A Differentiable Analogue of the `null()`Function

Etienne Grossmann and José Santos Victor

May 23, 2009

Abstract

Computing the nullspace of a matrix is a common operation in many fields of science and engineering. As it is done, this is a non-continuous operation. There exist situations in which it is desirable to define a function that computes the nullspace in a differentiable manner. The method that we propose is in many respects a differentiable analogue of the non-continuous nullspace computation that is commonly implemented in numerical software packages and will be written $n()$ in this article.

Given a differentiable real matrix-valued function $B(\Theta)$ whose rank is locally constant, we define locally a differentiable matrix-valued function $U(\Theta)$ whose columns form an orthonormal basis of the nullspace of $B(\Theta)$. A closed-form expressions are given to compute $U(\Theta)$ and its partial derivatives from $B(\Theta)$, its partial derivatives and $n(B(\Theta))$.

We illustrate the utility of the method by showing how to use it to solve a class of constrained optimization problems using general unconstrained optimization tools.

Keywords : Matrix differential calculus, Implicit function theorem, Differential of singular values, Repeated singular values.

1 Introduction

The nullspace of a matrix is commonly used in algorithms for the most varied purposes. Its computation, as provided by major numerical software packages [17, 30, 2, 10], is a non-continuous operation. In many situations, it would be convenient if it were continuous and differentiable.

For example, when optimizing a cost function, efficient optimization algorithms require the computation of derivatives of the cost. Thus, if a nullspace is computed as part of the evaluation of the cost function, it would be good that this operation be differentiable.

Also, the sensitivity of algorithms is often estimated by propagating an error from its input to its output. For this purpose, it is necessary to know the differentials of the output with respect to its input at each step of the algorithm. Again, if the nullspace is involved at some point, its differential is needed.

In this article, we show how to compute a differentiable real matrix-valued function whose columns form an orthonormal basis of the nullspace of a matrix-valued function. The computation of this "nullspace" function and of its differential only requires tools commonly found in linear algebra software, such as the non-continuous computation of the nullspace of a matrix.

Scope of this article

Before going more in detail, we discuss the scope of this article, by defining the conditions under which the proposed method can be applied. The main issue treated here, and the method for resolving it, arises from the following question :

> If one is given a real, possibly differentiable matrix-valued function $B(\theta)$, under what conditions can a matrix-valued differentiable function $U(\theta)$ be defined such that the columns of $U(\theta)$ form an orthonormal basis of $\mathcal{N}ull\,(B(\theta))$?

One first condition is that the rank of $B(\Theta)$ be constant – at least locally. Otherwise, the size of $U(\Theta)$ cannot be fixed and its continuity is hard to define.

The requirement of locally constant rank is fulfilled e.g. if $B(\Theta)$ is an analytic function [14] : in that case, the rank of $B(\Theta)$ is the greatest size of a minor whose determinant is not identically zero. Since such a determinant is itself analytic, its zeros are isolated and, for any point Θ in which the rank of $B(\Theta)$ is maximal, there is a neighborhood of Θ on which the rank of $B(\Theta)$ is constant.

For more general functions, the rank is not necessarily locally constant, as shown by the following 1×2 matrix-valued function, defined on R.

$$B(\theta) = \left[\begin{array}{cc} 0 & \theta^2 \sin(1/\theta) \end{array}\right] \text{ if } \theta \neq 0 \text{ and } B(0) = \left[\begin{array}{cc} 0 & 0 \end{array}\right].$$

In the neighborhood of $\theta = 0$, the rank of $B(\theta)$ is not constant, since it is zero if $\theta \in \{1/k\pi \,|\, k \in \mathrm{Z}\}$ and one otherwise.

One must thus assume that, on the domain on which $U(\Theta)$ is defined, the rank of $B(\theta)$ is constant; equivalently, $U(\Theta)$ can only be defined in a set on which the rank of $B(\theta)$ is constant.

The second important remark is that $U(\theta)$ is not unique when it has more than two columns, since right-multiplication by a non-identical unitary matrix yields a distinct function whose columns also form an orthonormal basis of the nullspace of $B(\theta)$.

Fixing the value of $B(\theta)$ in a point is not sufficient to remove this ambiguity, as shown in this example : consider the 1×3 matrix-valued function $B(\theta) = [\cos(\theta)\ \sin(\theta)\ 0]$ defined on R, and take $\theta_0 = 0$. Define :

$$U_1(\theta) = \left[\begin{array}{ccc} -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{array}\right]^\top \text{ and } U_2(\theta) = U_1(\theta) \left[\begin{array}{cc} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{array}\right],$$

so that both functions are differentiable and "are" orthonormal bases of the nullspace of $B(\theta)$. These two function only coincide on the set $\{2k\pi | k \in \mathrm{Z}\}$.

In order to define a unique function $U(\Theta)$, another *normalizing condition* should be chosen, beyond $U(\Theta)^\top U(\Theta) = I$ and fixing the value of $U(\Theta_0)$ for some point Θ_0. In the present work, we take the normalizing condition that $U(\theta)$ minimizes, amongst all matrices whose columns form orthonormal bases of $B(\Theta)$, the Frobenius norm $\|U(\theta) - U(\Theta_0)\|_F$.

Result

Having settled the issues of local rank constancy and choice of normalizing conditions, our main result is stated as :

Theorem 1 If one is given a differentiable function $B(\Theta)$ defined on an open set $\mathcal{D} \in \mathrm{R}^M$, whose images are $N \times P$ real matrices, and for all $\Theta \in \mathcal{D}$, the nullspace of $B(\Theta)$ has some fixed dimension Q, and one is given a point $\Theta_0 \in \mathcal{D}$ and a matrix U_0 whose columns form an orthonormal basis of $\mathcal{N}ull\,(B(\Theta_0))$,

then there exists a neighbourhood $\mathcal{F}$ of Θ_0 and a $P \times Q$ real matrix-valued differentiable function $U(\Theta)$ defined on $\mathcal{F}$ such that, for all $\Theta \in \mathcal{F}$, one has :

$$\begin{aligned} B(\Theta)\,U(\Theta) &= \mathrm{O}_{N\times Q}, & (1)\\ U(\Theta)^\top U(\Theta) &= I_Q \quad \text{and} & (2)\\ U(\Theta) &= \arg\min_U \left\{ \|\mathrm{U}_0 - U\|_F \,|\, U \,\text{verifies}\,(2,3) \right\}. & (3) \end{aligned}$$

The function $U(\Theta)$ is called the *nullspace function* of (the function) $B(\Theta)$. It is computed using :

$$U(\Theta) = n(B(\Theta))\,\mathcal{U}\mathcal{V}^\top, \tag{4}$$

where $n(B(\Theta))$ can be any matrix whose columns form an orthonormal basis of $\mathcal{N}ull\,(B(\Theta_0))$ and $\mathcal{U}$, $\mathcal{V}$ are given by the SVD decomposition of $n(B(\Theta))^\top U_0$: $n(B(\Theta))^\top U_0 \overset{\text{SVD}}{=} \mathcal{U}D\mathcal{V}^\top$. The differential of $U(\Theta)$ is

$$\underbrace{\frac{\partial}{\partial\Theta_i}U}_{U'_i} = -B^+ \underbrace{\frac{\partial}{\partial\Theta_i}B}_{B'_i}\,U + UC \tag{5}$$

where B^+ is the pseudo-inverse of B and C is the $Q \times Q$ skew-symmetric matrix defined by :

$$\begin{aligned} \mathrm{vecl}\,(C) = {} & \left(W_Q^\top \left(I_Q \otimes \left(U^\top U_0\right)\right) W_Q\right)^{-1} \\ & \mathrm{vecl}\left(-U_0^\top B^+ B'_i U + U^\top {B'}_i^\top B^{+\top} U_0\right), \end{aligned} \tag{6}$$

where $\mathrm{vecl}\,(C)$ is the vector of subdiagonal elements of C and W_Q^+ is the $(Q(Q-1)/2) \times Q^2$ matrix that selects the sub-diagonal elements (W_Q is defined in Section 2.1). At $U = U_0$, one has $C = 0$, so that (5) simplifies to :

$$U_i'(\Theta_0) = -B^+ B_i' U.$$

The domain on which $U(\Theta)$ is defined extends as far as the minimum in (3) is unique, which is the case iff $n(B(\Theta))^\top U_0$ has full rank. On this domain, the derivatives U_i' are also defined by (5) and (6).

One should note that the computation of $U(\Theta)$ and its differential only requires a matrix $n(B(\Theta))$ whose columns form an orthonormal basis of $\mathcal{N}ull(B(\Theta))$, a matrix inverse and pseudo-inverse, and a singular value decomposition. All these operations are available in the most common linear algebra software packages.

To our knowledge, the present article is the first to propose a method to compute a differentiable function that verifies (1-2). This is not to say that this is an entirely new result and will see in the next section how it relates with previous work.

Related work

One first difference with respect to previous work, with the exception of [22], is that we consider rectangular matrices instead of square matrices -or operators. In theory, this difference is not relevant, since the nullspace of a rectangular matrix $B(\Theta)$ coincides with that of the square matrix $A(\Theta) = B^\top(\Theta) B(\Theta)$. However, it may be inconvenient to compute this product, so that this difference can have some practical implications.

The work closest to ours is certainly that of Haviv and Avrachenkov [11] on the perturbation of the nullspace of an analytically perturbed square matrix $A(\epsilon) = A_0 + \epsilon A_1 + \epsilon^2 A_2 + \ldots$. While we assume the rank of $B(\Theta)$ is constant, the distinction is made in [11] between rank-preserving and non-rank-preserving perturbations and both situations are studied. In the former case, a matrix $V(\epsilon) = V_0 + \epsilon V_1 + \epsilon^2 V_2 + \ldots$ is defined whose columns form a basis of the nullspace of $A(\epsilon)$. The matrix $V(\epsilon)$ is thus defined by an infinite sum, whereas we propose a closed-form expression. Another important difference with our work is that, for $\epsilon \neq 0$, $V(\epsilon)$ is not orthonormal and verifies $V(\epsilon)^\top V_0 = I$ instead of (2) which is verified by our function $U(\Theta)$ and by the function $n(\Theta)$ commonly found in numerical software packages.

In terms of methodology Haviv and Avrachenkov cite [6] and [5, Theorem S6.1] to ensure the existence of analytic vector-valued functions that form a basis of the nullspace of an analytic matrix-valued function; in our case the implicit function theorem is used to prove that $U(\Theta)$ exists and is differentiable.

Looking at less closely related work, the study of the nullspace of a matrix is linked to that of the perturbation of eigenvalues and eigenvectors, which is studied e.g. by Kato [14, I.2], in the case of analytic functions, and by Lancaster [15][1] who uses the implicit function theorem. The important issue

[1]Earlier references exist, such as Vishik and Lyusternik [28] Kato [13], Grauert [9], Rellich [25] or Jacobi [12] and many more cited by Kato [14], to which we did not have access.

of the bifurcation of repeated eigenvalues that is present in these works is not relevent in the present article, since we assume that the rank of $B(\Theta)$ remains constant.

Computation procedures for the derivatives of eigenvalues and eigenvectors have been proposed in many areas of applied research, such as the study of mechanical systems in aeronautics [27, 19], acoustics [4], econometrics [16] or computer vision [22][2]. The procedures for isolated eigenvalues differ on such points as whether any eigenvalue derivative is be computed or only that of the greatest eigenvalue; whether the matrix is symmetric or general; real or complex; whether left and right eigenvectors are used and whether second derivatives can be computed. In the case of repeated eigenvalues, different procedures have been devised for the cases of differing eigenvalue derivatives [20, 21, 18], equal eigenvalue derivatives but differing second derivatives [4] etc. In all these cases, the derivative is computed at a given point, but no mean is given to define and compute the eigenvalues, eigenvectors and their derivatives in neighboring points.

Having discussed the differences of our work with the most closely related studies, Sections 2-4 give the proof of Theorem 1. Then, Section 5 shows how it can be used to solve a class of problems of constrained optimization and some concluding remarks are given in Section 6.

2 Existence and differentiability

We will rephrase (1-3) as a system of equations $h(\theta, U) = \mathrm{O}$, so that the implicit function theorem may be applied. We start (Section 2.1) by finding a set of non-redundant equations that is equivalent to (1-3). Then, existence and differentiability of $U(\Theta)$ is proved, first in a special case (Section 2.1), followed by the general case (Section 2.3). Finally, the computation of $U(\Theta)$ and its differential is treated in Sections 3 and 4.

2.1 Characterization

We now review the definition of $U(\Theta)$ with the aim of obtaining an equivalent set of independent equations.

The system of NQ equations (in U) $B(\theta)U = \mathrm{O}_{N\times Q}$ has rank LQ, so that, unless $L = N$, these equations are redundant. For convenience, it is assumed until Section 2.3 that $N = L$.

Since $U^\top U - I$ is symmetric, (2) is redundant. A non-redundant formulation of this equation is obtained by eliminating the supra-diagonal elements, which yields :

$$D_Q^\top \mathrm{vec}\left(U^\top U - I_Q\right) = \mathrm{O}_{(Q(Q+1)/2)\times 1},$$

where D_Q is the $Q^2 \times Q(Q+1)/2$ duplication matrix [8].

[2] Again, older references [3, 26, 23] exist that were not available to us.

Moreover, it is shown in Appendix A that a matrix U, verifying (1) and (2), minimises $\|\mathrm{U}_0 - U\|_F$ if and only if $U_0^\top U$ is symmetric, which is equivalently stated as : $U_0^\top U - U^\top U_0 = \mathrm{O}_{Q\times Q}$. Because the left-hand side of this equation is skew-symmetric, there are only $Q(Q-1)/2$ distinct equations :

$$W_Q^+ \mathrm{vec}\left(U_0^\top U - U^\top U_0\right) = \mathrm{O}_{(Q(Q-1)/2)\times 1}, \tag{7}$$

where W_Q^+ is the $(Q(Q-1)/2) \times Q^2$ matrix that selects the sub-diagonal elements : if $A$is a $Q \times Q$ skew-symmetric matrix, then one has :

$$\mathrm{vec}(A) = W_Q \mathrm{vecl}(A), \quad \mathrm{vecl}(A) = W_Q^+ \mathrm{vec}(A) \ \text{ and } \ W_Q^+ = \frac{1}{2} W_Q^\top,$$

where $\mathrm{vecl}(A) = [A_{2,1}, ..., A_{Q,1}, A_{3,2}, ..., A_{Q,Q-1}]^\top$ is the vector of sub-diagonal elements of A.

Returning to the characterization of $U(\Theta)$, (1-3) are equivalent to the system of equations :

$$h(\theta, U) = \begin{bmatrix} \mathrm{vec}(B(\theta)U) \\ D_Q^\top \mathrm{vec}\left(U^\top U - I_Q\right) \\ W_Q^\top \mathrm{vec}\left(U_0^\top U - U^\top U_0\right) \end{bmatrix} = \mathrm{O}_{QP\times 1}. \tag{8}$$

The three components in (8) have length NP, $Q(Q+1)/2$ and $Q(Q-1)/2$ respectively, which sum up to QP. It is clear that $h(\theta, U)$ is differentiable in both θ and U.

2.2 Existence and differentiability

In order to apply the implicit function theorem, one must show that $\frac{\partial}{\partial U} h$ is bijective. Using Appendix B and a little algebra, this differential takes the form of the $PQ \times PQ$ matrix :

$$\frac{\partial}{\partial U} h(\theta, U) = \begin{bmatrix} I_Q \otimes B(\theta) \\ 2D_Q^\top \left(I_Q \otimes U^\top\right) \\ 2W_Q^\top \left(I_Q \otimes U_0^\top\right) \end{bmatrix} \tag{9}$$

The invertibility of $\frac{\partial}{\partial U} h$ is shown in Appendix C, so that the implicit function theorem can be applied and guarantees the existence of a neighbourhood $\mathcal{F}$ of θ_0 and of a differentiable function $U(\theta)$ defined on $\mathcal{F}$, such that for all $\theta \in \mathcal{F}$, one has $h(\theta, U(\theta)) = \mathrm{O}$. The differential of $U(\theta)$ is :

$$\frac{\partial}{\partial \theta} U(\theta) = -\frac{\partial}{\partial U} h(\theta, U)^{-1} \frac{\partial}{\partial \theta} h(\theta, U), \tag{10}$$

and the differential of h with respect to θ is :

$$\frac{\partial}{\partial \theta} h(\theta, U) = \begin{bmatrix} K_{N,Q}\left(I_N \otimes U^\top\right) K_{N,P} \\ \mathrm{O}_{Q^2\times NP} \end{bmatrix} \frac{\partial}{\partial \theta} B(\theta), \tag{11}$$

where $K_{N,Q}$ is the commutation matrix [16, 8] such that for all $N \times Q$ matrix A, $K_{N,Q}\mathrm{vec}(A) = \mathrm{vec}\left(A^\top\right)$.

We have just proven theorem 1 in the special case $\mathrm{rank}(B(\theta)) = N$ and we now extend it to the case $N \geq \mathrm{rank}(B(\theta)) = L$.

2.3 Generalisation to rank-deficient $B(\theta)$

We now assume that $N \geq \operatorname{rank}(B(\theta)) = L$ and show that there exists a function $U(\theta)$ defined as in Theorem 1.

If there exists a differentiable function $\tilde{B}(\theta) \in \mathrm{R}^{L\times P}$ such that

$$\mathcal{S}pan\left(\tilde{B}(\theta)^\top\right) = \mathcal{S}pan\left(B(\theta)^\top\right)$$

for all θ (and thus $\tilde{B}(\theta)$ has rank L), then $B(\theta)$ and $\tilde{B}(\theta)$ would have same null space for all θ. As a consequence, their nullspace functions $U(\theta)$ and $\tilde{U}(\theta)$, if they exist, would be equal.

All that is needed is to prove the existence of a function $\tilde{B}(\theta) \in \mathrm{R}^{L\times P}$. We first define $\tilde{B}(\theta)$ locally, e.g. in a neighbourhood of any $\theta_1 \in \mathcal{D}$. Since there exists a subset of L independent rows of $B(\theta_1)$, there exists a $L \times N$ matrix of zeros and ones -$S1$- that selects these rows, so that $\tilde{B}_1(\theta_1) = S_1 B(\theta_1)$ has L rows and rank L. Moreover, there exists an open neighbourhood of θ_1 on which its rank does not vary. Then, $\mathcal{D}$ can be covered by such overlapping neighbourhoods, so that a $L \times P$ function $\tilde{B}(\theta)$ is defined. Note also that only finitely many neighbourhoods are needed, since there are finitely many selection matrices. Although $\tilde{B}(\theta)$ is not continuous, it is differentiable almost everywhere (everywhere except when "switching" neighbourhoods) and its nullspace varies in a "continuous" fashion. As a consequence, the nullspace function $U(\theta)$ can be defined everywhere and it is differentiable.

This function $\tilde{B}(\theta)$ is needed only for the purpose of the demonstration and is not used in the actual computation of $U(\Theta)$ and its differential, which we address in the following sections.

3 Computing $U(\theta)$

In practice, $U(\theta)$ is easily computed from any matrix $n(B(\Theta))$ whose columns form an orthonormal basis of $\mathcal{N}ull(B(\theta))$. Once a possible matrix $n(B(\Theta))$ is known, for example given by the singular value decomposition [7] of $B(\theta)$, one uses the well-known [7, p. 601] fact that the unitary matrix $U(\theta)$ with same span as $n(B(\Theta))$ that minimises $\|U(\theta) - U_0\|_F$ is :

$$U(\theta) = n(B(\Theta))\,\mathcal{U}\mathcal{V}^\top, \tag{4}$$

where $\mathcal{U}$, $\mathcal{V}$ are given by the SVD decomposition of $n(B(\Theta))^\top U_0$: $n(B(\Theta))^\top U_0 \overset{\mathrm{SVD}}{=} \mathcal{U}D\mathcal{V}^\top$. Note that this expression is valid even if $\operatorname{rank}(B(\theta)) < N$.

4 Computing $\frac{\partial}{\partial\theta}U(\theta)$

Equation (10) provides a straightforward way of computing the differential of U, but this computation can be done at a lower computational cost and without requiring that $B(\theta)$ have rank N. This is done, like in [16, 22], by using necessary conditions on the partial derivatives of $U(\theta)$ that completely characterise these partial derivatives.

For convenience, we will write the partial derivative of $B(\theta)$ with respect to the i^{th} component of θ :

$$B_i' = \frac{\partial}{\partial \theta_i} B(\theta), \text{ and similarly } U_i' = \frac{\partial}{\partial \theta_i} U(\theta).$$

Computing the derivative of (1) with respect to θ_i, one gets :

$$B_i' U + B U_i' = \mathrm{O}_{N \times Q},$$

which in turn implies that :

$$U_i' = -B^+ B_i' U + U C \tag{5}$$

where B^+ is the pseudo-inverse of B and C is a $Q \times Q$ matrix that is computed below. Then, the derivative of (2) with respect to θ_i yields :

$$U_i'^\top U + U^\top U_i' = \mathrm{O}. \tag{12}$$

Replacing (5) into (12) implies that $C = -C^\top$, that is, that C is skew-symmetric. Finally, the derivative of (7) with respect to θ_i is :

$$U_0^\top U_i' - {U'}_i^\top U_0 = \mathrm{O}_{Q \times Q}.$$

Using (5) in this last expression yields the following equation in C :

$$X + Y^\top C - C^\top Y = \mathrm{O}_{Q \times Q},$$

where $X = -U_0^\top B^+ B_i' U + U^\top {B'}_i^\top B^{+\top} U_0$ and $Y = U_0^\top U$. A little algebra shows that the skew-symmetric C that solves this equation is defined by :

$$\mathrm{vecl}(C) = \left(W_Q^\top \left(I_Q \otimes Y^\top \right) W_Q \right)^{-1} \mathrm{vecl}(X). \tag{6}$$

Where $\mathrm{vecl}(C)$ is the vector of subdiagonal elements of C. Finally, note that, when $U = U_0$, this equation reduces to $C = \mathrm{O}$, so that, at $U = U_0$, it takes the simple form :

$$U_i' = -B^+ B_i' U. \tag{13}$$

Two remarks are in order : first, that (6) is defined as long as $Y = U_0^\top U$ is invertible, which is the same condition that is required for U_i' to be defined by the implicit functions theorem (end remark of Annex C).

The second remark is that $U_i'(\Theta_0)$ verifies $U_0^\top U_i'(\Theta_0) = \mathrm{O}$, which is the normalizing equation used by Haviv and Avrachenkov [11] to define the derivative. Thus, at the origin Θ_0, both normalizing conditions yield the same derivative.

Finally, it is of practical importance to note that, just like in the previous section, the equations used to compute the differential of $U(\Theta)$ do not require that $B(\Theta)$ be full rank.

This concludes our proof of Theorem 1 and we may give an example of its applications.

5 Application to constrained optimization

We consider the problem of constrained optimization with respect to variables $\Theta \in \mathcal{D} \subset \mathrm{R}^M$ and $\Phi \in \mathrm{R}^P$:

$$\text{minimize } S(\Theta, \Phi) \text{ with } B(\Theta)\Phi = \mathrm{O}_{N\times 1}.$$

While there exist many methods of constrained optimization [1], methods of unconstrained optimization are easier to implement [24] and are more commonly found in software packages. It is thus useful to transform the above problem into an equivalent problem of *unconstrained optimization* which can then be solved with standard tools. We now show how this can be done by parameterizing the feasible set of the original problem.

The feasible set of the original problem,

$$\left\{(\Theta, \Phi) \mid \Theta \in \mathcal{D} \subset \mathrm{R}^M,\, \Phi \in \mathrm{R}^P,\, B(\Theta)\Phi = \mathrm{O}_{N\times 1}\right\},$$

is, at least locally, covered by :

$$\left\{(\Theta, U(\Theta)\Psi) \mid \Theta \in \mathcal{F},\, \Psi \in \mathrm{R}^Q\right\}.$$

The original optimization problem is thus transformed into the smaller unconstrained problem :

$$\text{minimize}\, S(\Theta, U(\Theta)\Psi) \text{ over } F \times R^Q.$$

Since the cost function is a differentiable function of Θ and Ψ, efficient algorithms that use the cost derivatives can be used.

One should note that $U(\Theta)$ is not necessarily defined on the whole domain $\mathcal{D}$, so that, if the optimal value $\hat{\Theta}$ does not belong to $\mathcal{F}$, it may be necessary to use many ovrerlapping local mappings. It is easy to pass from one mapping $U^1(\Theta)$, centered in Θ_0^1, to another mapping $U^2(\Theta)$, centered in Θ_0^2, because one may fix the value of $U^2\left(\Theta_1^2\right)$ to $U^2\left(\Theta_1^2\right) = U^1\left(\Theta_1^2\right)$, so that the two mappings coincide in Θ_0^2.

The origin Θ_0 of the mapping may be changed at each iteration of the optimization algorithm, or whenever the angle between the nullspace of $B(\Theta)$ and U_0 becomes smaller than a threshold.

The method we have just presented thus allows to transform a constrained optimization problem into a smaller one of unconstrained optimization which can be solved using well-known algorithms.

6 Conclusion

By using the implicit function theorem, it has been possible to provide a largely self-contained proof of Theorem 1, and its usefullness is shown by applying it to a class of problems of constrained optimization.

Many directions of expansion of the proposed result exist : the complex case is certainly of interest; also, one may want to study in the same manner the evolution of the linear subspace associated to non-zero repeated singular values, which can be expected to be very similar to that of the nullspace.

To summarize this article, We have shown how to define a differentiable matrix-valued function $U(\Theta)$ whose columns form an orthonormal basis of the nullspace of a real, possibly rectangular, matrix-valued function $B(\Theta)$ of constant rank. Since, contrarily to previous work, we use the normalizing condition $U(\Theta)^\top U(\Theta) = I$, this function is a differentiable analogue of the non-continuous function found in many software packages, where it is often called `null()`.

A Condition on the minimising U

It is well-known [7, p. 601] that, given $M \times N$ matrices U_0 and $n(B(\Theta))$, the orthogonal matrix R minimises $\|U_0 - n(B(\Theta))R\|_F$ is $R^* = \mathcal{V}\mathcal{U}^\top$, where $\mathcal{U}$ and $\mathcal{V}$ are given by the singular value decomposition of $U_0^\top n(B(\Theta)) \stackrel{\text{SVD}}{=} \mathcal{U}D\mathcal{V}^\top$. Moreover, one sees that $U_0^\top n(B(\Theta))R^*$ is symmetric, simply by computing $U_0^\top n(B(\Theta))R^* - R^{*\top} n(B(\Theta))^\top U_0$ and verifying that it is zero.

B Differentials of $f(\mathbf{u}, \mathbf{v}) = \mathbf{vec}\left(F^\top G\right)$

Let $\mathbf{u} \in \mathrm{R}^{PM}$, $\mathbf{v} \in \mathrm{R}^{PN}$ be vectors, $F = [\mathbf{u}_1 \dots \mathbf{u}_M] = \mathrm{mat}_{P,M}(\mathbf{u})$ and $G = [\mathbf{v}_1 \dots \mathbf{v}_N] = \mathrm{mat}_{P,N}(\mathbf{v})$ be $P \times M$ and $P \times N$ matrices such that $\mathrm{vec}(F) = \mathbf{u}$ and $\mathrm{vec}(\mathrm{G}) = \mathbf{v}$. Define $f(\mathbf{u}, \mathbf{v}) = \mathrm{vec}\left(F^\top G\right) \in \mathrm{R}^{MN}$. The differential of $f(\mathbf{u}, \mathbf{v})$ are given by :

Because $f(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} \mathbf{u}_1^\top \mathbf{v}_1 \\ \vdots \\ \mathbf{u}_M^\top \mathbf{v}_1 \\ \vdots \\ \mathbf{u}_M^\top \mathbf{v}_N \end{bmatrix} = \left(I_N \otimes F^\top\right)\mathbf{v}$, one has $\dfrac{\partial}{\partial \mathbf{v}} f(\mathbf{u}, \mathbf{v}) = \left(I_N \times F^\top\right)$.

Likewise,

$$f(\mathbf{u}, \mathbf{v}) = K_{M,N}\left(I_M \otimes G^\top\right)\mathbf{u}, \text{ so that } \frac{\partial}{\partial \mathbf{u}} f(\mathbf{u}, \mathbf{v}) = K_{M,N}\left(I_M \otimes G^\top\right).$$

Finally, if one defines $g(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$, one has $\dfrac{\partial}{\partial \mathbf{u}} g(\mathbf{u}) = (I_{M^2} + K_{M,N})\left(I_M \otimes F^\top\right)$.

C Invertibility of $\dfrac{\partial}{\partial U} h$

The invertibility of $\dfrac{\partial}{\partial U} h(\theta, U)$ is now shown by showing that any vector $\mathbf{w}^\top = \mathrm{vec}(W) \in \mathrm{R}^{PQ}$ such that :

$$\begin{bmatrix} I_Q \otimes B(\theta) \\ D_Q^\top \left(I_Q \otimes U^\top\right) \\ W_Q^\top \left(I_Q \otimes U_0^\top\right) \end{bmatrix} \mathbf{w} = \mathrm{O}_{PQ\times 1}.$$

is necessarily zero. This is done by considering successively the three blocks in this equation.

1. The equation $(I_Q \otimes B(\theta))\,\mathbf{w} = \mathrm{O}_{NQ\times 1}$ is equivalent to $B(\theta)\,W = \mathrm{O}_{N\times Q}$, which, by definition of U, implies that :
$$W = UV, \tag{14}$$
for some nonzero $Q \times Q$ matrix V.
2. Then, $D_Q^\top \left(I_Q \otimes U^\top\right)\mathbf{w} = \mathrm{O}$ holds iff $U^\top W$ is skew-symmetric, i.e., (because of (14)) iff $U^\top UV = V$ is skew-symmetric.
3. Finally, $W_Q^\top \left(I_Q \otimes U_0^\top\right)\mathbf{w} = \mathrm{O}$ holds iff $U_0^\top W = U_0^\top UV$ is symmetric, i.e. if $U_0^\top UV - V^\top U^\top U_0 = \mathrm{O}$. Since V is skew-symmetric, one has :
$$U_0^\top UV + VU^\top U_0 = \left(\left(U_0^\top U\right) \oplus \left(U_0^\top U\right)\right) \operatorname{vec}(V) = \mathrm{O}_{Q\times Q},$$
where $\oplus$ denotes is the Kronecker sum [8].

Since the eigenvalues of $\left(U_0^\top U\right) \oplus \left(U_0^\top U\right)$ are the sums of pairs of eigenvalues of $U_0^\top U$ [8, Ch. 2.4], all that is missing, in order to show that $\frac{\partial}{\partial U}h$ is invertible, is to show that $\left(U_0^\top U\right)$ is not singular.

Let's assume that there exists a nonzero $\mathbf{v} \in \mathrm{R}^Q$ such that $U_0^\top U\mathbf{v} = \mathrm{O}_{Q\times 1}$, so that $\mathcal{N}ull\left(U_0^\top\right) \cap \mathcal{S}pan\,(U) \neq \{\mathrm{O}_{P\times 1}\}$, or, equivalently, $\mathcal{S}pan\left(B_0^\top\right) \cap \mathcal{N}ull\,(B) \neq \{\mathrm{O}_{P\times 1}\}$., which is also equivalent to saying that there exists a nonzero $\mathbf{x} \in \mathrm{R}^N$ such that $BB_0^\top \mathbf{x} = \mathrm{O}_{N\times 1}$ (remember that $B_0^\top$, being full-rank, has rank N).

Now, $B(\theta)\,B_0^\top$ is a continuous function of θ, and, in $\theta = \theta_0$, $B(\theta_0)\,B_0^\top = B_0 B_0^\top$ has only positive eigenvalues (because B_0 is full-rank). The eigenvalues of smallest eigenvalue of $B(\theta)\,B_0^\top$ evolve continuously [29, Part 2][3], there is a neighbourhood of θ_0 the smallest eigenvalue of this matrix is necessarily positive. $\square$

This demonstration also shows that $\frac{\partial}{\partial U}h(\theta, U)$ is invertible as long as the span of U and that of B_0 have no nonzero vector in common.

References

[1] D.P.Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Acad. Press, NY, 1982.

[3] For any $\delta > 0$, one can find a $\epsilon > 0$ such that, for all Θ verifying $\|\Theta - \Theta_0\| < \epsilon$, all the eigenvalues of $B(\theta)\,B_0^\top$ are within distance δ of one eigenvalue of $B(\theta_0)\,B_0^\top$.

[2] J. W. Eaton. Octave—a high-level interactive language for numerical computations. Technical Report CCSR-93-003, U. of Texas. http://www.octave.org/.

[3] R.L. Fox and M.P. Kapoor. Rates of changes of eigenvalues and eigenvectors. *AIAA*, 6:2426–2429, 1968.

[4] M. I. Friswell. The derivatives of repeated eigenvalues and their associated eigenvectors. *Journal of Vibration and Acoustics*, 118:390–397, 1996.

[5] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials.* Academic Press, 1982.

[6] I. Gohberg and L. Rodman. Analytic matrix functions with prescribed local data. *Journal of Applied Mathematics*, 40:90–128, 1981.

[7] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins Studies in the Mathematical Sciences. The Johns Hopkins University Press, Baltimore, MD, USA, third edition, 1996.

[8] Alexander Graham. *Kronecker products and matrix calculus with applications.* Wiley, 1981.

[9] H. Grauert. Analytische fascrungen über holomorphvollständigen raumen. *Math. Ann.*, 135:263–273, 1958.

[10] The Scilab Group. Introduction to scilab. Technical report, INRIA, Rocquencourt, France, 1998.

[11] M. Haviv and K.E. Avrachenkov. Perturbation of null spaces with application to the eigenvalue problem and generalized inverses. *Accepted for publication in Linear Algebra and its Applications*, 2000.

[12] K.G.J. Jacobi. Uber ein leichtes Verfahren die in der Theorie der Saecularstoerungen vorkommenden Gleichungen numerisch aufzuloesen. *Zeitrschriift fur Reine und Angewandte Mathematik*, 30:51–59, 1846. Also, NASA TT.F-13,666 June 1971.

[13] T. Kato. Perturbation theory of matrices. *J. Assoc. Comput. Mach.*, 5(104), 1958.

[14] Tosio Kato. *Perturbation Theory for Linear Operators.* Springer, 1966.

[15] P. Lancaster. On eigenvalues of matrices dependent on a parameter. *Numerische Mathematik*, 6:377–387, 1964.

[16] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons, 1999.

[17] The MathWorks. *Matlab Reference Guide.* The MathWorks, Inc, Natick MA, 1992. http://www.mathworks.com/.

[18] W.C. Mills-Curran. Calculation of eigenvectors derivatives for structures with repeated eigenvalues. *AIAA Journal*, 26(7):867–871, 1988.

[19] Richard B. Nelson. Simplified calculation of eigenvectors derivatives. *AIAA Journal*, 14(9):1201–1205, 1976.

[20] I.U. Ojalvo. Gradients for large structural models with repeated frequencies. In Soc. for Automotive Engineers, editor, *Aerospace Technology Conference and Exposition*, pages Paper 86–1789, Warrendale, PA, October 1986.

[21] I.U. Ojalvo. Efficient computation of modal sensitivities for systems with repeated frequencies. *AIAA Journal*, 26(3):361–366, 1988.

[22] T. Papadopoulo and M.I.A. Lourakis. Estimating the jacobian of the singular value decomposition: Theory and applications. In *ECCV (1)*, pages 554–570, 2000.

[23] R.H. Plaut and K. Huseyin. Derivatives of eigenvalues and eigenvectors in non-sel-adjoint systems. *AIAA journal*, 11:250–251, 1973.

[24] W. H. Press, S. A. Teutolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes, The Art of Scientific Computing, 2nd edition*. Cambridge University Press, 1992.

[25] F. Rellich. Störungstheorie der Spektralzerlegung, I-V. *Math. Ann.*, (Numbers 113,116-118), 1937-1942.

[26] L.C. Rogers. Derivatives of eigenvalues and eigenvectors. *AIAA Journal*, 8:943–944, May 1970.

[27] Carl S. Rudisill. Derivatives of eigenvectors and eigenvalues of a general matrix. *AIAA Journal*, 12:721–722, 1974.

[28] M.I. Vishik and L.A. Lyusternik. The solution of some perturbation problems in the case of matrices and self-adjoint and non-self-adjoint differential equations. *Uspechi Mat. Nauk*, 15(3):3–80, 1960.

[29] J. H. Wilkinson. *The algebraic eigenvalue problem*. Oxford: Clarendon, 1965,1988.

[30] Stephen Wolfram. *Mathematica*. Addison Wesley, New York, 1988.