

Probability gradient estimation by set-valued calculus and applications in network design

Georg Ch. Pflug¹ and Heinz Weisshaupt²

Abstract. Let $\vartheta \mapsto P(\vartheta)$ be a set-valued mapping from \mathbb{R}^d into the family of closed compact polyhedra in \mathbb{R}^s . Let ξ be a \mathbb{R}^s valued random variable. Many stochastic optimization problems in computer networking, system reliability, transportation, telecommunication, finance etc. can be formulated as a problem to minimize (or maximize) the probability $\mathbb{P}\{\xi \in P(\vartheta)\}$ under some constraints on the decision variable ϑ . For a practical solution of such a problem, one has to approximate the objective function and its derivative by Monte Carlo simulation, since a closed analytical expression is only available in rare cases. In this paper, we present a new method of approximating the gradient of $\mathbb{P}\{\xi \in P(\vartheta)\}$ w.r.t ϑ by sampling, which is based on the concept of set-wise derivative. Quite surprisingly, it turns out that it is typically easier to approximate the derivative than the objective itself.

1 Introduction

In many applications in computer networking, systems reliability, transportation, telecommunication, insurance and financial optimization, the objective is to minimize the probability of a failure (or network breakdown, ruin, financial distress) under some appropriate constraints. These problems are stochastic optimization problems with a specific type of objective: the probability of an event depending on the decision parameter. They are closely related to chance constrained stochastic optimization problems, in which probabilities of unwanted events appear as constraints; see for instance [7], [4].

In this article, we concentrate on the essential part of a stochastic optimization problem with probability objective: To find the gradients of the objective w.r.t. the decision parameter. Once an accurate estimate of the gradient is found, any numerical algorithm for optimization, e.g. gradient projection may be applied.

The setup of our problem is as follows: Assume that $\vartheta \mapsto P(\vartheta)$ is a set-valued mapping from \mathbb{R}^d into the family of closed compact polyhedra in \mathbb{R}^s . Let ξ be a \mathbb{R}^s valued random variable with distribution μ and continuous density g . Let $p(\vartheta) = \mu(P(\vartheta)) = \mathbb{P}\{\xi \in P(\vartheta)\}$. If $P(\vartheta)$ is more complicated than a box or a disjoint union of boxes, then the only practical

¹Department of Statistics and Decision Support Systems, University of Vienna, Austria, A-1010 Wien, Universitätsstrasse 5

²Department of Statistics, University of Düsseldorf

method to calculate $p(\vartheta)$ is by Monte Carlo simulation. Tricky ways to include $P(\vartheta)$ in a union of boxes and/or to use importance sampling are well known. Even more complicated is to calculate the gradient $\nabla p(\vartheta)$ of $p(\vartheta)$ w.r.t ϑ .

Gradient estimation procedures based on sampling are known under the names of perturbation analysis [2], the score function method [10] and the method of weak derivatives [6]. Perturbation analysis does not work here. The push-out score function method depends on finding an appropriate reparameterization, which transforms the parameters of the polyhedron into parameters of the probability distribution, which works only in specific cases. We will present here a weak derivative method. Notice that differentiation formulas as integrals over the surface measure appear already in Raik [9], compare also the work of Uryasev [12].

This paper is organized as follows: Section 2 is devoted to a motivating example, in section 3 we present the algorithms. A numerical example is contained in section 4. Sections 5 and 6 are devoted to the elaboration of the set-valued weak derivative method in full mathematical strength. In section 7, we generalize the motivating problem and show how also the generalization may be treated. Some auxiliary results are collected in the Appendix.

Throughout the paper, we use the following notation: If C is a convex body in \mathbb{R}^s , then ∂C denotes its surface (the topological boundary) and $o_{\partial C}$ the surface area measure (see Proposition 2 in the Appendix).

2 A motivating example

We consider the well known problem of designing a communication network with random demand in an optimal way (see e.g. [8], [5]). Assume that the topology of the network is given, the decision is to be made about the optimal capacities of the links of the network under a budget constraint. The objective is to maximize the reliability, i.e. minimize the breakdown probability. A breakdown occurs, if the demand cannot be satisfied with the available capacities. The example was motivated by a project with the Austrian Telekom company.

The nodes of the network are denoted by $\mathcal{N} = \{1, \dots, N\}$. The potential links are pairs of nodes $\{(k, \ell), 1 \leq k \leq N, 1 \leq \ell \leq N\}$. Every link has a nonnegative capacity $\vartheta_{k,\ell}$. A zero capacity $\vartheta_{k,\ell} = 0$ means that the link (k, ℓ) does not exist.

The demand is given by a collection of random variables $(\xi_{m,n})$ indicating the traffic originating in node m and ending in node n , for every origin-destination (OD) relation $m \rightarrow n$. This demand is mapped onto the links through a routing system R .

We assume first that the routing is fixed. This assumption is relaxed in section 7. The proportion of traffic from m to n flowing through link (k, ℓ) is denoted by $r_{m,n,k,\ell}$. A routing must satisfy

$$\begin{aligned} \sum_{\ell} r_{m,n,m,\ell} &= 1 && \text{for all } m, n \\ \sum_{\ell} r_{m,n,\ell,k} &= \sum_{\ell} r_{m,n,k,\ell} && \text{for } k \neq m, n \end{aligned}$$

$$\sum_{\ell} r_{m,n,\ell,n} = 1 \quad \text{for all } m, n$$

The total flow on link (k, ℓ) is $\sum_{m,n} \xi_{m,n} r_{m,n,k,\ell}$. The capacity constraint reads that

$$\sum_{m,n} \xi_{m,n} r_{m,n,k,\ell} \leq \vartheta_{k,\ell}. \quad (1)$$

The network is reliable, if all capacity constraints are fulfilled.

In a compact matrix notation (after a new indexing of links and OD relationships by single indices) the constraint (1) reads

$$R\xi \leq \vartheta.$$

Denote by $P(\vartheta)$ the "reliable" polyhedron

$$P(\vartheta) = \{y \geq 0 : Ry \leq \vartheta\}.$$

The optimal network design problem is

$$\begin{aligned} & \text{Maximize } \mathbb{P}\{\xi \in P(\vartheta)\} \\ & h^T \vartheta \leq B \end{aligned} \quad (2)$$

Here h is a cost vector and B is the available budget (The constraints may be more complicated and nonlinear as well). The main challenge in solving (2) algorithmically is to calculate $p(\vartheta) = \mathbb{P}\{\xi \in P(\vartheta)\}$ and, even more importantly the gradient $\nabla \mathbb{P}\{\xi \in P(\vartheta)\} = (\partial p(\vartheta)/\partial \vartheta_1, \dots, \partial p(\vartheta)/\partial \vartheta_d)$.

In the next section, we present algorithms to estimate this gradient.

3 The main theorem and the algorithms

Let $\vartheta \in \mathbb{R}^d$ and let $(\mathbf{a}_i(\vartheta))_{i \in I}$ be row-vectors in \mathbb{R}^s , depending on ϑ . Let $(b_i(\vartheta))_{i \in I}$ be real numbers (the right hand sides). Let $P(\vartheta)$ be the polyhedron

$$P(\vartheta) = \{y \in \mathbb{R}^s : y^T \mathbf{a}_i(\vartheta) \leq b_i(\vartheta); \quad i \in I\}. \quad (3)$$

We suppose that $\mathbf{a}_i(\vartheta)$ and $b_i(\vartheta)$ are differentiable functions in ϑ and that the convex polyhedron $P(\vartheta)$ is compact and has nonempty interior for $\vartheta \in \Theta$, an open subset of \mathbb{R}^d .

Let μ be a probability measure on \mathbb{R}^s with continuous density g with respect to Lebesgue measure λ such that

$$\mathbb{P}\{\xi \in P(\vartheta)\} = \mu(P(\vartheta)) = \int_{P(\vartheta)} g(u) du.$$

Denote by $D_\tau \mu(P(\vartheta))$ the directional derivative from the right of $\mu(P(\vartheta))$ w.r.t. the direction τ , i.e.

$$D_\tau \mu(P(\vartheta)) = \lim_{h \downarrow 0} \frac{1}{h} [\mu(P(\vartheta + h\tau)) - \mu(P(\vartheta))].$$

Introduce in a similar manner the directional derivatives $D_\tau \mathbf{a}_i$ and $D_\tau b_i$.

It may happen that for the fixed ϑ under consideration, two hyperplanes $\{x : \mathbf{a}_{i_1}(\vartheta)x = b_{i_1}(\vartheta)\}$ and $\{x : \mathbf{a}_{i_2}(\vartheta)x = b_{i_2}(\vartheta)\}$ coincide but are not identical for all ϑ . For this reason, we introduce the equivalence classes

$$[i] := \{j \in I \mid \mathbf{a}_j(\vartheta) = \gamma_j \mathbf{a}_i(\vartheta), b_j(\vartheta) = \gamma_j b_i(\vartheta) \text{ for some } \gamma_j > 0\}, \quad J := \{[i] \mid i \in I\},$$

$$H_i := \{x \mid \mathbf{a}_i(\vartheta)x = b_i(\vartheta)\} \quad \text{and} \quad H_{[i]} := H_i$$

Our main theorem represents the desired directional derivative $D_\tau(\mu(P(\vartheta)))$ by an integral w.r.t. the surface area measure $o_{\partial P(\vartheta)}$ on $P(\vartheta)$.

Main Theorem.

Under the given assumptions,

$$\begin{aligned} & D_\tau \mu(P(\vartheta)) \\ &= \int_{\partial P(\vartheta)} g(x) \cdot \sum_{[j] \in J} \mathbf{1}_{P(\vartheta) \cap H_{[j]}} \cdot \min_{i \in [j]} \left\{ \frac{1}{\|\mathbf{a}_i\|} [D_\tau b_i(\vartheta) - x^T D_\tau \mathbf{a}_i(\vartheta)] \right\} d o_{\partial P(\vartheta)}(x). \end{aligned} \tag{4}$$

The proof of this theorem will be given in section 5.

Based on this theorem, the following algorithm implements a sampling method for getting an unbiased derivative estimate.

Algorithm 1

- Construct a random variable ζ with probability distribution proportional to $g(\cdot) o_{\partial P(\vartheta)}$.
- Sample x from ζ
- Calculate $f_i = \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - x^T D_\tau \mathbf{a}_i(\vartheta)]$ for all i with $x \in H_{[i]}$ and find $f = \min f_i$.
- The derivative estimate is $f \cdot \int_{\partial P(\vartheta)} g(u) d o_{\partial P(\vartheta)}(u)$

By virtue of our main Theorem, this algorithm produces an unbiased estimate of the derivative. The algorithm needs the construction of the random variable ζ . The construction of ζ is not easy in general. The problem is that we have to sample from a distribution on the boundary of $P(\vartheta)$. We replace now this sampling on the boundary by sampling (and if necessary discarding) points on the hyperplanes $H_{[i]}$. Algorithm 2 works only if the integrals $\int_{H_{[i]}} g(x) d o_{H_{[i]}}(x)$ are finite.

Algorithm 2

- Construct for all $[j] \in J$ random variables $\zeta_{[j]}$ with probability distributions proportional to $g(\cdot) o_{\partial H_{[j]}}$.
- Sample $x_{[j]}$ from $\zeta_{[j]}$.
- if $x_{[j]} \in P$ then
 - Calculate $f_i = \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - x_{[j]}^T D_\tau \mathbf{a}_i(\vartheta)]$ for all $i \in I$ and find for all $[j] \in J$ the value $f_{[j]} = \min_{i \in [j]} f_i$.
- else set $f_{[j]} = 0$
- The derivative estimate is

$$\sum_{[j] \in J} f_{[j]} \cdot \int_{H_{[j]}} g(u) do_{H_{[j]}}(u)$$

This algorithm needs the construction of random variables $\zeta_{[j]}$, i.e. sampling from a conditional distribution on hyperplanes, rather than from the distribution ζ itself. We will illustrate the method for the important case of ζ being a multivariate normal distribution, because then also the conditional distributions are multivariate normal and easy to sample.

Assuming that μ is multivariate normal, $\mu = N(m, \Sigma)$, recall that it may be transformed by a linear mapping to a standard normal distribution. For a positive definite Σ , let $\sqrt{\Sigma}$ be any matrix satisfying $\sqrt{\Sigma}^T \sqrt{\Sigma} = \Sigma$.

The algorithm specialized for multivariate normal reads

Algorithm 3

- Transform the problem so that the original distribution $N(m, \Sigma)$ becomes the standard normal distribution $N(0, I)$ on \mathbb{R}^s :
 - Let $A^{new} := A \cdot \sqrt{\Sigma}$ and let $b^{new} := b - A \cdot m$.
 - Change the notation to $A := A^{new}$ and $b := b^{new}$.
- Let $z_{[j]} \in H_{[j]}$ be the point in $H_{[j]}$ with minimal norm, i.e. $z_{[j]} = \frac{b_i \mathbf{a}_i}{\mathbf{a}_i^T \mathbf{a}_i}, i \in [j]$.
- Sample x from a standard normal distribution on \mathbb{R}^s .
- Calculate for all $[j] \in J$ the orthogonal projection $x_{[j]} = x - \frac{\mathbf{a}_i^T x - b_i}{\mathbf{a}_i^T \mathbf{a}_i} \mathbf{a}_i, i \in [j]$, of x onto $H_{[j]}$.
- if $x_{[j]} \in P(\vartheta)$ then
 - Calculate $f_i = \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - x_{[j]}^T D_\tau \mathbf{a}_i(\vartheta)]$ for all $i \in I$ and find for all $[j] \in J$ the value $f_{[j]} = \min_{i \in [j]} f_i$.
- else set $f_{[j]} = 0$
- The derivative estimate is

$$\sum_{[j] \in J} f_{[j]} \cdot \int_{H_{[j]}} \phi(u) do_{H_{[j]}}(u) = \sum_{[j] \in J} f_{[j]} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\|z_{[j]}\|^2}{2}}.$$

Here ϕ is the density of the standard normal distribution in \mathbb{R}^s .

4 A numerical example

We return to the network reliability problem discussed in section 2. Consider the network shown in Figure 1.

The set of nodes is

$$\mathcal{N} = \{1, 2, 3, 4, 5, 6\}.$$

The set of links is

$$\begin{aligned} \mathcal{L} = \{ & l_1 = (1, 2); l_2 = (1, 6); l_3 = (2, 3); l_4 = (2, 4); l_5 = (2, 6); l_6 = (3, 4); \\ & l_7 = (3, 5); l_8 = (4, 5); l_9 = (5, 6) \} \end{aligned}$$

The set of origin-destinations relations is

$$\begin{aligned} OD = \{ & (1, 2); (1, 3); (1, 4); (1, 5); (1, 6); (2, 3); (2, 4); (2, 5); (2, 6); (3, 4); (3, 5); \\ & (3, 6); (4, 5); (4, 6); (5, 6) \} \end{aligned}$$

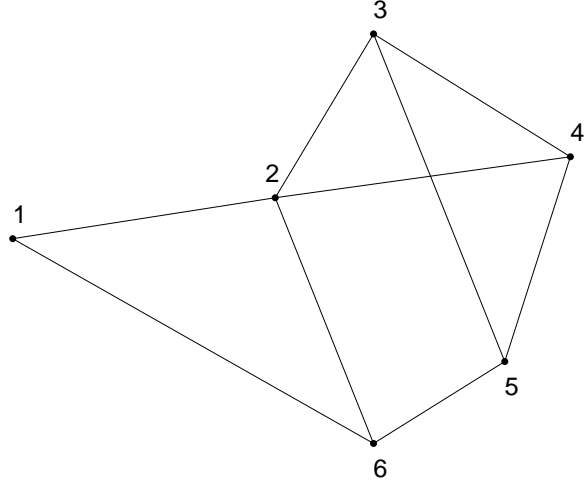


Figure 1: A network example

For simplicity, the traffic is considered undirected in this example. With the very same method, one could also treat the links as directed and thus double their number as well as the number of OD-relations.

The routing matrix is the fixed $[9 \times 15]$ matrix R given below. The ordering of the rows is according to the ordering of links in \mathcal{L} and the ordering of the columns is according to the ordering of the OD-relations above.

$$R = \begin{pmatrix} 1.0 & 1.0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 & 0 & 1.0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 1.0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 1.0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 1.0 & 1.0 \end{pmatrix}$$

We assume that the demand vector ξ follows a 15-dimensional normal distribution with mean value

$$\mathbb{E}(\xi) = (100; 200; 1500; 300; 80; 200; 1500; 300; 80; 3000; 600; 160; 4500; 1200; 240)$$

and covariance matrix

$$\Sigma = (\Sigma_1 \Sigma_2)$$

where

$$\Sigma_1 = \begin{pmatrix} 2.5 & 1.5 & 11.25 & 2.25 & 0.6 & 1.5 & 11.25 \\ 1.5 & 10.0 & 22.5 & 4.5 & 1.2 & 3.0 & 7.5 \\ 11.25 & 22.5 & 562.5 & 33.75 & 9.0 & 7.5 & 168.75 \\ 2.25 & 4.5 & 33.75 & 22.5 & 1.8 & 1.5 & 11.25 \\ 0.6 & 1.2 & 9.0 & 1.8 & 1.6 & 0.4 & 3.0 \\ 1.5 & 3.0 & 7.5 & 1.5 & 0.4 & 10.0 & 22.5 \\ 11.25 & 7.5 & 168.75 & 11.25 & 3.0 & 22.5 & 562.5 \\ 2.25 & 1.5 & 11.25 & 6.75 & 0.6 & 4.5 & 33.75 \\ 0.6 & 0.4 & 3.0 & 0.6 & 0.48 & 1.2 & 9.0 \\ 7.5 & 45.0 & 337.5 & 22.5 & 6.0 & 45.0 & 337.5 \\ 1.5 & 9.0 & 22.5 & 13.5 & 1.2 & 9.0 & 22.5 \\ 0.4 & 2.4 & 6.0 & 1.2 & 0.96 & 2.4 & 6.0 \\ 11.25 & 22.5 & 506.25 & 101.25 & 9.0 & 22.5 & 506.25 \\ 3.0 & 6.0 & 135.0 & 9.0 & 7.2 & 6.0 & 135.0 \\ 0.6 & 1.2 & 9.0 & 5.4 & 1.44 & 1.2 & 9.0 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 2.25 & 0.6 & 7.5 & 1.5 & 0.4 & 11.25 & 3.0 & 0.6 \\ 1.5 & 0.4 & 45.0 & 9.0 & 2.4 & 22.5 & 6.0 & 1.2 \\ 11.25 & 3.0 & 337.5 & 22.5 & 6.0 & 506.25 & 135.0 & 9.0 \\ 6.75 & 0.6 & 22.5 & 13.5 & 1.2 & 101.25 & 9.0 & 5.4 \\ 0.6 & 0.48 & 6.0 & 1.2 & 0.96 & 9.0 & 7.2 & 1.44 \\ 4.5 & 1.2 & 45.0 & 9.0 & 2.4 & 22.5 & 6.0 & 1.2 \\ 33.75 & 9.0 & 337.5 & 22.5 & 6.0 & 506.25 & 135.0 & 9.0 \\ 22.5 & 1.8 & 22.5 & 13.5 & 1.2 & 101.25 & 9.0 & 5.4 \\ 1.8 & 1.6 & 6.0 & 1.2 & 0.96 & 9.0 & 7.2 & 1.44 \\ 22.5 & 6.0 & 2250.0 & 135.0 & 36.0 & 1012.5 & 270.0 & 18.0 \\ 13.5 & 1.2 & 135.0 & 90.0 & 7.2 & 202.5 & 18.0 & 10.8 \\ 1.2 & 0.96 & 36.0 & 7.2 & 6.4 & 18.0 & 14.4 & 2.88 \\ 101.25 & 9.0 & 1012.5 & 202.5 & 18.0 & 5062.5 & 405.0 & 81.0 \\ 9.0 & 7.2 & 270.0 & 18.0 & 14.4 & 405.0 & 360.0 & 21.6 \\ 5.4 & 1.44 & 18.0 & 10.8 & 2.88 & 81.0 & 21.6 & 14.4 \end{pmatrix}$$

In this special case, the matrix A equals the routing matrix R not depending on ϑ . The right hand side is equal to ϑ , i.e., $b_i(\vartheta) = \vartheta$. Further the directions of the vectors \mathbf{a}_i are all different and so all hyperplanes of the form $\{x \mid \mathbf{a}_i^T \mathbf{x} = \vartheta\}$ are different. Thus the equivalence classes $[i]$ contain just the element i and the conditions concerning the formation of minimum in the different algorithms become vacuous. The derivative estimate in the direction e_i of the i -th coordinate-axis thus becomes in the special case of our example:

$$\begin{cases} \frac{1}{\|(R\sqrt{\Sigma})_{(i,:)}\|} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\|z_i\|_2^2}{2}} & \text{if } x \in H_i \cap P_\vartheta \\ 0 & \text{otherwise} \end{cases}$$

Here $(R\sqrt{\Sigma})_{(i,:)}$ is the i -th row of the transformed routing matrix $R\sqrt{\Sigma}$.

	gradient estimate	
	5000 iterations	50000 iterations
$\partial/\partial\vartheta_1$	$0.0281 * 10^{-3}$	$0.0282 * 10^{-3}$
$\partial/\partial\vartheta_2$	$0.0538 * 10^{-3}$	$0.0538 * 10^{-3}$
$\partial/\partial\vartheta_3$	$0.1394 * 10^{-3}$	$0.1388 * 10^{-3}$
$\partial/\partial\vartheta_4$	$0.0013 * 10^{-3}$	$0.0013 * 10^{-3}$
$\partial/\partial\vartheta_5$	$0.0000 * 10^{-3}$	$0.0000 * 10^{-3}$
$\partial/\partial\vartheta_6$	$0.0556 * 10^{-3}$	$0.0557 * 10^{-3}$
$\partial/\partial\vartheta_7$	$0.0784 * 10^{-3}$	$0.0786 * 10^{-3}$
$\partial/\partial\vartheta_8$	$0.0040 * 10^{-3}$	$0.0040 * 10^{-3}$
$\partial/\partial\vartheta_9$	$0.0009 * 10^{-3}$	$0.0009 * 10^{-3}$

Table 1: The values of the estimated gradient vector

Algorithm 3 was applied to this example to estimate $\nabla\mathbb{P}\{\xi \in P(\vartheta)\}$ at the point $\vartheta = (1890; 400; 500; 3310; 330; 3150; 715; 6150; 2070)$. The column vectors of estimates of the partial derivatives for a sample size of 5000 and 50000 are shown in Table 1.

We compare this result with the result obtained by estimating the gradient by an algorithm (Algorithm 4 below), which calculates an estimate of the directional derivative by calculating $\frac{\mathbb{P}\{\xi \in P(\vartheta + \Delta e_i) \setminus P(\vartheta)\}}{\Delta}$.

This algorithm is based on Monte Carlo simulation. Instead of sampling from the distribution of ξ it uses importance sampling from different distributions for each unit direction e_i . These distributions are chosen to improve the performance of this simple algorithm and to allow a fair comparison of our Algorithm 3 based on weak derivatives with this simple method.

	standard deviation	
	Algorithm 3	Algorithm 4
$\partial/\partial\vartheta_1$	$0.0627 * 10^{-4}$	$0.3158 * 10^{-3}$
$\partial/\partial\vartheta_2$	$0.0940 * 10^{-4}$	$0.2319 * 10^{-3}$
$\partial/\partial\vartheta_3$	$0.3196 * 10^{-4}$	$0.6393 * 10^{-3}$
$\partial/\partial\vartheta_4$	$0.0106 * 10^{-4}$	$0.0236 * 10^{-3}$
$\partial/\partial\vartheta_5$	$0.0000 * 10^{-4}$	$0.0000 * 10^{-3}$
$\partial/\partial\vartheta_6$	$0.0732 * 10^{-4}$	$0.8478 * 10^{-3}$
$\partial/\partial\vartheta_7$	$0.1520 * 10^{-4}$	$0.5134 * 10^{-3}$
$\partial/\partial\vartheta_8$	$0.0086 * 10^{-4}$	$0.0850 * 10^{-3}$
$\partial/\partial\vartheta_9$	$0.0038 * 10^{-4}$	$0.0104 * 10^{-3}$

Table 2: A comparison of the componentwise standard deviations

Algorithm 4

<ul style="list-style-type: none"> • Transform the problem so that the original distribution $N(m, \Sigma)$ becomes the standard normal distribution $N(0, I)$ on \mathbb{R}^s: <ul style="list-style-type: none"> – Let $A^{new} := A \cdot \sqrt{\Sigma}$ and let $b^{new} := b - A \cdot m$. – Change the notation to $A := A^{new}$ and $b := b^{new}$. • Let $z_{[j]} \in H_{[j]}$ be the point in $H_{[j]}$ with minimal norm, i.e. $z_{[j]} = \frac{b_i^T a_i}{a_i^T a_i}, i \in [j]$. • Sample x_i from a normal distribution on \mathbb{R}^s with mean-vector z_i and covariance matrix the identity. • if $x_i \in P(\vartheta + \Delta) \setminus P(\vartheta)$ then <ul style="list-style-type: none"> – let $f_i = 1$ • else set $f_i = 0$. • The estimate of the i-th directional derivative is $\frac{f_i}{\Delta} \cdot e^{\frac{-x_i^T \cdot x_i + (x_i - z_i)^T \cdot (x_i - z_i)}{2}}$
--

The vectors of estimated standard deviations of Algorithm 3 and Algorithm 4 (for $\Delta = 0.5$) are displayed in Table 2.

Considering these standard deviations we see that about only 0.25 percent of the sample size is needed in our algorithm to obtain an equally stable estimate compared to the direct Algorithm 4. Moreover, the estimate obtained by the direct algorithm is biased, while the

	gradient estimate	
	Algorithm 3 with $5 * 10^4$ iterations	Algorithm 4 with $5 * 10^5$ iterations and $\Delta = 0.5$
$\partial/\partial\vartheta_1$	$0.0281 * 10^{-3}$	$0.0270 * 10^{-3}$
$\partial/\partial\vartheta_2$	$0.0538 * 10^{-3}$	$0.0447 * 10^{-3}$
$\partial/\partial\vartheta_3$	$0.1394 * 10^{-3}$	$0.1191 * 10^{-3}$
$\partial/\partial\vartheta_4$	$0.0013 * 10^{-3}$	$0.0014 * 10^{-3}$
$\partial/\partial\vartheta_5$	$0.0000 * 10^{-3}$	$0.0000 * 10^{-3}$
$\partial/\partial\vartheta_6$	$0.0556 * 10^{-3}$	$0.0537 * 10^{-3}$
$\partial/\partial\vartheta_7$	$0.0784 * 10^{-3}$	$0.0712 * 10^{-3}$
$\partial/\partial\vartheta_8$	$0.0040 * 10^{-3}$	$0.0040 * 10^{-3}$
$\partial/\partial\vartheta_9$	$0.0009 * 10^{-3}$	$0.0009 * 10^{-3}$

Table 3: A comparison of the estimated gradient vectors

estimate based on the new algorithm is unbiased.

5 Weak derivatives of set-valued mappings

In this section, we discuss the notion of derivatives of set-valued mappings and give some properties. The concept will be introduced for general convex bodies and not only for convex polyhedra.

Let $\Theta \subseteq \mathbb{R}^d$ be open and let $\vartheta \mapsto C(\vartheta)$ be a set-valued mapping defined on Θ such that the values $C(\vartheta)$ are convex bodies (convex compact sets with nonempty interior) in \mathbb{R}^s . We call such mappings for short convex-valued mappings. Let μ be a probability measure on \mathbb{R}^s which possesses a continuous density g with respect to Lebesgue measure λ .

We denote by ∂C the boundary of the (convex) set C and by $o_{\partial C}$ the surface area measure on ∂C (see Appendix).

Given a measure μ we denote by $\mu|_A$ its restriction to S , i.e. $\mu|_A(B) := \mu(A \cap B)$.

We define now the important concept of the weak derivative of a set valued mapping $\vartheta \mapsto C(\vartheta)$ in direction $\tau \in \mathbb{R}^d$ with respect to a measure μ on \mathbb{R}^s .

Definition 1. (See also [6] and [13]) We say that a convex-valued mapping $\vartheta \mapsto C(\vartheta)$ is weakly differentiable from the right at ϑ in direction τ with respect to a measure μ if there exists a (possibly signed) measure $D_\tau^\mu C(\vartheta)$ such that for all continuous functions with compact support $\psi : \mathbb{R}^s \mapsto \mathbb{R}$ we have

$$\int \psi d(D_\tau^\mu C(\vartheta)) = \lim_{h \downarrow 0} \frac{1}{h} \int \psi [1_{C(\vartheta+\tau \cdot h)} - 1_{C(\vartheta)}] d\mu.$$

The (signed) measure $D_\tau^\mu C(\vartheta)$ is called the weak derivative.

From the definition of the weak derivative we get in the special case that $\psi = 1$ on a neighborhood of $C(\vartheta)$ that

$$D_\tau(\mu(C(\vartheta))) = \int_{\partial C(\vartheta)} d(D_\tau^\mu C(\vartheta)). \quad (5)$$

Suppose that $D_\tau^\lambda C(\vartheta) \ll o_{\partial C(\vartheta)}$. This implies for continuous $g := \frac{d\mu}{d\lambda}$ that $D_\tau^\mu C(\vartheta) \ll o_{\partial C(\vartheta)}$ and that

$$\frac{D_\tau^\mu C(\vartheta)}{do_{\partial C(\vartheta)}}(x) = g(x) \cdot \frac{dD_\tau^\lambda C(\vartheta)}{do_{\partial C(\vartheta)}}(x) \quad a.s. \quad do_{\partial C(\vartheta)}. \quad (6)$$

Sufficient condition (SC). If the derivative $D_\tau^\lambda C(\vartheta)$ exists, then a sufficient condition for $D_\tau^\lambda C(\vartheta) \ll o_{\partial C(\vartheta)}$ is (as can be easily derived from Definition 2 and Proposition 2 of the Appendix), that $\vartheta \mapsto C(\vartheta)$ fulfills a Lipschitz condition with respect to the Hausdorff distance *Haus*; i.e.

$$Haus(C(\vartheta), C(\vartheta + s)) \leq c \cdot \|s\|$$

for some constant c and sufficiently small $\|s\|$.

In the following, we will always assume that the sufficient condition SC is fulfilled. We will treat the case of Lebesgue measure λ in \mathbb{R}^s , since all other cases can be brought to this case by simple density correction. For short, we call a function $\vartheta \mapsto C(\vartheta)$ which fulfills Definition 1 for the Lebesgue measure a WDDCV (weakly directionally differentiable convex valued) function.

There is a calculus for WDDCV functions. The following Lemma shows two properties.

Lemma 1.

- (i) Let $\vartheta \mapsto C_1(\vartheta)$ be WDDCV function and let C_2 be a compact convex set such that $o_{\partial C_1(\vartheta)}(\partial C_2) = 0$. Then $\vartheta \mapsto C_1(\vartheta) \cap C_2$ is a WDDCV function with derivative $D_\tau^\lambda C_1(\vartheta) \upharpoonright_{C_2}$.
- (ii) Let $\vartheta \mapsto C_j(\vartheta)$ for $j \in J$, a finite index set, be WDDCV functions. Suppose that for all $j_1 \neq j_2$; $j_1, j_2 \in J$

$$\lim_{h \downarrow 0} \frac{1}{h} \lambda([C_{j_1}(\vartheta + h \cdot \tau) \Delta C_{j_1}(\vartheta)] \cap [C_{j_2}(\vartheta + h \cdot \tau) \Delta C_{j_2}(\vartheta)]) = 0 \quad (7)$$

where Δ denotes the symmetric difference. Then $\vartheta \mapsto \bigcap_{j \in J} C_j(\vartheta)$ is a WDDCV function with derivative $\sum_{j \in J} [D_\tau^\lambda C_j(\vartheta)] \upharpoonright_{\bigcap_{i \in J} C_i}$.

Proof. (i). For all continuous functions ψ with compact support

$$\int \psi d(D_\tau^\lambda C_1(\vartheta)) = \lim_{h \downarrow 0} \frac{1}{h} \int \psi [1_{C_1(\vartheta + \tau \cdot h)} - 1_{C_1(\vartheta)}] d\lambda.$$

This implies that the same property holds for all functions $\tilde{\psi}$, for which the set of discontinuities has zero measure w.r.t the limiting measure $D_\tau^\lambda C(\vartheta)$ (see Billingsley [1]). Since

$$\lim_{h \downarrow 0} \frac{1}{h} \int \psi [1_{C_1(\vartheta + \tau \cdot h) \cap C_2} - 1_{C_1(\vartheta) \cap C_2}] d\lambda = \lim_{h \downarrow 0} \frac{1}{h} \int \psi 1_{C_2} [1_{C_1(\vartheta + \tau \cdot h)} - 1_{C_1(\vartheta)}] d\lambda,$$

setting $\tilde{\psi} = \psi \cdot 1_{C_2}$ and noticing that the assumption implies that the set of discontinuities of $\tilde{\psi}$ have limiting measure zero, because they are contained in ∂C_2 , this implies the assertion.

(ii). We prove first the result for $J = \{1, 2\}$. Using (i), the assertion (ii) is shown if we establish that for all continuous ψ with compact support, bounded by 1,

$$\lim_{h \downarrow 0} \frac{1}{h} \int \psi(x) I(x) d\lambda(x) = 0 \quad (8)$$

where

$$\begin{aligned} I(x) = & [1_{C_1(\vartheta+h\cdot\tau)\cap C_2(\vartheta+h\cdot\tau)}(x) - 1_{C_1(\vartheta)\cap C_2(\vartheta)}(x)] - [1_{C_1(\vartheta+h\cdot\tau)\cap C_2(\vartheta)}(x) - 1_{C_1(\vartheta)\cap C_2(\vartheta)}(x)] \\ & - [1_{C_1(\vartheta)\cap C_2(\vartheta+h\cdot\tau)}(x) - 1_{C_1(\vartheta)\cap C_2(\vartheta)}(x)] \end{aligned}$$

i.e. the difference is equivalent to the sum of the cases when C_1 is treated as constant and C_2 is treated as constant – a form of a product rule of differentiation. The integral in (8) is bounded by

$$\frac{1}{h} \lambda([C_1(\vartheta + h \cdot \tau) \Delta C_1(\vartheta)] \cap [C_1(\vartheta + h \cdot \tau) \Delta C_1(\vartheta)]),$$

which by assumption (7) converges to zero.

For the general case one proceeds by induction. Suppose that the proof is established for $J = \{1, 2, \dots, k-1\}$. The condition needed to extend it for another index k is that

$$\lim_{h \downarrow 0} \frac{1}{h} \lambda \left(\left[\bigcap_{j \in J} C_j(\vartheta + h \cdot \tau) \Delta \bigcap_{j \in J} C_j(\vartheta) \right] \cap [C_k(\vartheta + h \cdot \tau) \Delta C_k(\vartheta)] \right) = 0.$$

This condition follows from the fact that

$$\begin{aligned} & \left[\bigcap_{j \in J} C_j(\vartheta + h \cdot \tau) \Delta \bigcap_{j \in J} C_j(\vartheta) \right] \cap [C_k(\vartheta + h \cdot \tau) \Delta C_k(\vartheta)] \\ \subseteq & \bigcup_{j \in J} [C_j(\vartheta + h \cdot \tau) \Delta C_j(\vartheta)] \cap [C_k(\vartheta + h \cdot \tau) \Delta C_k(\vartheta)] \end{aligned}$$

and the application of condition (7) for each of the sets

$$[C_j(\vartheta + h \cdot \tau) \Delta C_j(\vartheta)] \cap [C_k(\vartheta + h \cdot \tau) \Delta C_k(\vartheta)].$$

□

6 The proof of the main theorem

Let $C(\vartheta)$ be a convex body as in section 4. Later, we will specialize $C(\vartheta) = P(\vartheta)$, where P is the polyhedron defined in (3). From (5) and (6) we obtain

$$D_\tau \mu(C(\vartheta)) = \int_{\partial C(\vartheta)} g(x) \frac{dD_\tau^\lambda(C(\vartheta))}{do_{\partial C(\vartheta)}}(x) do_{\partial C(\vartheta)}(x) \quad (9)$$

Formula (9) entails the main formula (4), if we show that for the convex polyhedron P

$$\frac{dD_\tau^\mu(P(\vartheta))}{do_{\partial C(\vartheta)}}(x) = \sum_{[j] \in J} \mathbf{1}_{P(\vartheta) \cap H_{[j]}} \cdot \min_{i \in [j]} \left\{ \frac{1}{\|\mathbf{a}_i\|} [D_\tau b_i(\vartheta) - x^T D_\tau \mathbf{a}_i(\vartheta)] \right\}.$$

We will first prove the Main Theorem in the special case that $P(\vartheta)$ is a half space, i.e. that for all indices $i \in I$, the half-spaces coincide for some fixed ϑ , but are possibly different for all other values of ϑ . Having done this case, an application of Lemma 1 will provide the final proof.

In the following, we will alternatively use the notations $x^T y$ or $\langle x, y \rangle$ for the inner product in \mathbb{R}^s , whatever notation seems more appropriate.

Lemma 2. Let I be a finite index set and let for $i \in I$

$$\mathbf{a}_i(\tilde{\vartheta}) \in \mathbb{R}^s \text{ and } b_i(\tilde{\vartheta}) \in \mathbb{R}^+$$

with $\tilde{\vartheta} \mapsto \mathbf{a}_i(\tilde{\vartheta})$, $\tilde{\vartheta} \mapsto b_i(\tilde{\vartheta})$ differentiable. Let

$$P(\tilde{\vartheta}) = \bigcap_{i \in I} \{y : \langle \mathbf{a}_i(\tilde{\vartheta}), y \rangle \leq b_i(\tilde{\vartheta})\}$$

and suppose that for some fixed $\vartheta \in \Theta$ the half-spaces $\{y : \langle \mathbf{a}_i(\vartheta), y \rangle \leq b_i(\vartheta)\}$ are all identical. Then $P(\vartheta)$ is itself a half-space \mathcal{H} , the function $\tilde{\vartheta} \mapsto P(\tilde{\vartheta}) \cap R\mathbb{B}$ is weakly directional differentiable from the right at ϑ and its directional derivative in direction τ is given by $f_\tau \cdot o$ with o the surface area measure on $\partial \mathcal{H} \cap R\mathbb{B}$ and $f_\tau : \partial \mathcal{H} \mapsto \mathbb{R}$ given by

$$f_\tau(y) = \min_{i \in I} \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - y^T D_\tau \mathbf{a}_i(\vartheta)].$$

Proof. Let $\mathbf{u} = \frac{\mathbf{a}_i(\vartheta)}{\|\mathbf{a}_i(\vartheta)\|}$ denote the outward unit normal vector on \mathcal{H} , which – by assumption – is independent of i .

Introduce the half spaces $\mathcal{H}_i(h)$ as

$$\mathcal{H}_i(h) = \{x : \langle x, \mathbf{a}_i(\vartheta + h \cdot \tau) \rangle \leq b_i(\vartheta + h \cdot \tau)\}.$$

By assumption, $\mathcal{H} = \mathcal{H}_i(0)$ is independent of i and can be written as

$$\mathcal{H}_i(0) = \{x : \langle x, \mathbf{a} \rangle \leq b\} = \{y + \gamma \mathbf{u} : \gamma \leq 0; \langle y, \mathbf{a} \rangle = b\}$$

where $\mathbf{a} = \mathbf{a}_i(\vartheta)$, $b = b_i(\vartheta)$. Alternatively, we may write

$$\mathcal{H}_i(h) = \{y + \gamma \mathbf{u} : \gamma \leq \xi_{i,h}(y); \langle y, \mathbf{a} \rangle = b\}$$

where $\xi_{i,h}(y)$ for $y \in \partial \mathcal{H}$ is chosen such that

$$\langle y + \xi_{i,h}(y) \mathbf{u}, \mathbf{a}_i(\vartheta + h\tau) \rangle = b_i(\vartheta + h\tau).$$

Now

$$\bigcap_{i \in I} \mathcal{H}_i(h) = \{y + \gamma \mathbf{u} : \gamma \leq \min_i \xi_{i,h}(y); \langle y, \mathbf{a} \rangle = \mathbf{b}\}.$$

It is easy to show that in the definition of the weak derivative, the set of continuous functions with compact support may be replaced by any other dense set, for instance the set of Lipschitz continuous functions with compact support.

Let ψ be a Lipschitz continuous function with compact support. We have to show that

$$\lim_{h \downarrow 0} \frac{1}{h} \int \psi(x) [1_{\bigcap_{i \in I} \mathcal{H}_i(h)}(x) - 1_{\mathcal{H}(0)}(x)] d\lambda(x) = \int \psi(y) f_\tau(y) do_H(y) \quad (10)$$

where H is the hyperplane $H = \partial \mathcal{H}$. A change of integration variables gives

$$\begin{aligned} & \frac{1}{h} \int \psi(x) [1_{\bigcap_{i \in I} \mathcal{H}_i(h)}(x) - 1_{\mathcal{H}(0)}(x)] d\lambda(x) \\ &= \frac{1}{h} \int_H \int_{\mathbb{R}} \psi(y + \gamma \mathbf{u}) [1_{\bigcap_{i \in I} \mathcal{H}_i(h)}(y + \gamma \mathbf{u}) - 1_{\mathcal{H}(0)}(y + \gamma \mathbf{u})] d\gamma do_H(y) \\ &= \frac{1}{h} \int_H \int_{\mathbb{R}} [\psi(y) + O(\gamma)] [1_{0 \leq \gamma \leq \min_i \xi_{i,h}(y)} - 1_{\min_i \xi_{i,h}(y) \leq \gamma \leq 0}] d\gamma do_H(y) \\ &= \int_H \int_{\mathbb{R}} \psi(y) \frac{1}{h} \min_i \xi_{i,h}(y) do_H(y) + O\left(\int_H \frac{1}{h} [\min_i \xi_{i,h}(y)]^2 do_H(y)\right) \\ &= \int_H \int_{\mathbb{R}} \psi(y) \frac{1}{h} \min_i \xi_{i,h}(y) do_H(y) + o(1) \end{aligned}$$

as $h \downarrow 0$. Here we have used the Landau symbols $O(\cdot)$ and $o(\cdot)$. An easy calculation shows that

$$\lim_{h \downarrow 0} \frac{\xi_{i,h}(y)}{h} = \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - y^T D_\tau \mathbf{a}_i(\vartheta)]$$

and therefore

$$\lim_{h \downarrow 0} \min_i \frac{\xi_{i,h}(y)}{h} = \min_i \frac{1}{\|\mathbf{a}_i(\vartheta)\|} [D_\tau b_i(\vartheta) - y^T D_\tau \mathbf{a}_i(\vartheta)] = f_\tau(y).$$

Finally, by dominated convergence theorem

$$\lim_{h \downarrow 0} \frac{1}{h} \int \psi(x) [1_{\bigcap_{i \in I} \mathcal{H}_i(h)} - 1_{\mathcal{H}(0)}] d\lambda(x) = \int \psi(y) f_\tau(y) do_H(y).$$

□

We remark here that a similar result holds if $b_i(\vartheta) \in \mathbb{R}^-$.

The proof of the Main Theorem:

Consider a ball $R\mathbb{B}$ sufficiently big such that $P(\vartheta) \subset \text{int}(R\mathbb{B})$. Propositions 4 and 5 of the Appendix ensure that $\vartheta \mapsto P(\vartheta)$ is Lipschitz in the Hausdorff sense and that the sufficient condition is fulfilled. Moreover, for halfspaces the condition (7) is fulfilled. Thus we get by Lemma 2 with

$$C_{[j]} := R\mathbb{B} \cap \bigcap_{i \in [j]} \{x : \langle \mathbf{a}_i(\vartheta), x \rangle \leq b_i(\vartheta)\}$$

and Lemma 1 (ii) applied to the index set $J = \{[j]\}$ that

$$\frac{dD_\tau^\mu(P(\vartheta))}{do_{\partial C(\vartheta)}}(x) = \sum_{[j] \in J} \mathbf{1}_{P(\vartheta) \cap H_{[j]}} \cdot \min_{i \in [j]} \left\{ \frac{1}{\|\mathbf{a}_i\|} [D_\tau b_i(\vartheta) - x D_\tau \mathbf{a}_i(\vartheta)] \right\}.$$

Inserting this in equation (9) we obtain equation (4), i.e., the Main Theorem has been proved.

7 Adaptive routing

In this section we show that adaptive routing is included in the type of problems studied already. In the adaptive routing situation, the flows are calculated through an assignment problem of its own.

Let $z_{m,n,k,\ell}$ be the traffic from origin m to destination n , which flows over the link (k, ℓ) . These flows must satisfy

$$\begin{aligned} \sum_{\ell} z_{m,n,m,\ell} &= \xi_{m,n} && \text{for all } m, n \\ \sum_{\ell} z_{m,n,\ell,k} &= \sum_{\ell} z_{m,n,k,\ell} && \text{for } k \neq m, n \\ \sum_{\ell} z_{m,n,\ell,n} &= \xi_{n,m} && \text{for all } n, m \end{aligned} \tag{11}$$

The capacity constraints are

$$\sum_{m,n} z_{m,n,k,\ell} \leq \vartheta_{k,\ell} \tag{12}$$

We say that ξ is *admissible* if there is a $z \geq 0$ such that all equations (11) and the inequality (12) is fulfilled.

By introducing appropriate indices for links and OD-relations, introducing the appropriate matrices A and M , one gets a system of the following form:

ξ is admissible, if $\exists z \geq 0$ such that

$$Az = \xi, Bz = 0, Mz \leq \vartheta.$$

The admissible polyhedron is

$$P(\vartheta) = \{ \xi : \exists z \geq 0 : Az = \xi, Bz = 0, Mz \leq \vartheta \}.$$

This representation is not in the desired explicit form (3). Let's rewrite it as

$$\begin{aligned} P(\vartheta) &= \{ \xi : \exists z \geq 0 : Az = \xi, Bz = 0, Mz \leq \vartheta \} \\ &= \{ \xi : \exists z, y \geq 0 : Az = \xi, Bz = 0, Mz + y = \vartheta \} \\ &= \left\{ \xi : \exists x \geq 0 : Wx = \begin{pmatrix} \xi \\ 0 \\ \vartheta \end{pmatrix} \right\} \end{aligned}$$

Here we have set $x = \begin{pmatrix} z \\ y \end{pmatrix}$ and $W = \begin{pmatrix} A & 0 \\ B & 0 \\ M & I \end{pmatrix}$.

The set $\{d : \exists x \geq 0 : Wx = d\}$ is identical with the set $\{d : d^T e_i \geq 0\}$, where $\{e_i\}$ are the extremal rays of the polyhedral cone $\{u : W^T u \geq 0\}$. Let $\{(u_i, v_i, w_i), i \in I\}$ be the extremal rays of the polyhedral cone $\{(u, v, w) : A^T u + B^T v + M^T w \geq 0, w \geq 0\}$.

We see that the polyhedron $P(\vartheta)$ has the representation

$$P(\vartheta) = \{\xi : -\xi^T u_i \leq \vartheta^T w_i, i = 1, \dots, I\},$$

which is of the required form (3).

How can one calculate the extremal rays of a polyhedral cone? Let W be a $[n \times m]$ matrix of rank d . This matrix describes a linear mapping from \mathbb{R}^m onto a d -dimensional subspace of \mathbb{R}^n . Let k be the dimension of the kernel of W . We have that: $d = m - k$. The cone $\{u : Wu \geq 0\}$ is the set of all u 's whose image lies in the positive quadrant. We may find a basis (b_1, \dots, b_m) of \mathbb{R}^m such that (b_1, \dots, b_k) is a basis of the kernel and (b_{k+1}, \dots, b_m) is orthogonal to it. It is sufficient to intersect the linear combinations of the images of (b_{k+1}, \dots, b_m) with the facets of the positive orthant. The dimension of the image is $m - k$, thus more $n - m + k$ linear conditions are needed. We solve the homogeneous equation $\sum_{i=k+1}^m \lambda_i b_i = r$, where r is zero for $n - m + k$ indices and free elsewhere. If the solution is nonnegative, an extremal ray has been found. We have to test at most $\binom{n}{n-m+k}$ equations. Notice that the extremal rays are independent of ϑ and can be stored and kept once they were found.

Appendix

We denote by \mathbb{S} the unit sphere and by \mathbb{B} the unit ball in \mathbb{R}^s .

Proposition 1. Let C be a convex body and let for $h \geq 0$

$$C_h^- := C \cap (\partial C + h\mathbb{B}) \quad \text{and} \quad C_h^+ := (\partial C + h\mathbb{B}) \setminus \text{int}C.$$

Then there exists a measure $o_{\partial C}$ (the *surface area measure*) such that in weak sense

$$\lim_{h \downarrow 0} \frac{1}{h} \lambda|_{C_h^-} = \lim_{h \downarrow 0} \frac{1}{h} \lambda|_{C_h^+} = o_{\partial C}.$$

Definition 2. The *Hausdorff* distance $Haus$ of two convex bodies C, D is defined by

$$Haus(C, D) := \inf\{\gamma > 0 : D \subseteq C + \gamma\mathbb{B} \quad \text{and} \quad C \subseteq D + \gamma\mathbb{B}\}$$

Proposition 2.

- (i) If $Haus(C, D) = \epsilon$, then $C \setminus D \subseteq \partial D + \epsilon\mathbb{B}$ and $D \setminus C \subseteq \partial C + \epsilon\mathbb{B}$.
- (ii) $Haus(\partial C, \partial D) \leq Haus(C, D)$.

(iii) If $Haus(C, D) = \epsilon$, then $C\Delta D \subseteq \partial C + 2\epsilon\mathbb{B}$. Here Δ denotes the symmetric difference.

Proof. (i) follows directly from the definition. (ii): Let $\gamma = Haus(C, D)$ and let $x \in \partial C$. We have to find a $y \in \partial D$ such that $\|x - y\| \leq \gamma$. If $x \notin D$, then there is a $y \in \partial D$ such that $\|x - y\| \leq \gamma$ by assumption. If $x \in D$, then find a point $y \in \partial D$, such that the (convex) projection of y onto C is x . For this point, $\gamma \geq \inf\{\|y - v\| : v \in C\} = \|y - x\|$ and the assertion is shown. (iii) follows from (i) and (ii): $D \setminus C \subseteq \partial C + \epsilon\mathbb{B}$ and $C \setminus D \subseteq \partial D + \epsilon\mathbb{B} \subseteq \partial C + \epsilon\mathbb{B} + \epsilon\mathbb{B} \subseteq \partial C + 2\epsilon\mathbb{B}$. \square

Definition 3. For a convex body C with the property that $0 \in int(C)$, define for $x \in \mathbb{S}$ the radial function

$$\rho_C(x) = \sup\{\gamma : \gamma x \in C\}.$$

The radial distance between two convex bodies C, D with $0 \in int(C)$, $0 \in int(D)$ is

$$rad(C, D) := \sup_{x \in \mathbb{S}} |\rho_C(x) - \rho_D(x)|.$$

Proposition 3. If C, D are two convex bodies with $r\mathbb{B} \subseteq C, D \subseteq R\mathbb{B}$, then $\frac{r}{R}rad(C, D) \leq Haus(C, D) \leq rad(C, D)$. Thus rad and $Haus$ are equivalent on the space of convex bodies C with $r\mathbb{B} \subseteq C \subseteq R\mathbb{B}$.

Proof. The inequality $Haus(C, D) \leq rad(C, D)$ is obvious. Thus we show that $\frac{r}{R}rad(C, D) \leq Haus(C, D)$. Let $p \in \partial C$ and $q \in \partial D$ be points such that $\|p - q\| = rad(C, D)$, such that p and q lie on a ray originating at 0 and suppose without loss of generality that $p \in D$ and $q \notin C$. Let $v \in \partial C$ be such that $\|v - q\| = \min_{w \in \partial C} \|w - q\|$. Then $Haus(C, D) \geq \|v - q\|$. Let now $t \in r\mathbb{S}$ the point on the ray from 0 in direction $q - v$. We have $\|t\| = r$. Notice that p, q, v, t lie in one hyperplane. Let u the point where the line between 0 and q intersects the line between t and v . Notice that $\|u - q\| \geq \|p - q\|$ by the convexity of C . So

$$\frac{Haus(C, D)}{r} \geq \frac{\|v - q\|}{\|t\|} = \frac{\|u - q\|}{\|u\|} \geq \frac{\|p - q\|}{R} = \frac{rad(C, D)}{R}$$

\square

Notice that the radial functions satisfy $\rho_{C \cap D} = \min_i \rho_{C_i}$ and therefore the next proposition is an easy consequence of Proposition 3.

Proposition 4. Let $(C_i)_{i=1, \dots, l}$ and $(D_i)_{i=1, \dots, l}$ be convex sets with

$$r\mathbb{B} \subseteq C, D \subseteq R\mathbb{B} \quad \text{for all } i = 1, \dots, l. \quad (13)$$

Then

$$Haus\left(\bigcap_{i=1}^l C_i, \bigcap_{i=1}^l D_i\right) \leq \frac{R}{r} \max_{i=1, \dots, l} Haus(C_i, D_i).$$

Notice that the condition (13) can be weakened to the requirement that it is fulfilled for a certain translation of the sets C_i, D_i .

Proposition 5. Let $\Theta \subset \mathbb{R}^d$ and let $\mathbf{a} : \Theta \rightarrow \mathbb{R}^s$ and $b : \Theta \rightarrow \mathbb{R}$ be differentiable functions with $\mathbf{a}(\vartheta) \neq 0$ for all $\vartheta \in \Theta$. Let $H(\vartheta) = \{x \in \mathbb{R}^s \mid \mathbf{a}(\vartheta)^T x \leq b(\vartheta)\}$ and suppose that $H(\vartheta) \cap R\mathbb{B} \neq \emptyset$ for all $\vartheta \in \Theta$. Then, for all $R > 0$, there exists a $c > 0$ such that for sufficiently small $h > 0$

$$\text{Haus}(H(\vartheta) \cap R\mathbb{B}, H(\vartheta + h \cdot \tau) \cap R\mathbb{B}) \leq c \cdot h$$

i.e. the sufficient condition SC is fulfilled.

References

- [1] Billingsley P. (1968). *Convergence of probability measures*. J. Wiley and Sons, New York.
- [2] Glasserman P. (1991). *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, Norwell, USA.
- [3] Lieber, D.; Nemirovski, A.; Rubinstein, R.Y. (1996) A fast Monte Carlo Method for Evaluation of Reliability Indices, IEEE Transactions on Reliability Systems, Vol. 48, No 3, 256-261, 1999.
- [4] Kibzun A. I., Kan Yu. S. (2000). Stochastic Programming problems with Probability and Quantile Functions. J. Wiley and Sons, New York
- [5] Krivulin N. (1993). An analysis of gradient estimates in stochastic network optimization problems. In: Model-oriented data analysis (Petrodvorets, 1992), Physica, Heidelberg, pp. 227–240, 1993.
- [6] Pflug G. Ch. (1996). Optimization of Stochastic Models The Interface Between Simulation and Optimization, Kluwer Academic Publishers.
- [7] Prekopa A. Stochastic Programming. Mathematics and its Applications: Volume 324, Kluwer Academic Publishers, 1995.
- [8] Prekopa, A. Network planning using two-stage programming under uncertainty. Recent results in stochastic programming (Proc. Meeting, Oberwolfach, 1979), pp. 215–237, Lecture Notes in Econom. and Math. Systems, 179, Springer, Berlin-New York, 1980. 90C15
- [9] Raik, E. (1975). The differentiability in the parameter of the probability function and optimization via the stochastic pseudogradient method. Izvestiya Akad. Nauk Est. SSR, Phis. Math. 24 (1), 3 – 6 (in Russian).
- [10] Rubinstein, R.Y., Shapiro A. (1992). Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. John Wiley & Sons, New York.

- [11] Rubinstein R.Y. (1992). Sensitivity Analysis of Discrete Event Systems by the "Push Out" Method. *Ann. Oper. Res.* 39, 229 – 250.
- [12] Uryasev S. (1995). Derivatives of probability functions and integrals over sets given by inequalities. *J. Computational and Applied Mathematics* Vol. 56.
- [13] Weisshaupt H. A measure-valued approach to convex set-valued dynamics. *Set-Valued Anal.* 9 (2001), no. 4, 337–373.