MATHEMATICS OF OPERATIONS RESEARCH
Vol. 33, No. 1, February 2008, pp. 216–234
ISSN 0364-765X | EISSN 1526-5471 | 08 | 3301 | 0216
DOI 10.1287/moor.1070.0291

Alternating Projections on Manifolds

Adrian S. Lewis
School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853,
aslewis@orie.cornell.edu

Jérôme Malick
INRIA Rhone-Alpes, Montbonnot, 38334 St. Ismier Cedex, France, jerome.malick@inria.fr

We prove that if two smooth manifolds intersect transversally, then the method of alternating projections converges locally at a linear rate. We bound the speed of convergence in terms of the angle between the manifolds, which in turn we relate to the modulus of metric regularity for the intersection problem, a natural measure of conditioning. We discuss a variety of problem classes where the projections are computationally tractable, and we illustrate the method numerically on a problem of finding a low-rank solution of a matrix equation.

Key words: alternating projections; nonconvex; linear convergence; subspace angle; metric regularity; low-rank approximation; spectral set
MSC2000 subject classification: Primary: 65K10, 90C30; secondary: 49M99
OR/MS subject classification: Primary: nonlinear programming theory and algorithms
History: Received July 27, 2006; revised June 13, 2007.

1. Introduction. The method of alternating projections finds a point in the intersection of two closed convex sets by iteratively projecting a point first onto one set and then onto the other. Popular because of its simplicity and intuitive appeal, the method has been rediscovered many times in the literature. The survey article (Bauschke and Borwein [3]) covers much of the history; a careful development of the method appears in Deutsch [14]. Many practitioners have experimented with the method and its enhancements, in a wide variety of applications: typical examples are signal processing (Combettes [10]), finance (Higham [22]), and the perceptron algorithm in machine learning (see, for example, Widrow and Wallach [41]). The method extends in an obvious manner to find points in the intersection of several sets.

The attractive theory and extensive practice of alternating projections for convex feasibility makes it tempting to experiment with analogous heuristics for nonconvex feasibility problems. Inverse eigenvalue problems have been an area of particular interest in this respect (Chu [8], Chen and Chu [6]), and, in particular, pole placement (Yang and Orsi [42]) and the construction of tight frames in information theory (Tropp et al. [37]). Two very important application areas well suited to such techniques are low-order control design problems (see, for example, Grigoriadis and Skelton [20], Grigoriadis and Beran [19], Orsi et al. [31] for enhancements), and phase retrieval in image processing (see, for example, Weber and Allebach [40], Bauschke et al. [4]). For the specific problem of finding a matrix from a given affine family with a specified partial spectrum, Chen and Chu [6] prove a global descent result for a nonconvex alternating projection approach. However, existing general theory is sparse and much weaker than the convex case (Combettes and Trussell [11], Tropp et al. [37]), and has not explained some substantial practical successes with such methods. Our aim in this work is to enhance theoretical understanding of nonconvex alternating projections. We consider the simplest case, that of alternating projections onto two smooth manifolds, intersecting transversally. Locally, the manifolds can be approximated by affine subspaces and since in the case of subspaces the method of alternating projections converges linearly, one might also expect (as expressed in Orsi [30], supported by numerical evidence) linear convergence in the manifold case. Our main result is a proof of local linear convergence. For a more general and abstract study of conditions under which the method of alternating projections converges linearly, see Lewis et al. [27].

Some of the appeal of the alternating projection method for convex feasibility problems is the ease of the projection subproblem. If a closed set in a Euclidean space is convex, then any point has a unique nearest point in the set (and indeed the converse is also true, by the Motzkin-Bunt theorem; Borwein and Lewis [5]). Furthermore, providing the set is reasonably described, computing the projection is tractable computationally: modern interior point methods provide one avenue (Nesterov and Nemirovskii [28]).

By contrast, for nonconvex sets the projection mapping can no longer be single valued and may be hard to compute. Furthermore, even if two closed nonconvex sets have a nonempty intersection, very simple examples show we cannot expect alternating projections to converge in general. On the other hand, smooth manifolds belong to a large class of interesting sets that admit unique projections locally (specifically, prox-regular sets; Poliquin et al. [33]). Furthermore, for some fundamental nonconvex sets, the projection problem is computationally cheap. An obvious example is projection onto the unit sphere. More generally, projection onto any set defined by a single quadratic equation or inequality is an easy problem numerically. By arguments analogous to the theory for the classical trust region subproblem (Nocedal and Wright [29]), any optimal solution of the projection problem

$$\min_x \{\|x-a\|^2 \colon x^\top A x + b^\top x = c\}$$

associates with a (scalar) Lagrange multiplier solving the Lagrangian dual problem. This latter problem is a univariate maximization, solvable very efficiently by diagonalizing the symmetric matrix A and applying a specially designed Newton-type method. The approach for an inequality constraint is similar.

Some more complex sets have tractable projections. In particular, the paper by Tropp et al. [37] discusses projection onto the set of unit vectors satisfying an upper bound on their peak-to-average power ratio (or, equivalently, their infinity norm).

The singular-value decomposition provides some efficient and well-known nonconvex projection techniques. One example is the *orthogonal Procrustes problem*, which, for a given real rectangular matrix C, seeks the projection onto the set of matrices $\{CQ \colon Q^\top Q = I\}$: see §12.4.1 in Golub and Van Loan [18]. Another example is the problem of projecting a real rectangular matrix A onto the set of matrices with rank r or less (Horn and Johnson [24]). If A has singular value decomposition $UDV^\top$, where the matrices U and V are orthogonal, and the matrix D is nonnegative on its main diagonal and zero off it, then by replacing by zero all the main diagonal entries of D except the r largest, we obtain a nearest matrix to A (with respect to the Frobenius norm) from the set of matrices with rank no more than r. Notice that, since both the two sets above are not convex, the nearest matrices we construct in this fashion may not be unique: different singular value decompositions may correspond to different nearest matrices.

The spectral decomposition for symmetric matrices gives access to a broad range of projection techniques onto nonconvex spectral sets: that is, sets of matrices defined via inequalities of their eigenvalues. For example, given any symmetric matrix, a nearest matrix—with respect to the Frobenius norm, with given eigenvalues (and multiplicities)—is easy to compute. This observation is used in Orsi [30] in an alternating projection method to solve nonnegative inverse eigenvalue problems. Equally easy to compute is a nearest matrix from the set of matrices having largest eigenvalue multiplicity at least k. The (locally identical) set of matrices having largest eigenvalue multiplicity exactly k is a manifold, and Oustry [32] uses the corresponding projection as part of an eigenvalue optimization algorithm. We summarize general results about projections onto spectral sets of symmetric matrices in the appendix.

After outlining our notation in §2, we discuss the notion of angle between subspaces (and manifolds) in §3, a key idea both in the classical convergence theory for alternating projection on subspaces and for our extension to manifolds. We prove our main result—that the alternating projection method on transversal manifolds converges linearly locally—in §4. Just as in the classical theory for subspaces, the angle predicts the rate of linear convergence for the method. In §5, we relate this constant to a natural measure of the conditioning of the underlying feasibility problem. Finally in §6, we illustrate the theory with a numerical example, seeking a low-rank solution of a linear matrix equation. Our aim in this work is not to develop efficient numerical schemes. Indeed, even for the classical alternating projection method on subspaces many authors have observed the slow convergence of the raw method, and have experimented with enhancements. Our goal here is primarily to initiate a solid theoretical explanation for observed successes of heuristics based on nonconvex alternating projections.

2. Notation and basic results. We begin with elementary definitions. In this paper, we will consider a Euclidean space $\mathbb{E}$ (in other words, a finite-dimensional real space with inner product denoted $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$). We denote by $\mathbb{B}$ its (closed) unit ball, by $\mathbb{B}_\eta(x)$ the ball of radius η centered at the point $x \in \mathbb{E}$, and by $\mathbb{S}$ the unit sphere. A sequence $(x_k)_k$ in $\mathbb{E}$ *converges linearly with rate* $\kappa < 1$ to x if there is some constant α such that

$$\|x_k - x\| \le \alpha\kappa^k \quad \text{for } k \text{ large enough.}$$

More precisely, this inequality is *R-linear convergence* (Dennis and Schnabel [12]): the infimum of all possible constants κ, namely

$$\limsup_{k\to\infty} \|x_k - x\|^{1/k},$$

is the *rate* of R-linear convergence.

Manifolds. In this work we use only local inequalities of manifolds, so we follow as nontechnical an approach as possible. A smooth manifold in $\mathbb{E}$ is, loosely speaking, a set consisting locally of the solutions of some smooth equations. More precisely, we say (following Rockafellar and Wets [35]) that a set $\mathcal{M} \subset \mathbb{E}$ is a *C^k-manifold (of codimension d) around a point* $\bar{x} \in \mathcal{M}$ if there is an open set $U \subset \mathbb{E}$ containing $\bar{x}$ such that

$$\mathcal{M} \cap U = \{x \in U \colon F(x) = 0\},$$

where $F\colon U \to \mathbb{R}^d$ is a C^k function with surjective derivative throughout U. Note that k, the degree of smoothness of $\mathcal{M}$, will be omitted in statements if it is obvious or not useful.

We remark that, if $\mathcal{M}$ is a manifold around $\bar{x}$, then it is also a manifold around any nearby point in $\mathcal{M}$. We make some comments about the above definition. Details may be found in any text on elementary differential geometry: see for example Auslender [1]. The tangent space to $\mathcal{M}$ at $x \in \mathcal{M}$ is given by

$$\mathrm{T}_{\mathcal{M}}(x) = \ker \nabla F(x)$$

(which is actually independent of the choice of F). The normal space at $\mathcal{M}$ at x is then its orthogonal complement, namely

$$\mathrm{N}_{\mathcal{M}}(x) = \operatorname{range} \nabla F(x)^*.$$

Using the implicit function theorem (see Theorem I.1.3 in Auslender [1]), we can write down a useful equivalent definition of a manifold: a set $\mathcal{M} \subset \mathbb{E}$ is a C^k manifold (of codimension d) around a point $\bar{x} \in \mathcal{M}$ if there is an open set $U \subset \mathbb{E}$ containing $\bar{x}$, an open set $W \subset \mathbb{R}^{\dim \mathbb{E} - d}$, and a C^k function $G\colon W \to \mathbb{E}$ with injective derivative throughout W such that $G(W) = \mathcal{M} \cap U$.

Example 2.1 (Affine Manifold). Particularly easy examples of smooth manifolds are affine subspaces. If $\mathcal{M}$ is an affine subspace of $\mathbb{E}$, the equation $F(x) = 0$ can be taken to be affine: that is, of the form $\mathcal{A}(x) - b = 0$ with $\mathcal{A}\colon \mathbb{E} \to \mathbb{R}^m$ a linear map and a vector $b \in \mathbb{R}^d$. The tangent space $\ker \mathcal{A}$ is the same at any point in the affine subspace.

Example 2.2 (Fixed Rank Matrices). Let $\mathbb{E} = M_{n,m}(\mathbb{R})$ be the space of $n \times m$-matrices with the classical inner product $\langle A, B\rangle = \operatorname{trace}(A^\top B)$. Routine calculations show that the set of matrices with fixed rank r,

$$\mathcal{R}_r = \{X \in M_{n,m}(\mathbb{R}) \colon \operatorname{rank}(X) = r\},$$

is a smooth manifold around any matrix $A \in \mathcal{R}_r$. With the help of the singular value decomposition $A = UDV^\top$ (the two matrices $U = [u_1, u_2, \ldots, u_n]$ and $V = [v_1, v_2, \ldots, v_m]$ being orthogonal, and the diagonal entries in the diagonal matrix D being written in decreasing order) the tangent space at A to $\mathcal{R}_r$ is

$$\mathrm{T}_{\mathcal{R}_r}(A) = \{H \in M_{n,m}(\mathbb{R}) \colon u_i^\top H v_j = 0, \text{ for all } r < i \le n,\ r < j \le m\}.$$

This simple example will illustrate our forthcoming developments.

Let $\mathcal{M}$ and $\mathcal{N}$ be two C^k manifolds around $x \in \mathcal{M} \cap \mathcal{N}$. The classical sufficient assumption to ensure that the intersection $\mathcal{M} \cap \mathcal{N}$ is also a manifold around x is the following standard transversality assumption:

Definition 2.1 (Transversality). Suppose $\mathcal{M}$ and $\mathcal{N}$ are two C^k manifolds around a point $x \in \mathcal{M} \cap \mathcal{N}$. We say that $\mathcal{M}$ and $\mathcal{N}$ are *transverse* at x if

$$\mathrm{T}_{\mathcal{M}}(x) + \mathrm{T}_{\mathcal{N}}(x) = \mathbb{E}.$$

In this case, the intersection $\mathcal{M} \cap \mathcal{N}$ is a C^k manifold around x and there holds $\mathrm{T}_{\mathcal{M}\cap\mathcal{N}}(x) = \mathrm{T}_{\mathcal{M}}(x) \cap \mathrm{T}_{\mathcal{N}}(x)$.

Projections. The projection of an element $x \in \mathbb{E}$ onto a subset $M \subset \mathbb{E}$ is defined by

$$\mathrm{P}_M(x) := \operatorname{arg\,min}\{\|x - y\| \colon y \in M\}.$$

This set is nonempty providing M is nonempty and closed, and has at most one element if M is convex. Indeed, M is nonempty, closed, and convex if and only if the projector operator $\mathrm{P}_M\colon \mathbb{E} \to \mathbb{E}$ is everywhere defined and single valued. If furthermore the boundary of the closed convex M is a C^k manifold, the projection mapping P_M is C^{k-1} (Holmes [23]).

Despite the importance of convexity, many nonconvex sets have associated projections that are well behaved in some weaker sense. The following lemma and example illustrate this, and further examples appear in the appendix. The lemma is not new, but since it is a basic tool for us we discuss its proof in some detail.

Lemma 2.1 (Projection Onto a Manifold). *Let $\mathcal{M} \subset \mathbb{E}$ be a manifold of class C^k (with $k \geq 2$) around a point $\bar{x} \in \mathcal{M}$. Then each point x near $\bar{x}$ has a unique projection $\mathrm{P}_{\mathcal{M}}(x)$ on $\mathcal{M}$. Furthermore, the function $\mathrm{P}_{\mathcal{M}}$ is of class C^{k-1} around $\bar{x}$, with derivative*

$$\nabla \mathrm{P}_{\mathcal{M}}(\bar{x}) = \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x})}.$$

Proof. To see that the projection $\mathrm{P}_{\mathcal{M}}$ is single valued and continuous around the point $\bar{x}$, it suffices to note that $\mathcal{M}$ is prox-regular at $\bar{x}$ by Proposition 2.3 in Poliquin et al. [33], and then apply Theorem 1.3 in Poliquin et al. [33]. For a step-by-step proof, see §9.3 in Borwein and Lewis [5].

The technique we use for the remainder of the proof is standard: see, for example, Robinson [34]. In abstract language, given a point $y \in \mathbb{E}$ close to the point $\bar{x}$, we compare the solution $x \in \mathbb{E}$ of the generalized equation

$$y - x \in N_{\mathcal{M}}(x) \qquad \text{and} \qquad x \in \mathcal{M}$$

to the solution of a linearized version. For completeness, we give an elementary development.

Using the definition of a manifold, there is an open set $U \subset \mathbb{E}$ containing $\bar{x}$ such that

$$\mathcal{M} \cap U = \{x \in U \colon F(x) = 0\},$$

where $F\colon U \to \mathbb{R}^d$ is a C^k function with surjective derivative throughout U. By choosing a smaller neighborhood U if necessary, we can assume the corresponding restriction of the projection $\mathrm{P}_{\mathcal{M}}\colon U \to \mathcal{M}$ is everywhere single valued.

Now define a C^{k-1} function $G\colon U \times \mathbb{R}^d \to \mathbb{E} \times \mathbb{R}^d$ by

$$G(x, z) = (x + \nabla F(x)^* z, F(x)).$$

An easy calculation shows that the derivative $\nabla G(\bar{x}, 0)\colon \mathbb{E} \times \mathbb{R}^d \to \mathbb{E} \times \mathbb{R}^d$ is given by

$$\nabla G(\bar{x}, 0)(w, z) = (w + \nabla F(\bar{x})^* z, \nabla F(\bar{x}) w).$$

If $(w, z) \in \ker \nabla G(\bar{x}, 0)$, we deduce $\nabla F(\bar{x}) \nabla F(\bar{x})^* z = 0$, and then taking the inner product with z shows $\nabla F(\bar{x})^* z = 0$. Since $\nabla F(\bar{x})^*$ is injective, $z = 0$, and consequently $w = 0$. We have therefore shown that linear map $\nabla G(\bar{x}, 0)$ is invertible.

The inverse function theorem now applies to show that the map G has a C^{k-1} inverse $G^{-1}\colon V \times W \to U \times \mathbb{R}^d$, where the open set $V \subset \mathbb{E}$ contains the point $\bar{x}$ and the open set $W \subset \mathbb{R}^d$ contains zero. Specifically, the map $G^{-1} \circ G$ is the identity on a set $Q \times S$, where the open set $Q \subset U$ contains $\bar{x}$ and the open set $S \subset \mathbb{R}^d$ contains zero. Furthermore, we have

$$\nabla (G^{-1})(\bar{x}, 0) = \nabla G(\bar{x}, 0)^{-1}.$$

Every point $y \in U$ satisfies

$$y - \mathrm{P}_{\mathcal{M}}(y) \in \mathrm{N}_{\mathcal{M}}(\mathrm{P}_{\mathcal{M}}(y)) = \operatorname{range} \nabla F(\mathrm{P}_{\mathcal{M}}(y))^*.$$

Hence for any sequence $y_j \to \bar{x}$ in $\mathbb{E}$, there exists a sequence $z_j \in \mathbb{R}^d$ satisfying

$$y_j - \mathrm{P}_{\mathcal{M}}(y_j) = \nabla F(\mathrm{P}_{\mathcal{M}}(y_j))^* z_j$$

for all large j. By continuity, $\mathrm{P}_{\mathcal{M}}(y_j) \to \bar{x}$ and $\nabla F(\mathrm{P}_{\mathcal{M}}(y_j)) \to \nabla F(\bar{x})$. The sequence $(z_j)_j$ must be bounded. Otherwise, by restricting to a subsequence, we can suppose $\|z_j\| \to \infty$ and $\|z_j\|^{-1} z_j$ approaches a unit vector u, which must lie in $\ker \nabla F(\bar{x})^*$, contradicting the fact that $\nabla F(\bar{x})^*$ is injective. We claim $z_j \to 0$. If this fails, again by restricting to a subsequence, we can suppose z_j approaches a nonzero vector z, which again must lie in $\ker \nabla F(\bar{x})^*$, giving another contradiction.

Using the above argument, we deduce the following inequality: All points $y \in \mathbb{E}$ close to the point $\bar{x}$ satisfy $y \in V$ and $\mathrm{P}_{\mathcal{M}}(y) \in Q$, and furthermore there exists a vector $z \in S$ such that

$$G(\mathrm{P}_{\mathcal{M}}(y), z) = (\mathrm{P}_{\mathcal{M}}(y) + \nabla F(\mathrm{P}_{\mathcal{M}}(y))^* z, F(\mathrm{P}_{\mathcal{M}}(y))) = (y, 0).$$

Applying the inverse map G^{-1} to both sides then shows $(\mathrm{P}_{\mathcal{M}}(y), z) = G^{-1}(y, 0)$, so

$$\mathrm{P}_{\mathcal{M}}(y) = \mathrm{P}_{\mathbb{E}}(G^{-1}(y, 0)),$$

where $P_{\mathbb{E}}\colon \mathbb{E}\times\mathbb{R}^d \to \mathbb{E}$ is the canonical projection. Hence, on a neighborhood of $\bar{x}$, we have

$$P_{\mathcal{M}} = P_{\mathbb{E}} \circ G^{-1} \circ P_{\mathbb{E}}^*, \tag{1}$$

where the adjoint $P_{\mathbb{E}}^*\colon \mathbb{E}\to\mathbb{E}\times\mathbb{R}^d$ is just the embedding $y\mapsto(y,0)$. Since both $P_{\mathbb{E}}$ and its adjoint are linear, we deduce that $P_{\mathcal{M}}$ is C^{k-1} around $\bar{x}$, and furthermore

$$\nabla P_{\mathcal{M}}(\bar{x}) = P_{\mathbb{E}} \circ (\nabla(G^{-1})(\bar{x},0)) \circ P_{\mathbb{E}}^* = P_{\mathbb{E}} \circ \nabla G(\bar{x},0)^{-1} \circ P_{\mathbb{E}}^*. \tag{2}$$

We will conclude by proving that the right-hand side of (2) is locally the projection onto the tangent space. For any $y\in V$, set

$$w = (P_{\mathbb{E}} \circ \nabla G(\bar{x},0)^{-1} \circ P_{\mathbb{E}}^*)(y).$$

By construction, there exists $z\in\mathbb{R}^d$ such that

$$\nabla G(\bar{x},0)(w,z) = (y,0);$$

hence $w + \nabla F(\bar{x})^* z = y$ and $\nabla F(\bar{x})w = 0$. This results in

$$w = y - \nabla F(\bar{x})^*(\nabla F(\bar{x})\nabla F(\bar{x})^*)^{-1}\nabla F(\bar{x})y = P_{\ker \nabla F(\bar{x})} y = P_{T_{\mathcal{M}}(\bar{x})} y,$$

which finally yields

$$P_{T_{\mathcal{M}}(\bar{x})} = P_{\mathbb{E}} \circ \nabla G(\bar{x},0)^{-1} \circ P_{\mathbb{E}}^*.$$

Combining this with Equation (2) proves the final claim. □

Example 2.3 (Projection Onto Fixed Rank Matrices). The singular value decomposition gives an easy way to project any $n\times m$ matrix X onto the set of matrices with rank no more than r. Specifically, given any singular value decomposition $X = UDV^\top$, a nearest matrix with rank no more than r is

$$\widehat{X} = \sum_{i=1}^{r} \sigma_i u_i v_i^\top,$$

where the σ_i are the r largest singular values (see Horn and Johnson [24]). We remark that this nearest matrix may not be unique: different singular value decompositions may result in different nearest matrices.

The set of matrices with rank no more than r is not a manifold. However, locally this same technique allows us to project onto the manifold $\mathcal{R}_r$ we considered in Example 2.2, of matrices of rank exactly r. To see this, consider a matrix $\overline{X}\in\mathcal{R}_r$. Denote the distance from $\overline{X}$ to the set of matrices with rank strictly less than r by δ. The result in the preceding paragraph, with r replaced by $r-1$ and X replaced by $\overline{X}$, guarantees $\delta>0$. Now suppose the matrix X considered above satisfies $\|X-\overline{X}\|<\delta/2$. Since $\overline{X}$ has rank r, and $\widehat{X}$ is a nearest matrix with rank no more than r, we must have $\|\widehat{X}-X\| \le \|\overline{X}-X\| < \delta/2$. Hence by the triangle inequality $\|\widehat{X}-\overline{X}\|<\delta$, so $\widehat{X}$ has rank at least r, and therefore exactly r. Thus in fact $\widehat{X}$ is a nearest matrix to X in the manifold $\mathcal{R}_r$.

3. Angles between subspaces.

Let M and N be two subspaces of $\mathbb{E}$. Following Friedricks [17] and Deutsch [14], we define the *angle* between M and N as the angle between 0 and $\pi/2$ whose cosine is

$$\begin{aligned} c(M,N) := \max\{\langle x,y\rangle\colon\ & x\in\mathbb{S}\cap M\cap(M\cap N)^\perp \\ & y\in\mathbb{S}\cap N\cap(M\cap N)^\perp\}. \end{aligned} \tag{3}$$

The quantity $c(M,N)$ is well defined unless one subspace is a subspace of the other, in which case we set $c(M,N)=0$. The compactness ensures that the maximum is always attained, and this yields easily $c(M,N)<1$. Note also that there holds (see Deutsch [14, 9.5])

$$\|P_M P_N - P_{M\cap N}\| = c(M,N), \tag{4}$$

and more generally (Deutsch [14, Theorem 9.31]),

$$\|(P_M P_N)^n - P_{M\cap N}\| = c(M,N)^{2n-1} \tag{5}$$

for $n=1,2,\ldots$. Computation of the angle between two subspaces using the singular value decomposition is discussed in §12.4 of Golub and Van Loan [18].

3.1. Inequalities of the angle between two subspaces. We begin by developing some basic inequalities of the angle, useful in our later discussion of metric regularity. We start with a technical tool.

LEMMA 3.1. *Let M and N be two subspaces of the space $\mathbb{E}$. Consider two vectors $m \in \mathbb{S} \cap M \cap (M \cap N)^\perp$ and $n \in \mathbb{S} \cap N \cap (M \cap N)^\perp$ such that $c(M, N) = \langle m, n\rangle$. Then*

$$\mathrm{P}_M(n) = \mathrm{P}_{M\cap(M\cap N)^\perp}(n) = c(M, N)m. \tag{6}$$

PROOF. Consider first the decomposition

$$n = \mathrm{P}_M(n) + \mathrm{P}_{M^\perp}(n). \tag{7}$$

Observe from

$$n \in (M \cap N)^\perp \qquad \text{and} \qquad n - \mathrm{P}_M(n) = \mathrm{P}_{M^\perp}(n) \in M^\perp \subset (M \cap N)^\perp$$

that we have $\mathrm{P}_M(n) \in M \cap (M \cap N)^\perp$, and that we also have

$$\mathrm{P}_{M^\perp}(n) \in M^\perp \subset (M \cap (M \cap N)^\perp)^\perp.$$

Thus Equation (7) gives $\mathrm{P}_M(n) = \mathrm{P}_{M\cap(M\cap N)^\perp}(n)$. We consider now two cases. First, if $\mathrm{P}_M(n) = 0$, then $n \in M^\perp$, which yields $\langle n, m\rangle = 0$. Thus (6) holds. Second, if $\mathrm{P}_M(n) \neq 0$, then the minimum

$$\min_{x\in\mathbb{S}\cap M\cap(M\cap N)^\perp} \tfrac{1}{2}\|n - x\|^2$$

is attained at a unique point, namely $\mathrm{P}_M(n)/\|\mathrm{P}_M(n)\|$. Observe that this minimum can be rewritten

$$\min_{x\in\mathbb{S}\cap M\cap(M\cap N)^\perp} \left(\tfrac{1}{2}(\|n\|^2 + \|x\|^2) - \langle x, n\rangle\right) = 1 - \max_{x\in\mathbb{S}\cap M\cap(M\cap N)^\perp} \langle x, n\rangle.$$

By the choice of n and m, we thus have $m = \mathrm{P}_M(n)/\|\mathrm{P}_M(n)\|$. Hence

$$\|\mathrm{P}_M(n)\| = \langle m, \mathrm{P}_M(n)\rangle = \langle m, n\rangle,$$

and then (6) holds. The proof is complete. □

This result permits us to prove that the angle between two subspaces is equal to the angle between their orthogonal complements, as stated in the next lemma. A proof in the Hilbert space case appears in Deutsch [13], apparently for the first time in print, although the observation is attributed there to earlier work. We include a short finite-dimensional proof here.

LEMMA 3.2. *Let M and N be two subspaces of $\mathbb{E}$. Then*

$$c(M, N) = c(M^\perp, N^\perp).$$

PROOF. If one subspace is a subspace of the other, the result is immediate. Otherwise, denote $c = c(M, N)$, and consider vectors $m \in \mathbb{S} \cap M \cap (M \cap N)^\perp$ and $n \in \mathbb{S} \cap N \cap (M \cap N)^\perp$ such that $\langle m, n\rangle = c$. Then consider the two following vectors:

$$\bar{m} = \alpha(n - cm) \qquad \text{and} \qquad \bar{n} = \alpha(cn - m), \quad \text{with } \alpha = 1/\sqrt{1 - c^2}.$$

Let us check that $\bar{m} \in \mathbb{S} \cap M^\perp \cap (M^\perp \cap N^\perp)^\perp$. First, by definition, we have $\bar{m} \in N + M = (M^\perp \cap N^\perp)^\perp$. Second, Lemma 3.1 shows $\bar{m} = \alpha \mathrm{P}_{M^\perp}(n) \in M^\perp$. Third, we obtain

$$\|\bar{m}\|^2 = \alpha^2(\|n\|^2 + c^2\|m\|^2 - 2c\langle m, n\rangle) = 1.$$

Similarly, we obtain $\bar{n} \in \mathbb{S} \cap N^\perp \cap (M^\perp \cap N^\perp)^\perp$. Finally, we observe that

$$\langle \bar{m}, \bar{n}\rangle = \alpha^2\langle n - cm, cn - m\rangle = c\alpha^2\langle n - cm, n\rangle = c.$$

Thus by definition of the Angle (3), we see that

$$c(M^\perp, N^\perp) \geq \langle \bar{m}, \bar{n}\rangle = c = c(M, N).$$

Changing the roles of M and N with their orthogonal complements, we get the reverse inequality $c(M, N) \geq c(M^\perp, N^\perp)$ with the same argument. □

We also state the following technical result that will be useful afterward:

LEMMA 3.3. *Let M and N be two nontrivial subspaces of $\mathbb{E}$ such that $M \cap N = \{0\}$. Then*

$$1 - c(M, N) = \min_{x \in \mathbb{S},\, m \in M,\, n \in N} (\|x - m\|^2 + \|x - n\|^2).$$

PROOF. Denote by R the right-hand side of the equality to be proved. Developing the sum of squared norms, we get

$$\begin{aligned} R - 2 &= \min_{m \in M,\, n \in N} \Big(\|n\|^2 + \|m\|^2 - 2 \max_{x \in \mathbb{S}} \langle x, m + n \rangle \Big) \\ &= \min_{m \in M,\, n \in N} (\|n\|^2 + \|m\|^2 - 2\|m + n\|) \\ &= \min_{m \in M \cap \mathbb{S},\, n \in N \cap \mathbb{S}} \; \min_{\alpha,\, \beta \in \mathbb{R}} (\alpha^2 + \beta^2 - 2\|\alpha m + \beta n\|). \end{aligned}$$

Observe now that the function $f_{m,n}(\alpha, \beta) = \alpha^2 + \beta^2 - 2\|\alpha m + \beta n\|$ has compact lower-level sets, and is smooth on $\mathbb{R}^2 \backslash \{(0, 0)\}$ (since $M \cap N = \{0\}$), and that it has a local maximum at $(0, 0)$. Hence $f_{m,n}$ achieves its minimum at a critical point. Some algebra then gives

$$\min_{\alpha,\, \beta \in \mathbb{R}} f_{m,n}(\alpha, \beta) = -1 - |\langle m, n \rangle|.$$

So minimizing with respect to m and n gives

$$R - 2 = \min_{m \in M \cap \mathbb{S},\, n \in N \cap \mathbb{S}} (-1 - |\langle m, n \rangle|),$$

and changing signs if necessary (remember M and N are subspaces) yields

$$R - 2 = -1 - \max_{m \in M \cap \mathbb{S},\, n \in N \cap \mathbb{S}} \langle m, n \rangle = -1 - c(M, N),$$

the last equality holding again because of the assumption $M \cap N = \{0\}$. Finally, we thus get $R = 2 + (-1 - c(M, N)) = 1 - c(M, N)$ which is the targeted equality. □

3.2. Angle between two manifolds. Let us now generalize the previous framework: in view of the definition of the *angle* between two subspaces, the following definition makes sense:

DEFINITION 3.1 (ANGLE BETWEEN TWO MANIFOLDS). Let $\mathcal{M}$ and $\mathcal{N}$ be two manifolds in $\mathbb{E}$ around a point $x \in \mathcal{M} \cap \mathcal{N}$. The angle between $\mathcal{M}$ and $\mathcal{N}$ at x is the angle between the tangent spaces $\mathrm{T}_{\mathcal{M}}(x)$ and $\mathrm{T}_{\mathcal{N}}(x)$. In other words, it is the angle between 0 and $\pi/2$ with cosine

$$c(\mathcal{M}, \mathcal{N}, x) := c(\mathrm{T}_{\mathcal{M}}(x), \mathrm{T}_{\mathcal{N}}(x)).$$

If $\mathcal{M}$ and $\mathcal{N}$ are actually subspaces, it is clear that the angle between them does not depend on the point in their intersection, and that the two definitions coincide. Let us add a lemma that formalizes an obvious smoothness inequality.

LEMMA 3.4 (SMOOTHNESS OF THE ANGLE). *Let $\mathcal{M}$ and $\mathcal{N}$ be two transverse C^k manifolds in the space $\mathbb{E}$ (with $k \geq 2$) around the point $\bar{x} \in \mathcal{M} \cap \mathcal{N}$. Then the function*

$$c(\mathcal{M}, \mathcal{N}, \cdot)\colon \begin{cases} \mathcal{M} \cap \mathcal{N} \to [0, 1] \\ x \mapsto c(\mathcal{M}, \mathcal{N}, x) \end{cases}$$

is of class C^{k-1} around $\bar{x}$.

PROOF. From inequality (4) and Definition 3.1, we know for any point $x \in \mathcal{M} \cap \mathcal{N}$ that

$$c(\mathcal{M}, \mathcal{N}, x) = \|\mathrm{P}_{\mathrm{T}_{\mathcal{M}}(x)} \mathrm{P}_{\mathrm{T}_{\mathcal{N}}(x)} - \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(x) \cap \mathrm{T}_{\mathcal{N}}(x)}\|. \tag{8}$$

Moreover, the projectors $x \mapsto \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(x)}$ are C^{k-1}: the columns of the derivative of a local C^k parametrization of the manifold form a basis for the tangent space that is a C^{k-1} function of the base point, so the projectors (expressed with this basis) are also C^{k-1}. Inequality (8) now proves the result. □

4. Alternating projections onto manifolds. We are now ready to consider the alternating projection algorithm. We consider two manifolds $\mathcal{M}$ and $\mathcal{N}$ in the space $\mathbb{E}$, and study the *alternating projection sequence* defined iteratively as follows:

$$x_0 \in \mathbb{E} \text{ given}, \qquad x_{k+1} = \mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_k). \tag{9}$$

When $\mathcal{M}$ and $\mathcal{N}$ are actually affine subspaces, this algorithm is well defined and converges (von Neumann [39]), and its behavior is well understood (see Bauschke and Borwein [2]). In particular, we have the following theorem (see Smith et al. [36], Bauschke and Borwein [2, 4.11]):

Theorem 4.1 (Alternating Projections for Two Affine Subspaces). *Let M and N be two affine subspaces of the space $\mathbb{E}$. Then the alternating Projection Sequence (9) converges linearly with rate the cosine of the angle between the two subspaces, $c(M, N)$, independent of the starting point.*

When $\mathcal{M}$ and $\mathcal{N}$ are general smooth manifolds, we will see in Theorem 4.3 that Sequence (9) is also well defined in a neighborhood of a point $\bar{x}$ belonging to the intersection $\mathcal{M} \cap \mathcal{N}$ (assuming transversality), and that the previous convergence result generalizes. The next result gives the main tool.

Theorem 4.2 (Asymptotical Improvement). *Let $\mathcal{M}$ and $\mathcal{N}$ be two transverse C^2 manifolds around a point $\bar{x} \in \mathcal{M} \cap \mathcal{N}$. Then*

$$\limsup_{x \to \bar{x},\, x \notin \mathcal{M}\cap\mathcal{N}} \frac{\|\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x) - \mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|}{\|x - \mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|} \;\le\; c(\mathcal{M}, \mathcal{N}, \bar{x}).$$

Proof. Lemma 2.1 implies that there exists $\delta > 0$ such that the projection operators $\mathrm{P}_{\mathcal{M}}$, $\mathrm{P}_{\mathcal{N}}$, and $\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}$ are well defined and of class C^1, on the ball $B_\delta(\bar{x})$. Restricting further to points $x \in B_{\delta/2}(\bar{x})$, we have

$$\|\bar{x} - \mathrm{P}_{\mathcal{N}}(x)\| \le \|\bar{x} - x\| + \|x - \mathrm{P}_{\mathcal{N}}(x)\| \le 2\|x - \bar{x}\| \le \delta,$$

so $\mathrm{P}_{\mathcal{N}}(x) \in B_\delta(\bar{x})$, and therefore $\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}$ is also well defined and C^1 on $B_{\delta/2}(\bar{x})$. We thus ensure that the fraction in the result makes sense.

Let $(x_r)_r$ be an arbitrary sequence of points in $B_{\delta/2}(\bar{x}) \backslash (\mathcal{M} \cap \mathcal{N})$ tending to $\bar{x}$. To simplify notation, we use $\bar{x}_r = \mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x_r)$. Of course $\bar{x}_r \in \mathcal{M} \cap \mathcal{N}$, so

$$\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_r) - \bar{x}_r = \mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_r) - \mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(\bar{x}_r).$$

Observe also that the continuity of $\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}$ yields that $\bar{x}_r$ tends to $\bar{x}$, too, so the previous equation and continuous differentiability shows

$$\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_r) - \bar{x}_r = \nabla(\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}})(\bar{x}_r)(x_r - \bar{x}_r) + o(\|x_r - \bar{x}_r\|). \tag{10}$$

Using Lemma 2.1 and the chain rule we get

$$\nabla(\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}})(\bar{x}_r) = \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)}\mathrm{P}_{\mathrm{T}_{\mathcal{N}}(\bar{x}_r)}. \tag{11}$$

The transversality assumption now shows

$$\mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)\cap\mathrm{T}_{\mathcal{N}}(\bar{x}_r)}(x_r - \bar{x}_r) = \mathrm{P}_{\mathrm{T}_{\mathcal{M}\cap\mathcal{N}}(\bar{x}_r)}(x_r - \bar{x}_r) = 0,$$

since $x_r - \bar{x}_r \in \mathrm{N}_{\mathcal{M}\cap\mathcal{N}}(\bar{x}_r) = \mathrm{T}_{\mathcal{M}\cap\mathcal{N}}(\bar{x}_r)^\perp$. So we can write

$$\mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)}\mathrm{P}_{\mathrm{T}_{\mathcal{N}}(\bar{x}_r)}(x_r - \bar{x}_r) = \big(\mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)}\mathrm{P}_{\mathrm{T}_{\mathcal{N}}(\bar{x}_r)} - \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)\cap\mathrm{T}_{\mathcal{N}}(\bar{x}_r)}\big)(x_r - \bar{x}_r).$$

Combined with Equations (10) and (11), this gives

$$\frac{\|\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_r) - \bar{x}_r\|}{\|x_r - \bar{x}_r\|} \le \|\mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)}\mathrm{P}_{\mathrm{T}_{\mathcal{N}}(\bar{x}_r)} - \mathrm{P}_{\mathrm{T}_{\mathcal{M}}(\bar{x}_r)\cap\mathrm{T}_{\mathcal{N}}(\bar{x}_r)}\| + o(1),$$

that is,

$$\frac{\|\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_r) - \bar{x}_r\|}{\|x_r - \bar{x}_r\|} \le c(\mathcal{M}, \mathcal{N}, \bar{x}_r) + o(1),$$

by Equation (4) and Definition 3.1. Taking the lim sup in this inequality, the result now follows by Lemma 3.4. □

A refinement of the above argument, using Equation (5) in place of Equation (4), shows the generalization

$$\limsup_{x\to\bar{x},\, x\notin\mathcal{M}\cap\mathcal{N}} \frac{\|(\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}})^n(x)-\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|}{\|x-\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|} \;\le\; c(\mathcal{M},\mathcal{N},\bar{x})^{2n-1} \tag{12}$$

for $n=1,2,\ldots$.

Observe that, with the hypotheses of the above theorem, we have that, for all constants $c>c(\mathcal{M},\mathcal{N},\bar{x})$, there exists a radius $\eta>0$ such that

$$\forall\, x\in B_\eta(\bar{x}),\qquad \|\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x)-\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|\le c\|x-\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x)\|. \tag{13}$$

We can now prove our main result.

Theorem 4.3 (Linear Convergence). *In the space $\mathbb{E}$, let $\mathcal{M}$ and $\mathcal{N}$ be two transverse manifolds around a point $\bar{x}\in\mathcal{M}\cap\mathcal{N}$. If the initial point $x_0\in\mathbb{E}$ is close to $\bar{x}$, then the method of alternating projections*

$$x_{k+1}=\mathrm{P}_{\mathcal{M}}\mathrm{P}_{\mathcal{N}}(x_k)\quad (k=0,1,2,\ldots)$$

is well defined, and the distance $d_{\mathcal{M}\cap\mathcal{N}}(x_k)$ from the iterate x_k to the intersection $\mathcal{M}\cap\mathcal{N}$ decreases Q-linearly to zero. More precisely, given any constant c strictly larger than the cosine of the angle of intersection between the manifolds, $c(\mathcal{M},\mathcal{N},\bar{x})$, if x_0 is close to $\bar{x}$, then the iterates satisfy

$$d_{\mathcal{M}\cap\mathcal{N}}(x_{k+1})\le c\cdot d_{\mathcal{M}\cap\mathcal{N}}(x_k)\quad (k=0,1,2,\ldots). \tag{14}$$

Furthermore, x_k converges linearly to some point $x^*\in\mathcal{M}\cap\mathcal{N}$: for some constant $\alpha>0$,

$$\|x_k-x^*\|\le\alpha c^k\quad (k=0,1,2,\ldots). \tag{15}$$

Proof. Choose c such that $1>c>c(\mathcal{M},\mathcal{N},\bar{x})$ and $\eta>0$ such that (13) is satisfied. Set $\delta:=(1-c)\eta/4$ and choose any starting point $x_0\in B_\delta(\bar{x})$.

First step: *inequalities of x_k*. Let us prove by induction that the sequence of points x_k is well defined, and that both x_k and its projection $\bar{x}_k=\mathrm{P}_{\mathcal{M}\cap\mathcal{N}}(x_k)$ belong to the neighborhood $B_\eta(\bar{x})$ and satisfy the inequalities

$$\begin{aligned}
\|x_k-\bar{x}_{k-1}\| &\le \delta c^k, && \text{(H1)}\\
\|x_k-\bar{x}_k\| &\le \delta c^k, && \text{(H2)}\\
\|\bar{x}_k-\bar{x}_{k-1}\| &\le 2\delta c^k, && \text{(H3)}\\
\|\bar{x}_k-\bar{x}\| &\le 2\Big(\sum_{i=0}^k c^i\Big)\delta, && \text{(H4)}\\
\|x_k-\bar{x}\| &\le 2\Big(\sum_{i=0}^k c^i\Big)\delta. && \text{(H5)}
\end{aligned}$$

Setting $\bar{x}_{-1}=\bar{x}_0$ and using

$$\|x_0-\bar{x}_0\|\le\|x_0-\bar{x}\|\le\delta,$$

it is easy to see that inequalities (H1)–(H5) hold for $k=0$. Assume now that these inequalities hold for some $k\ge 0$: we prove they also hold with k replaced by $k+1$. Note that if x_k belongs to $\mathcal{M}\cap\mathcal{N}$, there is nothing to prove. Otherwise, since x_k belongs to $B_\eta(\bar{x})$, the next iterate x_{k+1} is well defined, and inequality (13) holds, so

$$d_{\mathcal{M}\cap\mathcal{N}}(x_{k+1})\le\|x_{k+1}-\bar{x}_k\|\le c\|x_k-\bar{x}_k\|=c\cdot d_{\mathcal{M}\cap\mathcal{N}}(x_k).$$

(H1) With the help of inequality (H2), the above inequality yields

$$\|x_{k+1}-\bar{x}_k\|\le\delta c^{k+1}. \tag{16}$$

(H2) Note that $\|x_{k+1}-\bar{x}_{k+1}\|\le\|x_{k+1}-\bar{x}_k\|$ by definition of $\bar{x}_{k+1}$. With inequality (16), this implies

$$\|x_{k+1}-\bar{x}_{k+1}\|\le\delta c^{k+1}. \tag{17}$$

(H3) We get inequality (H3) from inequalities (16) and (17) by observing

$$\|\bar{x}_{k+1}-\bar{x}_k\|\le\|\bar{x}_{k+1}-x_{k+1}\|+\|x_{k+1}-\bar{x}_k\|\le 2\delta c^{k+1}. \tag{18}$$

(H4) Finally, note that

$$\|\bar{x}_{k+1}-\bar{x}\| \le \|\bar{x}_{k+1}-\bar{x}_k\| + \|\bar{x}_k-\bar{x}\|,$$

so inequalities (18) and (H4) enable us to write

$$\|\bar{x}_{k+1}-\bar{x}\| \le 2\delta\ c^{k+1} + 2\delta\sum_{i=0}^{k} c^i \le 2\delta\sum_{i=0}^{k+1} c^i. \tag{19}$$

(H5) Similarly,

$$\|x_{k+1}-\bar{x}\| \le \|x_{k+1}-\bar{x}_k\| + \|\bar{x}_k-\bar{x}\|,$$

so we have from inequalities (16) and (H4)

$$\|x_{k+1}-\bar{x}\| \le \delta\ c^{k+1} + 2\delta\sum_{i=0}^{k} c^i \le 2\delta\sum_{i=0}^{k+1} c^i. \tag{20}$$

Observe now that inequality (19) yields

$$\|\bar{x}_{k+1}-\bar{x}\| \le 2\delta/(1-c) \le \eta/2,$$

and inequality (20) yields

$$\|x_{k+1}-\bar{x}\| \le \eta/2.$$

So $\bar{x}_{k+1}$ and x_{k+1} belong to $B_\eta(\bar{x})$ too. This ends the proof by induction.

Second step: *convergence.* We first prove the convergence of the sequence of projections $(\bar{x}_k)_k$: this sequence in $\mathcal{M}\cap\mathcal{N}\cap B_\eta(\bar{x})$ is Cauchy. To see this, use inequality (H3) to write, for all indices $k, p \ge 0$ with $p \ge k$,

$$\|\bar{x}_p-\bar{x}_k\| \le \sum_{i=k+1}^{p} \|\bar{x}_i-\bar{x}_{i-1}\| \le 2\delta \sum_{i=k+1}^{p} c^i \le \frac{2\delta}{1-c} c^{k+1}. \tag{21}$$

So $(\bar{x}_k)_k$ converges to an element x^* in $\mathcal{M}\cap\mathcal{N}$. Passing to the limit in p in inequality (21), we obtain

$$\|\bar{x}_k - x^*\| \le \frac{2\,\delta}{1-c}\, c^{k+1}.$$

With the help of inequality (H2), this implies

$$\|x_k - x^*\| \le \|x_k-\bar{x}_k\| + \|\bar{x}_k - x^*\| \le \left(1+\frac{2c}{1-c}\right)\delta\, c^k,$$

which yields inequality (15) and completes the proof. □

REMARK 4.1 (STRONGER BOUND). In fact, the distance d_k from the iterate x_k to the intersection of the two manifolds $\mathcal{M}\cap\mathcal{N}$ decreases to zero with R-linear rate $(c(\mathcal{M},\mathcal{N},x^*))^2$, a faster rate than predicted by inequality (14). To see this refinement, we argue as follows:

Fix any constant c in the interval $(c(\mathcal{M},\mathcal{N},x^*),1)$, and any integer $n>0$. We claim

$$\limsup_r d_r^{1/r} \le c^{2-1/n}. \tag{22}$$

Our result then follows by taking the infimum over c and n.

To verify the claim, note first that Theorem 4.3 and inequality (12) guarantee that there is an integer t_0 such that $d_{t+n} < c^{2n-1} d_t$ for all integers $t > t_0$, and hence by induction

$$d_{t+kn} < c^{k(2n-1)} d_t \quad \text{for all } t>t_0,\quad k=1,2,3,\ldots. \tag{23}$$

If inequality (22) fails, then there is a constant $\epsilon>0$ and a sequence of integers $r_1<r_2<r_3<\cdots$, all satisfying

$$\limsup_j d_{r_j}^{1/r_j} > c^{2-1/n}+\epsilon. \tag{24}$$

By considering the sequence (r_j) modulo n, and taking a further subsequence, we can suppose each r_j has the form $a+b_jn$ for some fixed integer a and sequence of integers $b_1<b_2<b_3<\cdots$. Choose any integer b satisfying $a+bn>t_0$. Then we have

$$d_{r_j} = d_{a+b_jn} = d_{a+bn+(b_j-b)n} < c^{(b_j-b)(2n-1)} d_{a+bn},$$

using inequality (23). We deduce

$$d_{r_j} < c^{(n^{-1}[r_j-a]-b)(2n-1)} d_{a+bn}.$$

Raising both sides to the power $1/r_j$ and letting $j \to \infty$ now contradicts inequality (24). This completes the proof of our claim (22), and the result follows.

Naturally, the convergence of Theorem 4.3 is only local, since the projections themselves are well defined only locally, in general. However by adding a convexity assumption, we can get global convergence while preserving the local rate. A result of this kind is the following:

Corollary 4.1. *Let A and B be two closed convex subsets of $\mathbb{E}$ such that the boundaries* $\operatorname{bd} A$ *and* $\operatorname{bd} B$ *are smooth manifolds. If the intersection $A \cap B$ is nonempty, then the alternating projection method*

$$x_0 \text{ given}, \quad x_{k+1} = \mathrm{P}_A \mathrm{P}_B(x_k)$$

is well defined and converges to a point $x^* \in A \cap B$. If furthermore $\operatorname{bd} A$ *and* $\operatorname{bd} B$ *are transversal at x^*, the sequence $(x_k)_k$ in fact converges linearly, with R-linear rate* $c(\operatorname{bd} A, \operatorname{bd} B, x^*)$.

Proof. Since A and B are closed and convex, the sequence $(x_k)_k$ is well defined for any starting point x_0. The classical theory of alternating projections (see for example Cheney and Goldstein [7]) implies global convergence to a point $x^* \in A \cap B$. Theorem 4.3 then gives the local linear convergence. $\square$

For a discussion of the classical linear convergence theory for alternating projections on convex sets, see Bauschke and Borwein [3].

5. Metric regularity and linear rate. Section 4 shows that the rate of convergence of the method of alternating projections for two transverse manifolds is related to the angle between the manifolds. The speed of basic algorithms is often closely associated with Lipschitzian inequalities of the underlying generalized equations, error bounds for these equations (see Facchinei and Pang [16], for example), and metric regularity (see Rockafellar and Wets [35]). Metric regularity, in turn, is related to the conditioning of a well-posed generalized equation, measured in terms of the size of allowable linear perturbations to the equation that preserve well posedness (see the discussion in Dontchev et al. [15]). In this section, we pursue this pattern in our context by relating the angle between the manifolds to the metric regularity of a natural associated generalized equation. To accomplish this, we use a variety of tools from variational analysis: we refer the reader to Rockafellar and Wets [35] for terminology.

5.1. Regular intersection. Given two Euclidean spaces $\mathbb{E}$ and $\mathbb{Y}$, a set-valued map $\Phi\colon \mathbb{Y} \rightrightarrows \mathbb{E}$ is *metrically regular* at a point $\bar{y} \in \mathbb{Y}$ for a point $\bar{z} \in \Phi(\bar{y})$ if there exists a constant κ such that for all points $y \in \mathbb{Y}$ near $\bar{y}$ and all points $z \in \mathbb{E}$ near $\bar{z}$ we have

$$d_{\Phi^{-1}(z)}(y) \le \kappa d_{\Phi(y)}(z).$$

This inequality is fundamental, in particular for our understanding of well-behaved systems of equations and inequalities. For example, given a vector z, the quantity on the left-hand side measures the distance from an approximate solution y of the generalized equation $z \in \Phi(y)$ to the set of exact solutions, whereas the quantity on the right-hand side is a multiple of the error when we substitute y into the generalized equation. The infimum of such constants κ is called the *regularity modulus*: the smaller this modulus is, the better the generalized equation behaves.

We consider the metric regularity of the problem of finding a point in the intersection of two closed sets M and N in the space $\mathbb{E}$. To use variational tools for this analysis, we introduce the multifunction $\phi\colon \mathbb{E}^2 \rightrightarrows \mathbb{E}$ defined by

$$\phi(x, y) = \begin{cases} \{x - y\} & \text{if } x \in M \text{ and } y \in N, \\ \varnothing & \text{otherwise.} \end{cases}$$

Thus we have

$$0 \in \phi(x, y) \iff x = y \in M \cap N.$$

Therefore we say M and N have *regular intersection* at x if ϕ is metrically regular at (x, x) for 0. In that case, we define the *regularity modulus* of the intersection at x via the regularity modulus of ϕ:

$$\operatorname{reg}_{M,N}(x) := \operatorname{reg} \phi((x, x) \mid 0). \tag{25}$$

Lemma 5.1 (Coderivatives of ϕ). *Let M and N be two closed sets, and let x be a point in the intersection $M\cap N$. Then the coderivative of the multifunction ϕ at the point (x,x) is related to the normal cones to M and N at x by*

$$\forall z\in\mathbb{E},\quad D^*\phi((x,x)\mid 0)(z)=(z+\mathrm{N}_M(x),\,-z+\mathrm{N}_N(x)). \tag{26}$$

Proof. Let us write ϕ as the sum

$$\phi(x,y)=x-y+\psi(x,y),$$

where the multifunction $\psi\colon \mathbb{E}^2\rightrightarrows\mathbb{E}$ is defined by

$$\psi(x,y)=\begin{cases}0 & \text{if } x\in M \quad\text{and}\quad y\in N,\\ \varnothing & \text{otherwise.}\end{cases}$$

Since the function $F\colon (x,y)\mapsto x-y$ is smooth, the calculus rule (Rockafellar and Wets [35, 10.43]) yields

$$D^*\phi((x,x)\mid 0)(z)=\nabla F(x,x)^*(z)+D^*\psi((x,x)\mid 0)(z),$$

so

$$D^*\phi((x,x)\mid 0)(z)=(z,-z)+D^*\psi((x,x)\mid 0)(z). \tag{27}$$

Thus we just have to compute $D^*\psi((x,x)|0)(z)$. Observe that

$$\operatorname{graph}\psi=M\times N\times\{0\}\subset\mathbb{E}^3,$$

and for $x\in M\cap N$ this yields

$$\mathrm{N}_{\operatorname{graph}\psi}(x,x,0)=\mathrm{N}_M(x)\times\mathrm{N}_N(x)\times\mathbb{E}.$$

Returning to the definition of coderivatives (Rockafellar and Wets [35, 8.33]), we compute

$$\begin{aligned}(u,v)\in D^*\psi((x,x)|0)(z)\quad&\Longleftrightarrow\quad (u,v,-z)\in\mathrm{N}_{\operatorname{graph}\psi}(x,x,0)\\ &\Longleftrightarrow\quad (u,v)\in\mathrm{N}_M(x)\times\mathrm{N}_N(x).\end{aligned}$$

Thus Equation (27) gives

$$D^*\phi((x,x)|0)(z)=(z,-z)+\mathrm{N}_M(x)\times\mathrm{N}_N(x),$$

which is exactly Equation (26). □

We can use this result to recognize regular intersections, as follows:

Theorem 5.1 (Condition for Regularity). *Two closed sets M and N have regular intersection at a point $x\in M\cap N$ if and only if*

$$-\mathrm{N}_M(x)\cap\mathrm{N}_N(x)=\{0\}.$$

In this case, we also have

$$\frac{1}{\operatorname{reg}_{M,N}(x)}=\min_{\|z\|=1}\sqrt{d(z,-\mathrm{N}_M(x))^2+d(z,\mathrm{N}_N(x))^2}. \tag{28}$$

Proof. We apply Rockafellar and Wets [35, 9.43]: the metric regularity of ϕ at (x,x) for 0 is equivalent to

$$(0,0)\in D^*\phi((x,x)|0)(z)\implies z=0.$$

In view of Lemma 5.1, this means

$$(0\in z+\mathrm{N}_M(x)\text{ and }0\in -z+\mathrm{N}_N(x))\implies z=0,$$

that is, $-\mathrm{N}_M(x)\cap\mathrm{N}_N(x)=\{0\}$. Now combining Rockafellar and Wets [35, 9.43] and the Mordukhovich criterion Rockfellar and Wets [35], we obtain

$$\frac{1}{\operatorname{reg}\phi((x,x)|0)}=\min_{\|z\|=1}d((0,0),D^*\phi((x,x)|0)(z)),$$

and Equation (28) follows, using Equations (25) and (26). □

5.2. Regular intersection of two manifolds. The regularity of the intersection is easier to grasp when dealing with manifolds. The following result proves that the nonsmooth regularity notion we introduced via metric regularity coincides with the regularity notion from smooth geometry, namely transversality.

Theorem 5.2 (Regularity for Two Manifolds). *Consider two manifolds $\mathcal{M}$ and $\mathcal{N}$ around a point $x \in \mathcal{M} \cap \mathcal{N}$. Then their intersection is regular at x if and only if they are transverse at x. In this case, the intersection $\mathcal{M} \cap \mathcal{N}$ is a smooth manifold around x, and, assuming that x does not lie in both the interiors of $\mathcal{M}$ and $\mathcal{N}$, the regularity modulus is related to the angle between them by*

$$\operatorname{reg}_{\mathcal{M},\mathcal{N}}(x) = \frac{1}{\sqrt{1 - c(\mathcal{M}, \mathcal{N}, x)}}. \tag{29}$$

Proof. The normal cone $\mathrm{N}_{\mathcal{M}}(x)$ is linear in this case, so the condition for regularity of Theorem 5.1 becomes $\{0\} = \mathrm{N}_{\mathcal{M}}(x) \cap \mathrm{N}_{\mathcal{N}}(x)$. Taking orthogonal complements, the condition is then

$$\mathbb{E} = (\mathrm{N}_{\mathcal{M}}(x) \cap \mathrm{N}_{\mathcal{N}}(x))^{\perp} = \mathrm{N}_{\mathcal{M}}(x)^{\perp} + \mathrm{N}_{\mathcal{N}}(x)^{\perp} = \mathrm{T}_{\mathcal{M}}(x) + \mathrm{T}_{\mathcal{N}}(x),$$

which is exactly the transversality assumption (Definition 2.1). This inequality yields in particular that $\mathcal{M} \cap \mathcal{N}$ is a smooth manifold around x.

Let us prove now Equation (29) when x does not lie in both the interiors of $\mathcal{M}$ and $\mathcal{N}$. The assumption guarantees that $\mathrm{N}_{\mathcal{M}}(x)$ and $\mathrm{N}_{\mathcal{N}}(x)$ are not reduced to $\{0\}$. From Equation (28), we first get

$$(\operatorname{reg}_{\mathcal{M},\mathcal{N}}(x))^{-2} = \min_{z \in \mathbb{S},\, m \in \mathrm{N}_{\mathcal{M}}(x),\, n \in \mathrm{N}_{\mathcal{N}}(x)} (\|z - m\|^2 + \|z - n\|^2).$$

Since we have $\mathrm{N}_{\mathcal{M}}(x) \cap \mathrm{N}_{\mathcal{N}}(x) = \{0\}$, we can invoke Lemma 3.3 and write

$$(\operatorname{reg}_{\mathcal{M},\mathcal{N}}(x))^{-2} = 1 - c(\mathrm{N}_{\mathcal{M}}(x), \mathrm{N}_{\mathcal{N}}(x)).$$

So Lemma 3.2 yields

$$(\operatorname{reg}_{\mathcal{M},\mathcal{N}}(x))^{-2} = 1 - c(\mathrm{T}_{\mathcal{M}}(x), \mathrm{T}_{\mathcal{N}}(x)) = 1 - c(\mathcal{M}, \mathcal{N}, x),$$

which completes the proof. □

Having this connection between the regularity modulus and the angle, the asymptotical rate of convergence of the alternating projection method of Theorem 4.3 can be written as

$$1 - (\operatorname{reg}_{\mathcal{M},\mathcal{N}}(\bar{x}))^{-2}.$$

6. A numerical illustration. In this section we give a numerical illustration showing the linear convergence of the alternating projection method. We focus on the following problem: using the notation of Example 2.2, we want to find an $n \times m$ matrix X of rank r, satisfying a linear system $\mathcal{A}(X) = b$. In other words, we seek a matrix in the intersection

$$\mathcal{R}_r \cap \{X \in M_{n,m}(\mathbb{R}) \colon \mathcal{A}(X) = b\}$$

for a given linear map $\mathcal{A}\colon M_{n,m}(\mathbb{R}) \to \mathbb{R}^d$ and vector $b \in \mathbb{R}^d$. This problem is a simple analogue of feasibility problems appearing in control, and treated by alternating projections in Grigoriadis and Beran [19].

General features of alternating projection methods are that they can be implemented easily and that usually the amount of calculation in one iteration is very small. In our example, a nearest point (not necessarily unique) in $\mathcal{R}_r$ is computed through a singular value decomposition (see Example 2.3). The projection onto the affine subspace $\mathcal{A}$ of equation $\mathcal{A}(X) = b$ is computed directly as

$$\mathrm{P}_{\mathcal{A}}(X) = X - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}(\mathcal{A}(X) - b),$$

with $\mathcal{A}\mathcal{A}^*$ and its LU factorization computed only one time at the beginning of the algorithm. So the work of each iteration is dominated by the singular value decomposition.

Experiments with MATLAB on randomly generated problems (that is, the operator $\mathcal{A}(X) = (\langle A_1, X\rangle, \ldots, \langle A_d, X\rangle)$ being constructed with random matrices, and the vector b being chosen so that $\mathcal{A}(X) = b$ has a rank r solution) always exhibit the linear convergence predicted by Theorem 4.3. For our experiments, we take, in general, matrix dimensions $m \geq n$, rank r rather small (lower than 10), and we pick the number of linear equations d satisfying

$$mr < d \leq r(m + n - r).$$

The left-hand inequality ensures we cannot solve the problem too easily, simply by setting all but r rows of the matrix X to zero, and the right-hand inequality ensures (by counting dimensions) that a random problem typically has a solution and transversality holds. Starting at a random initial matrix X_0, we compute

$$X_{k+1} = \mathrm{P}_{\mathcal{R}_r}(\mathrm{P}_{\mathcal{A}}(X_k))$$

and we stop the algorithm when the absolute error satisfies

$$\|\mathcal{A}(X_k) - b\| \leq 10^{-7}.$$

We illustrate with one typical case.

Example 6.1 (Linear Convergence). We take $n = 100$, $m = 110$, $r = 4$, and $d = 450$. The algorithm stops after 1,869 iterations (with around seven minutes of computing time on a standard PC). We give below a summary of the information printed at each iteration; that is,

- `iter` is the number of the iteration,
- `log|AX-b|` $= \log_{10}(\|\mathcal{A}(X_k) - b\|)$,
- `log|X-Xpre|` $= \log_{10}(\|X_k - X_{k-1}\|)$.

iter	log|AX-b|	log|X-Xpre|
1	−0.3010	0.4604
50	−1.3197	−2.5445
100	−1.6744	−3.0021
500	−3.1839	−4.6450
1,000	−4.6511	−6.1343
1,500	−6.0199	−7.5133
1,850	−6.9516	−8.4491
1,869	−7.0018	−8.4995

We plot in Figure 1 the value of $\log_{10}(\|\mathcal{A}(X_k) - b\|)$ at each iteration k. We see that after 200 iterations the quantity decreases linearly as expected, illustrating the linear convergence.

A second, very simple, example illustrates the relationship between the angle of intersection and the convergence rate.

Example 6.2 (Angle and Rate of Convergence). We repeat the same experiment, but with $d = 1$—that is, $\mathcal{A}(X) = \langle A, X \rangle$. In this case, we can simply compute the cosine of the angle at the intersection, $c(\mathcal{R}_r, \mathcal{A}, X^*)$. Indeed, Lemmas 3.1 and 3.2 show

$$c(\mathcal{R}_r, \mathcal{A}, X^*) = c(\mathrm{N}_{\mathcal{R}_r}(X^*), A) = \|\mathrm{P}_{\mathrm{N}_{\mathcal{R}_r}(X^*)}(A/\|A\|)\|. \tag{30}$$

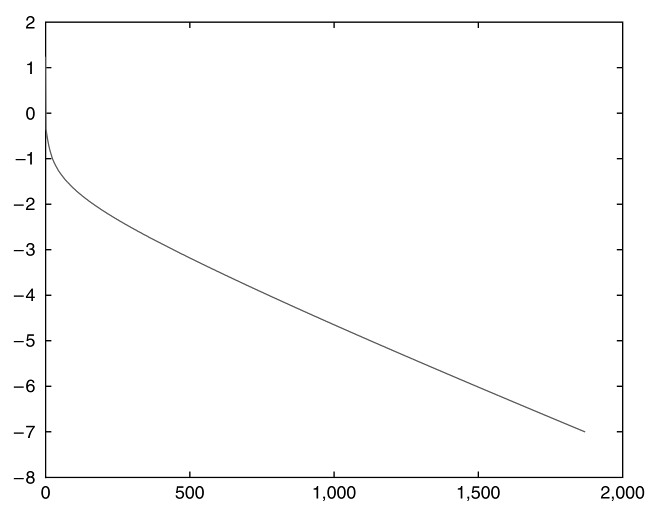

Figure 1. Plot of $\log_{10}(\|\mathcal{A}(X_k) - b\|)$ for each iteration k (with parameters $n = 100$, $m = 110$, $r = 4$, and $d = 450$).

In practice, this case is much easier and the algorithm stops after 178 iterations. Here is the information printed at some iterations:

iter	log|AX-b|	improv	log|X-Xpre|	c2
1	−1.1511	1.3350	0.4634	0.0004
50	−2.7735	0.9203	−5.1524	0.9197
100	−4.4291	0.9265	−6.8079	0.9265
150	−6.0846	0.9266	−8.4635	0.9266
178	−7.0117	0.9266	−9.3906	0.9266

At iteration k, the quantity

$$\left\langle \frac{x_k - x_{k+1}}{\|x_k - x_{k+1}\|}, \frac{\mathrm{P}_{\mathcal{A}}(x_k) - \mathrm{P}_{\mathcal{A}}(x_{k+1})}{\|\mathrm{P}_{\mathcal{A}}(x_k) - \mathrm{P}_{\mathcal{A}}(x_{k+1})\|} \right\rangle$$

provides an approximation of $c(\mathcal{R}_r, \mathcal{A}, X^*)$. We print at each iteration `c2`, the square of this quantity. We also get `improv`, the improvement at each iteration; that is

$$\frac{\|\mathcal{A}(X_k) - b\|}{\|\mathcal{A}(X_{k-1}) - b\|}.$$

We observe that the four quantities

- 10^s with s being the slope of the graph of $\log_{10}(\|\mathcal{A}(X_k) - b\|)$,
- $c(\mathcal{R}_r, \mathcal{A})^2$ the square cosine of the angle (computed with Equation (30)),
- the approximations `c2` (for the final iterations),
- the improvements `improv` (for the final iterations)

all coincide, and the common value is here around 0.9266. This illustrates that the asymptotic convergence rate is the square cosine of the angle, as predicted by Remark 4.1.

Appendix.

Projection onto spectral sets. In this appendix, we show that projection problems for spectral sets of symmetric matrices (that is, sets described solely by eigenvalue inequalities) are often easy. We begin with some basic ideas and notation, following Lewis [25, 26] and the references therein.

The space $\mathbf{S}^n$ of real symmetric n-by-n matrices, equipped with the trace inner product, is a Euclidean space. A subset T is *spectral* if, for every matrix $X \in T$ and every U in the group $\mathbf{O}^n$ of orthogonal matrices, we have $U^\top X U \in T$. The eigenvalue map λ: $\mathbf{S}^n \to \mathbb{R}^n$ maps any symmetric matrix X to its eigenvalues arranged in nonincreasing order: $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$. It is easy to see that any spectral set can be written in the form $\lambda^{-1}(K) = \{X\colon \lambda(X) \in K\}$, for some set $K \subset \mathbb{R}^n$, and that we can further restrict K to be *permutation-invariant*: for every vector $x \in K$ and every P in the group $\mathbf{P}^n$ of permutation matrices, we have $Px \in K$.

Projecting a matrix $Y \in \mathbf{S}^n$ onto a spectral set $\lambda^{-1}(K)$ (where the set $K \subset \mathbb{R}^n$ is permutation-invariant) is easy, providing we know how to project onto K. We proceed as follows: Calculate a spectral decomposition $Y = U^\top \operatorname{Diag}(y) U$, where the matrix U is orthogonal and $\operatorname{Diag}(y)$ denotes the diagonal matrix with diagonal entries the components of the vector $y \in \mathbb{R}^n$; next, find a nearest point $x \in K$ to y; now the matrix $U^\top \operatorname{Diag}(x) U$ is a nearest matrix to Y in $\lambda^{-1}(K)$.

This approach depends on the following classical result, which can be derived from the von Neumann trace inequality (von Neumann [38]):

$$\sup_{V \in \mathbf{O}^n} \operatorname{trace}\big(V^\top \operatorname{Diag}(z) V \operatorname{Diag}(y)\big) = \sup_{P \in \mathbf{P}^n} z^\top P y, \tag{31}$$

for any vectors $y, z \in \mathbb{R}^n$. See, for example, Lewis [25] for a discussion. We justify the projection procedure above in the following result:

Theorem A.1 (Spectral Projection). *If the point x in the permutation-invariant set $K \subset \mathbb{R}^n$ is a nearest point to the point $y \in \mathbb{R}^n$, then for any orthogonal matrix U the matrix $U^\top \operatorname{Diag}(x) U$ is a nearest matrix in the spectral set $\lambda^{-1}(K)$ to the matrix $U^\top \operatorname{Diag}(y) U$.*

PROOF. We can assume without loss of generality that the matrix U is the identity. Now using Equation (31), the permutation-invariance of the set K, and the assumption on the point x, we have

$$\begin{aligned}\inf_{X\in\lambda^{-1}(K)}\|X-\mathrm{Diag}(y)\|^2 &= \inf_{V\in\mathbf{O}^n,\, z\in K}\left\|V^\top \mathrm{Diag}(z)V-\mathrm{Diag}(y)\right\|^2\\ &= \inf_{V\in\mathbf{O}^n,\, z\in K}\{\|z\|^2+\|y\|^2-2\mathrm{trace}(V^\top \mathrm{Diag}(z)V\,\mathrm{Diag}(y))\}\\ &= \inf_{P\in\mathbf{P}^n,\, z\in K}\{\|z\|^2+\|y\|^2-2z^\top Py\}\\ &= \inf_{P\in\mathbf{P}^n,\, z\in K}\{\|P^\top z\|^2+\|y\|^2-2z^\top Py\}\\ &= \inf_{z\in K}\{\|z\|^2+\|y\|^2-2z^\top y\}\\ &= \inf_{z\in K}\|z-y\|^2\\ &= \|x-y\|^2.\end{aligned}$$

The first infimum is attained by $X=\mathrm{Diag}(x)$, completing the proof. □

We emphasize that the nearest points in the result above may not be unique.

A useful tool for projecting onto permutation-invariant sets is the following easy result. We denote the vectors in $\mathbb{R}^n$ with components in nonincreasing order by $\mathbb{R}^n_{\geq}$.

LEMMA A.1. *If the set $K\subset\mathbb{R}^n$ is permutation-invariant, then for any vector $y\in\mathbb{R}^n_{\geq}$ we have*

$$\inf_{x\in K}\|x-y\|=\inf_{x\in K\cap\mathbb{R}^n_{\geq}}\|x-y\|.$$

PROOF. A classical inequality (Hardy et al. [21]) shows that for any vector $x\in\mathbb{R}^n_{\geq}$ we have $\sup_{P\in\mathbf{P}^n} y^\top Px = y^\top x$. The permutation-invariance of the set K now shows

$$\begin{aligned}\inf_{x\in K}\|x-y\|^2 &= \inf_{x\in K\cap\mathbb{R}^n_{\geq},\, P\in\mathbf{P}^n}\|Px-y\|^2\\ &= \inf_{x\in K\cap\mathbb{R}^n_{\geq},\, P\in\mathbf{P}^n}\{\|x\|^2+\|y\|^2-2y^\top Px\}\\ &= \inf_{x\in K\cap\mathbb{R}^n_{\geq}}\{\|x\|^2+\|y\|^2-2y^\top x\}\\ &= \inf_{x\in K\cap\mathbb{R}^n_{\geq}}\|x-y\|^2,\end{aligned}$$

as desired. □

Our first example, showing how to project onto the *isospectral* set of all symmetric matrices with a given vector of eigenvalues, follows immediately. This result was observed in Chu and Driessel [9], Orsi [30].

EXAMPLE A.1 (ISOSPECTRAL PROJECTION). Suppose the matrix $Y\in\mathbf{S}^n$ has spectral decomposition $U^\top \mathrm{Diag}(\lambda(Y))U$. Then a nearest matrix to Y among all matrices with given eigenvalues $x_1\geq x_2\geq\cdots\geq x_n$ is the matrix $U^\top\mathrm{Diag}(x)U$. This follows by applying Lemma A.1 and Theorem A.1 to the set $K=\mathbf{P}^n x$.

The case $x=(\alpha,\alpha,\ldots,\alpha,0,0,\ldots,0)$ gives Theorem 3 in Tropp et al. [37], computing the projection onto a Grassmannian manifold.

The next example, useful in pole placement (Yang and Orsi [42]), generalizes the isospectral projection problem.

EXAMPLE A.2 (APPROXIMATE ISOSPECTRAL PROJECTION). Consider closed sets $C_1, C_2,\ldots,C_n\subset\mathbb{R}$. We ask for the closest matrix in $\mathbf{S}^n$ having a list of eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_n$ (not necessarily in decreasing order) satisfying $\lambda_i\in C_i$ for each i. Applying Theorem A.1 to the set

$$K=\mathbf{P}^n(C_1\times C_2\times\cdots\times C_n)$$

gives Theorem 8 in Yang and Orsi [42].

Another, more-challenging generalization of the isospectral projection problem appears in Chen and Chu [6]. That case concerns the set of all symmetric matrices with only some eigenvalues specified.

Example A.3 (Partial Isospectral Projection). Consider a partial list of eigenvalues $x_1 \geq x_2 \geq \cdots \geq x_k$ (where $k \leq n$). In order to solve the projection problem onto the corresponding set of symmetric matrices, we again apply Theorem A.1, using the permutation-invariant set

$$K = \{Pz\colon z \in \mathbb{R}^n,\ P \in \mathbf{P}^n,\ z_i = x_i\ (i = 1, 2, \ldots, k)\}.$$

To project a vector $y \in \mathbb{R}^n$ onto this set, we solve an assignment problem with cost matrix

$$c_{ij} = \begin{cases} (x_i - y_j)^2 & (i \leq k), \\ 0 & (i > k). \end{cases}$$

If an optimal permutation of $\{1, 2, \ldots, n\}$ for this problem is π, giving corresponding optimal value $\sum_{i=1}^{k} (x_i - y_{\pi(i)})^2$, then the point $z \in \mathbb{R}^n$, given by

$$z_j = \begin{cases} x_{\pi^{-1}(j)} & \text{if } j \in \pi\{1, 2, \ldots, k\}, \\ y_j & \text{otherwise}, \end{cases}$$

can serve as our desired projection, as discussed in Chen and Chu [6].

Another interesting case concerns projection onto the set of matrices with maximum eigenvalue having a given multiplicity. The following example completes a partial result of Oustry [32].

Example A.4 (Maximum Eigenvalue Multiplicity Projection). Suppose the matrix $Y \in \mathbf{S}^n$ has spectral decomposition $U^\top \operatorname{Diag}(\lambda(Y))U$. Then a nearest matrix to Y among all matrices with maximum eigenvalue having multiplicity at least k is the matrix $U^\top \operatorname{Diag}(x)U$, where

$$x_i = \begin{cases} k^{-1} \displaystyle\sum_{j=1}^{k} \lambda_j(Y) & (i \leq k), \\ \lambda_i(Y) & (i > k). \end{cases}$$

To see this result, we apply Lemma A.1 and Theorem A.1 to the set $K \subset \mathbb{R}^n$ consisting of all vectors whose k largest components are equal. Suppose we wish to project a point $y \in \mathbb{R}^n_{\geq}$ onto this set. By Lemma A.1, we need to solve the problem

$$\inf_{x \in K \cap \mathbb{R}^n_{\geq}} \|x - y\|,$$

and it is not hard to check that a solution is given by

$$x_i = \begin{cases} k^{-1} \displaystyle\sum_{j=1}^{k} y_j & (i \leq k), \\ y_i & (i > k). \end{cases}$$

The result then follows.

Analogous techniques hold for sets of matrices described by singular value inequalities. Specifically, we have the following results. We call a set $K \subset \mathbb{R}^n$ *absolutely permutation-invariant* if it is permutation-invariant; furthermore, whenever a point x lies in K, so do all the points

$$(\pm x_1, \pm x_2, \ldots, \pm x_3)^\top.$$

Consider the space of matrices $M_{n,m}(\mathbb{R})$, where $m \geq n$. The singular value map $\sigma\colon M_{n,m}(\mathbb{R}) \to \mathbb{R}^n$ maps a matrix X to its vector of singular values, arranged in decreasing order: $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X) \geq 0$. Given a vector $x \in \mathbb{R}^n$, the matrix $\operatorname{Diag}(x) \in M_{n,m}(\mathbb{R})$ has entries all zero except for its principal diagonal, which contains the entries from x.

Theorem A.2. *If the point x in the absolutely permutation-invariant set $K \subset \mathbb{R}^n$ is a nearest point to the point $y \in \mathbb{R}^n$, then for any orthogonal matrices U and V the matrix $U \operatorname{Diag}(x)V^\top$ is a nearest matrix in the set $\sigma^{-1}(K)$ to the matrix $U \operatorname{Diag}(y)V^\top$.*

LEMMA A.2. *If the set $K \subset \mathbb{R}^n$ is absolutely permutation-invariant, then for any nonnegative vector $y \in \mathbb{R}^n_{\geq}$ we have*

$$\inf_{x \in K} \|x - y\| = \inf_{0 \leq x \in K \cap \mathbb{R}^n_{\geq}} \|x - y\|.$$

Applying these results to the set K consisting of all vectors with at most r nonzero components gives Example 2.3 (projection onto fixed rank matrices). Considering instead the set K of vectors $[\pm\alpha, \pm\alpha, \ldots, \pm\alpha]^\top$ leads to Theorem 2 in Tropp et al. [37], which computes the projection onto a Stiefel manifold modeling the set of α-tight frames.

Acknowledgment. The first author's research was supported in part by National Science Foundation Grant DMS-0504032.

References

[1] Auslender, L. 1967. *Differential Geometry*. Harper and Row, New York.
[2] Bauschke, H. H., J. M. Borwein. 1993. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set Valued Anal.* **1**(2) 185–212.
[3] Bauschke, H. H., J. M. Borwein. 1996. On projection algorithms for solving convex feasibility problems. *SIAM Rev.* **38**(3) 367–426.
[4] Bauschke, H. H., P. L. Combettes, D. R. Luke. 2002. Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *J. Optical Soc. Amer.* **19**(7) 1334–1345.
[5] Borwein, J. M., A. S. Lewis. 2005. *Convex Analysis and Nonlinear Optimization*, 2nd ed. Springer-Verlag, New York.
[6] Chen, X., M. T. Chu. 1996. On the least squares solution of inverse eigenvalue problems. *SIAM J. Numer. Anal.* **33** 2417–2430.
[7] Cheney, W., A. Goldstein. 1959. Proximity maps for convex sets. *Proc. Amer. Math. Soc.* **10**(3) 448–450.
[8] Chu, M. T. 1995. Constructing a Hermitian matrix from its diagonal entries and eigenvalues. *SIAM J. Matrix Anal. Appl.* **16** 207–217.
[9] Chu, M. T., K. R. Driessel. 1990. The projected gradient method for least squares matrix approximations with spectral constraints. *SIAM J. Numer. Anal.* **27** 1050–1060.
[10] Combettes, P. L. 1993. Signal recovery by best feasible approximation. *IEEE Trans. Image Processing* **2**(2) 269–271.
[11] Combettes, P. L., H. J. Trussell. 1990. Method of successive projections for finding a common point of sets in metric spaces. *J. Optim. Theory Appl.* **67**(3) 487–507.
[12] Dennis, J. E., R. B. Schnabel. 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ.
[13] Deutsch, F. 1995. The angle between subspaces of a Hilbert space. S. P. Singh, ed. *Approximation Theory, Wavelets and Applications*. Kluwer, Dordrecht, Netherlands, 107–130.
[14] Deutsch, F. 2001. *Best Approximation in Inner Product Spaces*. Springer, New York.
[15] Dontchev, A. L., A. S. Lewis, T. Rockafellar. 2003. The radius of metric regularity. *Trans. Amer. Math. Soc.* **355**(2) 493–517.
[16] Facchinei, F., J.-S. Pang. 2003. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. II. Springer, New York.
[17] Friedricks, K. 1937. On certain inequalities and characteristic value problems for analytic functions and for functions of two variables. *Trans. Amer. Math. Soc.* **41** 321–364.
[18] Golub, G. H., C. F. Van Loan. 1995. *Matrix Computations*, 3rd ed. Johns Hopkins University Press, Baltimore.
[19] Grigoriadis, K. M., E. Beran. 2000. Alternating projection algorithms for linear matrix inequalities problems with rank constraints. *Adv. Linear Matrix Inequality Methods in Control, SIAM*, Philadelphia, 251–267.
[20] Grigoriadis, K. M., R. E. Skelton. 1996. Low-order control design for LMI problems using alternating projection methods. *Automatica* **32** 1117–1125.
[21] Hardy, G. H., J. E. Littlewood, G. Pólya. 1952. *Inequalities*. Cambridge University Press, Cambridge, UK.
[22] Higham, N. 2002. Computing a nearest symmetric correlation matrix: A problem from finance. *IMA J. Numer. Anal.* **22**(3) 329–343.
[23] Holmes, R. B. 1973. Smoothness of certain metric projections on Hilbert space. *Trans. Amer. Math. Soc.* **184** 87–100.
[24] Horn, R. A., Ch. R. Johnson. 1989. *Matrix Analysis*, New ed. 1999. Cambridge University Press, Cambridge, UK.
[25] Lewis, A. S. 1996. Convex analysis on the Hermitian matrices. *SIAM J. Optim.* **6** 164–177.
[26] Lewis, A. S. 1996. Group invariance and convex matrix analysis. *SIAM J. Matrix Anal. Appl.* **17** 927–949.
[27] Lewis, A. S., D. R. Luke, J. Malick. 2007. Local linear convergence of alternating and averaged projections. Preprint, Cornell University School of ORIE, Ithaca, NY.
[28] Nesterov, Y. E., A. S. Nemirovskii. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia.
[29] Nocedal, J., S. J. Wright. 1999. *Numerical Optimization*. Springer Verlag, New York.
[30] Orsi, R. 2006. Numerical methods for solving inverse eigenvalue problems for nonnegative matrices. *SIAM J. Matrix Anal. Appl.* **28** 190–212.
[31] Orsi, R., U. Helmke, J. Moore. 2006. A Newton-like method for solving rank constrained linear matrix inequalities. *Automatica* **42** 1875–1882.
[32] Oustry, F. 2000. A second-order bundle method to minimize the maximum eigenvalue function. *Math. Programming* **89** 1–34.
[33] Poliquin, R. A., R. T. Rockafellar, L. Thibault. 2000. Local differentiability of distance functions. *Trans. Amer. Math. Soc.* **352** 5231–5249.
[34] Robinson, S. B. 2003. Constraint nondegeneracy in variational analysis. *Math. Oper. Res.* **28** 201–232.
[35] Rockafellar, R. T., R. J.-B. Wets. 1998. *Variational Analysis*. Springer Verlag, Heidelberg, Germany.
[36] Smith, K. T., D. C. Solomon, S. L. Wagner. 1977. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bull. Amer. Math. Soc.* **83** 1227–1270.

[37] Tropp, J. A., I. S. Dhillon, R. W. Heath, T. Strohmer. 2005. Designing structured tight frames via an alternating projection method. *IEEE Trans. Inform. Theory* **51** 188–209.

[38] von Neumann, J. 1937. Some matrix inequalities and metrization of matric-space. *Tomsk University Rev.* **1** 286–300. *Collected Works*, Vol. IV. Pergamon, Oxford, UK, 1962, 205–218.

[39] von Neumann, J. 1950. Functional operators, Vol. II. The geometry of orthogonal spaces. *Annals of Mathematical Studies*, Vol. 22. Princeton University Press, Princeton, NJ. [Reprint of mimeographed notes first distributed in 1933.]

[40] Weber, C. A., J. P. Allebach. 1986. Reconstruction of frequency-offset Fourier data by alternating projection on constraint sets. *Proc. 24th Allerton Conf. Comm., Control and Comput.*, Monticello, IL, 194–203.

[41] Widrow, B., E. Wallach. 1996. *Adaptive Inverse Control*. Prentice-Hall, Englewood Cliffs, NJ.

[42] Yang, K., R. Orsi. 2006. Generalized pole placement via static output feedback: A methodology based on projections. *Automatica* **42** 2143–2150.