Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

S. Hosseini, D. R. Luke, A. Uschmajew

Tangent and normal cones for low-rank matrices

INS Preprint No. 1715

Oct 2017

TANGENT AND NORMAL CONES FOR LOW-RANK MATRICES

SEYEDEHSOMAYEH HOSSEINI*, D. RUSSELL LUKE†, AND ANDRÉ USCHMAJEW‡

ABSTRACT. In [D. R. Luke, *J. Math. Imaging Vision*, 47(3):231-238, 2013] the structure of the Mordukhovich normal cone to varieties of low rank matrices at rank-deficient points has been determined. A simplified proof of that result is presented here. As a corollary we obtain the corresponding Clarke normal cone. The results are put into context of first-order optimality conditions for low-rank matrix optimization problems.

1. INTRODUCTION

In continuous minimization problems, necessary conditions for local minima relate the first-order geometries of a constraint set $\mathcal{M}$ and a sublevel set of a cost function f with each other in order to give the geometric intuition that descent direction of f should point "away" from $\mathcal{M}$ a mathematically rigorous meaning. If $\mathcal{M}$ is a smooth submanifold in $\mathbb{R}^n$ and f is continuously differentiable in a neighborhood of $x \in \mathcal{M}$, then a necessary condition for x being a local minimum of f on $\mathcal{M}$ is that the anti-gradient $-\nabla f(x)$ belongs to the *normal space*, that is, the orthogonal complement of the tangent space $T_{\mathcal{M}}(x)$ at x:

$$-\nabla f(x) \in N_{\mathcal{M}}(x) := (T_{\mathcal{M}}(x))^{\perp}.$$

When the set $\mathcal{M}$ is just closed, but not necessarily smooth, the normal space in this optimality condition has to be replaced by the polar of the Bouligand tangent cone [4], which we will call the *Bouligand normal cone*:

$$-\nabla f(x) \in N^B_{\mathcal{M}}(x) := (T^B_{\mathcal{M}}(x))^{\circ}.$$

see Eq. (2.3) for the general definition of $T^B_{\mathcal{M}}(x)$.

More generally, when the function f is just locally Lipschitz continuous, but not necessarily differentiable at a critical point, generalized derivatives and subgradients need to be considered. First-order optimality conditions then take the form

$$0 \in N_{\mathcal{M}}(x) + \partial f(x), \tag{1.1}$$

where $N_{\mathcal{M}}(x)$ is a certain closed convex cone of "normal directions", and $\partial f(x)$ is the corresponding subdifferential containing all subgradients at x. Well-known normal cones are the *Clarke normal cone* $N^C_{\mathcal{M}}(x)$ and the *Mordukhovich normal cone* $N^M_{\mathcal{M}}(x)$; see Sec. 3.

Intuitively, the necessary first-order condition (1.1) becomes stronger the narrower the considered normal cone is, and hence gives more meaning to the notion of a critical point. In general it holds that

$$N^B_{\mathcal{M}}(x) \subseteq N^M_{\mathcal{M}}(x) \subseteq N^C_{\mathcal{M}}(x).$$

In this communication we show that for the set $\mathcal{M}_{\le k}$ of matrices of rank at most k these three cones are strictly included in each other at singular points where the rank is strictly less than k.

*Hausdorff Center for Mathematics & Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany (hosseini@ins.uni-bonn.de).

†Institute for Numerical and Applied Mathematics, University of Göttingen, 37083 Göttingen, Germany (r.luke@math.uni-goettingen.de).

‡Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany (uschmajew@mis.mpg.de).

Note that at points of rank k the set $\mathcal{M}_{\le k}$ is locally a smooth manifold and all normal cones equal to the orthogonal complement of the tangent space. This has interesting implications to necessary conditions in matrix optimization problems with low-rank constraints, as should be discussed at the end of this note.

2. Preliminaries

We consider the linear space $\mathbb{R}^{M\times N}$ of $M \times N$ matrices. The Euclidean structure in this space is provided by the Frobenius inner product. Given a real valued function f on $\mathbb{R}^{M\times N}$, we are interested in optimality conditions for *low-rank optimization problems* of the form

$$\min f(X), \qquad X \in \mathcal{M}_{\le k}, \tag{2.1}$$

where

$$\mathcal{M}_{\le k} = \{X \in \mathbb{R}^{M\times N} : \operatorname{rank}(X) \le k\}$$

is the set of matrices of rank at most k. These sets are real algebraic varieties and closed due to lower-semicontinuity of matrix rank. In the following we always assume

$$k \le k_{\max} := \min(M, N).$$

The characterization of necessary optimality conditions for (2.1) becomes nontrivial at rank-deficient points, which are the singular points of the variety $\mathcal{M}_{\le k}$. Therefore concepts from nonsmooth analysis have to be used.

In the following it will be frequently convenient to view matrices as elements of the tensor product of their column and row spaces. By this we mean that a matrix X admits a decomposition

$$X = U\Sigma V^T \text{ s.t. } \operatorname{range}(U) \subseteq \mathcal{U}, \operatorname{range}(V) \subseteq \mathcal{V} \quad \text{if and only if} \quad X \in \mathcal{U} \otimes \mathcal{V}. \tag{2.2}$$

Moreover, the smallest dimension that $\mathcal{U}$ and $\mathcal{V}$ can have is the same and equals the rank of X. In this case, the choice is unique, namely $\mathcal{U} = \operatorname{range}(X)$ has to be the *column space*, and $\mathcal{V} = \operatorname{range}(X^T)$ the *row space* of X. We recall that a singular value decomposition (SVD) of a matrix X of rank k is a decomposition of the form $X = \sum_{r=1}^{k} \sigma_r u_r v_r^T$, $\quad \sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_k > 0$, that is, $X = U\Sigma V^T$, where Σ is diagonal, and $U = [u_1, \dots, u_k]$ and $V = [v_1, \dots, v_k]$ are orthonormal bases for the column and row spaces of X, respectively. The rank-one terms in this decomposition are mutually orthogonal in the Frobenius inner product. A truncation $\sum_{r=1}^{s} \sigma_r u_r v_r^T$ of the SVD to rank $s < k$ yields the best approximation to X in the set $\mathcal{M}_{\le s}$ in the Frobenius norm. It is unique if and only if $\sigma_s > \sigma_{s+1}$.

2.1. Tangent and normal space of a fixed rank manifold.

The real algebraic variety $\mathcal{M}_{\le k}$ stratifies into the smooth manifolds

$$\mathcal{M}_s = \{X \in \mathbb{R}^{M\times N} : \operatorname{rank}(X) = s\}$$

of matrices of rank exactly s. Being a smooth manifold, every $\mathcal{M}_s$ is prox-regular in a neighborhood of each of its points; see [1]. Hence all usual notions of tangent cones coincide with the tangent space. Moreover, all notions of normal cones coincide with the normal space [3]. The tangent spaces to $\mathcal{M}_s$ are well known; see, e.g., [5, Ex. 14.16] or [6, Prop. 4.1].

Theorem 2.1. *Let X have rank s, column space $\mathcal{U}$, and row space $\mathcal{V}$. The tangent space to the manifold $\mathcal{M}_s$ at X admits the orthogonal decomposition*

$$T_{\mathcal{M}_s}(X) = (\mathcal{U} \otimes \mathcal{V}) \oplus (\mathcal{U} \otimes \mathcal{V}^\perp) \oplus (\mathcal{U}^\perp \otimes \mathcal{V}).$$

The normal space is

$$N_{\mathcal{M}_s}(X) = [T_{\mathcal{M}_s}(X)]^\perp = \mathcal{U}^\perp \otimes \mathcal{V}^\perp.$$

It is interesting to note that $T_{\mathcal{M}_s}(X)$ contains matrices of rank at most $2s$, while the maximum rank in $N_{\mathcal{M}_s}(X)$ is $k_{\max} - s$. Also note that $X \in T_{\mathcal{M}_s}(X)$ (which also follows from the fact that $\mathcal{M}_s$ is a cone). In applications s is much smaller than $k_{\max}$, and in this context it is important that orthogonal projections on $T_{\mathcal{M}_s}(X)$ and $N_{\mathcal{M}_s}(X)$ can be computed using only projections on the s-dimensional spaces $\mathcal{U}$ and $\mathcal{V}$. Namely,

$$P_{T_{\mathcal{M}_s}(X)}(Z) = P_{\mathcal{U}}Z + ZP_{\mathcal{V}} - P_{\mathcal{U}}ZP_{\mathcal{V}},$$

and

$$P_{N_{\mathcal{M}_s}(X)}(Z) = Z - P_{T_{\mathcal{M}_s}(X)}(Z).$$

2.2. Bouligand tangent and normal cone to $\mathcal{M}_{\leq k}$.

The general definition of the Bouligand tangent cone to a closed set $\mathcal{M} \subseteq \mathbb{R}^N$ is as follows:

$$T^B_{\mathcal{M}}(x) = \{\xi \in \mathbb{R}^N : \exists (x_n) \subseteq \mathcal{M},\ (a_n) \subseteq \mathbb{R}^+ \text{ s.t. } x_n \to x,\ a_n(x_n - x) \to \xi\}. \tag{2.3}$$

In the context of low-rank optimization, the Bouligand tangent cone to $\mathcal{M}_{\leq k}$ has been derived from this definition in [9]. In [2], an essentially equivalent definition based on derivatives of analytic curves has been used. The result is also well known in algebraic geometry [5, Ex. 20.5].

Theorem 2.2. *Let $X \in \mathcal{M}_{\leq k}$ have rank $s \leq k$. The Bouligand tangent cone to the closed variety $\mathcal{M}_{\leq k}$ at X is*

$$T^B_{\mathcal{M}_{\leq k}}(X) = T_{\mathcal{M}_s}(X) \oplus \{Y \in N_{\mathcal{M}_s}(X) : \operatorname{rank}(Y) \leq k - s\}.$$

As noted in [9], an element in the polar cone of $T^B_{\mathcal{M}_{\leq k}}(X)$ needs to be orthogonal to $T_{\mathcal{M}_s}(X)$ and to every rank $k - s$ matrix in $N_{\mathcal{M}_s}(X)$. When $s < k$, only the zero matrix fulfills this.

Corollary 2.3. *Let X have rank $s < k$. The Bouligand normal cone (defined as the polar of the Bouligand tangent cone) to $\mathcal{M}_{\leq k}$ at X is*

$$N^B_{\mathcal{M}_{\leq k}}(X) = \{0\}.$$

3. Clarke and Mordukhovich normal cones

Let $P_{\mathcal{M}}$ denote the (set-valued) metric projection (in the Euclidean norm $\|\cdot\|$) onto a closed subset $\mathcal{M} \subseteq \mathbb{R}^n$. Following [8, Theorem 1.6], the Mordukhovich normal cone $N^M_{\mathcal{M}}(x)$ to $\mathcal{M}$ at x can be defined as follows:

$$N^M_{\mathcal{M}}(x) = \{\eta \in \mathbb{R}^n : \text{there exist } (x_i) \subset \mathbb{R}^n \text{ and } (\eta_i) \subset \mathbb{R}^n \text{ such that } x_i \to x,\ \eta_i \to \eta, \text{ and } \eta_i \in \operatorname{cone}(x_i - P_{\mathcal{M}}(x_i)) \text{ for all } i \in \mathbb{N}\}. \tag{3.1}$$

The elements of $N^M_{\mathcal{M}}(x)$ are called *basic normal*, *limiting normal* or simply *normal vectors.*

It is proved in [8, Theorem 3.57] that the Clarke normal cone can be obtained from the Mordukhovich normal cone as its closed convex hull

$$N^C_{\mathcal{M}}(x) = \operatorname{cl}\operatorname{conv} N^M_{\mathcal{M}}(x). \tag{3.2}$$

We now consider the Mordukhovich normal cones to $\mathcal{M}_{\leq k}$. The following result is essentially due to Luke [7].[1]

Theorem 3.1. *Let $X \in \mathcal{M}_{\leq k}$ have rank $s \leq k$. The Mordukhovich normal cone to the closed variety $\mathcal{M}_{\leq k}$ at X is*

$$N^M_{\mathcal{M}_{\leq k}}(X) = \{Y \in N_{\mathcal{M}_s}(X) : \operatorname{rank}(Y) \leq k_{\max} - k\}. \tag{3.3}$$

[1]Some inaccuracies in the statement of Theorem 3.1 in [7] are corrected here. Also, the "⊆" part is proven by a more direct argument compared to [7].

Proof. The result is clear if $k = k_{\max}$, since in this case $\mathcal{M}_{\leq k_{\max}} = \mathbb{R}^{M\times N}$. Hence we consider $k < k_{\max}$. As before, let $\mathcal{U}$ and $\mathcal{V}$ denote the column and row space of X, respectively. Recall that $N_{\mathcal{M}_s}(X) = \mathcal{U}^\perp \otimes \mathcal{V}^\perp$.

Denoting the set on the right side of (3.3) by W, we first show $N^M_{\mathcal{M}_{\leq k}}(X) \supseteq W$. Let $Y \in W$, then by (2.2) there exist subspaces $\mathcal{U}_1 \subseteq \mathcal{U}^\perp$ and $\mathcal{V}_1 \subseteq \mathcal{V}^\perp$, both of dimension $k_{\max} - k$, such that $Y \in \mathcal{U}_1 \otimes \mathcal{V}_1$. Also there exist subspaces $\mathcal{U}_0 \subseteq \mathcal{U}^\perp \cap \mathcal{U}_1^\perp$ and $\mathcal{V}_0 \subseteq \mathcal{V}^\perp \cap \mathcal{V}_1^\perp$ of dimension $k - s$ both. Pick $Z \in \mathcal{U}_0 \otimes \mathcal{V}_0$ with $\operatorname{rank}(Z) = k - s$, and consider the sequence

$$X_i \coloneqq X + i^{-1/2} Z + i^{-1} Y \to X \quad (i \to \infty).$$

Using SVD and $\operatorname{rank}(Z) = k - s$, the mutual orthogonality of the column resp. row spaces of X, Z, and Y implies that for large enough $i \in \mathbb{N}$ the best approximation of rank at most k is

$$P_{\mathcal{M}_{\leq k}}(X_i) = X + i^{-1/2} Z.$$

Hence $Y \in \operatorname{cone}(X_i - P_{\mathcal{M}_{\leq k}}(X_i))$ for all i, which by (3.1) proves $Y \in N^M_{\mathcal{M}_{\leq k}}(X)$.

To prove $N^M_{\mathcal{M}_{\leq k}}(X) \subseteq W$, consider sequences $X_i \to X$ and $Y_i \to Y$ satisfying $Y_i = \alpha_i(X_i - Z_i)$ with $Z_i \in P_{\mathcal{M}_{\leq k}}(X_i)$. Note that $Z_i \to X$, since $\|Z_i - X\| \leq \|Z_i - X_i\| + \|X_i - X\| \leq 2\|X - X_i\|$ (the second inequality follows from $X \in \mathcal{M}_{\leq k}$). We have to show $Y \in W$. If $Y = 0$, this is clear. If $Y \neq 0$, it must hold $Y_i \neq 0$ and hence $\operatorname{rank}(X_i) > k$ for large enough i. As Z_i is a truncated SVD of X_i, we obtain

$$\operatorname{rank}(Y_i) = \operatorname{rank}(X_i - Z_i) = \operatorname{rank}(X_i) - \operatorname{rank}(Z_i) = \operatorname{rank}(X_i) - k \leq k_{\max} - k$$

for i large enough. From the lower semicontinuity of the rank function it now follows that $\operatorname{rank}(Y) \leq k_{\max} - k$. Another consequence of the fact that Z_i is a truncated SVD of X_i is that $Z_i Y_i^T = 0$ and $Z_i^T Y_i = 0$. In the limit we get XY^T and $X^T Y = 0$, which is equivalent to $Y \in \mathcal{U}^\perp \otimes \mathcal{V}^\perp$. In summary, we have shown $Y \in W$. □

As a corollary, we obtain the Clarke normal cone via (3.2).

Corollary 3.2. *Let $X \in \mathcal{M}_{\leq k}$ have rank $s \leq k$. The Clarke normal cone to the closed variety $\mathcal{M}_{\leq k}$ at X is*

$$N^C_{\mathcal{M}_{\leq k}}(X) = N_{\mathcal{M}_s}(X).$$

Further, we observe that the Mordukhovich normal cone provides the "missing" part in the Bouligand tangent cone to fill up the whole space.

Corollary 3.3. *Let X have rank $s \leq k$. Every matrix $Z \in \mathbb{R}^{M\times N}$ admits an orthogonal decomposition*

$$Z = Z_1 + Z_2, \qquad Z_1 \in T^B_{\mathcal{M}_{\leq k}}(X), \quad Z_2 \in N^M_{\mathcal{M}_{\leq k}}(X). \tag{3.4}$$

Proof. By Theorem 2.1, the normal space $N_{\mathcal{M}_s}(X)$ contains matrices of rank at most $k_{\max} - s$. Using SVD, every such matrix can be orthogonally decomposed into a matrix of rank $k_{\max} - k$ and another one of rank $k - s$. □

In this sense, the Mordukhovich normal cone can be seen as an appropriate "nonlinear orthogonal complement" to the Bouligand tangent cone. However, one should be aware that $N^C_{\mathcal{M}_{\leq k}}(X)$ is only complimentary to $T^B_{\mathcal{M}_{\leq k}}(X)$ if $\operatorname{rank}(X) = k$. Otherwise, their intersection contains all matrices in $N_{\mathcal{M}_s}(X)$ of rank at most $\min(k_{\max} - k, k - s)$. Further, the orthogonal decomposition (3.4) is then not unique.

4. Implications to necessary optimality conditions

The different necessary optimality conditions arising from different choices of normal cones are best illustrated for the case that the function f in (2.1) is continuously differentiable in a neighborhood of $\mathcal{M}_{\le k}$. Further, we consider matrices $X \in \mathcal{M}_{\le k}$ with $\operatorname{rank}(X) = s < k$, since otherwise all normal cones are the same and equal to $N_{\mathcal{M}_k}(X)$.

By Corollary 3.2, X will be a Clarke critical point on $\mathcal{M}_{\le k}$, if

$$-\nabla f(X) \in N_{\mathcal{M}_s}(X).$$

Hence this condition does not provide more information than stating that X is in particular a critical point on the smooth stratum $\mathcal{M}_s$. In this sense, the Clarke normal cone is "blind" to the fact that optimization is performed on $\mathcal{M}_{\le k}$ and that potentially a higher rank could be used. For example, if we imagine an optimization algorithm on $\mathcal{M}_{\le k}$ that is initialized with an optimal point $X \in \mathcal{M}_s$, it will terminate immediately without further improvement, if the Clarke normal cone would be used for checking optimality. Therefore, the Clarke normal cone is not the most suitable for optimization on the variety $\mathcal{M}_{\le k}$.

At the opposite extreme, a rank deficient point $X \in \mathcal{M}_s$, $s < k$, will be a critical for (2.1) in the Bouligand sense only if $\nabla f(X) = 0$. The optimality condition is thus the same as in the unconstrained case. Hence, if there do not exist points $X \in \mathcal{M}_{\le k}$ with $\nabla f(X) = 0$, then the problem (2.1) does not admit critical points that do not exploit the full possible rank k. In other words, even if $f(X)$ is minimal on $\mathcal{M}_s$, it is always possible to decrease the value of f on $\mathcal{M}_{\le k}$ in a neighborhood of X by increasing the rank. Indeed, the Euclidean projection of $-\nabla f(X)$ onto the Bouligand tangent cone $T^B_{\mathcal{M}_{\le k}}(X)$ provides a descent direction; cf. the discussion in [9].

Among the considered normal cones, the Mordukhovich normal cone is not convex. Its role is intermediate and best described by Corollary 3.3. On the one hand, the necessary condition

$$-\nabla f(X) \in N^M_{\mathcal{M}_{\le k}}(X) \subset N^C_{\mathcal{M}_{\le k}}(X)$$

implies $-\nabla f(X) \in N_{\mathcal{M}_s}(X)$, that is, X is critical on the smooth stratum $\mathcal{M}_s$. On the other hand, if $-\nabla f(X)$ is in $N_{\mathcal{M}_s}(X)$, but not in $N^C_{\mathcal{M}_{\le k}}(X)$ (then $s < k$), but an orthogonal decomposition

$$-\nabla f(X) = Z_1 + Z_2, \qquad Z_1 \in T^B_{\mathcal{M}_{\le k}}(X), \quad Z_2 \in N^M_{\mathcal{M}_{\le k}}(X)$$

with $Z_1 \neq 0$ is possible, that is, $-\nabla f(X)$ still contains a tangential component with respect to the variety $\mathcal{M}_{\le k}$.

References

[1] F. Bernard, L. Thibault, and N. Zlateva. Prox-regular sets and epigraphs in uniformly convex Banach spaces: various regularities and other properties. *Trans. Amer. Math. Soc.*, 363(4):2211–2247, 2011.

[2] T. P. Cason, P.-A. Absil, and P. Van Dooren. Iterative methods for low rank approximation of graph similarity matrices. *Linear Algebra Appl.*, 438(4):1863–1882, 2013.

[3] D. Drusvyatskiy and A. S. Lewis. Optimality, identifiability, and sensitivity. *Math. Program.*, 147(1-2, Ser. A):467–498, 2014.

[4] M. Guignard. Generalized Kuhn–Tucker conditions for mathematical programming problems in a Banach space. *SIAM J. Control*, 7:232–241, 1969.

[5] J. Harris. *Algebraic geometry. A First Course.* Springer-Verlag, New York, 1992.

[6] U. Helmke and M. A. Shayman. Critical points of matrix least squares distance functions. *Linear Algebra Appl.*, 215:1–19, 1995.

[7] D. R. Luke. Prox-regularity of rank constraints sets and implications for algorithms. *J. Math. Imaging Vision*, 47(3):231–238, 2013.

[8] B. S. Mordukhovich. *Variational analysis and generalized differentiation. I.* Springer-Verlag, Berlin, 2006.

[9] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM J. Optim.*, 25(1):622–646, 2015.