Differentiability of probability function involving non-linear mappings of Gaussian random vectors

W. van Ackooij[1]

[1] OSIRIS Department
EDF R&D
7 Boulevard Gaspard Monge; 9120 Palaiseau ; France

GDR Mascot, 13/05/2016

Outline

Motivation I

- A Probabilistic constraint is a constraint of the type

$$\varphi(x) := \mathbb{P}[g(x,\xi) \leq 0] \geq p, \tag{1}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ is a map, $\xi \in \mathbb{R}^m$ a (multi-variate) random variable

- Such constraints arise in many applications. For instance cascaded Reservoir management.

- We care for further understanding of differentiability of probability functions

Some differentiability properties of PCs I

- General differentiability statements exist and represent the gradient as an involved integral over a “surface” and “volume” . A key condition is that $\{z \in \mathbb{R}^m : g(x,z) \leq 0\}$ is bounded locally around a point x (e.g., [Uryas’ev(2009)]).

Some differentiability properties of PCs II

- Specific formulas such as the following, allow for efficient computation in practice:

Lemma ([Prékopa(1970), Prékopa(1995)])

Let ξ be an m-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^m$ and positive definite variance-covariance matrix Σ. Then the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is continuously differentiable and in any fixed $z \in \mathbb{R}^m$ the following holds:

$$\frac{\partial F_\xi}{\partial z_i}(z) = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m), i = 1, ..., m. \quad (2)$$

Here $\tilde{\xi}(z_i)$ is a Gaussian random variable with mean $\hat{\mu} \in \mathbb{R}^{m-1}$ and $(m-1) \times (m-1)$ positive definite covariance matrix $\hat{\Sigma}$. Let D^i_m denote the m-th order identity matrix from which the ith row has been deleted. Then $\hat{\mu} = D^i_m(\mu + \Sigma_{ii}^{-1}(z_i - \mu_i)\Sigma_i)$ and $\hat{\Sigma} = D^i_m(\Sigma - \Sigma_{ii}^{-1}\Sigma_i\Sigma_i^\mathsf{T})(D^i_m)^\mathsf{T}$, where Σ_i is the i-th column of Σ.

Some differentiability properties of PCs III

- $\varphi(x) := \mathbb{P}[\xi \leq x]$ ([Prékopa(1970)]) We have

$$\frac{\partial \varphi}{\partial x_i} = f_{\mu_i, \Sigma_{ii}}(x_i)\mathbb{P}[\tilde{\xi} \leq \tilde{x}]$$

- $\varphi(x) := \mathbb{P}[A(x)\xi \leq \alpha(x)]$ ([van Ackooij et al.(2011)])
- $\varphi(x) := \mathbb{P}[A\xi \leq \alpha(x)]$ ([Henrion and Möller(2012)])
- Other cases involve distribution functions of Dirichlet ([Szántai(1985), Gouda and Szántai(2010)]) and multi-variate Gamma ([Prékopa and Szántai(1979)]) random variables

Setting

- Consider the probabilistic constraint :

$$\varphi(x) := \mathbb{P}[g(x, \xi) \leq 0] \geq p, \tag{3}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is a continuously differentiable map (convex in the second argument), $\xi \sim \mathcal{N}(\mu, \Sigma)$ a (multi-variate) Gaussian random variable.

Motivation

- We would like to dispose of a gradient formulae for the case

$$\varphi(x) := \mathbb{P}[\langle c, \eta \rangle \leq h(x)],$$

where $c \geq 0, c \in \mathbb{R}^m$, and $\eta \in \mathbb{R}^m$ is a log-normal random variable

- We can cast this into the general case by defining the mapping

$$g(x, z) = \langle c, \exp(z) \rangle - h(x).$$

- Then $\varphi(x) = \mathbb{P}[g(x, \xi) \leq 0]$ with $\xi \sim \mathcal{N}(\mu, \Sigma)$.
- In fact by redefining g we may assume w.l.o.g. that $\xi \sim \mathcal{N}(0, R)$.

Inherent non-smoothness

- It is tempting to believe that "nice" properties of g carry forth to φ. For instance, if g is smooth enough, that φ will be at least continuously differentiable.
- Though "nasty laws" for ξ can be expected to have side-effects, nice laws may not.
- Let us first show that such considerations are dangerous.

Inherent non-smoothness: counterexample

Differentiability need not hold:

Proposition

Let $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ *be defined by*

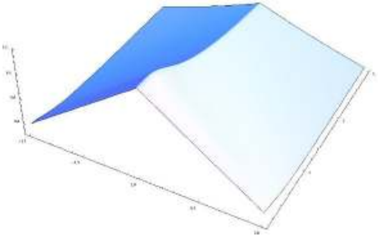
$$g(x_1, x_2, z_1, z_2) := x_1^2 e^{h(z_1)} + x_2 z_2 - 1, \quad \text{where} \quad h(t) := -1 - 2\log(1 - \Phi(t))$$

and Φ *is the cumulative distribution function of the one-dimensional standard Gaussian distribution. Let* $\xi \sim \mathcal{N}(0, I_2)$ *and* $\bar{x} = (0, 1)$*. Then, the following holds true:*

1. g *is continuously differentiable.*
2. g *is convex in the second argument.*
3. $g(\bar{x}, 0) = g(0, 1, 0, 0) < 0.$
4. φ *is not differentiable at* $\bar{x}$*.*

Inherent non-smoothness: counterexample

Graph of a non-differentiable probability function



Inherent non-smoothness: several components

Things may also go wrong when $p > 1$, i.e., g has several components:

Example

Let ξ have a one-dimensional standard Gaussian distribution and define

$$g(x_1, x_2, x_3, \xi) := (\xi - x_1, \xi - x_2, -\xi - x_3).$$

Then, with Φ referring to the one-dimensional standard Gaussian distribution function, one has that

$$\varphi(x_1, x_2) = \max\{\min\{\Phi(x_1), \Phi(x_2)\} - \Phi(x_3), 0\}.$$

Clearly φ fails to be differentiable at $\bar{x} := (0, 0, -1)$, while $\{z \ : g(\bar{x}, z) \leq 0\} = [-1, 0]$ is compact and satisfies Slater's condition in the description via g.

Inherent non-smoothness: the need for additional conditions

- From these discussion it is clear that some conditions needs to be appended in order to avoid some degeneracy
- Essentially two conditions are needed: bounded growth on $\nabla_x g$, some LICQ type of regularity.

Evaluating $\mathbb{P}$

- Let $\mathbb{S}^{m-1} := \left\{z \in \mathbb{R}^m \,\middle|\, \sum_{i=1}^m z_i^2 = 1\right\}$ be the euclidian unit-sphere of $\mathbb{R}^m$.
- Let $\xi \sim \mathcal{N}(0, R)$ be given and L be such that $R = LL^{\mathsf{T}}$.
- It is well known that $\xi = \eta L \zeta$, where η has a chi-distribution with m degrees of freedom and ζ is uniformly distributed over $\mathbb{S}^{m-1}$

Evaluating $\mathbb{P}$ II

- As a consequence if $M \subseteq \mathbb{R}^m$ is Lebesgue measurable
- We have

$$\mathbb{P}[\xi \in M] = \int_{v \in \mathbb{S}^{m-1}} \mu_\eta \left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right) d\mu_\zeta \tag{4}$$

- Efficient sampling schemes for such integrals are provided by [Deák(1986), Deák(2000)]
- In our case $M(x) = \{z \in \mathbb{R}^m \ : g(x,z) \leq 0\}$ is a convex (hence Lebesgue measurable) set.

Growth control

We cannot allow for unbounded growth of the mapping g. We thus define:

Definition

We say that g satisfies the **exponential growth condition** at x if there exist constants $\delta_0, C > 0$ and a neighbourhood $U(x)$ such that

$$\left\|\nabla_x g\left(x', z\right)\right\| \leq \delta_0 \exp(\|z\|) \quad \forall x' \in U(x)\ \forall z : \|z\| \geq C.$$

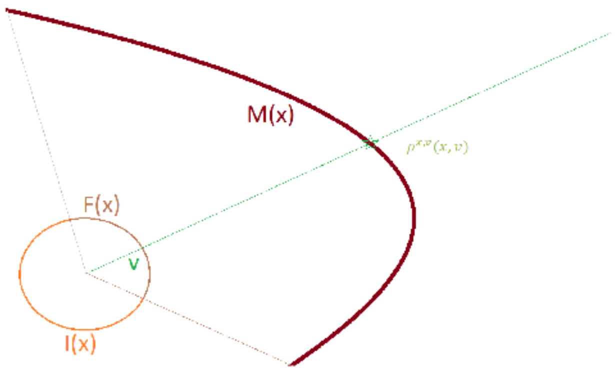
The case $p = 1$

- We define the sets of finite and infinite directions:

$$
\begin{aligned}
F(x) :=& \quad \left\{ v \in \mathbb{S}^{m-1} | \exists r > 0 : g\left(x, rLv\right) = 0 \right\} \\
I(x) :=& \quad \left\{ v \in \mathbb{S}^{m-1} | \forall r > 0 : g\left(x, rLv\right) \neq 0 \right\}.
\end{aligned}
$$

- For each $x \in \mathbb{R}^n$ with $g(x, 0) < 0$ and $v \in F(x)$ we can find a unique $\rho^{x,v}(x, v) > 0$ such that $g(x, \rho^{x,v}(x, v)Lv) = 0$.
- Numerically this value can be computed by a simple application of Newton-Rhapson.

The case $p = 1$: Illustration



The case $p = 1$: main result

Theorem ([van Ackooij and Henrion(2014)])

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function which is convex with respect to the second argument. Consider the probability function φ defined as $\varphi(x) = \mathbb{P}[g(x, \xi) \leq 0]$, where $\xi \sim \mathcal{N}(0, R)$ has a standard Gaussian distribution with correlation matrix R. Let the following assumptions be satisfied at some $\bar{x}$:

1. $g(\bar{x}, 0) < 0$.
2. *g satisfies the exponential growth condition at $\bar{x}$*

Then, φ is continuously differentiable on a neighbourhood U of $\bar{x}$ and it holds for all $x \in U$ that:

$$\nabla\varphi(x) = -\int_{v \in F(x)} \frac{\chi\left(\rho^{x,v}(x,v)\right) \nabla_x g\left(x, \rho^{x,v}(x,v) Lv\right)}{\left\langle \nabla_z g\left(x, \rho^{x,v}(x,v) Lv\right), Lv \right\rangle} d\mu_\zeta(v).$$

Theorem

The previous Theorem remains true if the growth condition is replaced by the condition that the set $\{z | g(\bar{x}, z) \leq 0\}$ is bounded. Then, the formula becomes

$$\nabla \varphi(x) = -\int_{v \in \mathbb{S}^{m-1}} \frac{\chi\left(\rho^{x,v}(x,v)\right) \nabla_x g\left(x, \rho^{x,v}(x,v) Lv\right)}{\left\langle \nabla_z g\left(x, \rho^{x,v}(x,v) Lv\right), Lv \right\rangle} d\mu_\zeta(v)$$

The case $p > 1$

- When $p > 1$ we can define

$$g^m(x,z) = \max_{j=1,\dots,p} g_j(x,z), \tag{5}$$

- Evidently, the probability function can be written as $\varphi(x) = \mathbb{P}(g^m(x,\xi) \leq 0)$.
- For each $x \in \mathbb{R}^n$ with $g(x,0) < 0$ and $v \in F(x)$ we can find a unique $\rho^{x,v}(x,v) > 0$ such that $g^m(x, \rho^{x,v}(x,v)Lv) = 0$. However this $\rho^{x,v}$ is no longer smooth!
- The sets of finite and infinite directions can be defined with respect to g^m or alternatively as unions (intersections) of their counterparts with respect to each component of g.

The case $p > 1$: main result

Theorem ([van Ackooij and Henrion(2016)])

Let the following conditions be satisfied at some fixed $\bar{x} \in \mathbb{R}^n$:

1. $g^m(\bar{x}, 0) < 0$.
2. g_j *satisfies the exponential growth condition at* $\bar{x}$ *for all* $j = 1, ..., p$.

Then, φ *is locally Lipschitz continuous on a neighbourhood* U *of* $\bar{x}$ *and it holds that*

$$\partial^c \varphi(x) \subseteq \int_{v \in F(x)} \mathrm{Co} \left\{ -\frac{\chi(\hat{\rho}(x,v)) \nabla_x g_j(x, \hat{\rho}(x,v) Lv)}{\langle \nabla_z g_j(x, \hat{\rho}(x,v) Lv), Lv \rangle} \middle| j \in \hat{\mathcal{J}}(x,v) \right\} d\mu_\zeta(v) \tag{6}$$

for all $x \in U$. *Here,*

$$\hat{\mathcal{J}}(x,v) := \{j \in \{1, \dots, p\} | g_j(x, \hat{\rho}(x,v) Lv) = 0\} \quad (v \in F(x))$$

The case $p > 1$: A first discussion

- Note that in the case $p > 1$, under the same conditions as for the case $p = 1$, we have a weaker results: local Lipschitz continuity and an outer estimate of the clarke-subdifferential
- The earlier example showed that this is inherent and not a weakness of the analysis.

The case $p > 1$: R2CQ

Definition

For any $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$ we denote by

$$\mathcal{I}(x,z) := \{j \in \{1,\ldots,p\} \,|\, g_j(x,z) = 0\} \tag{7}$$

the active index set of g at (x,z). We say that the inequality system $g(x,z) \leq 0$ satisfies the *Rank-2-Constraint Qualification* (*R2CQ*) at $x \in \mathbb{R}^n$ if

$$\text{rank} \; \{\nabla_z g_j(x,z), \nabla_z g_i(x,z)\} = 2 \quad \forall i,j \in \mathcal{I}(x,z), i \neq j \tag{8}$$

$$\forall z \in \mathbb{R}^m : g(x,z) \leq 0. \tag{9}$$

The case $p > 1$: R2CQ < LICQ

- Note that (*R2CQ*) is substantially weaker than the usual *Linear Independence Constraint Qualification (LICQ)* common in nonlinear optimization and requiring the linear independence of all gradients to active constraints.

The case $p > 1$: An auxiliary result

Lemma ([van Ackooij and Henrion(2016)])

Let $\bar{x} \in \mathbb{R}^n$ be given such that

1. $g^m(\bar{x}, 0) < 0$.
2. *g satisfies $(R2CQ)$ at $\bar{x}$.*

Then, $\mu_\zeta(M') = 0$ for $M' := \left\{v \in \mathbb{S}^{m-1} | \exists r > 0 : g(\bar{x}, rLv) \leq 0,\ \#\mathcal{I}(\bar{x}, rLv) \geq 2\right\}$, where L is the regular matrix in the decomposition $R = LL^T$.

The case $p > 1$: smoothness

Theorem ([van Ackooij and Henrion(2016)])

Let the following conditions be satisfied at some fixed $\bar{x} \in \mathbb{R}^n$:

1. $g^m(\bar{x}, 0) < 0$.
2. g_j *satisfies the exponential growth condition at* $\bar{x}$ *for all* $j = 1, ..., p$.
3. $(R2CQ)$ *is satisfied*

Then, φ *is Fréchet differentiable at* $\bar{x}$ *and the gradient formula:*

$$\nabla\varphi(\bar{x}) = -\int_{v \in F(\bar{x}), \#\hat{\mathcal{J}}(\bar{x}, v) = 1} \frac{\chi(\hat{\rho}(\bar{x}, v)) \nabla_x g_{j(v)}(\bar{x}, \hat{\rho}(\bar{x}, v) Lv)}{\langle \nabla_z g_{j(v)}(\bar{x}, \hat{\rho}(\bar{x}, v) Lv), Lv \rangle} d\mu_\zeta(v), \quad (10)$$

holds true.
If $(R2CQ)$ *is satisfied locally around* $\bar{x}$*, then,* φ *is continuously differentiable at* $\bar{x}$*.*

One last remark

The condition $g(x, 0) < 0$ is not very restrictive as the following result shows:

Lemma

With g and φ as before, let the following assumptions be satisfied at some $\bar{x}$:

1. *There exists some $\bar{z}$ such that $g(\bar{x}, \bar{z}) < 0$.*
2. $\varphi(\bar{x}) > 1/2$.

Then, $g(\bar{x}, 0) < 0$.

Motivation

- Let us consider the special case wherein φ results from

$$\varphi(x) := \mathbb{P}[B\xi \leq h(x)], \tag{11}$$

with $\xi \sim \mathcal{N}(\mu, \Sigma)$, $\Sigma \succ 0$.

- When B is of full rank then, $B^{\mathsf{T}}\Sigma B \succ 0$ too and differentiability follows from classic results.

- However in many applications B has more rows than columns (for instance when coming from Gale-Hoffmann inequalities): φ is no longer smooth.

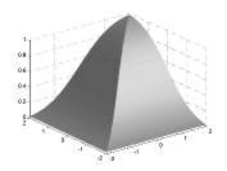
Motivation

Example

Let $m = 1$, $k = 2$, $\xi \sim \mathcal{N}(0,1)$ and B be given by

$$B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then it is readily observed that $\varphi(x) = \mathbb{P}[B\xi \leq x] = \mathbb{P}[\xi \leq \min\{x_1, x_2\}]$. As a consequence φ fails to be differentiable on the line $x_1 = x_2$ as is readily seen on the figure:



Setting

- Without loss of generality we concentrate on $\varphi(z) = \mathbb{P}[\xi \leq z]$, with $\xi \sim \mathcal{N}(0, \Sigma)$ and $\Sigma \succeq 0$.
- We may also assume that $\Sigma_{ii} = 1$ for all i without loss of generality (as otherwise either the system contains a redundant constraint (locally around z), or φ fails to be continuous in z).

Correlation graph

Definition

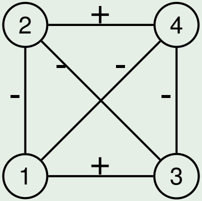
Let Σ be an $m \times m$ covariance matrix having all diagonal entries equal to 1. Let $G(\Sigma) = (V, E)$ denote the (undirected) graph on the vertex set $V = \{1, ..., m\}$ and with edge set $E = E^+ \cup E^- = \{(i,j) \ : i \neq j, \Sigma_{ji} = 1\} \cup \{(i,j) \ : i \neq j, \Sigma_{ji} = -1\}$. The graph $G(\Sigma)$ (which may contain isolated vertices) will be called the correlation graph associated with Σ.

Correlation graph: Example

Example

Consider the 4×4 covariance matrix Σ defined as follows:

$$\Sigma = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},$$

then the correlation graph:  is obtained

Correlation graph

- The correlation graph features Q connected components (each being either an isolated vertex or a complete subgraph (a clique)).
- Each connected component $G^q = (V^q, E^q)$ is bipartite and can be separated into a left and right side L^q, R^q: elements within L^q are positively correlated, elements in L^q are negatively correlated to those in R^q.

Correlation graph and z

Definition

Let $G(\Sigma) = (V, E)$ be a correlation graph:

- Given an arbitrary $z \in \mathbb{R}^m$, we will say that z is *auto-referenced* if there exists an arc $(i, j) \in E$ such that $z_j = \Sigma_{ji} z_i$ (in other words, such that $z_j = z_i$ if $(i, j) \in E^+$ or such that $z_j = -z_i$ if $(i, j) \in E^-$).
- An auto-referenced point $z \in \mathbb{R}^m$ will be called *changeable* if there exists $(i, j) \in E$ such that $z_k \geq z_i$ for all $(k, i) \in E^+$ and $z_k \geq -z_i$ for all $(k, i) \in E^-$.

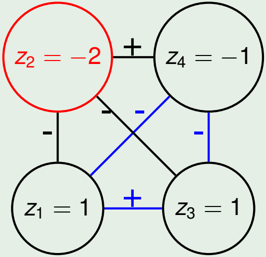
The arc $(i, j) \in E$ will occasionally be referred to as an auto-referencing (a changeable) arc with respect to z if z is auto-referenced (changeable).

Correlation graph: Example

Example

Consider again $z = (1, -2, 1, -1)$ and

$$\Sigma = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},$$



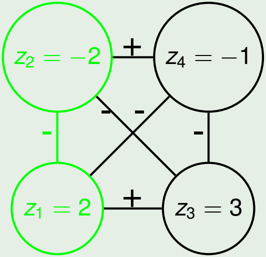
Then z is auto-referenced (blue), but not changeable ($-2 \geq -1$ is false).

Correlation graph: Example 2

Example

Consider again $z = (2, -2, 3, -1)$ and

$$\Sigma = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},$$



Then z is changeable (green) (argmins among the partitions L^q, R^q).

A first result

Theorem ([van Ackooij and Minoux(2015)])

Let ξ be an m-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^m$ and covariance matrix Σ having all diagonal entries equal to 1. Then for arbitrary not-changeable $z - \mu \in \mathbb{R}^m$, the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is locally Lipschitz at z and $\partial^c F_\xi(z) = \{v\}$, where for arbitrary i=1,...,m:

$$v_i = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m). \quad (12)$$

Here $\partial^c F_\xi(z)$ denotes the Clarke-subdifferential of F_ξ and $\tilde{\xi}(z_i)$ is an $m-1$ dimensional Gaussian random vector (familiar from classic results)

The familiar associated Gaussian

- f_{ξ_i} is the one dimensional Gaussian density of ξ_i
- $\tilde{\xi}(z_i) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$
- Let D^i_m denote the $(m-1) \times m$ matrix deduced from the $m \times m$ identity matrix by deleting the ith row.
- $\hat{\mu} = D^i_m(\mu + \Sigma_{ii}^{-1}(z_i - \mu_i)\Sigma_i)$
- $$\hat{\Sigma} = D^i_m(\Sigma - \Sigma_{ii}^{-1}\Sigma_i\Sigma_i^{\mathsf{T}})(D^i_m)^{\mathsf{T}},$$
  where Σ_i is the i-th column of Σ and Σ_{ii} is the i-th element of the main diagonal of Σ.

And changeable points?

Proposition ([van Ackooij and Minoux(2015)])

Let $G^q = (V^q, E^q)$, be the connected $q = 1, ..., Q$ components of the correlation graph and (L^q, R^q) be the associated bipartition. Let z be changeable. Define $J \subseteq \{1, ..., Q\}$ as the set of all q for which either $|V^q| = 1$ or no changeable arc exists in V^q. For each remaining $q \in \{1, ..., Q\} \setminus J$, pick $l^q \in L^q$, $r^q \in R^q$ such that $z_{l^q} \leq z_p$ for all $p \in L^q$ and $z_{r^q} \leq z_p$ for all $p \in R^q$. If R^q is empty, r^q should be interpreted as being "empty".
Then the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is locally Lipschitz at z and $v \in \partial^c F_\xi(z)$, where for arbitrary i=1,...,m:

$$v_i = \begin{cases} f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m) & \text{if} \quad i \in \cup_{j \in J} V^j \\ f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m) & \text{if} \quad \exists q \in \{1, ..., Q\} \setminus J, i \in \{l^q, r \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Moreover $\partial^c F_\xi(z)$ contains at least two elements.

A final definition

Definition

Let $z \in \mathbb{R}^m$ be arbitrary. Define the set $\mathcal{E}(z)$ as the set of all v defined according to previous formula, where we enumerate all possible choices of l^q, r^q for each q. For a specific q if V^q contains a changeable arc with one endpoint in L^q and the other endpoint in R^q we adjoin to this set of choices, $v \in \mathbb{R}^m$, with $v_p = 0$ for $p \in V^q$.

The main result

Theorem ([van Ackooij and Minoux(2015)])

Let ξ be an m-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^m$ and covariance matrix Σ having all diagonal entries equal to 1. Then the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is continuously differentiable if and only if $z - \mu$ is not changeable.
Moreover F_ξ is locally Lipschitz at z and

$$\partial^c F_\xi(z) = \mathrm{co}\,(\mathcal{E}(z)), \tag{14}$$

where $\mathrm{co}\,(B)$ denotes the convex hull of set $B \subseteq \mathbb{R}^m$.

Linear maps as a special case: another formula

When seeing the linear situation, i.e., $\varphi(x) := \mathbb{P}[A\xi \leq x]$, as a special case of nonlinear g we get:

Corollary

Let $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1, \ldots, p\}$. Finally assume that any two active rows of the matrix A are linearly independent:

$$Az \leq \bar{x},\ A_i z = \bar{x}_i,\ A_j z = \bar{x}_j,\ i \neq j \Longrightarrow \mathrm{rank}\ \{A_i, A_j\} = 2. \tag{15}$$

Then, φ is continuously differentiable at $\bar{x}$ and it holds that

$$\frac{\partial \varphi}{\partial x_j}(\bar{x}) = \int_{\left\{v \in \mathbb{S}^{m-1} \mid A_j L v > 0, \bar{x}_j = \hat{\rho}(v) A_j L v\right\}} \frac{\chi\left(\hat{\rho}\left(v\right)\right)}{A_j L v}\, d\mu_\zeta(v) \quad (j = 1, \ldots, p). \tag{16}$$

Some notation

- We introduce the following equivalence class within the index set $\{1, \dots, p\}$ of rows of the matrix A:

$$i \sim j \Longleftrightarrow \exists \lambda \in \mathbb{R} : A_i = \lambda A_j,\ \bar{x}_i = \lambda \bar{x}_j.$$

- By the assumption $\bar{x}_j > 0$ for all $j \in \{1, \dots, p\}$, $i \sim j$ implies that $\lambda > 0$ in the defining relation.
- Moreover $i \nsim j$ implies that rows A_i and A_j of A are linearly independent.

Some notation

- Denote by $\tilde{p} \leq p$ the number of different equivalence classes $[i]$.
- We may assume (w.l.o.g) that the first $\tilde{p}$ rows of A belong to different equivalence classes.
- Now, for any $i = 1, \ldots, \tilde{p}$ that

$$A_j z \leq x_j \quad \forall j \in [i] \iff A_i z \leq h_i(x) := \min_{j \in [i]} \lambda_j^{-1} x_j. \tag{17}$$

- We denote by $\tilde{A}$ the submatrix of first $\tilde{p}$ rows of A.

A fine characterization of the *M*-subdifferential

We can then show

Theorem ([van Ackooij and Henrion(2016)])

Let $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1, \dots, p\}$. Then, φ is locally Lipschitz continuous and its Mordukhovich subdifferential can be estimated from above by

$$\partial^M \varphi(\bar{x}) \subseteq \sum_{i=1}^{\tilde{p}} \int_{\mathfrak{S}} \frac{\chi(\hat{\rho}(v))}{\tilde{A}_i L v} d\mu_\zeta(v) \cdot \bigcup \left\{ \lambda_j^{-1} e_j | j \in [i] : \lambda_j^{-1} \bar{x}_j = h_i(\bar{x}) \right\},$$

where

$$\hat{\rho}(v) := \min \left\{ \bar{y}_j / (A_j L v) | j \in \{1, \dots, \tilde{p}\} : \tilde{A}_j L v > 0 \right\}.$$

and

$$\mathfrak{S} := \left\{ v \in \mathbb{S}^{m-1} | \tilde{A}_i L v > 0, \bar{y}_i = \hat{\rho}(v) \tilde{A}_i L v \right\}.$$

Summary

In this talk we have discussed several aspects related to differentiability of chance constraints

The references of the discussed works

- W. van Ackooij and R. Henrion. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. *SIAM Journal on Optimization*, 24(4):1864–1889, 2014
- W. van Ackooij and R. Henrion. (sub-) gradient formulae for probability functions of random inequality systems under gaussian distribution. *Submitted, WIAS preprint 2230*, pages 1–24, 2016
- W. van Ackooij and M. Minoux. A characterization of the subdifferential of singular Gaussian distribution functions. *Set Valued and Variational Analysis*, 23(3):465–483, 2015. doi: 10.1007/s11228-015-0317-8

Bibliography I

[[Deák(1986)]   ]I. Deák.
Computing probabilities of rectangles in case of multinormal distribution.
Journal of Statistical Computation and Simulation, 26(1-2):101–114, 1986.

[[Deák(2000)]   ]I. Deák.
Subroutines for computing normal probabilities of sets - computer experiences.
Annals of Operations Research, 100:103–122, 2000.

[[Gouda and Szántai(2010)]   ]A. Gouda and T. Szántai.
On numerical calculation of probabilities according to dirichlet distribution.
Annals of Operations Research, 177:185–200, 2010.

Bibliography II

[[Henrion and Möller(2012)]]R. Henrion and A. Möller.
A gradient formula for linear chance constraints under Gaussian distribution.
Mathematics of Operations Research, 37:475–488, 2012.

[[Prékopa(1970)]]A. Prékopa.
On probabilistic constrained programming.
In H.W. Kuhn, editor, *Proceedings of the Princeton Symposium on Math. Prog.*, volume 28, pages 113–138, 1970.

[[Prékopa(1995)]]A. Prékopa.
Stochastic Programming.
Kluwer, Dordrecht, 1995.

[[Prékopa and Szántai(1979)]]A. Prékopa and T. Szántai.
On optimal regulation of a storage level with application to the water level regulation of a lake.
European Journal of Operations Research, 3:175–189, 1979.

Bibliography III

[[Szántai(1985)]]T. Szántai.
Numerical evaluation of probabilities concerning multi-dimensional probability distributions.
PhD thesis, Hungarian Academy of Sciences, 1985.

[[Uryas'ev(2009)]]S. Uryas'ev.
Derivatives of probability and integral functions: General theory and examples.
In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization*, pages 658–663. Springer - Verlag, 2nd edition, 2009.

[[van Ackooij and Henrion(2014)]]W. van Ackooij and R. Henrion.
Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions.
SIAM Journal on Optimization, 24(4):1864–1889, 2014.

Bibliography IV

[[van Ackooij and Henrion(2016)]]W. van Ackooij and R. Henrion.
(sub-) gradient formulae for probability functions of random inequality systems under gaussian distribution.
Submitted, WIAS preprint 2230, pages 1–24, 2016.

[[van Ackooij and Minoux(2015)]]W. van Ackooij and M. Minoux.
A characterization of the subdifferential of singular Gaussian distribution functions.
Set Valued and Variational Analysis, 23(3):465–483, 2015.
doi: 10.1007/s11228-015-0317-8.

[[van Ackooij et al.(2011)]]W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati.
On joint probabilistic constraints with Gaussian Coefficient Matrix.
Operations Research Letters, 39:99–102, 2011.