Gradient Formulae for bilateral degenerate probabilistic constraints and applications

W. van Ackooij[1] R. Henrion[2] M. Minoux[3] P. Nzi[4]

[1] OSIRIS Department
EDF R&D
1 Avenue du Général de Gaulle; 92141 Clamart ; France

[2] Nonlinear Optimization and Inverse Problems
WIAS Berlin

[3] LIP6, Université Pierre et Marie Curie
France

[4] PGMO funded internship

Conférence PGMO, 2013

Outline

Motivation

- Consider the Probabilistic constraint :

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

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

- Such constraints (not necessarily with Gaussian random-variables) arise in many applications. For instance cascaded Reservoir management, Network optimization with random nodal injections, etc...
- Solving Optimization problems in practice with constraints (1) requires algorithms from non-linear programming.
- These Algorithms (mostly) require at least the knowledge of a gradient in trial points.

Motivation II

- Evaluating the probability (1) can not be done up to arbitrary precision (and is useless)
- In general it is not a good idea to use finite differences for computing an estimate of the gradient
- A criterion of success would be to obtain a formula with computation complexity of the same order of magnitude as that for computing the probability.

Known Special Cases

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

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

  where $\Sigma \succ 0$

- $\varphi(x) := \mathbb{P}[A(x)\xi \leq \alpha(x)]$ ([van Ackooij et al.(2011)]), $A(x)$ surjective
- $\varphi(x) := \mathbb{P}[g(x, \xi) \leq 0]$, $g \in C^1$, convex in the second argument. ([van Ackooij and Henrion(2013)]).
- $\varphi(x) := \mathbb{P}[A\xi \leq \alpha(x)]$ ([Henrion and Möller(2012)]). Here $\Sigma \succeq 0$ but not for all x.
- Other cases involve distribution functions of Dirichlet ([Szántai(1985), Gouda and Szántai(2010)]) and multi-variate Gamma ([Prékopa and Szántai(1979)]) random variables

A simple Example

- Consider a graph with $n-1$ nodes. Each arc can support a positive or negative flow.
- Each node can be either a source or a sink
- Each node has random injections x_j.

A simple Example II

- We add an n-th node to the network, representing the market to ensure a global equilibrium on the network. Let m be the number of arcs in the extended network.
- The flows through the network follows the two Kirchhoff's laws(dc-current approximation)
- We wish to dimension the capacity of the arcs in such a way that a feasible flow through the network exists with high enough probability. Capacity is assumed to have a linear cost.

A simple Example III

- Well known arguments allow us to establish the existence of a $m \times (n-1)$ matrix M, such that $y = Mx$, where y is the flow through each arc. Since x is assumed to be random, so is y.
- We thus obtain the following JCCP problem:

$$\begin{array}{cl} \min_{z \in \mathbb{R}^m_+} & c^{\mathsf{T}} z \\ s.t. & \mathbb{P}[y \in [-z, z]] \geq p \end{array}$$

- Here $z \in \mathbb{R}^m_+$ is the capacity of each arc (to be determined).

The probabilistic Constraint

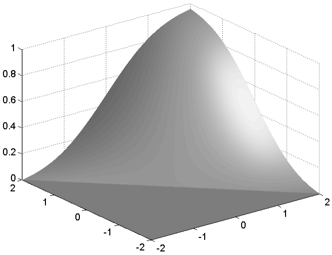
- Let $x \sim \mathcal{N}(\mu, \Sigma)$ and $\Sigma \succ 0$ then $y \sim \mathcal{N}(M\mu, M^{\mathsf{T}}\Sigma M)$.
- However in general $m > (n-1)$ and hence $\operatorname{rank} M < m$. As a consequence $\det(M^{\mathsf{T}}\Sigma M) = 0$.
- We can't use classical results on differentiability of $\varphi(z) := \mathbb{P}[y \in [-z, z]$, which would only hold for most points z.

An Example

- Let $y \sim \mathcal{N}(0, \Sigma), y \in \mathbb{R}^2$ and $\Sigma_{11} = \Sigma_{22} = 1, \Sigma_{12} = -1$. Consider $\varphi(z) = \mathbb{P}[y \leq z]$. It is then readily observed that

$$\varphi(z) = \begin{cases} \Phi(z_1) - \Phi(-z_2) & \text{when} \quad -z_2 \leq z_1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- Yet φ fails to be differentiable:



Some Differentiability results

Theorem ([Henrion and Möller(2012)])

Let ξ be an m-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^m$ and variance-covariance matrix Σ. Let $\Sigma = AA^\mathsf{T}$ be a Cholesky factorization of Σ, i.e., A is an $m \times k$ matrix, where $k = \operatorname{rank} \Sigma$. If $z \in \mathbb{R}^m$ is such that the polyhedron $P(A, z - \mu) := \left\{x \in \mathbb{R}^k \ : Ax \leq z - \mu\right\}$ is non-degenerate, then the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is continuously differentiable at z and the following holds for arbitrary i=1,...,m:

$$\frac{\partial F_\xi}{\partial z_i}(z) = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_m). \tag{3}$$

Here $\tilde{\xi}(z_i)$ is a Gaussian random variable with mean $\hat{\mu} \in \mathbb{R}^{m-1}$ and $(m-1) \times (m-1)$ 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.

A discussion of the Conditions

- The additional non-degeneracy assumption on the polyhedron $P(A, z)$ means that for each $x \in P(A, z)$ the lines of the active indices are linearly independent.
- Using the Theorem in practice would, in principle, mean that one has to check this condition.
- This amounts to enumerating all edges of a polyhedron which is, by nature, a lengthy procedure.

An Extension

Theorem ([van Ackooij and Minoux(2013)])

Let ξ be an m-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^m$ and variance-covariance matrix Σ. Then for arbitrary $z \in \mathbb{R}^m$, the distribution function $F_\xi(z) := \mathbb{P}[\xi \leq z]$ is locally Lipschitz at z and $v \in \partial^0 F_\xi(z)$, where for arbitrary i=1,...,m:

$$v_i = f_{\xi_i}(z_i) F_{\tilde{\xi}(z_i)}(z_1, ..., z_{i-1}, z_{i+1}, ..., z_m). \quad (4)$$

Here $\partial^0 F_\xi(z)$ denotes the Clarke-subdifferential of F_ξ, $\tilde{\xi}(z_i)$ is a Gaussian random variable with mean $\hat{\mu} \in \mathbb{R}^{m-1}$ and $(m-1) \times (m-1)$ covariance matrix $\hat{\Sigma}$. Let D^i_m denote the m-th order identity matrix from which the ith row has been deleted. Then $\hat{\mu} = D^i_m(\mu + \Sigma^{-1}_{ii}(z_i - \mu_i)\Sigma_i)$ and $\hat{\Sigma} = D^i_m(\Sigma - \Sigma^{-1}_{ii}\Sigma_i\Sigma_i^\mathsf{T})(D^i_m)^\mathsf{T}$, where Σ_i is the i-th column of Σ.

A Discussion of the Extension

- While φ fails to be differentiable on a set of zero Lebesgue measure, we can not exclude that an Algorithm produces such a trial point
- Since $c(x) := \log(p) - \log(\mathbb{P}[\xi \leq x]$ is convex, the extension allows us to compute a sub-gradient in those trial points.
- The network optimization problem can safely be handled by methods from non-smooth optimization (e.g., bundle approaches [Lemaréchal(1974)]).

Summary

- We have set up the tools for optimizing Network problems with random nodal injections
- Numerical experiences should still be conducted.

Bibliography I

[[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 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.

[[Lemaréchal(1974)]]C. Lemaréchal.
An algorithm for minimizing convex functions.
Information Processing, 1974:552–556, 1974.

[[Prékopa(1970)]]A. Prékopa.
On probabilistic constrained programming.
Proceedings of the Princeton Symposium on Math. Prog., 28:113–138, 1970.

[[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 II

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

[[van Ackooij and Henrion(2013)]]W. van Ackooij and R. Henrion.
Gradient formulae for nonlinear probabilistic constraints with gaussian and gaussian-like distributions.
Draft submitted, WIAS Preprint No. 1799, pages 1–18, 2013.

[[van Ackooij and Minoux(2013)]]W. van Ackooij and M. Minoux.
On a completion of the Henrion-Möller-Prékopa gradient formula for gaussian distribution functions.
draft submitted, pages 1–3, 2013.

[[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.