To appear in TOP (Journal of the Spanish Society
of Statistics and Operations Research), July 2013.

A VARIATIONAL APPROACH OF THE RANK FUNCTION

Jean-Baptiste Hiriart-Urruty and Hai Yen Le
Institute of Mathematics
Paul Sabatier University, Toulouse (France)
jbhu@math.univ-toulouse.fr, hyle@math.univ-toulouse.fr
http://www.math.univ-toulouse.fr/~jbhu/

Abstract

In the same spirit as the one of the paper [24] on positive semidefinite matrices, we survey several useful properties of the rank function (of a matrix) and add some new ones. Since the so-called rank minimization problems are the subject of intense studies, we adopt the viewpoint of variational analyis, that is the one considering all the properties useful for optimizing, approximating or regularizing the rank function.

Keywords: Rank of matrix. Optimization. Nonsmooth analysis. Moreau-Yosida regularization. Generalized subdifferentials.

Introduction

Associated with (square) matrices are some familiar notions like the *trace*, the *determinant*, *etc.* Their study from the variational point of view, via the usual differential calculus, is easy and well-known. We consider here another function of (not necessarily square) matrices, the **rank function**. The rank function has been studied for its properties in linear algebra (or matrix calculus), semi-algebraic geometry, *etc.* One task in the present paper is to survey some of the main properties of the rank in these classical areas of mathematics. But we are more interested in considering the rank function from the variational viewpoint. Actually, besides being integer-valued, the rank function is lower-semicontinuous; this is the only valuable topological property it enjoys. But, since modern nonsmooth analysis allows us to deal with non-differentiable (even discontinuous) functions, a second task we intend to carry out is to answer questions like: what is the generalized subdifferential (in whatever sense) of the rank function?

If we are interested in the rank function from the variational viewpoint, it is because the rank function appears as an objective (or constraint) function in various modern optimization problems. The archetype of the so-called **rank minimization problems** is as follows:

$$(\mathcal{P}) \quad \begin{cases} \text{Minimize } f(A) := \text{rank of } A \\ \text{subject to } A \in \mathcal{C}, \end{cases}$$

where $\mathcal{C}$ is a subset of $\mathcal{M}_{m,n}(\mathbb{R})$ (the vector space of m by n real matrices). The constraint set is usually rather "simple" (expressed as linear equalities, for example), the main difficulty lies in the objective function. Of course, we could have an optimization problem

$$(\mathcal{P}_1) \quad \begin{cases} \text{Minimize } g(A) \\ \text{subject to } A \in \mathcal{C} \text{ and rank } A \leq k, \end{cases}$$

with a rather simple objective function but a fairly complicated constraint set. Both problems $(\mathcal{P})$ and $(\mathcal{P}_1)$ suffer from the same intrinsic difficulty: the occurence of the rank function.

A related (or cousin) problem to $(\mathcal{P})$, actually equivalent in terms of difficulty, stated in $\mathbb{R}^n$ this time, consists in minimizing the so-called counting (or cardinality) function $x = (x_1, \ldots, x_n) \in \mathbb{R}^n \longmapsto c(x) :=$ number of nonzero components x_i of x:

$$(\mathcal{Q}) \quad \begin{cases} \text{Minimize } c(x) \\ \text{subject to } x \in S, \end{cases}$$

where S is a subset of $\mathbb{R}^n$. Often $c(x)$ is denoted as $\|x\|_0$, although it is not a norm.

Problems $(\mathcal{P})$ and $(\mathcal{Q})$ share some bizarre and/or interesting properties, from the optimization or variational viewpoint. We shall review some of them like those related to *relaxation*, *global optimization*, *Moreau-Yosida approximations*, *generalized subdifferentials*.

Here is the plan of our survey.

1. **The rank in linear algebra or matricial calculus.**

2. **The rank from the topological and semi-algebraic viewpoint.**

3. Best approximation in terms of rank.

4. Global minimization of the rank.

5. Relaxed forms of the rank function.

6. Surrogates of the rank function.

7. Regularization-approximation of the rank function.

8. Generalized subdifferentials of the rank function.

9. Further notions related to the rank: the spark, the rigidity, the cp-rank of a matrix.

1 The rank in linear algebra or matricial calculus

Let $p := \min(m,n)$, let

$$\begin{array}{rrcl} \text{rank}: & \mathcal{M}_{m,n}(\mathbb{R}) & \longrightarrow & \{0,1,\ldots,p\} \\ & A & \longmapsto & \text{rank } A. \end{array}$$

The properties of the rank function are well-known in the context of linear algebra or matricial calculus. Any book on these subjects presents them; there even are compilations of them (see [5] for example). We recall here the most basic ones:

(a) rank $A =$ rank A^T; rank $A =$ rank $(AA^T) =$ rank (A^TA).

(b) If the product AB can be done,

$$\text{rank } (AB) \leq \min(\text{rank } A, \text{rank } B) \quad (\text{S}\textsc{ylvester} \text{ inequality}).$$

As a general rule, when the proposed products of matrices can be done,

$$\text{rank } (A_1A_2\ldots A_k) \leq \min_{i=1,\ldots,k}(\text{rank } A_1, \text{rank } A_2, \ldots, \text{rank } A_k).$$

When $m = n$,

$$\text{rank } (A^k) \leq \text{rank } A.$$

(c) rank $A = 0$ if and only if $A = 0$; rank $(cA) =$ rank A for $c \neq 0$.

(d) $|\text{rank } A - \text{rank } B| \leq \text{rank } (A+B) \leq \text{rank } A + \text{rank } B$.

While properties like (b) are the ones used in solving systems of linear equations for example, properties and inequalities in (c)-(d) are those which are useful in optimization. The second property in (c) shows us that it is unnecessary to look at the rank of matrices lying outside some ball centered at 0 and of positive radius. Property (a) involves positive semidefinite (symmetric) matrices $A^T A$ and AA^T; therefore, *a priori* there should not be a loss of generality in considering positive semidefinite matrices, but it turns out that it is not the case in rank minimization problems (we have to treat of $A \in \mathcal{M}_{m,n}(\mathbb{R})$ as they are given in the data of the problems).

Since rank is a complicated integer-valued function, are there ways of bounding it by using more familiar and easier to handle functions like the trace, the determinant, the eigenvalue functions? Indeed, there are some, not very convincing however. Here are two prosaic ones.

Theorem 1. *(i) Let $A \in \mathcal{M}_n(\mathbb{R})$ be non-null diagonalizable, with all real eigenvalues (for example $A \in \mathcal{S}_n(\mathbb{R})$). Then*

$$\text{rank } A \geq \frac{(\text{tr } A)^2}{\text{tr } (A^2)}, \tag{1}$$

where tr B *denotes the trace of the matrix B.*

(ii) Let $A \in \mathcal{S}_n(\mathbb{R})$ be non-null and positive semidefinite. Then

$$\frac{\text{tr } A}{\lambda_{\max}(A)} \leq \text{rank } A \leq \frac{\text{tr } A}{\lambda_{\min}(A)}, \tag{2}$$

where $\lambda_{\max}(A)$ (resp. $\lambda_{\min}(A)$) stands for the maximal (resp. minimal) eigenvalue of A. (If $c > 0$, c divided by 0 is put equal to $+\infty$).

Proof. We content ourselves by sketching the proof of the inequality (1).
For given real numbers $\lambda_1, \ldots, \lambda_k$, let us denote by $\tilde{\lambda} := \frac{\lambda_1 + \cdots + \lambda_k}{k}$ their mean value. From the inequality $\sum_{i=1}^k (\lambda_i - \tilde{\lambda})^2 \geq 0$ comes the inequality below

$$\sum_{i=1}^k \lambda_i^2 \geq \frac{(\sum_{i=1}^k \lambda_i)^2}{k} = k\tilde{\lambda}^2. \tag{3}$$

Let thus $A \neq 0$ of rank k. If $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A, we may suppose, without loss of generality, that $\lambda_1 \neq 0, \ldots, \lambda_k \neq 0, \lambda_{k+1} = 0, \ldots, \lambda_n = 0$ (since $k \geq 1$).

Now, tr $A = \sum_{i=1}^{k} \lambda_i$, tr $(A^2) = \sum_{i=1}^{k} \lambda_i^2$. We then infer from (3)

$$k \geq \frac{(\sum_{i=1}^{k} \lambda_i)^2}{\sum_{i=1}^{k} \lambda_i^2} = \frac{(\text{tr } A)^2}{\text{tr } (A^2)}.$$

□

Comments and examples

- Having an equality in (1) requires too specific properties on A, like: there exists k such that $A^2 = kA$.

- The eigenvalues of A do not appear explicitly in the inequality (1). If the eigenvalues λ_i of A are available, a more informative inequality is

  $$\text{rank } A \geq \frac{(\sum_{i=1}^{k} |\lambda_i|)^2}{\text{tr}(A^2)}.$$

  This comes from the Cauchy-Schwarz inequality.

- Bounds (1) and (2), although rough, could be used to relax constraints on the rank of a parameterized matrix: rank $[A(x)] \leq k$ could be replaced by [tr $A(x)]^2 \leq k$ tr $[A(x)^2]$ or $\frac{\text{tr } [A(x)]}{\lambda_{\max}[A(x)]} \leq k$.

  To take a simple example, consider $x \in \mathbb{R} \longmapsto A(x) = \begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix} \in \mathcal{S}_2(\mathbb{R})$. Requiring that $A(x)$ is positive semidefinite amounts to having $x \in [-1; 1]$; requiring futhermore, that rank $[A(x)] \leq 1$ leads to $x = -1$ or $x = 1$. Now keeping the constraint "$A(x)$ is positive semidefinite", one relaxes the constraint rank $[A(x)] \leq 1$ with tr $[A(x)]^2 \leq$ tr $[A(x)^2]$. This leads to $x^2 \geq 1$, that is $x = -1$ or $x = 1$ again.

  Let us look at the second relaxation, the one coming from (2). Since tr $[A(x)] = 2$ and $\lambda_{\max}[A(x)] = 1 + |x|$, substituting the relaxed form $\frac{\text{tr } [A(x)]}{\lambda_{\max}[A(x)]} \leq 1$ for rank $[A(x)] \leq 1$ gives rives to $\frac{2}{1+|x|} \leq 1$, that is again $x = -1$ or $x = 1$.

- Note that $A \mapsto$ tr A is a linear function, $A \mapsto \lambda_{\max}(A)$ a convex one, $A \mapsto \lambda_{\min}(A)$ a concave one. The bounding functions in (2) are nonsmooth but continuous on the domains where they are defined.

2 The rank from the topological and semi-algebraic viewpoint

The only (useful) topological property of the rank function is that it is *lower-semicontinuous*: if $A_\nu \to A$ in $\mathcal{M}_{m,n}(\mathbb{R})$ when $\nu \to +\infty$, then

$$\liminf_{\nu \to +\infty} \operatorname{rank} A_\nu \geq \operatorname{rank} A. \tag{4}$$

This is easy to see if one thinks of rank A characterized as the maximal integer r such that a (r, r)-submatrix extracted from A is invertible; the continuity of the determinant function makes the rest.

Since the rank function is integer-valued, a consequence of the inequality (4) is that the rank function does not decrease in a sufficiently small neighborhood of any matrix A.

For $k \in \{0, 1, \ldots, p\}$, consider now the following two subsets of $\mathcal{M}_{m,n}(\mathbb{R})$:

$$S_k := \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \operatorname{rank} A \leq k\},$$

$$\Sigma_k := \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \operatorname{rank} A = k\}.$$

S_k is the sublevel-set (at level k) of the lower semicontinuous function rank; it is therefore closed. But, apart from the case $k = 0$ (where $S_0 = \Sigma_0$), what about the topological structure of Σ_k? The answer is given in the following statement.

Theorem 2. *(i) Σ_p is an open dense subset of $S_p = \mathcal{M}_{m,n}(\mathbb{R})$.*

(ii) If $k < p$, the interior of Σ_k is empty and its closure is S_k.

The singular value decomposition of matrices (see below) will show how intricate the subsets S_k and Σ_k may be. For example, *if* rank $A = k$*, in any neighborhood of A, there exist matrices of rank $k + 1, k + 2, \ldots, p$.*

Now, what about S_k from the algebraic geometry viewpoint? Consider $m = n$ for the sake of simplicity. The geometric structure of the S_k's is well understood (*cf.* [45]): since S_k is defined by the vanishing of all $(k+1, k+1)$-minors of A, it is thus a solution set of polynomial equations, therefore a so-called *semi-algebraic variety.* It is non-singular, except on those points (matrices) of rank less than or equal to $k - 1$. Its dimension is $(2n - k)k$. At a smooth point (matrix) A of S_k, the tangent space $T_{S_k}(A)$ can be made explicit, provided that one knows a singular value decomposition of A ([45, p.15]).

3 Best approximation in terms of rank

Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$ of rank r and an integer $k \leq r$, how are made the matrices in S_k closest to A? This best approximation problem needs firstly that a distance (via a norm for instance) be defined on $\mathcal{M}_{m,n}(\mathbb{R})$. Secondly, even if the existence of best approximants does not offer any difficulty (remember that S_k is closed), the question of uniqueness as well as that of an explicit form of best approximants remain posed. It turns out that there is a beautiful theorem answering these questions.

Before going further, we recall a technique of decomposition of matrices which is central in numerical matricial analysis and in statistics: the singular value decomposition (SVD). Here it is: *Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$, there is an (m,m) orthogonal matrix U, an (n,n) orthogonal matrix V, a "pseudo-diagonal" matrix D* [1] *of the same sizes as A, such that $A = UDV$* [2].

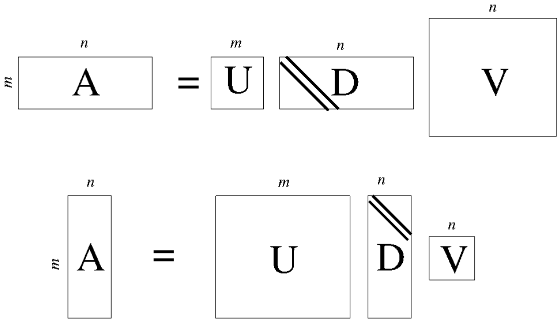

Figure 1: The singular value decomposition

The picture above helps to understand the decomposition.

- D has the same number of columns and rows as A, it is a sort of skeleton of A: all the "non-diagonal" entries of D are null; on the "diagonal" of D are the *singular values* $\sigma_1, \sigma_2, \dots, \sigma_p$ of A, that are the square roots of the eigenvalues of A^TA (or AA^T). By definition, all the σ_i's are nonnegative, and exactly r of them (if $r = \text{rank } A$) are positive. By changing the ordering in columns or rows in U and V, and without loss of generality,

[1] D "pseudo-diagonal" means that $d_{ij} = 0$ for $i \neq j$. One also uses the notation $\text{diag}_{m,n}[\sigma_1, \dots, \sigma_p]$ for D.

[2] The notation for the SVD is a bit nonstandard. Usually, a singular value decomposition takes the form $U\Sigma V^T$ (*i.e.*, V^T instead of V). But there is no difference from the mathematical point of view.

we can suppose that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_p = 0.$$

- U and V are orthogonal matrices of appropriate sizes (so that the product UDV of matrices can be performed). Of course, these U and V are not unique. We denote by $O(m,n)^A$ the set of (U,V) such that U and V are orthogonal matrices and $U\mathrm{diag}_{m,n}(\sigma_1(A),\ldots,\sigma_p(A))V^T$ is a singular value decomposition of A.

  Clearly, if one has a SVD of A, $A = UDV$, one also has one of A^T, namely $A^T = V^T D^T U^T$.

  The SVD helps to define some norms on $\mathcal{M}_{m,n}(\mathbb{R})$. The three most important ones are:

  - $A \longmapsto \max_{i=1,\ldots,p} \sigma_i(A)$, called the *spectral norm* of A, which we denote by $\|A\|_{sp}$ ($\|A\|_2$ is another much used notation);
  - $A \longmapsto \sqrt{\mathrm{tr}\ (A^T A)} = \sqrt{\sum_{i=1}^p \sigma_i^2(A)}$, called the *Frobenius-Schur norm* of A, which we denote by $\|A\|_F$. This norm is derived from the inner product $\langle\langle A,B\rangle\rangle := \mathrm{tr}\ (A^T B)$ on $\mathcal{M}_{m,n}(\mathbb{R})$; thus $(\mathcal{M}_{m,n}(\mathbb{R}), \|.\|_F)$ is an Euclidean space.
  - $A \longmapsto \sum_{i=1}^p \sigma_i(A)$, called the *nuclear norm* (or trace norm, or 1-norm, *etc.*) which we denote by $\|A\|_*$.

These three norms are the “matricial cousins” of the usual norms $\|.\|_\infty, \|.\|_2, \|.\|_1$ on $\mathbb{R}^p$.

The best approximation problem that we consider in this section is as follows: Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$ of rank r,

$$(\mathcal{A}_k) \qquad \begin{cases} \text{Minimize } \|A - M\| \\ M \in S_k \end{cases}$$

The problem of course depends on the choice of $\|.\|$; it is solved when $\|.\|$ is either the Frobenius-Schur norm or the spectral norm.

Theorem 3 (Eckart and Young, Mirsky [20])**.** *Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and let UDV be a SVD of A. Choose $\|.\|$ as either $\|.\|_F$ or $\|.\|_{sp}$. Then*

$$A_k := UD_kV,$$

(where D_k is obtained from D by keeping $\sigma_1,\ldots,\sigma_k$ and putting 0 in the place of $\sigma_{k+1},\ldots,\sigma_r$) is a solution of the best approximation problem $(\mathcal{A}_k)$. For the Frobenius-Schur norm case, A_k is the unique solution in $(\mathcal{A}_k)$ when $\sigma_k > \sigma_{k+1}$.

The optimal values in $(\mathcal{A}_k)$ *are as following:*

$$\min_{M \in S_k} \|A - M\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2};$$

$$\min_{M \in S_k} \|A - M\|_{sp} = \sigma_{k+1}.$$

This theorem is a classical result in numerical matricial analysis, proved for the Frobenius-Schur norm by Eckart and Young (1936). Mirsky (1960) showed that A_k is a minimizer in $(\mathcal{A}_k)$ for any unitary invariant norm (a norm $\|.\|$ on $\mathcal{M}_{m,n}(\mathbb{R})$ is called unitary invariant if $\|UAV\| = \|A\|$ for any orthogonal pair of matrices U and V).

From the historical point of view, there is however some controversy about the naming of Theorem 3; according to Stewart ([48]), the mathematician E.Schmidt should be credited for having given this approximation theorem, while studying integral equations, in a publication which dates back to 1907. We nevertheless stand by the usual appellation (in papers as well as in textbooks) which is "Eckart and Young theorem".

Remark 1. With the Frobenius-Schur norm, the objective function in $(\mathcal{A}_k)$ (taking its square actually, $\|A - M\|_F^2$) is convex and smooth. But, due to the non-convexity of the constraint set S_k, the optimization problem $(\mathcal{A}_k)$ is non-convex. It is therefore astonishing that one could provide (via Theorem 3) an explicit form of a solution of this problem. In short, since the S_k's are the sublevel-sets of the rank function, *one always has at one's disposal a "projection" of (an arbitrary) matrix A on the sublevel-sets S_k.*

Remark 2. From the differential geometry viewpoint, Σ_k is a smooth manifold around any matrix $A \in \Sigma_k$, while S_k is not. However, S_k enjoys a very specific property called *prox-regularity*[3]. Indeed, it is proved in various places like [38], [1], [43] that *S_k is prox-regularity at any $A \in S_k$ with* rank $A = k$. Moreover, the projection on S_k is also the projection on Σ_k, at least locally; more precisely, according to [1]: Let $\overline{A} \in \Sigma_k$, then for every A such that $\|A - \overline{A}\| \leq \sigma_k(\overline{A})/2$, the projection of A onto Σ_k exists, is unique, and is the projection of A onto S_k (given in the Eckart and Young theorem).

Remark 3. As stated earlier, if $\sigma(A) = (\sigma_1(A), \ldots, \sigma_p(A))$, the counting function c at $\sigma(A)$ is exactly the rank of A,

$$c[\sigma(A)] = \text{rank } A. \tag{5}$$

[3] A nonempty closed set S is an Euclidean space E is called *prox-regular* at a point $\overline{x} \in S$ if the projection set $P_S(x)$ of x on S is single-valued around $\overline{x}$.

The MATLAB software uses this (SVD decomposition) to compute the rank of a matrix. Among the many ways to compute the rank of a matrix, this one is maybe the most time consuming, but also the most reliable.

4 Global minimization of the rank

Consider the rank minimization problem $(\mathcal{P})$ that we alluded to in the beginning of the paper. The constraint set $\mathcal{C}$ does not contain the zero matrix; if it were so, the zero matrix would be trivially a global minimizer in $(\mathcal{P})$. If only local minimizers are sought, the problem does not make sense. As explained in [22], *every point (matrix) $A \in \mathcal{C}$ is a local minimizer in $(\mathcal{P})$* [4]. So, here, instead of tackling directly the hard problem $(\mathcal{P})$, a much used approach in optimization is to consider a relaxed form of $(\mathcal{P})$ by minimizing a more tractable objective function.

5 Relaxed forms of the rank function

Given the rank function rank : $\mathcal{M}_{m,n}(\mathbb{R}) \longrightarrow \mathbb{R}$, what is its closest "convex relative", that is to say its convex hull[5]? Posed as it is, this question has no interest; indeed, due to the properties (c) in Section 1, it is easy to understand that co(rank), the convex hull of the rank function, is the zero function (on $\mathcal{M}_{m,n}(\mathbb{R})$). So, we restrict the rank function to an appropriate ball, by assigning the value $+\infty$ out of this ball:

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto \operatorname{rank}_R(A) := \begin{cases} \text{rank of } A & \text{if } \|A\|_{sp} \leq R; \\ +\infty & \text{otherwise.} \end{cases} \tag{6}$$

We call this function the "restricted rank function". Now, an interesting question is: what is the convex hull of rank_R? To the best of our knowledge, this question was firstly answered in Fazel's PhD thesis ([18]).

[4] Such a phenomenon happens because the objective function is not continuous. Indeed, if $\mathcal{C}$ is a connected set, if $f : \mathcal{C} \longrightarrow \mathbb{R}$ is a continuous function, and if every point in $\mathcal{C}$ is a local minimizer or a local maximizer of f, then f is necessarily constant on $\mathcal{C}$ ([8]).

[5] The *convex hull* or *convex envelope* of a function f is the largest possible convex function below f. Its epigraph (*i.e.*, what is above the graph) is exactly the convex hull of the epigraph of f.

Theorem 4 (Fazel [18])**.** *We have:*

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

There are various ways of proving this theorem; we explain and comment three of them.

Proof 1. It consists in calculating the Legendre-Fenchel conjugate ϕ^* of $\phi := \mathrm{rank}_R$, and then the biconjugate ϕ^{**} $(:= (\phi^*)^*)$ of ϕ. We know from convex analysis that, since ϕ is bounded from below, ϕ^{**} is the closed convex hull of rank_R, that is: $\phi^{**} = \overline{\mathrm{co}}(\mathrm{rank}_R)$. But here, there is no difference between $\mathrm{co}(\mathrm{rank}_R)$ and its closure $\overline{\mathrm{co}}(\mathrm{rank}_R)$.

This is the way followed by Fazel ([18], Section 5.1.4).

Proof 2. Here, one begins by convexifying the restricted counting (or cardinality) function (on $\mathbb{R}^p$), and then one applies fine results by Lewis on how to get propertiers on rank_R from those on c_R.

The so-called restricted counting function c_R on $\mathbb{R}^p$ is defined as follows:

$$x \in \mathbb{R}^p \longmapsto c_R(x) := \begin{cases} c(x) & \text{if } \|x\|_\infty \leq R; \\ +\infty & \text{otherwise.} \end{cases}$$

The next statement gives the convex hull of c_R.

Theorem 5. *We have*

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

This result is part of the "folklore" in the areas where minimization problems involving counting functions appear (there are dozens of papers in signal recovery, compressed sensing, statistics, *etc.*). A direct proof is presented in [34].

A.Lewis ([36],[37]) showed that the Legendre-Fenchel conjugate of a function of matrices (satisfying some specific properties) could be obtained by just conjugating some associated function of the singular values of A. Using his results twice, one is able to calculate the Legendre-Fenchel biconjugate of the (restricted) rank function by calling on the biconjugate of the (restricted) counting function. Indeed, when dealing with matrices, we have: for $A \in \mathcal{M}_{m,n}(\mathbb{R})$,

$$\mathrm{rank}_R(A) = c_R[\sigma(A)], \tag{7}$$

where $\sigma(A) = (\sigma_1(A), \ldots, \sigma_p(A))$ is the vector of $\mathbb{R}^p$ made up with the singular values $\sigma_i(A)$ of A. It happens that the assumptions of Lewis' "transfer results" from $\mathbb{R}^p$ to $\mathcal{M}_{m,n}(\mathbb{R})$ are satisfied in our context (expressed in (7)). In doing so, ones retrieves Theorem 4.

Proof 3. Here we look at the problem with "geometrical glasses": instead of convexifying functions directly, we firstly convexify (appropriate) sets of matrices and, then, get convex hulls (or quasiconvex hulls) of functions as by-products. This is the path followed in ([25]).

For that purpose, let

$$\begin{aligned} S_k^R &:= S_k \cap \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \|A\|_{sp} \leq R\} \\ &= \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \text{rank } A \leq k \text{ and } \|A\|_{sp} \leq R\}. \end{aligned}$$

In this definition, $k \in \{0, 1, \ldots, p\}$ and $R \geq 0$; the restriction "$\|A\|_{sp} \leq R$" plays the role of a "moving wall". Although S_k^R is a very complicated set of matrices (due to the definition of S_k), its convex hull has a fairly simple expression.

Theorem 6 (Hiriart-Urruty and Le, [25])**.** *We have*

$$\text{co } S_k^R = \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \|A\|_{sp} \leq R \text{ and } \|A\|_* \leq Rk\}. \tag{8}$$

So, according to (8), the convex hull of S_k^R is the intersection of two balls, one for the nuclear norm $\|.\|_*$, the other one for the spectral norm (the bound on $\|.\|_{sp}$ is the same in co S_k^R as in S_k^R). Getting at such an explicit form of co S_k^R is due to the happy combination of these specific norms ($\|.\|_{sp}$ and $\|.\|_*$). If $\|.\|$ were any norm on $\mathcal{M}_{m,n}(\mathbb{R})$ and

$$\widehat{S}_k^R := \{A \in \mathcal{M}_{m,n}(\mathbb{R}) | \quad \text{rank } A \leq k \text{ and } \|A\| \leq R\},$$

due to the equivalence between the two norms $\|.\|$ and $\|.\|_{sp}$, we would get with (8) an inner estimate set and an outer estimate set of co $\widehat{S}_k^R$.

A function $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ is called *quasi-convex* if all its sublevel sets

$$[f \leq \alpha] := \{x \in X | \quad f(x) \leq \alpha\} \quad \text{(sublevel set of } f \text{ at the level } \alpha \in \mathbb{R})$$

are convex. Since the supremum of an arbitrary collection of quasiconvex functions is quasiconvex, one can define the largest quasiconvex function lower bounding f; it is called the *quasiconvex hull* of f and denoted by f_q. The way to construct f_q from the sublevel sets of f is easy

$$x \in X \longmapsto f_q(x) = \inf\{\alpha | \quad x \in \text{co}[f \leq \alpha]\}.$$

Following Theorem 6, we therefore get the explicit form of the quasiconvex hull of the (restricted) rank function.

Theorem 7 (Hiriart-Urruty and Le, [25])**.** *We have:*

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto (\mathrm{rank}_R)_q(A) = \begin{cases} \lceil \frac{1}{R}\|A\|_* \rceil & \text{if } \|A\|_{sp} \leq R, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\lceil a \rceil$ stands for the smallest integer which is larger than (or equal to) a.

A final step remains to be carried out: to go from the quasiconvex hull to the convex hull of rank_R. This is not difficult and has been done in [25].

Remark 4. If, in Theorem 4, the restriction "$\|A\|_{sp} \leq R$" is replaced by "$\|A\|_* \leq R$", the result (on the relaxed form) will be the same. But this does not hold true for a restriction "$\|A\| \leq R$" written for any matricial norm $\|.\|$.

Remark 5. Since Theorem 3 (of Eckart and Young) is a precise and powerful theorem solving the best approximation problem for *any* matrix to *any* sublevel-set of the rank function, we thought some time that it would be possible to derive Theorem 4 (on the convex hull of the restricted rank function) from Theorem 3; we did not succeed.

Remark 6. The question of the convex relaxation of the restricted rank function, solved in Theorem 4, could also be posed for tensors (or hypermatrices). Actually, provided that the notion of rank of a tensor and various norms of a tensor have been defined in an appropriate way, it has been proved in [32] that the nuclear norm of a tensor is a convex underestimator of the tensor rank on the unit ball for the Schatten ∞-norm (corresponding to the spectral norm for matrices). The question whether this nuclear norm is the *largest* convex underestimator of the rank (for tensors) was raised by the authors in [32]; they conjectured that was the case. In this context, contrary to the world of matrices, there is no SVD decomposition (for tensors) as a tool at our disposal.

6 Surrogates of the rank function

When A is positive semidefinite, some surrogates for rank A can be defined with the help of trace and determinant functions; these give rise to heuristics in rank minimization problems ([18], Chapter 5). For general matrices, there is no hope to give reliable and good approximations of the (discontinuous) rank function by smooth or even continuous functions involing only norms or inner product of matrices. We nevertheless mention some of them, the first one due to Flores ([19], Chapter II.3).

A proposal of a continuous surrogate of the rank is as follows:

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto g_{sf}(A) := \begin{cases} \dfrac{\|A\|_*^2}{\|A\|_F^2} & \text{if } A \neq 0, \\ 0 & \text{if } A = 0. \end{cases} \tag{9}$$

This function has a close relationship with the rank function, and also shares with it some similar properties.

Theorem 8 (Flores, [19]). *We have:*

(G1) $g_{sf}(cA) = g_{sf}(A)$ *for* $c \neq 0$.

(G2) $1 \leq g_{sf}(A) \leq \operatorname{rank} A$ *for* $A \neq 0$.

(G3) If all the non-zero singular values of A *are equal, then* $g_{sf}(A) = \operatorname{rank} A$.

(G4) $\operatorname{rank} A = 1$ *if and only if* $g_{sf}(A) = 1$.

Properties (G2) and (G4) above allow us to say that, for $A \neq 0$, the rank-one constraint could be replaced by a difference-of-convex (dc) inequality constraint; indeed, if $A \neq 0$,

$$\operatorname{rank} A = 1 \Longleftrightarrow \|A\|_* - \|A\|_F \leq 0. \tag{10}$$

A property like (G4) was also used by Malick ([44]) in SDP problems with rank-one constraints. Let

$$\mathcal{C} := \{A \in \mathcal{S}_n(\mathbb{R}) | \quad A \text{ positive semidefinite}, \operatorname{diag} A = (1, \dots, 1)\}$$

the convex set of the so-called correlation matrices. In ([44], Theorem 1), the following was proved for $A \in \mathcal{C}$:

$$\operatorname{rank} A = 1 \Longleftrightarrow \|A\|_F = n. \tag{11}$$

This property was at the origin of the designation “spherical constraint” as a subtitute for the rank-one constraint. The equality constraint $\|A\|_F^2 = n^2$, although nonlinear, can therefore be dualized; see [44] for more on that.

The underestimator g_{sf} of the rank funtion turns out to look similar to a large class of underestimators parameterized by matrices. The Theorem 9 below, due to Lokam ([42]), deserves to be better known, in our opinion. We provide here a proof for the convenience of the reader.

Theorem 9 (Lokam, [42])**.** *For any non-null $B \in \mathcal{M}_{m,n}(\mathbb{R})$, define*

$$A \in \mathcal{M}_{m,n}(\mathbb{R} \longmapsto g_B(A) := \begin{cases} \dfrac{|\langle\langle A, B\rangle\rangle|}{\|A\|_{sp}\|B\|_{sp}} & \text{if } A \neq 0, \\ 0 & \text{if } A = 0. \end{cases}$$

Then

$$g_B(A) \leq \text{rank } A \text{ for all } A \in \mathcal{M}_{m,n}(\mathbb{R}). \tag{12}$$

Proof. ([42, p.459]) According to a corollary of the Hoffman-Wielandt inequality ([29], Corollary 7.3.8), if $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$ and $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_p$ are the singular values of A and B respectively, one has

$$\sum_{i=1}^{p} (\sigma_i - \tau_i)^2 \leq \|A - B\|_F^2. \tag{13}$$

Developing the left-hand side of (13) gives

$$\sum_{i=1}^{p} \sigma_i^2 + \sum_{i=1}^{p} \tau_i^2 - 2\sum_{i=1}^{p} \sigma_i\tau_i = \|A\|_F^2 + \|B\|_F^2 - 2\sum_{i=1}^{p} \sigma_i\tau_i.$$

Now, if r denotes the rank of A (assumed non-null),

$$\sum_{i=1}^{p} \sigma_i\tau_i = \sum_{i=1}^{r} \sigma_i\tau_i \leq r\|A\|_{sp}\|B\|_{sp}.$$

Since the norm $\|.\|_F$ is derived from the inner product $\langle\langle ., .\rangle\rangle$, developing the right-hand side of (13) gives

$$\|A - B\|_F^2 = \|A\|_F^2 + \|B\|_F^2 - 2\langle\langle A, B\rangle\rangle.$$

Thus, combining all, a consequence of the inequality (13) is

$$\langle\langle A, B\rangle\rangle \leq r\|A\|_{sp}\|B\|_{sp}.$$

Changing B into $-B$ does not change $\|B\|_{sp}$. Whence the desired inequality

$$|\langle\langle A, B\rangle\rangle| \leq r\|A\|_{sp}\|B\|_{sp}$$

is proved. □

Remark 7. If B is chosen to be equal to A, Theorem 9 gives us an underestimator of the rank function,

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto g(A) = \frac{\|A\|_F^2}{\|A\|_{sp}^2},$$

which is of the same family as the g_{sf} function.

Remark 8. If $m = n$ and B is chosen to be the identity matrix, the inequality (12) reads as

$$\frac{|\mathrm{tr}\ A|}{\|A\|_{sp}} \leq \mathrm{rank}\ A. \tag{14}$$

7 Regularization-approximation of the rank function

Besides the underestimates of the rank function proposed in the previous section, it is conceivable -although difficult, as we already said it- to design smooth or just continuous approximations R_ε of the rank fucntion, depending on some parameter $\varepsilon > 0$. What we expect for R_ε is a general comparison result with the rank, for example: $R_\varepsilon(A) \leq \mathrm{rank}\ A$ for all $\varepsilon > 0$, as also a convergence result: $R_\varepsilon \to \mathrm{rank}\ A$ when $\varepsilon \to 0$. We propose here two classes of regularization-approximation of the rank function: the first one consists of smoothed versions of the rank, the second one relies on the so-called Moreau-Yosida technique, widely used in the context of variational analysis and optimization.

7.1 Smoothed versions of the rank function

The underlying idea is as follows: Since

$$\begin{aligned} \mathrm{rank}\ A &= c[\sigma_1(A), \ldots, \sigma_p(A)] \quad \text{(recall that c is the counting function on $\mathbb{R}^p$)} \\ &= \sum_{i=1}^{p} \theta[\sigma_i(A)], \end{aligned} \tag{15}$$

where $\theta(x) = 1$ if $x \neq 0$, $\theta(0) = 0$ (θ is the counting function on $\mathbb{R}$), we have to design some smooth approximation of the θ function. Because all the $\sigma_i(A)$ are nonnegative, θ acts only on the nonnegative part of the real line, so we can choose even approximations of the θ function (like θ which is itself even).

A first example was proposed in [23], it is as following: For $\varepsilon > 0$, let θ_ε be defined as

$$x \in \mathbb{R} \longmapsto \theta_\varepsilon(x) := 1 - e^{-x^2/\varepsilon}. \tag{16}$$

The resulting approximation of the rank function is

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto R_\varepsilon(A) := \sum_{i=1}^{p} [1 - e^{-\sigma_i^2(A)/\varepsilon}]. \tag{17}$$

An alternate expression of the R_ε function is

$$R_\varepsilon(A) = p - \operatorname{tr}(e^{-A^T A/\varepsilon}). \tag{18}$$

Then, R_ε is a C^∞ (even analytic) function of A. The properties of R_ε as an approximation of the rank function are summarized in the statement below.

Theorem 10. *We have*

(i) $R_\varepsilon(A) \leq \operatorname{rank} A$ *for all* $\varepsilon > 0$.

(ii) The sequence of functions $(R_\varepsilon)_{\varepsilon>0}$ *increases when* ε *decreases, and* $R_\varepsilon(A) \to \operatorname{rank} A$ *for all* A *when* $\varepsilon \to 0$.

(iii) If $A \neq 0$ *and* $r = \operatorname{rank} A$,

$$\operatorname{rank} A - R_\varepsilon(A) \leq \varepsilon \sum_{i=1}^{r} \frac{1}{\sigma_i^2(A)}, \tag{19}$$

as also

$$\operatorname{rank} A - R_\varepsilon(A) \leq \varepsilon^2 \sum_{i=1}^{r} \frac{1}{\sigma_i^4(A)}. \tag{20}$$

Sketch of the proof. The results in (i) and (ii) easily follow from the definitions (16) and (17).

The upper bounds (19) and (20), which are not comparable, come from the study of the function $x \mapsto e^{-x}x$ and $e^{-x}x^2$ on the half line $\mathbb{R}^+$. □

We do not know if such an approximation scheme is of any interest for solving rank minimization problems.

Another proposal for approximating the rank function, a quite recent one, is due to Zhao ([51]). It consists of using, for all $\varepsilon > 0$, the following even approximation of the θ function:

$$x \in \mathbb{R} \longmapsto \tau_\varepsilon(x) := \frac{x^2}{x^2 + \varepsilon}. \tag{21}$$

The resulting approximation of the rank function is

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) \longmapsto Z_\varepsilon(A) := \sum_{i=1}^{p} \frac{\sigma_i^2(A)}{\sigma_i^2(A)+\varepsilon}. \tag{22}$$

Alternate expressions of the Z_ε function are:

$$\begin{aligned} Z_\varepsilon(A) &= \operatorname{tr}[A(A^TA+\varepsilon I_n)^{-1}A^T] \\ &= n - \varepsilon \operatorname{tr}(A^TA+\varepsilon I_n)^{-1}. \end{aligned}$$

Here also, Z_ε is a C^∞ (even analytic) function of A. The properties of Z_ε as an approximation of the rank function are summarized in the next statement.

Theorem 11 (Zhao, [51])**.** *We have*

(i) $Z_\varepsilon(A) \leq \text{rank } A$ *for all* $\varepsilon > 0$*.*

(ii) The sequence of functions $(Z_\varepsilon)_{\epsilon>0}$ *increases when* ε *decreases, and* $Z_\varepsilon(A) \to \text{rank } A$ *for all* A *when* $\varepsilon \to 0$*.*

(iii) If $A \neq 0$ *and* $r = \text{rank } A$*,*

$$\text{rank } A - Z_\varepsilon(A) = \sum_{i=1}^{r} \frac{\varepsilon}{\sigma_i^2(A)+\varepsilon} \leq \varepsilon \sum_{i=1}^{r} \frac{1}{\sigma_i^2(A)}. \tag{23}$$

The use of this function Z_ε (instead of the rank function) in rank minimization problems as well as an application to solving a system of quadratic functions are discussed in ([51], Sections 3 and 4).

7.2 Moreau-Yosida approximation-regularization of the rank function

Although the rank function is a bumpy one, it is lower-semicontinuous and bounded from below; it therefore can be approximated-regularized in the so-called Moreau-Yosida way. Surprisingly enough, the rank function is amenable to such an approximation-regularization process, and we get at the end *explicit* forms of the approximated-regularized versions in terms of singular values (like in the previous subsection). Let us firstly recall what is known, as a general rule, for the Moreau-Yosida approximation-regularization technique in a nonconvex context (see [47, Section 1.G] for details, for example).

Let $(E, \|.\|)$ be an Euclidean space, let $f : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a lower-semicontinuous function, bounded from below on E. For a parameter value $\lambda > 0$, the Moreau-Yosida approximate (or Moreau envelope[6]) function f_λ and proximal set-valued mapping $\mathrm{Prox}_\lambda f$ are defined by

$$f_\lambda(x) := \inf_{u \in E} \{f(u) + \frac{1}{2\lambda}\|x - u\|^2\}, \tag{24}$$

$$\mathrm{Prox}_\lambda f(x) := \{\overline{u} \in E | \quad f(\overline{u}) + \frac{1}{2\lambda}\|x - \overline{u}\|^2 = f_\lambda(x)\}. \tag{25}$$

Then:

(i) f_λ is a finite-valued continuous function on E;

(ii) The sequence of functions $(f_\lambda)_{\lambda>0}$ increases when λ decreases, and $f_\lambda(x) \to f(x)$ for all x when $\lambda \to 0$;

(iii) The set $\mathrm{Prox}_\lambda f(x)$ is nonempty and compact.

(iv) The lower bounds of f and f_λ on E are equal:

$$\inf_{x \in E} f(x) = \inf_{x \in E} f_\lambda(x).$$

We now apply this process to the rank function (or restricted rank function). The context is therefore as following: $E = \mathcal{M}_{m,n}(\mathbb{R})$, $\|.\|_F$ is the Frobenius-Schur norm and $f : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ is the rank (or restricted rank) function. By definition,

$$(\mathrm{rank}_R)_\lambda(A) = \inf_{\substack{B \in \mathcal{M}_{m,n}(\mathbb{R}) \\ \|B\|_{sp \leq R}}} \left\{\mathrm{rank}\ B + \frac{1}{2\lambda}\|A - B\|_F^2\right\}. \tag{26}$$

The limiting case, *i.e.* with $R = +\infty$, is

$$(\mathrm{rank})_\lambda(A) = \inf_{B \in \mathcal{M}_{m,n}(\mathbb{R})} \left\{\mathrm{rank}\ B + \frac{1}{2\lambda}\|A - B\|_F^2\right\}. \tag{27}$$

The next two theorems are taken from [26] or [33, Chapter 3].

Theorem 12. *We have, for all $A \in \mathcal{M}_{m,n}(\mathbb{R})$ with rank $r \geq 1$:*

[6] J.-J. Moreau presented this approximation process as an example of inf-convolution or epigraphic addition of two functions; it was in 1962, exactly 50 years ago. This date marks the birth of modern convex analysis and optimization.

(i)

$$(\mathrm{rank})_\lambda(A) = \frac{1}{2\lambda}\|A\|_F^2 - \frac{1}{2\lambda}\sum_{i=1}^{r}[\sigma_i^2(A) - 2\lambda]^+. \tag{28}$$

(ii) One minimizer in (27), i.e. one element in $\mathrm{Prox}_\lambda(\mathrm{rank})(A)$*, is provided by* $B := U\Sigma_B V$*, where*

- $(U, V) \in O(m, n)^A$*, i.e.* U *and* V *are orthogonal matrices such that* $A = U\Sigma_A V$*, with* $\Sigma_A = \mathrm{diag}_{m,n}[\sigma_1(A), \dots, \sigma_r(A), 0, \dots, 0]$ *(a singular value decomposition of* A *with* $\sigma_1(A) \geq \dots \geq \sigma_r(A) > 0$*);*
-

$$\Sigma_B = \begin{cases} 0 & \text{if } \sigma_1 \leq \sqrt{2\lambda}, \\ \Sigma_A & \text{if } \sigma_r(A) \geq \sqrt{2\lambda}, \\ \mathrm{diag}_{m,n}[\sigma_1(A), \dots, \sigma_k(A), 0, \dots, 0] & \text{if there is an integer } k \text{ such that} \\ & \quad \sigma_k(A) \geq \sqrt{2\lambda} > \sigma_{k+1}(A). \end{cases} \tag{29}$$

We may complete the result (ii) in the theorem above by determining explicitly the whole set $\mathrm{Prox}_\lambda(\mathrm{rank})(A)$. Indeed, we have four cases to consider:

- If $\sigma_1(A) < \sqrt{2\lambda}$, then $\mathrm{Prox}_\lambda(\mathrm{rank})(A) = \{0\}$.
- If $\sigma_r(A) > \sqrt{2\lambda}$, then $\mathrm{Prox}_\lambda(\mathrm{rank})(A) = \{A\}$.
- If there is k such that $\sigma_k(A) > \sqrt{2\lambda} > \sigma_{k+1}(A)$, then for any (U_1, V_1) and (U_2, V_2) in $O(m, n)^A$,

$$U_1\mathrm{diag}_{m,n}[\sigma_1(A), \dots, \sigma_k(A), 0, \dots, 0]V_1 = U_2\mathrm{diag}_{m,n}[\sigma_1(A), \dots, \sigma_k(A), 0, \dots, 0]V_2$$

  So, the set $\mathrm{Prox}_\lambda(\mathrm{rank})(A)$ is a singleton and

$$\mathrm{Prox}_\lambda(\mathrm{rank})(A) = \{U\mathrm{diag}_{m,n}[\sigma_1(A), \dots, \sigma_k(A), 0, \dots, 0]V\}$$

  with $(U, V) \in O(m, n)^A$.

- Suppose there is k such that $\sigma_k(A) = \sqrt{2\lambda}$. We then define

$$k_0 := \min\{k| \quad \sigma_k(A) = \sqrt{2\lambda}\},$$

$$k_1 := \max\{k| \quad \sigma_k(A) = \sqrt{2\lambda}\}.$$

Then, $\mathrm{Prox}_\lambda(\mathrm{rank})(A)$ is the set of matrices of the form $U\mathrm{diag}_{m,n}(\tau_1, \dots, \tau_p)V$, where $(U, V) \in O(m, n)^A$ and

$$\tau_i = \sigma_i(A) \text{ if } i < k_0, \tau_i = 0 \text{ if } i > k_1;$$

$$\tau_i = 0 \text{ or } \sigma_i(A) \text{ if } k_0 \leq i \leq k_1.$$

Comments

1. One could wish to express $(\mathrm{rank})_\lambda(A)$ in terms of traces of matrices as this was done for the smoothed versions of the rank function in Section 7.1. Indeed, $A^TA - 2\lambda I_n$ is a symmetric matrix whose eigenvalues are $\sigma_1^2(A) - 2\lambda, \dots, \sigma_r^2(A) - 2\lambda, -2\lambda, \dots, -2\lambda$. Its projection on the cone $\mathcal{S}_n^+(\mathbb{R})$ of positive semidefinite matrices has eigenvalues $[\sigma_1^2(A) - 2\lambda]^+, \dots, [\sigma_r^2(A) - 2\lambda]^+, 0, \dots, 0$ ([24]). Thus, an alternate expression for $(\mathrm{rank})_\lambda(A)$ is:
$$(\mathrm{rank})_\lambda(A) = \frac{1}{2\lambda}\mathrm{tr}(A^TA) - \frac{1}{2\lambda}\mathrm{tr}[P_{\mathcal{S}_n^+(\mathbb{R})}(A^TA - 2\lambda I_n)]. \tag{30}$$

2. If λ is small enough, say if $\sqrt{2\lambda} \leq \sigma_r(A)$, then
$$(\mathrm{rank})_\lambda(A) = \mathrm{rank}\ A.$$
This easily comes from (28) since $\sigma_i^2(A) - 2\lambda \geq 0$ for all i and $\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2(A)$. Therefore, the general convergence result that is known for the Moreau-Yosida approximates f_λ of f is made much stronger here.

3. The counterpart of Theorem 12 for the counting function
$$x = (x_1, \dots, x_p) \in \mathbb{R}^p \longmapsto c(x) = \sum_{i=1}^p \theta(x_i) \quad \text{(see Section 7.1)}$$
is easier to obtain since c has a "separate" structure; it therefore suffices to calculate the Moreau envelope of the θ function. Actually,
   - $c_\lambda(x) = \frac{1}{2\lambda}\|x\|^2 - \frac{1}{2\lambda}\sum_{i=1}^p [x_i^2 - 2\lambda]^+$,
   - One element in $\mathrm{Prox}_\lambda c(x)$ is $p_\lambda(x) \in \mathbb{R}^p$, where
$$[p_\lambda(x)]_i = \begin{cases} x_i & \text{if } |x_i| \geq \sqrt{2\lambda}, \\ 0 & \text{otherwise.} \end{cases}$$

   These results were already observed in Example 5.4 of [2].

As we saw it when considering the relaxed forms of the rank function (in Section 5), what is more useful and interesting for applications is the restricted rank function rank_R. The calculations for its Moreau-Yosida approximates and proximal set-valued mappings are a bit more complicate than for the rank itself, of the same vein however. Here is the final and complete result.

Theorem 13. *We have, for all $A \in \mathcal{M}_{m,n}(\mathbb{R})$ of rank $r \geq 1$:*

(i)

$$(\mathrm{rank}_R)_\lambda(A) = \frac{1}{2\lambda}\|A\|_F^2 - \frac{1}{2\lambda}\sum_{i=1}^{r}\{\sigma_i^2(A) - [(\sigma_i(A) - R)^+]^2 - 2\lambda\}^+. \tag{31}$$

(ii) One minimizer in (31), i.e. one element in $\mathrm{Prox}_\lambda(\mathrm{rank}_R)(A)$, is provided by $B := U\Sigma_B V$ with $\Sigma_B = \mathrm{diag}_{m,n}[\sigma_1(B), \ldots, \sigma_p(B)]$, where

- *$(U, V) \in O(m,n)^A$, i.e. U and V are orthogonal matrices such that $A = U\Sigma_A V$, with $\Sigma_A = \mathrm{diag}_{m,n}[\sigma_1(A), \ldots, \sigma_r(A), 0, \ldots, 0]$ (a singular value decomposition of A with $\sigma_1(A) \geq \cdots \geq \sigma_r(A) > 0$);*
- *If $\sqrt{2\lambda} \geq R$*

$$\sigma_i(B) := \begin{cases} R & \text{if } \sigma_i(A) > \frac{2\lambda+\lambda^2}{2\lambda}, \\ 0 \text{ or } R & \text{if } \sigma_i(A) = \frac{2\lambda+\lambda^2}{2\lambda}, \\ 0 & \text{if } \sigma_i(A) < \frac{2\lambda+\lambda^2}{2\lambda}. \end{cases}$$

- *If $\sqrt{2\lambda} < R$*

$$\sigma_i(B) := \begin{cases} R & \text{if } \sigma_i(A) > R, \\ \sigma_i(A) & \text{if } \sqrt{2\lambda} < \sigma_i(A) \leq R, \\ 0 \text{ or } \sigma_i(A) & \text{if } \sqrt{2\lambda} = \sigma_i(A), \\ 0 & \text{if } \sigma_i(A) < \sqrt{2\lambda}. \end{cases}$$

Comments

1. As the positive parameter λ is supposed to approach 0 in the proximal approximation process, the second case of (ii) in the theorem above is more important than the first one.
2. When $\|A\|_{sp} = \max_{i=1,\ldots,p} \sigma_i(A) \leq R$, both Moreau-Yosida approximates $(\mathrm{rank})_\lambda$ and $(\mathrm{rank}_R)_\lambda$ coincide at A.

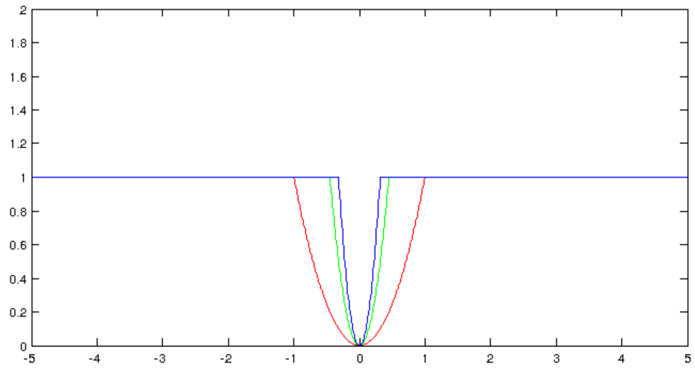

Figure 2: The Moreau-Yosida approximations of the rank.

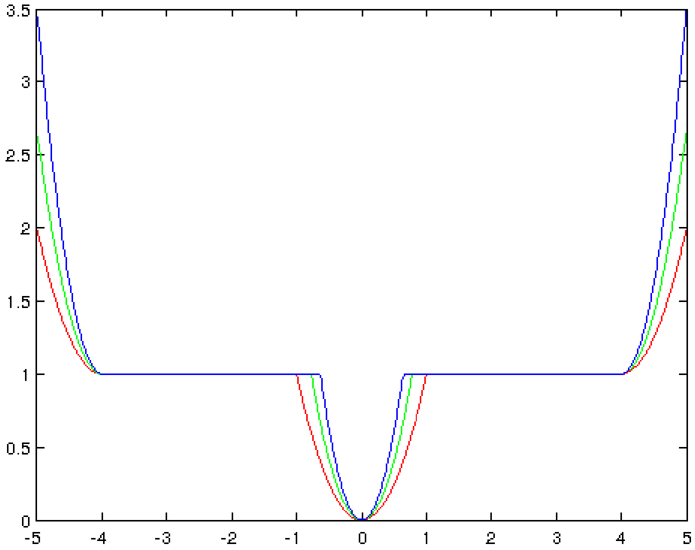

Figure 3: The Moreau-Yosida approximations of the restricted rank. $R = 4$ here.

The pictures in Figure 2 and Figure 3 show, in the one dimensional case, the behaviour of $(\text{rank})_\lambda := c_\lambda$ and $(\text{rank}_R)_\lambda = (c_R)_\lambda$ as $\lambda \to 0$.

We know from Section 5 that the convex relaxed form of the restricted rank function rank_R

is

$$\text{co}(\text{rank}_R) = \psi_R(A) = \begin{cases} \frac{1}{R}\|A\|_* & \text{if } \|A\|_{sp} \leq R, \\ +\infty & \text{otherwise.} \end{cases}$$

It is interesting to calculate explicitly the Moreau-Yosida approximations $(\psi_R)_\lambda$ of ψ_R, and to compare them with those of rank_R in Theorem 13. Here we are in a more familiar convex framework, so that calculations are easier to carry out. Since the proximal set-valued mapping $\text{Prox}_\lambda \psi_R$ is actually single-valued on $\mathcal{M}_{m,n}(\mathbb{R})$, we adopt the notation

$$\text{Prox}_\lambda \psi_R(A) = \{\text{prox}_\lambda \psi_R(A)\}.$$

Theorem 14. *Let U and V be orthogonal matrices such that $A = U\Sigma_A V$, with $\Sigma_A = \text{diag}_{m,n}[\sigma_1(A), \ldots, \sigma_r(A), 0, \ldots, 0]$ (a singular value decomposition of A with $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_r(A) > 0$). We set*

$$p_\lambda^R(A) = (y_1, \ldots, y_p),$$

with

$$y_i = \begin{cases} R & \text{if } \sigma_i(A) \geq \frac{\lambda}{R} + R, \\ \sigma_i(A) - \frac{\lambda}{R} & \text{if } \frac{\lambda}{R} \leq \sigma_i(A) < \frac{\lambda}{R} + R, \\ 0 & \text{if } \sigma_i(A) < \frac{\lambda}{R}. \end{cases} \tag{32}$$

Then, the proximal mapping and the Moreau envelope of ψ_R are described as following.

*(i) [*Proximal mapping*]*

$$\text{prox}_\lambda \psi_R(A) = U \text{diag}_{m,n}(y_1, \ldots, y_p) V; \tag{33}$$

*(ii) [*Moreau envelope*]*

We define, for $t \in \mathbb{R}$,

$$f^i_{\frac{\lambda}{R}}(t) := t^2 - 2\sigma_i(A)t + 2\frac{\lambda}{R}|t|,$$

and for $x = (x_1, \ldots, x_p)$

$$f_{\frac{\lambda}{R}}(x) := \sum_{i=1}^{p} f_i(x_i).$$

Then,

$$(\psi_R)_\lambda(A) = \frac{1}{2\lambda}\|A\|_F^2 - \frac{1}{2\lambda} f_{\frac{\lambda}{R}}[p_\lambda^R(A)]. \tag{34}$$

Moreover, for A such that $\|A\|_{sp} \leq R$,

$$(\psi_R)_\lambda(A) = \frac{1}{2\lambda}\|A\|_F^2 - \frac{1}{2\lambda}\|p_\lambda^R(A)\|^2. \tag{35}$$

The picture below shows, in the one dimensional case, the behaviour of $(\psi_R)_\lambda$ when $\lambda \to 0$, as well as how it compares with $(\mathrm{rank}_R)_\lambda$. It also illustrates the following fact: *the convex hull (or closed convex hull) of* $(\mathrm{rank}_R)_\lambda$ *is exactly* $(\psi_R)_\lambda$.

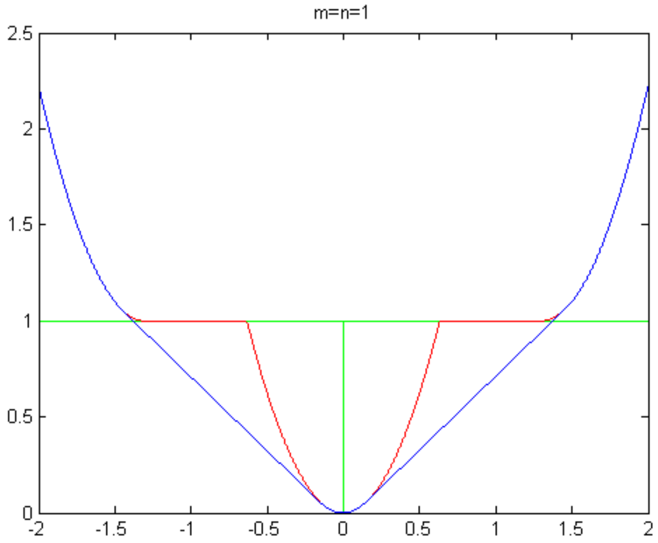

Figure 4: The Moreau-Yosida approximations of the restricted rank and nuclear norm.

The case where $R = 1$ deserves some additional comments. Recalling that

$$\psi_1(A) = \begin{cases} \|A\|_* & \text{if } \|A\|_{sp} \leq 1 \\ +\infty & \text{otherwise,} \end{cases}$$

we have

$$(\psi_1)_\lambda(A) = \frac{1}{2}\|A\|_F^2 - \frac{1}{2} f_\lambda[p_\lambda^1(A)], \tag{36}$$

where $p_\lambda^1(A) = (y_1, \ldots, y_p)$, with

$$y_i = \begin{cases} 1 & \text{if } \sigma_i(A) \geq \lambda + 1, \\ \sigma_i(A) - \lambda & \text{if } \lambda \leq \sigma_i(A) < \lambda + 1, \\ 0 & \text{if } \sigma_i(A) < \lambda. \end{cases} \tag{37}$$

In short,

$$y_i = [\sigma_i(A) - \lambda]^+ - [\sigma_i(A) - (\lambda + 1)]^+ \text{ for all } i = 1, \ldots, p,$$

so that

$$\operatorname{prox}_\lambda \psi_1(A) = U \operatorname{diag}_{m,n}\{([\sigma_i(A) - \lambda]^+ - [\sigma_i(A) - (\lambda + 1)]^+)_i\}V, \tag{38}$$

$$\begin{aligned}(\psi_1)_\lambda(A) &= \frac{1}{2\lambda}\|A\|_F^2 - \frac{1}{2\lambda}\sum_{i=1}^{p} f_\lambda^i(y_i) \\ &= \frac{1}{2\lambda}\sum_{i=1}^{r}\sigma_i^2(A) - \frac{1}{2\lambda}\sum_{i=1}^{r} f_\lambda^i([\sigma_i(A) - \lambda]^+ - [\sigma_i(A) - (\lambda + 1)]^+). \qquad (39)\end{aligned}$$

These formulas (38) and (39) should be put side by side with the expressions of $(\|.\|_*)_\lambda$ and $\operatorname{prox}_\lambda(\|.\|_*)$, such as given in [50] for example:

$$\operatorname{prox}_\lambda(\|.\|_*)(A) = U \operatorname{diag}_{m,n}[\sigma_i(A) - \lambda]^+ V, \tag{40}$$

$$\begin{aligned}(\|.\|_*)_\lambda(A) &= \frac{1}{2}\|A\|_F^2 - \frac{1}{2}\sum_{i=1}^{r}\{([\sigma_i(A) - \lambda]^+)_i\}^2. \\ &= \frac{1}{2}\sum_{i=1}^{r}\sigma_i^2(A) - \frac{1}{2}\sum_{i=1}^{r}\{[\sigma_i(A) - \lambda]^+\}^2. \qquad (41)\end{aligned}$$

As expected, since $\|.\|_* \leq \psi_1$, one has $(\|.\|_*)_\lambda \leq (\psi_1)_\lambda$ for all $\lambda > 0$. Also, for λ small enough, namely for $\lambda \leq \sigma_r(A)$,

$$(\|.\|_*)_\lambda(A) = (\psi_1)_\lambda(A).$$

Note however that the convex relaxed form of $(\text{rank})_\lambda$ is not $(\|.\|_*)_\lambda$; as said before, to compare the relaxed form of the rank function with $\|.\|_*$, as well as their corresponding Moreau-Yosida regularized forms, one has to consider their restricted versions on balls $\{A| \quad \|A\|_{sp \leq R}\}$.

The formulas (40) and (41) are used for designing a proximal point algorithm scheme for nuclear norm minimization ([50]).

8 The generalized subdifferentials of the rank function

Although the rank function is not differentiable, not even continuous, it is possible to calculate the so-called generalized subdifferentials of it (kinds of generalizations of gradients for differentiable functions). These mathematical objects are now much used in nonsmooth analysis and optimization, for theoretical purposes as well as for algorithmic ones. Dealing with the calculation of the generalized subdifferentials of the rank function directly is very difficult because:

- the behaviour of difference quotients like

$$\frac{\text{rank}(A' + tD) - \text{rank } A'}{t}$$

when $A' \to A$ and $t > 0$ goes to 0 is hard to control, and their limits impossible to determine;

- the singular values involved in the regularized versions of the rank function are themselves nonsmooth.

So, we adopted another strategy: to use the relationship between the counting and the rank function, as in [34]. We first calculate the generalized subdifferential(s) of the counting function; then, by using the results of Lewis and Sendov ([39], [40]), we obtain the generalized subdifferential(s) of the rank function. Good news: All the generalized subdifferentials turn out to be equal. Moreover, this (common) generalized subdifferential turns out to be a vector space. Certainly 0 always belongs to it. This was foreseen by the fact that "Every point is a local minimizer of the rank function" (see Section 4). Finally, with the help of an alternate expression of the common subdifferential as a tensor product of vector spaces, we provide its dimension. All the material of this section is excerpted from [35].

8.1 Definitions and preliminary properties

The definitions and properties of generalized subdifferentials have been developed in several works, beginning with the case of locally Lipschitz functions ([11]). Then, they have been generalized for lower-semicontinuous functions. In this section, we only consider the lower semicontinuous case, since it is that of the rank function. Our context is therefore the following:

- $f : E \to \mathbb{R} \cup \{+\infty\}$ is proper, lower-semicontinuous (l.s.c) and $\tilde{x}$ is a point at which f is finite-valued.

- $(E, \langle ., . \rangle)$ is an Euclidean space, for example $\mathbb{R}^p$ equipped with the usual inner product or $\mathcal{M}_{m,n}(\mathbb{R})$ equipped with $\langle\langle ., . \rangle\rangle$.

Definition 1. A vector $x^* \in E$ is a *F-subderivative* of f at $\tilde{x}$ if

$$\liminf_{y \to 0} \frac{f(\tilde{x} + y) - f(\tilde{x}) - \langle x^*, y \rangle}{\|y\|} \geq 0. \tag{42}$$

The set of all F-subderivatives of f at $\tilde{x}$ is called the *Fréchet subdifferential* of f at $\tilde{x}$, and denoted as $\partial^F f(\tilde{x})$.

Definition 2. A vector $x^* \in E$ is a *viscosity subderivative* of f at $\tilde{x}$ if there exists a C^1-function $g : E \to \mathbb{R}$ such that $\nabla g(\tilde{x}) = x^*$ and $f - g$ attains a local minimum at $\tilde{x}$.
If, in particular,

$$g(x) = \langle x^*, x - \tilde{x} \rangle - \sigma \|x - \tilde{x}\|^2$$

with some positive constant σ, then x^* is called a *proximal subgradient* of f at $\tilde{x}$.

The set of all viscosity subderivatives and proximal subgradients of f at $\tilde{x}$ are called the *viscosity subdifferential* and the *proximal subdifferential* of f at $\tilde{x}$ and denoted as $\partial^V f(\tilde{x})$ and $\partial^P f(\tilde{x})$, respectively.

In a finite dimensional context, the Fréchet and the vicosity subdifferentials coincide. And this common subdifferential is also called "regular subdifferential" in some other works (see [39], [40], [47]).

Definition 3. A vector $x^* \in E$ is a *limiting subgradient* of f at $\tilde{x}$ if there is a sequence of points x^ν in E approaching $\tilde{x}$ with values $f(x^\nu)$ approaching the finite value $f(\tilde{x})$, and a sequence of y^ν in $\partial^F f(x^\nu)$ approaching x^*.

The set of all limiting subgradients is called the *limiting subdifferential* and denoted as $\partial^L f(\tilde{x})$.

Definition 4. The *Clarke subdifferential* $\partial^C f(\tilde{x})$ of f at $\tilde{x}$ is the set of all $x^* \in E$ such that

$$\forall v \in E, \quad \langle x^*, v \rangle \leq f^0(\tilde{x}, v) := \lim_{\varepsilon \downarrow 0} \limsup_{\substack{y \downarrow_f \tilde{x} \\ t \downarrow 0}} \inf_{w \in v + \varepsilon B} \frac{f(y + tw) - f(y)}{t}, \tag{43}$$

where B is the unit ball in E and $y \downarrow_f \tilde{x}$ signifies that y and $f(y)$ converge to $\tilde{x}$ and $f(\tilde{x})$, respectively.

Remark 9. We can also define the Clarke subdifferential of f at $\tilde{x}$ as the set of x^* for which $(x^*, -1)$ lie in some appropriate normal cone to epi$f$ at $(\tilde{x}, f(\tilde{x}))$ (see [11]). Thus, $\partial^C f(\tilde{x})$ is closed and convex for every $\tilde{x}$ in E.

Moreover, since we are in a finite dimensional context, we have the next string of inclusions

$$\partial^P f(\tilde{x}) \subset \partial^V f(\tilde{x}) = \partial^F f(\tilde{x}) \subset \partial^L f(\tilde{x}) \subset \partial^C f(\tilde{x}). \tag{44}$$

8.2 The main results

Let us fix some notations:

- As previously, for $x \in \mathbb{R}^p$, let $\mathrm{diag}_{m,n}(x)$ denote an $m \times n$ matrix with entries $\mathrm{diag}_{m,n}(x)^{i,i} = x_i$ for all i, and $\mathrm{diag}_{m,n}(x)^{i,j} = 0$ for $i \neq j$.
- $O(n)$ is the set of orthogonal matrices in $\mathcal{M}_n(\mathbb{R})$.

Theorem 15. *All the generalized subdifferentials (proximal, Fréchet, viscosity, limiting, Clarke) of the rank function coincide. We denote $\partial(\mathrm{rank})$ the common subdifferential set-valued mapping. For $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $\partial(\mathrm{rank})(A)$ is constructed as follows:*

- *Consider the matrices $U \in O(m)$ and $V \in O(n)$ such that*

  $$U\mathrm{diag}_{m,n}(\sigma(A))V = A$$

  (in other words, we collect all the orthogonal matrices U and V which give a singular value decomposition of A; elsewhere in this paper, the set of all such (U, V) is denoted as $O(m, n)^A$).
- *Consider the "pseudo-diagonal" matrices $\mathrm{diag}_{m,n}(x^*)$, where $x^* \in \mathbb{R}^p$ is such that $x_i^* = 0$ for all $i = 1, \dots, r$ (recall that $r = \mathrm{rank}\ A$).*
- *Then, collect all the matrices of the form $U\mathrm{diag}_{m,n}(x^*)V$.*

In a single formula,

$$\begin{aligned}&\partial(\mathrm{rank})(A)\\ &= \{U\mathrm{diag}_{m,n}(x^*)V \mid \quad U \in O(m), V \in O(n) \textit{ such that } U\mathrm{diag}_{m,n}(\sigma(A))V = A,\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad x_i^* = 0 \textit{ for all } i = 1, \dots, r\}.\end{aligned}$$

A further interesting property of $\partial(\mathrm{rank})(A)$ is that it is not only a closed convex set but a vector space.

Proposition 1. *For $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $\partial(\mathrm{rank})(A)$ is a vector space.*

An alternate expression of $\partial(\mathrm{rank})(A)$, a more palpable one, is possible. The subdifferential of the rank function can be also represented as the tensor product of two vector spaces in $\mathbb{R}^m$ and $\mathbb{R}^n$, as indicated in the following proposition.

Proposition 2. *Let $N(A)$ and $N(A^T)$ denote the null spaces of matrices A and A^T, respectively. Then*

$$\partial(rank)(A) = N(A^T) \otimes N(A),$$

where $\otimes$ denotes the tensor product. In a more detailed form,

$$N(A^T) \otimes N(A) = \Big\{ \textstyle\sum_{i,j} a_{ij} \alpha_i \beta_j^T | \quad (\alpha_i) \textit{ is a basis of } N(A^T) \, , \\ (\beta_j) \textit{ is a basis of } N(A) \Big\} \, .$$

Consequentely, the dimension of $\partial(rank)(A)$ is $(m-r)(n-r)$, where r is the rank of A.

We now have a good knowledge of $\partial(\mathrm{rank})$; it is also possible to calculate $\partial(\mathrm{rank}_\lambda)$ and link Sections 7 and 8. The formulas obtained serve as an illustration or example of general results linking the generalized subdifferentials of f and of its Moreau envelope, like those displayed in

- [12, Section 5] on the so-called quadratic inf-convolutions (for example, properties at a point M where $\partial^P(\mathrm{rank}_\lambda)(M)$ is nonempty).
- [47, Section 10.32] on the "subsmoothness" of Moreau envelopes.
- [30, Section 5] on the limits (of various kinds) of the generalized subdifferentials of f_λ when $\lambda > 0$ goes to 0.

Remark 10. In a paper that we came aware of after a first writing of the present paper, LUKE (in [43, Proposition 3.6]) gives the common expression of various generalized normal cones to S_k at $A \in S_k$. Due to the existing relationship between generalized subdifferentials and generalized normal cones, via the indicator function, his results are consistent with our Theorem 15.

Remark 11. The rank function is an example of semi-algebraic function for which every point is a critical point (in the sense that 0 belongs to the (here common) generalized subdifferential of the function at any point). For semi-algebraic extended-real-valued functions defined on $\mathbb{R}^d$, the graphs of the generalized subdifferentials are semi-algebraic subsets of $\mathbb{R}^d \times \mathbb{R}^d$ whose appropriately defined dimensions (local and global) can be defined and studied ([13], [14]). In the case of the rank function defined on $\mathcal{M}_{m,n}(\mathbb{R}) \equiv \mathbb{R}^{m\times n}$, the dimension of the subdifferential graph (both in the local sense and in the global one) is equal to the dimension of the underlying space, namely $d = m \times n$[7].

9 Further notions related to the rank: the spark, the rigidity, the cp-rank of a matrix.

The notion of rank of a matrix dates back to the mathematician Sylvester (1814-1897); there have been since several further concepts whose definitions resemble that of rank, or whose definitions use that of rank. We mention here three of them: the spark, the rigidity and the cp-rank.

9.1 The spark of a matrix

Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$, the *spark* of A is the smallest positive integer k such that there exists a set of k columns of A which are linearly dependent. Remember that the rank of A is the largest number of columns of A which are linearly independent. The term spark seems to have been coined by Donoho and Elad in 2003 (see [10] and [17] for comments on it and some properties); it is also known as the *girth* in matroid theory, or even (within one unity) the *Kruskal rank* (see below).

Actually, the given definition of spark is a bit uncomplete: If A is of full column rank, *i.e.* if rank $A = n$, there is no set of k columns of A which are linearly dependent. In that case, we should adopt $+\infty$ as for the spark of A (the infimum over the empty set).

The other extreme case is when one column of A is a zero-column: then spark $A = 1$. In

[7] D.Drusvyatskiy, personal communication (August 2012).

short, *if A does not contain any zero-column and is not of full column rank,* we have

$$2 \leq \text{spark } A \leq \text{rank } A + 1. \tag{45}$$

Just for an example, consider

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \text{ so that } A^T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Here, $\text{rank}A = \text{rank } A^T = 2$ while spark $A = 2$. Note that spark $A^T = 3$, so that spark $A \neq$ spark A^T as a general rule.

The spark of A is much more difficult to calculate than the rank of A. There however is a variational formulation of spark A (*i.e.*, as the result of some minimization problem). Indeed, to have columns $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ to be linearly dependent means that there exists a non-zero $\lambda = (\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_k}) \in \mathbb{R}^k$ such that $\sum_{l=1}^{k} \lambda_{i_l} A_{i_l} = 0$; in other words

$$A\lambda = 0 \text{ for } \lambda = (0, \ldots, \lambda_{i_1}, \ldots, 0, \ldots, \lambda_{i_k}, \ldots, 0) \in \mathbb{R}^n.$$

To summarize in a single formula,

$$\text{spark } A = \min_{\substack{A\lambda = 0, \\ \lambda \neq 0}} c(\lambda). \tag{46}$$

Problem (46) is an example of rank minimization problem (see the Introduction), where the constraint set $C := \{\lambda \in \mathbb{R}^n, A\lambda = 0, \lambda \neq 0\}$ is nonempty when A is not of full column rank. A related notion is due to Kruskal (1977) who introduced it in the context of fitting trilinear arrays by simple models ([31]); this notion therefore has been called later the *Kruskal rank*, and denoted as K-rank. By definition, the Kruskal rank of A, K-rank A, is the maximal positive integer k such that any subset of k columns of A are linearly independent. For example, K-rank $A = n$ if A is of full column rank, and K-rank $A \geq 1$ except if $A = 0$ (in which case it is not defined). The relation between K-rank and spark is now clear: if $A \neq 0$ is not of full column rank,

$$\text{spark } A = \text{K-rank } A + 1. \tag{47}$$

9.2 The matrix rigidity

Suppose $m = n$. Then, the *rigidity* of $A \in \mathcal{M}_n(\mathbb{R})$, denoted by $R_A(k)$, is the smallest number of entries that need to be changed in order to reduce the rank of A below k. This notion take roots in computer science ([49]), and obtaining bounds on rigidity has a number of implications in complexity theory ([42]). However, computing the rigidity of a matrix is in general a NP-hard problem. For $A \in \mathcal{M}_n(\mathbb{R})$ and any integer k, one can check that

$$R_A(k) \leq (n-k)^2. \tag{48}$$

A variational formulation of $R_A(k)$ can easily be proposed as a rank minimization problem:

$$R_A(k) = \min_{M \in A + S_k} c(M), \tag{49}$$

where $c(M)$ denotes the number of nonzero entries in the matrix M (or the counting number of $(m_{11}, \ldots, m_{nn})$) and $S_k = \{S \in \mathcal{M}_n(\mathbb{R}) | \quad \text{rank } S \leq k\}$.

9.3 The cp-rank of a matrix

9.3.1 Definition and first properties

The cp-rank is defined for a specific class of symmetric matrices called *completely positive* matrices; the prefix "cp" is precisely there for recalling the wording "completely positive". Before defining completely positive matrices, we recall what copositive matrices are.

A matrix $M \in \mathcal{S}_n(\mathbb{R})$ is said to be copositive if

$$\langle Mx, x \rangle \geq 0 \text{ for all } x \in \mathbb{R}^n_+.$$

The set of all copositive matrices is a closed convex cone of $\mathcal{S}_n(\mathbb{R})$ which we denote by $C_n(\mathbb{R})$. A matrix $A \in \mathcal{S}_n(\mathbb{R})$ is then called *completely positive* if

$$\langle\langle A, M \rangle\rangle \geq 0 \text{ for all copositive matrices } M. \tag{50}$$

Thus, the set of all completely positive matrices is again a closed convex cone of $\mathcal{S}_n(\mathbb{R})$; we denote it by $CP_n(\mathbb{R})$. Hence, using the terminology of convex analysis, $CP_n(\mathbb{R})$ is the (positive) polar or dual cone of $C_n(\mathbb{R})$. Expressed in other ways,

$$\begin{aligned} CP_n(\mathbb{R}) &= \text{co}\{xx^T | \quad x \in \mathbb{R}^n_+\} && (51) \\ &= \{BB^T | \quad B \text{ is a nonnegative matrix}\}. && (52) \end{aligned}$$

A completely positive matrix A necessarily has nonnegative entries and is positive semidefinite.

Definition 5. The smallest number of columns of a nonnegative matrix B in a factorization $A = BB^T$ of a completely positive matrix A is called the cp-rank of A and denoted as cp-rank A.

The cp-rank of A is also the minimal number k of summands in a rank 1-decomposition of A as follows:

$$A = \sum_{i=1}^{k} b_i b_i^T, \text{ with } b_i \in \mathbb{R}_+^n. \tag{53}$$

For the zero matrix (which indeed lies in the cone $CP_n(\mathbb{R})$, we adopt by convention that its cp-rank is 0.

For matrices $A \in \mathcal{S}_n(\mathbb{R})$ which are not in $CP_n(\mathbb{R})$, a decomposition like (53) is not possible; as it is usual in variational analysis, we therefore pose cp-rank $A = +\infty$.

More information about copositive matrices and completely positive matrices can be found in survey papers like [28], [15], [9].

If one wished to compare the rank and the cp-rank of a matrix, one gets at the following inequality:

$$\text{rank } A \leq \text{cp-rank } A, \tag{54}$$

for all completely positive matrices, hence for all symmetric matrices.

Indeed, in some cases, the cp-rank of A is equal to the rank of A; for example this holds true when $n \leq 3$ or when rank $A \leq 2$. To give an example where the inequality in (54) is strict, consider

$$A = \begin{bmatrix} 6 & 3 & 3 & 0 \\ 3 & 5 & 1 & 3 \\ 3 & 1 & 5 & 3 \\ 0 & 3 & 3 & 6 \end{bmatrix} \in \mathcal{S}_4(\mathbb{R}).$$

Then, rank $A = 3$, while cp-rank $A = 4$ ([4, page 140]).

The cp-rank however enjoys some properties similar to those of the rank (see Section 1 and Section 2).

Theorem 16. *(i) If A and B are completely positive,*

$$\text{cp-rank } (A + B) \leq \text{cp-rank } A + \text{cp-rank } B. \tag{55}$$

(ii) If A is completely positive and $c > 0$,

$$\text{cp-rank } (cA) = \text{cp-rank } A. \tag{56}$$

(iii) The cp-rank is a lower-semicontinuous function: If $(A_\nu)_\nu$ is a sequence of completely positive matrices converging to A, then A is completely positive and

$$\liminf_{\nu \to +\infty} \text{cp-rank } A_\nu \geq \text{cp-rank } A. \tag{57}$$

9.3.2 Upper bounds for the cp-rank.

One of the most interesting challenges about the cp-rank is to provide an upper bound of it in terms of rank. In 1983, Hannay and Laffey ([21]) showed that the cp-rank of a completely positive matrix of rank k was necessarily less than or equal to $\frac{k(k+1)}{2}$. Later, in 2003, this upper bound was sharpened to $\frac{k(k+1)}{2} - 1$ by Barioli and Berman ([7]); they also proved that this upper bound could be achieved by a completely positive matrix of rank k. In short,

Theorem 17. *(i) For every completely positive matrix of rank $k \geq 2$,*

$$\text{cp-rank } A \leq \frac{k(k+1)}{2} - 1.$$

(ii) For every integer $k \geq 2$, there exists a completely positive matrix A whose rank is k and whose cp-rank is $\frac{k(k+1)}{2} - 1$.

By Theorem 17, if A is a $n \times n$ completely positive matrix, then

$$\text{cp-rank } A \leq \frac{n(n-1)}{2} - 1. \tag{58}$$

But is there a better upper bound on the cp-rank of $n \times n$ matrices? Drew, Johnson and Loewy proved that cp-rank $A \leq \frac{n^2}{4}$ for every completely positive matrix of order $n \geq 4$ whose graph is triangle free ([16]). The fact that the bound $\frac{n^2}{4}$ was also valid in all other known cases led the authors to wonder whether this holds for every completely positive matrix of order $n \geq 4$.

Conjecture (The DJL Conjecture) If A is an $n \times n$ completely positive matrix, with $n \geq 4$, then

$$\text{cp-rank } A \leq \frac{n^2}{4}. \tag{59}$$

The conjecture was proved by Berman and Shaked-Monderer ([3]) for matrices whose comparison matrices are M-matrices and by Loewy and Tam ([41]) for 5×5 matrices whose graph is not complete. But then, for the first time, Barioli announced (in [6]) an example of 7×7 completely positive matrix of rank 5 and cp-rank 14. Such matrix is a counter-example to the DJL Conjecture.

9.3.3 The convex relaxed form of the cp-rank.

Due to properties like (56), it is easy to see that the convex hull of the cp-rank function on $CP_n(\mathbb{R})$ is the zero function. So, like for the rank function (in Section 5), we restrict it to some appropriate ball. For $R > 0$, consider

$$A \in \mathcal{S}_n(\mathbb{R}) \mapsto \text{cp-rank}_R(A) := \begin{cases} \text{cp-rank } A & \text{if } A \text{ is completely positive and } \|A\|_* \leq R; \\ +\infty & \text{otherwise.} \end{cases} \tag{60}$$

We then have the following result, to be put in parallel with Theorem 4.

Theorem 18. *We have*

$$A \in \mathcal{S}_n(\mathbb{R}) \mapsto \text{co}(\text{cp-rank}_R)(A) := \begin{cases} \frac{1}{R}\|A\|_* & \text{if } A \text{ is completely positive and } \|A\|_* \leq R; \\ +\infty & \text{otherwise.} \end{cases} \tag{61}$$

Before going into the details of the proof, some observations are in order:

- Like for Theorem 4, due to the homogeneity properties of the functions involved, it suffices to prove Theorem 18 for $R = 1$.
- When a matrix A is positive semidefinite, which is the case for A completely positive,

$$\|A\|_* = \text{sum of the eigenvalues of } A = \text{tr } A. \tag{62}$$

Proof. (of Theorem 18)

Let ϕ denote the function defined in the right-hand side of (61) for $R = 1$, proposed as the convex hull of the function cp-rank$_1$. The domain of the function cp-rank$_1$, *i.e.* the set of matrices where it is finite-valued, and that of ϕ are equal: it is the compact convex set

$$CP_n(\mathbb{R}) \cap \{A \in \mathcal{S}_n(\mathbb{R}) | \quad \|A\|_* \leq 1\}.$$

So, if $A \in \mathcal{S}_n(\mathbb{R})$ lies out of the above set, the function ϕ and co(cp-rank$_1$) coincide at A, their common value is $+\infty$.

Let therefore now A be chosen completely positive, with $\|A\|_* \leq 1$. Firstly, if A is the zero matrix, it is clear that

$$\text{co}(\text{cp-rank}_1)(0) = 0 = \phi(0).$$

Secondly, let us assume that $A \neq 0$. We have to prove that

$$(\text{cp-rank}_1)(A) = \|A\|_*.$$

Because

$$\text{cp-rank} \geq \text{rank} \geq \|.\|_* \text{ on } \{A \in \mathcal{S}_n(\mathbb{R}) | \quad \|A\|_* \leq 1\},$$

we get a first inequality

$$\text{co}(\text{cp-rank}_1)(A) \geq \|A\|_*. \tag{63}$$

To get at the converse inequality, decompose A as follows:

$$A = \sum_{i=1}^{k} b_i b_i^T, b_i \in \mathbb{R}_+^n, b_i \neq 0, k \geq \text{cp-rank } A.$$

We "normalize" the decomposition above as:

$$A = \sum_{i=1}^{k} \alpha_i c_i c_i^T, \text{ with } c_i := \frac{b_i}{\|b_i\|} \text{ and } \alpha_i := \|b_i\|^2. \tag{64}$$

Now,

$$\text{tr } A = \sum_{i=1}^{k} \alpha_i \text{tr}(c_i c_i^T) = \sum_{i=1}^{k} \alpha_i \quad (\text{because tr}(c_i c_i^T) = \|c_i\|^2 = 1),$$

so that

$$\sum_{i=1}^{k} \alpha_i \leq 1 \quad (\text{because tr } A = \|A\|_* \leq 1).$$

We complete the decomposition (64) with the zero matrix:

$$A = \sum_{i=1}^{k} \alpha_i c_i c_i^T + (1 - \sum_{i=1}^{k} \alpha_i)0.$$

Thus, due to the convexity of co(cp-rank$_1$),

$$\begin{aligned}
\text{co(cp-rank}_1)(A) &\leq \sum_{i=1}^{k} \alpha_i \text{cp-rank}_1(c_i c_i^T) + (1 - \sum_{i=1}^{k} \alpha_i)\text{cp-rank}_1(0) \\
&\leq \sum_{i=1}^{k} \alpha_i \quad (\text{because cp-rank}_1(c_i c_i^T) = 1 \text{ and cp-rank}_1(0) = 0) \\
&\leq \text{tr } A = \|A\|_*. \qquad (65)
\end{aligned}$$

Putting together the inequalities (63) and (65) yield the desired result. □

Remark 12. As for the rank function, there is no difference between co(rank$_R$) and its closure $\overline{\text{co}}$(cp-rank$_R$).

Remark 13. On the set where both co(rank$_R$) and co(cp-rank$_R$) are finite, *i.e.* on $CP_n(\mathbb{R}) \cap \{A \in \mathcal{S}_n(\mathbb{R}) | \quad \|A\|_* \leq R\}$, they are *equal* and *linear*.

Like for the spark of A (*cf.* 9.1), the cp-rank of A is difficult to determine, much more than the rank of A.

Since the cp-rank is a lower-semicontinuous integer-valued function, like the rank function, it is tempting to ask ourselves: what are the generalized subdifferentials of the cp-rank function? The question was answered in Section 8 for the rank function, we did not succeed in answering it for the cp-rank function.

By way of conclusion

In this survey, we have painted a broad panorama of the main properties of the rank function (and of some related notions) in the context of variational analysis and optimization. The algorithmic aspects of tackling the rank minimization problems concern the convex relaxed forms of the rank function (such as seen in Section 5); there is indeed a huge specific literature devoted to this way of doing. It remains to see if, even if we deal with a very bumpy (but fascinating) function, the variational analysis properties of the rank function itself could be used for considering directly the rank minimization problems and not their relaxed forms.

Acknowledgments. We thank the various readers or referees who, after the first circulation of this survey paper (July 2012), provided comments, improvements and additional references.

Last but not least, we would like to thank Professor M.Goberna (University of Alicante) for his instrumental role in the publication of this survey paper.

References

[1] P.-A.Absil and J.Malick, *Projection-like retractions on matrix manifolds.* SIAMJ. Optim.22, No 1(2012), 135-158.

[2] H.Attouch, J.Bolte and B.F.Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel method*, Mathematical Programming (DOI: 10.1007/s10107-011-0484-9)(2011).

[3] A.Berman and N.Shaked-Monderer, *Remarks on completely positive matrices*, Linear and Multilinear Algebra, Vol 44(2) (1998), 149-163.

[4] A.Berman and N.Shaked-Monderer, *Completely positive matrices.* World Scientific Publishing Co. (2003).

[5] D.S. Bernstein, *Matrix mathematics: theory, facts, and formulas*, Princeton University Press (2005).

[6] F.Barioli, *Completely positive matrices of small and large accute sets of vectors*, Talk at the 10th ILAS conference in Auburn (2002).

[7] F.Barioli and A.Berman, *The maximal cp-rank of rank k completely positive matrices*, Linear Algebra and its Applications, Vol 363 (2003), 17-33.

[8] E. Behrends, S.Geschke and T.Natkaniec, *Functions for which all points are local extrema*, Real Anal. Exchange Vol 33(2) (2007), 467-470.

[9] I.Bomze, *Copositive optimization - recent developments and applications*, European Journal of Operational Research 216 (2012), 509-520.

[10] A.M.Bruckstein, D.L.Donoho and M.Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Review Vol 51(1) (2009), 34-81.

[11] F.H.Clarke, *Optimisation and nonsmooth analysis*, Wiley, NewYork (1983).

[12] F.H.Clarke, Y.S.Ledyaev, R.T.Stern and P.R.Wolenski, *Nonsmooth analysis and control theory*, Springer (1998).

[13] D.Drusvyatskiy, A.D.Ioffe and A.S.Lewis, *The dimension of semialgebraic subdifferential graphs*, Nonlinear Analysis 75(2012), 1231-1245.

[14] D.Drusvyatskiy and A.S.Lewis, *Semi-algebraic functions have small subdifferentials*, to appear in special issue of Mathematical Programming Ser.B in honor of C.Lemaréchal (2013).

[15] M.Dür, *Copositive Programming - a survey*, Recent Advances in Optimization and its Applications in Engineering 2010, Part 1, 3-20, DOI: 10.1007/978-3-642-12598-0-1.

[16] J.H.Drew, C.R.Johnson and R.Loewy, *Completely positive matrices associated with M-matrices*, Linear and Multilinear Algebra, Vol 37 (1994), 303-310.

[17] M.Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*, Springer (2010).

[18] M.Fazel, *Matrix rank minimization with applications*, PhD thesis, Stanford University (2002).

[19] S.Flores, *Problèmes d'optimisation globale en statistique robuste*, PhD thesis, Paul Sabatier University, Toulouse (2011).

[20] N.Higham, *Matrix nearness problems and applications*, In M.J.C Gover and S.Barnett, editors, *Applications of Matrix Theory*, Oxford University Press (1989), 1-27.

[21] J.Hanna and T.J.Laffey, *Nonnegative factorization of completely positive matrices*, Linear Algebra and Its Applications, Vol 55 (1983), 1-9.

[22] J.-B.Hiriart-Urruty, *When only global optimization matters*, J. of Global Optimization (DOI: 10.1007/s10898-011-9826-7) (2012).

[23] J.-B.Hiriart-Urruty, *Deux questions de rang*, Working Note, Paul Sabatier University (2009).

[24] J.-B.Hiriart-Urruty and J.Malick, *A fresh variational analysis look at the world of the positive semidefinite matrices*, J. of Optimization Theory and Applications, Vol 153(3) (2012), 551-577.

[25] J.-B.Hiriart-Urruty and H.Y.Le, *Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function*, Optimization Letters (DOI: 10.1007/s11590-011-0304-4)(2011).

[26] J.-B.Hiriart-Urruty and H.Y.Le, *From Eckart & Young approximation to Moreau envelopes and vice versa.* Preprint 2012. Submitted.

[27] J.-B.Hiriart-Urruty and C.Lemaréchal, *Convex analysis and minimization algorithms*, Springer (1993).

[28] J.-B.Hiriart-Urruty and A.Seeger, *A variational approach to copositive matrices*, SIAM Review Vol 54(4) (2010), 593-629.

[29] R.A.Horn and C.R.Johnson, *Matrix analysis*, Cambridge University Press (1975).

[30] A.Jourani, L.Thibault and D.Zagrodny, *Differential properties of Moreau envelope*, Preprint (2012).

[31] J.B.Kruskal *Three-way arrays: Rank and uniqueness of trilinear decomposition, with application to arithmetic complexity and statistics*, Linear Algebra and Its Applications, Vol 18(2) (1977), 95-138.

[32] L.-H. Lim and P.Common, *Multiarray signal processing: Tensor decomposition meets compressed sensing*, Preprint (2010).

[33] H.Y.Le, Ph.D Thesis, Paul Sabatier University. To be defended in 2013.

[34] H.Y.Le, *Confexifying the counting function on* $\mathbb{R}^p$ *for convexifying the rank function on* $\mathcal{M}_{m,n}(\mathbb{R})$, J. of Convex Analysis, Vol 19(2) (2012).

[35] H.Y.Le, *The generalized subdifferentials of the rank function*, Optimization Letters (2012), DOI: 10.1007/s11590-012-0456-x.

[36] A.S.Lewis, *Convex analysis on the Hermitian matrices*, SIAM Journal on Optimization, Vol 6 (1996), 164-177.

[37] A.S.Lewis, *The convex analysis of unitarily invariant matrix functions*, J. of Convex Analysis, Vol 2(1) (1995), 173-183.

[38] A.S.Lewis and J.Mallick, *Alternating projections on manifolds.* Math. Oper. Res 33(2008), 216-234.

[39] A.S.Lewis and H.S.Sendov, *Nonsmooth analysis of singular values. Part I: Theory.* Set-Valued Analysis, Vol 13 (2005), 213-241.

[40] A.S.Lewis and H.S.Sendov, *Nonsmooth analysis of singular values. Part II: Applications.* Set-Valued Analysis, Vol 13 (2005), 243-264.

[41] R.Loewy and B.-S.Tam, *CP-rank of completely positive matrices of order* 5, Linear Algebra and its Applications, Vol 363 (2003) , 161-176.

[42] S.V.Lokam, *Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity*, J. of Computer and System Sciences 63 (2001), 449-473.

[43] D.R.Luke, *Prox-regularity of rank contraint sets and implications for algorithms.* Preprint 2012. To appear in the Journal of Mathematical Imaging and Vision.

[44] J.Malick, *The spherical constraint in Boolean quadratic programs*, J. of Global Optimization Vol 39 (2007), 609-622.

[45] P.A.Parrilo, *The convex algebraic geometry of rank minimization*, Plenary talk at the International Symposium on Mathematical Programming, Chicago (2009).

[46] B.Recht, M.Fazel and P.A.Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Review Vol 52(3) (2010), 471-501.

[47] R.T.Rockafellar and R.J-B.Wets, *Variational analysis*, Springer (1998).

[48] G.W.Stewart, *On the early history of the singular value decomposition*, SIAM Review Vol 35(4) (1993), 551-556.

[49] L.G.Valiant, Graph theoretic arguments in low-level complexity, Lecture Notes in Computer Science, Springer, Berlin, Vol 53 (1977), 162-176.

[50] Y.-J.Liu, D.Sun and K.-C.Toh, *An implementable proximal point algorithmic framework for nuclear norm minimization*, Mathematical Programming, Vol 133 (2012), 399-436.

[51] Y.-B.Zhao, *An approximation theory of matrix rank minimization and its application to quadratic equations*, Linear Algebra and Its Applications (2012), 77-93.