of g.

In general we can distinguish two types of contraints:
- n constraints depend on the random vector Λ,
- $N_c - n$ constraints are deterministic:

$$g_p(x) \leq c_p\,, \quad p = n+1, ..., N_c\,.$$

The $N_c - n$ deterministic constraints will be dealt separately, using standard constrained optimization tools. Of course, it may happen that $N_c = n$, i.e. all constraints are random.

Let us fix x and consider the n constraints that depend on Λ. We shall restrict ourselves to cases where the constraints are linear in Λ, or to cases where the constraints can be linearized in Λ, in the sense that the approximation

$$g_p(x, \Lambda) \simeq g_p(x, 0) + \sum_{i=1}^{N_\Lambda} \frac{\partial g_p}{\partial \Lambda_i}(x, 0)\Lambda_i$$

is valid in the hypercube $\prod_{i=1}^{N_\Lambda}[-3\sigma_i, 3\sigma_i]$. Note that this is certainly the case when σ is small, and that the choice of the constant 3 is actually determined by the level $1 - \alpha$ of the admissible set. The set of constraints can then be rewritten in the form

$$G(x)\Lambda \leq C(x),$$

where $G(x)$ is a $n \times N_\Lambda$ deterministic matrix and $C(x)$ is a n-dimensional vector

$$\begin{aligned} G_{pi}(x) &= \frac{\partial g_p}{\partial \Lambda_i}(x, 0)\,, \quad p = 1, ..., n, \quad i = 1, ..., N_\Lambda, \\ C_p(x) &= c_p - g_p(x, 0)\,, \quad p = 1, ..., n. \end{aligned}$$

The computation of G requires $1 + 2N_\Lambda$ evaluations of g.

Now, let us fix $x \in \mathbb{R}^{N_x}$. This point is admissible if $P(x) \geq 1 - \alpha$, where the probabilistic constraint is

$$P(x) = \mathbb{P}(G(x)\Lambda \leq C(x)). \tag{17}$$

Our main goal is to compute $P(x)$, as well as its gradient and Hessian.

The random vector $G(x)\Lambda$ has distribution $\mathcal{N}(0, \Gamma(x))$ with $\Gamma(x)$ the real, symmetric, $n \times n$ matrix given by

$$\Gamma(x) = G(x) D_2 G(x)^t, \tag{18}$$

where D_2 is the covariance matrix of Λ.

We will assume in the following that $\Gamma(x)$ is an invertible matrix. This is the case in the applications we have considered so far, and this is one of the motivations for introducing the distinction between the n constraints that depend on Λ and the ones that do not depend on it. This hypothesis could be removed, but this would complicate the presentation.

Therefore, the probabilistic constraint can be expressed as the multi-dimensional integral

$$P(x) = \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_n(x)} p_{\Gamma(x)}(z_1, ..., z_n) dz_1 \cdots dz_n, \tag{19}$$

where p_Γ is the density of the n-dimensional distribution with mean 0 and covariance matrix Γ:

$$p_\Gamma(z) = \frac{1}{\sqrt{(2\pi)^n \det \Gamma}} \exp\left(-\frac{1}{2} z^t \Gamma^{-1} z\right).$$

If $\Gamma(x)$ were diagonal, then we would have simply

$$P(x) = \prod_{i=1}^{n} \Phi\left(\frac{C_i(x)}{\sqrt{\Gamma_{ii}(x)}}\right),$$

where Φ is the distribution function of the $\mathcal{N}(0,1)$ distribution. In this case, the probabilistic constraint can be evaluated with arbitrary precision, as well as its gradient and Hessian. Unfortunately, this is not the case in our applications, as well as in many other applications, so that the evaluation of the multi-dimensional integral (19) is necessary. The problem is to integrate a discontinuous function (the indicatrix function of the domain $(-\infty, C_1(x)] \times ... \times (-\infty, C_n(x)]$) in dimension n, which is usually large. Therefore, we propose to use a MC method. This method is cheap in our framework because it does not require any new call of the function g. It is reduced to the generation of an independent and identically distributed (i.i.d.) sequence of Gaussian random vectors with zero-mean and covariance matrix $\Gamma(x)$. Accordingly, $P(x)$ can be evaluated by MC method, and the estimator has the form:

$$P^{(M)}(x) = \frac{1}{M} \sum_{l=1}^{M} \mathbf{1}_{(-\infty, C_1(x)]}(Z_1^{(l)}) \cdots \mathbf{1}_{(-\infty, C_n(x)]}(Z_n^{(l)}), \tag{20}$$

where $Z^{(l)}$, $l = 1, ..., M$ is an i.i.d. sequence of random variables with density $p_{\Gamma(x)}$. The computational cost, i.e. the number of calls of the function g, is $1 + 2N_\Lambda$, which is necessary to evaluate $G(x)$, and therefore $\Gamma(x)$. Once this evaluation has been carried out the computational cost of the MC estimator is negligible.

5 The gradient of the probabilistic constraint

5.1 Probabilistic representations of gradients

The gradient of the probabilistic constraint is needed in optimization routines. We distinguish two methods:
Method 1. One can try to estimate the gradient from the MC estimations of $P(x)$. One should get estimates of $P(x)$ and $P(x+\delta x)$ and then use standard finite-dfferences to estimate the gradient. However, the fluctuations inherent to any MC method are dramatically amplified by the finite-differences scheme, so that a very high number M of simulations would be required. This method is usually prohibitive.
Method 2. It is possible to carry out MC simulations with an explicit expression of the gradient. The gradient of $P(x)$ has the form

$$\frac{\partial P}{\partial x_k}(x) = \sum_{i,j=1}^{n} A_{ij}(x)\frac{\partial \Gamma_{ij}(x)}{\partial x_k} + \sum_{i=1}^{n} B_i(x)\frac{\partial C_i(x)}{\partial x_k} \tag{21}$$

for $k = 1, ..., N_x$, where the matrix A and the vector B are given by:

$$\begin{aligned} A_{ij}(x) &= \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_n(x)} \frac{\partial p_\Gamma}{\partial \Gamma_{ij}}(z_1, ..., z_n)dz_1 \cdots dz_n \mid_{\Gamma=\Gamma(x)} \\ &= \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_n(x)} \frac{\partial \ln p_\Gamma}{\partial \Gamma_{ij}}(z)p_\Gamma(z)d^n z \mid_{\Gamma=\Gamma(x)}, \end{aligned} \tag{22}$$

$$\begin{aligned} B_i(x) &= \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_{i-1}(x)} \int_{-\infty}^{C_{i+1}(x)} \cdots \int_{-\infty}^{C_n(x)} \\ &\qquad \times p_{\Gamma(x)}(z_1, .., z_i = C_i(x), ..., z_n)dz_1 \cdots dz_{i-1}dz_{i+1} \cdots dz_n \\ &= \frac{1}{\sqrt{2\pi\Gamma_{ii}}} \exp\left(-\frac{C_i(x)^2}{2\Gamma_{ii}}\right) \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_{i-1}(x)} \int_{-\infty}^{C_{i+1}(x)} \cdots \int_{-\infty}^{C_n(x)} \\ &\qquad \times p_{\Gamma(x)}(z'|z_i = C_i(x))d^{n-1}z'. \end{aligned} \tag{23}$$

Here $p_\Gamma(z'|z_i)$ is the conditional density of the $(n-1)$-dimensional random vector $Z' = (Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n)$ given $Z_i = z_i$. The conditional distribution of the vector Z' given $Z_i = z_i$ is Gaussian with mean

$$\mu^{(i)} = \frac{1}{\Gamma_{ii}} z_i(\Gamma_{ji})_{j=1,...,i-1,i+1,...,n}$$

and $(n-1) \times (n-1)$ covariance matrix

$$\tilde{\Gamma}^{(i)} = \left(\Gamma_{kl} - \frac{1}{\Gamma_{ii}}\Gamma_{ki}\Gamma_{il}\right)_{k,l=1,...,i-1,i+1,...,n}.$$

Its density is given by

$$p_\Gamma(z'|z_i) = \frac{1}{\sqrt{(2\pi)^{n-1}\det\tilde{\Gamma}^{(i)}}} \exp\left(-\frac{1}{2}\left(z' - \mu^{(i)}\right)^t \left(\tilde{\Gamma}^{(i)}\right)^{-1} \left(z' - \mu^{(i)}\right)\right).$$

The logarithmic derivative of p_Γ has the form

$$\frac{\partial \ln p_\Gamma(z)}{\partial \Gamma_{ij}} = -\frac{1}{2\det\Gamma}\frac{\partial \det\Gamma}{\partial \Gamma_{ij}} - \frac{1}{2} z^t \frac{\partial \Gamma^{-1}}{\partial \Gamma_{ij}} z = -\frac{1}{2}(\Gamma^{-1})_{ji} + \frac{1}{2}(\Gamma^{-1}z)_i(\Gamma^{-1}z)_j. \tag{24}$$

5.2 Monte Carlo estimators

The goal of this subsection is to construct a MC estimator for the gradient of the probabilistic constraint. This estimator is based on the representation (21) and is given by (29) below.
Indeed, the matrix $A(x)$ can be evaluated by MC simulations, at the same time as $P(x)$ (i.e. with the same samples $Z^{(l)}$, $1 \le l \le M$):

$$\begin{aligned} A_{ij}^{(M)}(x) &= \frac{1}{M}\sum_{l=1}^{M} \mathbf{1}_{(-\infty,C_1(x)]}(Z_1^{(l)}) \cdots \mathbf{1}_{(-\infty,C_n(x)]}(Z_n^{(l)}) \frac{\partial \ln p_\Gamma}{\partial \Gamma_{ij}}(Z_1^{(l)},...,Z_n^{(l)}) \\ &= -\frac{1}{2}(\Gamma^{-1}(x))_{ji} P^{(M)}(x) \\ &\quad + \frac{1}{2M}\sum_{l=1}^{M} \mathbf{1}_{(-\infty,C_1(x)]}(Z_1^{(l)}) \cdots \mathbf{1}_{(-\infty,C_n(x)]}(Z_n^{(l)}) \\ &\qquad \times (\Gamma(x)^{-1}Z^{(l)})_i (\Gamma(x)^{-1}Z^{(l)})_j, \end{aligned} \tag{25}$$

where $Z^{(l)}, l = 1, ..., M$ is an i.i.d. sequence of n-dimensional Gaussian vectors with mean 0 and covariance matrix $\Gamma(x)$.

$B(x)$ can be evaluated by Monte Carlo as well, but this requires a specific computation for each $i = 1, ..., n$. The MC estimator for $B_i(x)$ is

$$\begin{aligned} B_i^{(M)}(x) &= \frac{1}{\sqrt{2\pi\Gamma_{ii}(x)}} \exp\left(-\frac{C_i(x)^2}{2\Gamma_{ii}(x)}\right) \\ &\quad \times \frac{1}{M}\sum_{l=1}^{M} \mathbf{1}_{(-\infty,C_1(x)]}({Z'_1}^{(l)}) \cdots \mathbf{1}_{(-\infty,C_{i-1}(x)]}({Z'_{i-1}}^{(l)}) \\ &\qquad \times \mathbf{1}_{(-\infty,C_{i+1}(x)]}({Z'_{i+1}}^{(l)}) \cdots \mathbf{1}_{(-\infty,C_n(x)]}({Z'_n}^{(l)}), \end{aligned} \tag{26}$$

where $({Z'_1}^{(l)}, ..., {Z'_{i-1}}^{(l)}, {Z'_{i+1}}^{(l)}, ..., {Z'_n}^{(l)})$, $l = 1, ..., M$ is an i.i.d. sequence of $(n-1)$-dimensional Gaussian vectors with mean

$$\mu^{(i)} = \left(\frac{\Gamma_{ji}(x)}{\Gamma_{ii}(x)}\right)_{j=1,...,i-1,i+1,...,n} C_i(x) \tag{27}$$

and covariance matrix

$$\tilde{\Gamma}^{(i)} = \left(\Gamma_{kl}(x) - \frac{\Gamma_{ki}(x)\Gamma_{il}(x)}{\Gamma_{ii}(x)}\right)_{k,l=1,...,i-1,i+1,...,n}. \tag{28}$$

In order to complete the computation of the gradient of $P(x)$ given by (21), one should also compute $(\partial \Gamma_{ij}(x))/(\partial x_k)$ and $(\partial C_i(x))/(\partial x_k)$. They are given by

$$\frac{\partial \Gamma_{ij}(x)}{\partial x_k} = \sum_{l,l'=1}^{N_\Lambda} \frac{\partial}{\partial x_k} \left[\frac{\partial g_i(x,0)}{\partial \Lambda_{l'}} D_{2ll'} \frac{\partial g_j(x,0)}{\partial \Lambda_l} \right],$$
$$\frac{\partial C_i(x)}{\partial x_k} = -\frac{\partial g_i(x,0)}{\partial x_k}.$$

Substituting these expressions into the expression (21) of the gradient, we obtain the MC estimator for the gradient of the probabilistic constraint:

$$\left(\frac{\partial P}{\partial x_k}(x) \right)^{(M)} = \sum_{i,j=1}^{n} A_{ij}^{(M)}(x) \frac{\partial \Gamma_{ij}(x)}{\partial x_k} - \sum_{i=1}^{n} B_i^{(M)}(x) \frac{\partial g_i(x,0)}{\partial x_k}, \qquad (29)$$

where $A_{ij}^{(M)}$ is the MC estimator (25) of A_{ij} and $B_i^{(M)}$ is the MC estimator (26) of B_i.

The computational cost for the evaluations of the terms $(\partial C_i(x))/(\partial x_k)$ is $2N_x$. The computational cost for the evaluations of the terms $(\partial^2 g_p(x,0))/(\partial x_k \partial \Lambda_i)$ is $4N_\Lambda N_x$ (by finite differences). This is the highest cost of the problem. However, the computation of the gradient of $P(x)$ can be dramatically simplified if the variances σ_i^2 are small enough, as we shall see in the next paragraph.

5.3 Simplified computation of the gradient

If the σ_i^2's are small, of the typical order σ^2, then the expression (21) of $(\partial P)/(\partial x_k)$ can be expanded in powers of σ^2. We can estimate the order of magnitudes of the four types of terms that appear in the sum (21). We can first note that the typical order of magnitude of $\Gamma(x)$ is σ^2, because it is linear in D_2. We have:
- $A_{ij}(x)$ is given by (22). It is of order σ^{-2}, because $(\partial \ln p_\Gamma)/(\partial \Gamma_{ij})$ is of order σ^{-2}.
- $(\partial \Gamma_{ij}(x))/(\partial x_k)$ is of oder σ^2, because it is linear in D_2.
- $B_i(x)$ is given by (23). It is of order σ^{-1}, because it is proportional to $1/\sqrt{\Gamma_{ii}}$.
- $(\partial C_i(x))/(\partial x_k)$ is of order 1, because it does not depend on D_2.

As a consequence, the dominant term in the expression (21) of $(\partial P)/(\partial x_k)$ is

$$\frac{\partial P}{\partial x_k}(x) \simeq \sum_{i=1}^{n} B_i(x) \frac{\partial C_i(x)}{\partial x_k}. \qquad (30)$$

This can be evaluated by the simplified MC estimator

$$\left(\frac{\partial P}{\partial x_k}(x) \right)^{(M)} \simeq -\sum_{i=1}^{n} B_i^{(M)}(x) \frac{\partial g_i(x,0)}{\partial x_k}, \qquad (31)$$

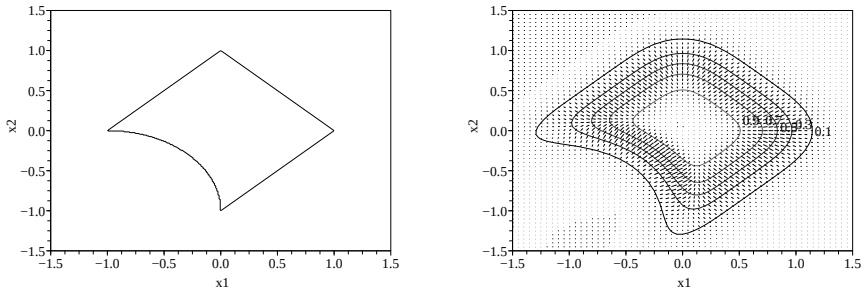

Fig. 3 Left picture: Deterministic admissible domain $\mathcal{A}$ for the model of page 14. Right picture: The solid lines plot the level sets $P(x) = 0.1$, 0.3, 0.5, 0.7 and 0.9 of the exact expression of the probabilistic constraint (32). The arrows represent the gradient field $\nabla P(x)$. Here $\sigma = 0.3$.

where $B_i^{(M)}$ is the MC estimator (26) of B_i. The advantage of this simplified expression is that it requires a small number of calls for g, namely $1+2(N_\Lambda + N_x)$, instead of $1 + 4N_\Lambda N_x$ for the exact one, because only the first-order derivatives of g with respect (x, Λ) at the point $(x, 0)$ are needed.

5.4 Numerical illustrations

We consider a particular example where the exact values of $P(x)$ and its gradient are known, so that the evaluation of the performance of the proposed simplified method is straightforward. We consider the case where $N_x = 2$, $N_\Lambda = 5$, $N_c = 5$, $g_p(x, \Lambda) = f_p(x) - \Lambda_p$, $c_p = 1$, with f_p given by (16). In the non-random case $\Lambda \equiv 0$, the admissible space is

$$\mathcal{A} = \left\{ x \in \mathbb{R}^2 \, : \, f_i(x) \leq 1 \, , \, i = 1, ..., 5 \right\},$$

which is the domain plotted in the left panel of Figure 3. It is in fact the limit of the domain $\mathcal{A}_{1-\alpha}$ when the variances of the Λ_i go to 0, whatever $\alpha \in (0, 1)$.

We consider the case where the Λ_p are i.i.d. zero-mean Gaussian variables with standard deviation $\sigma = 0.3$ (figures 3-4). The figures show that the simplified MC estimators for $P(x)$ and $\nabla P(x)$ are very accurate, even for $\sigma = 0.3$, while the method is derived in the asymptotic regime where σ is small. The exact expression of $P(x)$ is

$$P(x) = \prod_{i=1}^{5} \Phi \left(\frac{1 - f_i(x)}{\sigma} \right), \tag{32}$$

where Φ is the $\mathcal{N}(0, 1)$-distribution function.

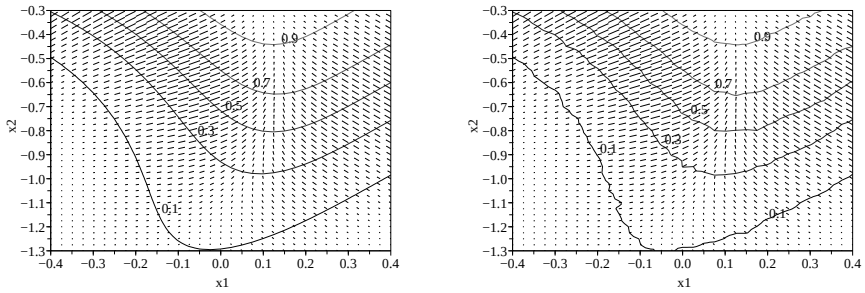

Fig. 4 The solid lines plot the level sets $P(x) = 0.1$, 0.3, 0.5, 0.7 and 0.9 of the model of page 14 with $\sigma = 0.3$. The arrows represent the gradient field $\nabla P(x)$. Left figure: the exact expression (32) of $P(x)$ is used. Right figure: the MC estimator (20) of $P(x)$ and the simplified MC estimator (31) of $\nabla P(x)$ are used, with $M = 5000$. The fluctuations of the MC estimation of $P(x)$ are of the order of the percent (the maximal absolute error on the 33×65 grid is 0.022). The fluctuations of the MC estimation of $\nabla P(x)$ are of the same order (the maximal value of ∇P is 2.68 and the maximal absolute error is 0.051).

6 The Hessian of the probabilistic constraint

The Hessian of the probabilistic constraint is needed (or can be useful) in some advanced optimization routines. Differentiating (21) with respect to x we obtain the complete expression of the Hessian. This expression contains a lot of terms, including terms that involve third-order derivatives of the form $(\partial^3 g_i)(\partial \Lambda_l \partial x_k \partial x_m)$, whose evaluations by finite-differences or any other method would be very costly. However, these terms have different orders of magnitude with respect to σ (assuming that σ is small). If we keep only the higher-order terms, we get the following simplified expression

$$\frac{\partial^2 P(x)}{\partial x_k \partial x_m} = \sum_{1 \le i \le j \le n} D_{ij}(x) \frac{\partial C_i}{\partial x_k} \frac{\partial C_j}{\partial x_m} \tag{33}$$

for $k, m = 1, ..., N_x$, where D_{ij} is of order σ^{-2} and $(\partial C_i)/(\partial x_k)$ is of order 1, so that the overall expression is of order σ^{-2}. Terms of order σ^{-1} and smaller have been neglected, as shown in the next subsection.

6.1 Derivation of the simplified expression of the Hessian

To get the expression (33), we have noted that the x-derivatives of the terms in the first sum in (21):

$$\sum_{i,j=1}^{n} A_{ij}(x) \frac{\partial \Gamma_{ij}(x)}{\partial x_k}$$

are of order σ^{-1}, and the x-derivatives of the terms of the second sum

$$\sum_{i=1}^{n} B_i(x) \frac{\partial C_i(x)}{\partial x_k}$$

have contributions of order σ^{-1}, namely

$$\sum_{i=1}^{n} B_i(x) \frac{\partial^2 C_i(x)}{\partial x_k \partial x_m},$$

and contributions of order σ^{-2}, namely

$$\sum_{i=1}^{n} \frac{\partial B_i(x)}{\partial x_m} \frac{\partial C_i(x)}{\partial x_k},$$

that give (33). Indeed:

$$\begin{aligned}\frac{\partial B_i(x)}{\partial x_m} = &\int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_{i-1}(x)} \int_{-\infty}^{C_{i+1}(x)} \cdots \int_{-\infty}^{C_n(x)} \frac{\partial p_{\Gamma(x)}(\check{z}_i, z_i = C_i(x))}{\partial x_m} d\check{z}_i \\ &+ \sum_{j \neq i} \int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_{i-1}(x)} \int_{-\infty}^{C_{i+1}(x)} \cdots \int_{-\infty}^{C_{j-1}(x)} \int_{-\infty}^{C_{j+1}(x)} \cdots \int_{-\infty}^{C_n(x)} \\ &\times p_{\Gamma(x)}(\check{z}_{i,j}, z_i = C_i(x), z_j = C_j(x)) d\check{z}_{i,j} \times \frac{\partial C_j(x)}{\partial x_m},\end{aligned}$$

where $\check{z}_i = (z_1, .., z_{i-1}, z_{i+1}, ..., z_n)$ and
$\check{z}_{i,j} = (z_1, .., z_{i-1}, z_{i+1}, ..., z_{j-1}, z_{j+1}, ..., z_n)$. The partial derivative of the probability density can be written as

$$\begin{aligned}\frac{\partial p_{\Gamma(x)}(\check{z}_i, z_i = C_i(x))}{\partial x_m} = &\sum_{k,l=1}^{n} \frac{\partial p_{\Gamma}(\check{z}_i, z_i = C_i(x))}{\partial \Gamma_{kl}} \mid_{\Gamma=\Gamma(x)} \frac{\partial \Gamma_{kl}(x)}{\partial x_m} \\ &+ \frac{\partial p_{\Gamma(x)}(\check{z}_i, z_i)}{\partial z_i} \mid_{z_i = C_i(x)} \frac{\partial C_i(x)}{\partial x_m}.\end{aligned}$$

The first sum is of order p_Γ because $(\partial p_\Gamma)/(\partial \Gamma_{kl})$ is of order $\sigma^{-2} p_\Gamma$ (see (24)) while $(\partial \Gamma_{kl})/(\partial x_m)$ is of order σ^2.
The second sum is of oder $\sigma^{-1} p_\Gamma$ because $(\partial C_i)/(\partial x_m)$ is of order 1 while $(\partial p_\Gamma)/(\partial z_i)$ is of order $\sigma^{-1} p_\Gamma$:

$$\frac{\partial \ln p_\Gamma(z)}{\partial z_i} = -\frac{1}{2} \frac{\partial}{\partial z_i} \left(z^t \Gamma^{-1} z \right) = -\frac{(z^t \Gamma^{-1})_i + (\Gamma^{-1} z)_i}{2} = -(\Gamma^{-1} z)_i.$$

We only keep the second sum in the simplified expression:

$$\begin{aligned}\frac{\partial B_i(x)}{\partial x_m} = &\int_{-\infty}^{C_1(x)} \cdots \int_{-\infty}^{C_{i-1}(x)} \int_{-\infty}^{C_{i+1}(x)} \cdots \int_{-\infty}^{C_n(x)} \\ &\times \frac{\ln p_{\Gamma(x)}}{\partial z_i}(\check{z}_i, z_i = C_i(x)) p_{\Gamma(x)}(\check{z}_i, z_i = C_i(x)) d\check{z}_i \times \frac{\partial C_i(x)}{\partial x_m}\end{aligned}$$

$$+\sum_{j\neq i}\int_{-\infty}^{C_1(x)}\cdots\int_{-\infty}^{C_{i-1}(x)}\int_{-\infty}^{C_{i+1}(x)}\cdots\int_{-\infty}^{C_{j-1}(x)}\int_{-\infty}^{C_{j+1}(x)}\cdots\int_{-\infty}^{C_n(x)}$$
$$\times p_{\Gamma(x)}(\check{z}_{i,j}, z_i = C_i(x), z_j = C_j(x))d\check{z}_{i,j}\times\frac{\partial C_j(x)}{\partial x_m}.$$

The explicit expressions of $D_{ij}(x)$ in (33) are the following ones. If $i<j$:

$$D_{ij}(x)=\frac{1}{\pi\sqrt{\Gamma_{ii}\Gamma_{jj}-\Gamma_{ij}^2}}\exp\left(-\frac{1}{2}\begin{pmatrix}C_i(x)\\C_j(x)\end{pmatrix}^t\begin{pmatrix}\Gamma_{ii}&\Gamma_{ij}\\\Gamma_{ij}&\Gamma_{jj}\end{pmatrix}^{-1}\begin{pmatrix}C_i(x)\\C_j(x)\end{pmatrix}\right)$$
$$\times\int_{-\infty}^{C_1(x)}\cdots\int_{-\infty}^{C_{i-1}(x)}\int_{-\infty}^{C_{i+1}(x)}\cdots\int_{-\infty}^{C_{j-1}(x)}\int_{-\infty}^{C_{j+1}(x)}\cdots\int_{-\infty}^{C_n(x)}$$
$$\times p_{\Gamma(x)}(z''|z_i=C_i(x),z_j=C_j(x))d^{n-2}z''. \tag{34}$$

The conditional distribution of the vector

$$Z''=(Z_1,...,Z_{i-1},Z_{i+1},...,Z_{j-1},Z_{j+1},...,Z_n)$$

given $(Z_i,Z_j)=(z_i,z_j)$ is Gaussian with mean $\mu^{(ij)}$, with

$$\mu_k^{(ij)}=\begin{pmatrix}\Gamma_{ki}\\\Gamma_{kj}\end{pmatrix}^t\begin{pmatrix}\Gamma_{ii}&\Gamma_{ij}\\\Gamma_{ij}&\Gamma_{jj}\end{pmatrix}^{-1}\begin{pmatrix}z_i\\z_j\end{pmatrix}$$

for $k=1,...,i-1,i+1,...,j-1,j+1,...,n$, and $(n-2)\times(n-2)$ covariance matrix $\tilde{\Gamma}^{(ij)}$ with

$$\tilde{\Gamma}_{kl}^{(ij)}=\Gamma_{kl}-\begin{pmatrix}\Gamma_{ki}\\\Gamma_{kj}\end{pmatrix}^t\begin{pmatrix}\Gamma_{ii}&\Gamma_{ij}\\\Gamma_{ij}&\Gamma_{jj}\end{pmatrix}^{-1}\begin{pmatrix}\Gamma_{li}\\\Gamma_{lj}\end{pmatrix}$$

for $k,l=1,...,i-1,i+1,...,j-1,j+1,...,n$.

If $i=j$

$$D_{ii}(x)=-\frac{1}{\sqrt{8\pi\Gamma_{ii}}}\exp\left(-\frac{C_i(x)^2}{2\Gamma_{ii}}\right)\int_{-\infty}^{C_1(x)}\cdots\int_{-\infty}^{C_{i-1}(x)}\int_{-\infty}^{C_{i+1}(x)}\cdots\int_{-\infty}^{C_n(x)}$$
$$\times p_{\Gamma(x)}(z'|z_i=C_i(x))\left((z^t\Gamma^{-1})_i+(\Gamma^{-1}z)_i\right)_{z_i=C_i(x)}d^{n-1}z'.$$

The expression of $D_{ii}(x)$ can also be written as

$$D_{ii}(x)=-(\Gamma^{-1})_{ii}C_i(x)B_i(x)$$
$$-\frac{1}{\sqrt{2\pi\Gamma_{ii}}}\exp\left(-\frac{C_i(x)^2}{2\Gamma_{ii}}\right)\int_{-\infty}^{C_1(x)}\cdots\int_{-\infty}^{C_{i-1}(x)}\int_{-\infty}^{C_{i+1}(x)}\cdots\int_{-\infty}^{C_n(x)}$$
$$\times p_{\Gamma(x)}(z'|z_i=C_i(x))\left(\sum_{k\neq i}(\Gamma^{-1})_{ik}z'_k\right)d^{n-1}z'. \tag{35}$$

Here $p_\Gamma(z'|z_i)$ is the conditional density of the $(n-1)$-dimensional random vector $Z'=(Z_1,...,Z_{i-1},Z_{i+1},...,Z_n)$ given $Z_i=z_i$ described in the previous section.

6.2 Monte Carlo estimators

The MC estimator for $D_{ij}(x)$, $i<j$, is

$$\begin{aligned} D_{ij}^{(M)}(x) = & \frac{1}{\pi\sqrt{\Gamma_{ii}\Gamma_{jj}(x)-\Gamma_{ij}^2(x)}} \\ & \times \exp\left(-\frac{1}{2}\begin{pmatrix} C_i(x) \\ C_j(x)\end{pmatrix}^t \begin{pmatrix} \Gamma_{ii}(x) & \Gamma_{ij}(x) \\ \Gamma_{ij}(x) & \Gamma_{jj}(x)\end{pmatrix}^{-1} \begin{pmatrix} C_i(x) \\ C_j(x)\end{pmatrix}\right) \\ & \times \frac{1}{M}\sum_{l=1}^M \mathbf{1}_{(-\infty,C_1(x)]}(Z_1''^{(l)})\cdots \mathbf{1}_{(-\infty,C_{i-1}(x)]}(Z_{i-1}''^{(l)}) \\ & \quad \times \mathbf{1}_{(-\infty,C_{i+1}(x)]}(Z_{i+1}''^{(l)})\cdots \mathbf{1}_{(-\infty,C_{j-1}(x)]}(Z_{j-1}''^{(l)}) \\ & \quad \times \mathbf{1}_{(-\infty,C_{j+1}(x)]}(Z_{j+1}''^{(l)})\cdots \mathbf{1}_{(-\infty,C_n(x)]}(Z_n''^{(l)}), \end{aligned} \quad (36)$$

where $(Z_1''^{(l)},...,Z_{i-1}''^{(l)},Z_{i+1}''^{(l)},...,Z_{j-1}''^{(l)},Z_{j+1}''^{(l)},...,Z_n''^{(l)})$, $l=1,...,M$ is an i.i.d. sequence of $(n-2)$-dimensional Gaussian vectors with mean $\mu^{(ij)}$,

$$\mu_k^{(ij)} = \begin{pmatrix} \Gamma_{ki}(x) \\ \Gamma_{kj}(x)\end{pmatrix}^t \begin{pmatrix} \Gamma_{ii}(x) & \Gamma_{ij}(x) \\ \Gamma_{ij}(x) & \Gamma_{jj}(x)\end{pmatrix}^{-1} \begin{pmatrix} C_i(x) \\ C_j(x)\end{pmatrix},$$

$k=1,...,i-1,i+1,...,j-1,j+1,...n$, and covariance matrix $\tilde{\Gamma}^{(ij)}$,

$$\tilde{\Gamma}_{kl}^{(ij)} = \Gamma_{kl} - \begin{pmatrix} \Gamma_{ki}(x) \\ \Gamma_{kj}(x)\end{pmatrix}^t \begin{pmatrix} \Gamma_{ii}(x) & \Gamma_{ij}(x) \\ \Gamma_{ij}(x) & \Gamma_{jj}(x)\end{pmatrix}^{-1} \begin{pmatrix} \Gamma_{li}(x) \\ \Gamma_{lj}(x)\end{pmatrix},$$

$k,l=1,...,i-1,i+1,...,j-1,j+1,...,n$.

The MC estimator for $D_{ii}(x)$ is

$$\begin{aligned} D_{ii}^{(M)}(x) = & -(\Gamma^{-1})_{ii}C_i(x)B_i^{(M)}(x) \\ & -\frac{1}{\sqrt{2\pi\Gamma_{ii}}}\exp\left(-\frac{C_i(x)^2}{2\Gamma_{ii}}\right) \\ & \times \frac{1}{M}\sum_{l=1}^M \mathbf{1}_{(-\infty,C_1(x)]}(Z_1'^{(l)})\cdots \mathbf{1}_{(-\infty,C_{i-1}(x)]}(Z_{i-1}'^{(l)}) \\ & \quad \times \mathbf{1}_{(-\infty,C_{i+1}(x)]}(Z_{i+1}'^{(l)})\cdots \mathbf{1}_{(-\infty,C_n(x)]}(Z_n'^{(l)}) \\ & \quad \times \left(\sum_{k\neq i}(\Gamma^{-1})_{ik}Z_k'^{(l)}\right), \end{aligned} \quad (37)$$

where $(Z_1'^{(l)},...,Z_{i-1}'^{(l)},Z_{i+1}'^{(l)},...,Z_n'^{(l)})$, $l=1,...,M$ is an i.i.d. sequence of $(n-1)$-dimensional Gaussian vectors with mean $\mu^{(i)}$ given by (27) and covariance matrix $\tilde{\Gamma}^{(i)}$ given by (28). The same sequence as the one used

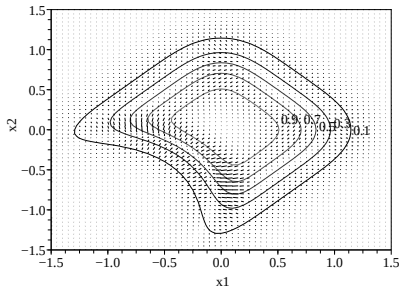

Fig. 5 The solid lines plot the level sets $P(x) = 0.1$, 0.3, 0.5, 0.7 and 0.9 of the model of page 14 with $\sigma = 0.3$. The arrows represent the diagonal Hessian field $(\partial^2_{x_1} P, \partial^2_{x_2} P)(x)$. Here the exact expression (32) of $P(x)$ is used.

to construct the estimator $B_i^{(M)}$ by (26) can be used here for the estimator $D_{ii}^{(M)}$. As a result, the MC estimator for the Hessian of $P(x)$ is:

$$\left(\frac{\partial^2 P(x)}{\partial x_k \partial x_m}\right)^{(M)} = \sum_{1 \le i \le j \le n} D_{ij}^{(M)}(x) \frac{\partial C_i}{\partial x_k} \frac{\partial C_j}{\partial x_m}. \tag{38}$$

To summarize, the computation of the Hessian (33) by the MC estimator (38) does not require any additional call to the function g, since only C, the gradient of C, and the matrix Γ are needed.

6.3 Numerical illustrations

We illustrate the results with the model of page 14 introduced in Subsection 5.4. Figure 6 shows that the simplified MC estimator for the Hessian of $P(x)$ is very accurate, even for $\sigma = 0.3$. This means that, with $1 + 2(N_x + N_\Lambda)$ calls of the function g, we can get accurate estimates of $P(x)$, its gradient and its Hessian.

7 Stochastic optimization

The goal is to solve

$$\min_{x \in \mathbb{R}^{N_x}} J(x) \text{ s. t. } \left\{ \begin{array}{l} P(x) \ge 1 - \alpha \text{ and} \\ g_p(x) \le c_p\,, \quad p = n+1, ..., N_c \end{array} \right\}. \tag{39}$$

The previous sections show how to compute efficiently $P(x)$, its gradient and its Hessian. Therefore, the problem has been reduced to a standard constrained optimization problem. Different techniques have been proposed for solving constrained optimization problems: reduced-gradient methods, sequential linear and quadratic programming methods, and methods based on

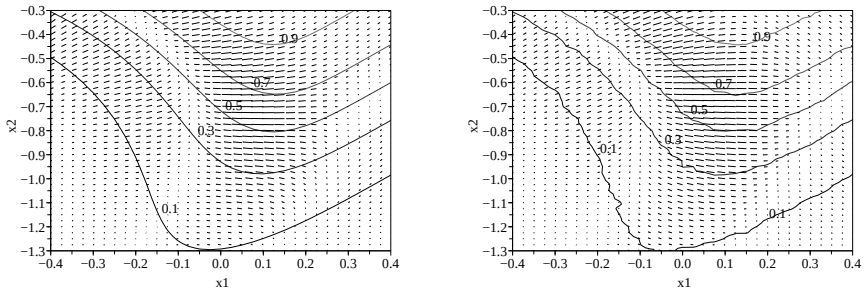

Fig. 6 The solid lines plot the level sets $P(x) = 0.1$, 0.3, 0.5, 0.7 and 0.9 of the model of page 14 with $\sigma = 0.3$. The arrows represent the diagonal Hessian field $(\partial_x^2 P, \partial_y^2 P)(x)$. Left figure: the exact expression (32) of $P(x)$ is used. Right figure: the MC estimator (20) of $P(x)$ and the simplified MC estimator (38) of $(\partial_{x_1}^2 P, \partial_{x_2}^2 P)(x)$ are used, with $M = 5000$.

augmented Lagrangians and exact penalty functions. Fletcher's book [8] contains discussions of constrained optimization theory and sequential quadratic programming. Gill, Murray, and Wright [9] discuss reduced-gradient methods. Bertsekas [3] presents augmented Lagrangian algorithms.

Penalization techniques are well-adapted to stochastic optimization [20]. The basic idea of the penalization approach is to convert the constrained optimization problem (39) into an unconstrained optimization problem of the form: minimize the function $\tilde{J}$ defined by

$$\tilde{J}(x) := J(x) + \rho Q(x),$$

where $Q : \mathbb{R}^{N_x} \to \mathbb{R}$ is the penalty function and ρ is a positive real number, referred to as the penalty parameter. Note that, in our case, we have $K = N_c - n + 1$, $G_1(x) = 1 - \alpha - P(x)$, and $G_k(x) = g_{k-n+1}(x) - c_{k-n+1}$, $k = 2, ..., K$. The q-th order penalization ($q \geq 2$) uses the penalty function [16]

$$Q(x) = \frac{1}{q} \sum_{k=1}^{K} \big(\max\{0, G_k(x)\} \big)^q,$$

which is constinuously differentiable with

$$\nabla Q(x) = \sum_{k=1}^{K} \max\{0, G_k(x)\}^{q-1} \nabla G_k(x).$$

In the following minimization examples, the unconstrained problem obtained by penalization is solved by the Scilab routine *optim*. This routine uses a gradient method optimization. At each iteration k, $k \geq 1$, a descent direction d_k of the cost function $\tilde{J}$ at the current state x_k is determined. This direction has the form $d_k = -W_k g_k$, where W_k is the current approximation

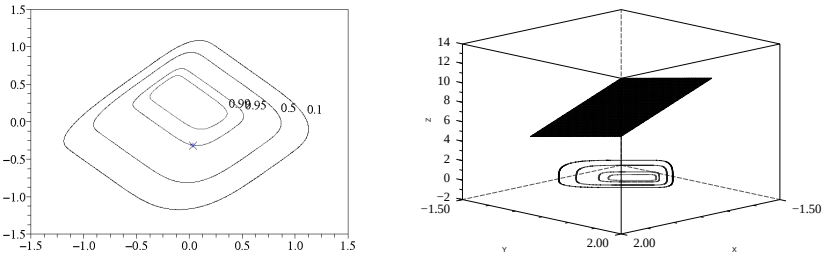

Fig. 7 Minimization of J_1 subject to G_1 for the probability level $1 - \alpha = 0.95$. The left picture shows that the minimum is obtained at the level set 0.95. The right picture plots the cost function and level set curves.

of the inverse Hessian of the cost function at x_k and g_k is the gradient of the cost function at x_k. The iteration formula has the form:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a step-size determined along the direction d_k by Wolf's conditions.

Example 3 (convex optimization) We consider the linear minimization problem:
Minimize the function

$$J_1(x) = x_1 + 4x_2 + 4,$$

subject to the constraints G1 given by (13) where the Λ_p are independent zero-mean Gaussian random variables with standard deviations $0.1p$, $p = 1, ..., 4$. Figure 7 shows the result of the optimization routine with quadratic penalization.

Example 4 (non-convex optimization) We consider the nonlinear minimization problem:
Minimize the function

$$J_2(x) = (x_1 - 1)^2 + 4x_2^2 + 4,$$

subject to the constraints G2 given by (15) where the Λ_p are independent zero-mean Gaussian random variables with standard deviations $0.1p$, $p = 1, ..., 5$. Figure 8 shows the result of the optimization routine with quadratic penalization.

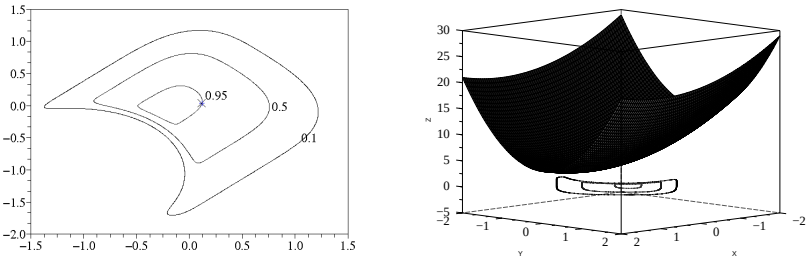

Fig. 8 Minimization of J_2 subject to G_2 for the probability level $1-\alpha = 0.95$. The left picture shows that the minimum is obtained at the level set 0.95. The right picture plots the cost function and level set curves.

8 Optimization of a random parameter

In this section we briefly address another stochastic optimization problem. Here the constraints are assumed to be known, as well as the cost function, but the input variable $x \in \mathbb{R}^{N_x}$ is known only approximately. This problem models the production of an artefact by a non-perfect machine and takes into account the unavoidable imperfections of the process. In this case the probabilistic constraint has the form

$$P(x) = \mathbb{P}\Big(f_p(X_x) \leq 0\,,\, p = 1, ..., N_c\Big), \tag{40}$$

where f_p, $p = 1, ..., N_c$ are constraint functions, X_x is a random variable with density $\phi(\cdot - x)$, and ϕ is the density of a zero-mean random variable that models the fluctuations of the process. The admissible domain $\mathcal{A}_{1-\alpha}$ is defined by

$$\mathcal{A}_{1-\alpha} = \big\{x \in \mathbb{R}^{N_x} \,:\, P(x) \geq 1-\alpha\big\}\,.$$

We can rewrite the probabilistic constraint as

$$P(x) = \mathbb{P}\Big(g_p(x, \Lambda) \leq 0\,,\, p = 1, ..., N_c\Big),$$

where $g_p(x, \Lambda) = f_p(x + \Lambda)$ and Λ a is a random vector with density f. This problem has the same form as the previous one. If ϕ is a multi-variate normal vector with zero mean and covariance matrix D_2 and the variations of f_p are small over the hypercube $\prod_{i=1}^{N_x}[-3\sigma_i, 3\sigma_i]$, then we can apply the same methodology. We first expand the constraints in Λ and obtain:

$$P(x) = \mathbb{P}(G(x)\Lambda \leq f(x)),$$

where the $N_x \times N_c$ matrix $G(x)$ is given by

$$G_{pk}(x) = \frac{\partial f_p(x)}{\partial x_k}\,, \quad p = 1, ..., N_c\,, \quad k = 1, ..., N_x,$$

and the random vector $G(x)\Lambda$ is Gaussian with zero-mean and covariance $\Gamma(x) = G(x)D_2G(x)^t$. By applying the strategy developed in the previous

section, we obtain that it is possible to compute accurately $P(x)$, its gradient and its Hessian if we have (finite-differences) estimates of $(f_p(x))_{p=1,...,N_c}$, its gradient and its Hessian. This means that an optimization problem with a probabilistic constraint of the form $P(x) \geq 1-\alpha$ with $P(x)$ given by (40) has the same computational cost as the same problem with the set of deterministic constraints $f_p(x) \leq 0$, $p = 1, ..., N_c$.

9 Conclusion

In this article we have investigated the possibility to solve a chance constrained problem. This problem consists in minimizing a deterministic cost function $J(x)$ over a domain $\mathcal{A}_{1-\alpha}$ defined as the set of points $x \in \mathbb{R}^{N_x}$ satisfying a collection of N_c random constraints $g_p(x, \Lambda) \leq 0$, $p = 1, ..., N_c$, with some prescribed probability $1-\alpha$. Here Λ is a random vector that models the uncertainty of the constraints. We have studied the probabilistic constraint

$$P(x) = \mathbb{P}(g_p(x, \Lambda) \leq 0\,,\, p = 1, ..., N_c),$$

that is the probability that the point x satisfies the constraints. We have analyzed the structural properties of the admissible domain

$$\mathcal{A}_{1-\alpha} = \left\{ x \in \mathbb{R}^{N_x} \,:\, P(x) \geq 1-\alpha \right\},$$

that is the set of points x that satisfy the constraints with probability $1-\alpha$. In particular we have presented general hypotheses about the constraints functions g_p and the distribution function of Λ ensuring the convexity of the domain.

The original contribution of this paper consists in a rapid and efficient MC method to estimate the probabilistic constraint $P(x)$, as well as its gradient vector and Hessian matrix. This method requires a very small number of calls of the constraint functions g_p, which makes it possible to implement an optimization routine for the chance constrained problem. The method is based on particular probabilistic representations of the gradient and Hessian of $P(x)$ which can be derived when the random vector Λ is multivariate normal. It should be possible to extend the results to cases where the conditional distributions of the vector Λ given one or two entries are explicit enough.

Acknowledgements Most of this work was carried out during the Summer Mathematical Research Center on Scientific Computing (whose french acronym is CEMRACS), which took place in Marseille in july and august 2006. This project was initiated and supported by the European Aeronautic Defence and Space (EADS) company. In particular, we thank R. Lebrun and F. Mangeant for stimulating and useful discussions.

References

1. Bastin, F., Cirillo, C., Toint, P. L.: Convergence theory for nonconvex stochastic programming with an application to mixed logit. Math. Program., Ser. A, Vol. 108, pp. 207-234 (2006).

2. Bergstrom, T., Bagnoli, M.: Log-concave probability and its applications. Econom. Theory, Vol. 26, pp 445-469 (2005).
3. Bertsekas, D.P.: Constrained optimization and Lagrange multiplier methods. Academic Press, New York (1982).
4. Birge, J. R.: Stochastic programming computation and applications. Informs. J. on Computing, Vol. 9, pp 111-133 (1997).
5. Birge, J. R., Louveaux, F.: Introduction to stochastic programming. Springer, New York (1997).
6. Borell, C.: Convex set functions in d-space. Period. Math. Hungar., Vol. 6, pp 111-136 (1975).
7. Dubi, A.: Monte Carlo applications in systems engineering. Wiley, Chichester (2000).
8. Fletcher, R.: Practical methods of optimization. Wiley, New York (1987).
9. Gill, P. E., Murray, W., Wright, M. H.: Practical optimization. Academic Press, New York (1981).
10. Henrion, R., Römisch, W.: Stability of solutions of chance constrained stochastic programs. In J. Guddat, R. Hirabayashi, H.Th. Jongen and F. Twilt (eds.): Parametric Optimization and Related Topics V, Peter Lang, Frankfurt A. M., pp. 95-114 (2000).
11. Henrion, R., Strugarek, C.: Convexity of chance constraints with independent random variables. Stochastic Programming E-Print Series (SPEPS) Vol. 9 (2006), to appear in: Computational Optimization and Applications.
12. Niederreiter, H.: Monte Carlo and Quasi-Monte Carlo Methods MCQMC 2002. Proceedings, Conference at the National University of Singapore, Republic of Singapore, Springer (2004).
13. Kall, P., Wallace, S.W.: Stochastic programming. Wiley, Chichester (1994).
14. Prékopa, A.: Logarithmic concave measures with application to stochastic programming. Acta Scientiarum Mathematicarum, Vol. 32, pp 301-316 (1971).
15. Prékopa, A.: Stochastic programming. Kluwer, Dordrecht (1995).
16. Ruszczyński, A.: Nonlinear optimization. Princeton University Press, Princeton (2006).
17. Ruszczyński, A., Shapiro, A.: Stochastic programming. Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam (2003).
18. Sen, S., Higle, J.L.: An introductory tutorial on stochastic linear programming models. Interfaces, Vol. 29, pp 33-61 (1999).
19. Shapiro, A.: Stochastic programming approach to optimization under uncertainty. To appear in Math. Program., Ser. B.
20. Wang, I. J., Spall, J. C.: Stochastic optimization with inequality constraints using simultaneous perturbations and penality functions. Proceedings of IEEE Conference on Decision and Control, Maui, HI, pp. 3808-3813 (2003).