

On joint probabilistic constraints with Gaussian coefficient matrix ^{*}

W. van Ackooij[†] R. Henrion[‡] A. Möller[‡] R. Zorgati[†]

October 9, 2010

Abstract

The paper deals with joint probabilistic constraints defined by a Gaussian coefficient matrix. It is shown how to explicitly reduce the computation of values and gradients of the underlying probability function to that of Gaussian distribution functions. This allows to employ existing efficient algorithms for calculating this latter class of function in order to solve probabilistically constrained optimization problems of the indicated type. Results are illustrated by an example from energy production.

An important class of probabilistic programming problems is given by

$$\min\{f(x) \mid \mathbb{P}(\Xi x \leq a) \geq p\}, \quad (1)$$

where f is an objective, x is a decision variable and Ξ is a random coefficient matrix. The probabilistic constraint in this problem expresses the wish to find a decision x which guarantees that the random inequality system $\Xi x \leq a$ is satisfied at least with a probability $p \in [0, 1]$. Problems of this type have abundant applications in engineering (e.g. blending problems) and finance (e.g. portfolio problems). In this paper we want to show how to deal with such probabilistic constraint under the assumption that the elements of Ξ have a joint Gaussian (multivariate normal) distribution. Observe first that the constraint $\Xi x \leq a$ can be equivalently written as $T(x)\xi \leq a$, where ξ refers to the row-wise vectorization of Ξ and $T(x)$ is a matrix having row i

$$T_i(x) = (0^T, \dots, 0^T, \underbrace{x^T}_i, 0^T, \dots, 0^T).$$

The use of this equivalent form of the constraint has some notational convenience and also allows us slightly to generalize the model (as needed, for instance in the

^{*}This work was supported by the OSIRIS Department of *Electricité de France R&D* and by the DFG Research Center MATHEON “Mathematics for key technologies” in Berlin

[†]Electricité de France R&D, 1 avenue du Général de Gaulle, 92141 Clamart, France

[‡]Weierstrass Institute, Mohrenstraße 39, 10117 Berlin, Germany

example presented below). Therefore, we are led to consider probabilistic constraints of the form $\varphi(x) \geq p$, where

$$\varphi(x) := \mathbb{P}(T(x)\xi \leq \alpha(x)) \quad (2)$$

for some s -dimensional Gaussian random vector ξ with expectation μ and (positive definite) covariance matrix Σ (notation: $\xi \sim \mathcal{N}(\mu, \Sigma)$) and for some continuously differentiable mappings $T : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times s}$ and $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Clearly, (2) provides a framework including the probabilistic constraint in (1) as a special case.

When dealing with inequality constraints like $\varphi(x) \geq p$ in algorithms of nonlinear optimization, it is necessary to have access to at least to values and gradients of φ . For simpler probabilistic constraints of the separated form $\mathbb{P}(\xi \leq T(x)) \geq p$ with Gaussian random vector ξ , successful numerical results are reported (e.g., [1]) which are based on an efficient computation of multivariate Gaussian distribution functions using a code by Genz [2]. A key issue here was the well-known possibility to reduce the computation of gradients for Gaussian distribution functions to such distribution functions again (see (9)), so that Genz's could be employed to calculating values and gradients of Gaussian distribution functions at a time. We shall see that, on a more involved level, it is also possible to completely break down the computation of φ and $\nabla\varphi$ in (2) to that of Gaussian distribution functions so that the same algorithmic approach for nonlinear optimization can be applied as in the case of the simpler probabilistic constraint. Our aim is to derive formulae for φ and $\nabla\varphi$ which are completely in terms of Gaussian distribution functions and of the problem data. Under the problem data of (2) we understand the mappings T, α and their gradients (to be provided) $\nabla T, \nabla\alpha$ as well as the distribution parameters μ, Σ of the random vector ξ . Since a closed form expression for φ and $\nabla\varphi$ in the desired explicit form is almost impossible to represent, we shall recursively break down parts of the corresponding formulae.

First observe that, by a well-known formula for parameter transformation of Gaussian distributions, our initial assumption $\xi \sim \mathcal{N}(\mu, \Sigma)$ leads to

$$T(x)\xi \sim \mathcal{N}(\mu(x), \Sigma(x)),$$

where

$$\mu(x) := T(x)\mu, \quad \Sigma(x) := T(x)\Sigma[T(x)]^T. \quad (3)$$

Putting

$$D(x) := \text{diag} \left(\Sigma_{ii}^{-1/2}(x) \right)_{i=1, \dots, m}; \quad R(x) := D(x)\Sigma(x)D(x). \quad (4)$$

one gets that the normalized vector $\eta(x) := D(x)[T(x)\xi - \mu(x)]$ is distributed according to $\eta(x) \sim \mathcal{N}(0, R(x))$. In particular, $R(x)$ is a correlation matrix by (4). Consequently, $\varphi(x) = \Phi^{R(x)}(\beta(x))$, where

$$\beta(x) := D(x)(\alpha(x) - \mu(x)) \quad (5)$$

and Φ^R denotes the distribution function of a standard Gaussian distribution with zero mean, unit variances and correlation matrix R . In this way, the first task is achieved, namely the evaluation of φ is reduced to that of a Gaussian distribution function applied to a mapping β which via (4) and (3) can be explicitly led back to the problem data. In order to do so for the gradient of φ as well, we define the mapping $\gamma(R, z) := \Phi^R(z)$. Note that γ will be differentiable provided that R is positive definite which can be ensured by requiring surjectivity of $T(x)$ (see (3) and (4)). Clearly, $\varphi(x) = \gamma(R(x), \beta(x))$ and deriving this expression at some arbitrarily fixed x yields

$$\nabla\varphi(x) = \sum_{i=1}^m \frac{\partial\gamma}{\partial z_i}(R(x), \beta(x)) \nabla\beta_i(x) + \sum_{i,j=1}^m \frac{\partial\gamma}{\partial R_{ij}}(R(x), \beta(x)) \nabla R_{ij}(x) \quad (6)$$

Let us focus first on the first term in this sum. By (5) and (4), we have

$$\nabla\beta_i(x) = (\alpha_i(x) - \mu_i(x)) \nabla\Sigma_{ii}^{-1/2}(x) + \Sigma_{ii}^{-1/2}(x) (\nabla\alpha_i(x) - \nabla\mu_i(x)). \quad (7)$$

The values for $\mu_i(x)$ and $\Sigma_{ii}^{-1/2}(x)$ are obtained from the model data via (3). Deriving these expressions, one gets

$$\nabla\mu_i(x) = \sum_{j=1}^s \mu_j \nabla T_{ij}(x); \quad \nabla\Sigma_{ii}^{-1/2}(x) = -\frac{1}{2} \Sigma_{ii}^{-3/2}(x) \nabla\Sigma_{ii}(x).$$

As far as the evaluation of $\nabla\Sigma_{ii}^{-1/2}(x)$ is concerned, the expression $\Sigma_{ii}^{-3/2}(x)$ follows immediately from (3), whereas for the gradient $\nabla\Sigma_{ii}(x)$ we make a forward reference to (11) in order to develop (7) explicitly in terms of $\alpha, \nabla\alpha, T, \nabla T, \mu, \Sigma$. Next, we observe that

$$\frac{\partial\gamma}{\partial z_i}(R(x), \beta(x)) = \frac{\partial\Phi^{R(x)}}{\partial z_i}(\beta(x)) \quad (i = 1, \dots, m). \quad (8)$$

Consequently, one is left with the task of calculating the gradient of a Gaussian distribution function (with correlation matrix fixed as $R(x)$). Here, one may rely on the well-known formula (see [4], p. 204)

$$\frac{\partial\Phi^R}{\partial z_i}(z) = h(z_i) \tilde{\Phi}^{\tilde{R}_i}(\tilde{z}_i). \quad (9)$$

for an arbitrary correlation matrix R . In the context of (8), we put $R := R(x)$ and $z := \beta(x)$ in (9). In this formula, h denotes the density of the one-dimensional standard Gaussian distribution. The argument $\tilde{z}_i \in \mathbb{R}^s$ is obtained from z and from R by

$$\tilde{z}_i = \left(\frac{z_1 - r_{1,i}z_i}{\sqrt{1 - r_{1,i}^2}}, \dots, \frac{z_{i-1} - r_{i-1,i}z_i}{\sqrt{1 - r_{i-1,i}^2}}, \frac{z_{i+1} - r_{i+1,i}z_i}{\sqrt{1 - r_{i+1,i}^2}}, \dots, \frac{z_s - r_{s,i}z_i}{\sqrt{1 - r_{s,i}^2}} \right), \quad (10)$$

where the $r_{i,j}$ refer to the entries of R . Moreover, $\tilde{\Phi}^{\tilde{R}_i}$ is the distribution function of the $(s-1)$ -dimensional standard Gaussian distribution with correlation matrix \tilde{R}_i which itself is obtained from R by deleting the i th row and column of the matrix \hat{R} whose entries are defined as

$$\hat{r}_{j,k} := \frac{r_{j,k} - r_{j,i}r_{k,i}}{\sqrt{1 - r_{j,i}^2}\sqrt{1 - r_{k,i}^2}} \quad (j, k = 1, \dots, s).$$

Summarizing, we have made the first term in the sum (6) fully explicit in terms of the model data. Turning to the second term, observe that by (4)

$$R_{ij}(x) = \frac{\Sigma_{ij}(x)}{\sqrt{\Sigma_{ii}(x)\Sigma_{jj}(x)}} \quad (i, j = 1, \dots, s).$$

Therefore, for $i, j = 1, \dots, s$,

$$\nabla R_{ij}(x) = \frac{\Sigma_{ii}(x)\Sigma_{jj}(x)\nabla\Sigma_{ij}(x) - \frac{1}{2}\Sigma_{ij}(x)[\Sigma_{jj}(x)\nabla\Sigma_{ii}(x) + \Sigma_{ii}(x)\nabla\Sigma_{jj}(x)]}{(\Sigma_{ii}(x)\Sigma_{jj}(x))^{3/2}}.$$

We can develop $\nabla R_{ij}(x)$ in terms of the model data upon noting that by (3)

$$\nabla\Sigma_{ij}(x) = \sum_{k,l=1}^s \Sigma_{kl} [T_{jl}(x)\nabla T_{ik}(x) + T_{ik}(x)\nabla T_{jl}(x)] \quad (i, j = 1, \dots, m). \quad (11)$$

It remains to analyze the expression $\frac{\partial\gamma}{\partial R_{ij}}(R(x), \beta(x))$ in (6). By definition,

$$\frac{\partial\gamma}{\partial R_{ij}}(R(x), \beta(x)) = \frac{\partial\Phi^{R(x)}}{\partial R_{ij}}(\beta(x)). \quad (12)$$

Hence, we have to know, how to calculate sensitivities of Gaussian distribution functions with respect to correlation coefficients. Here, we rely on Gupta's formula ([3], eq. (53)):

$$\begin{aligned} \frac{\partial\Phi^R}{\partial R_{ij}}(z) &= \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_s} \frac{\partial^2 g}{\partial z_i \partial z_j}(\zeta_1, \dots, \zeta_s) d\zeta_s \cdots d\zeta_1 \\ &= \frac{\partial^2}{\partial z_i \partial z_j} \left(\int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_s} g(\zeta_1, \dots, \zeta_s) d\zeta_s \cdots d\zeta_1 \right) \\ &= \frac{\partial^2 \Phi^R}{\partial z_i \partial z_j}(z) \quad (i, j = 1, \dots, s). \end{aligned}$$

To apply this formula in the context of (12), we have to set $R := R(x)$ and $z := \beta(x)$. In this formula, g is the density associated with Φ^R . The formula tells us, that the sensitivities with respect to correlation coefficients can be obtained as second partial derivatives of the Gaussian distribution function Φ^R . As we already know

how to calculate first order partial derivatives of Φ^R via (9) and this leads back to the calculus of (different) Gaussian distribution functions again, we may apply the same formula a second time in order to derive the desired second partial derivatives:

$$\frac{\partial^2 \Phi^R}{\partial z_i \partial z_j}(z) = \begin{cases} h'(z_i) \tilde{\Phi}^{\tilde{R}_i}(\tilde{z}_i) - h(z_i) \sum_{k=1, k \neq i}^s \frac{\partial \tilde{\Phi}^{\tilde{R}_i}}{\partial z_k}(\tilde{z}_i) \frac{r_{k,i}}{\sqrt{1-r_{k,i}^2}} & \text{if } i = j \\ h(z_i) \frac{\partial \tilde{\Phi}^{\tilde{R}_i}}{\partial z_j}(\tilde{z}_i) \frac{1}{\sqrt{1-r_{j,i}^2}} & \text{if } i \neq j \end{cases}$$

Here, we used corresponding inner derivatives of (10) with respect to z_j . Now, the partial derivatives $\frac{\partial \tilde{\Phi}^{\tilde{R}_i}}{\partial z_k}(\tilde{z}_i)$ occurring in the formula above can be calculated exactly in the same way as described in (9) for the partial derivatives $\frac{\partial \Phi^R}{\partial z_i}(z)$ but this time applied to the new correlation matrix \tilde{R}_i and the new argument \tilde{z}_i . Hence, we have provided a fully explicit way to calculate the gradient (6).

As an application we consider the following simple unit commitment problem: Given a time horizon of 12 months and stochastic electricity demands in each of the 12 time periods, we are looking for a cost minimal portfolio of 4 power generation units (nuclear, coal, fuel, gas) such that the demands are met at a probability of at least $p = 0.9$. Apart from the demands, also the availability of power plants is assumed to be random. This will be described by random availability coefficients (between zero and one) which have to be multiplied with the committed production in order to determine the actual production of a plant. Let us denote by $x_i := (x_{i1}, x_{i2}, x_{i3}, x_{i4})$ the production committed to units 1 to 4 in month i . Similarly, denote by $\hat{\xi}_i := (\hat{\xi}_{i1}, \hat{\xi}_{i2}, \hat{\xi}_{i3}, \hat{\xi}_{i4})$ the availability coefficients for units 1 to 4 in month i . Let the demand vector be given by $\tilde{\xi} := (\tilde{\xi}_1, \dots, \tilde{\xi}_{12})$. Then, demand satisfaction as a probabilistic constraint at level p can be modeled by the inequality

$$\mathbb{P}(x_i^T \hat{\xi}_i \geq \tilde{\xi}_i \quad (i = 1, \dots, 12)) \geq p.$$

This constraint fits to the model $\varphi(x) \geq p$ and φ defined as in (2) with the model data

$$T(x) := \begin{pmatrix} -x_1^T & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -x_{12}^T & 0 & \cdots & 1 \end{pmatrix}; \quad \xi := \begin{pmatrix} \tilde{\xi} \\ \tilde{\xi} \end{pmatrix}; \quad \alpha(x) := 0. \quad (13)$$

Observe that $T(x)$ is surjective for all x as required above for differentiability reasons. Apart from the probabilistic constraint, deterministic time-dependent upper levels are given for the production of each unit. We collect these upper levels in a vector \bar{x} obeying the same partition of components as x . Finally, production costs are assumed to depend just on the considered unit but not on time, hence we collect them in a 4-dimensional cost vector c . Summarizing, our optimization problem writes as

$$\min c^T \left(\sum_{i=1}^{12} x_i^T \right) \quad \text{subject to} \quad x \leq \bar{x}, \quad \mathbb{P}(T(x)\xi \leq 0) \geq p.$$

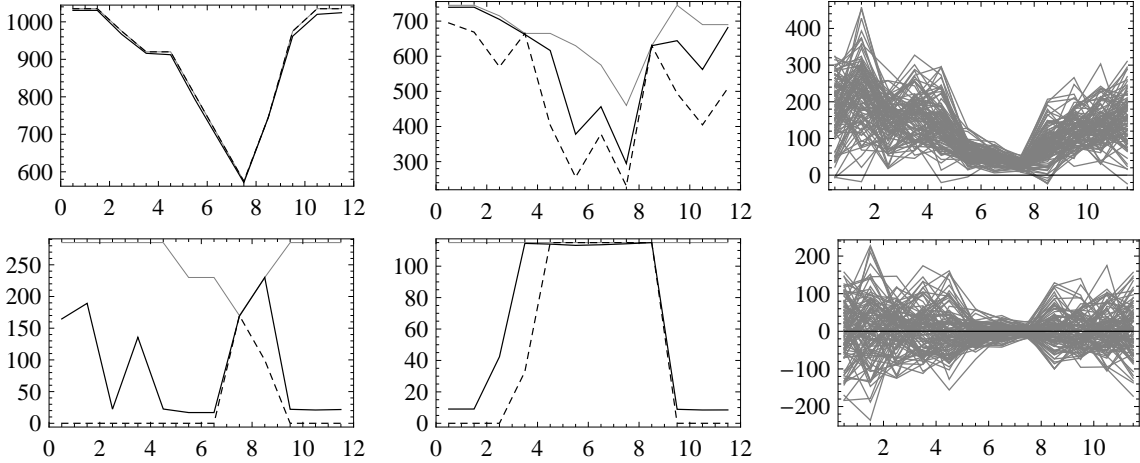


Figure 1: Illustration of results for an example (details see text)

The random vector is supposed to have a joint Gaussian distribution with parameters which we do not specify here. Components of $\hat{\xi}$ (availability coefficients for different units at different times) are assumed to be uncorrelated, whereas some correlation pattern is considered for the components of $\tilde{\xi}$ (demands at different times). The probability level is specified as $p = 0.9$. The dimensions of the problem are 48 for the decision and 60 for the random vector. The formulae for calculating ϕ and $\nabla\phi$ presented above are used to employ Genz's code inside numerical optimization. Results are presented in Figure . The first diagrams show the solutions for production profiles of the 4 units (left top: nuclear, middle top: coal, left bottom: fuel, middle bottom: gas). Black lines refer to the solutions of our probabilistic constrained problem, whereas dashed lines illustrate the corresponding solutions one would obtain upon replacing the random vector by its expectation (which allows a simple linear programming solution approach). Upper production limits are plotted as gray lines. While nuclear energy is fully employed in both solutions, the other units run on their maximum limits only for certain periods of time, where a certain complementarity between coal and gas becomes evident. Since the expected value solution does not take into account the random nature of availability coefficients and demands, it can afford lower production profiles leading to slightly smaller costs than the probabilistic solution (637.634 versus 738.915). On the other hand, there is a tremendous difference in the robustness of the solutions obtained. To illustrate this, 100 profiles for the overall random vector ξ were simulated according to the chosen distribution parameters and subjected to the optimal decisions obtained for the expected value and the probabilistic solution. The right diagrams illustrate the resulting differences between production and demand of energy (which should be non-negative). While, in good accordance with the chosen probability level of $p = 0.9$, there are only 12 out of 100 trajectories ever falling below zero during the whole time horizon in case of the probabilistic solution, all trajectories resulting from the expected value solution violate the required offer-demand relation at least

once in the considered period.

Acknowledgment: The authors would like to thank Prof. András Prékopa (RUTCOR, New Brunswick) for referring us to Gupta's formula which was crucial for our derivation.

References

- [1] W. van Ackooij, R. Henrion, A. Möller and R. Zorgati, On probabilistic constraints induced by rectangular sets and multivariate normal distributions, *Mathematical Methods of Operations Research*, 71 (2010) 535-549.
- [2] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics, Vol. 195, Springer, Dordrecht, 2009.
- [3] S. Gupta, Probability Integrals of Multivariate Normal and Multivariate t, *The Annals of Mathematical Statistics* 34 (1963), 792-828.
- [4] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.