Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

Website: http://AIMsciences.org

pp. **X–XX**

# GRADIENT ESTIMATES FOR GAUSSIAN DISTRIBUTION FUNCTIONS: APPLICATION TO PROBABILISTICALLY CONSTRAINED OPTIMIZATION PROBLEMS

René Henrion

Weierstrass Istitute Berlin
Mohrenstr. 39, 10117 Berlin, Germany

(Communicated by the associate editor name)

Abstract. We provide lower estimates for the norm of gradients of Gaussian distribution functions and apply the results obtained to a special class of probabilistically constrained optimization problems. In particular, it is shown how the precision of computing gradients in such problems can be controlled by the precision of function values for Gaussian distribution functions. Moreover, a sensitivity result for optimal values with respect to perturbations of the underlying random vector is derived. It is shown that the so-called maximal increasing slope of the optimal value with respect to the Kolmogorov distance between original and perturbed distribution can be estimated explicitly from the input data of the problem.

1. **Introduction.** Optimization problems with probabilistic constraints play an important role in engineering and operations research when decisions have to be taken under uncertainty affecting the given system of constraints. A typical class of such problems is given in the form

$$\min \left\{ c^T x | \mathbb{P}\left(Ax \geq \xi\right) \geq p \right\}. \tag{1}$$

Here, $x \in \mathbb{R}^n$ is a decision vector whose linear costs $c^T x$ are to be minimized subject to a probabilistic constraint. In this constraint, $\xi$ is some $s$-dimensional random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $A$ is a matrix of order $(s, n)$. In most cases, decisions $x$ have to be taken before realizations of the random vector $\xi$ are observed. Then, the random inequality system $Ax \geq \xi$ cannot serve as a constraint for a well defined optimization problem: whatever decision $x$ is taken, it may turn out to be unfeasible under a future realization of $\xi$ (in particular for unbounded distributions). Therefore, it is reasonable to declare $x$ to be feasible if the random constraint $Ax \geq \xi$ is satisfied under $x$ at a probability not smaller than a given level $p \in [0, 1]$. For introductory texts on probabilistically constrained optimization problems and their applications we refer to the monographs [14, 15, 17].

2000 *Mathematics Subject Classification.* 90C15; 90C31.

*Key words and phrases.* probabilistic constraints, chance constraints, stochastic optimization, Gaussian distribution function, sensitivity of optimal values.

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

For the numerical solution of (1) but also for theoretical issues like structure and stability it is essential not only to be able to calculate the probability function

$$x \mapsto \mathbb{P}(Ax \geq \xi) \tag{2}$$

but also its gradient. Note that, using the cumulative distribution function of $\xi$ defined as $F_\xi(y) := \mathbb{P}(y \geq \xi)$ one may rewrite (2) as the function $F_\xi(Ax)$, thus giving problem (1) the equivalent form

$$\min \left\{ c^T x | F_\xi(Ax) \geq p \right\}. \tag{3}$$

Evidently, for calculating values and gradients of $F_\xi(Ax)$ it is sufficient to be able to calculate values and gradients of $F_\xi$. As far as values are concerned, this can be done in moderate dimension $s$ at fairly satisfactroy precision for many prominent multivariate distributions, e.g. Gaussian, t-, Dirichlet, Gamma or Exponential dstribution (see, e.g., [1, 5, 12, 19, 20]). As far as gradients are concerned, several approaches have been proposed which allow, based on certain representations as surface or volume integrals, to approximate partial derivatives of probability functions via Monte Carlo and related techniques (e.g., [4, 11, 13, 21]). For a few special distributions (like Gaussian or Dirichlet) it is possible to establish an explicit relation between partial derivatives and function values of $F_\xi$ by means of an analytical formula. One may benefit from such comfortable situation in several respects:

- The same algorithm providing (approximate) values of $F_\xi$ can be employed for calculating (approximate) values of $\nabla F_\xi$.
- Higher order derivatives of $F_\xi$ can be obtained as well by recursively reducing their computation in an exact way to the computation of values of $F_\xi$.
- The precision of $\nabla F_\xi$ (and of higher order derivatives) can be controlled by that of $F_\xi$ itself.

The purpose of this paper is to show for the multivariate Gaussian distribution, how such reduction formula can be employed in order to estimate (independent of the concrete argument) the precision of $\nabla F_\xi$ by that of $F_\xi$ (Section 3) and, moreover, in order to derive sensitivity estimates for optimal values to problem (3) when perturbing the underlying distribution (Section 4). The latter issue is of interest, for instance, when replacing the theoretical, typically unknown distribution of $\xi$ by some empirical approximation). Both topics to be addresses rely essentially on the derivation of lower estimates for the norm of gradients of Gaussian distribution functions (Section 2).

The notation used is standard. $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and the maximum norm, respectively. The symbol $\xi \sim \mathcal{N}(\mu, \Sigma)$ expresses the fact that the random vector $\xi$ has a multivariate Gaussian distribution with mean vector $\mu$ and covariance matrix $\Sigma$.

## 2. Lower estimates for the norm of gradients of Gaussian distribution functions.

We commence this section by recalling a well-known gradient formula for Gaussian distribution functions. Without loss of generality, we may confine ourselves to the case of standard Gaussian distributions, i.e., where the mean vector equals zero and all variances (diagonal elements of the covariance matrix) are equal to one. Indeed, due to well-known transformation laws for Gaussian distributions, a general Gaussian distribution function $F_\xi$ with $\xi \sim \mathcal{N}(\mu, \Sigma)$, as it occurs, for instance, in (3), can be rewritten as

$$F_\xi(y) = \Phi^R\left(D^{-1/2}(y - \mu)\right), \tag{4}$$

where $\Phi^R$ is the distribution function of random vector distributed according to $\mathcal{N}(0,R)$, with $R$ being the correlation matrix associated with $\Sigma$. Moreover, $D$ denotes the diagonal part of $\Sigma$ (i.e., $D$ is the diagonal matrix composed of the variances of $\xi$). As a consequence, the probabilistically constrained optimization problem (1) can be equivalently rewriten via (3) and (4) as

$$\min\left\{c^T x | \Phi^R\left(D^{-1/2}\left(Ax-\mu\right)\right) \geq p\right\}. \tag{5}$$

In the present and in the following section, we shall assume our original problem (1) to be given in the form of (5). Consequently, in the vein of the previous discussion, the whole interest focusses on the gradient $\nabla\Phi^R$ of standard normal distribution functions.

For these reasons, let now $\Phi^R$ be the distribution function of an $s$-dimensional Gaussian random vector distributed according to $\mathcal{N}(0,R)$, with $R=(r_{i,j})_{i,j=1}^s$ and $r_{i,i}=1$ for $i=1,\ldots,s$. It is well known (see, e.g., [14], p.204) that $\Phi^R$ is a smooth function and that its partial derivatives calculate as

$$\frac{\partial\Phi^R}{\partial z_i}(z) = f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) \quad i=1,\ldots,s, \tag{6}$$

where $f$ is the density of the 1-dimensional standard Gaussian distribution,

$$\tilde{z}^{(i)} = \left(\frac{z_1 - r_{1,i}z_i}{\sqrt{1-r_{1,i}^2}},\cdots,\frac{z_{i-1}-r_{i-1,i}z_i}{\sqrt{1-r_{i-1,i}^2}},\frac{z_{i+1}-r_{i+1,i}z_i}{\sqrt{1-r_{i+1,i}^2}},\cdots,\frac{z_s-r_{s,i}z_i}{\sqrt{1-r_{s,i}^2}}\right) \tag{7}$$

and the correlation matrix $\tilde{R}^{(i)}$ of order $(s-1,s-1)$ has entries

$$\tilde{r}_{j,k}^{(i)} = \frac{r_{j',k'}-r_{j',i}r_{k',i}}{\sqrt{1-r_{j',i}^2}\sqrt{1-r_{k',i}^2}} \quad i=1,\ldots,s; \quad j,k=1,\ldots,s-1 \tag{8}$$

with

$$j' := \begin{cases} j \text{ if } j<i \\ j+1 \text{ if } j\geq i \end{cases}, \quad k' := \begin{cases} k \text{ if } k<i \\ k+1 \text{ if } k\geq i \end{cases}.$$

In the numerical solution of chance-constrained optimization problems, for instance when applying a supporting hyperplane method as in [18], one has to calculate gradients $\nabla\Phi^R(z)$ at arguments $z$ satisfying $\Phi^R(z)=p$ where $p\in[0,1]$ is some given probability level typically close to 1. Formula (6) permits an analytical reduction of gradients of Gaussian distribution functions to function values of the same type of distribution functions albeit with different parameters. As mentioned in the introduction, this fact has important consequences for numerics and sensitivity analysis in probabilistically constrained optimization problems. In this section, as a preparation of the following ones, we are interested in lower estimates for the norm of gradients of Gaussian distribution functions based on formula (6). More precisely, we are interested in estimates that depend only on the data ($R$,$p$,$s$) of the problem but not on the specific argument $z$ satisfying $\Phi^R(z)=p$ . Let us consider first the simple case of independently distributed components, i.e., $R=I_s$:

**Proposition 2.1.** *Let $p\geq 0.5$ and consider any point $z$ such that $\Phi^{I_s}(z)=p$. Then,*

$$\left\|\nabla\Phi^{I_s}(z)\right\|_\infty \geq f(q_{p^{1/s}})\cdot p^{1-1/s},$$

*where $f$ and $q_\alpha$ denote the density and the $\alpha$-quantile of the 1-dimensional standard Gaussian distribution, respectively.*

*Proof.* With $R = I_s$ one derives from (8) that $\tilde{R}^{(i)} = I_{s-1}$ for all $i = 1, \ldots, s$. Similarly, (7) yields that

$$\tilde{z}^{(i)} = (z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_s) \quad i = 1, \ldots, s.$$

Denoting by $F$ the cumulative 1-dimensional standard Gaussian distribution function, it follows that, for all $i = 1, \ldots, s$,

$$\begin{aligned} \Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) &= \Phi^{I_{s-1}}(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_s) = \prod_{j=1, j\neq i}^{s} F(z_j) \\ &= \Phi^{I_s}(z)/F(z_i) = p/F(z_i). \end{aligned}$$

As this estimate holds true for all $i = 1, \ldots, s$, we get from (8) that

$$\left\|\nabla \Phi^R(z)\right\|_\infty = \max_{i=1,\ldots,s} f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) = p \max_{i=1,\ldots,s} \frac{f(z_i)}{F(z_i)}. \tag{9}$$

Since $F$ is increasing as a distribution function and $f$ is decreasing for $p \geq 0.5$, one derives that $f/F$ is decreasing. Therefore,

$$\left\|\nabla \Phi^R(z)\right\|_\infty = p\frac{f(\tau)}{F(\tau)} \quad \text{with } \tau := \min_{i=1,\ldots,s} z_i.$$

We claim that $\tau \leq q_{p^{1/s}}$. Indeed, otherwise $z_i > q_{p^{1/s}}$ for all $i = 1, \ldots, s$, whence the contradiction

$$p = \Phi^{I_s}(z) = \prod_{i=1}^{s} F(z_i) > \prod_{i=1}^{s} F\left(q_{p^{1/s}}\right) = \prod_{i=1}^{s} p^{1/s} = p.$$

Now, exploiting once more the fact that $f/F$ is decreasing, we arrive at the desired lower estimate

$$\left\|\nabla \Phi^R(z)\right\|_\infty \geq p\frac{f\left(q_{p^{1/s}}\right)}{F\left(q_{p^{1/s}}\right)} = f(q_{p^{1/s}}) \cdot p^{1-1/s}.$$

□

The correlated case is less obvious and meaningful lower estimates for the precision of the gradient depending only on the data ($R$,$p$,$s$) seem to be very hard to derive for general correlation matrices. In the following, we identify a special class of correlation matrices for which such estimates are possible. To this aim, we associate with each correlation matrix $R$ the parameter

$$\Delta(R) := \rho_{\min} - \rho_1\rho_2,$$

where $\rho_1, \rho_2, \rho_{\min}$ denote the largest, second largest and smallest nondiagonal elements of $R$.

**Definition 2.2.** A correlation matrix is called amenable if $\Delta(R) \geq 0$.

We give examples for two subclasses of amenable correlation matrices:

**Example 2.3.** *Let $t \in [0, 1]$ be arbitrary. Assume that all nondiagonal entries of $R$ lie in the interval $[t^2, t]$. Then, $R$ is amenable. Indeed, by assumption, $\rho_{\min} \geq t^2$ and $\rho_1, \rho_2 \leq t$, whence $\Delta(R) \geq 0$. For example, $R$ is amenable if all nondiagonal correlation coefficients lie between 0.25 and 0.5.*

**Example 2.4.** *Assume that all or all but one nondiagonal entries of $R$ are constant and nonnegative. Then, $R$ is amenable. Indeed, by assumption, $\rho_{\min} = \rho_2$, whence $\Delta(R) = \rho_2 (1 - \rho_1) \geq 0$. For example, $R$ is amenable if one nondiagonal correlation coefficient equals 0.9 while all the others are equal to 0.1.*

**Proposition 2.5.** *If $R$ is amenable and positive definite, then $R \geq 0$.*

*Proof.* If not $R \geq 0$, then $\rho_{\min} < 0$. Assuming that $\rho_1, \rho_2$ have a common sign, one arrives at the contradiction $\Delta(R) < 0$. Otherwise, it follows that

$$1 > \rho_1 > 0 > \rho_2 \geq \rho_{\min},$$

where the first strict inequality relies on the fact that a correlation matrix cannot be positive definite if it contains an off-diagonal element equal to 1. From this chain of inequalities, we derive again the contradiction

$$\Delta(R) = \rho_{\min} - \rho_1 \rho_2 < \rho_{\min} - \rho_2 \leq 0.$$

□

The technical meaning of amenability is clarified in the following result:

**Lemma 2.6.** *If $R$ is an amenable, positive definite correlation matrix of order $(s, s)$, then $\tilde{R}^{(i)} \geq I_{s-1}$ (componentwise) holds true for the correlation matrices defined in (8) for all $i = 1, \ldots, s$.*

*Proof.* Fix an arbitrary $i \in \{1, \ldots, s\}$. By (8), $\tilde{r}^{(i)}_{j,j} = 1$ holds trivially true for all $j = 1, \ldots, s-1$. Next consider any off-diagonal element $\tilde{r}^{(i)}_{j,k}$ of the matrix $\tilde{R}^{(i)}$ (i.e., $j \neq k$). From the definition of the indices $j', k'$ in (8), it follows that $j' \neq k', j' \neq i$ and $k' \neq i$. Consequently, $r_{j',k'}, r_{j',i}$ and $r_{k',i}$ are off-diagonal elements of $R$. By definition, it holds that $r_{j',k'} \geq \rho_{\min}$. Due to Proposition 2.5, one has that $r_{j',i} \geq 0$ and $r_{k',i} \geq 0$. Moreover, by $j' \neq k'$, $r_{j',i}$ and $r_{k',i}$ are different elements of $R$. It follows that $r_{j',i} r_{k',i} \leq \rho_1 \rho_2$, whence, by amenability, $r_{j',k'} - r_{j',i} r_{k',i} \geq \Delta(R) \geq 0$. Therefore, $\tilde{r}^{(i)}_{j,k} \geq 0$ owing to (8). Summarizing, $\tilde{R}^{(i)} \geq I_{s-1}$. □

In order to prepare a result for the class of amenable correlation matrices, we define for each $i = 1, \ldots, s$ a correlation matrix $\Sigma^{(i)}$ of order $(i, i)$ having constant nondiagonal entries equal to $\rho_1$. Note that $\Sigma^{(i)}$ is positive semi-definite (as required for a correlation matrix) if $\rho_1 \geq 0$ (which is automatic if $R$ is amenable). From here we derive for each $i = 2, \ldots, s$ generalized quantiles of order $i$ as the unique values $\tau^{(i)}$ satisfying the equality

$$\Phi^{\Sigma^{(i)}} \left( \tau^{(i)}, \ldots, \tau^{(i)} \right) = p. \tag{10}$$

Here, uniqueness of the $\tau^{(i)}$ follows from the fact that a regular Gaussian distribution function is strongly increasing for arguments strictly inreasing in each componentNote that in the result of the following theorem the norm of the gradient is estimated by an expression depending exclusively on the original data $R$,$p$,$s$ but not on the specific argument $z$.

**Theorem 2.7.** *Let $R = (r_{i,j})_{i,j=1}^s$ be an amenable, positive definite correlation matrix with largest nondiagonal element $\rho_1$. Consider an arbitrary point $z$ such that $\Phi^R(z) = p \in [0.5, 1]$. Then,*

$$\left\|\nabla \Phi^R(z)\right\|_\infty \geq f(\tau) \prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right) \tag{11}$$

*where $f$ and $F$ refers to the density and distribution function, respectively, of the one-dimensional standard Gaussian distribution, the $\tau^{(j)}$ are defined in (10) and $\tau$ is the unique solution of the equation*

$$\Phi^R(\tau, \ldots, \tau) = p. \tag{12}$$

*Proof.* Define indices $i_1, \ldots, i_s$ in a way that $z_{i_1} \leq \cdots \leq z_{i_s}$. Since distribution functions are always dominated by their marginal distribution functions, it follows that

$$0.5 \leq p = \Phi^R(z) \leq F(z_i) \quad i = 1, \ldots, s.$$

Therefore,

$$0 \leq z_{i_1} \leq \cdots \leq z_{i_s}. \tag{13}$$

According to (7),

$$\tilde{z}^{(i_1)} = \left(\frac{z_1 - r_{1,i_1} z_{i_1}}{\sqrt{1-r_{1,i_1}^2}}, \cdots, \frac{z_{i_1-1} - r_{i_1-1,i_1} z_{i_1}}{\sqrt{1-r_{i_1-1,i_1}^2}}, \frac{z_{i_1+1} - r_{i_1+1,i_1} z_{i_1}}{\sqrt{1-r_{i_1+1,i_1}^2}}, \cdots, \frac{z_s - r_{s,i_1} z_{i_1}}{\sqrt{1-r_{s,i_1}^2}}\right).$$

Recall that $R \geq 0$ by Proposition 2.5. Now, (13) implies that

$$z_i - r_{i,i_1} z_{i_1} \geq (1 - r_{i,i_1}) z_i \quad \forall i \in \{1, \ldots, s\}.$$

Consequently,

$$\begin{aligned} \tilde{z}_i^{(i_1)} &\geq \frac{(1-r_{i,i_1}) z_i}{\sqrt{1-r_{i,i_1}^2}} \quad \forall i \in \{1, \ldots, i_1 - 1\} \\ \tilde{z}_i^{(i_1)} &\geq \frac{(1-r_{i+1,i_1}) z_{i+1}}{\sqrt{1-r_{i+1,i_1}^2}} \quad \forall i \in \{i_1, \ldots, s-1\}. \end{aligned}$$

Obviously, the correlation coefficients occuring in the above two expressions are off-diagonal, hence smaller than or equal to $\rho_1$. The function $t \mapsto (1-t)\left(1-t^2\right)^{-1/2}$ being decreasing, it follows that

$$\begin{aligned} \tilde{z}_i^{(i_1)} &\geq \frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}} z_i \quad \forall i \in \{1, \ldots, i_1 - 1\} \\ \tilde{z}_i^{(i_1)} &\geq \frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}} z_{i+1} \quad \forall i \in \{i_1, \ldots, s-1\}. \end{aligned}$$

$$\prod_{i=1}^{s-1} F\left(\tilde{z}_i^{(i_1)}\right) \geq \prod_{i=1, i\neq i_1}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}} z_i\right) = \prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}} z_{i_j}\right). \tag{14}$$

Now, let $j \in \{1, \ldots, s\}$ be arbitrarily given. Assume that $z_{i_j} < \tau^{(j)}$ with $\tau^{(j)}$ defined by (10). Then, by (13),

$$z_{i_k} < \tau^{(j)} \quad \forall k = 1, \ldots, j. \tag{15}$$

Denote by $\hat{R}$ the correlation matrix induced by components $i_1, \ldots, i_j$ (i.e., $\hat{R}$ is obtained from $R$ by selecting rows and columns corresponding to these indices). We exploit once more that distribution functions are dominated by their marginal distribution functions of any order and obtain that

$$p = \Phi^R(z) \leq \Phi^{\hat{R}}\left(z_{i_1}, \ldots, z_{i_j}\right) < \Phi^{\hat{R}}\left(\tau^{(j)}, \ldots, \tau^{(j)}\right), \tag{16}$$

where the strict inequality is a consequence of (15). Since $\hat{R} \leq \Sigma^{(j)}$ (componentwise) by definition of $\Sigma^{(j)}$, the well-known Slepian's inequality yields that

$$\Phi^{\hat{R}}\left(\tau^{(j)}, \ldots, \tau^{(j)}\right) \leq \Phi^{\Sigma^{(j)}}\left(\tau^{(j)}, \ldots, \tau^{(j)}\right) = p.$$

This is a contradiction with (16), whence

$$z_{i_j} \geq \tau^{(j)} \quad \forall j = 1, \ldots, s. \tag{17}$$

Combining this with (14), we have that

$$\tilde{z}_{i_j}^{(i_1)} \geq \frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)} \quad j = 2, \ldots, s.$$

Now, consider the matrix $\tilde{R}^{(i_1)}$ defined in (8). Since $\tilde{R}^{(i_1)} \geq I_{s-1}$ (componentwise) by Lemma 2.6, applying first Slepian's inequality again, and then taking into account (14) and (17), we arrive at

$$\Phi^{\tilde{R}^{(i_1)}}\left(\tilde{z}^{(i_1)}\right) \geq \Phi^{I_{s-1}}\left(\tilde{z}^{(i_1)}\right) = \prod_{i=1}^{s-1} F\left(\tilde{z}_i^{(i_1)}\right) \geq \prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right).$$

On the other hand, the standard Gaussian density being decreasing for positive arguments, (13) leads to

$$f\left(z_{i_1}\right) = \max_{i=1,\ldots,s} f(z_i),$$

whence

$$\begin{aligned} \left\|\nabla \Phi^R(z)\right\|_\infty &= \max_{i=1,\ldots,s} f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) \geq f\left(z_{i_1}\right)\Phi^{\tilde{R}^{(i_1)}}\left(\tilde{z}^{(i_1)}\right) \\ &\geq \left(\max_{i=1,\ldots,s} f(z_i)\right) \cdot \prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right). \end{aligned} \tag{18}$$

Next we claim that there exists some index $i^*$ such that $z_{i^*} \leq \tau$, where $\tau$ is defined in the assertion of our theorem. Indeed, otherwise $z_i > \tau$ for all $i$ and so we end up at the contradiction

$$p = \Phi^R(z) > \Phi^R(\tau, \ldots, \tau) = p.$$

Since $z_i \geq 0$ for all $i$ (see (13)) and $f$ is decreasing for nonnegative arguments, one derives that

$$\max_{i=1,\ldots,s} f(z_i) = f(\min_{i=1,\ldots,s} z_i) \geq f(\tau)$$

such that the assertion of the Theorem now follows from (18). □

3. **Precision of gradients of Gaussian distribution functions.** Although formula (6) establishes an analytical relation between partial derivatives and function values of Gaussian distribution functions, one has to take into account for numerical applications that function values themselves are already imprecise and can only be approximated by methods as described in [1, 5, 20]. Thanks to relation (6) one is not forced to further approximate partial derivatives by means of finite differences of function values already being imprecise. Rather, one may employ the same method to the computation of partial derivatives and thus avoid additional imprecision by the impact of another type of approximation. So, intuitively it is clear that in view of (6) the error of gradients can be controled by that of function values. Note that under the error of gradients we do not understand the norm of the absolute difference between theoretical and computed gradient but rather the norm of the difference between normalized theoretical and computed gradient:

$$\text{err}\ \left(\nabla\Phi^R(z)\right) := \left\| \frac{\nabla\Phi^R(z)}{\|\nabla\Phi^R(z)\|_\infty} - \frac{\left[\nabla\Phi^R(z)\right]^{\text{comp}}}{\left\|\left[\nabla\Phi^R(z)\right]^{\text{comp}}\right\|_\infty} \right\|_\infty .$$

(note that under the computed gradient $\left[\nabla\Phi^R(z)\right]^{\text{comp}}$ we understand the gradient obtained via the analytical relation (6) from computed function values $\left[\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right)\right]^{\text{comp}}$). The reason for this definition of error is that in a numerical context, the direction of a gradient is much more important than its norm. For instance, if gradients are used in a cutting plane method, the Walkup-Wets isometry theorem implies that the truncated Hausdorff distance between the cuts (halfspaces) defined by the gradients at a given point equals the truncated Hausdorff distance between the gradients themselves, which is exactly the error defined above. In other words, the error of gradients can be also interpreted as the error of the cuts induced by them. A first simple consequence of the triangle inequality yields the relation

$$\text{err}\ \left(\nabla\Phi^R(z)\right) \leq 2\frac{\left\|\nabla\Phi^R(z) - \left[\nabla\Phi^R(z)\right]^{\text{comp}}\right\|_\infty}{\|\nabla\Phi^R(z)\|_\infty}.$$

Using (6) and denoting by 'fv-err' the maximum absolute precision for the computation of function values for Gaussian distribution functions (predetermined by the user), one may continue via (6) as

$$\begin{aligned} \text{err}\ \left(\nabla\Phi^R(z)\right) &\leq 2\frac{\max\limits_{i=1,\ldots,s} f(z_i)\left|\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) - \left[\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right)\right]^{\text{comp}}\right|}{\max\limits_{i=1,\ldots,s} f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right)} \\ &\leq 2\frac{\text{fv-err}\cdot\max\limits_{i=1,\ldots,s} f(z_i)}{\max\limits_{i=1,\ldots,s} f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right)}. \end{aligned} \tag{19}$$

Now, we may invoke Proposition 2.1 in order to derive the following first estimate for the error of gradients in terms of the error of function values in the case of independent components:

**Corollary 3.1.** *Consider an arbitrary point $z$ such that $\Phi^{I_s}(z) = p$. Then,*

$$err\ \left(\nabla\Phi^{I_s}(z)\right) \leq 2\frac{fv\text{-}err}{p}. \tag{20}$$

*Proof.* One may continue (9) as

$$\max_{i=1,\dots,s} f(z_i)\Phi^{\tilde{R}^{(i)}}\left(\tilde{z}^{(i)}\right) \geq p \max_{i=1,\dots,s} f(z_i)$$

because of $F(z_i) \leq 1$. Now, the assertion follows from combining the previous relation with (19). □

Due to $p \approx 1$ in typical applications of chance constrained programming, the result of the Proposition yields that the error of the gradient is at most approximately double the chosen maximum error of function values in the case of independent components. Unfortuantely, this efficient estimate is of limited practical use because in the independent case both the function values and the gradients can be computed almost exactly as the calculus reduces to a multiplication of values of one-dimensional Gaussian distribution functions which is almost free of error. In the more interesting case of correlated components, we obtain the following Corollary by directly combining (18) with (19):

**Corollary 3.2.** *Consider an arbitrary point $z$ such that $\Phi^{I_s}(z) = p$. Then, under the assumptions of Theorem 2.7 it holds that*

$$err\left(\nabla\Phi^R(z)\right) \leq 2\frac{fv\text{-}err}{\prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right)}.$$

As an illustration, we consider the following example:

**Example 3.3.** *Let $p = 0.9$ and the following correlation matrix with band structure be given:*

$$R = \begin{pmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{pmatrix}.$$

*Then, $\rho_1 = 0.5$ and the correlation matrices $\Sigma^{(i)}$ defined before (10) get the form*

$$\Sigma^{(2)} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \Sigma^{(3)} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}, \Sigma^{(4)} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}.$$

*Now, the generalized quantiles defined in (10) and (12) are easily determined by using any code for the evaluation of multivariate Gaussian distribution function and bisecting on the values looked for as*

$$\tau^{(2)} = 1.577; \tau^{(3)} = 1.734; \tau^{(4)} = 1.838; \tau = 1.885.$$

*Then, $f(\tau) = 0.0675$ and*

$$\prod_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right) = 0.5896.$$

*Accordingly, Theorem 2.7 and Corollary 3.2 yield that*

$$\left\|\nabla\Phi^R(z)\right\|_\infty \geq 0.0398 \quad \text{and} \quad err\left(\nabla\Phi^R(z)\right) \leq 3.39 \cdot fv\text{-}err,$$

*respectively.*

4. **Sensitivity of optimal values to probabilistically constrained optimization problems.** When solving a probabilistically constrained optimization problem like (1), one is usually confronted with the fact that the underlying random vector $\xi$ and its distribution $\mathbb{P} \circ \xi^{-1}$ are not exactly known. Often, however, they may be approximated, for instance on the basis of historical observations, by some other random vector $\eta$ and then instead of the original problem (1) one solves the optimization problem

$$\min \left\{ c^T x | \mathbb{P}\left(Ax \geq \eta\right) \geq p \right\}. \tag{21}$$

This (inevitable) approximation step immediately raises the question about stability of the problem, in particular about sensitivity of its solution (set) and of its optimal value. For simplicity, let us confine ourselves in the following to the sensitivity of optimal values. Denote for an arbitrary random vector $\eta$ the optimal value of problem (21) by

$$\varphi\left(\eta\right) := \inf \left\{ c^T x | \mathbb{P}\left(Ax \geq \eta\right) \geq p \right\} \tag{22}$$

(note that an infimum occurs here because (21) does not necessarily have a solution). In particular, $\varphi\left(\xi\right)$ is the optimal value of the original problem (1). The stability of probabilistically constrained optimization problem has been analyzed in a fairly complete fashion in a series of papers [6, 7, 8, 9]. As pointed out in [16], stability of any stochastic optimization problem requires the choice of an appropriate probability metric adapted to the concrete problem structure and measuring the deviation between the underlying theoretical random vector $\xi$ and its approximation $\eta$. An appropriate choice in the context of our problem (1) (and its perturbation (21)) is the following discrepancy (see, e.g. [9]):

$$\Delta(\xi, \eta) := \sup_{z \in \mathbb{R}^s} \left| \mathbb{P}\left(z \geq \xi\right) - \mathbb{P}\left(z \geq \eta\right) \right| = \sup_{z \in \mathbb{R}^s} \left| F_\xi\left(z\right) - F_\eta\left(z\right) \right|, \tag{23}$$

where the second representation, adapted to the equivalent formulation (3) of the original problem, refers to the distribution function of the respective random vectors. This discrepancy is well-known as the Kolmogorov distance between the distributions $\mathbb{P} \circ \xi^{-1}$ and $\mathbb{P} \circ \eta^{-1}$ of $\xi$ and $\eta$. Note that the term distance relates just to the distributions but not to the random vectors themselves: $\Delta(\xi, \eta) = 0$ implies that $\mathbb{P} \circ \xi^{-1} = \mathbb{P} \circ \eta^{-1}$ (or $\xi \sim \eta$) but not necessarily $\xi = \eta$.

Using the discrepancy (23), the following more general result from [7] (Theorem 1) can be invoked for deriving the following result for our problem:

**Theorem 4.1.** *Let problem (1) be defined by a Gaussian random vector $\xi$ and have a nonempty and bounded solution set. Moreover assume that there exists a point $x^*$ such that $\mathbb{P}\left(Ax^* \geq \xi\right) > p$. Then, there exist constants $L, \delta > 0$ such that*

$$\left|\varphi\left(\xi\right) - \varphi\left(\eta\right)\right| \leq L \Delta(\xi, \eta)$$

*for all random vectors $\eta$ with $\Delta(\xi, \eta) < \delta$.*

Note, that all assumptions of this Theorem refer to the original problem and no assumptions are made on the perturbed problem. In particular, the perturbed random vector $\eta$ does not have to have a nice distribution like the Gaussian one (implying differentiability and convexity properties) but could follow, for instance, a discrete distribution coming from an empirical approximation of $\xi$. Although the result of the Theorem is nice in that it identifies a Lipschitz-like behavior of optimal values under perturbation, its practical use is limited due to the fact that the constant $L$ is hard to be eplicitly quantified from the data of the original problem. The following Theorem is, to the best of our knowledge, the first attempt to do so in

a concrete setting. In order to keep its presentation sufficiently simple, we confine ourselves to the weaker property of upper Lipschitz continuity of optimal values.

**Proposition 4.2.** *In problem (1), let $\xi$ be an $s$-dimensional Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$. Assume that the problem has a solution. Let $R$ be the correlation matrix associated with $\Sigma$. Finally, let $\varkappa > 0$ be given such that*

$$\left\|\nabla \Phi^R(y)\right\|_\infty \geq \varkappa \quad \forall y : \Phi^R(y) = p, \tag{24}$$

*where $\Phi^R$ is the distribution function according to $\mathcal{N}(0, R)$ (see (4)).*

*Then, for each $\varepsilon \in (0, \varkappa)$ there is some $\delta > 0$ with*

$$\varphi(\eta) \leq \varphi(\xi) + \frac{\|c\| \left\|A^T (AA^T)^{-1}\right\|_2}{\left(\varkappa \Big/ \max\limits_{i=1,\ldots,s} \sqrt{\Sigma_{ii}}\right) - \varepsilon} \cdot \Delta(\xi, \eta)$$

*for all random vectors $\eta$ with $\Delta(\xi, \eta) < \delta$. Here, $\|\cdot\|_2$ refers to the 2-norm of matrices.*

*Proof.* We consider problem (1) in its equivalent form (3). Let $\bar{x}$ be one of its solutions (whose existence is assumed). In particular,

$$\varphi(\xi) = c^T \bar{x}. \tag{25}$$

Since the objective of this problem is linear, it follows that $F_\xi(A\bar{x}) = p$. The correlation matrix $R$ associated with the covariance matrix $\Sigma$ computes according to $R = D^{-1/2} \Sigma D^{-1/2}$, where $D$ is a diagonal matrix with $D_{ii} = \Sigma_{ii}$. Hence, (4) implies that $\Phi^R\left(D^{-1/2}(A\bar{x} - \mu)\right) = p$. Then, (24) ensures that

$$\left\|\nabla \Phi^R(D^{-1/2}(A\bar{x} - \mu))\right\|_\infty \geq \varkappa. \tag{26}$$

Derivation of (4) yields that

$$\begin{aligned}
\|\nabla F_\xi(A\bar{x})\|^2 &= \nabla \Phi^R\left(D^{-1/2}(A\bar{x} - \mu)\right) D^{-1} \left[\nabla \Phi^R\left(D^{-1/2}(A\bar{x} - \mu)\right)\right]^T \\
&= \sum_{i=1}^{s} \frac{1}{D_{ii}} \left[\frac{\partial \Phi^R}{\partial y_i}\left(D^{-1/2}(A\bar{x} - \mu)\right)\right]^2 \\
&\geq \frac{1}{\max\limits_{i=1,\ldots,s} D_{ii}} \left\|\nabla \Phi^R\left(D^{-1/2}(A\bar{x} - \mu)\right)\right\|^2.
\end{aligned}$$

Taking into account (26), one arrives at

$$\|\nabla F_\xi(A\bar{x})\| \geq \tilde{\varkappa} := \frac{\varkappa}{\max\limits_{i=1,\ldots,s} \sqrt{\Sigma_{ii}}}. \tag{27}$$

For any $t \in \mathbb{R}$ we have that

$$\begin{aligned}
F_\xi(A\bar{x} + t\nabla F_\xi(A\bar{x})) &= F_\xi(A\bar{x}) + \langle \nabla F_\xi(A\bar{x}), t\nabla F_\xi(A\bar{x})\rangle + o(t) \\
&= p + t\|\nabla F_\xi(A\bar{x})\|^2 + o(t).
\end{aligned}$$

Now, let any $\varepsilon \in (0, \varkappa)$ be given. We choose $\tilde{\delta} > 0$ such that

$$o(t) \geq -\varepsilon \|\nabla F_\xi(A\bar{x})\| t \quad \forall t \in [0, \tilde{\delta}).$$

Then,

$$F_\xi(A\bar{x} + t\nabla F_\xi(A\bar{x})) \geq p + t\|\nabla F_\xi(A\bar{x})\| (\|\nabla F_\xi(A\bar{x})\| - \varepsilon) \quad \forall t \in [0, \tilde{\delta}). \tag{28}$$

We further put $\delta := \tilde{\varkappa}(\tilde{\varkappa} - \varepsilon)\tilde{\delta}$. Now, let any random vector $\eta$ be given such that $\Delta(\xi, \eta) < \delta$. From (27) it follows that

$$\bar{t} := \frac{\Delta(\xi, \eta)}{\|\nabla F_\xi(A\bar{x})\| \, (\|\nabla F_\xi(A\bar{x})\| - \varepsilon)} < \tilde{\delta}$$

Then, (28) yields that

$$F_\xi(A\bar{x} + \bar{t}\nabla F_\xi(A\bar{x})) \geq p + \Delta(\xi, \eta).$$

With

$$u := \bar{t}A^T(AA^T)^{-1}\nabla F_\xi(A\bar{x}) \tag{29}$$

we infer that

$$F_\xi(A(\bar{x} + u)) \geq p + \Delta(\xi, \eta),$$

whence by (23),

$$F_\eta(A(\bar{x} + u)) \geq F_\xi(A(\bar{x} + u)) - \Delta(\xi, \eta) \geq p.$$

Along with (22), this implies that $\varphi(\eta) \leq c^T(\bar{x} + u)$. Moreover, by (29), (27) and definition of $\bar{t}$, one has that

$$\|u\| \leq \bar{t} \left\|A^T(AA^T)^{-1}\right\|_2 \|\nabla F_\xi(A\bar{x})\| \leq \frac{\left\|A^T(AA^T)^{-1}\right\|_2}{\tilde{\varkappa} - \varepsilon}\Delta(\xi, \eta).$$

With (25) it follows that

$$\varphi(\eta) - \varphi(\xi) \leq c^T(\bar{x} + u) - c^T\bar{x} \leq \frac{\|c\| \left\|A^T(AA^T)^{-1}\right\|_2}{\tilde{\varkappa} - \varepsilon}\Delta(\xi, \eta).$$

Referring back to the definition of $\tilde{\varkappa}$ in (27), this proves the proposition. □

Observe that the result of this proposition makes a statement on the sensitivity of optimal values in problem (1) that depends on the norm $\|c\|$ of the linear objective functional although the solution of the problem itself is not affected by this norm but just by the direction of $c$. In order to arrive at a sensitivity result independent of this norm, one might either suppose without loss of generality, that $\|c\| = 1$ in problem (1) or alternatively pass to the sensitivity of relative (with respect to the solution) optimal values.

It is interesting to observe a close relation of the statement of Proposition 4.2 with the so-called *slope* of functions in metric spaces as introduced in [2]. More precisely for a metric space $M$, a function $h : M \to \mathbb{R}$ and a point $x \in M$, the slope of $h$ at $x$ is defined as

$$|\nabla h(x)| := \limsup_{u \to x, u \neq x} \frac{[h(x) - h(u)]_+}{d(x, u)}. \tag{30}$$

The notation is justified by the fact that in a differentible setting $|\nabla h(x)|$ would correspond to the norm of the gradient of $h$ at $x$. Of course, in a differentible setting one could replace $h$ by $-h$ without changing the norm of the gradient. In the nondiffernetiable case, however, it makes a big difference to do so. Then, (30) corresponds to a maximum local decrease of $h$ at $x$. Therefore, $|\nabla h(x)|$ is also sometimes referred to as the *maximal decreasing slope* and denoted by $|\nabla^- h(x)|$. This concept plays an important role in the theory of gradient flows or in stability (metric regularity, error bounds) of set-valued analysis (e.g., [3, 10]). On the other hand, as in the case of our sensitivity analysis of optimal values, it is often important to also have information about the *maximal increasing slope*, which would be defined in a natural way as

$$\left|\nabla^+ h(x)\right| := \limsup_{u\to x, u\neq x} \frac{[h(u)-h(x)]_+}{d(x,u)}.$$

Now, disregarding the negligible fact that (23) defines just a semi-metric but not a metric on the set of $s$-dimensional random vectors, it is natural to introduce the maximal increasing slope for the optimal value function $\varphi$ at $\xi$ as

$$\left|\nabla^+ \varphi(\xi)\right| := \limsup_{\Delta(\xi,\eta)\to 0, \xi\neq\eta} \frac{[\varphi(\eta)-\varphi(\xi)]_+}{\Delta(\xi,\eta)}.$$

Then, we derive the following result as an mmediate consequence of Proposition 4.2:

**Theorem 4.3.** *Under the assumptions of Proposition 4.2, one has that*

$$\left|\nabla^+ \varphi(\xi)\right| \leq \frac{\|c\| \left\|A^T(AA^T)^{-1}\right\|_2}{\left(\varkappa \Big/ \max\limits_{i=1,\ldots,s} \sqrt{\Sigma_{ii}}\right)}.$$

Now, in order to extract explicit quantitative information from Theorem 4.3, it is essential to have a concrete value of $\varkappa$ which can be directly computed from the input data of the problem. But, recalling from the statement of Proposition 4.2 that $\varkappa$ is a lower estimate for the (maximum) norm of the gradient $\nabla\Phi^R(y)$ at a point $y$ satisfying $\Phi^R(y) = p$, it is exactly the results from Section 2 (Proposition 2.1 and Theorem 2.7) which provide us which such desired concrete value of $\varkappa$. Combination with Theorem 4.3 yields the following two corollaries.

**Corollary 4.4.** *In problem (1), let $\xi$ be an $s$-dimensional Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$ where $\Sigma$ is diagonal (i.e., the components of $\xi$ are independent). Assume that the problem has a solution. Then,*

$$\left|\nabla^+ \varphi(\xi)\right| \leq \frac{\|c\| \left\|A^T(AA^T)^{-1}\right\|_2}{\left(f(q_{p^{1/s}}) \cdot p^{1-1/s} \Big/ \max\limits_{i=1,\ldots,s} \sqrt{\Sigma_{ii}}\right)},$$

*where $f$ and $q_\alpha$ denote the density and the $\alpha$-quantile of the 1-dimensional standard Gaussian distribution, respectively.*

**Corollary 4.5.** *In problem (1), let $\xi$ be an $s$-dimensional Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$ where the correlation matrix $R$ associated with $\Sigma$ is amenable in the sense of Definition 2.2. Assume that the problem has a solution. Then,*

$$\left|\nabla^+ \varphi(\xi)\right| \leq \frac{\|c\| \left\|A^T(AA^T)^{-1}\right\|_2}{\left(f(\tau) \prod\limits_{j=2}^{s} F\left(\frac{1-\rho_1}{\sqrt{1-(\rho_1)^2}}\tau^{(j)}\right) \Big/ \max\limits_{i=1,\ldots,s} \sqrt{\Sigma_{ii}}\right)},$$

*where the meaning of $f, F, \rho_1, \tau, \tau^{(j)}$ is as in Theorem 2.7.*

The following example illustrates Corollary 4.4:

**Example 4.6.** *In problem (1), let $s = 10, p = 0.9, \Sigma = I_s, A = I_n$ and $c = (1,\ldots,1)/\|1,\ldots,1\|$. Then it is easily seen that problem (1) has a solution and that*

$$\|c\| = \left\|A^T(AA^T)^{-1}\right\|_2 = \max_{i=1,\ldots,s} \sqrt{\Sigma_{ii}} = 1, p^{1-1/s} = 0.9095, q_{p^{1/s}} = 2.309, f(q_{p^{1/s}}) = 0.0277.$$

*Consequently, by Corollary 4.4,* $|\nabla^+\varphi(\xi)| \leq 39.6$.

A similar example could be constructed to illustrate Corollary 4.5 as well.

## REFERENCES


[1] I. Deák, *Subroutines for Computing Normal Probabilities of Sets - Computer Experiences*, Ann. Oper. Res., **100** (2000), 103-122.

[2] E. De Giorgi, A. Marino and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis .Mat. Natur. (8), **68** (1980), 180-187.

[3] M. Fabian, R. Henrion, A. Kruger and J. Outrata, *Error bounds: necessary and sufficient conditions*, Set-Valued Var. Anal., **18** (2010) 121-149.

[4] J. Garnier, A. Omrane and Y. Rouchdy, *Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations*, European J. Oper. Res., **198** (2009), 848-858.

[5] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statist., **195** (2009).

[6] R. Henrion and W. Römisch, *Metric regularity and quantitative stability in stochastic programs with probabilistic constraints* , Math. Program. **84** (1999), 55-88.

[7] R. Henrion, *Qualitative stability of convex programs with probabilistic constraints* , Lecture Notes in Econom. and Math. Systems, **481** (2000) 164-180.

[8] R. Henrion, *Perturbation Analysis of Chance-Constrained Programs under variation of all constraint data*, Lecture Notes in Econom. and Math. Systems, **532** (2004), 257-274.

[9] R. Henrion and W. Römisch,*Hölder and Lipschitz Stability of Solution Sets in Programs with Probabilistic Constraints*, Math. Program., **100** (2004), 589-611.

[10] A.D. Ioffe, *Metric regularity and subdifferential calculus*, Russian Math. Surveys **55** (2000), 501-558.

[11] K. Marti, *Differentiation of Probability Functions: The Transformation Method*, Comput. Math. Appl., **30** (1995), 361-382.

[12] N. Olieman and B. van Putten, *Estimation method of multivariate exponential probabilities based on a simplex coordinates transform*, J. Stat. Comput. Simul. **80** (2010), 355-361.

[13] G. Pflug and H. Weisshaupt, *Probability gradient estimation by set-valued calculus and applications in network design*, SIAM J. Optim. **15** (2005), 898-914.

[14] A. Prékopa, "Stochastic Programming," Kluwer, Dordrecht, 1995.

[15] A. Prékopa, *Probabilistic Programming*, in "Stochastic Programming" (eds. A. Ruszczyński and A. Shapiro), Handbooks in Operations Research and Management Science, **10**, Elsevier, (2003), 267-351.

[16] W. Römisch, *Stability of Stochastic Programming Problems*, in "Stochastic Programming" (eds. A. Ruszczyński and A. Shapiro), Handbooks in Operations Research and Management Science, **10**, Elsevier, (2003), 483-554.

[17] A. Shapiro, D. Dentcheva and A. Ruszczyński, "Lectures on Stochastic Programming", MPS/SIAM Ser. Optim., **9**, 2009.

[18] T. Szántai, *A Computer Code for Solution of Probabilistic Constrained Stochastic Programming Problems*, in "Numerical Techniques for Stochastic Optimization" (eds. Y. Ermoliev and R. J.-B. Wets), Springer-Verlag, (1988), 229–235.

[19] T. Szántai, *Evaluation of a special multivariate Gamma distribution*, Math. Program. Study **27** (1996), 1-16.

[20] T. Szántai, *Improved Bounds and Simulation Procedures on the Value of the Multivariate Normal Probability Distribution Function*, Ann. Oper. Res. **100** (2000), 85-101.

[21] S. Uryasev, *Derivatives of Probability Functions and some Applications*, Ann. Oper. Res. **56** (1995), 287-311.


Received xxxx 20xx; revised xxxx 20xx.

*E-mail address*: henrion@wias-berlin.de