ALTERNATING PROJECTIONS ON NON-TANGENTIAL MANIFOLDS.

FREDRIK ANDERSSON AND MARCUS CARLSSON

Abstract. We consider sequences \((B_k)_{k=0}^\infty\) of points obtained by projecting a given point \(B = B_0\) back and forth between two manifolds \(M_1\) and \(M_2\), and give conditions guaranteeing that the sequence converges to a limit \(B_\infty \in M_1 \cap M_2\). Our motivation is the study of algorithms based on finding the limit of such sequences, which have proven useful in a number of areas. The intersection is typically a set with desirable properties, but for which there is no efficient method for finding the closest point \(B_{opt}\) in \(M_1 \cap M_2\). Under appropriate conditions, 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.

1. Introduction

Let \(K\) be a finite dimensional Hilbert space over \(\mathbb{R}\) and let \(M\) be a manifold. Given two manifolds \(M_1, M_2 \subset K\) satisfying \(M = M_1 \cap M_2\) along with corresponding “projection operators”\(^1\) denoted by \(\pi, \pi_1\) and \(\pi_2\). Given a point \(B \in K\), suppose that \(\pi(B)\) is sought for and that \(\pi_1\) and \(\pi_2\) can be efficiently computed, whereas \(\pi\) cannot. A classical result by von Neumann [52] says that if \(M_1\) and \(M_2\) are affine linear manifolds, then the sequence of alternating projections

\begin{equation}
B_0 = B, \quad B_{k+1} = \begin{cases}
\pi_1(B_k), & \text{if } k \text{ is even}, \\
\pi_2(B_k), & \text{if } k \text{ is odd},
\end{cases}
\end{equation}

converges to \(\pi(B)\). Moreover, the convergence rate is linear, and determined by the angle between \(M_1\) and \(M_2\). This paper is concerned with extensions of this result to non-linear manifolds. Suppose for the moment that the limit point of a particular sequence \((B_k)_{k=0}^\infty\) exists and denote it by \(B_\infty\). In contrast to the case where each \(M_j\) (\(j = 1, 2\)) is an affine linear manifold, \(B_\infty\) is usually different from \(\pi(B)\). However, given that \(M_j\) behave nicely, we may expect \(B_\infty \approx \pi(B)\), which is one of the central results of the paper.

We begin with a brief review of related works. For a brief history of the early developments of von Neumann’s algorithm we refer to [29]. Alternating projection schemes for non-linear subsets have been used in a number of applications; cf.

\(^1\)If the manifold \(M\) is not convex, then there exist points with multiple closest points on the manifold. To define the projection onto \(M\), one thus needs to involve point to set maps. We will not use this formalism, but rather write \(\pi(B)\) to denote an arbitrarily chosen closest point to \(B\) on \(M\). This is done to simplify the presentation and because our results are stated in a local environment where the \(\pi's\) are shown to be well defined functions. See Proposition 2.3.
For instance, $\mathcal{K}$ can be the set $\mathbb{M}_{m,n}$ of $m \times n$-matrices and the sets $\mathcal{M}_j$ be subsets with a certain structure, e.g. matrices with a certain rank, self-adjoint matrices, Hankel or Toeplitz matrices etc. For a detailed overview of optimization methods on matrix manifolds, cf. [1]. Alternating projection schemes between several linear subsets was investigated in [3, 29]. Other applications concern e.g. the EM algorithm, see [8]. Much emphasis has been put towards the use of alternating projections for the case of convex sets $\mathcal{M}_1$ and $\mathcal{M}_2$, see for instance [5, 7, 12, 26]. It is worth pointing out that except for linear subspaces, there is no overlap between this theory and the one considered here. Dykstra’s alternating projection algorithm [6, 19] utilizes an additional dual variable and a slightly different projection scheme that is more efficient than the classical von Neumann method in the convex case. For connections between Dykstra’s alternating projection method and the alternating direction method of multipliers (ADMM), see [10].

However, for non-convex sets the field remained rather undeveloped until the 90’s. For example, in [13], Zangwill’s Global Convergence Theorem [55] is used to motivate the convergence of an alternating projection scheme. Zangwill’s theorem implies that if the sequence $(B_k)_{k=1}^\infty$ is bounded and the distance to $\mathcal{M}_1 \cap \mathcal{M}_2$ is strictly decreasing, then $(B_k)_{k=1}^\infty$ has a convergent subsequence to a point $B_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2$. This result is an easy consequence of the fact that any sequence in a compact set has a convergent subsequence. Thus, the use of Zangwill’s theorem in this context does not provide any information about whether the limit point of the entire sequence exists, or if so, whether it is close to the optimal point $\pi(B)$.

To our knowledge, the first rigorous attempt at dealing with the method of alternating projections for non-convex sets was made in [14]. However, the setting is rather general and the results are mainly concerned with convergence of the sequence. In particular, when this does happen, the results in [14] does not reveal anything about the size of the error $\|\pi(B) - B_\infty\|$.

Recently, A. Lewis and J. Malick presented stronger results [33], although valid under somewhat more restrictive conditions on $\mathcal{M}_1$ and $\mathcal{M}_2$. Before discussing their results in more detail, we give two simple examples that illustrate some of the difficulties that may arise when using alternating projections on non-linear manifolds.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{An example demonstrating that the algorithm can get stuck in loops where not even a subsequence converges to a point in the intersection (the black dots). However, starting closer to the intersection point, we do get convergence (the white dots).}
\end{figure}
Example 1.1. Let $K = \mathbb{R}^2$ and set $M_1 = \{(t, (t+1)(3-t)/4) : t \in \mathbb{R}\}$ and $M_2 = \mathbb{R} \times \{0\}$. It is easily seen that $\pi_1((1,0)) = (1,1)$ and $\pi_2((1,1)) = (1,0)$, and hence the sequence of alternating projections does not converge, cf. Figure 1. On the other hand, if $(1+\epsilon,0) \in M_2$, $\epsilon > 0$, is used as a starting point, the sequence of alternating projections will converge to $(3,0) \in M_1 \cap M_2$.

In general, it seems reasonable to assume that if the starting point is sufficiently close to an intersection point, then the sequence does converge to a point in the intersection. The next example shows that this is not always the case.

Figure 2. Alternating projections stuck in a loop.

Example 1.2. Without going into the details of the construction, we note that one can construct a $C^\infty$-function $f$ with $f(0) = 0$ and horizontal segments arbitrarily near 0. Figure 2 explains the idea. With $M_1 = \mathbb{R} \times \{0\}$ and $M_2 = \{(t,f(t)) : t \in \mathbb{R}\}$, the sequence of alternating projections can then get stuck in projecting back and forth between the same two points, even if the starting point is arbitrarily close to the intersection point $(0,0)$.

However, if we have $f'(0) \neq 0$ it is hard to imagine how to make a similar construction work. This is indeed impossible, which follows both from the present paper and [35]. In the terminology of the latter, the condition $f'(0) \neq 0$ implies that $M_1$ and $M_2$ are transversal at $(0,0)$. In general, given $C^1$-manifolds $M_1$, $M_2$ and a point $A \in M_1 \cap M_2$, we say that $A$ is transversal if

$$T_{M_1}(A) + T_{M_2}(A) = K,$$

where $T_{M_j}(A)$ denotes the tangent-space of $M_j$ at $A$, $j = 1, 2$. The main result in [35] is roughly the following:

Theorem 1.3. Let $M_1$ and $M_2$ be $C^3$-manifolds and let $A \in M_1 \cap M_2$ be transversal. If $B$ is close enough to $A$, then the sequence of alternating projections $(B_k)_{k=0}^\infty$ given by (1.1) converges at a linear rate to a point $B_\infty$ in $M_1 \cap M_2$.

Figure 3. Lack of transversality due to tangential curves.
The error $\|B_\infty - \pi(B)\|$ is not discussed in [35], but inspection of the proofs yield

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

The improvement over the previously mentioned results is thus that the entire sequence converges and that the limit $B_\infty$ is not too far away from $\pi(B)$, although in relative terms, i.e. comparing with $\text{dist}(B, M_1 \cap M_2) = \|B - \pi(B)\|$, it need not be particularly close either. Moreover, the assumption of transversality is rather restrictive. To demonstrate the essence of the transversality assumption we now present some examples.

**Example 1.4.** The manifolds in Example 1.2 are not transversal at $A = (0, 0)$, since $T_{M_1}(A) = T_{M_2}(A) = M_1$. With $M_1 = \mathbb{R} \times \{0\}$ and $M_2 = \{(t, t^2) : t \in \mathbb{R}\}$, we obtain another example of manifolds that are not transversal at $(0, 0)$, see Figure 3. In this case, the sequence of alternating projections do converge to $(0, 0)$, independent of the starting point, albeit extremely slowly.

**Example 1.5.** With $K = \mathbb{R}^3$ and $M_1 = \mathbb{R} \times \{0\}$ and $M_2 = \{(t, t, t^2) : t \in \mathbb{R}\}$, 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 are of too low dimension to satisfy the transversality assumption. In fact, if $K$ has dimension $n$ and $M_j$ has dimension $m_j$, $j = 1, 2$, the transversality (1.2) can never be satisfied if $m_1 + m_2 < n$, i.e. when the sum of the dimensions of the manifolds is less than that of $K$. This is, however, the case in many applications of practical interest. The papers [34] and [2] both consider alternating projections in a more general setting, but when dealing with manifolds, the assumptions also imply $m_1 + m_2 \geq n$ (see Theorem 5.16 and condition (5.17) in [34] and Section 4.5 in [2]).

We now present the main results of the present paper. We introduce the concept of a non-tangential intersection point, which impose no restriction on the dimensions of $M_1$ and $M_2$. Loosely speaking, a point $A$ in $M_1 \cap M_2$ is non-tangential if the two manifolds have a positive angle in directions perpendicular to $M_1 \cap M_2$. 

![Figure 4. Lack of transversality due to low dimensions.](image-url)
We will show that this happens if and only if
\[ T_{M_1}(A) \cap T_{M_2}(A) = T_{M_1 \cap M_2}(A), \]
which we can use as definition for now. In microlocal analysis this is called clean intersection [28]. We remark that non-tangential points are always transversal, since these do satisfy the above identity by Definition 3 in [35]. The point (0, 3) in Example 1.1 and the origin in Example 1.5 are both non-tangential, whereas the origins in Examples 1.2 and 1.4 are not. We now present a simplified version of the main result (Theorem 5.1). Figure 5 illustrates the idea.

**Theorem 1.6.** Let \( M_1, M_2 \) and \( M_1 \cap M_2 \) be \( C^2 \)-manifolds. Given a non-tangential point \( A \in M_1 \cap M_2 \) there exists an \( s > 0 \) such that the sequence of alternating projections (1.1) converges to a point \( B_\infty \in M_1 \cap M_2 \), given that \( \|B - A\| < s \). Moreover, given any \( \epsilon > 0 \) one can take \( s \) such that
\[
\|B_\infty - \pi(B)\| < \epsilon \|B - \pi(B)\|. 
\]

![Figure 5. Picture illustrating Theorem 1.6.](image)

\( B_\infty \) may differ from \( \pi(B) \), but the error is small compared to the distance \( d \) between \( B \) and the manifold \( M_1 \cap M_2 \), given that \( B \) is in the ball \( B_K(A, s) \).

The improvement over Theorem 1.3 mainly consists of two items. Primarily, the assumption that the surfaces are non-tangential is not at all restrictive, and in particular there is no implication about the dimensions of \( M_1 \) and \( M_2 \). For the applications we are aware of, the set of tangential points is very small, if it exists at all. Secondly, as has been highlighted before, we are usually interested not just in any point of \( M_1 \cap M_2 \), but the closest point \( \pi(B) \). Here the theorem says that in relative terms, i.e. after dividing with the distance to \( M_1 \cap M_2 \), the error is small and moreover improves if the distance from \( B \) to \( M_1 \cap M_2 \) decreases.

We now discuss the rate of convergence. When \( M_1 \) and \( M_2 \) are linear, we have
\[
\|B_k - B_\infty\| \leq \text{const} \cdot c^k
\]
where \( c = \cos \alpha \) and \( \alpha \) is the angle between the subspaces \( M_1 \) and \( M_2 \), which was shown in [26]. This is usually referred to as linear convergence. In the present setting, we show that (1.3) holds if \( B \) is sufficiently near a point whose angle \( \alpha \) satisfies \( c > \cos(\alpha) \). However, the definition of the angle between two manifolds needs some preparation, and is discussed in Section 3. For the precise results
see Definition 3.1 and Theorem 5.1. The corresponding statement for transversal manifolds has been proven in [35], and similar results can also be found in [5, 16, 18, 22, 29, 54].

In many applications, one has real algebraic varieties as opposed to manifolds, (of which the usual complex algebraic varieties are special cases). However, algebraic varieties are manifolds except at the singular set (singular locus). Moreover, the singular set is a variety of smaller dimension, and hence makes up a very small part of the original variety. Since Theorem 1.6 is a local statement, it is plausible that it applies at non-singular intersection points of two given varieties. In Section 6, we prove that this is indeed the case under mild assumptions. We also provide a number of concrete results for verifying these assumptions. Finally, the paper ends with a simple application where the alternating projection method is used for the problem of finding correlation matrices with prescribed rank from measurements with missing data.

The paper is organized as follows. Section 2 presents rudimentary results concerning manifolds, and Section 3 gives definitions and basic theory for determining angles between them. The technical part of the paper begins in Section 4, where the various projection operators are studied. With this at hand, the proof of the main theorem (given in Section 5) is fairly straightforward. Finally, Section 6 deals with real algebraic varieties and Section 7 presents the example. There are three appendices with proofs of the results in Sections 2, 6 and 7 respectively.

2. Manifolds

We provide a review of necessary concepts from differential geometry. Let $K$ be a Euclidean space, i.e. a Hilbert space of finite dimension $n \in \mathbb{N}$. Given $A \in K$ and $r > 0$ we write $B(A, r)$ or $B_K(A, r)$ for the open ball centered at $A$ with radius $r$. Since $K$ is finite-dimensional it has a unique Euclidean topology. Any subset $M$ of $K$ will be given the induced topology from $K$. Let $p \geq 1$ and let $M \subset K$ be an $m$-dimensional $C^p$-manifold. We recall that around each $A \in M$ there exists an injective $C^p$-immersion $\phi$ on an open set $U$ in $\mathbb{R}^m$ such that

\begin{equation}
M \cap B_K(A, s) = \text{Im } \phi \cap B_K(A, s),
\end{equation}

for some $s > 0$, where $\text{Im } \phi$ denotes the image of $\phi$. See e.g. Theorem 2.1.2 [9]. If $A = \phi(x_A)$, we define the tangent space $T_M(A)$ by $T_M(A) = \text{Ran } d\phi(x_A)$, where $d\phi$ denotes the Jacobian matrix. It is a standard fact from differential geometry that this definition is independent of $\phi$, (see e.g. Section 2.5 of [9]). Moreover, we set

$$\tilde{T}_M(A) = A + T_M(A),$$

i.e., $\tilde{T}_M(A)$ is the affine linear manifold which is tangent to $M$ at $A$. We now list some necessary properties of manifolds. Some statements are well-known, but for the convenience of the reader, we outline all proofs in Appendix A.

**Proposition 2.1.** Let $\alpha : K \to M$ be any $C^p$-map, where $M$ is a $C^p$-manifold and $p \geq 1$. With $\phi$ as above, the map $\phi^{-1} \circ \alpha$ is also $C^p$ (on its natural domain of definition).

\footnote{i.e., a $p$ times continuously differentiable function such that $d\phi(x)$ is injective for all $x \in B_{\mathbb{R}^m}(0, r)$}
We write $A \perp B$ to denote that $A, B \in \mathcal{K}$ are orthogonal. If $X, Y \subset \mathcal{K}$ are subsets such that $(A, B) = 0$ for all $A \in X$ and $B \in Y$, we also write $X \perp Y$. If $Y$ is the orthogonal complement of $X$, i.e. the maximal subspace with the above property, we shall denote it by $Y = \mathcal{K} \ominus X$. When $X$ is a subspace, we use $P_X : \mathcal{K} \to \mathcal{K}$ for the orthogonal projection onto $X$, i.e. if $A \in X$ and $B \in \mathcal{K} \ominus X$ we set $P_X(A + B) = A$.

**Proposition 2.2.** Let $\mathcal{M}$ be a $C^1$-manifold. Then $P_{T_{\mathcal{M}}(A)}$ is a continuous function of $A$.

The next proposition shows that projection on $\mathcal{M}$ is a locally well defined operation. Thus, as long as we start the alternating projections method sufficiently near the manifolds in question, there is no need to deal with point to set maps.

**Proposition 2.3.** Let $\mathcal{M}$ be a $C^2$-manifold and let $A \in \mathcal{M}$ be given. Then there exists an $s > 0$ such that for all $B \in \mathcal{B}_K(A, s)$ there exists a unique closest point in $\mathcal{M}$. Denoting this point by $\pi(B)$, the map $\pi : \mathcal{B}_K(A, s) \to \mathcal{M}$ is $C^2$. Moreover, $C \in \mathcal{M} \cap \mathcal{B}_K(A, s)$ equals $\pi(B)$ if and only if $B - C \perp T_{\mathcal{M}}(C)$.

When there is risk of confusion we will write $\pi_{\mathcal{M}}$ instead of $\pi$. We remark that it is easy to show that $\pi$ is $C^1$. The fact that it actually is $C^2$ is noted in [23] and further developed in [31]. It is however not true in general that $\pi$ is $C^1$ if $\mathcal{M}$ is a $C^1$-manifold, in fact, it does not even have to be single valued. The last line in the above proposition is an extension of Fermat’s principle, for similar statements in a more general environment we refer to [44]. We end with a proposition that basically says that the affine tangent-spaces are close to $\mathcal{M}$ locally.

**Proposition 2.4.** Let $\mathcal{M}$ be a $C^2$-manifold and $A \in \mathcal{M}$ be given. For each $\epsilon > 0$ there exists $s > 0$ such that for all $C \in \mathcal{B}_K(A, s) \cap \mathcal{M}$ we have

1. $\text{dist}(D, \hat{T}_\mathcal{M}(C)) < \epsilon \|D - C\|$, $\forall D \in \mathcal{B}(A, s) \cap \mathcal{M}$.
2. $\text{dist}(D, \mathcal{M}) < \epsilon \|D - C\|$, $\forall D \in \mathcal{B}(A, s) \cap \hat{T}_\mathcal{M}(C)$.

3. **Non-tangentiality**

Suppose now that we are given $C^1$-manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ such that their intersection $\mathcal{M}_1 \cap \mathcal{M}_2$ is itself a $C^1$-manifold. We will denote $\mathcal{M}_1 \cap \mathcal{M}_2$ by $\mathcal{M}$, and the objects from Section 2 associated to $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}$, e.g. the dimension, by $m_1$, $m_2$ and $m$ respectively. We thus omit subindex when dealing with $\mathcal{M}_1 \cap \mathcal{M}_2$. We now introduce the angle between $\mathcal{M}_1$ and $\mathcal{M}_2$ at an intersection point $A \in \mathcal{M}$. This issue is a bit delicate. We first discuss the case when $\mathcal{M}_1$ and $\mathcal{M}_2$ are linear subspaces. A good reference for various definitions of angles in this setting is [17].

We will here use the so called Friedrich’s angle, which is common in functional analysis [3, 42] and also the one appearing in (1.3). The Friedrich’s angle $\alpha$ between two subspaces is then given by

$$\alpha = \cos^{-1} (\|P_{\mathcal{M}_1}P_{\mathcal{M}_2} - P_{\mathcal{M}_1 \cap \mathcal{M}_2}\|).$$

(3.1)

To better understand this definition, set

$$\mathcal{F}_j = \{B_j \in \mathcal{M}_j \ominus (\mathcal{M}_1 \cap \mathcal{M}_2)\} \setminus \{0\}$$

and note that

$$\alpha = \cos^{-1} \left( \sup_{B_j \in \mathcal{F}_j} \frac{\langle B_1, B_2 \rangle}{\|B_1\| \|B_2\|} \right),$$

(3.2)
(assuming $F_1 \neq \emptyset \neq F_2$, in which case it becomes pointless to talk of angles). Hence, $\alpha$ is the minimal angle in directions perpendicular to the common intersection $M_1 \cap M_2$.

To introduce angles between manifolds one obviously needs to let the angle be dependent on the point of intersection $A \in M_1 \cap M_2$. Based on (3.1), one idea is to let the angle at $A$ be the angle of $T_{M_1}(A)$ and $T_{M_2}(A)$, i.e.

$$\alpha(A) = \cos^{-1}\left(\|P_{T_{M_1}(A)}P_{T_{M_2}(A)} - P_{T_{M_1}(A) \cap T_{M_2}(A)}\|\right).$$

This is indeed the definition adapted in [35], see Section 3.2. However, consider again the two manifolds in Example 1.4 and Figure 3. The expression (3.3) then leads to the counterintuitive conclusion that $\alpha((0,0)) = \cos^{-1}(0) = \pi/2$. In fact, it is easy to see from (3.2) that the angle between two subspaces never can be 0, whereas we argue that the angle should be 0 for the manifolds in Figure 3. For this reason, we adopt the following definition, generalizing (3.2).

**Definition 3.1.** Given $A \in M_1 \cap M_2$, set

$$F_j^r = \{B_j \in M_j \setminus A, \|B_j - A\| < r \text{ and } B_j - A \perp T_{M_j \cap M_2}(A)\}. $$

If $F_j^r \neq \emptyset$ for all $r > 0$ and $j = 1, 2$, we define the angle $\alpha(A)$ of $M_1$ and $M_2$ at $A$ as

$$\alpha(A) = \cos^{-1}(\sigma(A))$$

where

$$\sigma(A) = \lim_{r \to 0} \sup_{B_j \in F_j^r} \left\{ \frac{\langle B_1 - A, B_2 - A \rangle}{\|B_1 - A\| \|B_2 - A\|} \right\}. $$

Otherwise, we let both $\sigma(A)$ and $\alpha(A)$ be undefined.

This means that that the angle, if it is defined, is minimized locally in directions perpendicular to $T_{M_1 \cap M_2}(A)$, (as opposed to $T_{M_1}(A) \cap T_{M_2}(A)$). With Definition 3.1, angles can be 0, and it is readily verified that the angle is zero in Example 1.4. The next proposition shows that the angle is undefined only at points where it makes no sense to talk of angles.

**Proposition 3.2.** The angle of $M_1$ and $M_2$ at $A$ is undefined if and only if one of the manifolds is a subset of the other in a neighborhood of $A$, that is, if there exists an $s > 0$ such that $M_1 \cap \mathcal{B}_k(A, s) \subset M_2 \cap \mathcal{B}_k(A, s)$ or $M_2 \cap \mathcal{B}_k(A, s) \subset M_1 \cap \mathcal{B}_k(A, s)$.

**Proof.** There is clearly no restriction to assume that $A = 0$. Recall that a $C^1$-manifold $M$ containing 0 locally can be written as the graph of a $C^1$-function defined in a neighborhood of 0 in $T_M(0)$ with values in $(T_M(0))^\perp$, (this is a simple improvement of Theorem 2.1.2 (iv) in [9]). Now, either $T_{M_1 \cap M_2}(A) = T_{M_1}(A)$ or not. The fact just mentioned implies that in the first case $M_1 \cap M_2$ and $M_1$ coincide near 0, and in the second case $F_1^r \neq \emptyset$ for all $r > 0$. The same dichotomy obviously holds with the roles of 1 and 2 switched, and the proposition follows. \hfill $\square$

**Definition 3.3.** Points $A \in M_1 \cap M_2$ where the angle is defined will be called non-trivial intersection points. For such points, we say that $A$ is tangential if $\alpha(A) = 0$ and non-tangential if $\alpha(A) > 0$.

**Proposition 3.4.** $\sigma$ is continuous on $\text{supp } \sigma$ and

$$\sigma(A) = \|P_{T_{M_1}(A)}P_{T_{M_2}(A)} - P_{T_{M_1 \cap M_2}(A)}\|.$$
Proof. Let $A \in M_1 \cap M_2$ be non-trivial. It is easy to see that

$$\sigma(A) = \sup \left\{ \frac{\langle B_1 - A, B_2 - A \rangle}{\|B_1 - A\| \|B_2 - A\|} : B_j \in T_{M_1}(A) \cap T_{M_2}(A) \right\} =$$

$$= \|P_{T_{M_1}(A)} \cap T_{M_2}(A) P_{T_{M_2}(A)} \cap T_{M_1}(A)\| =$$

$$= \|P_{T_{M_1}(A)} P_{T_{M_2}(A)} - P_{T_{M_1 \cap M_2}(A)}\|.$$ (3.4)

The continuity of $\sigma$ now follows by Proposition 2.2. \hfill $\Box$

In particular, we get

$$\alpha(A) = \cos^{-1} \left( \|P_{T_{M_1}(A)} P_{T_{M_2}(A)} - P_{T_{M_1 \cap M_2}(A)}\| \right)$$

which should be compared with (3.3). Moreover, we infer that non-tangentiality is a stable property, i.e. if $A$ is non-tangential, then the same holds for all $B \in M_1 \cap M_2$ in a neighborhood of $A$. Our experience is that in practice, tangential and trivial intersection points are exceptional. For example, we will show in Section 6 that if $M_1$ and $M_2$ are also algebraic varieties, then non-tangentiality at one point in $M_1 \cap M_2$ implies non-tangentiality of the vast majority of points. The next proposition gives a useful criterion for checking non-tangentiality.

**Proposition 3.5.** Let $A \in M_1 \cap M_2$ be a non-trivial intersection point. Then $A$ is non-tangential if and only if

$$T_{M_1}(A) \cap T_{M_2}(A) = T_{M_1 \cap M_2}(A).$$ (3.6)

*Proof.* By (3.4) it is easy to see that $A$ is non-tangential if and only if

$$\left( T_{M_1}(A) \cap T_{M_1 \cap M_2}(A) \right) \cap \left( T_{M_2}(A) \cap T_{M_1 \cap M_2}(A) \right) = \{0\}$$

which in turn happens if and only if (3.6) holds. \hfill $\Box$

In particular, for non-trivial intersection points a simple criterion which implies non-tangentiality is

$$\dim(T_{M_1}(A) \cap T_{M_2}(A)) \leq m,$$

(recall that $m = \dim(M_1 \cap M_2)$), which is immediate by (3.6) and the inclusion $T_{M_1}(A) \cap T_{M_2}(A) \supset T_{M_1 \cap M_2}(A)$. (3.6) is usually referred to as *clean intersection* in microlocal analysis, (see Appendix C.3, [28]). We have chosen the terminology non-tangential since it is more intuitive. Note that by (3.5) and (3.6), the angle as defined in Definition 3.1 and that of (3.3) are equivalent whenever $A$ is non-tangential. Finally, Proposition 3.5 shows that non-tangentiality is a weaker concept than transversality, defined in (1.2). For if two manifolds are transversal at $A$, then (3.6) holds by Definition 3 in [35].

4. **Properties of the projection operators**

Let $M_1$, $M_2$ and $M = M_1 \cap M_2$ be $C^2$-manifolds and $A \in M_1 \cap M_2$ a non-tangential intersection point. This will be fixed throughout the section. We continue to use the convention of denoting objects related to $M_1 \cap M_2$ without subindex. As in the introduction we shall abbreviate $\pi_{M_1}$, $\pi_{M_2}$ and $\pi_M$ by $\pi_1$, $\pi_2$ and $\pi$. The following result will be the main tool for proving convergence of the alternating projections.
**Theorem 4.1.** For each $\varepsilon > 0$ there exists an $s > 0$ such that for all $B \in \mathcal{B}(A, s)$ we have
\[
\|\pi(\pi_j(B)) - \pi(B)\| < \varepsilon \|B - \pi(B)\|, \quad j = 1, 2.
\]

The proof is rather technical and depends on a long series of lemmas. The numbers $s$ and the map $\phi$ appearing in (2.1) clearly differ between $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}$. We denote the respective maps by $\phi_1$, $\phi_2$ and $\phi$, and let $s_{-1}$ be a number such that (2.1) holds in all three cases with $s = s_{-1}$. We will make repeated use of Propositions 2.3 and 2.4, applied to each of the manifolds $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_1 \cap \mathcal{M}_2$. The numbers $s$ appearing in these also depend on which manifold that is considered, and also on the auxiliary constant $\varepsilon$. We will choose a fixed value for $\varepsilon$ in the proof of Theorem 4.1. In the meantime, it will be treated as a fixed number. We let $s_0$ be a constant small enough that Proposition 2.3 applies to each of the manifolds $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_1 \cap \mathcal{M}_2$ in $\mathcal{B}(A, s_0)$. Since $\pi_1$, $\pi_2$ and $\pi$ then are continuous in $\mathcal{B}(A, s_0)$ and have $A$ as a fixed point, we may also assume that $s_0$ is small enough that their respective images are contained in $\mathcal{B}(A, s_{-1})$. Finally, we let $s_{\varepsilon} \leq s_0$ be such that also Proposition 2.4 applies for each of the 3 manifolds in $\mathcal{B}(A, s_{\varepsilon})$.

**Figure 6.** Illustration of the difference between $\rho_1$ and $\pi_1$. $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{F}$ appear in the proof of Proposition 4.3.

We introduce maps $\rho_j : \mathcal{B}(A, s_0) \rightarrow \mathcal{K}$, ($j = 1$ or $j = 2$), via
\[
\rho_j(B) = \pi_{\mathcal{M}_j}(\pi(B))(B).
\]
Thus, $\rho_j$ resemble $\pi_j$ but is slightly different. $\pi_j$ projects onto $\mathcal{M}_j$ whereas $\rho_j$ projects onto the tangent plane of $\mathcal{M}_j$ taken at the closest point to $B$ in $\mathcal{M}_1 \cap \mathcal{M}_2$, i.e. $\pi(B)$, (see Fig 6). An estimate of the difference between $\rho_j$ and $\pi_j$ is given in Proposition 4.3.

**Lemma 4.2.** The functions $\rho_1$ and $\rho_2$ are $C^1$-maps in $\mathcal{B}(A, s_0)$. Moreover, we can select a number $s_{\varepsilon}^1 < s_{\varepsilon}^0$ such that the image of $\mathcal{B}(A, s_{\varepsilon}^1)$ under $\rho_1$, $\rho_2$, $\pi$, $\pi_1$, $\pi_2$, as well as any composition of two of those maps, is contained in $\mathcal{B}(A, s_{\varepsilon}^0)$.
Proof. The second part is an immediate consequence of the continuity of the maps, and \( \pi, \pi_1 \) and \( \pi_2 \) are continuous by Proposition 2.3. Let \( j \) denote 1 or 2, and set \( M_j(B) = d\phi_j(\phi_j^{-1}(\pi(B))) \), which is well defined in \( B(A, s_0) \) by the choice of \( s_{-1} \) and \( s_0 \). By Proposition 2.1 and 2.3, we have that \( \phi_j^{-1}(\pi(B)) \) is a \( C^2 \)-map in \( B(A, s_0) \), and hence \( M_j \) is \( C^1 \) as \( \phi_j \) is \( C^2 \). Given any point \( B_0 \in M_j \) it is easy to see that

\[
\pi_{T_M_j}(B_0)(B) = B_0 + P_{TM_j}(B_0)(B - B_0)
\]

and hence, by the formula already used in the proof of Proposition 2.2, we have

\[
\rho_j(B) = \pi_{T_M_j}(\pi(B))(B) = \pi(B) + P_{TM_j}(\pi(B))(B - \pi(B)) = \\
= \pi(B) + M_j(B)(M_j^*(B)M_j(B))^{-1}M_j^*(B)(B - \pi(B))
\]

from which the result follows. \( \square \)

\[
\text{Figure 7. Proof illustration. } r = \|E_B\| + \epsilon\|D_B\|, r_1 = 2\epsilon(\|E_B\| + \|D_B\|) \text{ and } r_2 = \|E_B\| - h.
\]

**Proposition 4.3.** Suppose that \( \frac{1 + \epsilon}{\sqrt{1 - \epsilon^2}} < 2 \). Given any \( B \in B(A, s_0^1) \) and \( j = 1 \) or \( j = 2 \), we have

\[
\|\pi_j(B) - \rho_j(B)\| < 4\sqrt{\epsilon}\|B - \pi(B)\|.
\]

Proof. Lemma 4.2 implies that Proposition 2.4 applies to the point \( C = \pi(B) \). By a translation we can assume that \( \pi(B) = 0 \), which we do from now on. Denote \( \mathcal{D} = T_{M_j}(0), \mathcal{E} = \text{Span} \{B - \rho_j(B)\} \) and \( \mathcal{F} = \mathcal{K} \cap (\mathcal{D} + \mathcal{E}) \), (see Fig. 6). 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

\[
(4.1) \quad \|\pi_j(B) - D_B\| < 4\sqrt{\epsilon}\|B\|.
\]

First note that by Proposition 2.4 (ii) (applied with \( C = \pi(B) = 0 \) and \( D = D_B \)) the set \( M_j \cap B(D_B, \epsilon\|D_B\|) \) is not void. This set is included in the ball with center...
value in question is also larger than ∥B − DB∥ + ϵ∥DB∥ = ∥EB∥ + ϵ∥DB∥. Since πj(B) is the closest point to B on MJ, we conclude that

∥B − πj(B)∥ ≤ ∥EB∥ + ϵ∥DB∥.

If we write πj(B) = D + E + F (with D, E, F in the respective subspace D, E, F) this becomes

(4.2) ∥D − DB∥^2 + ∥E − EB∥^2 + ∥F∥^2 < (∥EB∥ + ϵ∥DB∥)^2.

However, by Proposition 2.4 (i) (applied with C = π(B) = 0 and D = πj(B)) we also get

(4.3) ∥E∥^2 + ∥F∥^2 < ϵ^2 (∥D∥^2 + ∥E∥^2 + ∥F∥^2).

The left hand side of (4.1) is thus dominated by the supremum of the function

δ(D, E, F) = ∥D + E + F − DB∥ = √∥D − DB∥^2 + ∥E∥^2 + ∥F∥^2,

subject to the conditions in (4.2) and (4.3). 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 = dDB and E = cEB where d, c ∈ R.

We now have a two-dimensional problem of circles and cones, and in the remainder of the proof we treat D, DB, E and EB as elements of R^2. Summing up, we want to maximize δ(D, E) = √∥D − DB∥^2 + ∥E∥^2 for D, E ∈ R^2 satisfying

(4.4) ∥D − DB∥^2 + ∥E − EB∥^2 < (∥EB∥ + ϵ∥DB∥)^2

and

(4.5) ∥E∥ < ϵ/√1 − ϵ^2 ∥D∥.

The first inequality gives ∥D∥ ≤ ∥D − DB∥ + ∥DB∥ < ∥EB∥ + ϵ∥DB∥ + ∥DB∥ which inserted in the second yields

∥E∥ < ϵ/√1 − ϵ^2 ((1 + ϵ)∥DB∥ + ∥EB∥).

As 1 + ϵ/√1 − ϵ^2 < 2, the following condition

(4.6) ∥E∥ < 2ϵ (∥DB∥ + ∥EB∥)

is less restrictive than (4.5) (when combined with (4.4), see Fig. 7). The function δ to be maximized is just the distance from D + E to DB, and clearly the maximal distance d (with constraints (4.4) and (4.6)) is obtained by the point x in the picture. Using the notation from the picture, we have

\[ d = ∥DB − x∥ = \sqrt{r_1^2 + h^2} = \sqrt{r^2 - r_1^2 + h^2} = \sqrt{∥EB∥^2 + ϵ∥DB∥^2 - (∥EB∥ - ϵ(∥EB∥ + ∥DB∥))^2 + (2ϵ∥EB∥ + ∥DB∥)^2} = \sqrt{2∥EB∥∥DB∥ + ϵ^2∥DB∥^2 + 4ϵ(∥EB∥^2 + ∥DB∥)^2}. \]

The assumption on ϵ clearly implies ϵ < 1 so ϵ^2 < ϵ. Moreover, ∥EB∥, ∥DB∥ ≤ ∥B∥ and hence the value obtained above is less than √11∥B∥^2 ≤ 4√∥B∥. Since the value in question is also larger than ∥πj(B) − ρj(B)∥ the proposition follows.

□
Lemma 4.4. Given \( B \in \mathcal{B}(A, s_t^1) \) we have
\[
\pi(B) = \pi(\rho_1(B)) = \pi(\rho_2(\rho_1(B))) .
\]

Proof. We begin with the first equality. By Lemma 4.2 we have \( \pi(B), \rho_1(B) \in \mathcal{B}(A, s_0) \), since \( s_t^0 < s_0 \), so Proposition 2.3 applies to give \( B - \pi(B) \perp T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\pi(B)) \). We obviously also have \( B - \rho_1(B) \perp T_{\mathcal{M}_1}(\pi(B)) \) so
\[
\rho_1(B) - \pi(B) = (\rho_1(B) - B) + (B - \pi(B)) \perp T_{\mathcal{M}_1 \cap \mathcal{M}_2}(\pi(B)).
\]
By Proposition 2.3 this implies that \( \pi(B) = \pi(\rho_1(B)) \), as desired. By Lemma 4.2, the above argument also applies to the point \( \rho_2(B) \) in place of \( B \), which yields \( \pi(\rho_2(B)) = \pi(\rho_1(\rho_2(B))) \). The desired second identity follows by reversing the roles of 1 and 2. \( \square \)

We finally have all the necessary ingredients.

Proof of Theorem 4.1. By the choice of \( s_0 \) and Proposition 2.3, \( \pi \) is \( C^2 \) on \( \mathcal{B}(A, s_0) \), and hence we can pick \( C > 0 \) such that
\[
(4.7) \quad \|\pi(B) - \pi(B')\| \leq C\|B - B'\|
\]
for all \( B, B' \in \mathcal{B}(A, s_0) \). Now fix \( \varepsilon \) such that \( C4\sqrt{\varepsilon} < \varepsilon \) and \( \frac{1 + \varepsilon}{\sqrt{1 + 4 \varepsilon}} < 2 \), and set \( s = s_t^1 \). By Lemma 4.4 we have \( \pi(B) = \pi(\rho_j(B)) \), and Lemma 4.2 guarantees that \( \pi_j(B), \rho_j(B) \in \mathcal{B}(A, s_0) \). By (4.7) and Proposition 4.3 we get
\[
\|\pi(\pi_j(B)) - \pi(B)\| = \|\pi(\pi_j(B)) - \pi(\rho_j(B))\| < C\|\pi_j(B) - \rho_j(B)\| \leq C4\sqrt{\varepsilon}\|B - \pi(B)\| < \varepsilon\|B - \pi(B)\| ,
\]
as desired. \( \square \)

Next we show that the distance to \( \mathcal{M}_1 \cap \mathcal{M}_2 \) is reduced in proportion to the angle, each time we project from \( \mathcal{M}_1 \) onto \( \mathcal{M}_2 \) or vice versa. Recall the function \( \sigma(A) \) from Definition 3.1.

Theorem 4.5. For each \( c > \sigma(A) \) there exists an \( s > 0 \) such that for all \( B \in \mathcal{M}_2 \cap \mathcal{B}(A, s) \) we have
\[
\|\pi_1(B) - \pi(B)\| < c\|B - \pi(B)\| .
\]
Moreover the same holds true with the roles of \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \) reversed.

Lemma 4.6. Let \( \hat{c} > 1 \) and \( \hat{c} > 0 \) be given. If \( E, F \in \mathcal{K} \) satisfies \( \|E\| > \hat{c}\|F\| \) and \( \|E - F\| < \hat{c} \), then
\[
\|E\| < \frac{\hat{c} + \hat{c}}{\hat{c} - 1} .
\]

Proof. If \( \|E\| < \hat{c} \) we are done, otherwise
\[
\hat{c} < \frac{\|E\|}{\|F\|} < \frac{\|E\|}{\|E\| - \hat{c}}
\]
which easily gives the desired estimate. \( \square \)

Proof of Theorem 4.5. Fix \( c_1 \) such that \( \sigma(A) < c_1 < c \) and pick an \( s_1 < s_0 \) such that
\[
\sup\{\sigma(C) : C \in \mathcal{M}_1 \cap \mathcal{M}_2 \cap \mathcal{B}(A, s_1)\} < c_1 ,
\]
which we can do since \( \sigma \) is continuous by Proposition 3.4. Let \( c_2 > 1 \) be such that \( c_2 c_1 < c \). By the choice of \( s_0 \) and Lemma 4.2, \( \rho_1 \) and \( \pi \) are \( C^1 \)-functions on \( B(A, s_0) \), and hence we can pick \( C > 0 \) such that
\[
\| \rho_j(B) - \rho_j(B') \| \leq C \| B - B' \| \quad \text{and} \quad \| \pi(B) - \pi(B') \| \leq C \| B - B' \|
\]
for all \( B, B' \in B(A, s_0) \). We now fix \( \epsilon \) such that \( \frac{1 + \epsilon}{\sqrt{1 - \epsilon^2}} < 2 \) and
\[
4\sqrt{\epsilon}(1 + C) \frac{c_2}{c_2 - 1} < c \quad \text{and} \quad (1 + 4\sqrt{\epsilon})c_2 c_1 < c.
\]
Then fix \( s < s_1 \) such that
\[
\pi(B(A, s)) \subset B(A, s_1),
\]
and let \( B \in \mathcal{M}_2 \cap B(A, s) \). There is no restriction to assume that \( \pi(B) = 0 \), which we do from now on. Note that \( \pi(B) \in \mathcal{M}_1 \cap \mathcal{M}_2 \cap B(A, s_1) \) so
\[
\sigma(0) = \sigma(\pi(B)) < c_1.
\]
Setting \( D = \pi_1(B) \), the desired inequality takes the form
\[
\| D \| / \| B \| < c.
\]
Put \( B' = \rho_2(B) \) and \( D' = \rho_1(B') \), (see Figure 8). Note that \( B \in B(A, s_1) \) whereas \( D, B', D' \in B(A, s_0) \) by Lemma 4.2. By Lemma 4.4 it then follows that \( 0 = \pi(\rho_2(B)) = \pi(B') \) and \( 0 = \pi(\rho_1(\rho_2(B))) = \pi(D') \) so
\[
D' = P_{T_{\mathcal{M}_1}(0)}(B') = P_{T_{\mathcal{M}_1}(0)}P_{T_{\mathcal{M}_2}(0)}(B'),
\]
since \( B' \in T_{\mathcal{M}_2}(0) \). By Proposition 2.3 we also have \( B' \perp T_{\mathcal{M}_1 \cap \mathcal{M}_2}(0) \) and hence
\[
D' = \left( P_{T_{\mathcal{M}_1}(0)}P_{T_{\mathcal{M}_2}(0)} - P_{T_{\mathcal{M}_1 \cap \mathcal{M}_2}(0)} \right) B'.
\]
By Proposition 3.4 and (4.10) we thus have
\[
\| D' \| / \| B' \| \leq \| P_{T_{\mathcal{M}_1}(0)}P_{T_{\mathcal{M}_2}(0)} - P_{T_{\mathcal{M}_1 \cap \mathcal{M}_2}(0)} \| = \sigma(0) < c_1,
\]

![Figure 8. Proof illustration](image-url)
which should be compared with (4.11). We now show that \( B' \approx B \) and \( D' \approx D \).

First, note that by Proposition 4.3

\[
\|B - B'\| = \|\pi_2(B) - \rho_2(B)\| < 4\sqrt{\epsilon}\|B\|.
\]

Moreover, by (4.8), (4.13) and Proposition 4.3 we have that

\[
\|D' - D\| = \|\rho_1(B') - \pi_1(B)\| \leq \|\rho_1(B') - \rho_1(B)\| + \|\rho_1(B) - \pi_1(B)\| \\
\leq C\|B' - B\| + 4\sqrt{\epsilon}\|B\| < 4\sqrt{\epsilon}(1 + C)\|B\|
\]

Apply Lemma 4.6 with \( E = D, F = D', \tilde{c} = 4\sqrt{\epsilon}(1 + C)\|B\| \) and \( \tilde{c} = c_2 \). We see that either

\[
\|D\| \leq c_2 \]

or \( \|D\| < 4\sqrt{\epsilon}(1 + C)\frac{c_2}{c_1}\|B\| \), in which case we are done since the constant is less than \( c \) by (4.9). We thus assume that (4.14) holds. Note that

\[
\|B'\| \leq 1 + 4\sqrt{\epsilon}
\]

by (4.13). Combining (4.12), (4.14) and (4.15) we get

\[
\frac{\|D\|}{\|B\|} = \frac{\|D\|\|B'\|\|D'\|}{\|B'\|\|B\|} < c_2(1 + 4\sqrt{\epsilon})c_1 < c,
\]

where the last inequality follows by (4.9). \( \square \)

5. Alternating projections

We are finally ready for the main theorem.

**Theorem 5.1.** Let \( \mathcal{M}_1, \mathcal{M}_2 \) and \( \mathcal{M}_1 \cap \mathcal{M}_2 \) be \( C^2 \)-manifolds and let \( A \in \mathcal{M}_1 \cap \mathcal{M}_2 \) be a non-tangential intersection point of \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \). Given \( \epsilon > 0 \) and \( 1 > c > \sigma(A) \), there exists an \( r > 0 \) such that for any \( B \in \mathcal{B}(A, r) \) the sequence of alternating projections

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

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

(ii) \( \|B_\infty - \pi(B)\| \leq \epsilon\|B - \pi(B)\| \)

(iii) \( \|B_\infty - B_k\| \leq \text{const} \cdot c^k\|B - \pi(B)\| \)

**Proof.** We may clearly assume that \( \epsilon < 1 \). Invoking Theorem 4.1 with

\[
\epsilon = \frac{1 - c}{3 - c},
\]

and Theorem 4.5 with \( c \) as above, gives us two possibly distinct radii. We let \( s \) denote the minimum of the two and pick

\[
r < \frac{s(1 - \epsilon)}{4(2 + \epsilon)}
\]

such that \( \pi(\mathcal{B}(A, r)) \subset \mathcal{B}(A, s/4) \). The latter condition ensures that

\[
\|\pi(B) - A\| < s/4.
\]
Let \( l = \|B - \pi(B)\| \) and observe that
\[
(5.4) \quad l \leq \|B - A\| + \|A - \pi(B)\| \leq r + s/4.
\]
As \( \pi(B) \in \mathcal{M}_1 \cap \mathcal{M}_2 \) we have \( \|B_0 - B\| = \|\pi_1(B) - B\| \leq \|\pi(B) - B\| = l \) and
\[
(5.5) \quad \|B_0 - \pi(B_0)\| \leq \|B_0 - \pi(B)\| \leq \|B_0 - B\| + \|B - \pi(B)\| \leq 2l.
\]
Applying Theorem 4.5 we get
\[
\|B_k - \pi(B_k)\| \leq \|B_k - \pi(B_{k-1})\| \leq c\|B_{k-1} - \pi(B_{k-1})\|
\]
as long as
\[
(5.6) \quad \{B_k\}_{k=0}^{k-1} \subset B(A, s),
\]
which will be established by induction. First note that
\[
\|B_0 - A\| \leq \|B_0 - B\| + \|B - A\| \leq l + r \leq 2r + s/4 \leq \frac{s(1 - \epsilon)}{2(2 + \epsilon)} + \frac{s}{4} < s
\]
by (5.2), (5.4) and (5.5), and hence (5.6) holds for \( k = 1 \). Now assume that it holds
\[
(5.7) \quad \|B_k - \pi(B_k)\| \leq c^k\|B_0 - \pi(B_0)\| \leq 2lc^k
\]
and by Theorem 4.1 we also have
\[
(5.8) \quad \|\pi(B_k) - \pi(B_{k-1})\| \leq \epsilon\|B_{k-1} - \pi(B_{k-1})\| \leq \epsilon(2lc^{k-1}).
\]
By the triangle inequality and Theorem 4.1 we get
\[
\|\pi(B_k) - \pi(B)\| \leq \|\pi(B) - \pi(B_0)\| + \sum_{j=1}^{k} \|\pi(B_j) - \pi(B_{j-1})\| \leq
\]
\[
(5.9) \quad \epsilon l + \sum_{j=1}^{k} \epsilon(2lc^{j-1}) < \epsilon l + \frac{2\epsilon l}{1 - c} = \frac{3 - \epsilon}{1 - c} \epsilon l = \epsilon l,
\]
where the last inequality follows by (5.1). Combining this with (5.3), (5.4), (5.7) we also have
\[
\|A - B_k\| \leq \|A - \pi(B)\| + \|\pi(B) - \pi(B_k)\| + \|\pi(B_k) - B_k\| < s/4 + \epsilon l + 2l \leq
\]
\[
\leq s/4 + \epsilon(r + s/4) + 2(r + s/4) = \frac{3 + \epsilon}{4}s + \frac{3 + \epsilon}{4}s + \frac{1 - \epsilon}{4}s = s,
\]
where the last inequality follows by (5.2). This shows that (5.6) holds for \( k := k + 1 \) and so (5.6) is true for all \( k \in \mathbb{N} \) by induction. In particular all the above estimates hold true for any \( k \in \mathbb{N} \).

By (5.8) we see that the sequence \( (\pi(B_k))_{k=1}^{\infty} \) is a Cauchy sequence, and hence it converges to some point \( B_\infty \). By (5.7) the sequence \( (B_k)_{k=1}^{\infty} \) must also converge, and the limit point is again \( B_\infty \), which thus satisfies \( B_\infty = \pi(B_\infty) \) since \( \pi \) is continuous. Hence \( (i) \) is established, and \( (ii) \) follows by taking the limit in (5.9). For \( (iii) \), note that by a similar calculation to (5.9) we have
\[
\|\pi(B_k) - B_\infty\| \leq \sum_{j=k+1}^{\infty} \|\pi(B_j) - \pi(B_{j-1})\| \leq \frac{\epsilon 2lc^k}{1 - c},
\]
which combined with (5.7) gives
\[
\|B_k - B_\infty\| \leq \|B_k - \pi(B_k)\| + \|\pi(B_k) - B_\infty\| \leq 2lc^k + \frac{\epsilon 2lc^k}{1 - c} = \frac{2 - 2c + 2\epsilon}{1 - c} c^k l,
\]
6. Non-tangentiality on real algebraic varieties

In this section we suppose that $\mathcal{K}$ has been given a basis and thus identify it with $\mathbb{R}^n$. In many applications the sets $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}$ are actually real algebraic varieties, that is, they can be defined by the vanishing of a set of polynomials on $\mathbb{R}^n$. To make the distinction explicit we denote them $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}$, respectively. This poses a few additional problems, but we shall see in the end of this section that these can be overcome and moreover, that non-tangentiality at one single intersection point implies non-tangentiality at the vast majority of points in $\mathcal{V}$, in a sense to be specified below. This result is interesting for applications since it says that the (a priori not known) closest point(s) to a given $B \in \mathbb{R}^n$ is likely to be non-tangential, so that Theorem 5.1 applies, given that we are close enough.

The first obstacle is that algebraic varieties have singular points, and are therefore not manifolds. This is true even for complex algebraic varieties, which of course can be regarded as real by separating the real and imaginary parts. Fortunately, we have the following theorem by H. Whitney [53].

**Theorem 6.1.** Given a real algebraic variety $\mathcal{V} \subset \mathbb{R}^n$, we can write $\mathcal{V} = \bigcup_{j=0}^{m} \mathcal{M}_j$ where each $\mathcal{M}_j$ is either void or a $C^\infty$-manifold of dimension $j$. Moreover, each $\mathcal{M}_j$, $j = 0, \ldots, m$ contains at most a finite number of connected components.

For example, consider the “Cartan umbrella” $\mathcal{V}$ given by $z(x^2 + y^2) - x^3 = 0$ in $\mathbb{R}^3$. The main part of the variety is a manifold of dimension 2, but it also contains the $z$-axis as a submanifold of complex dimension 1. Thus, in terms of Theorem 6.1, we have $\mathcal{V} = \mathcal{M}_1 \cup \mathcal{M}_2$ where

$$\mathcal{M}_2 = \{ (x, y, z) : (x, y) \neq 0 \text{ and } z = x^3/(x^2 + y^2) \}$$

and $\mathcal{M}_1 = \{0\} \times \{0\} \times \mathbb{R}$. For more interesting examples in this direction we refer to [4], Chapter 3.

Theorem 6.1 shows that the main part of $\mathcal{V}$ is a manifold, and we can apply the theory from the previous sections since the statements are local. We note that $\bigcup_{j=0}^{m-1} \mathcal{M}_j$ makes up a very small part of $\mathcal{V}$. We will not try to quantify this statement, but note that in the case $m = 1$, $\bigcup_{j=0}^{m-1} \mathcal{M}_j$ would correspond to a finite collection of points to be compared with the curve $\mathcal{M}_1$. Another example is obtained by taking $\mathcal{V} = \mathbb{R}^n$. The information we have about $\bigcup_{j=0}^{m-1} \mathcal{M}_j$ is then much stronger than e.g. saying that $\bigcup_{j=0}^{m-1} \mathcal{M}_j$ has Lebesgue measure zero, which is a common criterion for “negligible” sets, see for example Sard’s theorem [45].

Our next aim is to find conditions on two real algebraic varieties $\mathcal{V}_1$ and $\mathcal{V}_2$ which guarantee that Theorem 5.1 can be applied to “the majority” of points in $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2$, in the sense that the exceptional set is a union of manifolds of lower dimension with finitely many connected components. To do this, we need to connect Theorem 6.1 with classical algebraic geometry in $\mathbb{C}^n$. This is also done in the seminal paper [53], which has given rise to a rich theory on decomposition of various sets into manifolds, now known as (Whitney) stratification theory. For an overview and historical account we refer to [50]. However, although Propositions 6.2-6.4 below are implicitly shown in Sections 8-11 in [53], we have not been able to find a good reference for all the results we need here. Hence, for completeness...
we provide proofs in Appendix B. The material in this section is self-contained and it is not necessary to know algebraic geometry in order to apply the results below.

Usually, by an algebraic variety one refers to a subset of $\mathbb{C}^n$ defined as the common zeroes of a set of (analytic) polynomials (with complex coefficients). To distinguish between these varieties and real algebraic varieties, we will call them “complex algebraic varieties” or simply complex varieties. Note that all complex varieties are real varieties, but not conversely. However, if we identify $\mathbb{R}^n$ with a subset of $\mathbb{C}^n$ in the usual way, then each real variety $V$ has a related complex variety given by its Zariski closure $V_{Zar}$, defined as the subset in $\mathbb{C}^n$ of common zeroes to all polynomials that vanish on $V$. Given a real algebraic manifold $V$ we let $\mathbb{I}_R(V)$ denote the set of real polynomials that vanish on $V$.

There are several equivalent ways of defining the dimension of a complex algebraic variety, see e.g. [15, 47]. The below proposition basically states that these coincide with the maximal non-trivial dimension in any Whitney composition.

**Proposition 6.2.** Let $V$ be a real algebraic variety and let $V = \bigcup_{j=0}^{m} M_j$ be a decomposition as in Theorem 6.1, with $\mathcal{M}_m \neq \emptyset$. Then $m$ equals the algebraic dimension of $V_{Zar}$.

Hence, although Whitney decompositions are not unique (consider e.g. $\mathbb{R} = (\mathbb{R} \setminus \{0\}) \cup \{0\}$), the maximal dimension of its components is. The number $m$ will henceforth be called the dimension of $V$. For example, it can be verified that the Zariski closure of the Cartan umbrella equals the variety in $\mathbb{C}^3$ defined by the same equation, and that this variety has algebraic dimension 2, as predicted by the above proposition and (6.1).

We say that $V$ is irreducible if there does not exist any non-trivial decompositions of the form $V = V_1 \cup V_2$, where $V_1$ and $V_2$ are algebraic varieties. A simple condition which guarantees this (and involves no algebraic geometry) is given in Proposition 6.8. We say that a point $A \in V$ is non-singular if it is non-singular in the sense of algebraic geometry as an element of $V_{Zar}$. The theory becomes much simpler if we restrict attention to irreducible varieties, which we do from now on. Let $\nabla$ denote the gradient operator and set $N_V(A) = \{\nabla p(A) : p \in \mathbb{I}_R(V)\}$. Note that this is a linear subspace of $\mathbb{R}^n$.

**Proposition 6.3.** Let $V \subset \mathbb{R}^n$ be an irreducible real algebraic variety of dimension $m$. Then $\dim N_V(A) \leq n - m$ for all $A \in V$ and $A$ is non-singular if and only if
\[
\dim N_V(A) = n - m.
\]

The subset of $V$ of non-singular points will be denoted $V^{ns}$.

**Proposition 6.4.** Let $V$ be an irreducible real algebraic variety of dimension $m$. Then the decomposition $V = \bigcup_{j=0}^{m} M_j$ in Theorem 6.1 can be chosen such that $V^{ns} = M_m$. Moreover, given $A \in V^{ns}$ we have
\[
T_{V^{ns}}(A) = (N_V(A))^\perp.
\]

Note that the counterpart of (6.3) in the complex setting also holds, (this will follow in Appendix B). We remark that in this case, the right hand side is the definition of the tangent space of $V$ in algebraic geometry, so (6.3) implies that the “algebraic tangent space” coincides with the “differential geometry tangent space” at non-singular points.
Proposition 6.5. Suppose that $V_1$ and $V_2$ are irreducible real algebraic varieties and that $V = V_1 \cap V_2$ is irreducible and strictly smaller than both $V_1$ and $V_2$. Then each point in $V_{1s} \cap V_{2s} \cap V_{ns}$ is a non-trivial intersection point.

The next theorem shows that in this setting, the majority of points are non-tangential if one is. It also gives a simple criterion under which this is the case. Denote by $V_{ns,nt} \subset V$ the set of all points in $V_{ns} \cap V_{1s} \cap V_{2s}$ that are non-tangential with respect to the manifolds $V_{1s}$ and $V_{2s}$.

Theorem 6.6. Suppose that $V_1$ and $V_2$ are irreducible real algebraic varieties and that $V = V_1 \cap V_2$ is irreducible and strictly smaller than both $V_1$ and $V_2$. Let the dimension of $V$ be $m$. If $V_{ns,nt} \neq \emptyset$, then $V \setminus V_{ns,nt}$ is a real algebraic variety of dimension strictly less than $m$. A sufficient condition for this to happen is that there exists a point $A \in V_{1s} \cap V_{2s}$ such that

$$\dim (T_{V_{1s}}(A) \cap T_{V_{2s}}(A)) \leq m. \tag{6.4}$$

Combined with Theorem 6.1 and Proposition 6.2, the theorem implies that $V \setminus V_{ns,nt}$ is a union of finitely many disjoint connected $C^\infty$-manifolds of dimension strictly less than $m$. In other words, the “good points” make up the vast majority of points in $V$. Note that (6.4) is equivalent with

$$\dim (T_{V_{1s}}(A) + T_{V_{2s}}(A)) \geq m_1 + m_2 - m$$

by basic linear algebra.

Theorem 6.6 thus provides a concrete condition under which Theorem 5.1 applies to a majority of points in $V$. We now provide concrete means to verify the preconditions of Theorem 6.6.

Definition 6.7. Suppose we have given a $j \in \mathbb{N}$ and an index set $I$ such that for each $i \in I$, there exists an open connected $\Omega_i \subset \mathbb{R}^j$ and a real analytic map $\phi_i : \Omega_i \rightarrow V$. We will say that $V$ can be covered with analytic patches if, for each $A \in V$, there exists an $i \in I$ and a radius $r_A$ such that

$$V \cap B_{\mathbb{R}^n}(A, r_A) = \text{Im} \phi_i \cap B_{\mathbb{R}^n}(A, r_A).$$

Note that we put no restriction on the rank of $d\phi_i$ at any point, so that $\mathbb{R}^j$ may contain “redundant variables” and “degenerate points”.

Proposition 6.8. Let $V$ be a real algebraic variety. If $V$ is connected and can be covered with analytic patches, then $V$ is irreducible. The same conclusion holds if a dense subset of $V$ can be given as the image of one real analytic function $\phi$.

To apply Theorem 6.6, one needs to know the various dimensions involved. This can sometimes be tricky. For example, the curve $y^2 = x^2(x - 1)$ in $\mathbb{R}^2$ has dimension 1 but an isolated point at the origin. Thus a local analysis of the surface may lead to false conclusions concerning the dimension. A more complex example is provided by the “Cartan umbrella” discussed earlier. We end this section with a simple method
of determining the dimension. An alternative is to work with quotient manifolds, as explained in [1], but we will not pursue this here.

**Proposition 6.9.** Under either of the assumptions of Proposition 6.8, suppose in addition that an open subset of $V$ is the image of a bijective real analytic map defined on a subset of $\mathbb{R}^d$. Then $V$ has dimension $d$.

A final remark. Suppose that $K$ to begin with is a vector space over $\mathbb{C}$. In this case one may of course identify $\mathbb{C}$ with $\mathbb{R}^2$ and apply the results of this section. However, all subspaces considered above then become closed over $\mathbb{C}$ and all results have a natural complex counterpart where e.g. dimension refers to the dimension over $\mathbb{C}$. We leave these reformulations as well as their proofs to the reader.

7. An example

We provide a simple example in which Theorem 5.1 applies; Given an $n \times n$ symmetric matrix $B$ and an integer $k$, we seek the rank $k$ correlation matrix that lies closest to $A$ with respect to the Hilbert-Schmidt norm. For this problem, we choose $K$ to be the set of real symmetric $n \times n$-matrices, $V_1$ the affine subspace of matrices with 1's on the diagonal, and $V_2$ the subset of matrices with rank less than or equal to $k$, where $k < n$. This example is also considered in [11] as well as in [2], where the reformulation suggested in [25] is used so that it fits the framework of [2].

For the alternating projections method to be useful, we need fast implementations of the “projections” $\pi_1$ and $\pi_2$. Since $V_1$ is an affine subspace, $\pi_1$ is trivial to compute and well-defined everywhere. For $\pi_2$, a theorem essentially due to E. Schmidt [46], sometimes attributed to Eckart-Young [20], states that the closest rank $k$ (or less) approximation for a given $B$ is obtained by computing the singular value decomposition and then replace the $n - k$ smallest singular values by 0. This is well known and explained e.g. in [35], Example 5. Note that when the $k$th singular value has multiplicity higher than 1, the statement still applies but the best approximation is no longer unique. This is the reason why $\pi_2$ needs to be a point to set map when defined globally. However recall Proposition 2.3 and Proposition 6.4 which imply that $\pi_2$ is a well defined function near non-singular points of $V_2$.

We show that Theorem 5.1 applies to the majority of points in $V = V_1 \cap V_2$. First we need to obtain basic results concerning the structure of these sets.

**Proposition 7.1.** $K$ has dimension $\frac{n^2 + n}{2}$. $V_1$ is an affine subspace of dimension $\frac{n^2 - n}{2}$. $V_2$ is an irreducible real algebraic variety of dimension $\frac{2nk - k^2 + k}{2}$. $V$ is an irreducible real algebraic variety of dimension $\frac{2nk - k^2 + k}{2} - n$.

**Proposition 7.2.** $V_2^{ns}$ equals all matrices with rank equal to $k$ and

$V^{ns,nt} \neq \emptyset$.

Combining the above results with Proposition 6.5 and Theorem 6.6 we immediately obtain.

**Corollary 7.3.** The set $V^{ns,nt}$ of points on $V$ where Theorem 5.1 applies is an $\frac{2nk - k^2 + k}{2} - n$ dimensional manifold. Its complement $V \setminus V^{ns,nt}$ is a finite set of connected manifolds of lower dimension.
The proofs are given in Appendix C. We now present a concrete application. Consider random variables $X_1, \ldots, X_n$ that arise as

$$X_l = \sum_{j=1}^{k} c_{l,j} Y_j,$$

where $\{Y_j\}_{j=1}^{k}$ are independent random variables with a unit normal distribution, and where $\sum_{j=1}^{k} c_{l,j}^2 = 1$. Then the correlation matrix $C$ is given by

$$C(l, l') = \text{var}(X_l, X_{l'}) = \sum_{j=1}^{k} c_{l,j} c_{l',j},$$

which clearly is an $n \times n$-matrix in $\mathcal{K}$ with rank $k$ and ones on the diagonal, i.e. $C \in \mathcal{V}_1 \cap \mathcal{V}_2$. Let $N \in \mathbb{N}$ be given and $X \in \mathbb{M}_{N,n}$ a matrix where each row is a realization of $(X_i)_{i=1}^{n}$. If $N$ is large we should have $X^T X / N \approx C$. However, in real situations, for instance in financial applications [27], some of the values $X(j, l)$ are
not available. For the computation of the correlation matrix corresponding to (7.1),
it is then common to replace the missing elements by zeros. Given an index set \( I \)
that indicates the missing data points, let \( \tilde{X}_I \) denote the matrix corresponding
to \( X \) where the missing elements has been replaced by zeros, and let \(|I|\) denote
the number of missing elements. Generically, the matrix \( \tilde{X}_I^T \tilde{X}_I/N \) will no longer be
in \( \mathcal{V}_1 \cap \mathcal{V}_2 \). The distance between \( \tilde{X}_I^T \tilde{X}_I/N \) and \( \mathcal{V}_1 \cap \mathcal{V}_2 \) will tend to increase as the
number of missing data points increases. \( \pi(\tilde{X}_I^T \tilde{X}_I/N) \) is known to be computable
by using semi definite programming [49, Section 4.4]. We have used the SDPT3
implementation [48, 51].

Figure 9 shows results for 1000 simulations with varying number of missing
elements to generate initial matrices \( B = \tilde{X}_I^T \tilde{X}_I/N \). The plot displays logarithmic
values of the relative error \( \|B - \pi(B)\|/\|B - \pi(B)\| \) as a function of logarithmic values of the
distance \( \|B - \pi(B)\| \) to \( \mathcal{V}_1 \cap \mathcal{V}_2 \). The matrices \( X \) are of size \( 10^5 \times 50 \) and have
rank \( k = 10 \). The number of missing data points varied linearly from \( 10^3 \) to \( 10^6 \).
An increasing linear trend between the errors can clearly be seen. This should
be compared with Theorem 5.1, which predicts that we should see an increasing
function.

8. Appendix A

Recall the map \( \phi \) introduced in Section 2 and let \( A \) be a point in \( \mathcal{K} \). We will
throughout without restriction assume that \( \phi(0) = A \) and that the domain of
definition of \( \phi \) is \( B_{\mathbb{R}^m}(0, r) \) for some \( r > 0 \). We also let \( s \) be such that (2.1) holds.

**Proposition 2.1** Let \( \alpha : \mathcal{K} \to \mathcal{M} \) be any \( C^p \)-map, where \( \mathcal{M} \) is a \( C^p \)-manifold
and \( p \geq 1 \). Then the map \( \phi^{-1} \circ \alpha \) is also \( C^p \) (on its natural domain of definition).

**Proof.** Given \( x_0 \in B_{\mathbb{R}^m}(0, r) \) pick \( f_1, \ldots, f_{n-m} \in \mathcal{K} \) with the property that
\[
(T_{\mathcal{M}}(\phi(x_0)))^\perp = \text{Span} \{ f_1, \ldots, f_{n-m} \}
\]
and define \( \omega_{x_0} : B_{\mathbb{R}^m}(0, r) \times \mathbb{R}^{n-m} \to \mathcal{K} \) via
\[
\omega_{x_0}(x, y) = \phi(x) + \sum_{i=1}^{n-m} y_i f_i.
\]
By the inverse function theorem ([9] 0.2.22) \( \omega_{x_0} \) has a \( C^p \)-inverse in a neighborhood
of \((x_0, 0)\). The proposition now follows by noting that for values of \( \alpha \) near \( \phi(x_0) \)
we have \( \phi^{-1} \circ \alpha = \omega_{x_0}^{-1} \circ \alpha \).

**Proposition 2.2** Let \( \mathcal{M} \) be a \( C^1 \)-manifold. Then \( P_{T_{\mathcal{M}}(A)} \) is a continuous
function of \( A \).

**Proof.** Given \( A \in \text{Im } \phi \), set \( M = d\phi(\phi^{-1}(A)) \). It is easy to see that
\[
P_{T_{\mathcal{M}}(A)} = M(M^*M)^{-1}M^*.
\]
The conclusion now follows as \( d\phi \) and \( \phi^{-1} \) are continuous.

Propositions 2.3 and 2.4 are a bit harder. We begin with a lemma.

**Lemma 8.1.** If \( B \in \mathcal{K} \) is given and \( A \) is the closest point in \( \mathcal{M} \), then \( B - A \perp
T_{\mathcal{M}}(A) \). Moreover, \( \|\phi(x) - \phi(y)\|/\|x - y\| \) is uniformly bounded above and below
for \( x, y \) in any \( B(0, r') \), \( r' < r \).
Proof. Note that
\[ \phi(x) = A + d\phi(0)x + o(x) \]
where \( o \) stands for a function with the property that \( o(x)/\|x\| \) extends by continuity to 0 and takes the value 0 there. Thus we have
\[
\|\phi(x) - B\|^2 = \|A + d\phi(0)x + o(x) - B\|^2 = \\
\|A - B\|^2 + 2\langle A - B, d\phi(0)x + o(\|x\|) \rangle
\]
and hence the scalar product needs to be zero for all \( x \)'s. For the second claim, set \( w = (\phi(y) - \phi(x))/\|\phi(y) - \phi(x)\| \) and apply the mean value theorem to \( \gamma(t) = \langle \phi(x + (y - x)t) - \phi(x), w \rangle \) to conclude that
\[
\|\phi(y) - \phi(x)\| = \gamma(1) - \gamma(0) = \langle d\phi(z)(y - x), w \rangle
\]
for some \( z \) on the line between \( x \) and \( y \). Letting \( \sigma_1(d\phi(z)), \ldots, \sigma_m(d\phi(z)) \) denote the singular values of \( d\phi(z) \), we thus have
\[
\inf_{z \in B_{\mathbb{R}^m}} \{ \sigma_m(d\phi(z)) \} \|y - x\| \leq \|\phi(y) - \phi(x)\| \leq \sup_{z \in B_{\mathbb{R}^m}} \{ \sigma_1(d\phi(z)) \} \|y - x\|. 
\]
Now, \( d\phi(z) \) depends continuously on \( z \) and its singular values depend continuously on the matrix entries \( [9, p.19] \). Since \( \phi \) is an immersion we have \( \sigma_m(d\phi(x)) \neq 0 \) for all \( x \in B(0, r) \), so by compactness of \( cl(B(0, r')) \) we get that both the inf and sup amount to finite positive numbers, as desired.

Since \( s \) has been specified above, we work with \( s' \) in the below formulation.

Proposition 2.3. Let \( M \) be a \( C^2 \)-manifold. Given any fixed \( A \in M \), there exists \( s' > 0 \) and a \( C^2 \)-map
\[
\pi : B_{\mathbb{K}}(A, s') \to M
\]
such that for all \( B \in B_{\mathbb{K}}(A, s') \) there exists a unique closest point in \( M \) which is given by \( \pi(B) \). Moreover, \( C \in \mathcal{M} \cap B_{\mathbb{K}}(A, s') \) equals \( \pi(B) \) if and only if \( B - C \perp T_{\mathcal{M}}(C) \).

Proof. We repeat the standard construction of a tubular neighborhood of \( M \), (see e.g., \([9, p.40]\)). By standard differential geometry there exists an \( r_0 < r \) and \( C^1 \)-functions \( f_1, \ldots, f_{n-m} : \mathbb{R}^n \to K \) with the property that
\[
(T_{\mathcal{M}}(\phi(x)))^\perp = \text{Span} \{ f_1(x), \ldots, f_{n-m}(x) \}
\]
for all \( x \in B_{\mathbb{R}^n}(0, r_0) \). Moreover, applying the Gram-Schmidt process we may assume that \( \{ f_1(x), \ldots, f_{n-m}(x) \} \) is an orthonormal set for all \( x \in B_{\mathbb{R}^n}(0, r_0) \).

Define \( \tau : B_{\mathbb{R}^n}(0, r_0) \times \mathbb{R}^{n-m} \to K \) via
\[
\tau(x, y) = \phi(x) + \sum_{i=1}^{n-m} y_if_i(x).
\]
\( \tau \) is \( C^1 \) by construction, so the inverse function theorem implies that there exists \( r_1 \) such that \( \tau \) is a diffeomorphism from \( B_{\mathbb{R}^n}(0, r_1) \) onto a neighborhood of \( A \). Choose \( s' \) such that \( s' < r_1/4 \), \( 2s' < s \) and
\[
B_{\mathbb{K}}(A, 2s') \subset \tau(B(0, r_1/2)).
\]
Given any \( B \in \mathcal{B}_K(A,s') \) there thus exists a unique \( (x_B,y_B) \) such that \( B = \tau((x_B,y_B)) \) and \( \| (x_B,y_B) \| \leq r_1/2 \). We define
\[
\pi(B) = \phi(x_B).
\]

To see that \( \pi \) is a \( C^1 \)-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 on \( \mathcal{B}(A,s') \) we have
\[
\pi = \phi \circ \theta \circ (\tau|_{\mathcal{B}_K(0,r_1)})^{-1}.
\]

For the fact that \( \pi \) is actually \( C^2 \), (which is not needed in this paper), we refer to [31] or Sec. 14.6 in [23]. We now show that \( \pi(B) \) have the desired properties. Suppose \( C \in \mathcal{M} \) is a closest point to \( B \). Since \( \| A - B \| < s' \) we clearly must have \( \| C - A \| < 2s' \) so by (2.1) and (8.1) there exists a \( x_C \in B_{\mathbb{R}^m}(0,r_1/2) \) with \( \phi(x_C) = C \). Moreover, by Lemma 8.1 \( B - C \perp T_{\mathcal{M}}(C) \) so (by the orthonormality of the \( f \)'s at \( x_C \)) there exists a \( y \in \mathbb{R}^{n-m} \) with \( \tau(x_C,y) = B \) and \( \| y \| = \| B - C \| < s' < r_1/4 \).

Since \( \| (x_C,y) \| \leq r_1/2 + r_1/4 \) and \( \tau \) is a bijection in \( B(0,r_1) \), we conclude that \( (x_C,y) = (x_B,y_B) \). Hence \( \pi(B) = \phi(x_B) = \phi(x_C) = C \). This establishes the first part of the proposition.

By the construction, \( \pi(B) - B \in \text{Span} \{ f_1(x_B), \ldots, f_{n-m}(x_B) \} \) is orthogonal to \( T_{\mathcal{M}}(\phi(x_B)) = T_{\mathcal{M}}(\pi(B)) \). Conversely, let \( C \) be as in the second part of the proposition. As above we have \( C = \phi(x_C) \) with \( \| x_C \| < r_1/2 \) and there exists a \( y \) with \( B = \tau((x_C,y)) \) and \( \| y \| = \| B - C \| < 2s' < r_1/2 \). As earlier, this implies \( C = \pi(B) \), as desired.

\[
\square
\]

**Proposition 2.4** Let \( \mathcal{M} \) be a locally \( C^2 \)-manifold at \( A \). For each \( \epsilon > 0 \) there exists \( s_\epsilon > 0 \) such that for all \( C \in \mathcal{B}_K(A,s_\epsilon) \cap \mathcal{M} \) we have
\[
(i) \text{ dist}(D,\tilde{T}_{\mathcal{M}}(C)) < \epsilon \| D - C \|, \quad D \in \mathcal{B}(A,s_\epsilon) \cap \mathcal{M},
\]
\[
(ii) \text{ dist}(D,\mathcal{M}) < \epsilon \| D - C \|, \quad D \in \mathcal{B}(A,s_\epsilon) \cap \tilde{T}_{\mathcal{M}}(C).
\]

**Proof.** We first prove (i). We may clearly assume that \( s_\epsilon < s \). Set \( w = D - (C + d\phi(x_C)(x_D - x_C)) \) and note that \( \text{dist}(D,\tilde{T}_{\mathcal{M}}(C)) \leq \| w \| \). Apply the mean value theorem to the function
\[
\gamma(t) = \left( \phi\left((x_D - x_C)t + x_C\right) - \phi(x_C) - d\phi(x_C)(x_D - x_C)t, \frac{w}{\| w \|} \right)
\]
to conclude that there exists a \( y \) on the line \([x_C,x_D]\) between \( x_C \) and \( x_D \) such that
\[
\| w \| = \gamma(1) - \gamma(0) = \left( \left( d\phi(y + x_C) - d\phi(x_C) \right)(x_D - x_C), \frac{w}{\| w \|} \right).
\]

By the Cauchy-Schwartz's inequality we get
\[
\text{dist}(D,\tilde{T}_{\mathcal{M}}(C)) \leq \| d\phi(y + x_C) - d\phi(x_C) \| \| x_D - x_C \|
\]
for some \( y \in [x_C,x_D] \). By Lemma 8.1 and the equicontinuity of continuous functions on compact sets, we can choose an \( r_\epsilon < r \) such that \( \text{dist}(D,\tilde{T}_{\mathcal{M}}(C)) < \epsilon \| D - C \| \) for all \( x_D, x_C \in B(0,r_\epsilon) \). As in the proof of Proposition 2.3 we can now define \( \tau \) and use it to pick an \( s_\epsilon \) such that \( x_C, x_D \in B(0,r_\epsilon) \) whenever \( C, D \in B(A,s_\epsilon) \). Hence (i) holds.

The proof of (ii) is similar. We assume without restriction that \( C \in \text{Im} \phi \) and let \( y \) be such that \( D = C + d\phi(x_C)y \). We may chose \( s_\epsilon \) small enough that \( \| x_C + y \| < r \),
and then clearly \( \text{dist}(D, \mathcal{M}) \leq \| \phi(x_C + y) - D \|. \) Setting \( w = \phi(x_C + y) - D \) and applying the mean value theorem to

\[
\gamma(t) = \left( \phi(x_C + yt) - C - d\phi(x_C)yt, \frac{w}{\|w\|} \right)
\]

we easily obtain

\[
\text{dist}(D, \mathcal{M}) \leq \|w\| \leq \|d\phi(x_C + t_0 y) - d\phi(x_C)\| \|y\|
\]

for some \( t_0 \in [0, 1] \). We omit the remaining details.

\[
\square
\]

9. Appendix B

When dealing with concepts such as dimension or the ideal generated by a variety, we will use \( \mathbb{R} \) and \( \mathbb{C} \) as subscripts when there can be confusion about which field is involved. To begin, a short argument shows that

\[
\mathbb{I}_C(\mathcal{V}_{\text{Zar}}) = \mathbb{I}_R(\mathcal{V}) + i\mathbb{I}_R(\mathcal{V}),
\]

from which \( (6.2) \) easily follows. In case we work over \( \mathbb{C} \), we define \( \nabla \) as \( \nabla = (\partial_{z_1}, \ldots, \partial_{z_n}) \) where \( \partial_z \) refers to the formal partial derivatives, or, which is the same, the standard derivatives for analytic functions (see e.g. Definition 3.3, Ch. II of [30]). We need the following classical result from algebraic geometry. See e.g. Theorem 3 and the comments following it, Ch. II, Sec. 1. [47].

**Theorem 9.1.** Let \( \mathcal{V} \) be an irreducible complex algebraic variety with algebraic dimension \( d \). Then \( \dim \mathbb{C}\text{Span} \{\nabla p(A) : p \in \mathbb{I}_C(\mathcal{V})\} < n - d \) for all singular points \( A \in \mathcal{V} \) and \( \dim \mathbb{C}\text{Span} \{\nabla p(A) : p \in \mathbb{I}_C(\mathcal{V})\} = n - d \) for all non-singular points \( A \in \mathcal{V} \). Moreover the set of singular points form a complex variety of lower dimension.

If \( P \) is a set of polynomials, we will write \( \mathcal{V}(P) \) for the variety of common zeroes (in \( \mathbb{R}^n \) or \( \mathbb{C}^n \) depending on the context). Given two sets \( S_1 \) and \( S_2 \) and a common point \( A \), we will write that \( S_1 \equiv S_2 \) near \( A \), if there exists an open set \( U \) containing \( A \) such that \( S_1 \cap U = S_2 \cap U \). Theorem 9.1 shows that for an irreducible variety, the non-singular points coincide with those of maximal rank in the work of H. Whitney [33]. This observation combined with Sections 6 and 7 of that paper yields the following key result.

**Theorem 9.2.** Let \( \mathcal{V} \) be an irreducible complex algebraic manifold and \( A \) a non-singular point. Given any \( p_1, \ldots, p_{n-d} \in \mathbb{I}(\mathcal{V}) \) such that \( \nabla p_1(A), \ldots, \nabla p_{n-d}(A) \) are linearly independent, we have \( \mathcal{V} \equiv \mathcal{V}((p_1, \ldots, p_{n-d})) \) near \( A \).

The remark after Proposition 6.4 now easily follows from the above results and the complex version of the implicit function theorem (see e.g. Theorem 3.5, Ch. II, [30]). In order to transfer these results to the real setting, we first need a lemma.

**Lemma 9.3.** A real algebraic variety \( \mathcal{V} \) is irreducible in \( \mathbb{R}^n \) if and only if \( \mathcal{V}_{\text{Zar}} \) is irreducible in \( \mathbb{C}^n \).

**Proof.** If \( \mathcal{V} \) has a nontrivial decomposition in \( \mathbb{R}^n \) into smaller varieties as \( \mathcal{U}_1 \cup \mathcal{U}_2 \), then clearly \( (\mathcal{U}_1)_{\text{Zar}} \cup (\mathcal{U}_2)_{\text{Zar}} \) is a nontrivial decomposition of \( \mathcal{V}_{\text{Zar}} \). Conversely, let \( \mathcal{V}_{\text{Zar}} = \mathcal{U}_1 \cup \mathcal{U}_2 \) be a nontrivial decomposition of \( \mathcal{V}_{\text{Zar}} \) in \( \mathbb{C}^n \). Let \( \mathcal{I}_C(\mathcal{U}_i) \) denote the respective complex ideals, \( i = 1, 2 \), and set \( \mathcal{I}_i = (\text{Re} \mathcal{I}_C(\mathcal{U}_i)) \cup (\text{Im} \mathcal{I}_C(\mathcal{U}_i)) \). The real variety corresponding to each \( \mathcal{I}_i \) clearly coincides with \( \mathcal{U}_i \cap \mathbb{R}^n \), which thus is
a variety in $\mathbb{R}^n$. Since $\mathcal{V}_{\text{Zar}}$ is the smallest complex variety including $\mathcal{V}$, $\mathcal{V}$ is not a subset of either $U_1$ or $U_2$, and hence $\mathcal{V} = (U_1 \cap \mathbb{R}^n) \cup (U_2 \cap \mathbb{R}^n)$ is a nontrivial real decomposition of $\mathcal{V}$. □

**Lemma 9.4.** Let $\mathcal{V}$ be an irreducible real algebraic variety such that $\mathcal{V}_{\text{Zar}}$ has algebraic dimension $d$. Then $\dim_{\mathbb{R}} N_{\mathcal{V}}(A) < n - d$ for all singular points $A \in \mathcal{V}$ and $\dim_{\mathbb{R}} N_{\mathcal{V}}(A) = n - d$ for all non-singular points $A \in \mathcal{V}$.

**Proof.** First note that $\mathcal{V}_{\text{Zar}}$ is irreducible, by Lemma 9.3, and hence Theorem 9.1 applies to $\mathcal{V}_{\text{Zar}}$. Given any real vector space $V$, it follows by linear algebra that $\dim_{\mathbb{R}} V = \dim_{\mathbb{C}} (V + iV)$. By (9.1) we thus get

$$\dim_{\mathbb{R}} \text{Span}_{\mathbb{R}} \{ \nabla p(A) : p \in \mathbb{I}_\mathbb{R}(\mathcal{V}) \} = \dim_{\mathbb{C}} \text{Span}_{\mathbb{C}} \{ \nabla p(A) : p \in \mathbb{I}_\mathbb{C}(\mathcal{V}) \}$$

for all $A \in \mathbb{R}^n$, and hence the lemma follows by Theorem 9.1. □

**Lemma 9.5.** Let $\mathcal{V}$ be an irreducible real algebraic variety such that $\mathcal{V}_{\text{Zar}}$ has algebraic dimension $d$. Then $\mathcal{V}^\text{ns}$ is a $C^\infty$-manifold of dimension $d$ and $\mathcal{V} \setminus \mathcal{V}^\text{ns}$ is a real algebraic manifold such that $(\mathcal{V} \setminus \mathcal{V}^\text{ns})_{\text{Zar}}$ has algebraic dimension strictly less than $d$.

**Proof.** By Lemma 9.3, $\mathcal{V}_{\text{Zar}}$ is irreducible. By Theorem 9.1, we have that the singular points of $\mathcal{V}_{\text{Zar}}$ form a proper subvariety of dimension strictly less than $d$. Hence $\mathcal{V} \setminus \mathcal{V}^\text{ns}$ is included in a complex subvariety of lower dimension than $d$. Since the algebraic dimension decreases when taking subvarieties, we conclude that $\dim (\mathcal{V} \setminus \mathcal{V}^\text{ns})_{\text{Zar}} < d$. If $\mathcal{V}^\text{ns}$ would be empty, $(\mathcal{V} \setminus \mathcal{V}^\text{ns})_{\text{Zar}}$ would include all of $\mathcal{V}$, contradicting the definition of $\mathcal{V}_{\text{Zar}}$ as the smallest complex variety including $\mathcal{V}$. Finally, given $A \in \mathcal{V}^\text{ns}$, Theorem 9.2 and Lemma 9.4 imply that we can find $p_1, \ldots, p_{n-d} \in \mathbb{I}_\mathbb{R}(\mathcal{V})$ with linearly independent derivatives at $A$ such that

$$\mathcal{V}^\text{ns} \overset{\text{loc}}{=} \mathcal{V}(\{p_1, \ldots, p_{n-d}\})$$

near $A$. The fact that $\mathcal{V}^\text{ns}$ is a $d$-dimensional $C^\infty$-manifold now follows directly from Theorem 2.1.2 (ii) in [9]. □

We are now ready to prove the results in Section 6.

**Proof of Proposition 6.2 and 6.4.** We begin with Proposition 6.2. Let $\mathcal{V}_{\text{Zar}}$ have algebraic dimension $d$. First assume that $\mathcal{V}$ is irreducible. By Lemma 9.5, $\mathcal{V}^\text{ns}$ is a non-empty manifold of dimension $d$, and $(\mathcal{V} \setminus \mathcal{V}^\text{ns})_{\text{Zar}}$ has algebraic dimension strictly less than $d$. Thus $(\mathcal{V} \setminus \mathcal{V}^\text{ns})_{\text{Zar}}$ can be decomposed into finitely many irreducible components of dimension strictly less than $d$, (see Section 4.6 and 9.4 in [15]). Lemma 9.5 can then be applied to the real part of each such component.

Continuing like this, the dimension drops at each step and hence the process must terminate. This process will give us a decomposition $\mathcal{V} = \bigcup_{j=0}^d M_j$ where each $M_j$ has dimension $j$. Now, such decompositions are not unique, but basic differential geometry implies that the number $d$ is an invariant of $\mathcal{V}$. Indeed, let $\mathcal{V} = \bigcup_{j=0}^m M_j$ be another such decomposition with $M_m \neq 0$, and suppose that $m > d$. Let $\phi$ be a chart covering a patch of $M_m$, as in (2.1). $\phi$ is then defined on an open subset $U$ of $\mathbb{R}^m$, and the subsets $\phi^{-1}(M_j)$ are manifolds of dimension strictly less than $m$. Hence each one has Lebesgue measure zero, which is not compatible with that their union should equal $U$. Reversing the roles of $m$ and $d$, Proposition 6.2 follows...
in the irreducible case. Incidentally, we have also shown the first part of Proposition 6.4.

If \( \mathcal{V} \) is not irreducible, we can apply the above argument to each of its irreducible components. Since the dimension of \( \mathcal{V}_{Zar} \) is the maximum of the dimension of its components (Proposition 8, Sec. 9.6, [15]), Proposition 6.2 follows as above.

Finally, with \( A \in \mathcal{V}_{ns} \), the identity (6.3) follows by (9.2) and the implicit function theorem. This establishes the second part of Proposition 6.4 and we are done. □

**Proof of Proposition 6.3.** This is now immediate by Lemma 9.4. □

**Proof of Proposition 6.5.** \( \mathcal{V}_{Zar} \) is a strict submanifold of both \((\mathcal{V}_{1})_{Zar}\) and \((\mathcal{V}_{2})_{Zar}\), and all three are irreducible by assumption and Lemma 9.3. Hence \( \mathcal{V}_{Zar} \) has strictly lower dimension than the other two, by basic algebraic geometry, see e.g. Theorem 1, Chap I, Sec. 6, in [47]. By Lemma 9.5 the manifold \( \mathcal{V}_{ns} \) has strictly lower dimension than both \( \mathcal{V}_{1}^{ns} \) and \( \mathcal{V}_{2}^{ns} \). This means that these have a proper intersection at any intersection point \( A \in \mathcal{V}_{ns} \cap \mathcal{V}_{1}^{ns} \cap \mathcal{V}_{2}^{ns} \), which by Proposition 3.2 means that the angle at \( A \) is defined, as desired.

**Proof of Theorem 6.6.** By Proposition 6.5, all points \( A \in \mathcal{V}_{ns} \cap \mathcal{V}_{1}^{ns} \cap \mathcal{V}_{2}^{ns} \) are non-trivial, (i.e. the angle between \( \mathcal{V}_{1}^{ns} \) and \( \mathcal{V}_{2}^{ns} \) exists). We first prove the latter statement. By Proposition 6.4, (6.4) implies that

\[
\dim (N_{V_{1}}(A) + N_{V_{2}}(A)) \geq n - m.
\]

Since obviously

\[
N_{V}(A) \supset N_{V_{1}}(A) + N_{V_{2}}(A),
\]

we have \( \dim(N_{V}(A)) \geq n - m \), which combined with Lemma 9.4 and Proposition 6.2 shows that in fact we must have \( A \in \mathcal{V}_{ns} \) and \( \dim(N_{V}(A)) = n - m \). Moreover, combined with (9.3) and (9.4) this implies that \( N_{V}(A) = N_{V_{1}}(A) + N_{V_{2}}(A) \) which, upon taking the complement and recalling (6.3), yields

\[
T_{\mathcal{V}_{ns}}(A) \cap T_{\mathcal{V}_{ns}}^{1}(A) = T_{\mathcal{V}_{ns}}(A).
\]

By Proposition 3.5 we conclude that \( A \) is non-tangential, so \( A \in \mathcal{V}_{ns,nt} \). For the first statement, we note that \( \mathcal{V} \setminus (\mathcal{V}_{1}^{ns} \cap \mathcal{V}_{2}^{ns} \cap \mathcal{V}_{ns}) = (\mathcal{V} \setminus \mathcal{V}_{1}^{ns}) \cup (\mathcal{V} \setminus \mathcal{V}_{2}^{ns}) \cup (\mathcal{V} \setminus \mathcal{V}_{ns}) \) and \( \mathcal{V} \setminus \mathcal{V}_{i}^{ns} = \mathcal{V} \setminus (\mathcal{V}_{i} \setminus \mathcal{V}_{i}^{ns}) \) for \( i = 1, 2 \). Hence \( \mathcal{V} \setminus (\mathcal{V}_{1}^{ns} \cap \mathcal{V}_{2}^{ns} \cap \mathcal{V}_{ns}) \) is a real algebraic variety by Lemma 9.5 and the trivial fact that unions and intersections of varieties yield new varieties. Now, suppose \( A \in \mathcal{V}_{1}^{ns} \cap \mathcal{V}_{2}^{ns} \cap \mathcal{V}_{ns} \) is not in \( \mathcal{V}_{ns,nt} \).

Then it is tangential, which by the earlier arguments happens if and only if

\[
\dim(N_{V_{1}}(A) + N_{V_{2}}(A)) < n - m.
\]

By Hilbert’s basis theorem we can pick finite sets \( \{p_{j,1}\} \) and \( \{p_{j,2}\} \) such that \( \mathbb{I}_{R}(\mathcal{V}_{i}) \)

is generated by these sets for \( l = 1, 2 \). Let \( M \) be the matrix with the gradients of each \( \{p_{j,1}\}_{1,1}^{l} \) as columns. The condition (9.5) can then be reformulated as the vanishing of the determinants of all \( (n - m) \times (n - m) \) submatrices of \( M \). Since each such determinant is a polynomial, we conclude that \( \mathcal{V} \setminus \mathcal{V}_{ns,nt} \) is defined by the vanishing of a finite number of polynomials, so it is a real algebraic variety. Finally, if \( \mathcal{V}_{ns,nt} \) is not void, then \((\mathcal{V} \setminus \mathcal{V}_{ns,nt})_{Zar} \) is a proper subvariety of \( \mathcal{V}_{Zar} \). The latter is irreducible by Lemma 9.3, and hence \((\mathcal{V} \setminus \mathcal{V}_{ns,nt})_{Zar} \) has lower dimension than \( \mathcal{V}_{Zar} \) by standard algebraic geometry, (see e.g. Theorem 1, Ch. I, Sec. 6, [47]). □

**Proof of Proposition 6.8.** It is well-known that \( \mathcal{V} \) is irreducible if and only if \( \mathbb{I}_{R}(\mathcal{V}) \) is prime, i.e. if and only if \( fg \in \mathbb{I}_{R}(\mathcal{V}) \) implies that either \( f \in \mathbb{I}_{R}(\mathcal{V}) \) or
Clearly these can be ordered such that \( B \) is compact and other point and let \( \gamma \) be a continuous path connecting \( A \) with \( B \). The image \( \text{Im} \gamma \) is compact and \( \{ B_{r,C}(C, r_C) \}_{C \in \text{Im} \gamma} \) an open covering, where \( r_C \) is as in Definition 6.7. We pick a finite subcovering at the points \( \{ C_l \}_{l=0}^L \) with \( C_0 = A \) and \( C_L = B \). Clearly these can be ordered such that \( \mathcal{V} \cap B(C_i, r_{C_i}) \cap B(C_{i+1}, r_{C_{i+1}}) \neq \emptyset \). Also let \( i_0 \) be the index of the covering \( \phi_{i_0} \) of \( B(C_0, r_{C_0}) \) in accordance with Definition 6.7, where \( i_0 \) already has been chosen. Let \( D \in \mathcal{V} \cap B(C_0, r_{C_0}) \cap B(C_1, r_{C_1}) \) be given and let \( R \) be a radius such that \( B(D, R) \subset B(C_0, r_{C_0}) \cap B(C_1, r_{C_1}) \). By assumption, \( f \) vanishes on all points of \( \cup C \mathcal{V} \cap B(C_0, r_{C_0}) \), and hence \( f \circ \phi_{i_0} \) vanishes on the open set \( \phi_{i_0}^{-1}(B(D, R)) \), which by analyticity means that it vanishes identically on \( \Omega_{i_0} \) (since it is assumed to be connected). By induction it follows that \( f(B) = 0 \), and the first part is proved. The second is a simple consequence of continuity. We omit the details.

**Proof of Proposition 6.9.** By Proposition 6.8, \( \mathcal{V} \) is irreducible. We first show that \( \mathcal{V}^n_{ns} \) is dense in \( \mathcal{V} \). Consider the case when \( \mathcal{V} \) is covered by analytic patches, (the proof in the second case is easier and will be omitted). Since \( \mathcal{V} \setminus \mathcal{V}^{ns} \) is a nontrivial subvariety by Lemma 9.5, there exists a polynomial \( f \) which vanishes on \( \mathcal{V} \setminus \mathcal{V}^{ns} \) but not on \( \mathcal{V} \). However, if \( \mathcal{V} \setminus \mathcal{V}^{ns} \) contains an open set, then the argument in Proposition 6.8 shows that \( f \equiv 0 \) on \( \mathcal{V} \), a contradiction.

Now, let \( \theta : U \to V \) be the bijection in question, where \( U \subset \mathbb{R}^d \) and \( V \subset \mathcal{V} \) are open. Let \( m \) be the dimension of \( \mathcal{V} \) and pick any \( A \in \mathcal{V}^{ns} \cap V \). By Proposition 6.4 and (2.1), there exists open sets \( \hat{U} \subset \mathbb{R}^m \) and \( \hat{V} \subset V \) containing \( A \) and a \( C^\infty \)-bijection \( \phi : \hat{U} \to \hat{V} \). Moreover, by Proposition 2.1, \( \phi^{-1} \circ \theta \) is bijective and differentiable between the open sets \( \theta^{-1}(\hat{V}) \subset \mathbb{R}^d \) and \( \hat{U} \subset \mathbb{R}^d \). That \( m = d \) is now a well-known consequence of the implicit function theorem.

\[ \square \]

10. **Appendix C**

If we were to write out all the details, this section would get rather long. Considering that it is just an illustration, we will be a bit brief.

**Proof of Proposition 7.1.** Given a matrix \( A \in \mathcal{K} \), we can consider all elements above and on the diagonal as variables, and the remaining to be determined by \( A^T = A \). It follows that \( \mathcal{K} \) is a linear space of dimension \( n(n+1)/2 \). The statements concerning \( \mathcal{V}_1 \) are now immediate. To see that \( \mathcal{V}_2 \) is a real algebraic variety, note that \( B \) has rank greater than \( k \) if and only if one can find a non-zero \((k+1) \times (k+1)\) invertible minor (that is, a matrix obtained by deleting \( n - (k + 1) \) rows and columns). The determinant of each such minor is a polynomial, (more precisely, the determinant composed with the map that identifies \( \mathcal{K} \) with \( \mathbb{R}^{(n^2+n)/2} \), and \( \mathcal{V}_2 \) is clearly the variety obtained from the collection of such polynomials. Thus \( \mathcal{V}_2 \) is a real algebraic variety. The same is true for \( \mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2 \) since it is obtained by adding the algebraic equations \( \{ B_{j,j} = 1 \}_{j=1}^n \) to those defining \( \mathcal{V}_2 \). We now study \( \mathcal{V}_2 \). We denote by \( \mathbb{M}_{i,j} \) the set of \( i \times j \) matrices with real entries. By the spectral
variables, so we can identify the set of such matrices with $R$ patches. Let $\sigma$ be the open set corresponding to the matrices with strictly positive diagonal elements. The inverse of this identification by $B(10.2)$ consider all $U$ diagonal matrix. Then $\sigma$ be the solution whose sign coincides with $\tau$ $U$ values of $\tilde{\sigma}$ by $\tilde{\tau}$ be the set of all possible triples $\sigma, \tau, \Omega$. Define $\phi : \Omega \rightarrow \mathcal{V}_2$ by

\begin{equation}
\phi(y) = \iota(y)(\iota(y))^T.
\end{equation}

It is easy to see that $\phi$ is in bijective correspondence with an open set including $B_0$, and moreover $\phi$ is a polynomial. Thus Proposition 6.9 implies that $\mathcal{V}_2$ has dimension $\frac{2nk-k^2+k}{2}$, as desired.

We turn our attention to $\mathcal{V}$ and first prove that it can be covered with analytic patches. Let $\sigma : \{1, \ldots, n\} \rightarrow \{1, \ldots, k\}$ and $\tau : \{1, \ldots, n\} \rightarrow \{-1, 1\}$ be given and consider all $U \in M_{n,k}$ where $U_{j,\sigma_j} = x_j$ is an undetermined variable whereas all other values are fixed. Denote the $j$th row of $U$ by $U_j$, and let $\Sigma \in M_{k,k}$ be a fixed diagonal matrix. Then $U_j\Sigma U_j^T = 1$ is a quadratic equation with $x_j$ as unknown, which may have 0, 1, 2 or infinitely many real solutions. Suppose the remaining values of $U$ are such that this has two solutions for all $1 \leq j \leq n$, and fix $x_j$ to be the solution whose sign coincides with $\tau(j)$. Denote the corresponding matrix by $\tilde{U}$ and note that it is a real analytic function if we now consider the remaining values of $\tilde{U}$ as variables. These variables and the values on the diagonal of $\Sigma$ are $n(k-1)+k$ in number, and so can be identified with points $y$ in an open subset of $R^{nk-n+k}$. Let $\Omega$ be a particular connected component of this open set. Consider $\tilde{U}$ and $\Sigma$ as functions of $y$ on $\Omega$, in the obvious way, and set

\begin{equation}
\psi_{\sigma, \tau, \Omega}(y) = \tilde{U}(y)\Sigma(y)(\tilde{U}(y))^T, \quad y \in \Omega.
\end{equation}

Let $I$ be the set of all possible triples $\sigma, \tau, \Omega$. By the spectral theorem one easily sees that each $B \in \mathcal{V}$ is in the image of at least one of these maps $\psi_i$, $i \in I$. It is now not hard to see that $\{\psi_i\}_{i \in I}$ is a covering with analytic patches of $\mathcal{V}$. We wish to use Proposition 6.8 to conclude that $\mathcal{V}$ is irreducible, but first we need to show that $\mathcal{V}$ is connected. The proof gets very lengthy and technical, so we will only outline the details. The idea is that any $B \in \mathcal{V}$ is path connected with the matrix $I$ with all elements equal to 1. To see this, first note that the subset of $\mathcal{V}$ that can be represented as $(10.1)$ with all elements of $U$ non-zero, is dense in $\mathcal{V}$. Hence it suffices to find a path from such an element to $1$. Let $B$ be fixed. Now,
all values in \( \Sigma \) are not negative, for then the diagonal values of \( B \) would be as well. We can assume that the diagonal elements of \( \Sigma \) are ordered decreasingly and that \( \Sigma_{1,1} = 1 \). Pick \( \sigma \) such that \( \sigma(j) = k \) for all \( j \), and choose \( \tau \) and \( \Omega \) such that the representation (10.1) can be written in the form (10.3). Now, if the second diagonal value in \( \Sigma \) is negative, we may continuously change it until it is not, without leaving \( \Omega \). Then the values of \( y \) corresponding to the first and second column of \( U \) can be continuously moved until all elements of the first column are positive. At this point, we can reduce all values of \( U \) except the first column to zero, increasing the first value of each row whenever necessary to stay in \( \Omega \). Then we can move \( y \) so that the values in the first column become the same. Finally, we can let these values increase simultaneously until they reach 1. We have now obtained the matrix \( 1 \), and conclude that \( \mathcal{V} \) is connected and hence also irreducible, as desired.

Finally, we shall determine the dimension of \( \mathcal{V} \). Consider again the map \( \iota \) introduced earlier, with the difference that this time the last lower diagonal element in each row is not a variable, but instead determined by the other variables in that row and the constraint that it be strictly positive and that the norm of the row be 1. The number of free variables is thus \( \frac{2nk-k^2+k}{2} - n \), and the above construction thus naturally defines a real analytic map on an open subset \( \Xi \) of \( \mathbb{R}^{\frac{2nk-k^2+k}{2} - n} \). Denote this map by \( \theta \) and define \( \psi : \Xi \rightarrow \mathcal{V} \) by

\[
\psi(y) = \theta(y)(\theta(y))^T.
\]

It is not hard to see that \( \psi \) is a bijection with an open subset of \( \mathcal{V} \), so by Proposition 6.9 the dimension of \( \mathcal{V} \) is \( \frac{2nk-k^2+k}{2} - n \), as desired.

\begin{proof}[Proof of Proposition 7.2] By Proposition 6.3 and 7.1 we need to show that

\[
\dim \mathcal{V}_2(A) = (n^2 + n)/2 - (2nk - k^2 + k)/2 = (n - k + 1)(n - k)/2
\]

if and only if \( \operatorname{Rank} A = k \). By the same propositions we already have that \( \dim \mathcal{V}_2(A) \leq (n - k + 1)(n - k)/2 \), so it suffices to show that this inequality is strict when \( \operatorname{Rank} A < k \) and that the reverse inequality holds when \( \operatorname{Rank} A = k \). Based on the representation (10.1), it is easily seen that each \( A \in \mathcal{V}_2 \) can be written

\[
A = U\Sigma U^T
\]

where now both \( \Sigma \) and \( U \) lie in \( \mathbb{M}_{n,n} \), \( U \) is unitary and \( \Sigma \) is diagonal with only 0's after the \( k \)th element. In this proof the particular identification of \( \mathcal{K} \) with \( \mathbb{R}^{(n^2+n)/2} \) is important. We define \( \omega : \mathbb{R}^{(n^2+n)/2} \rightarrow \mathcal{K} \) by letting the first \( n \) entries be placed on the diagonal and the remaining ones be distributed over the upper triangular part, but multiplied by the factor \( 1/\sqrt{2} \). Finally, the lower triangular part is defined by \( A^T = A \). This identification will be implicit. For example, if \( p \) is a polynomial on \( \mathcal{K} \) then we will write \( \nabla p \) instead of the correct \( \omega(\nabla(p \circ \omega)) \). Note however that \( \nabla p \) depends on \( \omega \). Now, given a polynomial \( p \in \mathbb{I}(\mathcal{V}_2) \) and \( C \in \mathbb{M}_{n,n} \), \( q_C(x) = p\left(CT \cdot C\right) \) is clearly also in \( \mathbb{I}(\mathcal{V}_2) \). Due to the particular choice of \( \omega \), we have that \( \nabla q_C(B) = C^T \nabla p(C^T \cdot C) C \) as can be verified by direct computation. Letting \( A \) be fixed of rank \( j \leq k \), it is easy to use (10.4) to produce an invertible matrix \( C \) such that \( CT AC = I_j \), where \( I_j \in \mathcal{K} \) is the diagonal matrix whose first \( j \) diagonal values are 1 and 0 elsewhere. In particular

\[
\nabla q_C(A) = C^T \nabla p(I_j) C,
\]
which implies that \( \dim N_{V_2}(A) = \dim N_{V_2}(I_j) \). Now, all \( M_{k+1,k+1} \) subdeterminants of \( K \) form polynomials in \( \mathbb{I}(V_2) \) and their derivatives at \( I_k \) are easily computed by hand. In this way one easily gets
\[
\dim N_{V_2}(I_k) \geq (n - k + 1)(n - k)/2,
\]
which proves that any rank \( k \) element of \( V_2 \) is non-singular. Conversely, if \( j < k \), consider a fixed \( u \in \mathbb{R}^n \) as a row-vector and define the map \( \theta_u : \mathbb{R} \to V_2 \) via \( \theta_u(x) = I_j + xu^T u \). Letting \( \{e_l\}_{l=1}^n \) be the standard basis of \( \mathbb{R}^n \) and considering \( \theta_{e_l} \) as well as all differences \( \theta_{e_{l}+e_{l'}} - \theta_{e_{l'}} \), one easily sees that
\[
\text{Span} \left\{ \frac{d}{dx}\theta_u(0) : u \in \mathbb{R}^n \right\} = K
\]
and hence \( \dim N_{V_2}(I_j) = 0 \). This shows that \( I_j \) is singular, by the remarks in the beginning of the proof.

Finally, we shall establish that \( V_{n,\text{ns},\text{nt}} \) is not void. By Proposition 6.5, Theorem 6.6, Proposition 7.1 and the first part of this proposition, it suffices to show that
\[(10.5) \quad \dim(T_{V_1}(A) \cap T_{V_2}(A)) \leq \frac{2nk - k^2 + k}{2} - n \]
for some point \( A \) which has rank \( k \). We choose the point \( A = U U^T \) where \( U \in M_{n \times k} \) has a 1 on the last lower diagonal element (i.e. with index \( (j, j) \) for \( j < k \) and \( (j, k) \) for \( j \leq k \)) of each row, and zeroes elsewhere. Recall the map \( \phi \) in (10.2) and let \( y \in \mathbb{R}^{\frac{n(n-1)}{2} - \frac{(n-k+1)(n-k)}{2}} \) be such that \( U = \phi(y) \). Given \( (i, j) \) let \( E_{(i,j)} \) be the matrix with a 1 on positions \( (i, j) \) and \( (j, i) \) and zeroes elsewhere. The partial derivatives of \( \phi \) at \( y \) contain multiples of all \( E_{(i,j)} \) with \( i \leq j < k \), as well as \( n - k + 1 \) derivatives related to the last row of \( U \), which we denote by \( F_1, \ldots, F_{n-k+1} \). These are not hard to compute, but we only need the fact each \( F_i \) has precisely one non-zero diagonal value on one of the \( n-k+1 \) last elements on the diagonal, (with distinct positions for distinct \( i \)'s). Since \( T_{V_1}(A) = \text{Span} \{ E_{i,j} : i \neq j \} \), it is easily seen that
\[
\text{Span} \{ T_{V_1}(A) \cap T_{V_2}(A) \} = \text{Span} \{ E_{i,j} : i < j < k \}.
\]
These are \( \frac{n(n-1)}{2} - \frac{(n-k+1)(n-k)}{2} \) in number, and (10.5) follows.

\[\Box\]

11. ACKNOWLEDGEMENTS

This work was supported by the Swedish Research Council and the Swedish Foundation for International Cooperation in Research and Higher Education, as well as Dirección de Investigación Científica y Tecnológica del Universidad de Santiago de Chile, Chile. Part of the work was conducted while Marcus Carlsson was employed at Universidad de Santiago de Chile. We thank Arne Meurman for fruitful discussions. We also would like to thank one of the reviewers for the constructive criticism and useful suggestions that have helped to improve the quality of the paper.

REFERENCES


Centre for Mathematical Sciences, Lund University, Sweden  
E-mail address: fa@maths.lth.se

Centre for Mathematical Sciences, Lund University, Sweden  
E-mail address: marcus.carlsson@math.lu.se