The viscosity subdifferential of the rank function via the corresponding subdifferential of its Moreau envelopes

Jean-Baptiste Hiriart-Urruty[1] and Hai Yen Le[2]

Dedicated to Lionel Thibault at the occasion of his 65^{th} birthday and the honorary degree (doctorate honoris causa) conferred by the university of Chili at Santiago.

Abstract

We derive the so-called viscosity subdifferential of the rank function via a limiting process applied to the Moreau envelopes of the rank function. Before that, we obtain the explicit expressions of all the generalized subdifferentials of the Moreau envelopes of the rank function.

Keywords. Rank function, Singular values of a matrix, Moreau envelopes, Generalized subdifferentials, Viscosity subdifferential.
2010 Mathematics Subject Classification. 15A, 46N10, 65K10, 90C.

1 Introduction

The rank of a matrix is a basic notion in matricial calculus. The so-called *rank minimization problems* (*i.e.*, problems where the rank function appears as an objective function or as a constraint) are a hot subject in modern optimization. However, the rank function is a very "bumpy" one, it is just lower-semicontinuous (as a function of matrices). The questions are thus: What kind of generalized differentiability could we expect for it? What are its generalized subdifferentials? These questions were answered by H.Y.Le in [7]. Here we propose another approach or strategy to calculate the generalized viscosity (or Fréchet) subdifferential of the rank function: first calculate the generalized viscosity subdifferential of the Moreau envelopes of the rank function, then use a limiting process and a theorem by Jourani ([6]). In doing so, we also calculate all the other generalized subdifferentials (proximal, viscosity, Fréchet, limiting, Clarke) of the Moreau envelopes of the rank function.

The plan of our paper is as follows: In Section 2, we give all the preliminaries: the singular value decomposition of a matrix, the Moreau envelopes of the rank function,

[1]Institute of Mathematics, Paul Sabatier University, Toulouse (France)
http://www.math.univ-toulouse.fr/~jbhu/
Email: jbhu@math.univ-toulouse.fr
[2]Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi (Vietnam)
Email: lhyen@math.ac.vn

the various definitions of the generalized subdifferentials of a discontinuous function. In Section 3, we determine all the generalized subdifferentials of the Moreau envelopes of the rank function. This is done in an indirect way: we first determine the corresponding generalized subdifferentials of the so-called counting function on $\mathbb{R}^p$, and then apply results by Lewis and Sendov ([8],[9]) about the nonsmooth analysis of functions of singular values. In Section 4, we finally derive the generalized viscosity subdifferential of the rank function.

For general results on the rank function from the variational viewpoint, we refer the reader to our survey paper [5].

2 Preliminaries

2.1 Moreau envelopes of the rank in terms of singular values

Let $\mathcal{M}_{m,n}(\mathbb{R})$ denote the set of real matrices with m columns and n rows, let $p = \min(m, n)$. We consider the rank function

$$\begin{array}{rccl} \text{rank}: & \mathcal{M}_{m,n}(\mathbb{R}) & \longrightarrow & \{0, 1, \ldots, p\} \\ & A & \mapsto & \text{rank } A (= \text{ rank of the matrix } A). \end{array}$$

The rank function can also be defined as a function of singular values of a matrix. For that, we firstly recall here a technique of decomposition of matrices which is central in numerical matricial analysis and statistics: the singular value decomposition (SVD in short).

Theorem 1. *For $A \in \mathcal{M}_{m,n}(\mathbb{R})$, there exist an (m, m) orthogonal matrix U, an (n, n) orthogonal matrix V and a (m, n) "pseudo-diagonal"* [3] *matrix Σ with nonnegative real numbers on the diagonal, such that:*

$$A = U \Sigma V^T.$$

The nonnegative real numbers on the diagonal of the Σ matrix are the so-called singular values $\sigma_1(A), \sigma_2(A), \ldots, \sigma_p(A)$ of A. They are the square roots of the eigenvalues of the symmetric matrix $A^T A$ (or AA^T). Without loss of generality, we can suppose that

$$\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_p(A).$$

If we denote by $c(x)$ the number of non-zero components x_i of a vector $x = (x_1, x_2, \ldots, x_p)$ in $\mathbb{R}^p$, then the rank of A is exactly given by

$$\text{rank } A = c(\sigma_1(A), \sigma_2(A), \ldots, \sigma_p(A)).$$

In some sense, the rank function in the "matricial cousin" of the c-function. They share many properties such as lower-semicontinuity, sub-additivity, etc. In some papers, the c-function is called the 0-norm (although it is not a norm) and denoted by $\|.\|_0$. In our present note, we call c the *counting function*.

[3] Σ "pseudo-diagonal" means that $\Sigma_{ij} = 0$ for $i \neq j$.

In order to solve the rank minimization problems, several underestimates of the rank function have been proposed in recent years ([2], [5], [13], etc.). In this note, we only consider one way of approximating the rank function, the "most variational one": using the approximation-regularization technique of Moreau. All the details of this way of doing and resulting expressions can be found in [4]. For $\varepsilon > 0$, the Moreau envelope rank_ε of the rank function is defined as

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

where $\|.\|_F$ denotes the Frobenius matricial norm (actually, $\|A\|_F^2 = \text{tr}(A^T A) = \sum_{i=1}^p \sigma_i^2(A)$). The result of the minimization problem (1) is

$$\text{rank}_\varepsilon(A) = \frac{1}{\varepsilon} \|A\|_F^2 - \frac{1}{\varepsilon} \sum_{i=1}^p \left[\sigma_i^2(A) - \varepsilon \right]^+, \tag{2}$$

where $[a]^+ = \max(a, 0)$.

The Moreau envelopes of the counting function c are much easier to compute explicitly: for $x = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$ and $\varepsilon > 0$,

$$c_\varepsilon(x) = \frac{1}{\varepsilon} \|x\|^2 - \frac{1}{\varepsilon} \sum_{i=1}^p \left[x_i^2 - \varepsilon \right]^+. \tag{3}$$

From (2) and (3), it is clear that, denoting the vector $(\sigma_1(A), \sigma_2(A), \ldots, \sigma_p(A))$ as $\sigma(A)$,

$$\text{rank}_\varepsilon(A) = c_\varepsilon(\sigma(A)). \tag{4}$$

This formula will be our key-ingredient for our Section 3.

2.2 The various generalized subdifferentials of a discontinuous function

We recall here various notions of generalized subdifferentials of a discontinuous (in fact lower-semicontinuous) function: the proximal subdifferential, the viscosity subdifferential, the Fréchet subdifferential, the limiting subdifferential, the Clarke subdifferential. The concepts appear under various names in the literature, witness the three most recent books defining and using them ([1], [3], [10]). Fortunately, in our context, all the different notions will boil down more or less to the same mathematical object.

Let $f : \mathbb{R}^p \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a lower-semicontinuos (l.s.c) function, and let $\tilde{x} \in \mathbb{R}^p$ be a point at which f is finite.

Definition 1. A vector $x^* \in \mathbb{R}^p$ is called a *F-subgradient* of f at $\tilde{x}$ if

$$\liminf_{d \to 0} \frac{f(\tilde{x} + d) - f(\tilde{x}) - \langle x^*, d \rangle}{\|d\|} \geq 0. \tag{5}$$

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

A more palpable notion is given in the next definition.

Definition 2. A vector $x^* \in \mathbb{R}^p$ is called a *viscosity subgradient* of f at $\tilde{x}$ if there exists a C^1-function $g : \mathbb{R}^p \to \mathbb{R}$ such that $\nabla g(\tilde{x}) = x^*$ and $f \geq g$ in a neighborhood of $\tilde{x}$.

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 subgradients 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.

It turns out that in a finite dimensional context (which is the case in our paper), the Fréchet and the viscosity subdifferentials coincide; but both definitions (Definition 1 and Definition 2) are useful in calculations. This common subdifferential may bear other names in the literature, for example, it is called "regular subdifferential" in some works ([8], [9], [12]).

A further notion, defined via the previous ones, is that of limiting subgradient.

Definition 3. A vector $x^* \in \mathbb{R}^p$ is called a *limiting subgradient* of f at $\tilde{x}$ if one can find a sequence of points (x^ν) converging to $\tilde{x}$ with values $f(x^\nu)$ converging to $f(\tilde{x})$, and a sequence of (Fréchet subgradients) $x^{*\nu} \in \partial^F f(x^\nu)$ converging to x^*.

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

Finally, the most complicated one to define, but also one of the most useful ones in variational analysis and optimization, is Clarke's subdifferential.

Definition 4. A vector $x^* \in \mathbb{R}^p$ is called a *Clarke subgradient* of f at $\tilde{x}$ if

$$\langle x^*, d \rangle \leq f^o(\tilde{x}, d) \quad \text{for all } d \in \mathbb{R}^p,$$

where

$$f^o(\tilde{x}, d) = \lim_{\varepsilon \to 0} \limsup_{\substack{x \downarrow_f \tilde{x} \\ t \to 0}} \inf_{d' \in d + \varepsilon B} \frac{f(x + td') - f(x)}{t} \tag{6}$$

(called Clarke's generalized directional derivative of f at $\tilde{x}$ in the d direction), where B is the unit ball in $\mathbb{R}^p$ and $x \downarrow_f \tilde{x}$ means that x converges to $\tilde{x}$ and $f(x)$ converges to $f(\tilde{x})$.

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

In short, to compare all these generalized subdifferentials, we have the following string of inclusions:

$$\partial^P f(\tilde{x}) \subset \partial^F f(\tilde{x}) = \partial^V f(\tilde{x}) \subset \partial^L f(\tilde{x}) \subset \partial^C f(\tilde{x}). \tag{7}$$

Calculus rules on the various generalized subdifferentials presented above are well discussed in the literature, for example in the following books ([11], [12]). We just recall here some of them, to be used in the next sections.

Theorem 2 (Adding a C^1 function)**.** *Suppose that f_1 is lower-semicontinuous and that f_2 is continuously differentiable in a neighborhood of $\tilde{x}$. Then*

$$\partial^F(f_1 + f_2)(\tilde{x}) = \partial^F f_1(\tilde{x}) + \nabla f_2(\tilde{x}),$$

$$\partial^L(f_1 + f_2)(\tilde{x}) = \partial^L f_1(\tilde{x}) + \nabla f_2(\tilde{x}),$$

$$\partial^C f_1 + f_2)(\tilde{x}) = \partial^C f_1(\tilde{x}) + \nabla f_2(\tilde{x}).$$

Theorem 3 (Subdifferentiation of a separable function)**.** *Let f be defined as $f(x) = f_1(x_1) + \cdots + f_p(x_p)$ for some lower-semicontinuous functions $f_i : \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}$, where $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$. Then, at $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_p)$,*

$$\partial^F f(\tilde{x}) = \partial^F f_1(\tilde{x}_1) \times \cdots \times \partial^F f_p(\tilde{x}_p),$$

$$\partial^L f(\tilde{x}) = \partial^L f_1(\tilde{x}_1) \times \cdots \times \partial^L f_p(\tilde{x}_p),$$

$$\partial^C f(\tilde{x}) = \partial^C f_1(\tilde{x}_1) \times \cdots \times \partial^C f_p(\tilde{x}_p).$$

3 The generalized subdifferentials of the Moreau envelopes of the rank function

In this section, we calculate explicitly (all) the generalized subdifferentials of the Moreau envelopes of the rank function. The process adopted for that purpose is the following: Firstly, compute the generalized subdifferentials of the Moreau envelopes c_ε of the counting function c; then apply fine results by Lewis and Sendov ([8],[9]) on the nonsmooth analysis of functions of singular values.

Theorem 4. *Let x be a vector in $\mathbb{R}^p$ for which*

$$x_1 \geq x_2 \geq \cdots \geq x_p \geq 0.$$

The generalized subdifferentials of c_ε at x are then expressed as follows:

- *If $x_1 < \sqrt{\varepsilon}$, then*

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \left\{ \left(\frac{2x_1}{\varepsilon}, \ldots, \frac{2x_p}{\varepsilon} \right) \right\}.$$

- *If $x_p > \sqrt{\varepsilon}$, then*

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \{(0, \ldots, 0)\}.$$

- *If there exists k such that $x_k > \sqrt{\varepsilon} > x_{k+1}$, then*

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \left\{ \left(0, \ldots, 0, \frac{2x_{k+1}}{\varepsilon}, \ldots, \frac{2x_p}{\varepsilon} \right) \right\}.$$

- *If there exists k such that $x_k = \sqrt{\varepsilon}$, then*

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \emptyset.$$

If we denote

$$k_0 = \min\{k| \quad x_k = \sqrt{\varepsilon}\},$$
$$k_1 = \max\{k| \quad x_k = \sqrt{\varepsilon}\},$$

then,

$$\partial^L c_\varepsilon(x) = \{0\} \times \dots \{0\} \times \{0; \frac{2x_{k_0}}{\varepsilon}\} \times \dots \times \{0; \frac{2x_{k_1}}{\varepsilon}\} \times \{\frac{2x_{k_1+1}}{\varepsilon}\} \times \dots \{\frac{2x_p}{\varepsilon}\};$$

$$\partial^C c_\varepsilon(x) = \{0\} \times \dots \{0\} \times \left[0; \frac{2x_{k_0}}{\varepsilon}\right] \times \dots \times \left[0; \frac{2x_{k_1}}{\varepsilon}\right] \times \{\frac{2x_{k_1+1}}{\varepsilon}\} \times \dots \{\frac{2x_p}{\varepsilon}\}.$$

Since c_ε has a "separable" structure, $c_\varepsilon(x_1, \dots, x_p) = \frac{1}{\varepsilon}\sum_{i=1}^{p}\left[x_i^2 - (x_i^2 - \varepsilon)^+\right]$, our first task is to calculate the generalized subdifferentials of functions like $x_i \in \mathbb{R} \mapsto (x_i^2 - \varepsilon)^+$

Lemma 1. *For $\varepsilon > 0$, we define h as follows*

$$\begin{array}{rccc} h: & \mathbb{R} & \to & \mathbb{R} \\ & x & \mapsto & -\frac{1}{\varepsilon}(x^2 - \varepsilon)^+. \end{array}$$

We then have:

- *If $x^2 < \varepsilon$*

$$\partial^P h(x) = \partial^V h(x) = \partial^L h(x) = \partial^C h(x) = \{\nabla h(x)\} = \{0\}.$$

- *If $x^2 > \varepsilon$*

$$\partial^P h(x) = \partial^V h(x) = \partial^L h(x) = \partial^C h(x) = \{\nabla h(x)\} = \{-\frac{2x}{\varepsilon}\}.$$

- *If $x^2 = \varepsilon$*

$$\partial^P h(x) = \partial^V h(x) = \emptyset,$$
$$\partial^L h(x) = \left\{0; -\frac{2x}{\varepsilon}\right\},$$
$$\partial^C h(x) = \begin{cases} \left[0; -\frac{2x}{\varepsilon}\right] & \text{if } x = -\sqrt{\varepsilon}, \\ \left[-\frac{2x}{\varepsilon}; 0\right] & \text{if } x = \sqrt{\varepsilon}. \end{cases}$$

Proof. By definition,

$$h(x) = \begin{cases} 0 & \text{if } x^2 < \varepsilon \\ -\frac{1}{\varepsilon}(x^2 - \varepsilon) & \text{if } x^2 \geq \varepsilon \end{cases}.$$

Thus, the function h is differentiable at any $x \notin \{-\sqrt{\varepsilon}; \sqrt{\varepsilon}\}$ with

$$h'(x) = 0 \quad \text{if } x^2 < \varepsilon,$$

$$h'(x) = -\frac{2x}{\varepsilon} \quad \text{if } x^2 > \varepsilon.$$

For $x = -\sqrt{\varepsilon}$, $h(x) = 0$. Then, $x^* \in \partial^F h(-\sqrt{\varepsilon})$ if and only if

$$\liminf_{y \to 0} \frac{h(y - \sqrt{\varepsilon}) - x^* y}{|y|} \geq 0.$$

This is equivalent to

$$\liminf_{y \to 0^+} \frac{h(y - \sqrt{\varepsilon}) - x^* y}{|y|} \geq 0, \tag{8}$$

and

$$\liminf_{y \to 0^-} \frac{h(y - \sqrt{\varepsilon}) - x^* y}{|y|} \geq 0. \tag{9}$$

When $y > 0$ is close to 0, the value of h at $y - \sqrt{\varepsilon}$ is 0. Thus (8) becomes

$$x^* \leq 0.$$

When $y < 0$ is close to 0, the value of h at $y - \sqrt{\varepsilon}$ is $-\frac{1}{\varepsilon}(y^2 - 2\sqrt{\varepsilon} y)$. Thus (9) becomes

$$x^* \geq \frac{2\sqrt{\varepsilon}}{\varepsilon} > 0.$$

This means that $\partial^F h(-\sqrt{\varepsilon})$ has no element, that is to say

$$\partial^F h(-\sqrt{\varepsilon}) = \emptyset.$$

We also prove in the same way that

$$\partial^F h(\sqrt{\varepsilon}) = \emptyset.$$

From the fact that $\partial^P h(x)$ is a subset of $\partial^F h(x)$, we infer that, for $x = \pm\sqrt{\varepsilon}$,

$$\partial^V h(x) = \emptyset.$$

Now, from the definition of the limiting subdifferential, we obtain

$$\partial^L h(x) = \left\{0; -\frac{2x}{\varepsilon}\right\},$$

for $x = \pm\sqrt{\varepsilon}$.

Because the Clarke subdifferential of h at any $x \in \mathbb{R}$ is the closed convex hull of the limiting subdifferential of h at x, we get

$$\partial^C h(x) = \begin{cases} \left[0; -\frac{2x}{\varepsilon}\right] & \text{if } x = -\sqrt{\varepsilon}, \\ \left[-\frac{2x}{\varepsilon}; 0\right] & \text{if } x = \sqrt{\varepsilon}. \end{cases}$$

□

Proof. (of Theorem 4)

The Moreau envelope of the counting function is given by:

$$c_\varepsilon(x) = \frac{1}{\varepsilon}\|x\|^2 - \frac{1}{\varepsilon}\sum_{i=1}^{p}(x_i^2 - \varepsilon)^+.$$

We can rewrite c_ε as the sum of two functions c_1 and c_2 where $c_1(x) = \frac{1}{\varepsilon}\|x\|^2$ and $c_2(x) = -\frac{1}{\varepsilon}\sum_{i=1}^{p}(x_i^2 - \varepsilon)^+$.

It is clear that c_1 is a C^1 function and $\nabla c_1(x) = \frac{2x}{\varepsilon}$. Because $c_2(x) = -\frac{1}{\varepsilon}\sum_{i=1}^{p}(x_i^2 - \varepsilon)^+ = \sum_{i=1}^{p} h(x_i)$, the Fréchet subdifferential of c_2 at x can be expressed as the product of the ones of h at x_i (*cf.* Theorem 3). By applying Theorem 2 for the two functions c_1 and c_2, we obtain

$$\partial^F c_\varepsilon(x) = \nabla c_1(x) + \partial^F c_2(x),$$
$$\partial^L c_\varepsilon(x) = \nabla c_1(x) + \partial^L c_2(x),$$
$$\partial^C c_\varphi(x) = \nabla c_1(x) + \partial^C c_2(x).$$

Thus,

$$\partial^F c_\varepsilon(x) = \frac{2x}{\varepsilon} + \prod_{i=1}^{p} \partial^F h(x_i),$$
$$\partial^L c_\varepsilon(x) = \frac{2x}{\varepsilon} + \prod_{i=1}^{p} \partial^L h(x_i),$$
$$\partial^C c_\varepsilon(x) = \frac{2x}{\varepsilon} + \prod_{i=1}^{p} \partial^C h(x_i).$$

For $x = (x_1, \dots, x_p)$ such that $x_1 \geq \dots \geq x_p \geq 0$ and $\varepsilon > 0$, it remains to consider four cases:

- If $x_1 < \sqrt{\varepsilon}$, then

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \left\{\left(\frac{2x_1}{\varepsilon}, \dots, \frac{2x_p}{\varepsilon}\right)\right\}.$$

- If $x_p > \sqrt{\varepsilon}$ then

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \{(0, \dots, 0)\}.$$

- If there exists k such that $x_k > \sqrt{\varepsilon} > x_{k+1}$, then c_ε is differentiable at x and

$$\partial^P c_\varepsilon(x) = \partial^F c_\varepsilon(x) = \partial^L c_\varepsilon(x) = \partial^C c_\varepsilon(x) = \{\nabla c_\varepsilon(x)\} = \{(0,\ldots,0,\frac{2x_{k+1}}{\varepsilon},\ldots,\frac{2x_p}{\varepsilon})\}.$$

- If there exists k satisfying

$$x_k = \sqrt{\varepsilon},$$

by denoting

$$k_0 = \min\{k| \quad x_k = \sqrt{\varepsilon}\},$$
$$k_1 = \max\{k| \quad x_k = \sqrt{\varepsilon}\},$$

we get

$$\partial^L c_\varepsilon(x) = \{0\} \times \ldots \{0\} \times \{0; \frac{2x_{k_0}}{\varepsilon}\} \times \cdots \times \{0; \frac{2x_{k_1}}{\varepsilon}\} \times \{\frac{2x_{k_1+1}}{\varepsilon}\} \times \ldots \{\frac{2x_p}{\varepsilon}\};$$

$$\partial^C c_\varepsilon(x) = \{0\} \times \ldots \{0\} \times \left[0; \frac{2x_{k_0}}{\varepsilon}\right] \times \cdots \times \left[0; \frac{2x_{k_1}}{\varepsilon}\right] \times \{\frac{2x_{k_1+1}}{\varepsilon}\} \times \ldots \{\frac{2x_p}{\varepsilon}\}.$$

□

Before passing from the results on c_ε to those concerning rank_ε, we need to recall two results on the generalized subdifferentiation of nonsmooth functions of singular values of matrices.

Recall that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *absolutely symmetric* when

$$f(x_1,\ldots,x_p) = f(\hat{x}_1,\ldots,\hat{x}_p) \quad \text{for all } x = (x_1,\ldots,x_p) \in \mathbb{R}^p,$$

where $\hat{x} = (\hat{x}_1,\ldots,\hat{x}_p)$ is the vector, built up from $x = (x_1,\ldots,x_p)$, whose components are the $|x_i|$'s arranged in a decreasing order ($|\hat{x}_1| \geq |\hat{x}_2| \geq \cdots \geq |\hat{x}_p|$).

For a matrix $A \in \mathcal{M}_{m,n}(\mathbb{R})$, we denoted by $O(m,n)^A$ the set of pair (U,V) of orthogonal matrices which give a singular value decomposition of A, *i.e.*

$$A = U\Sigma V^T.$$

Theorem 5 ([9]). *If $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and if f is an absolutely symmetric function, lower-semicontinuous around $\sigma(A) = (\sigma_1(A),\ldots,\sigma_p(A))$, then $f \circ \sigma$ is lower-semicontinuous around A and*

$$\begin{aligned}
\partial^C(f\circ\sigma)(A) &= O(m,n)^A.\text{diag}_{m,n}\partial^C(f(\sigma(A)) \\
&= \{U.\text{diag}_{m,n}(y).V^T| \quad y \in \partial^C(f(\sigma(A)), (U,V) \in O(m,n)^A\} \\
\partial^P(f\circ\sigma)(A) &= O(m,n)^A.\text{diag}_{m,n}\partial^P(f(\sigma(A)) \\
&= \{U.\text{diag}_{m,n}(y).V^T| \quad y \in \partial^F(f(\sigma(A)), (U,V) \in O(m,n)^A\}
\end{aligned}.$$

Theorem 6 ([8]). *If $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and if f is an absolutely symmetric function, lower-semicontinuous around $\sigma(A)$, then the proximal subdifferential of any singular value function $f \circ \sigma$ at A is given by the formula*

$$\begin{aligned}\partial^F(f \circ \sigma)(A) &= O(m,n)^A.\mathrm{diag}_{m,n}\partial^F(f(\sigma(A)) \\ &= \{U.\mathrm{diag}_{m,n}(y).V^T | \quad y \in \partial^P(f(\sigma(A)), (U,V) \in O(m,n)^A\}.\end{aligned}$$

Our function c_ε is absolutely symmetric and continuous on $\mathbb{R}^p$. Then we apply the two theorems above in order to obtain the generalized subdifferentials of the Moreau envelopes $\mathrm{rank}_\varepsilon$ of the rank function. This is the main result of this Section 3.

Theorem 7. *Let A be a matrix in $\mathcal{M}_{m,n}(\mathbb{R})$ and let $\sigma_1(A) \geq \cdots \geq \sigma_p(A)$ be the singular values of A. Then the generalized subdifferentials of $\mathrm{rank}_\varepsilon$ at A can be expressed as follows:*

- *If $\sigma_1(A) < \sqrt{\varepsilon}$, then*

$$\partial^P \mathrm{rank}_\varepsilon(A) = \partial^C \mathrm{rank}_\varepsilon(A) = \left\{ U \mathrm{diag}\left(\frac{2\sigma_1(A)}{\varepsilon}, \ldots, \frac{2\sigma_p(A)}{\varepsilon} \right) V^T \right\}.$$

- *If $\sigma_p(A) > \sqrt{\varepsilon}$, then*

$$\partial^P \mathrm{rank}_\varepsilon(A) = \partial^C \mathrm{rank}_\varepsilon(A) = \{0\}.$$

- *If there exists k such that $\sigma_k(A) > \sqrt{\varepsilon} > \sigma_{k+1}(A)$, then*

$$\partial^P \mathrm{rank}_\varepsilon(A) = \partial^C \mathrm{rank}_\varepsilon(A) = \left\{ U \mathrm{diag}\left(0, \ldots, 0, \frac{2\sigma_{k+1}(A)}{\varepsilon}, \ldots, \frac{2\sigma_p(A)}{\varepsilon} \right) V^T \right\}.$$

- *If there exists k such that $\sigma_k(A) = \sqrt{\varepsilon}$, then*

$$\partial^P \mathrm{rank}_\varepsilon(A) = \partial^F \mathrm{rank}_\varepsilon(A) = \emptyset.$$

  If we denote

$$\begin{aligned} k_0 &= \min\{k | \quad \sigma_k(A) = \sqrt{\varepsilon}\}, \\ k_1 &= \max\{k | \quad \sigma_k(A) = \sqrt{\varepsilon}\}, \end{aligned}$$

  then,

$$\partial^L \mathrm{rank}_\varepsilon(A) = \left\{ U \mathrm{diag}(y) V^T | \quad y \in \partial^L c_\varepsilon(\sigma(A)); (U,V) \in O(m,n)^A \right\};$$

$$\partial^C \mathrm{rank}_\varepsilon(A) = \left\{ U \mathrm{diag}(y) V^T | \quad y \in \partial^C c_\varepsilon(\sigma(A)); (U,V) \in O(m,n)^A \right\}.$$

4 The viscosity subdifferential of the rank function

To get the viscosity subdifferential of the rank function from that of its Moreau envelopes, it remains a final step to carry out. This can be done with the help of a theorem by Jourani ([6]). Such a result exists for the viscosity (or Fréchet) subdifferential, we are not aware of any similar result for the other generalized subdifferentials. Recall that in our finite-dimensional context, $\partial^V = \partial^F$.

Theorem 8 ([6])**.** *Let X be a finite-dimensional space and let $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a lower-semicontinuous function. We suppose that f is bounded from below by a nonnegative quadratic function, that is to say:*

$$\exists a > 0, \exists b > 0, \exists \overline{x} \in X \text{ such that } f(x) \geq -a\|x - \overline{x}\|^2 - b \text{ for all } x \in X.$$

Then, at a point x_0 where f is finite,

$$\partial^V f(x_0) = \text{seq} - \limsup_{\substack{\varepsilon \to O^+ \\ u \to x_0 \\ f_\varepsilon(u) \to f(x_0)}} \partial^V f_\varepsilon(u),$$

where f_ε denotes the Moreau envelope of f (with coefficient $\varepsilon > 0$) and

$$\text{seq} - \limsup_{\substack{\varepsilon \to O^+ \\ u \to x_0 \\ f_{\varepsilon_k}(u) \to f(x_0)}} \partial^V f_\varepsilon(u) = \left\{ \begin{array}{ll} x^*| & \exists u_k^* \in \partial^V f_{\varepsilon_k}(u_k) \to x^*, \text{ with } \varepsilon_k \to 0^+, u_k \to x_0 \\ & \text{and } f_{\varepsilon_k}(u_k) \to f(x_0) \end{array} \right\}.$$

We now state the final result of our paper.

Theorem 9 (The viscosity subdifferential of the rank function)**.** *For $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $\partial^V(\text{rank})(A)$ is constructed as follows:*

- *Consider the pairs of matrices $(U, V) \in O(m, n)^A$, i.e.*

$$U \text{diag}_{m,n}(\sigma(A))V^T = A.$$

- *Consider the "pseudo-diagonal" matrices $\text{diag}_{m,n}(x^*)$, where $x^* \in \mathbb{R}^p$ is such that $x_i^* = 0$ for all $i = 1, \ldots, r$ (recall that $r = \text{rank } A$).*

- *Then, collect all the matrices of the form $U \text{diag}_{m,n}(x^*)V^T$.*

In a single formula,

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

Proof. We begin by getting the viscosity subdifferential of the counting function c from that of its Moreau envelopes c_ε.

Let $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$ be satisfying

$$x_1 \geq \cdots \geq x_p \geq 0.$$

Thanks to Theorem 8, we have

$$\partial^V c(x) = seq - \limsup_{\substack{\varepsilon \to 0^+ \\ u \to x \\ c_\varepsilon(u) \to c(x)}} \partial^V c_\varepsilon(u).$$

Let $\{\varepsilon_k\}_k$ be a sequence converging to 0 and let $\{u^k\}_k$ be a sequence converging to x.

Let $r = c(x)$. For $\epsilon > 0$ small enough, there exist K_1 and K_2 such that

$$\forall k \geq K_1 \quad \forall i = 1, \ldots, r, \quad u_i^k \geq x_i - \epsilon,$$

$$\forall k \geq K_2, \quad \sqrt{\varepsilon_k} < x_r - \epsilon.$$

Then, if $K_0 = \max(K_1, K_2)$, we have

$$\forall k \geq K_0 \quad \forall i = 1, \ldots, r, \quad u_i^k > \sqrt{\varepsilon_k}.$$

By Theorem 4, we have

$$\forall k \geq K_0 \quad \partial^V c_{\varepsilon_k}(x_k) \subset \{0\}^r \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

Hence, $\partial^V c(x) \subset \{0\}^r \times \mathbb{R} \times \cdots \times \mathbb{R}$.

On the other hand, any vector in $\mathbb{R}^p$ whose first r components are 0 belongs to $\partial^V c(x)$.

Indeed, for $a = (0, \ldots, 0, a_{r+1}, \ldots, a_p)$ and $\varepsilon_k \to 0^+$, we take

$$y_k = (x_1, \ldots, x_r, \varepsilon_k a_{r+1}, \ldots, \varepsilon_k a_p) \to x.$$

Because $\varepsilon_k \to 0$, there exists K_3 such that

$$\forall k \geq K_3 \quad \forall i = r+1, \ldots, p, \quad |\varepsilon_k a_i| < \sqrt{2\varepsilon_k}.$$

Then, by using Theorem 4, for all $k \geq K_3$

$$\partial^V c_{\varepsilon_k}(y_k) = a.$$

Consequently,

$$\partial^V c(x) = \{0\}^r \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

Now, thanks to Theorem 6, we get at the announced expression of the viscosity subdifferential of the rank function. □

Final remark. It happens that all the generalized subdifferentials of the rank function do coincide; see [7] for that. Our approach in the present paper allowed to retrieve the viscosity (or Fréchet) subdifferential of the rank function, not the other ones.

Acknowledgment. We would like to thank Prof. A.Jourani (University of Bourgogne, Dijon) for drawing our attention to this possible way of getting at the Fréchet generalized subdifferential of the rank function (ALEL meeting in Castro-Urdiales, June 2011).

References

[1] F.H. Clarke, *Functional analysis, calculus of variations and optimal control.* Springer (2013).

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

[3] J.-B. Hiriart-Urruty, *Bases, outils et principes pour l'analyse variationnelle.* Springer (2012).

[4] J.-B. Hiriart-Urruty and H.Y. Le, *From Eckart-Young approximation to Moreau envelopes and vice versa.* RAIRO - Operations Research, Vol 47 (2013), no.3, 299-310.

[5] J.-B. Hiriart-Urruty and H.Y. Le, *A variational approach of the rank function*, TOP (Journal of the Spanish Society of Statistics and Operations Research), Vol 21(2013), no.2, 207-240.

[6] A. Jourani, *Limit superior of subdifferentials of uniformly convergent functions*, Positivity Vol 3(1999), 33-47.

[7] H.Y. Le, *The generalized subdifferentials of the rank function*, Optimization Letters, Vol 7 (2013), no. 4, 731-743.

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

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

[10] J.-P. Penot, *Calculus without derivatives.* Graduate Texts in Mathematics, Vol. 266, Springer (2013).

[11] W. Schirotzek, *Nonsmooth analysis.* Springer(2007).

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

[13] Y.-B. Zhao, *Approximation theory of matrix rank minimization and its application to quadratic equations*, Linear Algebra and Its Applications, Vol 437 (2012), no.1, 77-93.