Journal of Convex Analysis
Volume 19 (2012), No. 2, 519–524

Confexifying the Counting Function on $\mathbb{R}^p$ for Convexifying the Rank Function on $\mathcal{M}_{m,n}(\mathbb{R})$

Hai Yen Le
Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 09, France
yenlhp@yahoo.com

Received: January 10, 2011

We calculate the convex hull of the so-called counting function restricted to a ball of $\mathbb{R}^p$, and we then use it, with a result of A. S. Lewis, to recover the convex hull of the rank function restricted to a ball of $\mathcal{M}_{m,n}(\mathbb{R})$.

Keywords: Counting function, convex hull, rank of a matrix

1. Introduction

Some optimization problems on $\mathbb{R}^p$ involve (in the objective function or in the functions defining the constraints) the **counting function** $c : \mathbb{R}^p \to \mathbb{R}$ as follows:

$$\forall x = (x_1, \ldots, x_p) \in \mathbb{R}^p, \quad c(x) := \text{the number of i's for which } x_i \neq 0.$$

Sometimes, $c(x)$ is denoted as $\|x\|_0$, a misleading notation since $c(x)$ is not a norm on $\mathbb{R}^p$. Note however that, if $\|x\|_k$ denote $(\sum_{i=1}^p |x_i|^k)^{1/k}$ as usual, $(\|x\|_k)^k \to c(x)$ when $k \to 0^+$ (but $\|x\|_k$ does not converge to 0 when $p \to 0^+$, as it is stated sometimes). The function c gives rise to the so-called *Hamming* distance d (used in coding theory), defined on $\mathbb{R}^p$ as:

$$d(x, y) := c(x - y).$$

It is known that most of the optimization problems involving the counting funtion are NP-hard. So, what is usually done is to appeal to some "relaxed" form of c. The way of relaxing c that we consider in the present note is **to convexify** c. We get that the convex hull of the c function, restricted to some ball $\|x\|_\infty \leq r$, is simply $\frac{1}{r}\|.\|_1$, the scaled l_1 norm on $\mathbb{R}^p$.

When dealing with matrices $A \in \mathcal{M}_{m,n}(\mathbb{R})$, we know that:

- for $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, $\operatorname{rank}[\operatorname{diag}(x_1, \ldots, x_p)] = c(x)$;
- for $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $\operatorname{rank} A = c[\sigma(A)]$, where $\sigma(A) = (\sigma_1(A), \ldots, \sigma_p(A))$ is the vector made up with the singular values $\sigma_i(A)$ of A.

A. S. Lewis ([3], [4]) showed that the *Legendre-Fenchel* conjugate of a function of matrices (satisfying some specific properties) could be obtained by just conjugating

ISSN 0944-6532 / $ 2.50

some associated function of the singular values of A. Using his results twice, we are able to calculate the *Legendre-Fenchel* biconjugate of the rank function (that is the convex hull of the rank function) by calling on the biconjugate of the c function. In doing so, we retrieve Fazel's relaxation theorem ([1, p. 54–60]): the convex hull of the rank function restricted to the ball $\{A|\sigma_1(A) \leq r\}$ is $\frac{1}{r}\|A\|_* = \frac{1}{r}\sum_{i=1}^{p}\sigma_i(A)$, that is to say, within the factor $\frac{1}{r}$, the so-called nuclear norm (or trace norm) of A.

There is a "dictionary" between counting fuction minimization and rank minimization. For that and various examples and motivations of rank minimization problems, see [5].

2. Convexifying the counting function on $\mathbb{R}^p$

c is an integer-valued, subadditive, lower-semicontinous function on $\mathbb{R}^p$. Since $c(\alpha x) = c(x)$ for all $\alpha \neq 0$, there is no hope to get anything interesting by convexifying (i.e., taking the convex hull of) the function c. So, we consider it on some appropriate ball, namely, for $r > 0$:

$$c_r(x) := \begin{cases} c(x) & \text{if } \|x\|_\infty \leq r; \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

Taking the convex hull and the closed convex hull of c amount to the same here; so we just note $\mathrm{co}(c_r)$ the convexified form of c (i.e., the largest convex function minorizing c_r). Here is the result of this section.

Theorem 2.1. *We have:*

$$\forall x \in \mathbb{R}^p, \quad \mathrm{co}(c_r)(x) = \begin{cases} \frac{1}{r}\|x\|_1 & \text{if } \|x\|_\infty \leq r; \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. The basic properties of the convexifying operation (see [2, Chap. X] for example) show that the domain of $\mathrm{co}(c_r)$, i.e. the set on which this function is finite-valued, is just the convex hull of the domain of c_r. So, in our particular instance, the domain of $\mathrm{co}(c_r)$ is that of c_r, which is the convex set $\{x|\|x\|_\infty \leq r\}$.

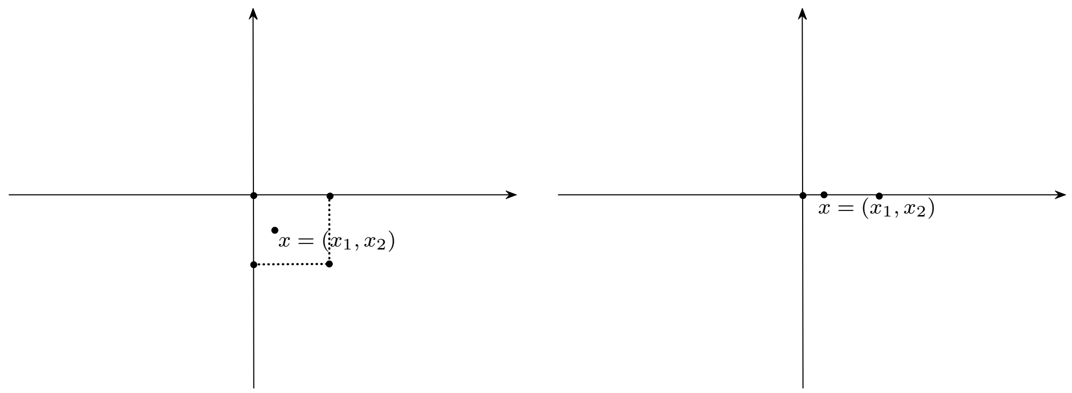
We therefor have to prove that $\mathrm{co}(c_r)(x) = \frac{1}{r}\|x\|_1$ whenever $\|x\|_\infty \leq r$.

First point. $\mathrm{co}(c_r)(x) \geq \frac{1}{r}\|x\|_1$ for x satisfying $\|x\|_\infty \leq r$.
If $\|x\|_\infty \leq r$,

$$c_r(x) = c(x) \geq \sum_{i=1}^{p} \frac{|x_i|}{\max_i |x_i|} = \frac{1}{\max_i |x_i|} \sum_{i=1}^{p} |x_i| \geq \frac{1}{r}\|x\|_1.$$

Second point. $\frac{1}{r}\|x\|_1 \geq \mathrm{co}(c_r)(x)$ for x satisfying $\|x\|_\infty \leq r$.

Let x satisfy $\|x\|_\infty \leq r$. For such an $x = (x_1, \ldots, x_p)$, we define vectors $y =$



(a) $c(x) = 2$ (b) $c(x) = 1$

Figure 2.1: $\square_x$ for $x = (x_1, x_2) \in \mathbb{R}^2$

$(y_1, \ldots, y_p)$ according to the following rule:

$$\begin{cases} \text{if } x_i = 0, \text{ then } y_i = 0; \\ \text{if } x_i > 0, \text{ then } y_i = 0 \text{ or } r; \\ \text{if } x_i < 0, \text{ then } y_i = 0 \text{ or } -r; \end{cases} \tag{2}$$

In doing so, we get at a "net on a box" $\square_x$ of $\mathbb{R}^p$:

$$\square_x := \{(y_1, \ldots, y_p)|\ y_i \text{ designed according to the rule (2)}\}$$
(see Figure 2.1, with $p = 2$).

$\square_x$ has $2^{c(x)}$ vectors, which are the vertices of a box containing x (this has been done for that!). In other words, x lies in the convex hull of $\square_x$: there exist real numbers $\alpha_1, \ldots, \alpha_k$ and $y_1, \ldots, y_k$ in $\square_x$ such that:

$$\begin{cases} \alpha_i \geq 0 \text{ for all } i \\ \sum_{i=1}^k \alpha_i = 1 \\ x = \sum_{i=1}^k \alpha_i y^i. \end{cases}$$

Consider now an arbitrary convex function h minorizing c_r. Then, due to the convexity of h,

$$h(x) = h\left(\sum_{i=1}^k \alpha_i y^i\right) \leq \sum_{i=1}^k \alpha_i h(y^i) \tag{3}$$

But, when $y \in \square_x$,

$$\begin{aligned} c_r(y) = &\text{ number of j's for which } y_j \neq 0 \\ = &\sum_{\{j|y_j \neq 0\}} \frac{|y_j|}{r} \qquad \text{(because } |y_j| = r \text{ whenever } y_j \neq 0) \\ = &\frac{1}{r} \sum_{\{j|y_j \neq 0\}} |y_j| = \frac{1}{r}\|y\|_1. \end{aligned}$$

So, with all the y^i lying in $\square_x$, we get from (3):

$$h(x) \leq \sum_{i=1}^{k} \alpha_i h(y^i) \leq \sum_{i=1}^{k} \alpha_i c_r(y^i) = \frac{1}{r} \sum_{i=1}^{k} \alpha_i \|y^i\|_1. \tag{4}$$

On the other hand, we have

$$x_j = \sum_{i=1}^{k} \alpha_i (y^i)_j \text{ for all } j = 1, \ldots, p.$$

Thus, due to the specific correspondence between the signs of x_j and $(y^i)_j$ (cf. (2)),

$$|x_j| = \sum_{i=1}^{k} \alpha_i |(y^i)_j| \text{ for all } j = 1, \ldots, p$$

so that:

$$\|x\|_1 = \sum_{i=1}^{k} \alpha_i \|y^i\|_1.$$

Consequently, we derive from (4):

$$h(x) \leq \frac{1}{r}\|x\|_1.$$

Finally,

$$\begin{aligned} \operatorname{co}(c_r)(x) = &\sup\{h(x)|h \text{ convex function minorizing } c_r\} \\ \leq &\frac{1}{r}\|x\|_1. \end{aligned}$$

Altogether (First point and Second point), we have proved that $\operatorname{co}(c_r)(x) = \frac{1}{r}\|x\|_1$ whenever $\|x\|_\infty \leq r$. $\square$

Comment 2.2. The result of Theorem 2.1 is part of the "folklore" in the areas where minimizing counting function appears (there are numerous papers in signal recovery, compressed sensing, statistics, etc.). We did not find any reference where it was stated in a clear-cut manner. That was the reason for a *direct* proof here.

Comment 2.3. Another convexification result, similar to Theorem 2.1, easy to prove, is as follows: Consider the function $\|.\|_k$ with $0 < k < 1$ (no more a norm), restricted to the ball $\{x|\ \|x\|_1 \leq 1\}$; then its convex hull is still the l_1 norm $\|.\|_1$ (restricted to the same ball).

3. Convexifying the rank function

For $A \in \mathcal{M}_{m,n}(\mathbb{R})$, let $p := \min(m,n)$ and $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_p(A)$ the singular values of A arranged in a decreasing order. So, if k is the rank of A, the first k singular values are positive, while the $p-k$ remaining ones are null. We make use of two matricial norms:

$$A \mapsto \|A\|_{sp} := \sigma_1(A), \text{ the largest singular value of } A;$$

$$A \mapsto \|A\|_* := \sum_{i=1}^{p} \sigma_i(A), \text{ the sum of all singular values of } A.$$

$\|.\|_{sp}$ is called the **spectral norm**, and $\|.\|_*$ the **nuclear norm** (although other various names are also used for $\|.\|_*$).

Consider the following function on $\mathcal{M}_{m,n}(\mathbb{R})$, it is just the "matricial cousin" of the c_r function in Section 2:

$$\operatorname{rank}_r(A) := \begin{cases} \text{rank of } A & \text{if } \|A\|_{sp} \leq r; \\ +\infty & \text{otherwise.} \end{cases}$$

Here also, convexifying the rank function on the whole space $\mathcal{M}_{m,n}(\mathbb{R})$ does not make sense: we just get at the null function.

But, for the rank function restricted to the ball $\{A \mid \|A\|_{sp} \leq r\}$, one gets an explicit form of its convex hull. Here is the "matricial cousin" of Theorem 2.1.

Theorem 3.1 (M. Fazel). *We have:*

$$\forall A \in \mathcal{M}_{m,n}(\mathbb{R}), \quad \operatorname{co}(\operatorname{rank}_r)(A) = \begin{cases} \frac{1}{r}\|A\|_* & \text{if } \|A\|_{sp} \leq r; \\ +\infty & \text{otherwise.} \end{cases}$$

M. Fazel ([1, p. 54–60]) proved Theorem 3.1 by calculating the *Legendre-Fenchel* conjugate $(\operatorname{rank}_r)^*$ of the rank_r function, and conjugating again, the biconjugate of the rank_r function; this biconjugate function $(\operatorname{rank}_r)^{**}$ is also the closed convex hull of rank_r (or just the convex hull of rank_r, since both coincide here). We propose here another path to prove Theorem 3.1: apply A. S. Lewis' fine results (of conjugation), such as displayed in [3] and [4]. Let us recall them briefly.

A function $f : \mathbb{R}^p \to \mathbb{R}$ is called absolutely symmetric if, for all $x \in \mathbb{R}^p$,

$$f(x_1, \ldots, x_p) = f(\hat{x}_1, \ldots, \hat{x}_p),$$

where $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_p)$ is the vector, built up from $x = (x_1, \ldots, x_p)$, whose components are the $|x_i|$'s arranged in a decreasing order. Associated with f is the function $F : \mathcal{M}_{m,n}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ defined as follows:

$$\forall A \in \mathcal{M}_{m,n}(\mathbb{R}), \quad F(A) := f[\sigma_1(A), \ldots, \sigma_p(A)].$$

A. S. Lewis' conjugacy rule is now:

Theorem 3.2 ([3], [4]). *With f satisfying the symmetry property above, we have:*

$$\forall X \in \mathcal{M}_{m,n}(\mathbb{R}), \quad F^*(X) = f^*[\sigma_1(X), \dots, \sigma_p(X)].$$

Proof of Theorem 3.1. Since f^* is in turn absolutely symmetric, we can apply Lewis' theorem twice, so that:

$$\forall A \in \mathcal{M}_{m,n}(\mathbb{R}), \quad F^{**}(A) = f^{**}[\sigma_1(A), \dots, \sigma_p(A)]. \tag{5}$$

In our particular setting, we choose:

$$f = c_r, \quad \text{so that } F = \text{rank}_r \,.$$

The biconjugate of f (resp. of F) is its (closed) convex hull $\text{co}(c_r)$ (resp. $\text{co}(\text{rank}_r)$). Whence Fazel's theorem follows from (5). □

4. Conclusion

With our main result, giving the convex hull of the (restricted) counting function, we have derived Fazel's convexification result on the (restricted) rank function. The way of doing we have adopted is in the lines of the following: when working with matrices (for question of convexity, differentiability, conjugation, rank, etc.), work instead on the "X-ray traces" provided by the singular values of such matrices; the results on $\mathbb{R}^p$ are then transferred to $\mathcal{M}_{m,n}(\mathbb{R})$.

Acknowledgements. We thank J.-B. Hiriart-Urruty, our Ph.D supervisor, who suggested the idea of the approach developed in the paper.

References

[1] M. Fazel: Matrix Rank Minimization with Applications, Ph.D. Thesis, Stanford University (2002).

[2] J.-B. Hiriart-Urruty, C. Lemarechal: Convex Analysis and Minimization Algorithms, Vol. 2, Springer, Berlin (1996) 35–82.

[3] A. S. Lewis: The convex analysis of unitarily invariant matrix functions, J. Convex Analysis 2 (1995) 173–183.

[4] A. S. Lewis: Convex analysis on the Hermitian matrices, SIAM J. Optim. 6 (1996) 164–177.

[5] B. Recht, M. Fazel, P. Parrilo: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52(3) (2010) 471–501.