Applied Mathematics Letters 24 (2011) 1120–1123

Contents lists available at ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

Derivative of an integral over a convex polytope

C. Müller [a], A. Feuer [b,*], G.C. Goodwin [a]

[a] School of Electrical Engineering & Computer Science, University of Newcastle, NSW 2308, Australia

[b] Department of Electrical Engineering, Technion, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 29 October 2009
Accepted 20 January 2011

Keywords:
Integral over polytope

ABSTRACT

Many problems in optimization lead to a cost function in the form of an integral over a polytope. To find the gradient or Hessian of such costs one needs to take the derivative of an integral over a convex polytope with respect to the parameters defining the polytope. The contribution of the current paper is to present a formula for such derivatives.

1. Introduction

A variety of optimization problems have a cost function in the form of an integral over a polytope in $\mathbb{R}^n$. For example, this problem arises in Universal Barrier Functions [1,2] and Optimal Vector Quantization [3,4]. Specifically, the problem of finding an optimal vector quantizer is a central problem in source coding (see e.g. [4]). The cost function in these problems consists of a sum of integrals over a set of polytopes. Optimization is performed with respect to the parameters defining these polytopes. The development of effective algorithms for this class of problems requires that one must be able to generate an expression for the derivative of these integrals. The current contribution addresses this problem in a general setting. In particular, our expression enables the calculation of first and second derivatives. These, in turn, enable the development of Newton type algorithms.

In our formulation, the polytopes are described by a set of linear inequalities and we derive a formula for the derivative of an integral cost function with respect to the coefficients describing the inequalities. Our result extends earlier results [5] where derivatives were obtained with respect to parallel shifts in the polytope boundaries.

2. The results

Let $A : \mathbb{R} \to \mathbb{R}^{m\times n}$ and $\mathbf{b} : \mathbb{R} \to \mathbb{R}^m$ be functions with coefficients $A_{i,j} : \mathbb{R} \to \mathbb{R}$, $b_i : \mathbb{R} \to \mathbb{R}$, respectively. Then for fixed Θ, the set $\Omega(\Theta) = \{\mathbf{x} \in \mathbb{R}^n \mid A(\Theta)\mathbf{x} \le \mathbf{b}(\Theta)\}$ is a convex polytope in $\mathbb{R}^n$. We are interested in developing formulae for the derivatives of the function

$$g(\Theta) = \int_{\Omega(\Theta)} f(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

where $f : \mathbb{R}^n \to \mathbb{R}$. We also present conditions for the existence of these derivatives. We assume that $J : \mathbb{R} \to \mathbb{R}^{m\times n} \times \mathbb{R}^m$ given by $J(\Theta) = (A(\Theta), \mathbf{b}(\Theta))$ is differentiable. Then the chain rule allows us to concentrate on the derivative of a cost function

$$G : \mathbb{R}^{m\times n} \times \mathbb{R}^m = (\mathbb{R}^m)^{n+1} \to \mathbb{R}$$
$$G(A, \mathbf{b}) = \int_{A\mathbf{x}\le\mathbf{b}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

* Corresponding address: Israel Institute of Technology, Department of Electrical Engineering, 32000, Technion, Haifa, Israel.
E-mail addresses: claus.mueller@newcastle.edu.au (C. Müller), feuer@ee.technion.ac.il (A. Feuer), graham.goodwin@newcastle.edu.au (G.C. Goodwin).

0893-9659/$ – see front matter
doi:10.1016/j.aml.2011.01.037

The formula for $\frac{\partial G}{\partial b_i}$ has been presented earlier [5] where f is assumed to be positively homogeneous. We will present the formulae for both $\frac{\partial G}{\partial b_i}$ and $\frac{\partial G}{\partial A_{i,j}}$.

Let A_i be the ith row of A, and define the set $H(A_i, b_i) = \{\mathbf{x} \in \mathbb{R}^n \mid A_i\mathbf{x} \le b_i\}$, which is a half space if $A_i \neq 0$. We denote the boundary of $H(A_i, b_i)$ by $\partial H(A_i, b_i) = \{\mathbf{x} \in \mathbb{R}^n \mid A_i\mathbf{x} = b_i\}$.

Theorem 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be integrable, and assume that there are constants $c, d > 0$ and $p \in (n, \infty)$ such that*

$$|f(\mathbf{x})| \le \frac{c}{\|\mathbf{x}\|^p + d} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{1}$$

Assume that for some $1 \le i \le m$ the following conditions are satisfied:

(a) $A_i \neq 0$.
(b) *f is almost surely continuous on $\partial H(A_i, b_i)$. That is, the set*

$$\left\{\mathbf{z} \in \partial H(A_i, b_i) \mid \lim_{\mathbf{z}\in\mathbb{R}^n,\ \mathbf{z}\to\mathbf{x}} f(\mathbf{z}) \neq f(\mathbf{x})\right\}$$

is a null set with respect to the $n-1$-dimensional Lebesgue measure on $\partial H(A_i, b_i)$.
(c) *If $q \neq i$, then $\partial H(A_q, b_q) \neq \partial H(A_i, b_i)$.*

Then

$$\frac{\partial G}{\partial b_i}(A, \mathbf{b}) = \frac{1}{\|A_i\|} \int_{\mathcal{F}_i} f(\mathbf{x})\, \mathrm{d}\mathcal{F}_i \tag{2}$$

and

$$\frac{\partial G}{\partial A_{i,j}}(A, \mathbf{b}) = -\frac{1}{\|A_i\|} \int_{\mathcal{F}_i} x_j f(\mathbf{x})\, \mathrm{d}\mathcal{F}_i \quad \text{for all } 1 \le j \le n, \tag{3}$$

where

$$\mathcal{F}_i = \{\mathbf{x} \in \mathbb{R}^n \mid A_i\mathbf{x} = b_i,\ A_q\mathbf{x} \le b_q \text{ for all } q \neq i\}.$$

Proof. We note that, if $\Psi : M \subset \mathbb{R}^{n-1} \to \mathcal{F}_i$ is some C^1-parametrization of $\mathcal{F}_i$, then

$$\int_{\mathcal{F}_i} f(\mathbf{x})\, \mathrm{d}\mathcal{F}_i = \int_M f(\Psi(\mathbf{z})) \left| \det \left(\frac{\partial \Psi}{\partial z_1}(\mathbf{z}), \ldots, \frac{\partial \Psi}{\partial z_{k-1}}(\mathbf{z}), \frac{A_i}{\|A_i\|} \right) \right| \mathrm{d}\mathbf{z}.$$

If we define $M_i = \bigcap_{q\neq i} H(A_q, b_q)$ and $f_i(\mathbf{x}) = f(\mathbf{x}) \cdot \mathbf{1}_{M_i}(\mathbf{x})$ we obtain

$$G(A, \mathbf{b} + h\mathbf{e}_i) - G(A, \mathbf{b}) = \int_{H(A_i, b_i+h)} f_i(\mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{H(A_i, b_i)} f_i(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

where $\mathbf{e}_i$ is the ith unit vector in $\mathbb{R}^m$ and $\mathbf{1}_{M_i}$ is the characteristic function of M_i. We next apply the coordinate transformation

$$\mathbf{x} = \left[T \ \ \frac{A_i^T}{\|A_i\|} \right] \begin{pmatrix} \widetilde{x} \\ \widetilde{x}_n \end{pmatrix} + b_i \frac{A_i^T}{\|A_i\|^2} = \varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n),$$

where the columns of $[T \ \ \frac{A_i^T}{\|A_i\|}]$ are any orthonormal basis of $\mathbb{R}^n$. We then obtain

$$G(A, \mathbf{b} + h\mathbf{e}_i) - G(A, \mathbf{b}) = \int_{\mathbb{R}^{n-1}} \int_0^{h/\|A_i\|} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n \mathrm{d}\widetilde{\mathbf{x}}.$$

Using assumption (1) and the substitution $u = \widetilde{x}_n + b_i/\|A_i\|$ we obtain

$$\begin{aligned} \left| \frac{1}{h} \int_0^{h/\|A_i\|} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n \right| &\le \frac{1}{|h|} \int_{-|h|/\|A_i\|}^{|h|/\|A_i\|} \frac{c}{\|\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n)\|^p + d}\, \mathrm{d}\widetilde{x}_n \\ &= \frac{1}{|h|} \int_{(b_i-|h|)/\|A_i\|}^{(b_i+|h|)/\|A_i\|} \frac{c}{\left\| T\widetilde{\mathbf{x}} + u\frac{A_i^T}{\|A_i\|} \right\|^p + d}\, \mathrm{d}u \\ &\le \frac{2c}{\|A_i\|(\|\widetilde{\mathbf{x}}\|^p + d)} \\ &= K(\widetilde{\mathbf{x}}). \end{aligned}$$

Since $p > n - 1$ we see that K is integrable over $\mathbb{R}^{n-1}$. We can thus apply the dominated convergence theorem to obtain

$$\lim_{h\to 0} \frac{G(A, \mathbf{b} + h\mathbf{e}_i) - G(A, \mathbf{b})}{h} = \int_{\mathbb{R}^{n-1}} \lim_{h\to 0} \frac{1}{h} \int_0^{h/\|A_i\|} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n\, \mathrm{d}\widetilde{\mathbf{x}} = \frac{1}{\|A_i\|} \int_{\mathbb{R}^{n-1}} f_i(\varphi_i(\widetilde{\mathbf{x}}, 0))\, \mathrm{d}\widetilde{\mathbf{x}},$$

since, by assumptions (b) and (c), $\lim_{h\to 0} \frac{1}{h} \int_0^{h/\|A_i\|} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n = \frac{1}{\|A_i\|} f_i(\varphi_i(\widetilde{\mathbf{x}}, 0))$ for almost all $\widetilde{\mathbf{x}} \in \mathbb{R}^{n-1}$. Thus

$$\lim_{h\to 0} \frac{G(A, \mathbf{b} + h\mathbf{e}_i) - G(A, b_i)}{h} = \frac{1}{\|A_i\|} \int_{\mathcal{F}_i} f(\mathbf{x})\, \mathrm{d}\mathcal{F}_i.$$

Similarly, if $|h|$ is sufficiently small,

$$G(A + h\mathbf{e}_i\mathbf{e}_j^T, \mathbf{b}) - G(A, \mathbf{b}) = \int_{H(A_i + h\mathbf{e}_j^T, b_i)} f_i(\mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{H(A_i, b_i)} f_i(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^{n-1}} \int_0^{\eta(\widetilde{\mathbf{x}}, h)} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n\, \mathrm{d}\widetilde{\mathbf{x}},$$

where

$$\eta(\widetilde{\mathbf{x}}, h) = -h \frac{\mathbf{e}_j^T \left(T\widetilde{\mathbf{x}} + b_i \frac{A_i^T}{\|A_i\|^2} \right)}{\|A_i\| + h \frac{\mathbf{e}_j^T A_i^T}{\|A_i\|}}.$$

Again using assumption (1) and the substitution $u = \widetilde{x}_n + b_i/\|A_i\|$ we obtain

$$\left| \frac{1}{h} \int_0^{\eta(\widetilde{\mathbf{x}}, h)} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n \right| \le \frac{1}{|h|} \frac{2c|\eta(\widetilde{\mathbf{x}}, h)|}{\|\mathbf{x}\|^p + d}$$

which is integrable over $\mathbb{R}^{n-1}$ since $p > n$. Arguing as above, we obtain

$$\begin{aligned} \frac{\partial G}{\partial A_{i,j}}(A, \mathbf{b}) &= \int_{\mathbb{R}^{n-1}} \lim_{h\to 0} \frac{1}{h} \int_0^{\eta(\widetilde{\mathbf{x}}, h)} f_i(\varphi_i(\widetilde{\mathbf{x}}, \widetilde{x}_n))\, \mathrm{d}\widetilde{x}_n\, \mathrm{d}\widetilde{\mathbf{x}} \\ &= \int_{\mathbb{R}^{n-1}} f_i(\varphi_i(\widetilde{\mathbf{x}}, 0)) \cdot \lim_{h\to 0} \frac{\eta(\widetilde{\mathbf{x}}, h)}{h}\, \mathrm{d}\widetilde{\mathbf{x}} \\ &= -\frac{1}{\|A_i\|} \int_{\mathbb{R}^{n-1}} f_i(\varphi_i(\widetilde{\mathbf{x}}, 0)) \cdot \mathbf{e}_j^T \varphi_i(\widetilde{\mathbf{x}}, 0)\, \mathrm{d}\widetilde{\mathbf{x}} \\ &= -\frac{1}{\|A_i\|} \int_{\mathcal{F}_i} x_j f(\mathbf{x})\, \mathrm{d}\mathcal{F}_i. \end{aligned}$$

This completes the proof. □

Remark 2. (i) Condition (b) is satisfied if f is continuous, or if f is almost surely continuous on every hyperplane.
(ii) Condition (c) is equivalent to $q \neq i \Rightarrow (b_q A_i \neq b_i A_q$ or A_i, A_q are linear independent).

Remark 3. We note that (2) and (3) can be rewritten as

$$\frac{\partial G}{\partial b_i}(A, \mathbf{b}) = \frac{1}{\|A_i\|} \int_{\widetilde{A}\mathbf{x} \le \widetilde{\mathbf{b}}} \widetilde{f}_i(\widetilde{\mathbf{x}})\, \mathrm{d}\widetilde{\mathbf{x}} \tag{4}$$

and

$$\frac{\partial G}{\partial A_{i,j}}(A, \mathbf{b}) = -\frac{1}{\|A_i\|} \int_{\widetilde{A}\mathbf{x} \le \widetilde{\mathbf{b}}} (e_j^T AT\widetilde{\mathbf{x}}) \widetilde{f}_i(\widetilde{\mathbf{x}})\, \mathrm{d}\widetilde{\mathbf{x}} \quad \text{for all } 1 \le j \le n, \tag{5}$$

where $\widetilde{\mathbf{x}} \in \mathbb{R}^{n-1}$

$$\begin{aligned} \widetilde{A} &= E_i AT \\ \widetilde{\mathbf{b}} &= E_i \mathbf{b} \\ \widetilde{f}_i(\widetilde{\mathbf{x}}) &= f(\varphi_i(\widetilde{\mathbf{x}}, 0)). \end{aligned}$$

T and $\varphi_i(\widetilde{\mathbf{x}}, 0)$ are as in the proof of Theorem 1. The significance of this observation is that the derivative turns out to be an integral over a polytope in $\mathbb{R}^{n-1}$. There are a number of packages available for calculating such integrals (see e.g. the website [6]). $\square$

We next give the conditions under which G is continuously differentiable on some open subset of $\mathbb{R}^{m\times n} \times \mathbb{R}^m$. To this end, we first note that the set

$$\mathcal{Z} = \{(A, \mathbf{b}) \in \mathbb{R}^{m\times n} \times \mathbb{R}^m \mid A_i \neq 0 \text{ for all } i, \text{ and } \partial H(A_q, b_q) \neq \partial H(A_i, b_i),\ q \neq i\}$$

is open (and dense) in $\mathbb{R}^{m\times n} \times \mathbb{R}^m$. This follows from the following facts, which are not hard to prove:

(1) If $i \neq q$ then $T_{i,q} = \{(A, \mathbf{b}) \mid b_q A_i \neq b_i A_q\}$ is open and dense.
(2) $L_{i,q} = \{(A, \mathbf{b}) \mid A_i$ and A_q are linear independent$\}$ is open.
(3) $S_i = \{(A, \mathbf{b}) \mid A_i \neq 0\}$ is open and dense.
(4) Since $\mathcal{Z} = (\bigcap_{i=1}^m S_i) \cap \bigcap_{i\neq q}(T_{i,q} \cup L_{i,q})$ it is then clear that this set is open and dense.

Theorem 4. *If f is integrable and almost surely continuous on every hyperplane such that condition (1) is satisfied then G is continuously differentiable on $\mathcal{Z}$.*

Proof. We show that the partial derivatives are continuous in $(A, \mathbf{b}) \in \mathcal{Z}$, i.e. for all i, j we have $\lim_{B\to 0,\mathbf{v}\to 0} \frac{\partial G}{\partial A_{i,j}}(A + B, \mathbf{b} + \mathbf{v}) = \frac{\partial G}{\partial A_{i,j}}(A, \mathbf{b})$ and $\lim_{B\to 0,\mathbf{v}\to 0} \frac{\partial G}{\partial b_i}(A + B, \mathbf{b} + \mathbf{v}) = \frac{\partial G}{\partial b_i}(A, \mathbf{b})$.

To this end, we fix i and define the sets $M_{i,B,\mathbf{v}} = \bigcap_{q\neq i} H(A_q + B_q, b_q + v_q)$ and the transformations $\varphi_{i,B,\mathbf{v}}(\widetilde{\mathbf{x}}, \widetilde{x}_n) = [T_B \ \ \frac{A_i^T + B_i^T}{\|A_i + B_i\|}] \binom{\widetilde{x}}{\widetilde{x}_n} + (b_i + v_i) \frac{A_i^T + B_i^T}{\|A_i + B_i\|^2}$, where the columns of $[T_B \ \ \frac{A_i^T + B_i^T}{\|A_i + B_i\|}]$ are any orthonormal basis of $\mathbb{R}^n$ and $\lim_{B\to 0,\mathbf{v}\to 0} \varphi_{i,B,\mathbf{v}} = \varphi_i$ pointwise.

Then, if $f_{i,B,\mathbf{v}} = f \cdot \mathbf{1}_{M_{i,B,\mathbf{v}}}$,

$$\frac{\partial G}{\partial b_i}(A + B, \mathbf{b} + \mathbf{v}) = \frac{1}{\|A_i + B_i\|} \int_{\mathbb{R}^{n-1}} f_{i,B,\mathbf{v}}(\varphi_{i,B,\mathbf{v}}(\widetilde{\mathbf{x}}, 0))\, \mathrm{d}\widetilde{\mathbf{x}}.$$

Application of the dominated convergence theorem in a similar way as in the proof of Theorem 1, establishes the assertion for $\frac{\partial G}{\partial b_i}$. The rest of the proof is straightforward. $\square$

References

[1] Yu.E. Nesterov, A.S. Nemirovskii, Interior Point Polynomial Methods in Convex Programming, SIAM, Philadelphia, PA, 1994.
[2] O. Guler, On the self-concordance of the universal barrier function, SIAM J. Optim. 7 (2) (1997) 295–303.
[3] S.P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory IT-28 (2) (1982) 129–137.
[4] A. Gersho, R.M. Gray, Vector Quantization and Signal Compression, Springer, Netherlands, 1992.
[5] J.B. Lasserre, Integration on a convex polytope, Proc. Amer. Math. Soc. 126 (8) (1998) 2433–2441.
[6] http://library.wolfram.com/infocenter/MathSource/3652/.