ANNALES
POLONICI MATHEMATICI
LXVII.3 (1997)

# Smoothing a polyhedral convex function via cumulant transformation and homogenization

by Alberto Seeger (Avignon)

**Abstract.** Given a polyhedral convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, it is always possible to construct a family $\{g_t\}_{t>0}$ which converges pointwise to $g$ and such that each $g_t : \mathbb{R}^n \to \mathbb{R}$ is convex and infinitely often differentiable. The construction of such a family $\{g_t\}_{t>0}$ involves the concept of cumulant transformation and a standard homogenization procedure.

**1. Introduction.** A broad class of nonsmooth optimization problems can be written in the composite form

$$(P) \qquad \text{Minimize } \{g(M(\xi)) : \xi \in \Xi\},$$

where $M$ is a mapping from some normed space $\Xi$ to the Euclidean space $\mathbb{R}^n$, and $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a *polyhedral convex function*, i.e.

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

is a polyhedral convex set [11, p. 172]. As a general rule $M$ is a smooth mapping (say, of class $C^k$, for some $k \ge 1$), but the composite function $g \circ M$ is nonsmooth. This fact leads us to consider an approximate version

$$(P)_t \qquad \text{Minimize } \{g_t(M(\xi)) : \xi \in \Xi\}$$

for the original problem $(P)$. A fundamental question which is addressed in this note is thus:

$$(1.1) \qquad \begin{cases} \text{How to construct a family } \{g_t\}_{t>0} \text{ of} \\ \text{smooth convex functions } g_t : \mathbb{R}^n \to \mathbb{R} \text{ such that} \\ \lim\limits_{t\to\infty} g_t(x) = g(x) \text{ for all } x \in \mathbb{R}^n\ ? \end{cases}$$

The above mentioned question has an interest which goes far beyond the

1991 *Mathematics Subject Classification*: Primary 41A30; Secondary 52B70, 60E10.
*Key words and phrases*: polyhedral convex function, smooth approximation, Laplace transformation, cumulant transformation, homogenization, recession function.

context of the composite optimization problem $(P)$. This note will concentrate on this question and will discuss some related issues.

As a first attempt to answer (1.1), one may consider the classical Moreau–Yosida approximation technique:

$$g_t(x) = \inf_{u\in\mathbb{R}^n} \left\{ g(x-u) + \frac{t}{2}\|u\|^2 \right\}.$$

The disadvantage of such an approach is twofold: first of all, the evaluation of $g_t(x)$ is not straightforward since it requires solving a minimization problem. Secondly, the convex function $g_t$ is of class $C^1$, but its degree of smoothness is not higher than one (unless one imposes additional assumptions on $g$; cf. Lemaréchal and Sagastizabal [7]). The same remark applies to the rolling ball approximation technique [12]:

$$g_t(x) = \inf_{\|u\|\le t^{-1}} \{g(x-u) - [t^{-2} - \|u\|^2]^{1/2}\}.$$

The approach suggested in this note is completely different: it uses homogenization technique applied to the Laplace transform and to the cumulant transform of some discrete measures associated with the function $g$.

**2. Smoothing a polyhedral supporting function.** To start with, consider the case in which $g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is the *supporting function* of a nonempty polyhedral convex set $\Omega \subset \mathbb{R}^n$, i.e.

$$g(x) = \Psi_\Omega^*(x) := \sup_{w\in\Omega} \langle w, x\rangle. \tag{2.1}$$

One may think of $g$ as the recession function ([11, p. 66])

$$g(x) = [\operatorname{rec} f](x) := \lim_{t\to\infty} f(tx)/t \tag{2.2}$$

of some convex lower-semicontinuous function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ that is finite at $0 \in \mathbb{R}^n$. As observed by Ben-Tal and Teboulle [2], the usefulness of the approximation scheme (2.2) lies in the fact that frequently $f$ is a smooth function, in which case the convex function

$$\mathbb{R}^n \ni x \mapsto g_t(x) := f(tx)/t$$

is also smooth. Ben-Tal and Teboulle [2] provided the examples

$$g(x) = \max\{x_1,\ldots,x_n\}, \qquad f(x) = \log\Big[\sum_{j=1}^n e^{x_j}\Big] \tag{2.3}$$

and

$$g(x) = \sum_{j=1}^n |x_j|, \qquad f(x) = \sum_{j=1}^n [1+x_j^2]^{1/2} \tag{2.4}$$

to justify their observation, but they did not give a method for finding a smooth function $f$ in other cases. Examples (2.3) and (2.4) were inspired by particular approximation techniques suggested by Bertsekas [3] and El-Attar *et al.* [5], respectively.

The aim of this section is to provide the reader with a simple and elegant method for constructing a smooth function $f$ in the case in which $g$ is an arbitrary polyhedral supporting function. Our approximation mechanism relies on the following basic assumption:

(2.5) *$\Omega \subset \mathbb{R}^n$ is a polyhedral convex set which admits at least one extreme point.*

As is well known, such a set $\Omega$ can be represented in the form of a Minkowski sum:

$$\Omega = \mathrm{co}[\mathrm{extr}\,\Omega] + \mathrm{rec}\,\Omega. \tag{2.6}$$

Here

$$\mathrm{extr}\,\Omega = \{w^1, \ldots, w^k\} \tag{2.7}$$

denotes the set of extreme points of $\Omega$, and $\mathrm{rec}\,\Omega$ refers to the recession cone of $\Omega$ ([11, p. 61]). Since $\Omega$ is a polyhedral convex set, $\mathrm{rec}\,\Omega$ can be represented in terms of a set $\{a^1, \ldots, a^m\} \subset \mathbb{R}^n$ of generating directions:

$$\mathrm{rec}\,\Omega = \Big\{ \sum_{i=1}^{m} \lambda_i a^i : \lambda_i \ge 0 \ \forall i = 1, \ldots, m \Big\}. \tag{2.8}$$

Without loss of generality one may suppose that the set $\{a^1, \ldots, a^m\}$ is minimal in the sense that none of these directions can be expressed as a nonnegative linear combination of the others. It is not difficult to show that $g = \Psi_\Omega^*$ takes the form

$$g(x) = \begin{cases} \max\{\langle w^1, x\rangle, \ldots, \langle w^k, x\rangle\} & \text{if } x \in K, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.9}$$

where

$$K = \{v \in \mathbb{R}^n : \langle a^i, v\rangle \le 0 \ \forall i = 1, \ldots, m\}.$$

Now we are ready to state:

THEOREM 2.1. *Let $\Omega$ be as in* (2.5) *and let $g$ be the supporting function of $\Omega$. Then there exists a convex function $f : \mathbb{R}^n \to \mathbb{R}$ of class $C^\infty$ such that*

$$\lim_{t\to\infty} f(tx)/t = g(x) \quad \textit{for all } x \in \mathbb{R}^n. \tag{2.10}$$

*An example of such a function is given by*

$$f(x) = \log \Big[ \int_{\mathbb{R}^n} e^{\langle w,x\rangle}\, d\mu(w)\Big] + \int_{\mathbb{R}^n} e^{\langle w,x\rangle}\, d\nu(w), \tag{2.11}$$

*where $\mu$ is any discrete measure concentrated on* $\operatorname{extr}\Omega$, *and $\nu$ is any discrete measure concentrated on a minimal set of generating directions for* $\operatorname{rec}\Omega$. *If the information* (2.7)–(2.8) *is available*, *then one has the more explicit example*

$$f(x) = \log\Big[\sum_{j=1}^{k} e^{\langle w^j, x\rangle}\Big] + \sum_{i=1}^{m} e^{\langle a^i, x\rangle}. \tag{2.12}$$

P r o o f. The function $f$ in (2.11) involves the Laplace transform

$$\mathbb{R}^n \ni x \mapsto L_\nu(x) := \int_{\mathbb{R}^n} e^{\langle w, x\rangle}\, d\nu(w)$$

of the measure $\nu$, and the cumulant transform

$$\mathbb{R}^n \ni x \mapsto K_\mu(x) := \log L_\mu(x)$$

of the measure $\mu$. Since $\mu$ and $\nu$ are discrete, both transforms are finite-valued. By invoking some classical results (cf. [1, Theorem 4.1], [6, Theorem 7.5.1]), one can show that $L_\nu$ and $K_\mu$ are $C^\infty$ convex functions. It just remains to prove the convergence property (2.10). That $\mu$ is concentrated on $\operatorname{extr}\Omega$ means simply

$$\mu(\{w\}) > 0 \quad \text{iff} \quad w \in \operatorname{extr}\Omega.$$

Denote by $\mu_j = \mu(\{w^j\})$ the mass of the extreme point $w^j$. Then

$$K_\mu(x) = \log\Big[\sum_{j=1}^{k} \mu_j e^{\langle w^j, x\rangle}\Big]$$

and

$$\lim_{t\to\infty} K_\mu(tx)/t = \max\{\langle w^1, x\rangle, \ldots, \langle w^k, x\rangle\},$$

regardless of the values of the $\mu_j$'s. Similarly, denote by $\nu_i = \nu(\{a^i\})$ the mass of the generating direction $a^i$. In this case

$$L_\nu(x) = \sum_{i=1}^{m} \nu_i e^{\langle a^i, x\rangle}$$

and

$$\lim_{t\to\infty} L_\nu(tx)/t = \begin{cases} 0 & \text{if } \langle a^i, x\rangle \le 0\ \forall i = 1,\ldots,m,\\ +\infty & \text{otherwise,}\end{cases}$$

regardless of the values of the $\nu_i$'s. This completes the proof of (2.10). Finally, observe that (2.12) corresponds to the particular case in which the masses of $\mu$ and $\nu$ are uniformly distributed.

R e m a r k 2.1. Instead of (2.10), one can write the equality

$$\lim_{t\to\infty} \frac{f(tx) - f(0)}{t} = g(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{2.13}$$

The advantage of the approximation scheme (2.13) over (2.10) is that the quotient $[f(tx) - f(0)]/t$ converges monotonically upward to $g(x)$ as $t$ goes to $\infty$. Of course, one can always normalize $f$ so that $f(0)=0$. It suffices to subtract the constant $\nu(\mathbb{R}^n) + \log \mu(\mathbb{R}^n)$ from the expression appearing on the right-hand side of (2.11).

Remark 2.2. The measure $\mu$ used in (2.11) can be concentrated on a set which is larger than $\operatorname{extr} \Omega$, but it cannot assign a positive mass to a point which is outside the polytope $\operatorname{co}(\operatorname{extr} \Omega)$. Similarly, $\nu$ can be concentrated on a set which is larger than a minimal set of generating directions for $\operatorname{rec} \Omega$. However, $\nu$ should not assign a positive mass to a direction which is not in $\operatorname{rec} \Omega$.

Remark 2.3. The function $f$ given by (2.11) can also be used to approximate

$$\mathbb{R}^n \ni x \mapsto \inf_{w \in \Omega} \langle w, x \rangle.$$

Indeed,

$$\begin{aligned}\lim_{t \to -\infty} f(tx)/t &= \lim_{t \to \infty} f(-tx)/(-t) \\ &= -\lim_{t \to \infty} f(t(-x))/t = -\sup_{w \in \Omega} \langle w, -x \rangle = \inf_{w \in \Omega} \langle w, x \rangle.\end{aligned}$$

Of course, for $t < 0$, the function $x \mapsto f(tx)/t$ is concave.

Theorem 2.1 can be illustrated with an example.

Example 2.1. Consider the function $g : \mathbb{R}^8 \to \mathbb{R}$ given by

$$g(x) = \max\{x_1, x_2, x_3\} + |x_4| + |x_5| + \max\{0, x_6\} + \max\{|x_7|, |x_8|\}.$$

The first term corresponds to the supporting function of the set $\{u \in \mathbb{R}^3_+ : u_1 + u_2 + u_3 = 1\}$, whose extreme points are the canonical vectors $(1,0,0), (0,1,0)$, and $(0,0,1)$. Thus, $\max\{x_1, x_2, x_3\}$ can be approximated with the help of $\log(e^{x_1} + e^{x_2} + e^{x_3})$. The absolute value function $|\cdot|$ corresponds to the supporting function of the interval $[-1,1]$, whose extreme points are $-1$ and 1. Thus, $|x_4|$ and $|x_5|$ can be approximated by using $\log(\cosh x_4)$ and $\log(\cosh x_5)$, respectively. Similarly, $\max\{0, \cdot\}$ is the supporting function of the interval $[0,1]$, and therefore it can be approximated by using $\log(1+e^{(\cdot)})$. The last term of $g$ corresponds to the supporting function of a set whose extreme points are $(1,0), (-1,0), (0,1)$, and $(0,-1)$. Thus, $\max\{|x_7|, |x_8|\}$ can be approximated with the help of $\log(\cosh x_7 + \cosh x_8)$. Summarizing,

$$\begin{aligned}f(x) = {} & \log(e^{x_1} + e^{x_2} + e^{x_3}) + \log(\cosh x_4) + \log(\cosh x_5) \\ & + \log(1 + e^{x_6}) + \log(\cosh x_7 + \cosh x_8).\end{aligned}$$

Of course, $g$ can be regarded as the supporting function of some polytope in $\mathbb{R}^8$. In this case, however, the identification of the extreme points is a more cumbersome task.

**3. Smoothing a polyhedral convex function.** The approximation technique developed in Section 2 can be extended to the case in which $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is an arbitrary polyhedral convex function. It suffices to use the representation

$$g(x) = \Psi^*_{\operatorname{epi} g^*}(x, -1) = \sup_{(w,\beta)\in\operatorname{epi} g^*} \{\langle w, x\rangle - \beta\}, \tag{3.1}$$

where $g^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ stands for the Legendre–Fenchel conjugate of $g$. Since $g$ is a polyhedral convex function, it follows that the epigraph of $g^*$ is a polyhedral convex set ([11, Theorem 19.2]).

THEOREM 3.1. *Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a polyhedral convex function such that* $\operatorname{epi} g^*$ *has at least one extreme point. Then there exists a convex function $\mathcal{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ which is of class $C^\infty$ and such that*

$$\lim_{t\to\infty} \mathcal{F}(tx, -t)/t = g(x) \quad \textit{for all } x \in \mathbb{R}^n. \tag{3.2}$$

*An example of such a function is*

$$\mathcal{F}(x, \alpha) = \log\Big[\int_{\mathbb{R}^n\times\mathbb{R}} e^{\langle w,x\rangle+\beta\alpha}\, d\mu(w, \beta)\Big] + \int_{\mathbb{R}^n\times\mathbb{R}} e^{\langle w,x\rangle+\beta\alpha}\, d\nu(w, \beta), \tag{3.3}$$

*where $\mu$ is any discrete measure concentrated on* $\operatorname{extr}(\operatorname{epi} g^*)$, *and $\nu$ is any discrete measure concentrated on a minimal set of generating directions for* $\operatorname{rec}(\operatorname{epi} g^*)$.

Proof. Observe that the function $\mathcal{F}$ is given simply by

$$\mathcal{F}(x, \alpha) = K_\mu(x, \alpha) + L_\nu(x, \alpha).$$

According to Theorem 2.1, $\mathcal{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a convex function of class $C^\infty$ such that

$$\lim_{t\to\infty} \mathcal{F}(t(x, \alpha))/t = \Psi^*_{\operatorname{epi} g^*}(x, \alpha).$$

This and the representation formula (3.1) yield the convergence result (3.2). ■

It should be clear that, for each $t > 0$, the function

$$\mathbb{R}^n \ni x \mapsto g_t(x) = \mathcal{F}(tx, -t)/t$$

is convex and of class $C^\infty$. Also, $g_t(x)$ converges toward $g(x)$ as $t$ goes to $\infty$. Thus, Theorem 3.1 answers completely the question stated in (1.1). Of course, if one seeks a more explicit formula for the function $\mathcal{F}$, then more information is needed regarding the structure of $g$, namely one needs to know the extreme points $\{(w^1,\beta^1),\ldots,(w^k,\beta^k)\}$ of $\operatorname{epi} g^*$ and a minimal set $\{(a^1,\gamma^1),\ldots,(a^m,\gamma^m)\}$ of generating directions for $\operatorname{rec}(\operatorname{epi} g^*)$. This amounts to representing $g$ in the following "canonical" form:

$$g(x)=\begin{cases}\max\{\langle w^1,x\rangle-\beta^1,\ldots,\langle w^k,x\rangle-\beta^k\} & \text{if } x\in K,\\ +\infty & \text{otherwise,}\end{cases}$$

with

$$K=\{v\in\mathbb{R}^n:\langle a^i,v\rangle\le\gamma^i\ \forall i=1,\ldots,m\}.$$

If this representation is available, then one can take $\mathcal{F}$ simply as

$$\mathcal{F}(x,\alpha)=\log\Big[\sum_{j=1}^{k}e^{\langle w^j,x\rangle+\beta^j\alpha}\Big]+\sum_{i=1}^{m}e^{\langle a^i,x\rangle+\gamma^i\alpha}.$$

Example 3.1. If one wishes to approximate

$$\mathbb{R}^2\ni x\mapsto g(x)=\begin{cases}\max\{6x_1-x_2+4,x_1+x_2-2\} & \text{if } x_1\ge 3,\\ +\infty & \text{otherwise,}\end{cases}$$

then it suffices to take

$$\mathcal{F}(x,\alpha)=\log[e^{6x_1-x_2-4\alpha}+e^{x_1+x_2+2\alpha}]+e^{-x_1-3\alpha}.$$

**4. Application: smoothing a spectrally defined matrix function.** Consider the case of a function $\Phi:S_n\to\mathbb{R}\cup\{+\infty\}$ defined over the space $S_n$ of $n\times n$ real symmetric matrices. Such a function $\Phi$ is said to be *spectral* (or spectrally defined) if there is a symmetric function $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ such that

$$\Phi(A)=g(\lambda(A))\quad\text{for all } A\in S_n, \tag{4.1}$$

where $\lambda(A)=(\lambda_1(A),\ldots,\lambda_n(A))^\top$ is the vector of eigenvalues of $A$ in nonincreasing order. The symmetric function $g$ is necessarily unique. In fact, it is given by

$$g(x)=\Phi(\operatorname{diag}x)\quad\text{for all } x\in\mathbb{R}^n,$$

where $\operatorname{diag} x$ stands for the diagonal matrix whose entries on the diagonal are the components of $x$. For a detailed account on spectral functions, see, for instance, [13] (also [8], [9]). Examples of spectral functions include:

$\Phi(A)=\lambda_1(A)=$ largest eigenvalue of $A$;

$\Phi(A)=\lambda_1(A)+\ldots+\lambda_p(A)=$ sum of the $p$ largest eigenvalues of $A$;

$\Phi(A)=\lambda_1(A)-\lambda_n(A)=$ width of the spectrum of $A$;

$\Phi(A)=\max\{\lambda_1(A),-\lambda_n(A)\}=$ spectral radius of $A$.

In connection with these examples, two comments deserve to be made: first, none of the above functions is differentiable; and, second, all the above functions can be written in the form

$$\Phi(A) = \Psi_\Omega^*(\lambda(A)),$$

with $\Omega \subset \mathbb{R}^n$ being a symmetric convex polytope. The symmetry property means that

$$w \in \Omega \Rightarrow \Pi w \in \Omega \text{ for any } n \times n \text{ permutation matrix } \Pi.$$

These facts lead us to establishing the following approximation result.

THEOREM 4.1. *Let $\Omega \subset \mathbb{R}^n$ be a symmetric convex polytope. Then $\Phi = \Psi_\Omega^* \circ \lambda$ is a convex spectral function. Moreover, there exists a convex spectral function $F : S_n \to \mathbb{R}$ of class $C^\infty$ such that*

$$\Phi(A) = \lim_{t\to\infty} F(tA)/t \quad \textit{for all } A \in S_n. \tag{4.2}$$

*An example of such a function is $F = f \circ \lambda$, where*

$$f(x) = \log \Big[ \int_{\mathbb{R}^n} e^{\langle w,x\rangle}\, d\mu(w) \Big] \tag{4.3}$$

*is defined in terms of a discrete measure $\mu$ which distributes uniformly its total mass among all the extreme points of $\Omega$.*

Proof. $\Phi$ is a convex spectral function because $\Psi_\Omega^*$ is a symmetric convex function (cf. Davis [4]). Since the convex polytope $\Omega$ is symmetric, so is the set $\operatorname{extr} \Omega$. From this, and the fact that the discrete measure $\mu$ distributes uniformly its total mass over $\operatorname{extr} \Omega$, one deduces that the convex function $f$ is symmetric. Hence, $F = f \circ \lambda$ is a convex spectral function. Since $f$ is of class $C^\infty$, so is $F$ (even if $\lambda : S_n \to \mathbb{R}^n$ is not differentiable). Finally, observe that, for all $A \in S_n$, one has

$$\lim_{t\to\infty} F(tA)/t = \lim_{t\to\infty} (f \circ \lambda)(tA)/t = \lim_{t\to\infty} f(t\lambda(A))/t = \Psi_\Omega^*(\lambda(A)).$$

This proves the convergence property (4.2). ■

EXAMPLE 4.1. The largest eigenvalue function $S_n \ni A \mapsto \lambda_1(A)$ corresponds to the composition of $\mathbb{R}^n \ni x \to g(x) = \max\{x_1, \ldots, x_n\}$ and $\lambda : S_n \to \mathbb{R}^n$. Thus

$$\lambda_1(A) = \lim_{t\to\infty} F(tA)/t \quad \text{for all } A \in S_n,$$

with

$$F(A) = \log \Big[ \sum_{j=1}^{n} e^{\lambda_j(A)} \Big] = \log[\operatorname{trace} e^A].$$

Observe that the smallest eigenvalue function is given by

$$\lambda_n(A) = \lim_{t\to-\infty} F(tA)/t.$$

EXAMPLE 4.2. Consider the function

$$S_n \ni A \mapsto \Phi(A) = \text{sum of the } p \text{ largest eigenvalues of } A.$$

In this case $\Phi = g \circ \lambda$, with

$$g(x) = \text{sum of the } p \text{ largest components of } x.$$

According to Overton and Womersley [10], $g$ is the supporting function of the set

$$\Omega = \Big\{u \in [0,1]^n : \sum_{j=1}^{n} u_j = p\Big\},$$

whose extreme points $u \in \Omega$ are given by

$$u_i = \begin{cases} 1 & \text{for exactly } p \text{ of the indices } 1,\ldots,n, \\ 0 & \text{otherwise.} \end{cases}$$

If one denotes by $w^1,\ldots,w^k$ the $k = n!/(p!(n-p)!)$ extreme points of $\Omega$, then one can approximate $\Phi(A)$ with the help of

$$F(A) = \log\Big[\sum_{j=1}^{k} e^{\langle w^j, \lambda(A)\rangle}\Big].$$

EXAMPLE 4.3. The spectral radius function

$$S_n \ni A \mapsto \Phi(A) = \max\{\lambda_1(A), -\lambda_n(A)\}$$

corresponds to the case $\Phi = \Psi_\Omega^* \circ \lambda$ with

$$\Omega = \Big\{u \in \mathbb{R}^n : \sum_{j=1}^{n} |u_j| \le 1\Big\}.$$

The extreme points of $\Omega$ are the canonical vectors of $\mathbb{R}^n$ and their opposite vectors. Thus, $\Psi_\Omega^*(x) = \max\{|x_1|,\ldots,|x_n|\}$ can be approximated with the help of

$$f(x) = \log(\cosh x_1 + \ldots + \cosh x_n).$$

This leads to the expression

$$F(A) = \log\Big(\sum_{j=1}^{n} \cosh \lambda_j(A)\Big) = \log[\operatorname{trace}(\cosh A)].$$

## References

[1] O. Barndorff-Nielsen, *Exponential families: exact theory*, Various Publ. Ser. 19, Inst. of Math., Univ. of Aarhus, Denmark, 1970.

[2] A. Ben-Tal and M. Teboulle, *A smoothing technique for nondifferentiable optimization problems*, in: Lecture Notes in Math. 1405, S. Dolecki (ed.), Springer, Berlin, 1989, 1–11.

[3] D. Bertsekas, *Constrained Optimization and Lagrangian Multiplier Methods*, Academic Press, New York, 1982.
[4] C. Davis, *All convex invariant functions of hermitian matrices*, Arch. Math. (Basel) 8 (1957), 276–278.
[5] R. A. El-Attar, M. Vidyasagar, and S. R. K. Dutta, *An algorithm for $\ell_1$-norm minimization with application to nonlinear $\ell_1$-approximation*, SIAM J. Numer. Anal. 16 (1979), 70–86.
[6] R. Ellis, *Entropy*, *Large Deviations and Statistical Mechanics*, Springer, Berlin, 1985.
[7] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau–Yosida regularization*: *theoretical preliminaries*, SIAM J. Optim. 7 (1997), 367–385.
[8] A. S. Lewis, *Convex analysis on the Hermitian matrices*, ibid. 6 (1996), 164–177.
[9] J. E. Martinez-Legaz, *On convex and quasiconvex spectral functions*, in: Proc. 2nd Catalan Days on Appl. Math., M. Sofonea and J. N. Corvellec (eds.), Presses Univ. de Perpignan, Perpignan, 1995, 199–208.
[10] M. L. Overton and R. S. Womersley, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Math. Programming 62 (1993), 321–357.
[11] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.
[12] A. Seeger, *Smoothing a nondifferentiable convex function*: *the technique of the rolling ball*, Technical Report 165, Dep. of Mathematical Sciences, King Fahd Univ. of Petroleum and Minerals, Dhahran, Saudi Arabia, October 1994.
[13] —, *Convex analysis of spectrally defined matrix functions*, SIAM J. Optim. 7 (1997), 679–696.

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

*Reçu par la Rédaction le 2.4.1996*
*Révisé le 17.10.1996*