Proceedings of the Project Review, Geo-Mathematical Imaging Group (Purdue University, West Lafayette IN), Vol. 1 (2011) pp. 37-56.

ALTERNATING PROJECTIONS ON LOW-DIMENSIONAL MANIFOLDS

FREDRIK ANDERSSON* AND MARCUS CARLSSON†

Abstract. Let B_0 be a point in some space, which in the applications we have in mind could be a signal or image. We consider sequences $(B_k)_{k=0}^{\infty}$ of points obtained by projecting back and forth between two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, and give conditions guaranteeing that the sequence converge to a limit $B_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2$. Our motivation is the study of algorithms based on finding the limit of such sequences, which have proven useful in a number of areas, [3]. The intersection is typically a set with desirable properties, but for which there is no efficient method of finding the closest point B_{opt} in $\mathcal{M}_1 \cap \mathcal{M}_2$. We prove not only that the sequence of alternating projections converges, but that the limit point is fairly close to B_{opt}, in a manner relative to the distance $\|B_0 - B_{opt}\|$, thereby significantly improving earlier results in the field. A concrete example with applications to frequency estimation of signals is also presented.

1. Introduction. Let $\mathcal{K}$ be a finite dimensional Hilbert space over $\mathbb{R}$ and let $\mathcal{M}_1, \mathcal{M}_2 \subset \mathcal{K}$ be manifolds. Suppose that for any $B \in \mathcal{K}$ the closest point on $\mathcal{M}_j$, $j = 1$ or $j = 2$, is well defined and lets denote it by $\pi_j(B)$. A problem of great practical interest is how to find the closest point in $\mathcal{M}_1 \cap \mathcal{M}_2$, which we denote by $\pi(B)$. It is a classical result by von Neumann that if $\mathcal{M}_1$ and $\mathcal{M}_2$ are affine linear manifolds, then the sequence of alternating projections

$$\pi_1(B),\ \pi_2(\pi_1(B)),\ \pi_1(\pi_2(\pi_1(B))),\ \pi_2(\pi_1(\pi_2(\pi_1(B)))), \dots \tag{1.1}$$

converges to $\pi(B)$, and moreover the speed of the convergence is determined by the angle between $\mathcal{M}_1$ and $\mathcal{M}_2$. This paper is concerned with extensions of this result to non-linear manifolds.

The research is motivated by applications to algorithms with the goal of, given $B \in \mathcal{K}$, find a point B_∞ in the common intersection $\mathcal{M}_1 \cap \mathcal{M}_2$, by applying alternating projections as above and hope that the so arising sequence converges. That is, we set $B_1 = \pi_1(B)$ and then

$$B_{k+1} = \begin{cases} \pi_1(B_k) & k \text{ is even} \\ \pi_2(B_k) & k \text{ is odd} \end{cases} \tag{1.2}$$

Ideally, one would of course like that B_∞ is the closest point in $\mathcal{M}_1 \cap \mathcal{M}_2$ to B (denoted $\pi(B)$), or at least close to it. Such algorithms have many applications in signal/image processing, where $\mathcal{K}$ typically is the set $\mathbb{M}_{m,n}$ of $m \times n$-matrices, and the manifolds are subsets with a certain structure, e.g. matrices with a certain rank, self-adjoint matrices, Hankel or Toeplitz matrices etc. We refer to [3] for a number of interesting applications of this type. In this paper we will consider a concrete example where $\mathcal{M}_1$ is the set of matrices with a certain rank and M_2 is the set of Hankel matrices, with applications to the noise elimination of signals and the issue of finding the "best" approximation to a given function (on an interval) by sums of exponential functions.

Such applications are not new, but has appeared in various articles in the applied area. In this context, Zangwill's Global Convergence Theorem [6] is usually recalled to motivate convergence. However, this is a very weak motivation. Zangwill's theorem in this setting basically says that if the sequence $(B_k)_{k=1}^{\infty}$ is bounded and the distance to $\mathcal{M}_1 \cap \mathcal{M}_2$ is always decreasing, then $(B_k)_{k=1}^{\infty}$ has a convergent subsequence to a point $B_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2$. But this is an almost immediate application of the fact that any bounded sequence in a compact set has a convergent subsequence. Moreover, there is nothing indicating that so arising limit B_∞ is at all close to $\pi(B)$, which is the point we seek. An improvement to the above conclusions, which applies in much more restrictive cases, has recently been discovered by A. Lewis and J. Malick. We present their results in a moment, but first two simple examples showing the limitations of the above procedure in the general setting.

*Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden and Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA

†Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA, Now at Universidad de Santiago de Chile, Avenida Bernardo OHiggins 3363, Estacion Central, Santiago, Chile, and Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden

Example 1.1. *Consider* $\mathcal{K} = \mathbb{R}^2$ *and set* $\mathcal{M}_1 = \{(t, (t+1)(3-t)/4) : t \in \mathbb{R}\}$ *and* $\mathcal{M}_2 = \mathbb{R} \times \{0\}$. *It is not hard to check that* $\pi_1((1,0)) = (1,1)$ *and* $\pi_2((1,1)) = (1,0)$, *and hence the sequence of alternating projections does not converge, see Figure 1. On the other hand, if we start at the point* $(1+\epsilon, 0) \in \mathcal{M}_2$, $\epsilon > 0$, *we see that the sequence of alternating projections will converge to* $(3,0) \in \mathcal{M}_1 \cap \mathcal{M}_2$.

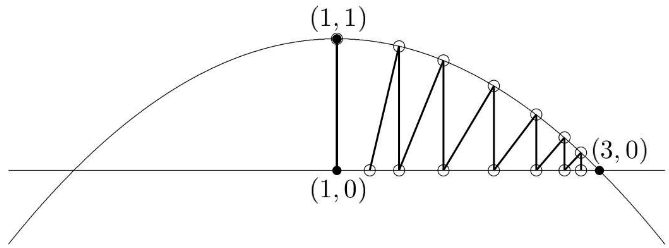


Fig. 1. *An example demonstrating that the algorithm can get stuck in loops where even no subsequence converges to a point in the intersection.*

In general, it seems reasonable to assume that if we start with a point close enough to the an intersection point, then the sequence does converge a point in the intersection. The next example shows that this is not the case if we have limited smoothness.

Example 1.2. *Without going in to the details of the construction, we note that one can construct a* C^1*-function* f *such that, with* $\mathcal{M}_1 = \mathbb{R} \times \{0\}$ *and* $\mathcal{M}_2 = \{(t, f(t)) : t \in \mathbb{R}\}$, *the sequence of alternating projections can get stuck in projecting back and forth between the same two points. Figure 2 explains the idea.*

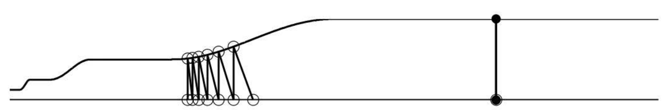


Fig. 2. *The "stuck in loop"-issue can occur arbitrarily close to an intersection point. We suppose that the two curves meet at* $t = 0$.

However, if we have $f \in C^1$ and $f'(0) \neq 0$, it is hard to imagine how to make a similar construction work. Indeed, by results in this paper it will follow that in this case the alternating projections will converge, given that we start close enough to the intersection point $(0,0)$. In the setting of Lewis and Malick [5], the condition $f'(0) \neq 0$ implies that $\mathcal{M}_1$ and $\mathcal{M}_2$ are *transversal* at $(0,0)$. Transversality is a central concept in their work, which we now present.

Given in the general setting considered initially, suppose that $\mathcal{M}_1$ and $\mathcal{M}_2$ are at least C^3-manifolds, and let $A \in \mathcal{M}_1 \cap \mathcal{M}_2$ be given. A is called *transversal* if

$$T_{\mathcal{M}_1}(A) + T_{\mathcal{M}_2}(A) = \mathcal{K}, \tag{1.3}$$

where $T_{\mathcal{M}_j}(A)$ denotes the tangent-space of $\mathcal{M}_j$ at A, $j = 1, 2$. The main result of [5] is roughly the following:

Theorem 1.3. *If A is transversal and B is close enough to A, then the sequence of alternating projections* $(B_k)_{k=1}^{\infty}$ *given by (1.2) converges to a point* B_∞ *in* $\mathcal{M}_1 \cap \mathcal{M}_2$. *Moreover,*

$$\|B_\infty - \pi(B)\| \leq 2\|A - B\|.$$

The improvement over Zangwill's theorem is thus that we get that the entire sequence converges and we do not have to assume boundedness. Moreover the limit B_∞ is not too far off from $\pi(B)$, (although in relative terms, i.e. comparing with $\mathsf{dist}(B, \mathcal{M}_1 \cap \mathcal{M}_2) = \|B - \pi(B)\|$, it need not be particularly close either). On the other hand, we have to assume that we are close to $\mathcal{M}_1 \cap \mathcal{M}_2$ and moreover that there is a transversal point there. To demonstrate the essence of the transversality assumption, we now present two cases which are not covered by the above result, but where the conclusion still holds.

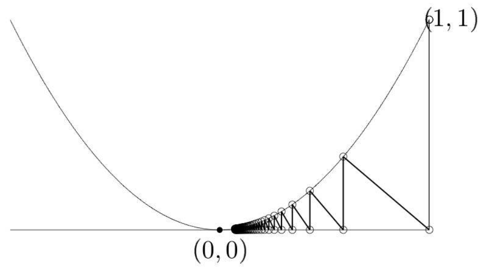


FIG. 3. *Transversality fails due to the curves being "tangential", but the sequence still converges.*

EXAMPLE 1.4. *With* $\mathcal{M}_1 = \mathbb{R} \times \{0\}$ *and* $\mathcal{M}_2 = \{(t, t^2) : t \in \mathbb{R}\}$*, transversality is not satisfied at* $A = (0,0)$ *(since* $T_{\mathcal{M}_1}(A) = T_{\mathcal{M}_2}(A) = \mathcal{M}_1$*), but it is not hard to see that the sequence of alternating projections still converges to* $(0,0)$*, (see Figure 3).*

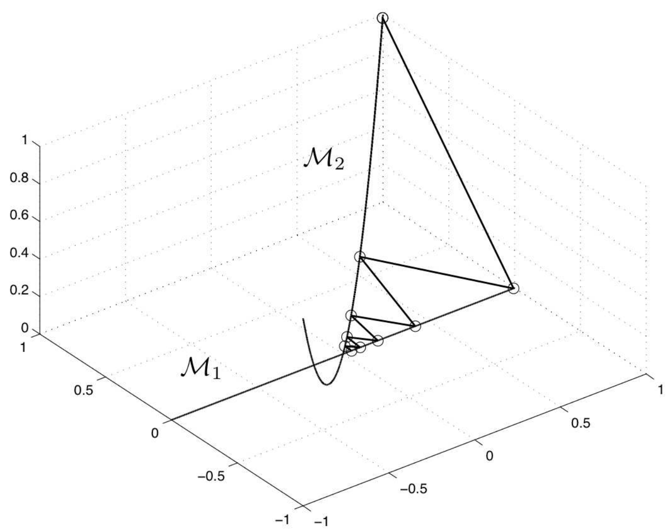


FIG. 4. *Transversality fails due to the curves having "too low dimension", but the sequence still converges.*

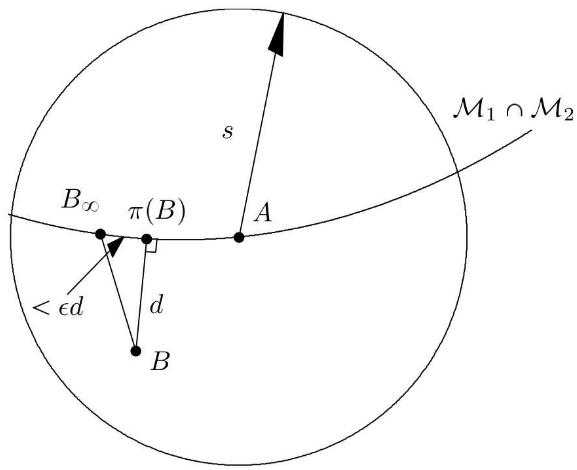
The difference between Example 1.2 and 1.4 is that in the latter case the manifolds are more regular.

EXAMPLE 1.5. *With* $\mathcal{K} = \mathbb{R}^3$ *and* $\mathcal{M}_1 = \mathbb{R} \times \{0\}^2$ *and* $\mathcal{M}_2 = \{(t, t, t^2) : t \in \mathbb{R}\}$*, transversality is not satisfied at* $A = (0,0)$*, but again it seems plausible that the sequence of alternating projections converges to* $(0,0)$*. See Figure 4.*

In Example 1.5, the two manifolds clearly sit at a positive angle, but this situation is not covered by Theorem 1.3, since the manifolds have too low dimension to satisfy the transversality assumption. In fact, if $\mathcal{K}$ has dimension n and $\mathcal{M}_j$ has dimension m_j, $j = 1, 2$, the transversality (1.3) can never be satisfied if $m_1 + m_2 < n$. But for many applications of practical interest, one has $m_1 + m_2 << n$. We will introduce a concept which we call *non-tangential*, which loosely speaking says that the manifolds should have a positive angle in directions perpendicular to $\mathcal{M}_1 \cap \mathcal{M}_2$. The point $(0,0)$ in Example 1.5 is non-tangential, whereas the same point in Example 1.4 is not. Loosely speaking, the main result goes as follows.

THEOREM 1.6. *Given a non-tangential point $A \in \mathcal{M}_1 \cap \mathcal{M}_2$ there exists a $s > 0$ such that the sequence of alternating projections (1.2) converges to a point $B_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2$, given that $\|B - A\| < s$. Moreover, given any $\epsilon > 0$ one can take s such that*

$$\|B_\infty - \pi(B)\| < \epsilon \|B - \pi(B)\|.$$



The improvement over Theorem 1.3 mainly consists of two items. Primarily, the assumption that the surfaces be non-tangential is not at all restrictive, and in particular there is no implication on the dimensions of $\mathcal{M}_1$ and $\mathcal{M}_2$. (Despite the title, the results apply to manifolds with high dimension as well. For the applications we are aware of, the set of non-non-tangential points is very very small, if it exists at all.) Secondly, as has been highlighted before, we are usually interested not just in any point of $\mathcal{M}_1 \cap \mathcal{M}_2$, but the closest point $\pi(B)$. Here the theorem says that in relative terms, i.e. after dividing with the distance to $\mathcal{M}_1 \cap \mathcal{M}_2$, we can get arbitrary accuracy if we start close enough to $\mathcal{M}_1 \cap \mathcal{M}_2$.

In order to motivate the need for certain technicalities in the abstract setting, as well as the need for a theorem concerning low-dimensional manifolds, we will develop the theory in parallel with an example, namely that when $\mathcal{K}$ is the space $\mathbb{M}_{n,n}(\mathbb{C})$ of complex $n \times n$-matrices, $\mathcal{M}_1$ is the set of Hankel matrices and $\mathcal{M}_2$ the set of matrices of rank at most k. The former is a subspace of real dimension $2(n-1)$, whereas the latter is not actually a manifold. However, it is "locally" a manifold of real dimension $2(2nk - k^2)$ at all matrices A with rank precisely k, (which clearly constitutes the majority of matrices in the set). If k is small and A is a non-tangential intersection point, then $\dim(\mathcal{M}_1) + \dim(\mathcal{M}_2) = 2(2nk - k^2 + n - 1)$, which clearly is much less than $\dim \mathcal{K} = 2n^2$, unless k is close to n. Another thing that does not match between this example and the theory outlined above is that the projection π_2 is not uniquely defined at certain points. This is common in algorithmic theory and can be dealt with by dealing with point-to-set maps, following Zangwill [6]. However, in our case, this is not a theoretical complication to deal with, but it complicates the notation and moreover the set of points where ambiguity appears is vanishingly small, and *never encountered in applications.* It thus seems silly to work in this general setting, and since our theorem is of local nature anyway, we will simply assume that the projections are well defined locally.

For the above example, the interest in $\mathcal{M}_1 \cap \mathcal{M}_2$ lies in the fact that such sequences are samplings of functions which are sums of k exponential functions. Given a function f on an interval the alternating projections algorithm can then be used to find approximations of f by sums of k exponential functions, which is a problem of great practical interest. We demonstrate the idea with a picture, see Figure 5. More thorough examples are conducted [1].

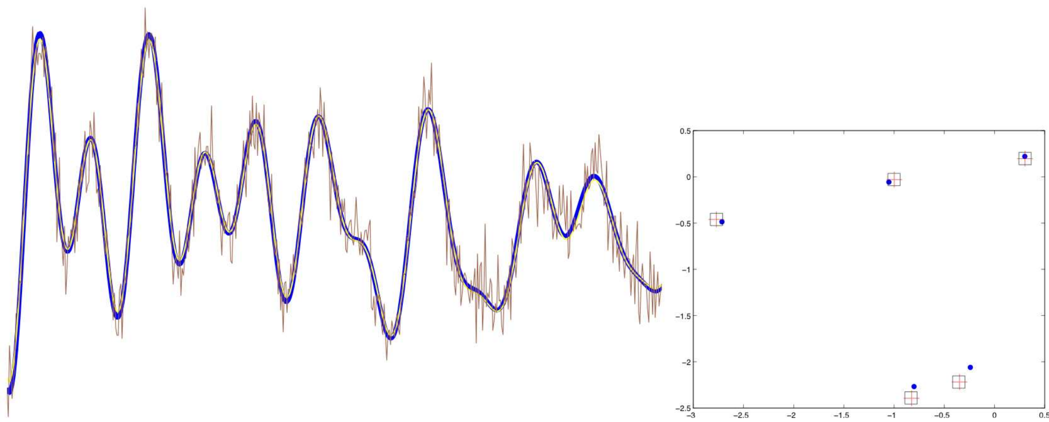


FIG. 5. *Signal with five exponentials. Left panel: original signal (yellow); signal with white noise (brown); and the reconstruction using the alternating Hankel projection method (blue). Right panel: original exponential nodes (blue dots), estimated nodes from noisy signal (black squares); and estimated nodes from noisy orthogonalized signal (red plus).*

2. Case study; rank k matrices versus Hankel matrices I. In the setting considered by Lewis and Malick, the fact that $M_1 \cap \mathcal{M}_2$ is itself a manifold (around a point A, say) follows by the tranversality assumption (1.3) and standard differential geometry. In the present setting, this is not the case, but the example developed in this section shows that this still can happen in natural circumstances. In fact, it is not hard to see that it will always happen for algebraic manifolds, which is the case for all of the various applications presented in [3].

Let $n \in \mathbb{N}$ be fixed, let $\mathcal{K} = \mathbb{M}_{n,n}(\mathbb{C})$ and let $\mathcal{H} \subset \mathcal{K}$ be the set of Hankel matrices, e.g. matrices of the form

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_2 & a_3 & \ddots & a_n & a_{n+1} \\ a_3 & \ddots & \ddots & \ddots & \vdots \\ \vdots & a_n & \ddots & \ddots & a_{2n-2} \\ a_n & a_{n+1} & \cdots & a_{2n-2} & a_{2n-1} \end{pmatrix} = \sum_{j=1}^{2n-1} a_j E_j, \tag{2.1}$$

where $E_j \in \mathcal{H}$ is the natural basis such that

$$E_j(i_1, i_2) = \begin{cases} 1 & \text{if } i_1 + i_2 = j + 1 \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{H}$ is a linear subspace and hence is a manifold at each point, of (real) dimension $2(2n-1)$. Denoting the matrix in (2.1) by $H(a)$ where $a = (a_1, \ldots, a_{2n-1})$, we obtain a natural chart for $\mathcal{H}$, (by identifying $\mathbb{C}$ with $\mathbb{R}^2$ in the obvious way).

Given $k < n$ we denote by $\mathcal{R}_k \subset \mathcal{K}$ the set of matrices of rank less than or equal to k. We wish to do alternating projections as outlined in the introduction between $\mathcal{H}$ and $\mathcal{R}_k$, which is slightly complicated by the fact that $\mathcal{R}_k$ is not a manifold. However, except for some exceptional points, $\mathcal{R}_k$ is locally a manifold of (real) dimension $2(2nk - k^2)$, which we now show. Since this is not the main topic of the paper, the exposition will be a bit imprecise. Suppose $A \in \mathcal{R}_k$, and use the singular value decomposition of A to find $\sigma_A \in (\mathbb{R}^+)^k$ and $U_A, V_A \in \mathbb{M}_{n,k}$ such that $U_A^* U_A = V_A^* V_A = I$

(where I is the identity matrix) and

$$A = V_A \begin{pmatrix} \sigma_{A,1} & 0 & \cdots & 0 \\ 0 & \sigma_{A,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{A,k} \end{pmatrix} U_A^* = V_A I_\sigma U_A^*. \tag{2.2}$$

The vast majority of matrices in $\mathcal{R}_k$ satisfy

$$\sigma_{A,1} > \sigma_{A,2} > \ldots > \sigma_{A,k} > 0. \tag{2.3}$$

The subset of $\mathcal{R}_k$ satisfying (2.3) will be denoted $\mathcal{R}_k^d$, where d stands for distinct. If $\mathcal{M}$ is a manifold and its closure $\overline{\mathcal{M}}$ is such that $\overline{\mathcal{M}} \setminus \mathcal{M}$ is a union of manifolds of lower dimension than $\mathcal{M}$, we will say that $\overline{\mathcal{M}} \setminus \mathcal{M}$ is *thin.*

PROPOSITION 2.1. *$\mathcal{R}_k^d$ is a manifold of (real) dimension $2(2nk - k^2)$. Moreover, $\mathcal{R}_k = \overline{\mathcal{R}_k^d}$ and $\mathcal{R}_k \setminus \mathcal{R}_k^d$ is thin.*

Proof. To avoid silly exceptions, we assume that all entries in U_A and V_A are non-zero. In order for the representation (2.2) to be unique, we can for example demand that $V_A(1,:) \in \mathbb{R}^k$, (where $V_A(1,:) = (V(1,1), \ldots, V(1,k))$). Letting the entries in U_A, V_A and σ_A be free variables, we obtain the set $\mathcal{R}_k$ around A as the image of the map defined by the right hand side of (2.2). However, the map is not injective and thus not good enough to be a chart. To remedy this, one can by bare hands see that the set of $\mathbb{M}_{n,k}-$matrices satisfying $U^*U = I$ is a $2nk - k^2$-dimensional manifold, (or consult Theorem 2.1.2 and 2.1.6.4 in [2]). Let $\mathfrak{U} : \mathbb{R}^{2nk-k^2} \to \mathbb{M}_{n,k}$ be a chart containing U_A in its image. Moreover let $\mathfrak{V} : \mathbb{R}^{2nk-k^2-k} \to \mathbb{M}_{n,k}$ be an analogous chart for orthogonal V's with the additional restriction $V(1,:) \in \mathbb{R}^k$. Assuming that $A \in \mathcal{R}_k^d$ and considering $\sigma = \sigma_A$ as a free variable in $(\mathbb{R}^+)^k$, it is now easy to show (using the uniqueness of 2.2 with the restriction on V) that $\mathfrak{R} = \mathfrak{V} I_\sigma \mathfrak{U}$ is a chart for $\mathcal{R}_k^d$ around A, which thus has dimension

$$(2nk - k^2) + k + (2nk - k^2 - k) = 2(2nk - k^2).$$

We omit a proof of the remaining statements, which can be obtained by standard differential geometry.□

The "bare-hands" construction of $\mathfrak{U}$ and $\mathfrak{V}$ actually yields an $\mathfrak{R}$ that cover all but some small exceptional subset of $\mathcal{R}_k$. However, if $\sigma_k = 0$ or $\sigma_j = \sigma_{j+1}$ for some $j < k$, the manifold structure brakes down, because $\mathfrak{R}$ is then not injective and hence not a chart. For example, if $\sigma_{A,k} = 0$ in (2.2) then the last column in U_A and V_A do not affect A. (The structure of the set $\mathcal{R}_k$ around such points can be rather complicated, and although the alternating projection scheme could theoretically approach such a point, we have not done any analysis of convergence properties in this setting. This seems to be a hard problem.)

We note that by the Eckart-Young theorem [4] it is easy to project any matrix $B \in \mathbb{M}_{n,n}$ onto its closest point in $\mathcal{R}_k$. In fact, the theorem says that if $B = V_B I_\sigma U_B$ is a singular value decomposition of B, then the closest point in $\mathcal{R}_k$ is given by replacing I_σ with I_τ where $\tau = (\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$. The closest point is thus unique as long as the singular values are different, which is always the case when working with real data, so we will for simplicity treat the projection onto $\mathcal{R}_k$ as a well defined map (which it will be locally in the neighborhoods we working with). We denote this map by $\pi_{\mathcal{R}_k}$.

The last manifold that needs to be discussed is of course

$$\mathcal{H}_k = \mathcal{R}_k \cap \mathcal{H},$$

i.e. the set of hankel matrices with rank $\leq k$. It is easily seen that, given any $\alpha \in \mathbb{C}$, the matrix

$$H(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & \alpha^2 & \ddots & \alpha^{n-1} & \alpha^n \\ \alpha^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \alpha^{n-1} & \ddots & \ddots & \alpha^{2n-3} \\ \alpha^{n-1} & \alpha^n & \cdots & \alpha^{2n-3} & \alpha^{2n-2} \end{pmatrix} \tag{2.4}$$

defines a rank 1 hankel matrix. Moreover, Kronecker's theorem states that "most" rank k Hankel matrices are of the form

$$\mathfrak{H}(c,\alpha) = \sum_{j=1}^{k} c_j H(\alpha_j). \tag{2.5}$$

In fact, the image of the above chart is dense in $\mathcal{H}_k$, which thus for the most part is a $4k$-dimensional manifold. Lets denote the image of $\mathfrak{H}$ intersected with $\mathcal{R}_k^d$ by $\mathcal{H}_k^n$, where n stands for "nice". Again, the analysis is complicated by some exceptional points. For example, $\frac{d^j}{d\alpha^j}H(\alpha)$ is a Hankel matrix of rank j which is not covered by $\mathfrak{H}$. Moreover rank k Hankel matrices containing summands of this type are usually inside $\mathcal{H}_k$, but despite that, some investigations show that the manifold structure of $\mathcal{H}_k$ collapses around such points. This despite that such points are most of the time in $\mathcal{R}_k^d$, and hence the sup-script "n"$(=$ *nice*) really is more restrictive than "d"$(=$ *distinct*). Nevertheless, $\mathcal{H}_k \setminus \mathcal{H}_k^n$ is thin and never encountered in practice, so we omit a study of such cases. The following proposition sums up the conclusions of this section, and a few other observations that we state without proof.

Proposition 2.2.

$\mathcal{H}$ is a $2(2n-1)$-dimensional linear subspace,

$\mathcal{H}_k^n$ is a $4k$-dimensional manifold which is dense in $\mathcal{H}_k$, its complement is thin.

The map $\pi_{\mathcal{R}_k}$ is well defined on all points in $\mathbb{M}_{n,n}$ except a for thin subset.

We will continue this example in Section 7.

3. Preliminaries. $\mathcal{K}$ is a Hilbert space of dimension $n \in \mathbb{N}$. Given $A \in \mathcal{K}$ and $r > 0$ we write $\mathcal{B}(A,r)$ or $\mathcal{B}_{\mathcal{K}}(A,r)$ for the open ball centered at A with radius r. Since $\mathcal{K}$ is finite-dimensional it has a unique Euklidean topology. Any subset $\mathcal{M}$ of $\mathcal{K}$ will be given the induced topology.

Definition 3.1. *We say that $\mathcal{M} \subset \mathcal{K}$ is locally an m-dimensional C^p-manifold around $A \in \mathcal{M}$ if there exists $r_1 > 0$ and a C^p-map $\phi : \mathcal{B}_{\mathbb{R}^m}(0,r_1) \to \mathcal{K}$ with the following properties:*

- *$d\phi(x)$ is injective for all $x \in \mathcal{B}_{\mathbb{R}^m}(0,r_1)$,*
- *$\phi(0) = A$,*
- *ϕ is a homeomorphism onto an open neighborhood of A in $\mathcal{M}$.*

This is in line with the standard definition of C^p-manifolds, see Theorem 2.1.2 [2] for a number of equivalent definitions. As a consequence of the homeomorphism condition, note that there exists an $s_1 > 0$ such that

$$\mathcal{M} \cap \mathcal{B}_{\mathcal{K}}(A,s_1) = \mathsf{Im}\phi \cap \mathcal{B}_{\mathcal{K}}(A,s_1), \tag{3.1}$$

where $\mathsf{Im}\phi$ denotes the image of ϕ. Given $B \in \mathsf{Im}\phi$, there exists a unique $x \in \mathcal{B}(0,r_1)$ such that $B = \phi(x)$. We will without comment denote this x by x_B. All the manifolds considered in this paper are at least C^1, and hence we have

$$\phi(x) = C + d\phi(x_C)(x - x_C) + o(x - x_C) \tag{3.2}$$

where o stands for "little ordo".[1] We define the tangent space $T_{\mathcal{M}}(B)$ by $T_{\mathcal{M}}(B) = \mathsf{Ran}\ d\phi(x_B)$. It is a standard fact from differential geometry that this definition is independent of ϕ. Moreover we

[1] i.e. it stands for a function with the property that $o(x)/\|x\|$ extends by continuity to 0 and takes the value 0 there.

set

$$\tilde{T}_{\mathcal{M}}(B) = B + T_{\mathcal{M}}(B),$$

i.e. $\tilde{T}_{\mathcal{M}}(B)$ is the affine linear manifold which is tangent to $\mathcal{M}$ at B. Throughout this section, $\mathcal{M}$ will be a locally C^p-manifold at A, where $p \geq 1$ and we associate with it r_1 and s_1 as in Definition 3.1 and (3.1). The following proposition basically says that the affine tangent-spaces are close to $\mathcal{M}$ locally.

Proposition 3.2. *Let $\mathcal{M}$ be a locally C^1-manifold at A. For each $\epsilon_2 > 0$ there exists s_2, $0 < s_2 < s_1$, such that for all $C \in \mathcal{B}(A, s_2) \cap \mathcal{M}$ we have*

(i) $\mathsf{dist}(B, \tilde{T}_{\mathcal{M}}(C)) \leq \epsilon_2 \|B - C\|, \quad B \in \mathcal{B}(A, s_2) \cap \mathcal{M}.$

(ii) $\mathsf{dist}(B, \mathcal{M}) \leq \epsilon_2 \|B - C\|, \qquad B \in \mathcal{B}(A, s_2) \cap T_{\mathcal{M}}(C).$

Proof.

Given $r < r_1$, we first show that

$$\|\phi(x) - \phi(y)\| \asymp \|x - y\| \tag{3.3}$$

in $\in \mathcal{B}_{\mathbb{R}^m}(0, r)$, i.e. that $\|\phi(x) - \phi(y)\|/\|x - y\|$ is uniformly bounded above and below for $x, y \in \mathcal{B}_{\mathbb{R}^m}(0, r)$. By (3.2) and the mean value theorem we have

$$\|\phi(y) - \phi(x)\| = \|d\phi(z)(y - x)\|$$

for some z on the line between x and y. Now, $d\phi(z)$ depends continuously on z and its singular values $\sigma_1(d\phi(z)), \ldots, \sigma_m(d\phi(z))$ depend continuously on the matrix entries ([4]), hence

$$\inf_{z \in \mathcal{B}_{\mathbb{R}^m}(0,r)} \sigma_n(d\phi(z))\|y - x\| \leq \|\phi(y) - \phi(x)\| = \sup_{z \in \mathcal{B}_{\mathbb{R}^m}(0,r)} \sigma_1(d\phi(z))\|y - x\|.$$

By Definition 3.1, $\sigma_m(d\phi(x))$ is never zero and $cl(\mathcal{B}_{\mathbb{R}^m}(0, r))$ is compact, so both the inf and sup amount to finite positive numbers, as desired. (cl denotes the closure).

We now prove (i); by (3.2) and the mean value theorem we have

$$\begin{aligned}\mathsf{dist}(B, \tilde{T}_{\mathcal{M}}(C)) &\leq \|B - (C + d\phi(x_C)(x_B - x_C))\| = \|\phi(x_B) - \phi(x_C) - d\phi(x_C)(x_B - x_C))\| \leq \\ &\leq \sup\Big\{\|d\phi(y) - d\phi(x_C)\| : y \in \mathbb{R}^m \text{ such that } \|y - x_C\| \leq \|x_B - x_C\|\Big\}\|x_B - x_C\|.\end{aligned}$$

Given $r < r_1$, note that $d\phi$ is continuous on the compact set $cl(\mathcal{B}_{\mathbb{R}^m}(0, r))$. Thus for each $\epsilon > 0$ we can pick a $\delta > 0$ such that

$$\sup\big\{\|d\phi(y) - d\phi(x)\| : x, y \in \mathcal{B}_{\mathbb{R}^m}(0, r) \text{ such that } \|y - x\| \leq \delta\big\} < \epsilon.$$

Set $r < \min(\delta_\epsilon/2, r_1)$ and let s, $0 < s < s_1$, be such that $\|B\| < s$ implies $\|x_B\| < r$, which we can do since ϕ is a homeomorphism. For $B, C \in \mathcal{B}_{\mathcal{K}}(0, s)$, we then have

$$\mathsf{dist}(B, \tilde{T}_{\mathcal{M}}(C)) \leq \epsilon\|x_B - x_C\| \asymp \|\phi(x_B) - \phi(x_C)\| = \|B - C\|,$$

as desired.

It remains to prove (ii), which can be done in a similar fashion, we omit the details. □

Proposition 3.3. *Let $\mathcal{M}$ be a locally C^p-manifold at A with $p \geq 1$. Then there exists s_3 and a C^{p-1} map*

$$\pi : \mathcal{B}_{\mathcal{K}}(A, s_3) \to \mathcal{M}$$

such that for all $B \in \mathcal{B}_{\mathcal{K}}(A, s_3)$ there exists a unique closest point in $\mathcal{M}$ which is given by $\pi(B)$. Moreover, $C \in \mathcal{M} \cap \mathcal{B}_{\mathcal{K}}(A, s_3)$ equals $\pi(B)$ if and only if $B - C \perp T_{\mathcal{M}}(C)$.

Proof. By standard differential geometry there exists an $r_2 < r_1$ and C^{p-1}-functions $f_1, \ldots, f_{n-m} : \mathcal{B}_{\mathbb{R}^m}(0, r_2)$ with the property that

$$\left(T_{\mathcal{M}}(\phi(x))\right)^{\perp} = \mathsf{Span}\ \{f_1(x), \ldots, f_{n-m}(x)\}$$

for all $x \in \mathcal{B}_{\mathbb{R}^m}(0, r_2)$. Define $\sigma : \mathcal{B}_{\mathbb{R}^m}(0, r_2) \times \mathbb{R}^{n-m} \to \mathcal{K}$ via

$$\sigma(x, y) = \phi(x) + \sum_{i=1}^{n-m} y_i f_i(x).$$

Consider the set $\mathcal{S} \subset \mathcal{K}$ of points whose multiplicity under σ is greater than 1, (i.e. all points hit more than once by σ). By the inverse function theorem, the set $\sigma^{-1}(\mathcal{S})$ can not have 0 as an accumulation-point, for it says that there exists a $r < r_2$ such that σ restricted to $\mathcal{B}_{\mathbb{R}^n}(0, r)$ is a diffeomorphism onto its image. Pick $r_3 < \min\left(r, \mathsf{dist}(\sigma^{-1}(\mathcal{S}), 0)\right)$ and pick s_3 such that

$$\mathcal{M} \cap \mathcal{B}_{\mathcal{K}}(A, 2s_3) = \phi(\mathcal{B}_{\mathbb{R}^m}(0, r_3)) \cap \mathcal{B}_{\mathcal{K}}(A, 2s_3), \tag{3.4}$$

and

$$\mathcal{B}_{\mathcal{K}}(A, s_3) \subset \sigma(\mathcal{B}_{\mathbb{R}^n}(0, r_3)). \tag{3.5}$$

Note that, given $B \in \mathcal{B}_{\mathcal{K}}(A, s_3)$ there exists a unique (x_B, y_B)such that $B = \sigma((x_B, y_B))$ and moreover $\|(x_B, y_B)\| \leq r_3$ by (3.5). We define

$$\pi(B) = \phi(x_B).$$

To see that π is a C^{p-1}-map, let $\theta : \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}^m$ be given by $\theta((x, y)) = x$ and note that

$$\pi = \phi \circ \theta \circ (\sigma|_{\mathcal{B}(0,r_3)})^{-1}.$$

We now show that $\pi(B)$ have the desired properties. By the construction,

$$\pi(B) - B \in \mathsf{Span}\ \{f_1(x_B), \ldots, f_{n-m}(x_B)\} \perp T_{\mathcal{M}}(\phi(x_B)) = T_{\mathcal{M}}(\pi(B)).$$

Now suppose $C \in \mathcal{M}$ is a closest point to B. Since $\|A-B\| < s_3$ we clearly must have $\|C-A\| < 2s_3$ so by (3.4) there exists a $x_C \in \mathcal{B}_{\mathbb{R}^m}(0, r_3)$ with $\phi(x_C) = C$. Since $r_3 < r_1$ we know that $\mathcal{M}$ is completely determined by ϕ in the vicinity of C. In particular, it makes sense to talk about $T_{\mathcal{M}}(C)$ and it is easily seen that $B - C \perp T_{\mathcal{M}}(C)$, for by Proposition 3.2 we have

$$\begin{aligned}\|\phi(x) - B\|^2 &= \|C + d\phi(x_C)(x - x_C) + o(x - x_C) - B\| = \\ &= \|C - B\|^2 + 2\mathsf{Re}\ \langle C - B, d\phi(x_C)(x - x_C)\rangle + o(\|x - x_C\|)\end{aligned}$$

and hence the scalar product needs to be zero for all x's. Thus there is a y such that $B = \sigma((x_C, y))$. But since (x_B, y_B) is the unique point with this property, we deduce that $x_B = x_C$ and hence $\pi(B) = \phi(x_B) = \phi(x_C) = C$. This establishes the first part of the proposition. Now let C be as in the second part of the proposition. As above we have $C = \phi(x_C)$ with $\|x_C\| < r_3$ and the orthogonality implies that there exists a y with $B = \sigma((x_C, y))$ and again this implies $C = \pi(B)$, as desired. □

4. Non-tangentiality. Suppose now that we are given closed sets $\mathcal{M}_1$ and $\mathcal{M}_2$ as well as an intersection point $A \in \mathcal{M}_1 \cap \mathcal{M}_2$.

Definition 4.1. *A is called regular if there are numbers m_1, m_2 and m such that*

- *$p \geq 1$ and $\mathcal{M}_j$ is locally a m_j-dimensional C^p-manifold at A, $j = 1, 2$.*
- *$\mathcal{M}_1 \cap \mathcal{M}_2$ is locally a m dimensional manifold at A.*

Note that the set of regular points is clearly a relatively open set in $\mathcal{M}_1\cap\mathcal{M}_2$, (with the topology induced from $\mathcal{K}$). Next, we introduce angles. For more information on angles, we refer to [5].

Definition 4.2. *For any regular point A, we define the angle $\alpha(A)$ of $\mathcal{M}_1$ and $\mathcal{M}_2$ at A to be the $\cos^{-1}$ of the number*

$$c(A)=\lim_{r\to 0}\sup\left\{\frac{\langle B_1-A,B_2-A\rangle}{\|B_1-A\|\|B_2-A\|}:B_j\in\mathcal{M}_j,\|B_j-A\|<r \text{ and } B_j-A\perp T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)\right\}.$$

It is easy to see that it suffices to compute the above lim sup with $B_1\in T_{\mathcal{M}_1}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)$ and $B_2\in T_{\mathcal{M}_2}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)$. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are hyperplanes through the origin and $P_{\mathcal{M}_j}$, $(j=1,2)$, denotes the orthogonal projection onto $\mathcal{M}_j$ then it is easy to see that the above definition coincides with the classical definition;

$$\cos\alpha_{clas}(A)=c_{clas}(A)=\|P_{\mathcal{M}_1}P_{\mathcal{M}_2}-P_{\mathcal{M}_1\cap\mathcal{M}_2}\|. \tag{4.1}$$

It is important to note that $\alpha_{clas}(A,T_{\mathcal{M}_1(A)},T_{\mathcal{M}_2(A)})$ and $\alpha(A)$ are not necessarily the same. For example, take $\mathcal{M}_1=\{(x_1,x_2,0):x\in\mathbb{R}^2\}$ and $\mathcal{M}_2=\{(x_1,x_2,x_1^2):x\in\mathbb{R}^2\}$. Then $\alpha_{clas}(0,T_{\mathcal{M}_1(0)},T_{\mathcal{M}_2(0)})=\pi/2$, whereas $\alpha(0)=0$. However, what goes wrong above is clearly that the two surfaces are tangential to each other, and it is intuitively clear that when this is not the case, the two concepts should coincide. To avoid such obstacles, we therefore introduce:

Definition 4.3. *$\mathcal{M}_1$ and $\mathcal{M}_2$ are said to be non-tangential at A if A is regular and they have a positive angle at A, i.e. if $c(A)<1$. We will often simply say that A is non-tangential.* Note that the angle is always a number between 0 and $\pi/2$, so in an intuitive sense only exceptional points do not satisfy non-tangentiality. We have no intention of formalizing this statement, but point out already that for the example considered in Section 2, this is indeed the case, which will be shown in Section 7. Pathological examples do exist. For example, take $\mathcal{M}_1=\{(x_1,x_2,0):x\in\mathbb{R}^2\}$ and $\mathcal{M}_2=\{(x_1,x_2,x_1^2+x_2^2):x\in\mathbb{R}^2\}$. However, when $\mathcal{M}_1$, $\mathcal{M}_2$ and are defined by polynomials, one can use similar methods as in Section 7 to show that under mild conditions, if non-tangentiality holds at one point, then it holds everywhere except for a thin set. We will not pursue this.

By the remark following Definition 4.2, it is easy to see that A is non-tangential if and only if

$$\Big(T_{\mathcal{M}_1}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)\Big)\cap\Big(T_{\mathcal{M}_2}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)\Big)=\{0\} \tag{4.2}$$

which in turn happens if and only if

$$T_{\mathcal{M}_1}(A)\cap T_{\mathcal{M}_2}(A)=T_{\mathcal{M}_1\cap\mathcal{M}_2}(A). \tag{4.3}$$

Proposition 4.4. *A transversal point A is also non-tangential.*

Proof. Applying the implicit function theorem to $\phi_1-\phi_2$ easily yields that $\mathcal{M}_1\cap\mathcal{M}_2$ is a manifold of dimension m_1+m_2-n at A. Thus $T_{\mathcal{M}_1}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)$ has dimension $n-m_2$ and $T_{\mathcal{M}_2}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A)$ has dimension $n-m_1$. Thus

$$\dim(T_{\mathcal{M}_1}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A))+\dim(T_{\mathcal{M}_2}(A)\ominus T_{\mathcal{M}_1\cap\mathcal{M}_2}(A))+\dim(T_{\mathcal{M}_1\cap\mathcal{M}_2}(A))=n.$$

But by transversality the sum of the above subspaces equals $\mathcal{K}$, which can only happen if (4.2) is satisfied. □

From now on we assume without saying that A is a non-tangential point. We will denote $\mathcal{M}_1\cap\mathcal{M}_2$ by $\mathcal{M}$, and the objects from Section 3 associated to $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}$, e.g. ϕ, by ϕ_1, ϕ_2 and ϕ respectively. We thus omit subindex when dealing with $\mathcal{M}_1\cap\mathcal{M}_2$. We now prove that for non non-tangential points, the angle as defined here and the classical angle of the respective tangent spaces coincide.

Proposition 4.5. *If A is non-tangential, then*

$$\alpha_{clas}(A,T_{\mathcal{M}_1(A)},T_{\mathcal{M}_2(A)})=\alpha(A).$$

Proof. $\alpha(A)$ is easily seen to be the angle (as in Definition 4.2) between $T_{\mathcal{M}_1} \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}$ and $T_{\mathcal{M}_2} \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}$. Moreover, it is easy to see that

$$\begin{aligned} c(A) &= \|P_{T_{\mathcal{M}_1} \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}} P_{T_{\mathcal{M}_2} \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}}\| = \|P_{T_{\mathcal{M}_1}} P_{T_{\mathcal{M}_2}} - P_{T_{\mathcal{M}_1 \cap \mathcal{M}_2}}\| = \\ &= \|P_{T_{\mathcal{M}_1}} P_{T_{\mathcal{M}_2}} - P_{T_{\mathcal{M}_1} \cap T_{\mathcal{M}_2}}\| = c_{clas}(A, T_{\mathcal{M}_1}, P_{T_{\mathcal{M}_2}}), \end{aligned}$$

where we used (4.3) in the crucial step. □

PROPOSITION 4.6. *The function c in Definition 4.2 is C^{p-1} (on the set of regular points in $\mathcal{M}_1 \cap \mathcal{M}_2$). In particular, non-tangentiality is a local property, i.e. if A is non-tangential, then the same holds for all $B \in \mathcal{M}_1 \cap \mathcal{M}_2$ in a neighborhood of A.*

Proof. It suffices to show that c is C^p at the point A. It is a standard fact from differential geometry (see e.g. [2]) that ϕ is a diffeomorphism. Thus it suffices to show that $c \circ \phi$ is continuous. By the same book, there exists C^{p-1}-functions $f_1, \ldots, f_{n-m}$ defined on the domain of ϕ such that

$$\left(T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\phi(x))\right)^{\perp} = \mathsf{Span}\ \{f_1(x), \ldots, f_{n-m}(x)\}.$$

It is easy to see that ($j = 1$ or $j = 2$)

$$T_{\mathcal{M}_j}(\phi(x)) \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\phi(x)) = \mathsf{Span}\ \{P_{T_{\mathcal{M}_j}(\phi(x))} f_1(x), \ldots P_{T_{\mathcal{M}_j}(\phi(x))} f_1(x)\}. \tag{4.4}$$

Since the dimension of $T_{\mathcal{M}_j}(\phi(x))$ is constant m_j, the functions on the right are continuous. Moreover, since the dimension of these space on the left in (4.4) is constant $m_j - m$, we can pick $m_j - m$ of the functions on the right and put them as columns of an injective matrix $M_j(0)$ such that

$$T_{\mathcal{M}_j}(\phi(0)) \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\phi(0)) = \mathsf{Ran}\ M_j(0).$$

The above identity can then be extended to all x in a neighborhood of 0, (since injectivity is a local property and the dimension is independent of x). Locally at 0 we thus have

$$T_{\mathcal{M}_j}(\phi(x)) \ominus T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\phi(x)) = \mathsf{Ran}\ M_j(x).$$

As was noted after Definition 4.2, $\cos^{-1}(c(\phi(x)))$ equals the angle of the two spaces above, and hence by standard facts of angles between subspaces (see e.g. [5]) we have for x around 0 that

$$c(\phi(x)) = \|P_{\mathsf{Ran}\ M_1(x)} P_{\mathsf{Ran}\ M_2(x)}\|. \tag{4.5}$$

(Note that this is not true with the classical definition when the two ranges have a non-trivial intersection, but with the definition here it works.) Moreover, it is easy to see that

$$P_{\mathsf{Ran}\ M_j} = M_j (M_j^* M_j)^{-1} M_j^*,$$

where x has been omitted for readability. Combining this with (4.5) it is clear that c is C^{p-1} at 0, as desired. □

5. π's and ρ's. Let $\mathcal{M}$, $\mathcal{M}_1$, $\mathcal{M}_2$, ϕ etc., be as stated before Proposition 4.5. In particular, $A \in \mathcal{M}$ is a non-tangential point. In Propositions 3.2 and 3.3 there appears s_2 and s_3, which are not necessarily the same and who depend on an auxiliary constant ϵ_2 in Proposition 3.2. In this section we let ϵ_2 be a fixed number which we will determine later, and we let s_4 denote the minimum of all possible s's from Section 3 related to the 3 manifolds. Thus we can apply any result from that section to either of the manifolds considered here. Moreover, letting j denote either $1, 2$ or nothing, we let $r_4 > 0$ be such that

$$\mathcal{M}_j \cap \mathcal{B}_{\mathcal{K}}(A, s_4) = \mathsf{Im}\phi_j \cap \mathcal{B}_{\mathcal{K}}(A, s_4),$$

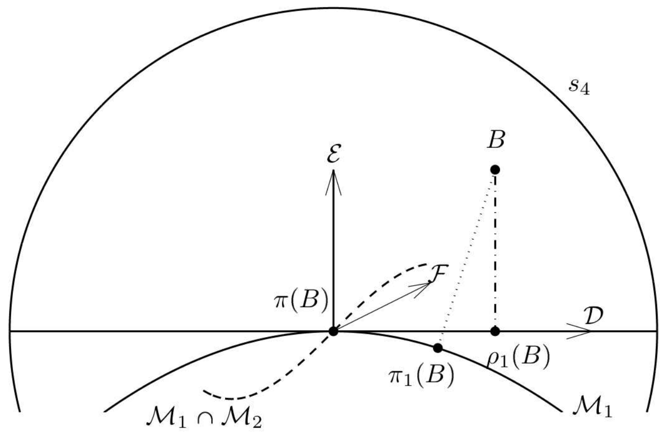


FIG. 6. *Illustration of the difference between ρ_1 and π_1. $\mathcal{D}$ and $\mathcal{E}$ appear in the proof of Proposition 5.2.*

and such that the results of Section 3 applies to each $\phi_j(x)$ with $\|x\| < r_4$. We also assume that $\epsilon_2 < 1$.

Given an affine linear manifold $\mathcal{N} \subset \mathcal{K}$ we denote by $P_{\mathcal{N}}$ the orthogonal projection onto $\mathcal{N}$. We introduce maps $\rho_j : \mathcal{B}(A, s_4) \to \mathcal{K}$, ($j = 1$ or $j = 2$), via

$$\rho_j(B) = P_{\tilde{T}_{\mathcal{M}_j}(\pi(B))}(B).$$

Thus ρ_j is similar to π_j but slightly different. π_j projects onto $\mathcal{M}_j$ whereas ρ_j projects onto the tangent plane of $\mathcal{M}_j$ taken at the closest point to B in $\mathcal{M}_1 \cap \mathcal{M}_2$. A proper estimate of the difference is given in Proposition 5.2.

LEMMA 5.1. *ρ_1 and ρ_2 are C^{p-1}-maps in $\mathcal{B}_{\mathcal{K}}(A, s_4)$. Moreover, we can pick a $s_5 < s_4$ such that the image of $\mathcal{B}(A, s_5)$ under $\rho_1, \rho_2, \pi, \pi_1, \pi_2$, as well as any composition of two of those maps, is contained in $\mathcal{B}(A, s_4)$.*

Proof. The second part is an immediate consequence of the continuity of the maps. Let j denote 1 or 2 and set $M_j(B) = d\phi_j(\pi(B))$. Note that M_j is a C^{p-1}-map in $\mathcal{B}_{\mathcal{K}}(A, s_4)$ by Proposition 3.3 and the choice of s_4. Moreover,

$$\rho_j(B) = P_{\tilde{T}_{\mathcal{M}_j}(\pi(B))}(B) = P_{T_{\mathcal{M}_j}(\pi(B))}(B - \pi(B)) = M_j(B)\big(M_j^*(B)M_j(B)\big)^{-1}M_j^*(B)\big(B - \pi(B)\big),$$

from which the result follows. □

PROPOSITION 5.2. *Given any $B \in \mathcal{B}(A, s_4)$ and $j = 1$ or $j = 2$, we have*

$$\|\pi_j(B) - \rho_j(B)\| < 5\epsilon_2\|B - \pi(B)\|.$$

Proof. By Lemma 5.1 we have that Proposition 3.2 applies to the point $C = \pi(B)$. It is no restriction to assume that $\pi(B) = 0$, which we now do. Denote $\mathcal{D} = T_{\mathcal{M}_j}(0)$, $\mathcal{E} = \mathsf{Span}\,\{B - \rho_j(B)\}$ and $\mathcal{F} = \mathcal{K} \ominus (\mathcal{D} \oplus \mathcal{E})$, (see Fig. 5). Let D_B and E_B be elements of $\mathcal{D}$ and $\mathcal{E}$ such that $B = D_B + E_B$, and note that

$$\rho_j(B) = D_B.$$

We thus have to show that

$$\|\pi_j(B) - D_B\| < 4\epsilon_2\|B\|. \tag{5.1}$$

First note that by Proposition 3.2 (*ii*) there exists a point in $\mathcal{M}$ in $\mathcal{B}(D_B, \epsilon_2\|D_B\|)$. Thus

$$\|B - \pi_j(B)\| \leq \|B - D_B\| + \epsilon_2\|D_B\| = \|E_B\| + \epsilon_2\|D_B\|.$$

Combining this with Proposition 3.2 (*i*) we have that $\pi_j(B)$ lies in the intersection of the sets

$$\Big\{(D, E, F) : \|D - D_B\|^2 + \|E - E_B\|^2 + \|F\|^2 < (\|E_B\| + \epsilon_2\|D_B\|)^2\Big\} \tag{5.2}$$

and

$$\Big\{(D, E, F) : \|E\|^2 + \|F\|^2 < \epsilon_2^2\|D\|^2\Big\}. \tag{5.3}$$

The left hand side of (5.1) is thus dominated by the supremum of the function

$$f(D, E, F) = \|D - D_B\|^2 + \|E\|^2 + \|F\|^2$$

subject to the conditions in (5.2) and (5.3). Either by geometrical considerations or the method of Lagrange multipliers, it is not hard to deduce that this supremum is attained for $F = 0$ and D, E of the form $D = dD_B$ and $E = eE_B$ where $d, e \in \mathbb{R}$. We now have a two-dimensional problem of circles and cones, and in the remainder of the proof we treat D and E as elements of $\mathbb{R}^2$. Given $(D, E, 0)$ in the intersection of (5.2) and (5.3), it is easily seen (see Fig. 7) that

$$\|E\| \leq \epsilon_2(\|D_B\| + \|E_B\| + \epsilon_2\|D_B\|).$$

The problem becomes simpler if we replace the cone in (5.3) by the following strip:

$$\Big\{E : \|E\| < \epsilon_2((1 + \epsilon_2)\|D_B\| + \|E_B\|)\Big\}. \tag{5.4}$$

Looking a while at Figure 7 and recalling some freshman formulas, one realizes that the sought

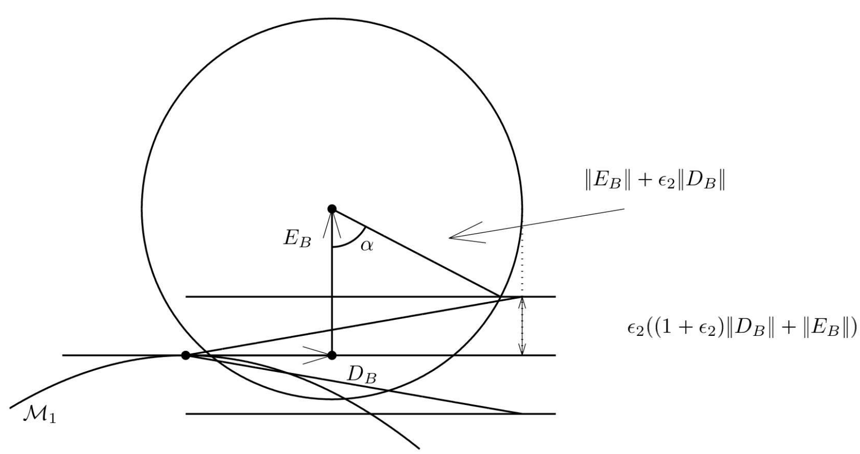


FIG. 7. *Illustration of what is going on.*

supremum is dominated by $\alpha(\|E_B\| + \epsilon_2\|D_B\|)$, where α is the angle given by

$$\alpha = \cos^{-1}\left(\frac{\|E_B\| - \epsilon_2((1 + \epsilon_2)\|D_B\| + \|E_B\|)}{(\|E_B\| + \epsilon_2\|D_B\|)}\right),$$

where we define $\cos^{-1}(t) = \pi/2$ for all $t > 1$. Since we have assumed that $\epsilon_2 < 1$, the proposition thus follows if we establish that

$$\cos^{-1}\left(\frac{\|E_B\| - \epsilon_2(2\|D_B\| + \|E_B\|)}{(\|E_B\| + \epsilon_2\|D_B\|)}\right)\frac{\|E_B\| + \epsilon_2\|D_B\|}{\sqrt{\|E_B\|^2 + \|D_B\|^2}} < 4\epsilon_2$$

Dividing here and there with $\|E_B\|$ one sees that this is equivalent to

$$\sup_{t>0}\left\{\cos^{-1}\left(\frac{1-2\epsilon_2 t-\epsilon_2}{1+\epsilon_2 t}\right)\frac{1+\epsilon_2 t}{\sqrt{1+t^2}}\right\}<4\epsilon_2$$

Using more freshman calculus we obtain

$$\cos^{-1}\left(\frac{1-2\epsilon_2 t-\epsilon_2}{1+\epsilon_2 t}\right)\frac{1+\epsilon_2 t}{\sqrt{1+t^2}}=\cos^{-1}\left(1-\frac{3\epsilon_2 t+\epsilon_2}{1+\epsilon_2 t}\right)\frac{1+\epsilon_2 t}{\sqrt{1+t^2}}\le$$
$$\le\frac{\pi}{2}\frac{3\epsilon_2 t+\epsilon_2}{1+\epsilon_2 t}\frac{1+\epsilon_2 t}{\sqrt{1+t^2}}=\epsilon_2\frac{\pi(3t+1)}{2\sqrt{1+t^2}}\le\epsilon_2\frac{\pi}{2}\frac{10}{\sqrt{10}}<5\epsilon_2,$$

since it is readily verified that $\frac{3t+1}{\sqrt{1+t^2}}$ attains its supremum at $t=3$ and $\frac{\pi}{2}\frac{10}{\sqrt{10}}<5$. □

LEMMA 5.3. *Given $B\in\mathcal{B}(A,s_4)$ we have*

$$\pi(B)=\pi(\rho_1(B))=\pi(\rho_2(\rho_1(B))).$$

Proof. If we prove the first equality the second follows by reversing the roles of 1 and 2 and applying the first equality to $\rho_1(B)$. To see the first equality, note that $B-\rho_1(B)\perp T_{\mathcal{M}_1}(\pi(B))$ and obviously $B-\pi(B)\perp T_{\mathcal{M}_1\cap M_2}(\pi(B))$ so

$$\rho_1(B)-\pi(B)=(\rho_1(B)-B)+(B-\pi(B))\perp T_{\mathcal{M}_1\cap M_2}(\pi(B)),$$

which by Proposition 3.3 implies $\pi(B)=\pi(\rho_1(B))$, as desired. □

LEMMA 5.4. *Let $c_4>1$ and $\epsilon_4>0$ be given. If $E,F\in\mathcal{K}$ satisfies $\|E\|>c_4\|F\|$ and $\|E-F\|<\epsilon_4$, then*

$$\|E\|<\epsilon_4\frac{c_4}{c_4-1}.$$

Proof. If $\|E\|<\epsilon_4$ we are done, otherwise

$$c_4<\frac{\|E\|}{\|F\|}<\frac{\|E\|}{\|E\|-\epsilon_4}$$

which easily gives the desired estimate. □

The next result will be the main tool for proving convergence of the alternating projections.

THEOREM 5.5. *Let $A\in\mathcal{M}_1\cap\mathcal{M}_2$ be a non-tangential point and assume that $\mathcal{M}_1$ and $\mathcal{M}_2$ are C^2-manifolds. Then for each $c_6>c(A)$ and each $\epsilon_6>0$ there exists a positive $s_6<s_5$ such that for all $B\in\mathcal{M}_2\cap\mathcal{B}(A,s_6)$ we have*

(i) $\|\pi_1(B)-\pi(B)\|<c_6\|B-\pi(B)\|$

(ii) $\|\pi(\pi_1(B))-\pi(B)\|<\epsilon_6\|B-\pi(B)\|$

Moreover the same holds true with the roles of $\mathcal{M}_1$ and $\mathcal{M}_2$ reversed.

Proof. Fix c_1 such that $c(A)<c_1<c_6$ and pick an $s_7<s_5$ such that

$$\sup\{c(C):C\in\mathcal{M}_1\cap\mathcal{M}_2\cap\mathcal{B}(A,s_7)\}<c_6, \tag{5.5}$$

which we can do since c_6 is continuous. Let $c_2>1$ be such that $c_2c_1<c_6$. By Lemma 5.1 and Proposition 3.3, ρ_1 and π are C^1-functions, and hence we can pick $c_3>0$ such that

$$\|\rho_j(B)-\rho_j(B')\|\le c_3\|B-B'\| \text{ and } \|\pi(B)-\pi(B')\|\le c_3\|B-B'\| \tag{5.6}$$

for all $B,B'\in\mathcal{B}(A,s_4)$, (recall that s_4 was chosen in the beginning of this section). Finally, we fix ϵ_2 such that

$$5\epsilon_2(1+c_3)\frac{c_2}{c_2-1}<c \text{ and } 5\epsilon_2(c_3+c_3^2)<\epsilon_6 \text{ and } (1+5\epsilon_2)c_2c_1<c_6. \tag{5.7}$$

This may seem like a circle argument, because c_1, c_2 and c_3 depends on c_6 which depends on ϵ_2 via s_5, and the c_j's appears in (5.7). However, this is easily circumvented by first choosing s_5 with $\epsilon_2 = 1$, say, and pick values of c_1, c_2, c_3. Then, once the real ϵ_2 has been chosen via (5.7) we can redefine s_4, s_5 and s_7 accordingly without violating (5.5).

Now, pick an s_8 such that

$$\pi(\mathcal{B}(A, s_8)) \subset \mathcal{B}(A, s_7),$$

and let $B \in \mathcal{M}_2 \cap \mathcal{B}(A, s_7)$. Denote $C = \pi(B)$ and $D = \pi_1(B)$ and note that $C \in \mathcal{M}_2 \cap \mathcal{B}(A, s_8)$ so $c(C) < c_4$. There is no restriction to assume that $C = 0$, which we do from now on. Put $B' = \rho_2(B)$ and $D' = \rho_1(B')$. First note that by Proposition 5.2

$$\|B - B'\| = \|\pi_2(B) - \rho_2(B)\| < 5\epsilon_2\|B\| \tag{5.8}$$

and moreover by (5.7) and Lemma 5.1 we have that

$$\begin{aligned} \|D - D'\| = \|\rho_1(B') - \pi_1(B)\| &\le \|\rho_1(B') - \rho_1(B)\| + \|\rho_1(B) - \pi_1(B)\| \le \\ &\le c_3\|B' - B\| + 5\epsilon_2\|B\| < 5\epsilon_2(1 + c_3)\|B\| \end{aligned} \tag{5.9}$$

By Lemma 5.3 we have $\pi(B) = \pi(D')$ so part (ii) follows by (5.6), (5.9) and the calculation

$$\|\pi(\pi_1(B)) - \pi(B)\| < \|\pi(D) - \pi(D')\| < c_3(5\epsilon_2(1 + c_3)\|B\|) < \epsilon_6\|B\|.$$

We turn to part (i). By Lemma 5.3 we have

$$B' \in T_{\mathcal{M}_2}(0) \text{ and } D' \in T_{\mathcal{M}_1}(0).$$

Thus $\|D'\|/\|B'\| < c(0) < c_1$ whereas (i) amounts to showing that $\|D\|/\|B\| < c_6$. Recall (5.7) and (5.9). Applying Lemma 5.4 with $E = D$, $F = D'$, $\epsilon_4 = 5\epsilon_2(1 + c_3)\|B\|$ and $c_4 = c_2$ we see that either

$$\frac{\|D\|}{\|D'\|} \le c_2 \tag{5.10}$$

or $\|D\| < 5\epsilon_2(1 + c_3)\frac{c_1}{c_1 - 1}\|B\|$, in which case we are done since the constant is less than c_6 by (5.7). We thus assume that (5.10) holds. Note that

$$\frac{\|B'\|}{\|B\|} \le 1 + 5\epsilon_2$$

by (5.8). In this case, we have

$$\frac{\|D\|}{\|B\|} = \frac{\|D\|}{\|D'\|}\frac{\|B'\|}{\|B\|}\frac{\|D'\|}{\|B'\|} < c_2(1 + 5\epsilon_2)c_1 < c_6$$

by (5.7). $\square$

6. Alternating projections. We are finally ready for the main result of this paper.

THEOREM 6.1. *Let $A \in \mathcal{M}_1 \cap \mathcal{M}_2$ be regular and non-tangential, and let $\epsilon_7 > 0$ and $1 > c_7 > c(A)$ be given. Then there exists a $s_7 > 0$ such that the sequence of alternating projections*

$$B_0 = \pi_1(B),\ B_1 = \pi_2(\pi_1(B)),\ B_2 = \pi_1(\pi_2(\pi_1(B))),\ B_3 = \pi_2(\pi_1(\pi_2(\pi_1(B)))), \ldots$$

(i) converges to a point $B_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2$

(ii) $\|B_\infty - \pi(B)\| < \epsilon_7 \|B - \pi(B)\|$

(iii) $\|B_\infty - B_k\| < c_7^k \|B - \pi(B)\|$

Proof. Let c_6 and ϵ_6 in Theorem 5.5 be given by $c_6 = c_7$ and

$$\epsilon_6 = (1 - c_7)\epsilon_7/2. \tag{6.1}$$

Let s_6 be given by Theorem 5.5 and pick

$$s_7 < \frac{s_6(1-\epsilon_7)}{4(2+\epsilon_7)} \tag{6.2}$$

such that $\pi(\mathcal{B}(A, s_7)) \subset \mathcal{B}(A, s_6/4)$, (recall that $\epsilon_7 \leq \epsilon_2 < 1$, by assumption). The latter condition ensures that

$$\|\pi(B) - A\| < s_6/4. \tag{6.3}$$

Let $l = \|B - \pi(B)\|$ and note that

$$l \leq \|B - A\| + \|A - \pi(B)\| \leq s_7 + s_6/4. \tag{6.4}$$

First note that $\|B_0 - B\| \leq l$ and that $\pi(B) = \pi(B_0)$ by Lemma 5.3, so

$$\|B_0 - \pi(B_0)\| \leq \|B_0 - B\| + \|B - \pi(B)\| \leq 2l.$$

Applying Theorem 5.5 we get

$$\|B_{k+1} - \pi(B_{k+1})\| \leq \|B_{k+1} - \pi(\mathcal{B}_k)\| \leq c_7 \|B_{k+1} - \pi(\mathcal{B}_k)\|,$$

as long as

$$B_k \in \mathcal{B}(A, s_6). \tag{6.5}$$

Assuming this for the moment we get

$$\|B_k - \pi(B_k)\| \leq 2lc_7^k \tag{6.6}$$

and

$$\|\pi(B_{k+1}) - \pi(B_k)\| \leq \epsilon_6(2lc_7^k). \tag{6.7}$$

The sequence $(\pi(B_k))_{k=1}^\infty$ is thus a Cauchy sequence, and hence converges to some point B_∞. By (6.6) the same is then true for $(B_k)_{k=1}^\infty$, and the limit point is again B_∞, which thus satisfies $B_\infty = \pi(B_\infty)$ since π, is continuous. By the triangle inequality, (6.1) and (6.7) we have

$$\|\pi(B_k) - \pi(B)\| < \frac{2\epsilon_6 l}{1 - c_7} < \epsilon_7 l,$$

and combining this with (6.3), (6.4), (6.6) we also have

$$\begin{aligned} \|A - B_k\| &\leq \|A - \pi(B)\| + \|\pi(B) - \pi(B_k)\| + \|\pi(B_k) - B_k\| < s_6/4 + \epsilon_7 l + 2l \leq \\ &\leq s_6/4 + \epsilon_7(s_7 + s_6/4) + 2(s_7 + s_6/4) < s_6, \end{aligned}$$

where the last inequality follows by (6.2). With these estimates at hand, it is easy to turn the above argument into a proper induction proof in which (6.5) is verified at each step. We omit the details.
$\square$

7. Case study; rank k matrices versus Hankel matrices II. We now continue the example in Section 2. In order to apply Theorem 6.1 we need to check that an arbitrary point in $\mathcal{H}_k^n$ is non-tangential. We believe that this is true, but can only prove it for all but an exceptional set.

THEOREM 7.1. *The set of non-tangential points in $\mathcal{H}_k^n$ is thin in $\mathcal{H}_k$.* By Proposition 2.2 we immediately get that all points in $\mathcal{H}_k^d$ are regular. To verify that $A \in \mathcal{H}_k^n$ is non-tangential, it thus suffices to establish (4.3). Clearly

$$T_{\mathcal{H}}(A) \cap T_{\mathcal{R}_k^d}(A) \supset T_{\mathcal{H}\cap\mathcal{R}_k^d}(A). \tag{7.1}$$

By Proposition 2.2 and the fact that A is regular we have $\dim(T_{\mathcal{H}\cap\mathcal{R}_k^d}(A)) = 4k$ and

$$\begin{aligned}\dim(T_{\mathcal{H}}(A) \cap T_{\mathcal{R}_k^d}(A)) &= \dim(T_{\mathcal{H}}(A)) + \dim(T_{\mathcal{R}_k^d}(A)) - \dim(T_{\mathcal{H}}(A) + T_{\mathcal{R}_k^d}(A)) = \\ &= 2(2n-1) + 2(2kn - k^2) - \dim(T_{\mathcal{H}}(A) + T_{\mathcal{R}_k^d}(A)).\end{aligned}$$

To establish the reverse inclusion to (7.1), it thus suffices to show that $\dim(T_{\mathcal{H}}(A) \cap T_{\mathcal{R}_k^d}(A)) \leq 4k$, or equivalently

$$\dim(T_{\mathcal{H}}(A) + T_{\mathcal{R}_k^d}(A)) \geq 2(2n-1) + 2(2kn - k^2) - 4k.$$

Moreover, since both subspaces are closed under multiplication by $\mathbb{C}$, it suffices to verify

$$\dim_{\mathbb{C}}(T_{\mathcal{H}}(A) + T_{\mathcal{R}_k^d}(A)) \geq 2n - 1 + 2kn - k^2 - 2k, \tag{7.2}$$

where $\dim_{\mathbb{C}}$ denotes the dimension over $\mathbb{C}$. To this end, note that the map $\mathfrak{W} : (\mathbb{M}_{n,k})^2 \to \mathbb{M}_{n,n}$ given by

$$\mathfrak{W}(U, V) = VU^* = \sum_{j=1}^{k} \mathbf{v}_j \mathbf{u}_j^*,$$

(where $\mathbf{u}_j$, $\mathbf{v}_j$ denote the columns of U and V respectively), is an *immersion* onto $\mathcal{R}_k$. By this we mean that for each $A \in \mathcal{R}_k$ there exists U_A, V_A such that $A = \mathfrak{W}(U_A, V_A)$ and, if $A \in \mathcal{R}_k^d$, then

$$T_{\mathcal{R}_k^d}(A) = \mathsf{Ran}\ \partial\mathfrak{W},$$

where $\partial\mathfrak{W}$ denotes the derivative of $\mathfrak{W}$. In this section we define $\mathfrak{U} : \mathbb{C}^n \to \mathbb{M}_{n,k}$ and $\mathfrak{V} : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{M}_{n,k}$ via

$$\mathfrak{U}(\alpha) = \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_k \\ \alpha_1^2 & \cdots & \alpha_k^2 \\ \vdots & \vdots & \vdots \\ \alpha_1^n & \cdots & \alpha_k^n \end{pmatrix}, \qquad \mathfrak{V}(c, \alpha) = \begin{pmatrix} c_1 & \cdots & c_k \\ c_1\alpha_1 & \cdots & c_k\alpha_k \\ c_1\alpha_1^2 & \cdots & c_k\alpha_k^2 \\ \vdots & \vdots & \vdots \\ c_1\alpha_1^n & \cdots & c_k\alpha_k^n \end{pmatrix}.$$

Recall the definition of $\mathfrak{H}$; (2.5). It is clear that

$$\mathfrak{H}(c, \alpha) = \mathfrak{W}(\mathfrak{U}(\alpha), \mathfrak{V}(c, \alpha)).$$

Thus, whenever $A = \mathfrak{H}(c, \alpha) \in \mathcal{H}_k^d$, we have

$$T_{\mathcal{R}_k^d}(A) = \mathsf{Ran}\ \partial\mathfrak{W}(\mathfrak{U}(\alpha), \mathfrak{V}(c, \alpha)). \tag{7.3}$$

Now, it is not hard to see that $\partial\mathfrak{W}(\mathfrak{U}(\alpha), \mathfrak{V}(c, \alpha))$ is a polynomial in the variables c and α. To visualize, say that $n = 3$ and $k = 2$. Then the right hand side is given as the span of the 12 matrices

$$\begin{pmatrix} 1 & \alpha_j & \alpha_j^2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & \alpha_j & \alpha_j^2 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & \alpha_j & \alpha_j^2 \end{pmatrix}, \qquad (j = 1, 2),$$

and

$$\begin{pmatrix} c_j & 0 & 0 \\ c_j\alpha_j & 0 & 0 \\ c_j\alpha_j^2 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & c_j & 0 \\ 0 & c_j\alpha_j & 0 \\ 0 & c_j\alpha_j^2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & c_j \\ 0 & 0 & c_j\alpha_j \\ 0 & 0 & c_j\alpha_j^2 \end{pmatrix}, \qquad (j = 1, 2).$$

Moreover, picking a basis for $\mathbb{M}_{n,n}$ (for example, the standard one which we order lexicographically), the right hand side of (7.3) can be identified with the range of a matrix with polynomial entries which we denote by $\widetilde{\partial\mathfrak{W}}(\mathfrak{U}(\alpha), \mathfrak{V}(c, \alpha))$. To continue the example, we get

$$\widetilde{\partial\mathfrak{W}}(\cdots) = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & c_1 & 0 & 0 & c_2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & c_1\alpha_1 & 0 & 0 & c_2\alpha_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & c_1\alpha_1^2 & 0 & 0 & c_2\alpha_2^2 & 0 & 0 \\ \alpha_1 & 0 & 0 & \alpha_2 & 0 & 0 & 0 & c_1 & 0 & 0 & c_2 & 0 \\ 0 & \alpha_1 & 0 & 0 & \alpha_2 & 0 & 0 & c_1\alpha_1 & 0 & 0 & c_2\alpha_2 & 0 \\ 0 & 0 & \alpha_1 & 0 & 0 & \alpha_2 & 0 & c_1\alpha_1^2 & 0 & 0 & c_2\alpha_2^2 & 0 \\ \alpha_1^2 & 0 & 0 & \alpha_2^2 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & c_2 \\ 0 & \alpha_1^2 & 0 & 0 & \alpha_2^2 & 0 & 0 & 0 & c_1\alpha_1 & 0 & 0 & c_2\alpha_2 \\ 0 & 0 & \alpha_1^2 & 0 & 0 & \alpha_2^2 & 0 & 0 & c_1\alpha_1^2 & 0 & 0 & c_2\alpha_2^2 \end{pmatrix} \tag{7.4}$$

In the same basis, the Hankel matrices are spanned by $\widetilde{E}_1, \ldots, \widetilde{E}_{2n-1}$, where the E_j's are defined in (2.1) and the notation is self-explanatory. In our example we get

$$\tilde{\mathcal{H}} = \mathsf{Ran} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.5}$$

Let us denote the matrix obtained by adjoining $\widetilde{\partial\mathfrak{W}}(\cdots)$ and $\widetilde{\mathcal{H}}$ by $[\widetilde{\partial\mathfrak{W}\ \mathcal{H}}]$. To verify (7.2), it thus suffices to show that

$$\mathsf{Rank}\ [\widetilde{\partial\mathfrak{W}\ \mathcal{H}}] \geq 2n - 1 + 2kn - k^2 - 2k. \tag{7.6}$$

Note that

1. If we can establish (7.6) for one point $A = \mathfrak{H}(c, \alpha)$, then it easily follows that (7.6) holds at all but a thin set of points A. To see this, set $q = 2n - 1 + 2kn - k^2 - 2k$ and first note that if this is the case, then we can pick a $q \times q$ submatrix of $[\widetilde{\partial\mathfrak{W}\ \mathcal{H}}]$ whose determinant is a non-zero polynomial. Thus, by standard algebraic geometry, the set of points (c, α) where the determinant is zero is thin in $\mathbb{C}^{2k}$. Finally, it is also clear that the image of a thin set under a chart, in this case $\mathfrak{H}$, is again thin.
2. Let $B = \mathfrak{W}(U_B, V_B)$ be a point such that $B \in \mathcal{H}_k$ and

$$\mathsf{Rank}\ \partial\mathfrak{W}(U_B, V_B) = 2kn - k^2, \tag{7.7}$$

but where (U_B, V_B) is not in the closure of the range of $(\mathfrak{U}, \mathfrak{V})$. We claim that in order to establish 1., it suffices to establish (7.6) at the point (U_B, V_B). To see this, first note that by (7.7), $\mathcal{R}_k$ is locally a manifold (of dimension $2kn - k^2$) around B, and we can take an affine subspace of $\mathcal{N} \subset \mathbb{M}^2_{n,k}$ containing (U_B, V_B) such that $\mathfrak{W}|_{\mathcal{N}}$ becomes a local chart for $\mathcal{R}_k$. If (7.6) holds for (U_B, V_B), then arguing as above with determinants, it holds in a neighborhood of (U_B, V_B). By Proposition 2.2, $\mathcal{H}_k^n$ is dense in $\mathcal{H}_k$, so in particular we can pick a $C \in \mathcal{H}_k^n$ and corresponding $U_C,\ V_C,\ c_C, \alpha_C$ such that $C = \mathfrak{W}(U_C, V_C) = \mathfrak{H}(c_C, \alpha_C)$ and (7.6) is satisfied for $[\widetilde{\partial\mathfrak{W}(U_C, V_C)\ \mathcal{H}}]$. By (7.3) and (7.7) we then have

$$[\partial\mathfrak{W}(U_C, V_C)\ \tilde{\mathcal{H}}] = T_{\mathcal{R}_k^d}(C) + T_{\mathcal{H}}(C) = [\partial\mathfrak{W}(\mathfrak{U}(\alpha_C), \mathfrak{V}(c_C, \alpha_C))\ \tilde{\mathcal{H}}]$$

which does the job.

So, it remains to verify (7.6) for (U_B, V_B) as in point 2. above. In terms of our example, we pick

$$U_B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad V_B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$$

so that B becomes the rank 2 Hankel operator

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $\partial\mathfrak{W}(U_B, V_B)$ is spanned by the 6 "V-derivatives";

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and the 6 "U-derivatives;

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

This is clearly an 8-dimensional space not including

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which happens to be E_5 in the basis for $T_{\mathcal{H}}$ (recall (2.1)), and thus

$$\mathsf{Rank}\ [\widetilde{\partial\mathfrak{W}(U_B, V_B)\ \mathcal{H}}] = 9 = 2*3 - 1 + 2*2*3 - 2^2 - 2*2, \tag{7.8}$$

establishing (7.6) in this particular case. It is easy to generalize this example to arbitrary k, n. The "V-derivatives" will span the first k columns of $\mathbb{M}_{n,n}$, whereas the U-derivatives will span the first k rows. Thus $\mathsf{Rank}\ \partial\mathfrak{W}(U_B, V_B) = 2kn - k^2$, as required in (7.7). Moreover, it is easy to see that $\{E_1, \ldots, E_{2k}\}$ is a subset of $\mathsf{Ran}\ \partial\mathfrak{W}(U_B, V_B)$, whereas $\{E_{2k+1}, \ldots, E_{2n-1}\}$ form a basis for a disjoint subspace, except for the point zero. In general we thus get

$$\mathsf{Rank}\ [\partial\mathfrak{W}(U_B, V_B),\ \ T_{\mathcal{H}}] = 2kn - k^2 + 2n - 1 - 2k,$$

as desired.

Acknowledgements. This work was supported in part by the Swedish Research Council Grant 2008-23883-61232-34

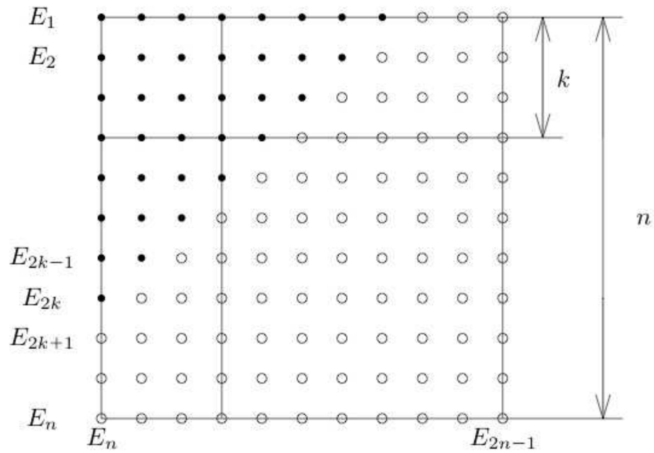


FIG. 8. *Rank illustration. The filled dots represents the Hankel basis elements which are included in* $\mathsf{Ran}\ \partial\mathfrak{W}(U_B, V_B)$.

REFERENCES

[1] ANDERSSON, F., CARLSSON, M., *A fast alternating projection method for complex frequency estimation.* GMIG-report, this issue.

[2] BERGER, M., GOSTIAUX, B., *Differential Geometry: Manifolds, Curves and Surfaces.* Springer Verlag, 1988.

[3] CADZOW, J., *Signal Enhancement-A Composite Property Mapping Algorithm.* IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 36, no. 1, January 1988.

[4] HORN, R. A., JOHNSSON, C., R,. *Matrix Analysis.* Cambridge University Press, 1985.

[5] LEWIS, A. S., MALICK, J., *Alternating projections on manifolds.* Math. Oper. Res. 33 (2008), no. 1, 216234.

[6] ZANGWILL, W. *Nonlinear Programming, A Unified Approach.* Englewood-Cliffs, NJ, Prentice-Hall, 1969.