Optim Lett
DOI 10.1007/s11590-011-0304-4

ORIGINAL PAPER

Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function

Jean-Baptiste Hiriart-Urruty · Hai Yen Le

Received: 14 September 2010 / Accepted: 8 March 2011

Abstract We provide an explicit description of the convex hull of the set of matrices of bounded rank, restricted to balls for the spectral norm. As applications, we deduce two relaxed forms of the rank function restricted to balls for the spectral norm: one is the quasiconvex hull of this rank function, another one is the convex hull of the rank function, thus retrieving Fazel's theorem (Matrix rank minimization with applications, 2002).

Keywords Rank of a matrix · Spectral norm · Trace (or nuclear) norm · Quasiconvex hull of a function

1 Introduction

In optimization problems, when the objective function or the constraint set are too difficult to handle directly, one usually appeals to some "relaxed" forms of them. This is especially true with optimization problems dealing with (rectangular) matrices. The set and the function (of matrices) we consider in the present paper are defined via the **rank function**. Before going further, let us fix some notations.

- $\mathcal{M}_{m,n}(\mathbb{R})$: the set of real matrices with m columns and n rows.
- Let $p := min(m, n)$. For $M \in \mathcal{M}_{m,n}(\mathbb{R})$, let $\sigma_1(M) \geq \cdots \geq \sigma_p(M)$ denote the singular values of M, arranged in the decreasing order; if r stands for the rank of M, the first r singular values are non-null, the remaining ones are null.

J.-B. Hiriart-Urruty · H. Y. Le (✉)
Institut de mathématiques de Toulouse, Université Paul Sabatier (Toulouse III),
118, route de Narbonne, 31062 Toulouse Cedex 09, France
e-mail: yenlhp@gmail.com

J.-B. Hiriart-Urruty
e-mail: jbhu@cict.fr

Published online: 22 March 2011

- Three matricial norms on $\mathcal{M}_{m,n}(\mathbb{R})$ will be used: for $M \in \mathcal{M}_{m,n}(\mathbb{R})$,
  $\|M\|_F := \sqrt{tr(M^T M)} = \sqrt{\sum_{i=1}^{p} \sigma_i(M)^2}$, the Frobenius norm of M;
  $\|M\|_{sp} := \sup_{x \neq 0} \frac{\|Mx\|_2}{\|x\|_2} = \sigma_1(M)$, the spectral norm of M;
  $\|M\|_* := \sum_{i=1}^{p} \sigma_i(M)$, the nuclear (or trace) norm of M.
  $\|.\|_F$ is a "smooth" norm since it derives from an inner product on $\mathcal{M}_{m,n}(\mathbb{R})$, namely $\ll M, N \gg := tr(M^T N)$. It is therefore its own dual, while the spectral norm and the nuclear norm are mutually dual (one is the dual norm of the other). These are classical results in matricial analysis. For variational characterizations of this duality relationship as semidefinite programs, see [9, Proposition 2.1].
- If $k \in \{0, 1, \ldots, p\}$ and $r \geq 0$,

$$S_k := \{M \in \mathcal{M}_{m,n}(\mathbb{R}) |\ \text{rank}\ M \leq k\},$$
$$S_k^r := S_k \cap \{M \in \mathcal{M}_{m,n}(\mathbb{R}) |\ \|M\|_{sp} \leq r\}.$$

  For $m = n$, S_k is an algebraic variety of dimension $(2n - k)k$.

Convexifying the set S_k is not of any use since the convex hull of S_k, denoted as co S_k, is the whole space $\mathcal{M}_{m,n}(\mathbb{R})$; indeed this comes from the singular value decomposition technique. Thus:

$$\forall k = 1, \ldots, p \quad \text{co}\ S_k = \mathcal{M}_{m,n}(\mathbb{R}).$$

The question becomes of some interest if we add some "moving wall" $\|M\|_{sp} \leq r$, like in the definition of S_k^r. So, we will give an explicit description of co S_k^r. As applications, we deduce two relaxed forms of the following (restricted) rank function:

$$\text{rank}_r(M) := \begin{cases} \text{rank of } M & \text{if } \|M\|_{sp} \leq r, \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

The first relaxed form is the so-called quasiconvex hull of rank_r, i.e., the largest quasi-convex function minorizing it. Then, as an ultimate step, we retrieve Fazel's theorem [3] on the convex hull (or biconjugate) of the rank_r function.

2 The main result

Theorem 1 *We have:*

$$co\ S_k^r = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) |\ \|M\|_{sp} \leq r \text{ and } \|M\|_* \leq rk\}. \tag{2}$$

Proof For either $k = 0$ or $r = 0$ there is nothing to prove. We therefore suppose that k is a positive integer and $r > 0$. Moreover, since rank (M/r) = rank M, and the norms are positively homogeneous functions ($\|M/r\| = \|M\|/r$), it suffices to prove (2) for $r = 1$.

First inclusion

$$\text{co } S_k^1 \subset \{M \in \mathcal{M}_{m,n}(\mathbb{R}) | \ \|M\|_{sp} \leq 1 \text{ and } \|M\|_* \leq k\}. \tag{3}$$

Let $M \in S_k^1$; by definition of S_k^1, we have $\|M\|_{sp} = \sigma_1(M) \leq 1$ and rank $M \leq k$. Consequently, all the non-null singular values of M—they are less than k—are majorized by 1; hence

$$\|M\|_* = \sum_{i=1}^{\text{rank } M} \sigma_i(M) \leq k.$$

Since the right-hand side of (3) is convex (as an intersection of sublevel sets of two norms), we derive the inclusion (3).

Reverse inclusion

$$\text{co } S_k^1 \supset \{M \in \mathcal{M}_{m,n}(\mathbb{R}) | \ \|M\|_{sp} \leq 1 \text{ and } \|M\|_* \leq k\}. \tag{4}$$

This is the tricky part of the proof. We first begin with a technical lemma on a specific convex polyhedron in $\mathbb{R}^p$; its proof can be found in [5, Exercises V.4 and V.15].

Lemma 1 *For $k = 1, \ldots, p$, let*

$$C := \left\{ x = (x_1, \ldots, x_p) \in \mathbb{R}^p | \ 0 \leq x_i \leq 1 \textit{ for all } i, \sum_{i=1}^p x_i \leq k \right\},$$

$$\Omega := \left\{ x = (x_1, \ldots, x_p) \in \mathbb{R}^p | \ x_i \in \{0, 1\} \textit{ for all } i, \sum_{i=1}^p x_i = k \right\}.$$

Then, $C = co\ \Omega$.

This result holds true because k is an integer. A picture in $\mathbb{R}^p$ helps to understand its meaning.

Let now M be satisfying $\|M\|_{sp} \leq 1$ and $\|M\|_* \leq k$. Consider a singular value decomposition of M:

$$M = U \Sigma V^T, \tag{5}$$

where U and V are orthogonal matrices of appropriate size and Σ, of the same type as M, with $\sigma_1(M), \ldots, \sigma_p(M)$ on the "diagonal" and 0 elsewhere. We write $\Sigma = diag_{m,n}(\sigma_1(M), \ldots, \sigma_p(M))$.

Because $0 \leq \sigma_i(M) \leq 1$ for all i and $\sum_{i=1}^p \sigma_i(M) \leq k$, according to the lemma recalled above, the vector $(\sigma_1(M), \ldots, \sigma_p(M))$ can be expressed as a convex combination of elements in Ω: there exist real numbers $\alpha_1, \ldots, \alpha_q$, vectors $\beta^1, \ldots, \beta^q$ in Ω such that:

$$\begin{cases} \alpha_j \in [0, 1] \text{ for all } j, \sum_{j=1}^p \alpha_j = 1 \\ (\sigma_1(M), \ldots, \sigma_p(M)) = \sum_{j=1}^p \alpha_j \beta^j. \end{cases} \tag{6}$$

For $\beta^j = (\beta_1^j, \ldots, \beta_p^j)$, we set

$$Y^j = diag_{m,n}(\beta_1^j, \ldots, \beta_p^j), \quad B^j = UY^jV^T. \tag{7}$$

Because $\beta^j \in \Omega$, we have:

$$\|B^j\|_{sp} = \|Y^j\|_{sp} \leq 1, \quad \text{rank } B^j = \text{rank } Y^j \leq k.$$

Moreover, in view of (6) and (7), we derive from (5):

$$\Sigma = \sum_{j=1}^{q} \alpha_j Y^j, \quad M = \sum_{j=1}^{q} \alpha_j B^j.$$

Hence, M is a convex combination of matrices in S_k^1. □

Remarks

1. Although S_k^r is a fairly complicated set of matrices (due to the definition of S_k), its convex hull is simple: according to (2), it is the intersection of two balls, one for the spectral norm, the other one for the nuclear norm. Getting at such an explicit form of co S_k^r is due to the happy combination of these specific norms. If $\|.\|$ were any norm on $\mathcal{M}_{m,n}(\mathbb{R})$ and

$$\hat{S}_k^r = \{M | \text{ rank } M \leq k \text{ and } \|M\| \leq r\},$$

due to the equivalence between the norm $\|.\|$ and $\|.\|_{sp}$, we would get with (2) an inner estimate and an outer estimate of co $\hat{S}_k^r$.
2. A particular case. Let $r = 1$ and $k = 1$ in the result of Theorem 1. We get that

$$\text{co } \{M | \text{rank } M \leq 1 \text{ and } \sigma_1(M) \leq 1\} = \left\{ M \left| \sum_{i=1}^{p} \sigma_i(M) \leq 1 \right. \right\}. \tag{8}$$

Remember that maximizing a linear form (of matrices) on both sets in (8) yields the same optimal value.
3. There are quite a few examples where the convex hull of a set of matrices can be expressed explicity. We mention here one of them, a very recent result indeed (see [4,7]). For $m \leq n$, let

$$T_m^n := \{M \in \mathcal{M}_{m,n}(\mathbb{R}) | M^T M = I_m\}.$$

T_m^n is called the Stiefel manifold. For $m = n$, T_n^n is just the set orthogonal (n, n) matrices. According to [7, p. 531] (see also [4]), the support function of T_m^n is $\|.\|_*$, hence:

$$\text{co } T_m^n = \{M | \|M\|_{sp} \leq 1\}. \tag{9}$$

For the particular case when $m = n$, a rather well-known (and interesting) exercise consists in proving that:

- $T_n^n \subset \{M | \|M\|_F = \sqrt{n}\}$;
- the support function of T_n^n is $M \mapsto \|M\|_* (= tr(M^T M)^{1/2})$.

As a result, the convex hull of T_n^n (whose support function is still $\|.\|_*$) is contained in the closed ball $\{M | \|M\|_F \le \sqrt{n}\}$ (whose support function is $\sqrt{n}\|.\|_F$). This is explained in [1, Exercise 59] for example.

3 A first application: the quasiconvex hull of the restricted rank function

The function we are going to relax here is the (restricted) rank function rank_r, such as defined in (1). Before going into this, we recall some basic facts about quasiconvex functions and the quasiconvexification of functions.

3.1 Quasiconvexity: a digest

Let X be a general topological vector space (for us, X will just be $\mathcal{M}_{m,n}(\mathbb{R})$) and $f : X \to \mathbb{R} \cup \{+\infty\}$. Such a f is said to be quasiconvex (in the sense used in optimization and mathematical economy, different from the one used in the calculus of variations) when:

$$\forall x_1, x_2 \in X, \forall \lambda \in [0, 1], f[\lambda x_1 + (1 - \lambda)x_2] \le \max[f(x_1), f(x_2)].$$

Besides this analytical definition, there is a geometrical characterization. We set the notation $[f \le \alpha]$ for $\{x \in X | \ f(x) \le \alpha\}$ (the sublevel set of f at the level $\alpha \in \mathbb{R}$). Then f is quasiconvex if and only if $[f \le \alpha]$ is convex for all $\alpha \in \mathbb{R}$.

Since the supremum of an arbitrary collection of quasiconvex functions is quasiconvex, we can define the largest quasiconvex function minorizing f; it is called the quasiconvex hull of f and denoted by f_q. Since the function f we are going to consider has compact sublevel sets, we do not add here the definition of the closed quasiconvex hull of f: f_q will be a closed function in our example.

How to construct f_q from the sublevel sets of f? The answer is easy:

$$\forall x \in X, \ f_q(x) = \inf\{\alpha | \ x \in \text{co}\,[f \le \alpha]\}. \tag{10}$$

All these results date back to J.-P. Crouzeix's works [2].

3.2 The explicit form of the quasiconvex hull of the rank function

Theorem 2 *The quasiconvex hull $rank_{r,q}$ of the function $rank_r$ is given as following:*

$$M \mapsto rank_{r,q}(M) = \begin{cases} \left\lceil \frac{1}{r}\|M\|_* \right\rceil & \text{if } \|M\|_{sp} \le r, \\ +\infty & \text{otherwise}, \end{cases} \tag{11}$$

wher $\lceil a \rceil$ stands for the smallest integer which is larger than a.

Proof Since the domain of the function rank_r (i.e., the set of M at which $\text{rank}_r(M)$ is finite-valued) is the (convex compact) ball $\{M \mid \|M\|_{sp} \leq r\}$, the quasiconvex hull $\text{rank}_{r,q}$ will have the same domain. In short,

$$\text{rank}_{r,q}(M) = +\infty \quad \text{if } \|M\|_{sp} > r.$$

Let $\alpha \geq 0$. Since the rank is an integer, one obviously has

$$[\text{rank}_r \leq \alpha] = [\text{rank}_r \leq \lfloor \alpha \rfloor],$$

where $\lfloor \alpha \rfloor$ denotes the integer part of α. So, by application of Theorem 1,

$$\begin{aligned} \text{co}\,[\text{rank}_r \leq \alpha] &= \text{co}\,[\,\text{rank}_r \leq \lfloor \alpha \rfloor] \\ &= \{M \mid \|M\|_{sp} \leq r \text{ and } \|M\|_* \leq r\lfloor \alpha \rfloor\}. \end{aligned}$$

Now, following the construction recalled in (10), we have: for all M such that $\|M\|_{sp} \leq r$,

$$\begin{aligned} \text{rank}_{r,q}(M) &= \inf\{\alpha \mid \|M\|_* \leq r\lfloor \alpha \rfloor\} \\ &= \inf\left\{\alpha \,\middle|\, \frac{\|M\|_*}{r} \leq \lfloor \alpha \rfloor\right\} = \left\lceil \frac{1}{r}\|M\|_* \right\rceil. \end{aligned}$$

□

4 A further step: the convex hull of the restricted rank function

A further relaxation of the function rank_r consists in taking its convex hull (or its closed convex hull, since it amounts to the same here). The convex hull $\text{co}(\text{rank}_r)$ of rank_r is the largest convex function minorizing rank_r; in geometrical words, the epigraph of $\text{co}(\text{rank}_r)$ is the convex hull of the epigraph of rank_r. This construction as well as the link with the Legendre–Fenchel transform can be found in any book on convex analysis and optimization (see [6], for example).

As a further and ultimate step from Theorem 2, we easily get at Fazel's theorem on the convex hull of rank_r.

Theorem 3 (M. Fazel [3]). *We have:*

$$M \mapsto co(rank_r)(M) = \begin{cases} \frac{1}{r}\|M\|_* & \text{if } \|M\|_{sp} \leq r, \\ +\infty & \text{otherwise.} \end{cases} \tag{12}$$

Proof Again, when $\|M\|_{sp} > r$, there is nothing special to say:

$$\text{rank}_r(M) = \text{co}(\text{rank}_r)(M) = +\infty.$$

We just have to prove that $\text{co}(\text{rank}_r)(M) = \frac{1}{r}\|M\|_*$ whenever $\|M\|_{sp} \leq r$. Consider therefore such an M.

First of all, observe that, since any convex function is quasiconvex, taking the hulls for both notions yield that

$$\mathrm{co}(\mathrm{rank}_r) \leq \mathrm{rank}_{r,q},$$

thus

$$\mathrm{co}(\mathrm{rank}_r) \leq \mathrm{co}(\mathrm{rank}_{r,q}). \tag{13}$$

As in the second part of the proof of Theorem 1, we set:

$$\begin{aligned} \Gamma &:= \{x = (x_1, \ldots, x_p) \in \mathbb{R}^p \mid x_i \in \{0, r\} \text{ for all } i\}, \\ \mathcal{M} &:= \{X \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \sigma_i(X) \in \{0, r\} \text{ for all } i\}. \end{aligned}$$

Since $(\sigma_1(M), \ldots, \sigma_p(M)) \in \mathrm{co}\,\Gamma$, M lies in $\mathrm{co}\,\mathcal{M}$. There therefore exists real numbers $\alpha_1, \ldots, \alpha_l$, matrices $X_1, \ldots, X_l$ in $\mathcal{M}$ (constructed like the matrices B^j in the proof of Theorem 1) such that:

$$\begin{cases} \alpha_i \in [0, 1] \text{ for all } i, \sum_{j=1}^{l} \alpha_i = 1 \\ M = \sum_{j=1}^{l} \alpha_j X_j. \end{cases} \tag{14}$$

Now, since $X_j \in \mathcal{M}$, it comes from Theorem 2 that

$$\mathrm{rank}_{r,q}(X_j) = \left\lceil \frac{1}{r} \|X_j\|_* \right\rceil = \frac{1}{r} \|X_j\|_*.$$

Consequently,

$$\begin{aligned} \mathrm{co}(\mathrm{rank}_{r,q})(M) &= \mathrm{co}(\mathrm{rank}_{r,q}) \left(\sum_{j=1}^{l} \alpha_j X_j \right) \\ &\leq \sum_{j=1}^{l} \alpha_j \mathrm{co}(\mathrm{rank}_{r,q})(X_j) \\ &\leq \sum_{j=1}^{l} \alpha_j \mathrm{rank}_{r,q}(X_j) \\ &= \sum_{j=1}^{l} \frac{1}{r} \alpha_j \|X_j\|_* = \frac{1}{r} \|M\|_*. \end{aligned}$$

Thus, $\mathrm{co}(\mathrm{rank}_{r,q})(M) \leq \frac{1}{r} \|M\|_*$.

On the other hand, because $\text{rank}_r(M) \geq \frac{1}{r}\|M\|_*$ for all $M \in \mathcal{M}_{m,n}(\mathbb{R})$ and $\frac{1}{r}\|.\|_*$ is convex, we have that:

$$\text{co}(\text{rank}_r)(M) \geq \frac{1}{r}\|M\|_*.$$

So we have proved that

$$\text{co}(\text{rank}_{r,q})(M) = \text{co}(\text{rank}_r) = \frac{1}{r}\|M\|_*.$$

□

Remarks

1. M. Fazel ([3, pp. 54-60]) proved Theorem 3 by calculating the Legendre–Fenchel conjugate $(\text{rank}_r)^*$ of the rank_r function, and conjugating again, the biconjugate of the rank_r function; this biconjugate function rank_r^{**} is the closed convex hull of rank_r (or just the convex hull of rank_r, since both coincide here).
   Another path is followed by Le in [8]: first convexify the counting (or cardinal) function on $\mathbb{R}^p$

$$x = (x_1, \ldots, x_p) \in \mathbb{R}^p \mapsto c_r(x) := \begin{cases} \text{the numbers of } i\text{'s} & \\ \text{ for which } x_i \neq 0 & \text{if } \|x\|_\infty \leq r, \\ +\infty & \text{otherwise,} \end{cases}$$

   and then apply A. Lewis' fine results on how to conjugate (or biconjugate) functions of matrices.
2. A way of recovering the convex hull of the c_r function via Fazel's theorem is to note that

$$c_r(x) = \text{rank}_r[diag(x_1, \ldots, x_p)]$$

   and apply the calculus rules on the convex hull of composed functions (see [6], for example).
   In brief, Fazel's theorem and the one giving that the convex hull of c_r is

$$\text{co}(c_r)(x) = \begin{cases} \frac{1}{r}\|x\|_1 & \text{if } \|x\|_\infty \leq r; \\ +\infty & \text{otherwise,} \end{cases}$$

   proved directly in [8], are equivalent.

References

1. Azé, D., Hiriart-Urruty, J.-B.: Analyse variationnelle et optimisation. Eléments de cours, exercices et problèmes corrigés. Editions Cépaduès, Toulouse (2010)

2. Crouzeix, J.-P.: Contributions à l'étude des fonctions quasiconvexes. Thèse de doctorat ès sciences. Université de Clermont-Ferrand II (1977)
3. Fazel, M.: Matrix rank minimization with applications. Ph.D Thesis. Stanford University (2002)
4. Gallivan, K.A., Absil, P.-A.: Note on the convex hull of the Stiefeld manifold. Technical note (2010)
5. Hiriart-Urruty, J.-B.: Optimisation et analyse convexe. Exercices corrigés (1998). Reprinted by EDP Sciences (2009)
6. Hiriart-Urruty, J.-B., Lemarechal, C.: Convex analysis and minimization algorithms, vol. 2. pp. 35–82. Springer, Berlin (1993)
7. Journée, M., Nesterov, Y., Richtárik, P., Sepulchre, R.: Generalized power method for sparse principal component analysis. J. Mach. Learn. Res. **11**, 517–553 (2010)
8. Le, H.Y.: Convexifying the counting function on $\mathbb{R}^p$ for convexifying the rank function on $\mathcal{M}_{m,n}(\mathbb{R})$ (2010, submitted)
9. Recht, B., Fazel, M., Parrilo, P.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. (2011, in press)