Probability constraints: analytical properties and a discussion of dedicated algorithms

W. van Ackooij[1]

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

EJC Grenoble, 06/07/2016

Outline

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

Motivation I

- A Probabilistic constraint is a constraint of the type

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

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.

Motivation II

- When considering k constraints of the type

$$\varphi_i(x) := \mathbb{P}[g_i(x,\xi) \geq 0] \geq p, i = 1, ..., k \tag{2}$$

  we speak of individual probabilistic constraints. The case of (1) is a joint probabilistic constraint.

- Individual PCs offer easier numerical treatment, but obviously lack robustness.

Why PCs

- In many applications, one encounters residual uncertainty, i.e., after making a decision a random outcome is observed.
- Such uncertainty may occur in constraints. In Unit-Commitment problems one encounters the following cases:

$$\begin{aligned} V_{min} &\leq V_0 - Ax + \xi \leq V_{max} \\ s^d &\leq D - A'x \leq s^u, \end{aligned} \tag{3}$$

  where x models the turbining/pumping policy in cascaded reservoir management, unit-commitment schedule respectively.

- Here x is decided upon before observing ξ (inflows) or D (net customer load).
- PCs are a way to give a meaning to (3)

What do PCs do

- Adding a probabilistic constraint, e.g.,

$$\begin{aligned} \mathbb{P}[V_{min} \leq V_0 - Ax + \xi \leq V_{max}] &\geq p \\ \mathbb{P}[s^d \leq D - A'x \leq s^u] &\geq p, \end{aligned} \tag{4}$$

  restrains the set of feasible solutions. Since x is decided upon before observing uncertainty, a posteriori violated inequalities are not arbitrarily "bad" .

- From a programming perspective: $\varphi(x) \geq p$, with $\varphi(x) := \mathbb{P}[g(x, \xi) \geq 0]$ is "just" a non-linear constraint

- In most cases φ is only known implicitly.

- The mapping φ is (usually) not concave, but could have generalized (e.g., log-concavity) properties.

Properties of PCs

- What mathematical properties can φ be expected to have ? Continuity, differentiability ?
- What properties does the set $M(p) := \{x \in \mathbb{R}^n \ : \varphi(x) \geq p\}$ have ? Connectedness, convexity
- What properties could problems of the type

$$\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f(x) \\ s.t. & x \in X \\ & \varphi(x) \geq p, \end{array} \tag{5}$$

  have if f is a convex mapping, X a convex set ? Stability ?
- Studying properties of φ, $M(p)$ or problems (5) is important for efficient numerical treatment of problems (5).

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

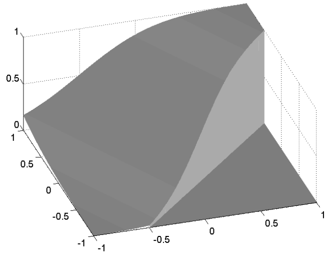
An example

- Let $\xi \sim \mathcal{N}(0,1)$ be given and consider

$$\varphi(x) := \mathbb{P}[Qx + L\xi \geq b], \tag{6}$$

with

$$Q = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}, L = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, b = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$$



- Then φ is not continuous

An example II

- The reason of this discontinuity is because of the presence of "deterministic" constraints $-1x_1 + x_2 \geq -\frac{1}{2}$ inside the probability constraint.
- Alternatively stated, we have a situation wherein the set $\{z \in \mathbb{R}^m \ : g(x, z) = 0\}$ is not of zero measure (at some x's).
- This shows the need for appropriate conditions (or a better model)
- Still continuity holds in a great many situations

Lower semi continuity

Lemma (e.g., [Henrion(2010)])

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ be (jointly) lower semi-continuous and assume that the sets $N_x = \{z \in \mathbb{R}^m : g(x,z) = 0\}$ are $\mathbb{P}$-null sets for all $x \in \mathbb{R}^n$. Let $\xi \in \mathbb{R}^m$ be a random variable. Then the mapping $\varphi(x) := \mathbb{P}[g(x,\xi) \geq 0]$ is also lower semi continuous.

Upper semi continuity

Lemma (e.g., [Henrion(2010)])

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ *be (jointly) upper semi-continuous and let* $\xi \in \mathbb{R}^m$ *be a random variable. Then the mapping* $\varphi(x) := \mathbb{P}[g(x,\xi) \geq 0]$ *is also upper semi continuous.*

Continuity

Lemma

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ *be (jointly) continuous and let* $\xi \in \mathbb{R}^m$ *be a random variable admitting a density with respect to the Lesbesgue measure in* $\mathbb{R}^m$*. Assume that the sets* $N_x = \{z \in \mathbb{R}^m : g(x,z) = 0\}$ *are Lesbesgue-null sets for all* $x \in \mathbb{R}^n$*. Then the mapping* $\varphi(x) := \mathbb{P}[g(x,\xi) \geq 0]$ *is continuous.*

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

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 $\{z \in \mathbb{R}^m : g(x,z) \geq 0\}$ 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:

$$\frac{\partial F_\xi}{\partial z_i}(z) = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m), i = 1, ..., m. \tag{7}$$

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)van Ackooij, Henrion, Möller, a

- $\varphi(x) := \mathbb{P}[A\xi \leq \alpha(x)]$ ([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 :

$$\varphi(x) := \mathbb{P}[g(x, \xi) \leq 0] \geq p, \tag{8}$$

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*

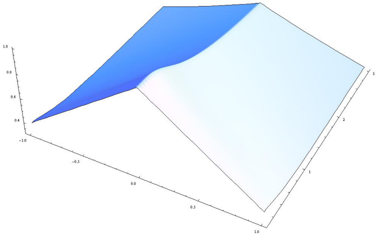
$$g(x_1, x_2, z_1, z_2) := x_1^2 e^{h(z_1)} + x_2 z_2 - 1, \quad \textit{where} \quad h(t) := -1 - 2\log(1 - \Phi(t))$$

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

$$\mathbb{P}[\xi \in M] = \int_{v \in \mathbb{S}^{m-1}} \mu_\eta \left(\{ r \geq 0 : rLv \cap M \neq \emptyset \} \right) d\mu_\zeta \tag{9}$$

- 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 $p = 1$

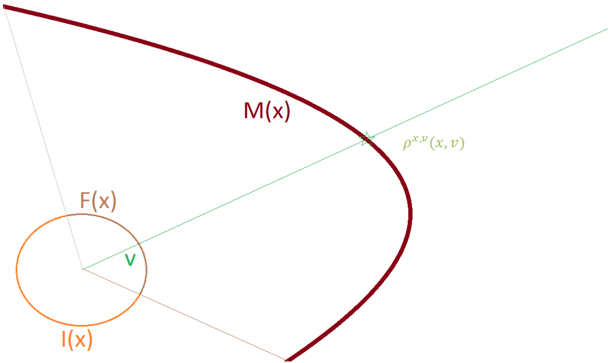
- We define the sets of finite and infinite directions:

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

- 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*

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

The case $p > 1$

- When $p > 1$ we can define
$$g^m(x,z) = \max_{j=1,\dots,p} g_j(x,z), \tag{10}$$
- 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

$$\partial^c \varphi(x) \subseteq \int_{v \in F(x)} \mathrm{Co}\left\{ -\frac{\chi(\hat{\rho}(x,v)) \nabla_x g_j(x, \hat{\rho}(x,v) Lv)}{\langle \nabla_z g_j(x, \hat{\rho}(x,v) Lv), Lv \rangle} \middle| j \in \hat{\mathcal{J}}(x,v) \right\} d\mu_\zeta(v) \tag{11}$$

for all $x \in U$. Here,

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

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

$$\mathcal{I}(x,z) := \{j \in \{1,\ldots,p\} \,|\, g_j(x,z) = 0\} \tag{12}$$

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

$$\operatorname{rank} \{\nabla_z g_j(x,z), \nabla_z g_i(x,z)\} = 2 \quad \forall i,j \in \mathcal{I}(x,z), i \neq j \tag{13}$$

$$\forall z \in \mathbb{R}^m : g(x,z) \leq 0. \tag{14}$$

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:*

$$\nabla\varphi(\bar{x}) = -\int_{v \in F(\bar{x}), \#\hat{\mathcal{J}}(\bar{x},v)=1} \frac{\chi\left(\hat{\rho}\left(\bar{x}, v\right)\right) \nabla_x g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right)}{\left\langle \nabla_z g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right), Lv\right\rangle} d\mu_\zeta(v), \quad (15)$$

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

$$\varphi(x) := \mathbb{P}[B\xi \leq h(x)], \tag{16}$$

  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.

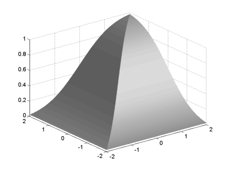
Motivation

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

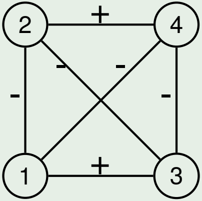
Let Σ be an $m \times m$ covariance matrix having all diagonal entries equal to 1. Let $G(\Sigma) = (V, E)$ denote the (undirected) graph on the vertex set $V = \{1, ..., m\}$ and with edge set $E = E^+ \cup E^- = \{(i,j) \ : i \neq j, \Sigma_{ji} = 1\} \cup \{(i,j) \ : i \neq j, \Sigma_{ji} = -1\}$.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

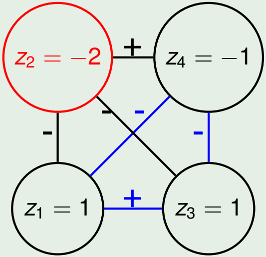
The arc $(i, j) \in E$ will occasionally be referred to as an auto-referencing (a changeable) arc with respect to z if z is auto-referenced (changeable).

Correlation graph: Example

Example

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

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



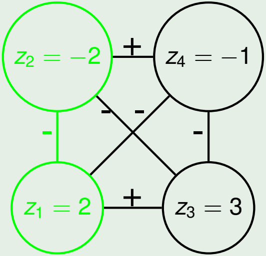
Then z is auto-referenced (blue), but not changeable ($-2 \geq -1$ is false).

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:

$$v_i = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m). \tag{17}$$

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 $q = 1, \dots, Q$ components of the correlation graph and (L^q, R^q) be the associated bipartition. Let z be changeable. Define $J \subseteq \{1, \dots, Q\}$ as the set of all q for which either $|V^q| = 1$ or no changeable arc exists in V^q. For each remaining $q \in \{1, \dots, Q\} \setminus J$, 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:

$$v_i = \begin{cases} f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_m) & \text{if} \quad i \in \cup_{j \in J} V^j \\ f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_m) & \text{if} \quad \exists q \in \{1, \dots, Q\} \setminus J, i \in \{l^q, r \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

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

$$\partial^c F_\xi(z) = \mathrm{co}\,(\mathcal{E}(z)), \tag{19}$$

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

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

Definition of Generalized Concavity I

- The following mapping plays a key role in the definition of generalized concavity:

Definition

Let $\alpha \in [-\infty, \infty]$ and $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \to \mathbb{R}$ be defined as follows

$$m_\alpha(a, b, \lambda) = 0 \text{ if } ab = 0, \tag{20}$$

for $a > 0$, $b > 0$, $\lambda \in [0,1]$:

$$m_\alpha(a, b, \lambda) = \begin{cases} a^\lambda b^{1-\lambda} & \text{if} \quad \alpha = 0 \\ \max\{a, b\} & \text{if} \quad \alpha = \infty \\ \min\{a, b\} & \text{if} \quad \alpha = -\infty \\ (\lambda a^\alpha + (1-\lambda) b^\alpha)^{\frac{1}{\alpha}} & \text{else} \end{cases} \tag{21}$$

Definition of Generalized Concavity II

- We can now define generalized concavity of a mapping f:

Definition

A non-negative function f defined on some convex set $C \subseteq \mathbb{R}^n$ is called α-concave ($\alpha \in [-\infty, \infty]$) if and only if for all $x, y \in C, \lambda \in [0, 1]$:

$$f(\lambda x + (1-\lambda)y) \geq m_\alpha(f(x), f(y), \lambda). \tag{22}$$

- For $\alpha = 1$ this is just the definition of concavity. For $\alpha = 0$, f is log-concave and satisfies $f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$.

Some Properties of PCs I

- Convexity of $M(p)$ can be asserted under general conditions on g, ξ regardless of p

Theorem ([Prékopa(1972), Prékopa(1973), Tamm(1977), Borell(1975), Brascamp and Lieb(1976)])

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ be a (jointly) quasi-concave function and let $\xi \in \mathbb{R}^m$ be a random variable inducing an α-concave probability distribution $\mathbb{P}$. Then the mapping $x \in \mathbb{R}^n \mapsto G(x) := \mathbb{P}[g(x,\xi) \geq 0]$ is an α-concave function on the set $D = \{x \in \mathbb{R}^n \ : \ \exists z \in \mathbb{R}^m \text{ with } g(x,z) \geq 0\}$.

An example of another type of convexity

Let us consider the following example, wherein $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined as follows:

$$g(x,z) := z^{\mathsf{T}} W(x) z + 2 \sum_{i=1}^{n} x_i w_i^{\mathsf{T}} z + b, \tag{23}$$

where $W : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^m$ a positive semi-definite matrix valued mapping.

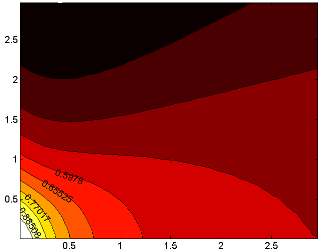
$W(x) = x_1 W_1 + x_2 W_2$, where

$$W_1 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \text{ and } W_2 = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}.$$

Moreover the correlation matrix R is taken to be:

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

Finally we take $w_1 = (-1, 1)$, $w_2 = (2, 3)$ and $b = -3$



Some Properties of PCs II

- Eventual convexity is defined as convexity of $M(p)$ for all $p > p^*$. A classic result:

Lemma ([Kataoka(1963)])

Consider the constraint of the form $\varphi(x) \geq p$ where $k = 1$, $g(x,z) = z^{\top}x - b$ and $\xi \in \mathbb{R}^m$ is a multivariate Gaussian random variable. Then the feasible set $M(p)$ is convex for all $p > \frac{1}{2}$.

- Recent important eventual convexity results for $M(p)$ involving specially structured probabilistic constraints have been derived by [Henrion and Strugarek(2008)], [Henrion and Strugarek(2011)].

Setting

- In practice, an important question concerns convexity of the set

$$M(p) := \left\{ x \in \mathbb{R}^n \ : \mathbb{P}[\xi \leq h(x)] \geq p \right\}. \tag{24}$$

- When h has weaker concavity properties (e.g., only log-concave), the classic results can't be applied (directly).
- We are interested in identifying a computable threshold p^* such that $M(p)$ can be shown to be convex provided $p \geq p^*$: eventual convexity

Structure

- We assume that $\varphi(x) := \mathbb{P}[\xi \leq h(x)]$ can be cast into the following form:

$$\varphi(x) := C(F_1(h_1(x)), ..., F_m(h_m(x))), \tag{25}$$

  where $C : [0, 1]^m \to [0, 1]$ is a Copula.

- The component ξ_i is assumed to have one dimensional distribution function $z \in \mathbb{R} \mapsto F_i(z) := \mathbb{P}[\xi_i \leq z]$, $i = 1, ..., m$.
- A copula is the distribution function of a multi-variate random variable with uniformly distributed marginals.
- According to Sklar's Theorem, every joint probability distribution can be associated with a Copula

Definition

- We introduce a family of copulae:

Definition

- Let $\gamma \in \mathbb{R}$ be given, and let the set $X(\gamma)$ be defined as $X(\gamma) = [0,1]^m$ for $\gamma > 0$, $X(0) = (-\infty, 0]^m$ and $X(\gamma) = [1, \infty)^m$ for $\gamma < 0$.
- Let $\delta \in [-\infty, \infty]$ be equally given.

We call a Copula $C : [0,1]^m \to [0,1]$ δ-γ-concave if the mapping $u \in X(\gamma) \mapsto C(u^{\frac{1}{\gamma}})$ is δ-concave, whenever $\gamma \neq 0$ and $u \in X(0) \mapsto C(e^u)$ is δ-concave whenever $\gamma = 0$.

Relation

- A logexp-concave copulae C ([Henrion and Strugarek(2011)]) is 0-0-concave. The concept of δ-γ-concavity is a direct extension.
- A quasi-concave copula is $-\infty$-1-concave.
- It is sufficient for δ-γ-concavity to hold locally (Not shown here for notational convenience).

Structure I

- The γ parameter has an ascending effect:

Lemma

Let C $: [0,1]^m \to [0,1]$ *be a* δ-β*-concave Copula and let* $\alpha \in \mathbb{R}$ *be given such that* $\beta \leq \alpha$*. Then C is also* δ-α*-concave.*

Structure II

- The δ parameter has a descending effect as is well-known.

Corollary

Let $C : [0,1]^m \to [0,1]$ be a δ-γ-concave Copula and let $\alpha \geq \gamma$ and $\beta \leq \delta$ be given. Then C is also β-α-concave.

- This shows that the strongest characterization is obtained when γ is smallest and δ highest.

General observation

- The family contains all Archimedian copulae

Proposition ([van Ackooij and de Oliveira(2016)])

Let $C : [0,1]^m \to [0,1]$ *be an Archimedian copula, and* $\psi : (0,1] \to [0,\infty)$ *be its generator. Then* C *is a* $-\infty$-1*-concave copula.*

Stronger characterizations

- The independent, maximum and Gumbel Copula are 0-0-concave ([Henrion and Strugarek(2011)])
- The Clayton copula can also be characterized in a stronger way:

Lemma ([van Ackooij(2015)])

Let $\theta > 0$ be the parameter of the strict generator $\psi : [0,1] \to \mathbb{R}_+$, $\psi(t) = \theta^{-1}(t^{-\theta} - 1)$ of the Clayton Copula. This Copula is δ-γ-concave for all $\gamma > 0$ provided that $\delta \leq -\theta < 0$.

Main result I

Theorem ([van Ackooij(2015)])

Assume that we can find $\alpha_i \in \mathbb{R}$, such that the functions h_i are α_i-concave and a second set of parameters $\gamma_i \in (-\infty, \infty]$, $b_i > 0$ such that either one of the following conditions holds:

1. $\alpha_i < 0$ *and* $z \mapsto F_i(z^{\frac{1}{\alpha_i}})$ *is* γ_i*-concave on* $(0, b_i^{\alpha_i}]$
2. $\alpha_i = 0$ *and* $z \mapsto F_i(\exp z)$ *is* γ_i*-concave on* $[\log b_i, \infty)$
3. $\alpha_i > 0$ *and* $z \mapsto F_i(z^{\frac{1}{\alpha_i}})$ *is* γ_i*-concave on* $[b_i^{\alpha_i}, \infty)$,

where $i \in \{1, ..., m\}$ *is arbitrary. If the Copula is* δ-γ*-concave for* $\gamma \leq \gamma_i \leq \infty$, $i = 1, ..., m$, *then the set*

$$M(p) := \{x \in \mathbb{R}^n \ : \mathbb{P}[\xi \leq h(x)] \geq p\}$$

is convex for all $p > p^* := \max_{i=1,\dots,m} F_i(b_i)$. *Convexity can moreover be derived for all* $p \geq p^*$ *if each individual distribution function* F_i, $i = 1, ..., m$ *is strictly increasing. In the specific case that* $\alpha_i \geq 0$, γ_i*-concavity of the distribution functions holding everywhere, for all* $i \in \{1, ..., m\}$ *and* C *being a* δ-γ*-concave Copula, the set* $M(p)$ *is convex for all* p.

Some comments

- An “unfortunate” effect in the previous result is that p^* depends somehow on the "worst" distribution function F_i, but is only needed for a single inequality.
- Generalized concavity of the mappings h_i need only hold on specific level sets $\{x \in \mathbb{R}^n \; : h_i(x) \geq b_i\}$.

Main result II

- These concerns are addressed in the following result:

Theorem ([van Ackooij(2015)])

Define the set $D := \{x \in \mathbb{R}^n : h_i(x) \geq b_i, \forall i = 1, ..., m\}$, *where* b_i *is as defined in the previous Theorem and we make the same assumptions on* ξ, F_i *and the Copula. Then the set* D *is convex and* $D \cap M(p)$ *is convex for all* $p \geq p^* = C(F_1(b_1), ..., F_n(b_n))$.

Discussion

The conditions of the theorem can be shown to hold in many situations:

- Generalized concavity properties of mappings h_i are known from data; Implicit from some underlying nominal "deterministic" problem involving constraints $h_i(x) \geq b_i$.
- The requests on the marginal distribution functions follow from results in [Henrion and Strugarek(2008)] for nearly all choices
- The class of δ-γ-concave copula cover at least all Archimedean copula.

1 Introduction
- Motivation

2 Continuity
- Introductory discussion
- Continuity statements

3 Differentiability
- Classics
- Consideration of non-linear mappings
- More components
- A further characterization of Clarke's sub-differential

4 Eventual convexity
- Introduction
- The tools: special family of copulæ
- Main Results

5 Algorithms
- Introduction
- Level bundle methods
- Dedicated method and results

Cutting plane models

- Consider the problem of minimizing a convex mapping $f : \mathbb{R}^n \to \mathbb{R}$
- f is only known partially through a "black box" called oracle. Given an entry x_i, it returns $f(x_i)$ and $g_i \in \partial f(x_i)$.
- With a set of points $x_1, ..., x_k$, we can build the cutting plane model for f:

$$\check{f}_k(x) := \max_{j=1,\dots,k} \{f(x_j) + \langle g_j, x - x_j \rangle\} \tag{26}$$

- Convexity yields: $\check{f}_k(x) \leq f(x)$ for all x (and $k \geq 1$).

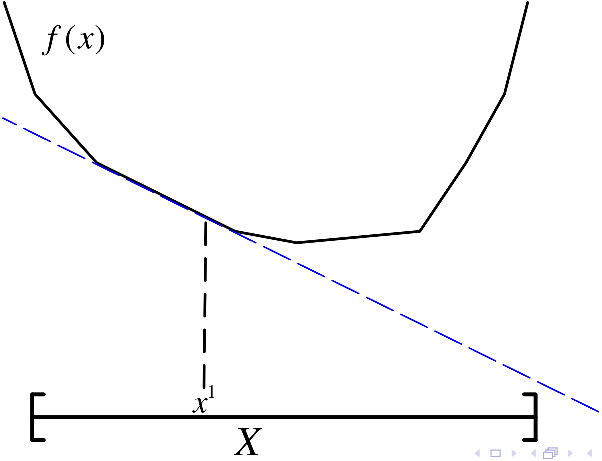
Cutting plane methods

- So instead of minimizing f over a "simple" set X, we solve

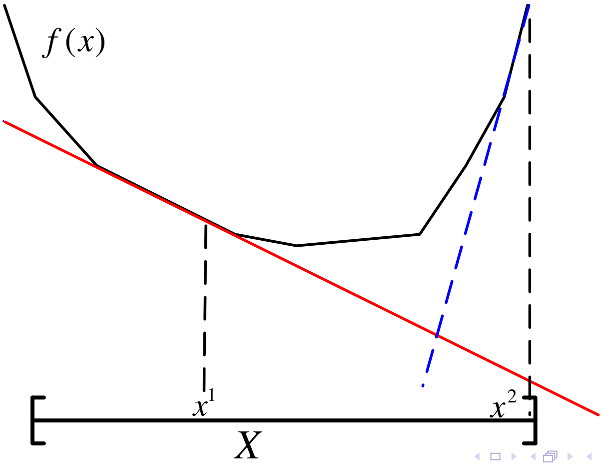
$$x_{k+1} = \operatorname*{argmin}_{x \in X} \check{f}_k(x). \tag{27}$$

- when X is polyhedral this is a linear program.

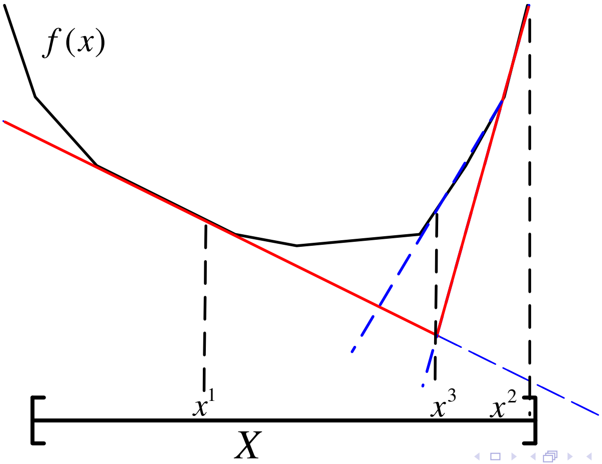
Cutting-plane method: illustration



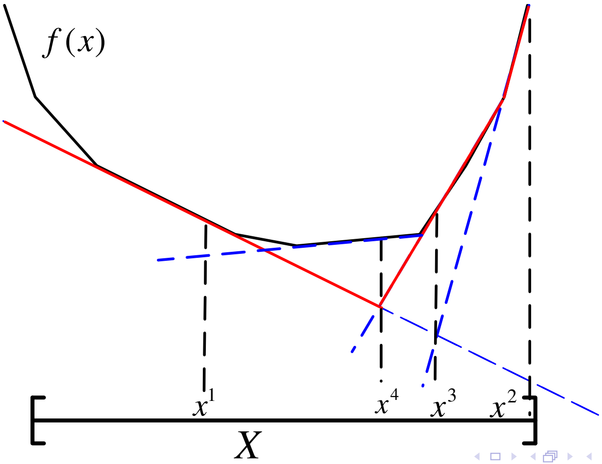
Cutting-plane method: illustration



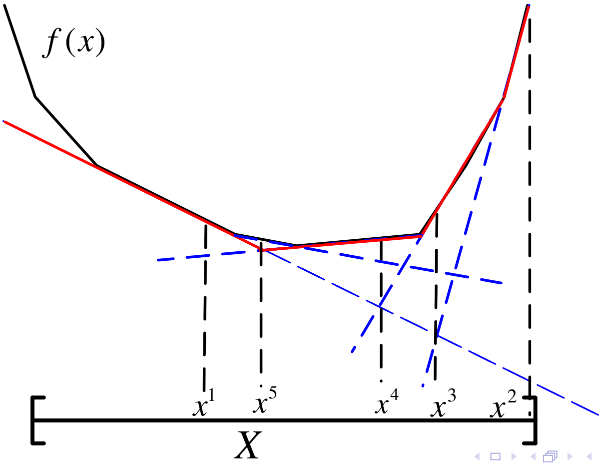
Cutting-plane method: illustration



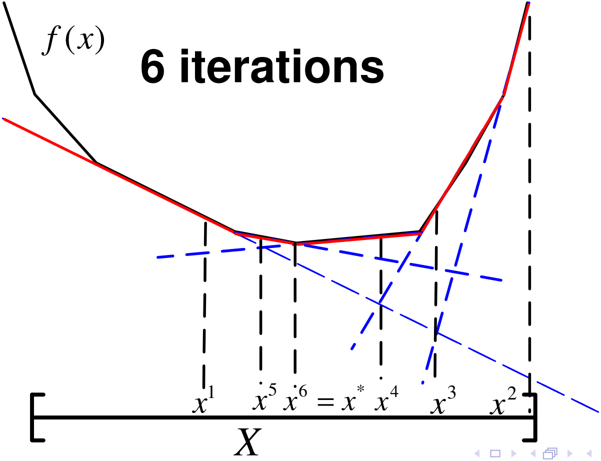
Cutting-plane method: illustration



Cutting-plane method: illustration



Cutting-plane method: illustration



Convex constrained problems

- Consider the problem

$$\min_{x \in X} \{f(x) \ : \text{s.t. } \varphi(x) := \mathbb{P}[g(x,\xi) \leq 0] \geq p\} \tag{28}$$

- Then under appropriate assumptions, $x \mapsto \varphi(x)$ has convex level sets, e.g., φ could be log-concave
- Now, $c(x) = \log(p) - \log(\varphi(x))$ is a convex map.
- the problem is a convex constrained problem (under the appropriate assumptions)

Supporting hyperplane method

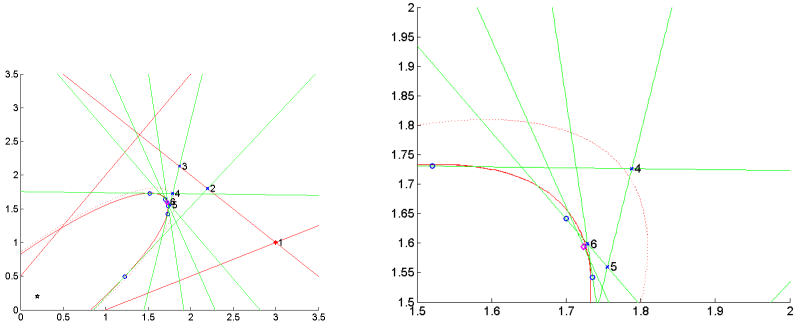
- A classic method in chance constrained programming.
- We suppose available a slater point x^s, i.e., such that $\varphi(x^s) > p$.
- At iteration k we solve $\min_{x \in X} \left\{ \check{f}_k(x) \ : \check{c}_k(x) \leq 0 \right\}$ to find $\tilde{x}_{k+1}$.
- Typically $\tilde{x}_{k+1}$ is not feasible, so we compute the largest $\lambda \in [0,1]$ such that $x_{k+1} = \lambda x^s + (1-\lambda)\tilde{x}_{k+1}$ satisfies $\varphi(x_{k+1}) = p$.
- We have upper and lower bounds on the optimal value and stop whenever these are close enough.

Are chance constraints just plain non-linear constraints?

- In a way yes, but the mapping $\varphi(c)$ is not known up to arbitrary precision (or would be unreasonably costly). A (sub-)gradient of $\varphi(c)$ also suffers from numerical imprecision. Here we make use of the earlier derived formula allowing for efficient and precise computations. Again with a trade-off cost/Efficiency
- So then is $\check{c}$ a true cutting plane model for c underestimating it ?

Special methods for chance constraints

- An example shows that cutting planes may locally over-estimate the map (or set):



Upper-oracle

- We can set up an “upper"-oracle for constraints of type c and specially structured probability constraints. These may provide a cutting planes model with cutting planes:

$$\begin{aligned} c_x &= c(x) - \eta_c^x \\ c(y) &\geq c_x + \langle g_c^x, y - x \rangle - \varepsilon_c^x, \end{aligned}$$

  having $\varepsilon_c^x > 0$.

- ε_c^x can be shown to have a link with the precision used in evaluating probabilities $\mathbb{P}[g(x,\xi) \geq 0]$.

Special methods

- So we need a method capable of handling inaccuracy of $\varphi(c)$ explicitly
- A method able to account for the flaws of cutting plane methods: (oscillation, slow convergence (for high accuracy solutions))
- Lets look at special bundle methods

Level bundle method: main ideas

What characterizes a level bundle method are essentially:

- [(i)] a convex model $\check{f}_k(x) \leq f(x)$;
- [(ii)] a stability center $\hat{x}_k$;
- [(iii)] a parameter $f_k^{\texttt{lev}}$ to be updated at each iteration k.

The new iterate x_{k+1} is obtained by solving a projection problem

$$x_{k+1} := \operatorname{argmin}\left\{\frac{1}{2}\|x - \hat{x}_k\|^2 : \ \check{f}_k(x) \leq f_k^{\texttt{lev}}, \ x \in X\right\}$$

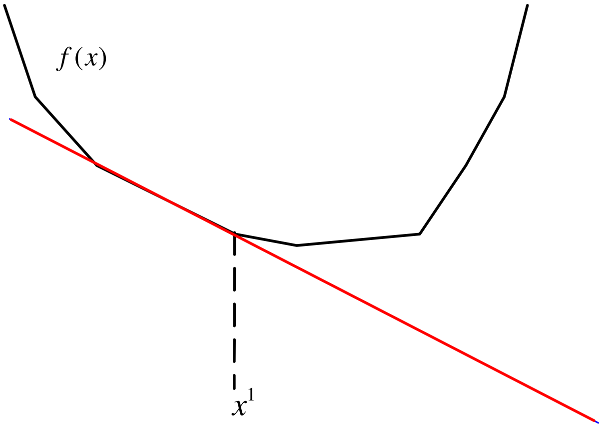
Level bundle method: some elements

Definitions

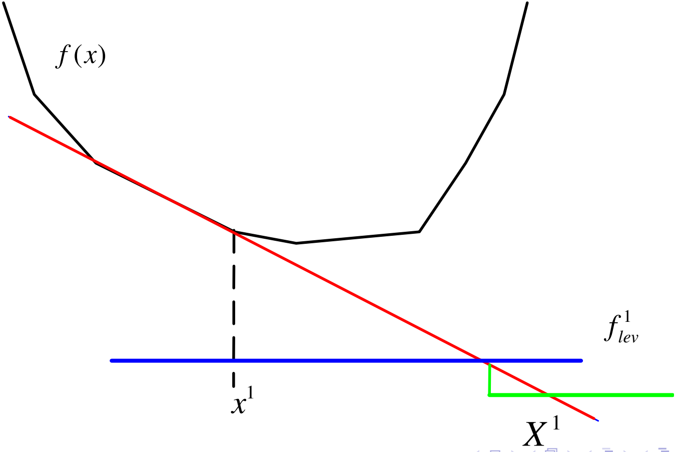
$f_k^{\mathtt{up}}$	$:= \min_{1\le j\le k} f(x_j)$	is an upper bound for f_*
$f_k^{\mathtt{low}}$	$:= \min_{x\in X} \check{f}_k(x)$	is a lower bound for f_*
$f_k^{\mathtt{lev}}$	$:= \lambda f_k^{\mathtt{up}} + (1-\lambda) f_k^{\mathtt{low}}$	is the level parameter, for $\lambda \in (0,1)$
$\mathbb{X}_k$	$:= \{x \in X : \check{f}_k(x) \le f_k^{\mathtt{lev}}\}$	is the level set of $\check{f}_k$
Δ_k	$:= f_k^{\mathtt{up}} - f_k^{\mathtt{low}}$	is an optimality gap

- Solving the LP defining $f^{\mathtt{low}}$ is optional

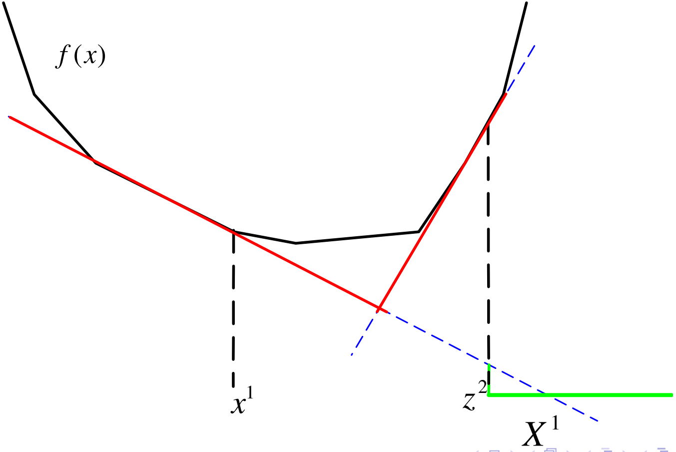
Level bundle method: illustration



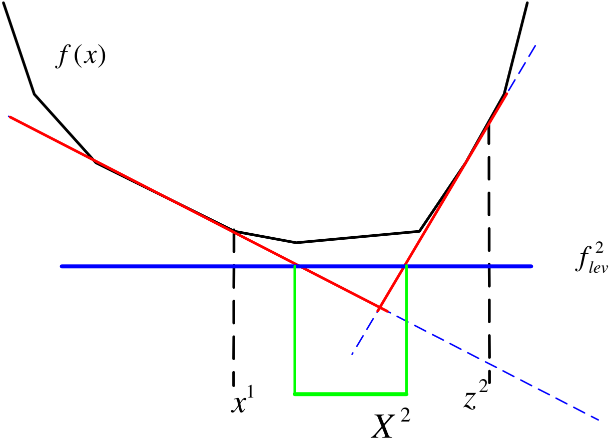
Level bundle method: illustration



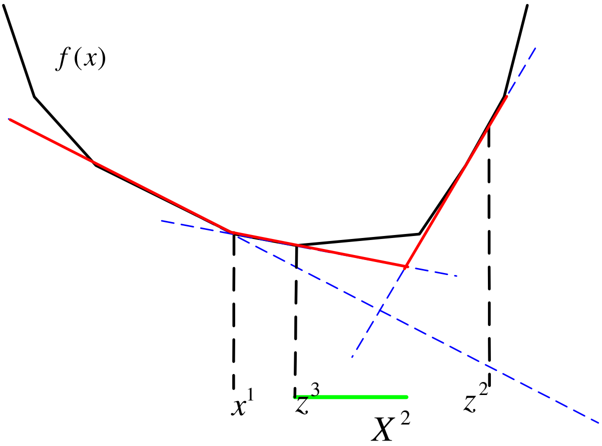
Level bundle method: illustration



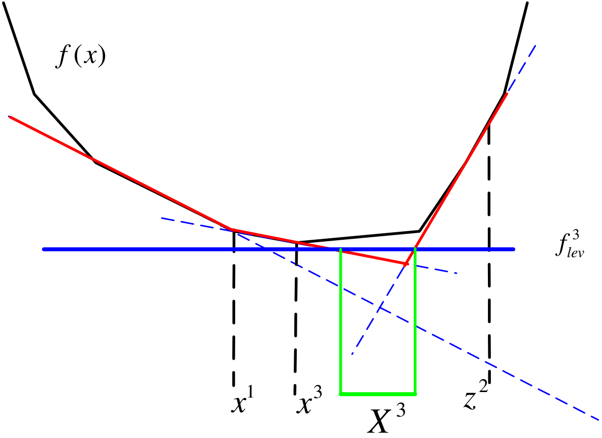
Level bundle method: illustration



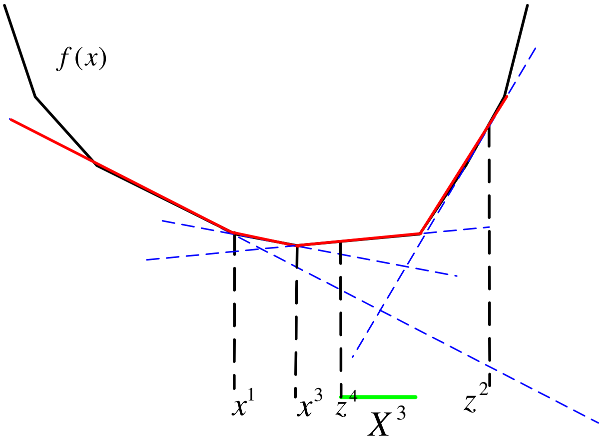
Level bundle method: illustration



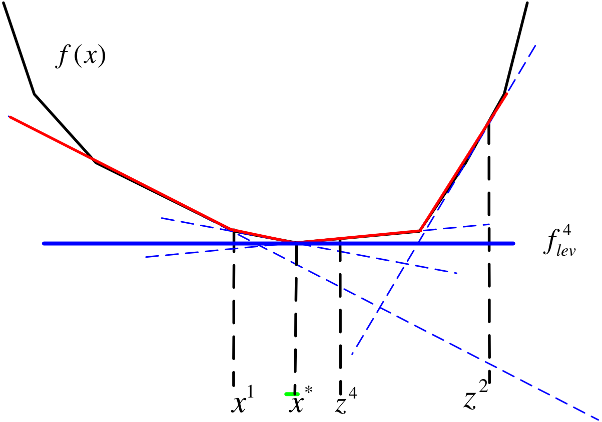
Level bundle method: illustration



Level bundle method: illustration



Level bundle method: illustration



Measure of optimality

- We define the improvement function $h(x; f_k^{\mathtt{low}}) := \max\left\{f(x) - f_k^{\mathtt{low}}, c(x)\right\}$.
- The optimality measure is

$$h_k^{\mathtt{rec}} := \begin{cases} h(x_0, f_0^{\mathtt{low}}) & \text{if} \quad k = 0 \\ \min\left\{\left(\min_j h(x_j, f_k^{\mathtt{low}})\right), h_{k-1}^{\mathtt{rec}}\right\} & \text{if} \quad k > 0 \end{cases} \tag{29}$$

- $x_k^{\mathtt{rec}}$ is the past iterate such that $h(x_k^{\mathtt{rec}}; f_k^{\mathtt{low}}) = h_k^{\mathtt{rec}}$. It is the sequence of best solutions.

The algorithm

After initialization, the algorithm moves through the following steps

- (Best minimizer) Update $x_k^{\mathtt{rec}}$ and $h_k^{\mathtt{rec}}$
- (Stopping test) If $h_k^{\mathtt{rec}} < \delta$ is sufficiently small, then stop
- (Level update): Compute $f_k^{\mathtt{lev}}$
- (Projection problem): Compute x_{k+1}
- (Oracle): call the oracle to update the models
- (Bundle Management): Optionally remove old linearizations

The algorithm: convergence, optimality certificate

Lemma ([van Ackooij and de Oliveira(2014)])

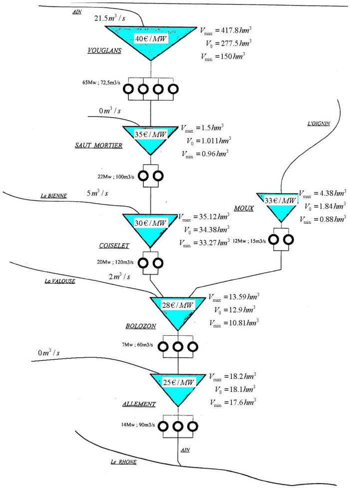
If $\lim_k h_k^{\mathtt{rec}} \leq 0$, then any cluster point of the sequence $x_k^{\mathtt{rec}}$ is an η-optimal solution to the problem, with $\eta := \max\{\eta_f + \varepsilon_f, \eta_c\}$. In particular if $h_k^{\mathtt{rec}} \leq 0$ for some (finite) k, $x_k^{\mathtt{rec}}$ is an η-optimal solution.

The algorithm: convergence

Theorem ([van Ackooij and de Oliveira(2014)])

The algorithm with an upper oracle (and with $\delta = 0$) will either stop or generate a sequence of points such that $\lim_k h_k^{\mathtt{rec}} \leq 0$.

Numerical example : cascaded reservoir management



- A network flow problem when the valley acts on a price signal, the latter typically comes from a Lagrangian dual.
- Water values, i.e., costs are assumed pre-computed and can be volume dependent
- Constraints imply:
  - Production level bounds
  - Reservoir bounds have to be satisfied
- When Inflows are deterministic, the above problem is linear. Inflows are however stochastic

Problem Structure

- Reservoir bound constraints need to be interpreted "somehow" in a stochastic setting. We will use (joint-)chance constrained programming in order to do so. This gives the problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n_+} \quad & c^{\mathsf{T}} x \\ s.t. \quad & Ax \leq b \\ & p \leq \mathbb{P}[a^r + A^r x \leq \xi \leq b^r + A^r x], \end{aligned} \tag{30}$$

- If Inflows follow a Causal time series model with Gaussian innovations, then ξ above is Gaussian as well. In particular problem (30) has a convex feasible set.

Numerical example: benchmark I

Table: In all computations: precision of oracle $\varepsilon^g = 5e^{-4}$

Instance	method	Obj. Value	$\mathbb{P}$	Nb. Iter. [InfeasQP]	CPU time (mins)	parameters
lsr48	Alg.[Prékopa(2003)]	-175031	0.799975	35	10.5	$\delta_{\texttt{Tol}} = 1e^{-4}$
lsr48	Alg.[Kiwiel(2008)]	-175042	0.799885	49	7.5	$K = 1e^4$, $\delta_{\texttt{Tol}} = 1e^{-5}$
lsr48	Alg.PB	-175043	0.799145	88	11.2	$K = 1e^5$,$\mu_0 = 1e^{-5}$, $\delta_{\texttt{Tol}} = \frac{1}{2}$
lsr48	Alg.PB	-175042	0.799536	69	8.3	$K = 1e^5$,$\mu_0 = 1e^{-6}$, $\delta_{\texttt{Tol}} = \frac{1}{2}$
lsr48	Alg.PB	-175042	0.799588	31	4.5	$K = 1e^5$,$\mu_0 = 1e^{-8}$, $\delta_{\texttt{Tol}} = \frac{1}{2}$
lsr48	Alg.LB	-175039	0.800041	66	10.0	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\texttt{Tol}} = 5$
lsr48	Alg.LB	-175040	0.799755	38	5.4	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\texttt{Tol}} = 5$
lsr48	Alg.LB	-175040	0.799855	63 [3]	8.6	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\texttt{Tol}} = 5$, $[\neg\texttt{LP}]$
lsr48	Alg.LB	-175037	0.79966	38 [4]	5.2	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\texttt{Tol}} = 5$, $[\neg\texttt{LP}]$

Numerical example: benchmark II

- Bundle Methods offer computational advantages over Cutting Planes methods, mainly if the instance is hard (e.g., case of Ain48, Isr96, Isr168). It does not show much for Isr48.
- The Level Method has an easier parameter setup “globally” .
- The Proximal Method (see [van Ackooij and Sagastizábal(2014)]) produces feasible solutions quickly

Numerical example: benchmark III

Instance	method	Obj. Value	$\mathbb{P}$	Nb. Iter. [InfeasQP]	CPU time (mins)	parameters
Isr96	Alg.[Prékopa(2003)]	-175708	0.799492	143	217.4	$\delta_{\mathtt{Tol}} = 1e^{-4}$
Isr96	Alg.[Kiwiel(2008)]	-175713	0.799541	127	86.8	$K = 1e^4$, $\delta_{\mathtt{Tol}} = 1e^{-5}$
Isr96	Alg.PB	-175715	0.799413	159	110.9	$K = 1e^5$,$\mu_0 = 1e^{-5}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr96	Alg.PB	-175715	0.799406	177	123.5	$K = 1e^5$,$\mu_0 = 1e^{-6}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr96	Alg.PB	-175713	0.799346	95	66.5	$K = 1e^5$,$\mu_0 = 1e^{-8}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr96	Alg.LB	-175713	0.799874	122	82.5	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$
Isr96	Alg.LB	-175713	0.799599	94	48.4	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$
Isr96	Alg.LB	-175710	0.799809	115 [3]	75.3	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$, [$\neg$LP]
Isr96	Alg.LB	-175697	0.799866	76 [4]	44.3	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$, [$\neg$LP]
Isr168	Alg.[Prékopa(2003)]	-175222	0.799511	190	1504.7	$\delta_{\mathtt{Tol}} = 1e^{-4}$
Isr168	Alg.[Kiwiel(2008)]	-175237	0.799394	204	627.3	$K = 1e^4$, $\delta_{\mathtt{Tol}} = 1e^{-5}$
Isr168	Alg.PB	-175237	0.799408	219	687.4	$K = 1e^5$,$\mu_0 = 1e^{-5}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr168	Alg.PB	-175237	0.799418	188	573.5	$K = 1e^5$,$\mu_0 = 1e^{-6}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr168	Alg.PB	-175236	0.799503	133	343.4	$K = 1e^5$,$\mu_0 = 1e^{-8}$, $\delta_{\mathtt{Tol}} = \frac{1}{2}$
Isr168	Alg.LB	-175235	0.799854	161	529.6	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$
Isr168	Alg.LB	-175232	0.799717	110	352.3	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$
Isr168	Alg.LB	-175235	0.799604	165 [3]	423.2	$K = 1e^5$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$, [$\neg$LP]
Isr168	Alg.LB	-175220	0.799423	127 [5]	353.5	$K = 1e^4$, $\gamma = 0.8$, $\delta_{\mathtt{Tol}} = 5$, [$\neg$LP]

Summary

In this talk we have discussed several aspects of chance constraints

- Differentiability
- Convexity
- Algorithms

Thank you for your attention!

- Time for questions
- Special thanks to Welington de Oliveira for the cutting-plane / bundle illustrations.

Some references

Obviously the main reference is

- A. Prékopa. *Stochastic Programming.* Kluwer, Dordrecht, 1995

Some further references I

The works cited are:

- 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

Some further references II

The works cited are:

- W. van Ackooij. Eventual convexity of chance constrained feasible sets. *Optimization (A Journal of Math. Programming and Operations Research)*, 64(5):1263–1284, 2015.
  doi: 10.1080/02331934.2013.855211

- W. van Ackooij and W. de Oliveira. Level bundle methods for constrained convex optimization with various oracles.
  Computation Optimization and Applications, 57(3):555–597, 2014

- W. van Ackooij and C. Sagastizábal. Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems.
  SIAM Journal on Optimization, 24(2):733–765, 2014

Bibliography I

[[Borell(1975)]]C. Borell.
Convex set functions in d-space.
Periodica Mathematica Hungarica, 6:111–136, 1975.

[[Brascamp and Lieb(1976)]]H.J. Brascamp and E.H. Lieb.
On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log-concave functions and with an application to the diffusion equations.
Journal of Functional Analysis, 22:366–389, 1976.

[[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.

Bibliography II

[[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.

[[Henrion(2010)]]R. Henrion.
Optimierungsprobleme mit wahrscheinlichkeitsrestriktionen: Modelle, struktur, numerik.
Lecture Notes, page 43, 2010.

Bibliography III

[[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.

[[Henrion and Strugarek(2008)]]R. Henrion and C. Strugarek.
Convexity of chance constraints with independent random variables.
Computational Optimization and Applications, 41:263–276, 2008.

[[Henrion and Strugarek(2011)]]R. Henrion and C. Strugarek.
Convexity of chance constraints with dependent random variables: the use of copulae.
In M. Bertocchi, G. Consigli, and M.A.H. Dempster, editors, *Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies*, International Series in Operations Research and Management Science, pages 427–439. Springer-Verlag New York, 2011.

Bibliography IV

[[Kataoka(1963)]]]S. Kataoka.
A stochastic programming model.
Econometrica, 31:181–196, 1963.

[[Kiwiel(2008)]]]K.C. Kiwiel.
A method of centers with approximate subgradient linearizations for non-smooth convex optimization.
SIAM Journal on Optimization, 18(4):1467–1489, 2008.

[[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(1972)]]]A. Prékopa.
A class of stochastic programming decision problems.
Matematische Operations forschung und Statistik, 3:349–354, 1972.

Bibliography V

[[Prékopa(1973)]]A. Prékopa.
On logarithmic concave measures and functions.
Acta Scientiarium Mathematicarum (Szeged), 34:335–343, 1973.

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

[[Prékopa(2003)]]A. Prékopa.
Probabilistic programming.
In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267–351. Elsevier, Amsterdam, 2003.

[[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 VI

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

[[Tamm(1977)]]E. Tamm.
On *g*-concave functions and probability measures (russian).
Eesti NSV Teaduste Akademia Toimetised, Füüsika-Matemaatika, 28:17–24, 1977.

[[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.

Bibliography VII

[[van Ackooij(2015)]]W. van Ackooij.
Eventual convexity of chance constrained feasible sets.
Optimization (A Journal of Math. Programming and Operations Research), 64(5):1263–1284, 2015.
doi: 10.1080/02331934.2013.855211.

[[van Ackooij and de Oliveira(2014)]]W. van Ackooij and W. de Oliveira.
Level bundle methods for constrained convex optimization with various oracles.
Computation Optimization and Applications, 57(3):555–597, 2014.

[[van Ackooij and de Oliveira(2016)]]W. van Ackooij and W. de Oliveira.
Convexity and optimization with copulæ structured probabilistic constraints.
Optimization: A Journal of Mathematical Programming and Operations Research, 65(7):1349–1376, 2016.
doi: 10.1080/02331934.2016.1179302.

Bibliography VIII

[[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.

[[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.

Bibliography IX

[[van Ackooij and Sagastizábal(2014)]]W. van Ackooij and C. Sagastizábal.
Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems.
SIAM Journal on Optimization, 24(2):733–765, 2014.

[[van Ackooij et al.(2011)van Ackooij, Henrion, Möller, and Zorgati]]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.