

Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions*

René Henrion

Weierstrass Institute Berlin, 10117 Berlin, Germany

Werner Römisch

Humboldt-University Berlin, Institute of Mathematics,
10099 Berlin, Germany

January 7, 2005

Abstract

The paper provides a condition for differentiability as well as an equivalent criterion for Lipschitz continuity of singular normal distributions. Such distributions are of interest, for instance, in stochastic optimization problems with probabilistic constraints, where a comparatively small (nondegenerate-) normally distributed random vector induces a large number of linear inequality constraints (e.g. networks with stochastic demands). The criterion for Lipschitz continuity is established for the class of quasi-concave distributions which the singular normal distribution belongs to.

Key Words. quasi-concave measures, singular normal distributions, Lipschitz continuity, differentiability, stochastic optimization, probabilistic constraints

AMS Classification. 90C15

1 Introduction

An m -dimensional random vector η is said to have a singular normal distribution if there exists some s -dimensional random vector ξ having a nondegenerate normal distribution such that

$$\eta = A\xi + b,$$

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

where A is an (m, s) -matrix with rank smaller than m and b is an m -vector. In particular, one may choose $A = 0$ to see that the Dirac measure, placing mass one at the point b , has a singular normal distribution. More generally, singular normal distributions are those normal distributions whose covariance matrix has a rank strictly smaller than the dimension of the random vector.

Such seemingly artificial distributions arise in a natural way in problems of stochastic optimization, where a relatively small (nondegenerate-) normally distributed random vector induces a large number of linear inequality constraints. As an example, consider the problem of optimal capacity expansion in a network with stochastic demands (see [9], p. 453). Let the random vector ξ represent the demands in the nodes of the network and let x be a vector of capacities for the arcs in the network. The costs of installing these capacities are to be minimized as a function of x under the constraint that there exists a flow through the network which is feasible at high probability, i.e., which satisfies both the capacity restrictions along the arcs and the random demands in the nodes (at high probability). Using the *Gale-Hoffman* theorem, feasibility can be modeled as a linear relation

$$A\xi \leq Bx.$$

Taking into account the random character of ξ , it makes sense to require feasibility in a probabilistic sense:

$$P(A\xi \leq Bx) \geq p,$$

where P denotes probability and $p \in [0, 1]$ is some chosen level of reliability. In general, the sizes of A and B can be drastically reduced by eliminating redundancy etc. Nevertheless, even the reduced systems may contain a number of inequalities which is considerably larger than the dimension of ξ (number of nodes). Passing to the transformed random vector $\eta = A\xi$, the probabilistic constraint obtained above can be rewritten as

$$\Phi(Bx) \geq p,$$

where Φ is the distribution function of η . However, since A may have more rows than columns, we have to expect that η has a singular normal distribution even though ξ had a regular normal distribution.

The example shows that, in order to cope with certain types of probabilistic constraints, it is important to be able to calculate values and gradients of singular normal distribution functions. As the latter need not exist in general, it is of interest to characterize differentiability of such functions. If differentiability fails to hold, one could rely on more general tools from nonsmooth optimization (both for algorithmic purposes and optimality conditions). In such constellation, local or global Lipschitz continuity is a favorable property. Whether a singular normal distribution function is discontinuous or not does not depend on the rank of the covariance matrix. Figure 1 shows (from the left to the right) the distribution functions of 2-dimensional normal distributions with zero mean and covariance matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

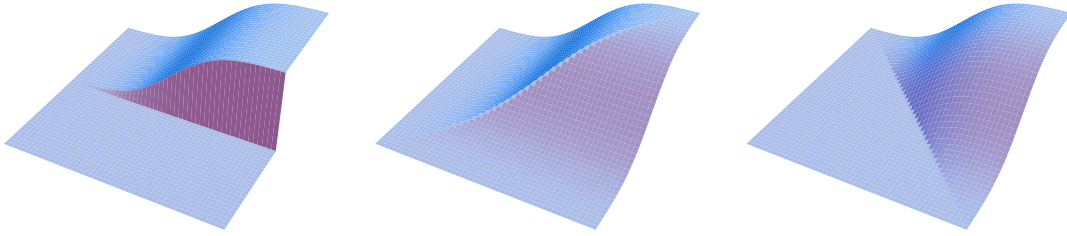


Figure 1: Distribution functions of 2-dimensional singular normal distributions with covariance matrix having rank one (see text).

all of which have rank one. Note that, in the first case, the distribution function is discontinuous whereas it is Lipschitz continuous (piecewise selection of smooth functions of min- and max-type, respectively) in the remaining cases.

The paper provides a condition for differentiability as well as an equivalent criterion for Lipschitz continuity of singular normal distribution functions. The criterion for Lipschitz continuity can be obtained for the general class of quasi-concave distributions which singular normal distributions belong to.

2 Lipschitz continuity of quasi-concave distributions

We start this section by introducing the class of quasi-concave probability measures (see [9]). By $\mathcal{P}(\mathbb{R}^s)$ we denote the set of probability measures on \mathbb{R}^s .

Definition 2.1 *A probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ is called quasi-concave whenever*

$$\mu(\lambda A + (1 - \lambda)B) \geq \min\{\mu(A), \mu(B)\}$$

holds true for all convex and Borel measurable subsets $A, B \subseteq \mathbb{R}^s$ and all $\lambda \in [0, 1]$ such that $\lambda A + (1 - \lambda)B$ is Borel measurable.

It is well known that a large class of prominent multivariate distributions shares the property of being quasi-concave. Among those are the multivariate normal distribution (nondegenerate or singular), the Dirichlet-, Pareto-, Gamma-, Log-normal distributions (possibly with a restricted range of parameters) as well as uniform distributions over compact, convex subsets of \mathbb{R}^s (see [9], [3]). Consequently, all future statements in this section apply in particular to singular normal distributions.

For the proof of our Lipschitz criterion, we shall make use of the following three propositions:

Proposition 2.1 *A quasiconcave measure $\mu \in \mathcal{P}(\mathbb{R})$ has either a density or coincides with some Dirac measure, i.e. $\mu = \delta_x$ for some $x \in \mathbb{R}$.*

Proof. Follows immediately from Theorem 3.2 in [3] ■

Proposition 2.2 *If for all marginal distributions μ_i of $\mu \in \mathcal{P}(\mathbb{R}^s)$ there exist bounded densities on \mathbb{R} , then the distribution function F_μ of μ is Lipschitz continuous.*

Proof. See Prop. 3.8 in [10] ■

Proposition 2.3 *If $\mu \in \mathcal{P}(\mathbb{R})$ is a quasiconcave measure with density f_μ , then f_μ is bounded.*

Proof. According to Theorem 3.2 in [3], the possibly extended-valued function $1/f_\mu$ is convex and the support of μ is a convex subset of \mathbb{R} . Assuming that f_μ is unbounded, there exists a sequence $\{x_n\} \subseteq \mathbb{R}$ such that $f_\mu(x_n) \geq n$. If $\{x_n\}$ is unbounded, then, without loss of generality, it is increasing, hence $[x_1, \infty) \subseteq \text{supp } \mu$ and $\{1/f_\mu(x_n)\}$ is decreasing. Since $1/f_\mu$ is convex, it follows that $1/f_\mu$ is decreasing on $[x_1, \infty)$. Therefore, f_μ is increasing on $[x_1, \infty)$ which contradicts the fact that f_μ is a density. Now, assume that $\{x_n\}$ is bounded, hence $x_n \rightarrow \bar{x}$ upon passing to some subsequence. Then, $1/f_\mu(\bar{x}) = 0$. Indeed, this follows in case of $\bar{x} \in \text{int supp } \mu$ from the continuity of the convex function $1/f_\mu$ on the interior of its domain. In case that \bar{x} belongs to the boundary of $\text{supp } \mu$, we may redefine $f_\mu(\bar{x}) := \infty$ without changing the measure μ and without affecting the convexity of $1/f_\mu$ (due to $1/f_\mu(x_n) \rightarrow 0$). Now, from $1/f_\mu \geq 0$ being convex and satisfying $1/f_\mu(\bar{x}) = 0$, it follows that $1/f_\mu(\bar{x} + h)$ is nondecreasing for $h > 0$ and that the difference quotients

$$h \mapsto h^{-1}(1/f_\mu(\bar{x} + h) - 1/f_\mu(\bar{x}))$$

are nondecreasing in h . Consequently, one has for $h_2 \geq h_1 > 0$

$$f_\mu(\bar{x} + h_1) \geq f_\mu(\bar{x} + h_2) \tag{1}$$

$$f_\mu(\bar{x} + h_1)h_1 \geq f_\mu(\bar{x} + h_2)h_2 \tag{2}$$

We assume that either $\bar{x} \in \text{int supp } \mu$ or that \bar{x} belongs to the left boundary of $\text{supp } \mu$ (the proof running analogously in case that \bar{x} belongs to the right boundary of $\text{supp } \mu$). In both cases there exists some $\delta > 0$ such that $f_\mu(\bar{x} + \delta) > 0$. It follows for arbitrary $n \in \mathbb{N}$ that

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_\mu(x) dx \geq \int_{\bar{x}+2^{-n}\delta}^{\bar{x}+\delta} f_\mu(x) dx = \sum_{j=0}^{n-1} \int_{\bar{x}+2^{-(j+1)}\delta}^{\bar{x}+2^{-j}\delta} f_\mu(x) dx \\ &\geq \sum_{j=0}^{n-1} f_\mu(\bar{x} + 2^{-j}\delta) 2^{-j} \frac{\delta}{2} \quad (\text{by (1)}) \\ &\geq \sum_{j=0}^{n-1} f_\mu(\bar{x} + \delta) \frac{\delta}{2} \quad (\text{by (2)}) \\ &= n \frac{\delta}{2} f_\mu(\bar{x} + \delta). \end{aligned}$$

This, however, is a contradiction to

$$n \frac{\delta}{2} f_{\mu}(\bar{x} + \delta) \rightarrow_n \infty.$$

■

For the narrower class of log-concave measures, Proposition 2.3 is a (1-dimensional) special case of a Theorem by Barndorff-Nielsen ([2]).

Definition 2.2 We call a subset $H \subseteq \mathbb{R}^s$ a canonical hyperplane if there exist $t \in \mathbb{R}$ and $i \in \{1, \dots, s\}$ such that

$$H = \mathbb{R}^1 \times \dots \times \mathbb{R}^{i-1} \times \{t\} \times \mathbb{R}^{i+1} \times \dots \times \mathbb{R}^s.$$

Now, we are in a position to formulate the desired criterion for Lipschitz continuity of distribution functions in the considered class of distributions:

Theorem 2.1 A quasiconcave probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ has a Lipschitz continuous distribution function F_{μ} if and only if the support of μ is not contained in a canonical hyperplane of \mathbb{R}^s .

Proof. We denote by $\mu_i \in \mathcal{P}(\mathbb{R})$ the i -th marginal distribution of μ . Clearly, the μ_i are quasiconcave on \mathbb{R} . With T being the support of μ and δ_t referring to the one-dimensional Dirac measure placed at $t \in \mathbb{R}$, the following chain of equivalences results:

$$\begin{aligned} T \text{ is contained in a canonical hyperplane of } \mathbb{R}^s &\iff \\ \exists t \in \mathbb{R} \exists i \in \{1, \dots, s\} : \mu(\mathbb{R}^1 \times \dots \times \mathbb{R}^{i-1} \times \{t\} \times \mathbb{R}^{i+1} \times \dots \times \mathbb{R}^s) = 1 &\iff \\ \exists t \in \mathbb{R} \exists i \in \{1, \dots, s\} : \mu_i(\{t\}) = 1 &\iff \\ \exists t \in \mathbb{R} \exists i \in \{1, \dots, s\} : \mu_i = \delta_t &\iff \\ \exists i \in \{1, \dots, s\} : \mu_i \text{ does not have a density.} & \end{aligned}$$

Here, the last equivalence is implied by Prop. 2.1. Contraposition gives the following chain of implications with the second and third one following from Propositions 2.3 and 2.2, respectively.

$$\begin{aligned} T \text{ is not contained in any canonical hyperplane of } \mathbb{R}^s &\implies \\ \mu_i \text{ has a density } f_{\mu_i} \text{ for all } i \in \{1, \dots, s\} &\implies \\ f_{\mu_i} \text{ is bounded for all } i \in \{1, \dots, s\} &\implies \\ F_{\mu} \text{ is globally Lipschitzian.} & \end{aligned}$$

Now, this chain of implications proves the 'if'-part of the theorem. For the reverse direction, assume that T is contained in a canonical hyperplane of \mathbb{R}^s . Then, the above chain of equivalences shows that

$$\mu(\mathbb{R}^1 \times \dots \times \mathbb{R}^{i-1} \times \{t\} \times \mathbb{R}^{i+1} \times \dots \times \mathbb{R}^s) = 1 \quad \text{for some } t \in \mathbb{R} \text{ and } i \in \{1, \dots, s\}.$$

Consequently, one may choose some $\tau \in \mathbb{R}$ large enough such that

$$F_\mu(\tau, \dots, \tau, t, \tau, \dots, \tau) > 0.$$

On the other hand, $F_\mu(\tau, \dots, \tau, t', \tau, \dots, \tau) = 0$ for any $t' < t$, hence F_μ is not continuous (much less it is Lipschitz continuous). ■

The last argument in the proof of Theorem 2.1 shows that the failure of Lipschitz continuity entails the failure of continuity, so we get the following useful observation:

Corollary 2.1 *The distribution function of some quasiconcave probability measure is Lipschitz continuous if and only if it is continuous. In particular, the distribution function of some quasiconcave probability measure with density is Lipschitz continuous.*

Concerning the second statement of the last corollary, we emphasize that in general even the existence of a bounded and continuous density does not imply the Lipschitz continuity of the distribution function (for a counterexample see [7], Ex. 9). A slightly more illustrative reformulation of Theorem 2.1 is:

Theorem 2.2 *Let ξ be an s -dimensional random vector with quasi-concave distribution $\mu \in \mathcal{P}(\mathbb{R}^s)$. Then, the distribution function of ξ is Lipschitz continuous if and only if none of the components ξ_i has zero variance.*

As an application of Theorem 2.2 we come back to the singular normal distributions with the three covariance matrices mentioned in the introduction. The first covariance matrix contains a zero diagonal element whereas the second and third ones do not. This explains why the first distribution function depicted in Figure 1 is discontinuous whereas the second and third ones are Lipschitz continuous.

At the end of this section, we consider an application to probability functions

$$\varphi(x) = P(A\xi \leq h(x)), \quad (3)$$

where, ξ is an s -dimensional random vector, A is an (m, s) -matrix, $x \in \mathbb{R}^n$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Recall that such type of probability functions arises in the context of probabilistic constraints $\varphi(x) \geq p$ as presented in the introduction.

Corollary 2.2 *In (3), assume that h is locally Lipschitzian and that ξ has a quasi-concave distribution with some covariance matrix Σ . Then, φ is locally Lipschitzian under the condition*

$$a_i \notin \text{Ker } \Sigma \quad \forall i \in \{1, \dots, m\}, \quad (4)$$

where the a_i denote the rows of A .

Proof. The transformed random vector $\eta := A\xi$ inherits a quasi-concave distribution from that of ξ . With F_η being the distribution function of η , one may write $\varphi = F_\eta \circ h$. The i th component of η has variance $a_i^T \Sigma a_i$. Since this variance is larger than zero according to (4), Theorem 2.2 provides that F_η is Lipschitz continuous. Hence, φ is locally Lipschitzian as a composition of two such mappings. ■

3 Differentiability of singular normal distribution functions

Although the 3 examples of singular normal distribution functions presented in the introduction and depicted in Figure 1 fail to be differentiable in a global sense, they are differentiable almost everywhere. In order to establish a condition for differentiability, we shall introduce some concepts related with systems of linear inequalities. More precisely, let A is an (m, s) -matrix and $b \in \mathbb{R}^m$. We shall briefly speak of the system (A, b) to refer to the family

$$\langle a_i, z \rangle \leq b_i \quad (i = 1, \dots, m)$$

of linear inequalities in \mathbb{R}^s induces by the rows a_i of A and the components b_i of b . With the system (A, b) we associate a family of index sets defined by

$$I(A, b) := \{I \subseteq \{1, \dots, m\} \mid \exists z \in \mathbb{R}^s : \begin{array}{l} \langle a_i, z \rangle = b_i \quad (i \in I), \\ \langle a_i, z \rangle < b_i \quad (i \in \{1, \dots, m\} \setminus I) \end{array}\}.$$

Observe that $\emptyset \in I(A, b)$ if and only if the system (A, b) admits a Slater point, i.e., a point \bar{x} with

$$\langle a_i, \bar{x} \rangle < b_i \quad (i = 1, \dots, m).$$

The system (A, b) is said to be *nondegenerate*, if

$$\text{rank}\{a_i \mid i \in I\} = \#I \quad \forall I \in I(A, b).$$

In the language of optimization theory, the system (A, b) is nondegenerate if and only if it satisfies the *Linear Independence Constraint Qualification* (LICQ).

Proposition 3.1 *If the system (A, b) is nondegenerate, then*

$$I' \in I(A, b) \quad \forall I' \subseteq I \quad \forall I \in I(A, b).$$

In particular, $\emptyset \in I(A, b)$, hence the system (A, b) admits a Slater point.

Proof. Let $I \in I(A, b)$ and $I' \subseteq I$ be arbitrary. By definition, there is some $x \in \mathbb{R}^s$ with

$$\langle a_i, x \rangle = b_i \quad (i \in I), \quad \langle a_i, x \rangle < b_i \quad (i \in \{1, \dots, m\} \setminus I).$$

Regularity of the system (A, b) ensures that the a_i with $i \in I$ are linearly independent and, hence, the system of equations

$$\langle a_i, h \rangle = 0, \quad (i \in I'), \quad \langle a_i, h \rangle = -1 \quad (i \in I \setminus I')$$

has a solution h . For small enough $t > 0$, one has that

$$\begin{aligned} \langle a_i, x + th \rangle &= b_i, & (i \in I') \\ \langle a_i, x + th \rangle &= b_i - t < b_i, & (i \in I \setminus I') \\ \langle a_i, x + th \rangle &= \langle a_i, x \rangle + t \langle a_i, h \rangle < b_i & (i \in \{1, \dots, m\} \setminus I). \end{aligned}$$

In other words, $I' \in I(A, b)$. ■

Our differentiability result will basically rely on the following formula for the probability of polyhedra proved in [8] by means of the so-called abstract-tube theory (a recent proof based on more elementary arguments like duality of linear programming can be found in [4]):

Theorem 3.1 *Let ξ be an s -dimensional random vector. If the system (A, b) is nondegenerate, then the probability of the polyhedron induced by (A, b) equals*

$$P(\langle a_i, \xi \rangle \leq b_i \quad (i = 1, \dots, m)) = \sum_{I \in I(A, b)} (-1)^{\#I} P(\langle a_i, \xi \rangle > b_i \quad (i \in I)).$$

By Proposition 3.1, the assumed nondegeneracy implies that $\emptyset \in I(A, b)$. In this case, by formal argumentation, the corresponding term in the sum above becomes

$$(-1)^{\#\emptyset} P(\langle a_i, \xi \rangle > b_i \quad (i \in \emptyset)) = P(\mathbb{R}^s) = 1.$$

Next we need the following technical result:

Proposition 3.2 *Suppose that the system (A, b) is nondegenerate. Then, there exists a neighborhood U of b such that for all $b' \in U$ the systems (A, b') are nondegenerate too and $I(A, b') = I(A, b)$.*

Proof. According to the definition of nondegeneracy, the first assertion is an immediate consequence of the second one. We show first that there is a neighborhood U of b such that

$$I(A, b) \subseteq I(A, b') \quad \forall b' \in U. \tag{5}$$

Let $I \in I(A, b)$ be arbitrary. By definition, there is some $z \in \mathbb{R}^s$ with

$$\langle a_i, z \rangle = b_i \quad (i \in I), \quad \langle a_i, z \rangle < b_i \quad (i \in \{1, \dots, m\} \setminus I).$$

Let U, V be neighborhoods of b and z , respectively, such that

$$\langle a_i, z' \rangle < b'_i \quad (i \in \{1, \dots, m\} \setminus I) \quad \forall z' \in V \quad \forall b' \in U.$$

Due to the nondegeneracy of the system (A, b) , we have that

$$\text{rank}\{a_i | i \in I\} = \#I.$$

This implies that, choosing U small enough, for all $b' \in U$ the linear equations

$$\langle a_i, z' \rangle = b'_i \quad (i \in I)$$

possess solutions $z' \in V$ (this can be seen by working with the right inverse of the submatrix of A defined by the rows a_i for $i \in I$). Consequently, for all $b' \in U$, there exists some z' satisfying the relations

$$\langle a_i, z' \rangle = b'_i \quad (i \in I), \quad \langle a_i, z' \rangle < b'_i \quad (i \in \{1, \dots, m\} \setminus I).$$

This amounts to $I \in I(A, b')$, whence the inclusion (5).

Now, we show that there is a neighborhood U of b such that

$$I(A, b') \subseteq I(A, b) \quad \forall b' \in U. \quad (6)$$

Choosing the intersection of this neighborhood U with the one found for (5) will prove the assertion of the proposition. It is well-known (see, e.g., [1], Theorem 3.4.1) that the multifunction M which assigns to each b' the solution of the system (A, b') , can be decomposed as

$$M(b') = K(b') + U,$$

where K is a Hausdorff-continuous multifunction such that the $K(b')$ are convex, compact polyhedra for all b' , and where

$$U = \{u \mid \langle a_i, u \rangle \leq 0 \quad (i = 1, \dots, m)\}.$$

Now, negating (6) would imply the existence of sequences x_k and $b^{(k)} \rightarrow b$ as well as of an index set $I \subseteq \{1, \dots, m\} \setminus I(A, b)$ such that

$$\langle a_i, x_k \rangle = b_i^{(k)} \quad (i \in I), \quad \langle a_i, x_k \rangle < b_i^{(k)} \quad (i \in \{1, \dots, m\} \setminus I).$$

(Note that there are only finitely many subsets of $\{1, \dots, m\}$, so instead of considering a sequence of index sets I_k which would result by formal negation of (6), one may restrict the considerations to a fixed index set I upon passing to subsequences). Clearly, $x_k \in M(b^{(k)})$, hence there are sequences $y_k \in K(b^{(k)})$ and $u_k \in U$ with $y_k = x_k - u_k$. By the Hausdorff continuity of K and the compactness of $K(b)$ it follows that y_k is bounded, hence without loss of generality, we assume that $y_k \rightarrow \bar{y}$ for some $\bar{y} \in K(b)$ (again by Hausdorff continuity of K). For any $i \in I$, one has that

$$\langle a_i, \bar{y} \rangle = \lim \{\langle a_i, x_k \rangle - \langle a_i, u_k \rangle\} \geq \lim b_i^{(k)} = b_i.$$

On the other hand, since $0 \in U$, we know that $\bar{y} \in M(b)$, whence $\langle a_i, \bar{y} \rangle \leq b_i$. Summarizing, $\langle a_i, \bar{y} \rangle = b_i$ for all $i \in I$. Since \bar{y} solves the system (A, b) , there is some index set $I' \supseteq I$ such that

$$\langle a_i, \bar{y} \rangle = b_i \quad (i \in I'), \quad \langle a_i, \bar{y} \rangle < b_i \quad (i \in \{1, \dots, m\} \setminus I').$$

In other words, $I' \in I(A, b)$. ■

Recall from the introduction that a singular normal distribution can always be obtained as a linear transformation of some nondegenerate normal distribution. If this linear transformation is not explicitly given but just the covariance matrix Ξ and the mean vector γ of the singular normal distribution are known, this transformation can be found as follows: First decompose the (possibly degenerate) covariance matrix Ξ as $\Xi = AA^T$ such that A has full rank. Let ξ be a random vector whose dimension coincides with the number of columns of A and which has independent normally distributed components with zero mean. Then, the transformation $A\xi + \gamma$ generates a random vector with covariance matrix $AA^T = \Xi$ and mean γ , i.e., A and γ define the desired linear transformation.

Now, we state the main result of this section.

Theorem 3.2 *Let ξ be an s -dimensional random vector having a nondegenerate normal distribution. Denote by Φ_η the distribution function of the linearly transformed random vector $\eta = A\xi + b$, where A is an (m, s) -matrix and $b \in \mathbb{R}^m$. Then, Φ_η is smooth (infinitely many times differentiable) at any point $\bar{x} \in \mathbb{R}^m$ for which the system $(A, \bar{x} - b)$ is nondegenerate.*

Proof. By Proposition 3.2, there exists a neighborhood U of \bar{x} such that the system $(A, x - b)$ is nondegenerate and $I(A, x - b) = I(A, \bar{x} - b)$ for all $x \in U$. By definition, one has that

$$\Phi_\eta(x) = P(\eta \leq x) = P(\langle a_i, \xi \rangle \leq x_i - b_i \quad (i = 1, \dots, m)).$$

Application of Theorem 3.1 to the systems $(A, x - b)$ yields that, for all $x \in U$:

$$\Phi_\eta(x) = \sum_{I \in I(A, x-b)} (-1)^{\#I} P(\langle a_i, \xi \rangle > x_i - b_i \quad (i \in I)).$$

We note that in the last relation, one may pass to a non-strict inequality due to ξ having a nondegenerate normal distribution. Indeed, since all a_i are different from zero as a consequence of nondegeneracy, the difference between the strict and non-strict inequality signs above is given by events which are contained in a hyperplane and thus in a set of Lebesgue measure zero. Since ξ has a density, passing to non-strict inequalities will not change the probability:

$$\Phi_\eta(x) = \sum_{I \in I(A, x-b)} (-1)^{\#I} P(\langle a_i, \xi \rangle \geq x_i - b_i \quad (i \in I)).$$

Denote by A^I the submatrix of A whose rows are given by the selection a_i for $i \in I$. Similarly, denote by v^I a subvector of some $v \in \mathbb{R}^m$ such that the components of v^I realize the selection $i \in I$ of the components of v . Define $\eta^I := -A^I\xi$. Then, with the inequality sign understood componentwise, one has for all $x \in U$ that

$$\Phi_\eta(x) = \sum_{I \in I(A, x-b)} (-1)^{\#I} P(\eta^I \leq b^I - x^I) = \sum_{I \in I(A, x-b)} (-1)^{\#I} F^I(b^I - x^I). \quad (7)$$

Here, F^I refers to the distribution function of η^I . Obviously, η^I has a normal distribution with covariance matrix $A^I (A^I)^T$. Due to nondegeneracy of the systems $(A, x - b)$, we know that A^I has full rank for all $I \in I(A, x - b)$ and all $x \in U$. Consequently, the covariance matrix itself has full rank, which means that η^I has a nondegenerate normal distribution. As a consequence, all distribution functions F^I are (globally) smooth. We are tempted now, to differentiate the sum in (7) all terms of which are differentiable. This would imply the desired smoothness of Φ_η at \bar{x} . However, care has to be taken since the number of terms given by the cardinality of $I(A, x - b)$ does formally depend on x , hence certain terms could suddenly disappear or appear in addition, when moving away from \bar{x} . Fortunately, we know (see beginning of the proof) that $I(A, x - b) = I(A, \bar{x} - b)$ for all $x \in U$. This allows to write Φ_η locally around \bar{x} as a sum of a fixed number of smooth functions which implies smoothness of Φ_η at \bar{x} :

$$\Phi_\eta(x) = \sum_{I \in I(A, \bar{x} - b)} (-1)^{\#I} F^I(b^I - x^I) \quad \forall x \in U. \quad (8)$$

■

Note that Theorem 3.2 does not just make a theoretical statement on smoothness of singular normal distribution functions, but even provides a formula how to calculate their derivatives. Indeed, one may use (8) in order to calculate the gradient (or higher order derivatives) of Φ_η on the basis of the same objects for nondegenerate (!) normal distribution functions (the F^I). As first and higher order derivatives of nondegenerate normal distribution functions can be analytically reduced to functional values themselves (see, e.g., [9]), everything boils down to the mere calculation of nondegenerate normal distribution functions. This can be carried out by several existing algorithms (see, e.g., [5], [6], [11]).

Finally, we want to illustrate Theorem 3.2 by applying it to the singular normal distribution with zero mean vector and covariance matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = AA^T \quad \text{with } A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(see second picture in Fig. 1). Such distribution is realized by a random vector $\eta = A\xi$, where ξ has a one-dimensional standard normal distribution (compare remark in front of Theorem 3.2). We have to check, for which vectors $x \in \mathbb{R}^2$ the system (A, x) is nondegenerate. Concerning the calculation of the index family $I(A, x)$, one has to distinguish three cases:

$$\begin{aligned} x_1 < x_2 &\implies I(A, x) = \{\emptyset, \{1\}\} \\ x_1 > x_2 &\implies I(A, x) = \{\emptyset, \{2\}\} \\ x_1 = x_2 &\implies I(A, x) = \{\emptyset, \{1, 2\}\}. \end{aligned}$$

Obviously, nondegeneracy holds true in the first two cases since both 'rows' of A (which reduce to real numbers here) are different from zero. Consequently, Theorem

3.2 guarantees differentiability of the distribution function of η whenever $x_1 \neq x_2$ (this can be verified from Fig. 1). On the other hand, the two 'rows' of A cannot be linearly independent, hence nondegeneracy is lost in case of $x_1 = x_2$. This harmonizes with the fact that the distribution function of η is not differentiable on the bisectrix $x_1 = x_2$ (see Fig. 1).

References

- [1] B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer: *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.
- [2] O.E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, Wiley, Chichester, 1978.
- [3] C. Borell, Convex Sets in d-Space, *Periodica Mathematica Hungarica* 6 (1975), 111-136.
- [4] J. Bukszár, R. Henrion, M. Hujter and T. Szántai, Polyhedral Inclusion-Exclusion, Weierstrass Institute Berlin, Preprint No. 913, 2004.
- [5] H.I. Gassmann, I. Deák, T. Szántai, Computing multivariate normal probabilities: A new look, *J. Comput. Graph. Statist.* 11 (2002) 920-949.
- [6] A. Genz, Numerical computation of multivariate normal probabilities, *J. Comput. Graph. Statist.* 1 (1992) 141-149.
- [7] R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Mathematical Programming* 84 (1999), 55-88.
- [8] D.Q. Naiman and H.P. Wynn, Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling, *Annals of Statistics* 25 (1997), 1954-1983.
- [9] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.
- [10] W. Römisch and R. Schultz, Stability of solutions for stochastic programs with complete recourse, *Mathematics of Operations Research* 18 (1993), 590-609.
- [11] T. Szántai, Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function, *Ann. Oper. Res.* 100 (2000) 85-101.