Journal of Convex Analysis
Volume 8 (2001), No. 1, 109–126

A Higher-Order Smoothing Technique for Polyhedral Convex Functions: Geometric and Probabilistic Considerations

Sophie Guillaume
Department of Mathematics, University of Avignon,
33, rue Louis Pasteur, 84000 Avignon, France.
e-mail: sophie.guillaume@univ-avignon.fr

Alberto Seeger
Department of Mathematics, University of Avignon,
33, rue Louis Pasteur, 84000 Avignon, France.
e-mail: alberto.seeger@univ-avignon.fr

Received July 8, 1999
Revised manuscript received October 2, 2000

Let $\mathbb{R}^n$ denote the usual n-dimensional Euclidean space. A polyhedral convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ can always be seen as the pointwise limit of a certain family $\{f^t\}_{t>0}$ of C^∞ convex functions. An explicit construction of this family $\{f^t\}_{t>0}$ can be found in a previous paper by the second author. The aim of the present work is to further explore this C^∞-approximation scheme. In particular, one shows how the family $\{f^t\}_{t>0}$ yields first and second-order information on the behavior of f. Links to linear programming and Legendre-Fenchel duality theory are also discussed.

Keywords: Polyhedral convex function, smooth approximation, subgradient, linear programming

1991 Mathematics Subject Classification: 41A30, 52B70, 60E10

1. Introduction

Throughout this note $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is assumed to be a polyhedral convex function in the sense that its epigraph

$$\operatorname{epi} f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}$$

is a polyhedral convex set [6, p. 172]. An important consequence of the polyhedrality assumption is that a finite number of vectors in $\mathbb{R}^n \times \mathbb{R}$ are enough to determine completely the function f. Indeed, it is possible to represent f in the following "canonical" form:

$$f(x) = \begin{cases} \operatorname{Max}\{\langle w^1, x\rangle - \beta^1, \cdots, \langle w^p, x\rangle - \beta^p\} & \text{if } x \in K \ , \\ +\infty & \text{if } x \notin K \ , \end{cases} \tag{1}$$

with

$$K = \{x \in \mathbb{R}^n : \langle a^i, x\rangle - \gamma^i \leq 0 \qquad \forall i = 1, \cdots, q\} \ . \tag{2}$$

The symbol $\langle \cdot, \cdot \rangle$ denotes here the usual Euclidean product in the space $\mathbb{R}^n$. Two comments concerning the data

$$\{(w^j, \beta^j)\}_{j=1}^p \ , \ \ \{(a^i, \gamma^i)\}_{i=1}^q$$

ISSN 0944-6532 / \$ 2.50

are in order. On the one hand side, the finite term

$$h(x) = \mathrm{Max}\{\langle w^1, x\rangle - \beta^1, \cdots, \langle w^p, x\rangle - \beta^p\} \tag{3}$$

can be written in the form

$$h(x) = \sigma_P(x, -1)\ ,$$

where $\sigma_P : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ stands for the support function of the polytope

$$\begin{aligned} P &= \mathrm{conv}\{(w^1, \beta^1), \cdots, (w^p, \beta^p)\} \\ &= \{\sum_{j=1}^{p} \mu_j (w^j, \beta^j) : \sum_{j=1}^{p} \mu_j = 1\ ,\ \mu_j \geq 0 \qquad \forall j = 1, \cdots, p\}\ . \end{aligned}$$

If a particular vector (w^j, β^j) is not an extreme point of P, then the corresponding affine function $\langle w^j, \cdot\rangle - \beta^j$ can be deleted from (3) without affecting the values of h. On the other hand, the set K admits the alternative expression

$$K = \{x \in \mathbb{R}^n : \sigma_Q(x, -1) \leq 0\}$$

where $\sigma_Q : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ stands for the support function of the polyhedral convex cone

$$\begin{aligned} Q &= \mathrm{cone}\{(a^1, \gamma^1), \cdots, (a^q, \gamma^q)\} \\ &= \{\sum_{i=1}^{q} \nu_i (a^i, \gamma^i) : \nu_i \geq 0 \qquad \forall i = 1, \cdots, q\}\ . \end{aligned}$$

If a particular vector (a^i, γ^i) is a positive linear combination of the others, then the deletion of the corresponding affine function $\langle a^i, \cdot\rangle - \gamma^i$ from (2) does not affect the set K.

As shown recently by Seeger [7, Theorem 3.1], one can always construct a family $\{f^t\}_{t>0}$ such that

$$\begin{cases} \text{each } f^t : \mathbb{R}^n \to \mathbb{R} \text{ is convex and infinitely often differentiable ;} \\ \{f^t\}_{t>0} \text{ converges pointwise to } f \text{ as the parameter } t \text{ goes to } \infty\ . \end{cases} \tag{4}$$

Thus, when $t > 0$ is sufficiently large, f^t can be viewed as a C^∞-approximation of f. As example of such a family $\{f^t\}_{t>0}$, one may consider

$$f^t(x) := \frac{M(tx, -t)}{t} \qquad \forall x \in \mathbb{R}^n\ , \tag{5}$$

with

$$M(x, \alpha) := c + \int_{\mathbb{R}^n \times \mathbb{R}} e^{\langle a, x\rangle + \gamma\alpha} d\nu(a, \gamma) + \log\left[\int_{\mathbb{R}^n \times \mathbb{R}} e^{\langle w, x\rangle + \beta\alpha} d\mu(w, \beta)\right]. \tag{6}$$

Here ν is any discrete measure concentrated on $\{(a^1, \gamma^1), \cdots, (a^q, \gamma^q)\}$, and μ is any discrete measure concentrated on $\{(w^1, \beta^1), \cdots, (w^p, \beta^p)\}$. The constant $c \in \mathbb{R}$ in (6) is irrelevant, but for notational convenience, one chooses the value of c that yields the normalization condition $M(0,0) = 0$. The consequence of this particular choice is that $f^t(x)$ is nondecreasing with respect to the parameter t, and therefore $\{f^t\}_{t>0}$ converges monotonically upwards to f as t goes to ∞.

If ν_i denotes the weight of (a^i, γ^i) and μ_j denotes the weight of (w^j, β^j), then the function M takes the form

$$M(x,\alpha) = c + \sum_{i=1}^{q} \nu_i e^{\langle a^i, x\rangle + \gamma^i \alpha} + \log\left[\sum_{j=1}^{p} \mu_j e^{\langle w^j, x\rangle + \beta^j \alpha}\right], \tag{7}$$

with

$$c = -\sum_{i=1}^{q} \nu_i - \log[\sum_{j=1}^{p} \mu_j] .$$

For the sake of the exposition, $M : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ will be referred to as a generating function for f. As a trivial justification for this terminology, one may invoke the formula

$$f(x) = \lim_{t\to\infty} \frac{M(tx,-t)}{t} = \sup_{t>0} \frac{M(t(x,-1))}{t} \tag{8}$$

yielding the values of f. There is however a deeper justification. The purpose of this note is to show that M provides a wealth of information on the behavior of f.

2. Limiting formula for subgradients

Consider a reference point $\bar{x} \in \mathbb{R}^n$ at which the polyhedral convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is finite. The first-order behavior of f around $\bar{x}$ is reflected by the nonempty set

$$\partial f(\bar{x}) := \{y \in \mathbb{R}^n : f(x) \geq f(\bar{x}) + \langle y, x - \bar{x}\rangle \qquad \forall x \in \mathbb{R}^n\} . \tag{9}$$

This set is known as the subdifferential of f at $\bar{x}$, and each of its elements is called a subgradient of f at $\bar{x}$ (see Rockafellar [6] or Moreau [5]).

A natural question to ask is whether the subgradients of f can be obtained by computing classical gradients of smooth approximations of f. More precisely, if f^t denotes the C^∞-approximation of f given by (5)–(6), what happens with the gradient

$$\nabla f^t(\bar{x}) = (\frac{\partial f^t}{\partial x_1}(\bar{x}), \cdots, \frac{\partial f^t}{\partial x_n}(\bar{x}))^T$$

when the parameter t goes to ∞ ? Does one get closer to an element of $\partial f(\bar{x})$? Which one is the subgradient that is being approached in this way ?

All these questions will be answered in a clear-cut manner, but first one needs to introduce some adjustments in the notation. First of all, the generating function M, as given by the expression (7), depends on the weight vector

$$\xi := (\nu_1, \cdots, \nu_q \ , \ \mu_1, \cdots, \mu_p) \in \mathbb{R}^q_+ \times \mathbb{R}^p_+ .$$

To underline the role of ξ, we shall write

$$M_\xi(x,\alpha) = \sum_{i=1}^{q} \nu_i \{e^{\langle a^i, x\rangle + \gamma^i \alpha} - 1\} + \log\{\sum_{i=1}^{p} \frac{\mu_j}{\langle \mu\rangle} e^{\langle w^j, x\rangle + \beta^j \alpha}\} .$$

The above expression is obtained by getting rid of the constant c, and by setting $\langle\mu\rangle = \mu_1+\cdots+\mu_p$. Of course, the case $\langle\mu\rangle = 0$ has been ruled out. Without loss of generality, one may suppose that $\langle\mu\rangle = 1$ (otherwise, one works with the new coefficients $\mu_j' := \mu_j/\langle\mu\rangle$). In short,

$$M_\xi(x,\alpha) = \sum_{i=1}^{q} \nu_i\{e^{\langle a^i,x\rangle+\gamma^i\alpha} - 1\} + \log\{\sum_{j=1}^{p} \mu_j e^{\langle w^j,x\rangle+\beta^j\alpha}\}\ , \tag{10}$$

with a weight vector ξ ranging over

$$\Xi := \{(\nu_1,\cdots,\nu_q\ ,\ \mu_1,\cdots,\mu_p) \in \mathbb{R}^q_+ \times \mathbb{R}^p_+ : \sum_{j=1}^{p} \mu_j = 1\}\ .$$

The same notational consideration leads us to write

$$f^t_\xi(x) = \frac{M_\xi(tx,-t)}{t}\ , \tag{11}$$

or in full extent

$$f^t_\xi(x) = \frac{1}{t}\sum_{i=1}^{q} \nu_i\{e^{t[\langle a^i,x\rangle-\gamma^i]} - 1\} + \frac{1}{t}\ \log\{\sum_{j=1}^{p}\ \mu_j\ e^{t[\langle w^j,x\rangle-\beta^j]}\}\ . \tag{12}$$

Finally, we introduce the notation

$$I(\bar{x}) := \{i \in \{1,\cdots,q\} : \langle a^i,\bar{x}\rangle - \gamma^i = 0\}\ ,$$
$$J(\bar{x}) := \{j \in \{1,\cdots,p\} : \langle w^j,\bar{x}\rangle - \beta^j = h(\bar{x})\}\ ,$$
$$\Xi(\bar{x}) = \{(\nu_1,\cdots,\nu_q\ ,\ \mu_1,\cdots,\mu_p) \in \Xi :\ \nu_i = 0 \text{ for } i \notin I(\bar{x}) \ \text{ and } \ \ \mu_j = 0 \text{ for } j \notin J(\bar{x})\}\ .$$

Without further ado, we state:

Theorem 2.1. *Consider a polyhedral convex function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ given in the canonical form (1). Let $\bar{x} \in \mathbb{R}^n$ be a point at which f is finite. Then, the following statements are equivalent:*

(a) *$\bar{y}$ is a subgradient of f at $\bar{x}$;*

(b) *$\bar{y} = \sum_{i\in I(\bar{x})} \nu_i a^i + \sum_{j\in J(\bar{x})} \mu_j w^j$ for suitable coefficients $\{\nu_i\}_{i\in I(\bar{x})}$ and $\{\mu_j\}_{j\in J(\bar{x})}$ satisfying*

$$\nu_i \geq 0 \quad \forall i \in I(\bar{x})\ ,\ \mu_j \geq 0 \quad \forall j \in J(\bar{x})\ ,\ \sum_{j\in J(\bar{x})} \mu_j = 1\ ;$$

(c) *there is a weight vector $\xi \in \Xi(\bar{x})$ such that*

$$\bar{y} = \lim_{t\to\infty} \nabla f^t_\xi(\bar{x})\ .$$

Proof. The equivalence between (a) and (b) is well known, and is obtained by applying standard calculus rules for computing subdifferentials (see, for instance, [4]). The novelty of this theorem lies in the condition (c). A straightforward computation shows that

$$\nabla f_\xi^t(\bar{x}) = \sum_{i=1}^{q} \nu_i e^{t[\langle a^i,\bar{x}\rangle - \gamma^i]} a^i \; + \; \frac{\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j]} w^j}{\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j]}} \; . \tag{13}$$

Observe that

$$\lim_{t\to\infty} \nu_i e^{t[\langle a^i,\bar{x}\rangle - \gamma^i]} a^i = \begin{cases} 0 & \text{if } \langle a^i,\bar{x}\rangle - \gamma^i < 0 \;, \\ \nu_i a^i & \text{if } \langle a^i,\bar{x}\rangle - \gamma^i = 0 \;. \end{cases}$$

Therefore,

$$\lim_{t\to\infty} \sum_{i=1}^{q} \nu_i e^{t[\langle a^i,\bar{x}\rangle - \gamma^i]} a^i = \sum_{i\in I(\bar{x})} \nu_i a^i \; .$$

On the other hand,

$$\begin{aligned} v_t &:= \Big\{\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j]}\Big\}^{-1} \sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j]} w^j \\ &= \Big\{\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j - h(\bar{x})]}\Big\}^{-1} \sum_{j=1}^{p} \mu_j e^{t[\langle w^j,\bar{x}\rangle - \beta^j - h(\bar{x})]} w^j \; . \end{aligned}$$

Hence,

$$\lim_{t\to\infty} v_t = \frac{\sum_{j\in J(\bar{x})} \mu_j w^j}{\sum_{j\in J(\bar{x})} \mu_j} \; .$$

Putting all the pieces together, it is not hard to see that (b) $\Longleftrightarrow$ (c). The details of the proof are left aside. □

3. Link to linear programming

Before passing to the second-order analysis of f, we make a pause in our way and provide a further interpretation of the approximating function f_ξ^t. Let us have a closer look at the terms

$$g_\nu^t(x) := \frac{1}{t} \sum_{i=1}^{q} \nu_i \{e^{t[\langle a^i,x\rangle - \gamma^i]} - 1\} \; ,$$

$$h_\mu^t(x) := \frac{1}{t} \log\Big\{\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,x\rangle - \beta^j]}\Big\}$$

appearing in the decomposition of f_ξ^t. Both terms have an interesting interpretation in the context of linear programming. Recall that the standard form of a linear programming problem is

$$\begin{cases} \text{Minimize} \langle \gamma, r\rangle \\ \text{subject to } \; Ar = z \; , \; r \in \mathbb{R}_+^q \; , \end{cases} \tag{14}$$

where $\gamma = (\gamma^1, \cdots, \gamma^q)^T \in \mathbb{R}^q$ and $z = (z_1, \cdots, z_n)^T \in \mathbb{R}^n$ are given vectors, and

$$A = \left[a^1 \vdots a^2 \vdots \quad \vdots a^q \right]$$

is a matrix of appropriate dimensions. In full extent, this problem is expressed as

$$\begin{cases} \text{Minimize} \sum_{i=1}^q \gamma^i r_i \\ \text{subject to } \sum_{i=1}^q r_i a^i = z \ , \ r_i \geq 0 \quad \forall i = 1, \cdots, q \ . \end{cases}$$

The positivity constraints in (14) can be incorporated in the objective function through the penalty term

$$Q_\nu(r) := \begin{cases} \sum_{i=1}^q \{\nu_i + r_i[\log(\frac{r_i}{\nu_i}) - 1]\} & \text{if } r \in \mathbb{R}^q_+ \\ +\infty & \text{otherwise} \ , \end{cases} \tag{15}$$

where the convention $0 \log 0 = 0$ is in force. Thus,

$$\begin{cases} \text{Minimize} \langle \gamma, r \rangle + \dfrac{1}{t} Q_\nu(r) \\ \text{subject to } Ar = z \end{cases} \tag{16}$$

corresponds to a penalized version of (14). This specific penalization scheme has been studied in depth by Cominetti and San Martin [1], at least from the point of view of the asymptotic analysis of optimal trajectories. Observe that (16) falls within the framework of convex analysis but it is no longer a linear programming problem.

As shown next, Legendre-Fenchel conjugation theory [6, Section 16] provides a framework for the interpretation of g^t_ν:

Proposition 3.1. *The functions g^t_ν and*

$$z \in \mathbb{R}^n \longmapsto v^t_\nu(z) := \operatorname*{Inf}_{Ar=z} \{\langle \gamma, r \rangle + \frac{1}{t} \ Q_\nu(r)\}$$

are mutually conjugate.

Proof. The proposition says that

$$v^t_\nu(z) = (g^t_\nu)^*(z) := \operatorname*{Sup}_{x \in \mathbb{R}^n} \{\langle z, x \rangle - g^t_\nu(x)\} \qquad \forall z \in \mathbb{R}^n$$

and

$$g^t_\nu(x) = (v^t_\nu)^*(x) := \operatorname*{Sup}_{z \in \mathbb{R}^n} \{\langle z, x \rangle - v^t_\nu(z)\} \qquad \forall x \in \mathbb{R}^n \ .$$

Since $g^t_\nu : \mathbb{R}^n \to \mathbb{R}$ is convex, it is enough to prove the first equality ([6, Theorem 12.2]). Observe that

$$g^t_\nu(x) = \frac{1}{t} \ M_\nu(t(A^T x - \gamma)) \qquad \forall x \in \mathbb{R}^n \ , \tag{17}$$

with A^T denoting the transpose of A and $M_\nu : \mathbb{R}^q \to \mathbb{R}$ being defined by

$$M_\nu(c) = \sum_{i=1}^{q} \nu_i(e^{c_i} - 1) \ .$$

A direct calculation shows that

$$M_\nu^*(r) := \operatorname*{Sup}_{c\in\mathbb{R}^q} \left\{ \langle r, c\rangle - \sum_{i=1}^{q} \nu_i(e^{c_i} - 1) \right\} = \sum_{i=1}^{q} \operatorname*{Sup}_{\tau\in\mathbb{R}} \left\{ r_i\tau - \nu_i(e^{\tau} - 1) \right\} = Q_\nu(r) \ .$$

On the other hand

$$(g_\nu^t)^*_{(z)} = \operatorname*{Sup}_{x\in\mathbb{R}^n} \left\{ \langle z, x\rangle - \frac{1}{t} M_\nu\big(t(A^T x - \gamma)\big) \right\} = \frac{1}{t} \operatorname*{Sup}_{x\in\mathbb{R}^n} \left\{ \langle z, x\rangle - M_\nu\big(A^T x - t\gamma\big) \right\} \ .$$

To conclude one just needs to apply a well known formula for the conjugate of the composition of a real-valued convex function and an affine mapping (see [3]). □

The interpretation of h_μ^t follows the same pattern as before, but this time one has to start with a linear programming problem written in Karmarkar's form:

$$\begin{cases} \text{Minimize} \langle \beta, s\rangle \\ \text{subject to } \ Ws = z \ , \ \langle \mathbf{1}, s\rangle = 1 \ , \ s \in \mathbb{R}^p_+ \ . \end{cases} \tag{18}$$

Here $\beta = (\beta^1, \cdots, \beta^p)^T \in \mathbb{R}^p$, $z = (z_1, \cdots, z_n)^T \in \mathbb{R}^n$, and $\mathbf{1} = (1, \cdots, 1)^T \in \mathbb{R}^p$ are given vectors, and

$$W = \left[w^1 \vdots w^2 \vdots \quad \vdots w^p \right]$$

is a matrix of appropriate dimensions. Karmarkar's formulation stipulates that the minimization variables add up to 1. This fact leads us to consider a penalty term of the form

$$P_\mu(s) = \begin{cases} \sum_{j=1}^{p} \ s_j \log(\dfrac{s_j}{\mu_j}) & \text{if } s \in \mathbb{R}^p_+ \ , \ \displaystyle\sum_{j=1}^{p} s_j = 1 \\ +\infty & \text{otherwise .} \end{cases} \tag{19}$$

As shown in the next proposition, the smoothing function h_μ^t turns out to be related to the penalized version

$$\begin{cases} \text{Minimize} \langle \beta, s\rangle + \dfrac{1}{t} \ P_\mu(s) \\ \text{subject to } \ Ws = z \end{cases} \tag{20}$$

of the linear programming problem (18). The opposite of P_μ corresponds to the Kullback-Leibler entropy function associated to μ. For this reason, some authors refer to (20) as an entropic regularization method (cf. Fang and Li [2], Sheu and Wu [8], and the references therein). Of course, (20) is no longer a linear programming problem.

Proposition 3.2. *The functions* h_μ^t *and*

$$z \in \mathbb{R}^n \longmapsto \ell_\mu^t(z) := \operatorname*{Inf}_{Ws=z} \{\langle \beta, s\rangle + \frac{1}{t}\, P_\mu(s)\}$$

are mutually conjugate.

Proof. It is similar to the proof of Proposition 3.1. It is based on the representation

$$h_\mu^t(x) = \frac{1}{t}\, N_\mu(t(W^T x - \beta)) \qquad \forall x \in \mathbb{R}^n\ ,$$

and the fact that P_μ is the conjugate function of

$$d \in \mathbb{R}^p \longmapsto N_\mu(d) = \log(\sum_{j=1}^{p} \mu_j e^{d_j})\ .$$

□

By combining the standard formulation (14) and Karmarkar's formulation (18), one obtains the mixed form of a linear programming problem:

$$\begin{cases} \text{Minimize} \displaystyle\sum_{i=1}^{q} \gamma^i r_i \ + \ \sum_{j=1}^{p} \beta^j s_j \\ \text{subject to } \displaystyle\sum_{i=1}^{q} r_i a^i \ + \ \sum_{j=1}^{p} s_j w^j = z\ , \\ r_i \geq 0 \ \ \forall i = 1, \cdots, q\ , \quad s_j \geq 0 \ \ \forall j = 1, \cdots, p\ ,\ \sum_{j=1}^{p} s_j = 1\ . \end{cases}$$

For this problem to be feasible, it is necessary and sufficient that

$$z \in \Omega := \operatorname{cone}\{a^1, \cdots, a^q\} \ + \ \operatorname{conv}\{w^1, \cdots, w^p\}\ .$$

The mixed linear programming problem can be written also in a more compact way, namely

$$\begin{cases} \text{Minimize} \langle \begin{bmatrix} \gamma \\ \beta \end{bmatrix} , \begin{bmatrix} r \\ s \end{bmatrix} \rangle \\ \text{subject to } \ C \begin{bmatrix} r \\ s \end{bmatrix} = z\ ,\ \langle \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} , \begin{bmatrix} r \\ s \end{bmatrix} \rangle = 1\ ,\ \begin{bmatrix} r \\ s \end{bmatrix} \in \mathbb{R}_+^{q+p}\ , \end{cases} \tag{21}$$

with

$$C = \left[a^1 \vdots a^2 \vdots \ \ \vdots a^q \vdots w^1 \vdots w^2 \vdots \ \ \vdots w^p\right]\ .$$

As a penalization term for the basic constraints

$$r \in \mathbb{R}_+^q\ ,\ s \in \mathbb{R}_+^p\ ,\ \langle \mathbf{1}, s\rangle = 1\ ,$$

one may consider the expression

$$R_\xi(r,s) = Q_\nu(r) + P_\mu(s) \ ,$$

yielding in this way the penalized version

$$\begin{cases} \text{Minimize}\langle \begin{bmatrix} \gamma \\ \beta \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} \rangle + \dfrac{1}{t} R_\xi(r,s) \\[2em] \text{subject to} \quad C \begin{bmatrix} r \\ s \end{bmatrix} = z \end{cases} \tag{22}$$

of the linear programming problem (21). The infimal-value function

$$z \in \mathbb{R}^n \longmapsto m_\xi^t(z) := \operatorname*{Inf}_{C \left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right] = z} \left\{ \langle \begin{bmatrix} \gamma \\ \beta \end{bmatrix} , \begin{bmatrix} r \\ s \end{bmatrix} \rangle + \frac{1}{t} R_\xi(r,s) \right\}$$

is closely related to the smoothing function f_ξ^t. At this stage of our exposition, it should be no surprise that f_ξ^t and m_ξ^t are related through conjugacy:

Theorem 3.3. *The function f_ξ^t and m_ξ^t are mutually conjugate.*

Proof. According with Propositions 3.1 and 3.2, the conjugate function of f_ξ^t is given by

$$(f_\xi^t)^* = (g_\nu^t + h_\mu^t)^* = [(v_\nu^t)^* + (\ell_\mu^t)^*]^* \ .$$

From a general result of convex analysis [6, Theorem 16.4], it follows that

$$(f_\xi^t)^* = v_\nu^t \Box \ell_\mu^t \ ,$$

where

$$z \in \mathbb{R}^n \longmapsto [v_\nu^t \Box \ell_\mu^t](z) = \operatorname*{Inf}_{\theta \in \mathbb{R}^n} \{ v_\nu^t(z-\theta) + \ell_\mu^t(\theta) \}$$

stands for the infimal-convolution of v_ν^t and ℓ_μ^t. A straightforward calculation shows that

$$\begin{aligned} [v_\nu^t \Box \ell_\mu^t](z) &= \operatorname*{Inf}_{\theta \in \mathbb{R}^n} \left[\operatorname*{Inf}_{Ar = z-\theta} \{ \langle \gamma, r \rangle + \frac{1}{t} Q_\nu(r) \} + \operatorname*{Inf}_{Ws=\theta} \{ \langle \beta, s \rangle + \frac{1}{t} P_\mu(s) \} \right] \\ &= \operatorname*{Inf}_{Ar+Ws=z} \{ \langle \gamma, r \rangle + \langle \beta, s \rangle + \frac{1}{t} \Big[Q_\nu(r) + P_\mu(s) \Big] \} \\ &= m_\xi^t(z) \ . \end{aligned}$$

This completes the proof. □

As immediate consequence of Theorem 3.3, one sees that the upward convergence of $\{f_\xi^t\}_{t>0}$ toward f is equivalent to the downward convergence of $\{m_\xi^t\}_{t>0}$ toward the conjugate of f. This observation is consistent with the fact that the conjugate of f evaluated at z is precisely the infimal-value $m(z)$ of the mixed linear programming problem (21).

Table 3.1: Smoothing, penalization, and duality

form of LP program	standard	Karmarkar	mixed
minimization variables	$r \in \mathbb{R}^q_+$	$s \in \mathbb{R}^p_+, \langle \mathbf{1}, s\rangle = 1$	$\left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right] \in \mathbb{R}^{q+p}_+, \langle \left[\begin{smallmatrix} 0 \\ \mathbf{1} \end{smallmatrix}\right], \left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right]\rangle = 1$
penalty term	$Q_\nu(r)$	$P_\mu(s)$	$R_\xi(r,s) = Q_\nu(r) + P_\mu(s)$
inf-value of penalized program	v^t_ν	ℓ^t_μ	$m^t_\xi = v^t_\nu \square \ell^t_\mu$
conjugate of inf-value function	g^t_ν	h^t_μ	$f^t_\xi = g^t_\nu + h^t_\mu$

4. Second-order information

Following the pattern initiated in Section 2, one may ask now whether the Hessian matrix

$$\nabla^2 f^t_\xi(\bar{x}) = \left[\frac{\partial^2 f^t_\xi}{\partial x_k \partial x_\ell}(\bar{x}) \right]_{k,\ell=1,\cdots,n}$$

converges somewhere as the parameter t goes to ∞. Preliminary computations show that $\nabla^2 f^t_\xi(\bar{x})$ converges to the zero matrix if the function f happens to be affine. More precisely,

Proposition 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a polyhedral convex function given in the canonical form* (1). *For a point $\bar{x} \in \mathbb{R}^n$ at which f is finite, the following three conditions are equivalent*

(a) *f is affine on a neighborhood of $\bar{x}$;*
(b) *$I(\bar{x})$ is empty and $J(\bar{x})$ is a singleton ;*
(c) *for each $\xi \in \Xi$, $\lim_{t\to\infty} \nabla^2 f^t_\xi(\bar{x}) = 0$ (the zero matrix of order $n \times n$).*

Proof. The equivalence between (a) and (b) is evident. Condition (c) can be established after computing explicitly the Hessian matrix of f^t_ξ. This will be done in the proof of the next theorem. □

The result established above is, of course, not surprising. The limiting behavior of $\nabla^2 f^t_\xi(\bar{x})$ is more interesting when f behaves "genuinely" as a polyhedral function around $\bar{x}$, that is to say, when $I(\bar{x})$ is nonempty (i.e. at least one of the constraints defining K becomes active at $\bar{x}$), or $J(\bar{x})$ contains more than one index (i.e. at least two of the affine functions defining h achieve at $\bar{x}$ the value $h(\bar{x})$). In this genuine polyhedral setting, $\nabla^2 f^t_\xi(\bar{x})$ may not remain bounded as t goes to ∞. In other words, it may well happen that

$$\lim_{t\to\infty} \|\nabla^2 f^t_\xi(\bar{x})\| = \infty .$$

However, if one multiplies $\nabla^2 f^t_\xi(\bar{x})$ by a suitable scaling factor, then one does obtain convergence. More precisely, the limiting matrix

$$H_\xi(\bar{x}) = \lim_{t\to\infty} \frac{1}{t} \nabla^2 f^t_\xi(\bar{x})$$

does exist. What kind of information is contained in $H_\xi(\bar{x})$? Since our geometric intuition does not help us to clarify this matter, we rely on our analytic computations. The outcome of our analysis is a result that has an interesting probabilistic interpretation. Recall that for a (discrete) measure λ on $\mathbb{R}^n$, the $n \times n$ matrix

$$m_2[\lambda] := \int_{R^n} uu^T d\lambda(u)$$

is referred to as the second-order moment of λ. As usual, the symbol "T" denotes transposition. The notation

$$\mathrm{var}[\lambda] := \frac{1}{\lambda(\mathbb{R}^n)} \int_{\mathbb{R}^n} uu^T d\lambda(u) - [\int_{\mathbb{R}^n} ud\lambda(u)]\,[\int_{\mathbb{R}^n} ud\lambda(u)]^T$$

stands for the covariance matrix of λ.

Theorem 4.2. *Consider a polyhedral convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ given in the canonical form* (1). *Let $\bar{x} \in \mathbb{R}^n$ be a point at which f is finite. For each weight vector $\xi \in \Xi$, one has*

$$\lim_{t\to\infty} \frac{1}{t} \nabla^2 f_\xi^t(\bar{x}) = m_2[\nu^{\bar{x}}] + \mathrm{var}[\mu^{\bar{x}}] \,. \tag{23}$$

Here $\nu^{\bar{x}}$ is the discrete measure concentrated on $\{a^i : i \in I(\bar{x})\}$ given by

$$\nu^{\bar{x}}(\{a^i\}) = \nu_i \qquad \forall i \in I(\bar{x}) \,,$$

and $\mu^{\bar{x}}$ is the discrete measure concentrated on $\{w^j : j \in J(\bar{x})\}$ given by

$$\mu^{\bar{x}}(\{w^j\}) = \frac{\mu_j}{\sum_{k\in J(\bar{x})} \mu_k} \qquad \forall j \in J(\bar{x}) \,.$$

Proof. Consider an arbitrary $\xi \in \Xi$. We have seen already that

$$\frac{\partial f_\xi^t}{\partial x_k}(x) = \sum_{i=1}^{q} \nu_i e^{t[\langle a^i,x\rangle-\gamma^i]}(a^i)_k \;+\; \sum_{j=1}^{p} \lambda_j^t(x)(w^j)_k \qquad \forall x \in \mathbb{R}^n \,,$$

where $(u)_k$ denotes the k-th component of a vector $u \in \mathbb{R}^n$, and

$$\lambda_j^t(x) := \frac{\mu_j e^{t[\langle w^j,x\rangle-\beta^j]}}{\sum_{k=1}^{p} \mu_k e^{t[\langle w^k,x\rangle-\beta^k]}} \,.$$

Observe, incidentally, that

$$\sum_{j=1}^{p} \lambda_j^t(x) = 1 \,,\ \lambda_j^t(x) \geq 0 \qquad \forall j = 1, \cdots, p \,.$$

By differentiating once again, one arrives at the expression

$$\frac{\partial^2 f_\xi^t}{\partial x_k \partial x_\ell}(x) = t \sum_{i=1}^{q} \nu_i e^{t[\langle a^i,x\rangle-\gamma^i]}(a^i)_k\,(a^i)_\ell \;+\; \sum_{j=1}^{p} \frac{\partial \lambda_j^t}{\partial x_\ell}(x)(w^j)_k \,,$$

with

$$\frac{\partial \lambda_j^t}{\partial x_\ell}(x) = t\lambda_j^t(x)\left[(w^j)_\ell - \sum_{r=1}^p \lambda_r^t(x)(w^r)_\ell\right].$$

In other words,

$$\frac{1}{t}\,\frac{\partial^2 f_\xi^t}{\partial x_k \partial x_\ell}(x) = \sum_{i=1}^q \nu_i e^{t[\langle a^i, x\rangle - \gamma^i]}(a^i)_k\,(a^i)_\ell + \sum_{j=1}^p \lambda_j^t(x)(w^j)_k(w^j)_\ell$$
$$- \left[\sum_{j=1}^p \lambda_j^t(x)(w^j)_k\right]\left[\sum_{j=1}^p \lambda_j^t(x)(w^j)_\ell\right].$$

Hence, up to a scaling factor, the Hessian of f_ξ^t evaluated at $\bar{x}$ admits the following expression:

$$\frac{1}{t}\nabla^2 f_\xi^t(\bar{x}) = \sum_{i=1}^q \nu_i e^{t[\langle a^i, \bar{x}\rangle - \gamma^i]} a^i (a^i)^T + \sum_{j=1}^p \lambda_j^t(\bar{x}) w^j (w^j)^T$$
$$- \left[\sum_{j=1}^p \lambda_j^t(\bar{x}) w^j\right]\left[\sum_{j=1}^p \lambda_j^t(\bar{x}) w^j\right]^T.$$

This shows that

$$\lim_{t\to\infty} \frac{1}{t}\nabla^2 f_\xi^t(\bar{x}) = \sum_{i\in I(\bar{x})} \nu_i a^i (a^i)^T + \sum_{j\in J(\bar{x})} \tilde{\mu}_j w^j (w^j)^T - \left[\sum_{j\in J(\bar{x})} \tilde{\mu}_j w^j\right]\left[\sum_{j\in J(\bar{x})} \tilde{\mu}_j w^j\right]^T,$$

with

$$\tilde{\mu}_j = \frac{\mu_j}{\sum_{k\in J(\bar{x})} \mu_k} \qquad \forall j \in J(\bar{x}).$$

This is, of course, the desired conclusion. □

5. Differentiation with respect to the smoothing parameter

Next on our agenda is the analysis of $f_\xi^t(x)$ as a function of the parameter t. For the sake of notational convenience, we drop again the reference to the weight vector ξ and write simply

$$F(t, x) := f^t(x) \qquad \forall (t, x) \in \,]0, \infty[\, \times \mathbb{R}^n. \tag{24}$$

The function $F : \,]0, \infty[\, \times \mathbb{R}^n \to \mathbb{R}$ is infinitely often differentiable. The first and second-order partial derivatives of F with respect to the "state" vector x have been studied at lenght in Sections 2 and 4, respectively. From a purely formal point of view, the smoothing or approximation parameter t can be interpreted as a time variable.

A direct differentiation of F with respect to t yields

$$\frac{\partial F}{\partial t}(t,x) = -\frac{1}{t^2}\left[\log\Big(\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,x\rangle-\beta^j]}\Big) + \sum_{i=1}^{q} \nu_i\big(e^{t[\langle a^i,x\rangle-\gamma^i]}-1\big)\right] \tag{25}$$

$$+\ \frac{1}{t}\left[\frac{1}{\sum_{j=1}^{p}\mu_j e^{t[\langle w^j,x\rangle-\beta^j]}}\ \sum_{j=1}^{p}\mu_j[\langle w^j,x\rangle-\beta^j]e^{t[\langle w^j,x\rangle-\beta^j]} + \sum_{i=1}^{q}\nu_i\,[\langle a^i,x\rangle-\gamma^i]\,e^{t[\langle a^i,x\rangle-\gamma^i]}\right].$$

Some terms in the above equality are easily recognizable, some others are not. Anyway, observe that

$$t\,\frac{\partial F}{\partial t}\,(t,x) = -F(t,x) + \Gamma^t(x)\,, \tag{26}$$

with

$$\Gamma^t(x) = \frac{1}{\sum_{j=1}^{p}\mu_j e^{t[\langle w^j,x\rangle-\beta^j]}}\ \sum_{j=1}^{p}\mu_j[\langle w^j,x\rangle-\beta^j]e^{t[\langle w^j,x\rangle-\beta^j]} + \sum_{i=1}^{q}\nu_i\,[\langle a^i,x\rangle-\gamma^i]\,e^{t[\langle a^i,x\rangle-\gamma^i]}\,.$$

Theorem 5.1. *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be the polyhedral convex function with canonical representation given by* (1)*. Then,*

(a) *the family $\{\Gamma^t\}_{t>0}$ converges pointwise to f as t goes to ∞, i.e. f admits the approximation formula*

$$f(x) = \lim_{t\to\infty}\Gamma^t(x) \qquad \forall x\in\mathbb{R}^n\,; \tag{27}$$

(b) $\dfrac{\partial F}{\partial t}$ *and* $t\,\dfrac{\partial F}{\partial t}$ *converge to the indicator function of K, i.e.*

$$\lim_{t\to\infty}\frac{\partial F}{\partial t}\,(t,x) = \lim_{t\to\infty} t\,\frac{\partial F}{\partial t}\,(t,x) = \psi_K(x) := \begin{cases} 0 & \text{if } x\in K\,,\\ +\infty & \text{if } x\in K\,.\end{cases} \tag{28}$$

(c) $t^2\,\dfrac{\partial F}{\partial t}$ *behaves asymptotically according with the rule*

$$\lim_{t\to\infty} t^2\,\frac{\partial F}{\partial t}\,(t,x) = \begin{cases} \theta(x) & \text{if } x\in K\,,\\ +\infty & \text{if } x\notin K\,.\end{cases} \tag{29}$$

with

$$\theta(x) := \sum_{i\notin I(x)}\nu_i - \log\Big[\sum_{j\in J(x)}\mu_j\Big]\,.$$

Proof. To prove the part (a), it suffices to write Γ^t in the form

$$\Gamma^t(x) = \frac{1}{\sum_{j=1}^{p}\mu_j e^{t[\langle w^j,x\rangle-\beta^j-h(x)]}}\sum_{j=1}^{p}\ \mu_j\,[\langle w^j,x\rangle-\beta^j]\,e^{t[\langle w^j,x\rangle-\beta^j-h(x)]}$$

$$+\sum_{i=1}^{q}\nu_i\,[\langle a^i,x\rangle-\gamma^i]\,e^{t[\langle a^i,x\rangle-\gamma^i]}\,. \tag{30}$$

The first term on the right-hand side of (30) clearly converges to

$$\frac{1}{\sum\limits_{j\in J(x)} \mu_j} \sum_{j\in J(x)} \mu_j[\langle w^j, x\rangle - \beta^j] = h(x) \ .$$

The second term on the right-hand side of (30) converges to

$$\psi_K(x) = \begin{cases} 0 & \text{if } \langle a^i, x\rangle - \gamma^i \le 0 \qquad \forall i \in \{1, \cdots, q\} \ , \\ +\infty & \text{otherwise} \ . \end{cases}$$

To summarize,

$$\lim_{t\to\infty} \Gamma^t(x) = h(x) + \psi_K(x) = f(x) \qquad \forall x \in \mathbb{R}^n \ .$$

This result together with the equation (26) yield

$$\lim_{t\to\infty} t\, \frac{\partial F}{\partial t}\,(t,x) = 0 \qquad \forall x \in K \ ,$$

which, in turns, implies that

$$\lim_{t\to\infty} \frac{\partial F}{\partial t}\,(t,x) = 0 \qquad \forall x \in K \ .$$

The latter equality can be derived also from the fact that

$$\lim_{t\to\infty} F(t,x) = h(x) \in \mathbb{R} \qquad \forall x \in K \ .$$

The case $x \notin K$ must be handled with more care, because passing to the limit on the right-hand side of (26) leads to the undetermined expression $-\infty\ +\infty$. So, let us rearrange (25) to get

$$\frac{\partial F}{\partial t}\,(t,x) = \frac{1}{t}\,\Big\{ -h^t_\mu(x) + \frac{1}{\sum\limits_{j=1}^{p} \mu_j e^{t[\langle w^j,x\rangle - \beta^j - h(x)]}} \sum_{j=1}^{p} \mu_j[\langle w^j, x\rangle - \beta^j] e^{t[\langle w^j,x\rangle - \beta^j - h(x)]} \Big\}$$
$$+ \frac{1}{t^2} \sum_{i=1}^{q} \nu_i + \frac{1}{t} \sum_{i=1}^{p} \nu_i\, [\langle a^i, x\rangle - \gamma^i - \frac{1}{t}]\; e^{t[\langle a^i,x\rangle - \gamma^i]} \ .$$

The term between brackets goes 0, so does the next term. But, the last term goes to ∞ because

$$\langle a^i, x\rangle - \gamma^i > 0 \quad \text{for at least one index } i \ .$$

This shows that

$$\lim_{t\to\infty} \frac{\partial F}{\partial t}\,(t,x) = \infty \qquad \forall x \notin K \ ,$$

which, in turns, implies that

$$\lim_{t\to\infty} t\, \frac{\partial F}{\partial t}\,(t,x) = \lim_{t\to\infty} t^2\, \frac{\partial F}{\partial t}\,(t,x) = \infty \qquad \forall x \notin K \ ,$$

To complete the proof it remains to show that

$$\lim_{t\to\infty} t^2 \, \frac{\partial F}{\partial t}\,(t,x) = \theta(x) \qquad \forall x \in K \ ,$$

but this equality is left as exercise. It follows from a careful examination of (25). □

An important comment on Theorem 5.1 is in order: formula (27) tells us that the polyhedral convex function f can be seen as the pointwise limit of the family $\{\Gamma^t\}_{t>0}$, with each $\Gamma^t : \mathbb{R}^n \to \mathbb{R}$ being infinitely often differentiable. So, we have obtained yet another C^∞-approximation method for polyhedral convex functions. The link between f^t and Γ^t is expressed by the formula

$$\Gamma^t(x) = \frac{\partial}{\partial t}\,[t f^t(x)] \ , \tag{31}$$

which is another way of writing the equation (26). In contrast with f^t, the function Γ^t is not convex in general. Let us illustrate this point with a simple example:

Example 5.2. Let f be the absolute value function on $\mathbb{R}$, i.e.

$$f(x) = |x| = \mathrm{Max}\{x, -x\} \qquad \forall x \in \mathbb{R} \ .$$

In this case

$$\begin{aligned} f^t(x) &= \frac{1}{t}\, \log(\mu_1 e^{tx} \ + \ \mu_2 e^{-tx}) \ , \\ \Gamma^t(x) &= \left[\frac{\mu_1 e^{tx}}{\mu_1 e^{tx} + \mu_2 e^{-tx}}\right] x \ + \ \left[\frac{\mu_2 e^{-tx}}{\mu_1 e^{tx} + \mu_2 e^{-tx}}\right] (-x) \ . \end{aligned}$$

Regardless of the coefficients $\mu_1 > 0$ and $\mu_2 > 0$, one obtains

$$|x| = \lim_{t\to\infty} f^t(x) = \lim_{t\to\infty} \Gamma^t(x) \qquad \forall x \in \mathbb{R} \ .$$

Both functions f^t and Γ^t are of class C^∞, but only f^t is convex. The lack of convexity of Γ^t can be checked by using the second-derivative test.

6. Euler-type PDE with source term

For the sake of the exposition, consider momentarily the case in which

$$\beta^j = 0 \quad \forall j = 1, \cdots, p \qquad \text{and} \qquad \gamma^i = 0 \quad \forall i = 1, \cdots, q \ . \tag{32}$$

In this particular setting, $F :]0, \infty[\times \mathbb{R}^n \to \mathbb{R}$ can be written in the form

$$F(t,x) = u(\frac{1}{t}\,, x) \ ,$$

with

$$(\tau, x) \in]0,\infty[\times \mathbb{R}^n \longmapsto u(\tau,x) = \tau \sum_{i=1}^{q} \ \nu_i(e^{\frac{1}{\tau}\langle a^i, x\rangle} - 1) \ + \ \tau \log\{\sum_{j=1}^{p} \ \mu_j \ e^{\frac{1}{\tau}\langle w^j, x\rangle}\}$$

being a homogeneous function of degree 1. Due to its homogeneity, u satisfies Euler's equation

$$\langle \nabla_x u(\tau,x), x\rangle + \tau \frac{\partial u}{\partial \tau}(\tau,x) - u(\tau,x) = 0 \,,$$

and therefore F satisfies

$$\langle \nabla_x F(t,x), x\rangle - t \frac{\partial F}{\partial t}(t,x) - F(t,x) = 0 \,. \tag{33}$$

The latter equation is referred to as an Euler-type PDE because it is obtained from the former one by a simple change of variables.

If one drops the simplificatory assumption (32), then there is a "source" term that shows up on the right-hand side of (33), namely

$$S(t,x) = \frac{1}{\displaystyle\sum_{j=1}^{p} \mu_j e^{t[\langle w^j,x\rangle - \beta^j]}} \sum_{j=1}^{p} \mu_j \beta^j e^{t[\langle w^j,x\rangle - \beta^j]} + \sum_{i=1}^{q} \nu_i \gamma^i e^{t[\langle a^i,x\rangle - \gamma^i]} \,. \tag{34}$$

Proposition 6.1. *Consider a polyhedral convex function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *given in the canonical form* (1). *Then,* $(t,x) \in \,]0,\infty[\, \times \mathbb{R}^n \longmapsto f^t(x)$ *is a solution to the PDE*

$$\langle \nabla_x F(t,x), x\rangle - t\, \frac{\partial F}{\partial t}\, (t,x) - F(t,x) = S(t,x) \,, \tag{35}$$

with "boundary" condition

$$\lim_{t\to\infty} F(t,x) = f(x) \,.$$

Proof. The Euler-type PDE with source term S is a consequence of (26) and the fact that

$$\Gamma^t(x) = \langle \nabla_x F(t,x), x\rangle - S(t,x) \,,$$

the latter equality being obtained with the help of (13). □

7. Conclusions

As seen in this paper, a substantial amount of information on the behavior of the polyhedral convex function (1) can be recovered directly from its generating function (10). Rather than review the full ramifications and applications of our work, we prefer to dress a short list of significant results:

(a) Selection principle for subgradients: it has been shown that the subdifferential $\partial f(x)$ of the polyhedral convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ can be obtained by computing classical gradients $\nabla f^t_\xi(x)$ of smooth approximations of f. Each subgradient of f corresponds to a particular choice of the weight vector ξ.

(b) Selection principle for optimal solutions: this principle is a by-product of the previous one. A minimization problem with polyhedral convex data can always be cast in the form

$$\text{Minimize}\{m(y) : y \in \mathbb{R}^n\} \,, \tag{36}$$

with m being the conjugate function of (1). The solution set of (36) is known to be

$$\arg\min m = \partial m^*(0) = \partial f(0) \quad \text{(cf. [6, Theorem 27.1])} .$$

As a consequence, each minimum $y \in \mathbb{R}^n$ of the function m can be expressed as

$$y = \lim_{t\to\infty} \nabla f_\xi^t(0) .$$

Observe that $y_\xi^t := \nabla f_\xi^t(0)$ corresponds to the unique solution to the approximated version

$$\text{Minimize}\{m_\xi^t(z) : z \in \mathbb{R}^n\} \tag{37}$$

of the original problem (36). In short,

$$\arg\min m = \{\lim_{t\to\infty} \arg\min m_\xi^t : \xi \in \Xi(0)\} .$$

(c) C^∞-approximation and duality: it has been shown that the C^∞-approximation f_ξ^t and the infimal-value function m_ξ^t are mutually conjugate. The standard machinery of Legendre-Fenchel conjugation theory can be put now into practice in order to exploit this result.

(d) Second-order statistical information: the generating function (10) allows us to recover also second-order information on the sample

$$\{a^i : i \in I(\bar{x})\} , \ \{w^j : j \in J(\bar{x})\} .$$

The details are discussed in Section 4.

(e) C^∞-approximation via Gibbs functionals: let $\mathcal{P}_\mu : \mathbb{R}^p \to \mathbb{R}^p$ and $\mathcal{Q}_\nu : \mathbb{R}^q \to \mathbb{R}^q$ be defined respectively by

$$\mathcal{P}_\mu(s) = \left(\frac{\mu_1 e^{s_1}}{\sum_{j=1}^p \mu_j e^{s_j}} , \cdots , \frac{\mu_p e^{s_p}}{\sum_{j=1}^p \mu_j e^{s_j}} \right)^T , \quad \mathcal{Q}_\nu(r) = (\nu_1 e^{r_1} , \cdots , \nu_q e^{r_q})^T .$$

Some authors refer $\mathcal{P}_\mu$ as the Gibbs functional associated with μ, but $\mathcal{Q}_\nu$ does not seem to have a proper terminology. Anyway, both functionals have played an important role in our work. Observe that $\mathcal{P}_\mu$ takes values in the elementary simplex of $\mathbb{R}^p$ and $\mathcal{Q}_\nu$ takes values in the positive orthant of $\mathbb{R}^q$. As a consequence,

$$x \in \mathbb{R}^n \longmapsto \Gamma^t(x) = \langle \mathcal{P}_\mu(t(W^T x - \beta)) , \ W^T x - \beta \rangle \ + \ \langle \mathcal{Q}_\nu(t(A^T x - \gamma)) , \ A^T x - \gamma \rangle$$

corresponds to a convex combination of the affine mappings $\{\langle w^j, \cdot\rangle - \beta^j\}_{j=1}^p$ and a conic combination of the affine mappings $\{\langle a^i, \cdot\rangle - \gamma^i\}_{i=1}^q$. Recall that Γ^t was shown to be a C^∞-approximation of f.

(f) PDE approach: the expression $f_\xi^t(x)$ is (infinitely often) differentiable with respect to the couple (t, x). It has been shown that $(t, x) \in \,]0, \infty[\, \times \mathbb{R}^n \longmapsto f_\xi^t(x)$ is a solution to the first-order PDE (35). A natural question that arises in this context is whether it is possible to obtain other approximation schemes by changing the “source” term in (35). To interact with the reader, we leave open for discussion this challenging question.

References

[1] R. Cominetti, J. San Martin: Asymptotic analysis of the exponential penalty trajectory in linear programming, Math. Programming 67 (1994) 169–187.

[2] S. C. Fang, X. S. Li: On the entropic regularization method for solving min-max problems with applications, Math. Methods of Operations Research 46 (1997) 119–130.

[3] J. B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms, Vol. II, Grundlehren der Mathematischen Wissenschaften 306, Springer-Verlag, Berlin, 1993.

[4] S. S. Kutateladze: Convex operators, Russian Math. Surveys 34 (1979) 181–214.

[5] J. J. Moreau: Fonctionelles Convexes, Lecture Notes, Séminaire "Equations aux Dérivées Partielles", Collège de France, Paris, 1966.

[6] R. T. Rockafellar: Convex Analysis, Princeton Univ. Press, Princeton, N. J., 1970.

[7] A. Seeger: Smoothing a polyhedral convex function via cumulant transformation and homogenization, Annales Polinici Mathematici 67 (1997) 259–268.

[8] R. L. Sheu, S. Y. Wu: Combined entropic regularization and path-following method for solving finite convex min-max problems subject to infinitely many linear constraints, J. Optim. Theory and Appl. 101 (1999) 167–190.