

Alternating projections and coupling slope

D. Drusvyatskiy* A.D. Ioffe† A.S. Lewis‡

January 28, 2014

Abstract

We consider the method of alternating projections for finding a point in the intersection of two possibly nonconvex closed sets. We present a local linear convergence result that makes no regularity assumptions on either set (unlike previous results), while at the same time weakening standard transversal intersection assumptions. The proof grows out of a study of the slope of a natural nonsmooth coupling function. When the two sets are semi-algebraic and bounded, we also prove subsequence convergence to the intersection with no transversality assumption.

Key words: alternating projections, linear convergence, variational analysis, slope, transversality

AMS 2000 Subject Classification: 49M20, 65K10, 90C30

1 Introduction

Finding a point in the intersection of two closed, convex subsets X and Y of \mathbf{R}^n is a ubiquitous problem in computational mathematics. A conceptually simple and widely used method for solving feasibility problems of this form is the *method of alternating projections* — discovered and rediscovered by a number of authors, notably von Neumann [30]. The method presupposes that the nearest point mappings

*Department of Mathematics, University of Washington, Seattle, WA 98195; Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; people.orie.cornell.edu/dd379.

†Mathematics Department, Technion-Israel Institute of Technology, Haifa, Israel 32000; ioffe@math.technion.ac.il. Research supported in part by the US-Israel Binational Science Foundation Grant 2008261.

‡ORIE, Cornell University, Ithaca, NY 14853, U.S.A. adrian.lewis@cornell.edu people.orie.cornell.edu/aslewis. Research supported in part by National Science Foundation Grant DMS-1208338 and by the US-Israel Binational Science Foundation Grant 2008261.

to X and to Y can be easily computed — a reasonable assumption in a variety of applications. The algorithm then proceeds by projecting a starting point onto the first set X , then projecting the resulting point onto Y , and then projecting back onto X , and so on and so forth. Sublinear convergence of the method for two intersecting closed convex sets is always guaranteed [10], while linear convergence holds whenever the relative interiors of X and Y intersect [3].

The method of alternating projections makes sense even for nonconvex feasibility problems and has been used extensively, notably for inverse eigenvalue problems [11, 12], pole placement [32], information theory [29], control design [15, 16, 27], and phase retrieval [5, 31]. Aiming to explain the success of the method in nonconvex settings, the authors of [24, 25] established the first (linear) convergence guarantees for the method of alternating projections in absence of convexity. More recently, their results have been further refined and extended in a series of papers [6, 7, 18]. To date, convergence proofs in the literature combine two distinct ingredients:

- (i) an appeal to a transversality-like condition of the intersection, lower-bounding the “angle” between the two sets at a point of the intersection
- (ii) a convexity-like property of one of the sets.

Easy examples immediately suggest that some condition on the intersection of the two sets is necessary to guarantee linear convergence. The second ingredient, on the other hand, seems more an artifact of the proofs rather than being inherently tied to the algorithm. In this work, we entirely dispense with the second ingredient, using an interesting new argument deviating sharply from the previous techniques. Moreover, we show that if the two sets X and Y are semi-algebraic (meaning that they are composed of finitely many sets each defined by finitely many polynomial inequalities [8, 13]) and one of the sets is compact, then the method of alternating projections initiated sufficiently close to the intersection always generates iterates whose distance to the intersection tends to zero. A result of the same flavor for averaged projections appears in [2]; it does not subsume the result we present here.

The outline of the manuscript is as follows. In Section 2, we record the relevant background about the method of alternating projections and we summarize the main results of the manuscript. Section 3 contains the proof of our basic linear convergence result. In Section 4, we compare various notions of transversality needed to guarantee linear convergence of alternating projections. In Sections 5 and 6, we study the slope and the coupling function, key ingredients in our convergence argument. Finally, in Section 7 we consider alternating projections in the context of semi-algebraic intersections. We mention in passing that all results of the latter section hold much more generally for tame functions; see [22] for the role of such functions in nonsmooth optimization.

2 Background on alternating projections

We consider a Euclidean space \mathbf{E} with inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$. For any point $y \in \mathbf{E}$ and any set $X \subset \mathbf{E}$, the *distance* from y to X and the *projection* of y onto X are

$$\begin{aligned} d(y, X) &= \inf\{|y - x| : x \in X\}, \\ P_X(y) &= \operatorname{argmin}\{|y - x| : x \in X\}, \end{aligned}$$

respectively. For any point $x \in X$, vectors in the cone

$$N_X^p(x) = \{\lambda u : \lambda \in \mathbf{R}_+, x \in P_X(x + u)\}$$

are called *proximal normals* to X at x . Limits of proximal normals to X at points $x_n \in X$ approaching x are called *limiting normals*, and comprise the *limiting normal cone* $N_X(x)$.

The *method of alternating projections* for finding a point in the intersection of two nonempty closed sets $X, Y \subset \mathbf{E}$ iterates the following pair of steps: given a current point $x_n \in X$,

$$\begin{aligned} \text{choose } y_n &\in P_Y(x_n) \\ \text{choose } x_{n+1} &\in P_X(y_n). \end{aligned}$$

Convergence proofs typically combine two distinct ingredients, each involving angles (which we always consider to lie in the interval $[0, \pi]$), along the following lines; see Figure 1 for an illustration.

- Appeal to a transversality-like condition on the intersection of X and Y to show that the angle between the vectors

$$x_n - y_n \in N_Y^p(y_n) \quad \text{and} \quad x_{n+1} - y_n \in -N_X^p(x_{n+1})$$

is always larger than some constant angle $\theta > 0$.

- Then appeal to a convexity-like property of the set X to show that the angle between the vectors

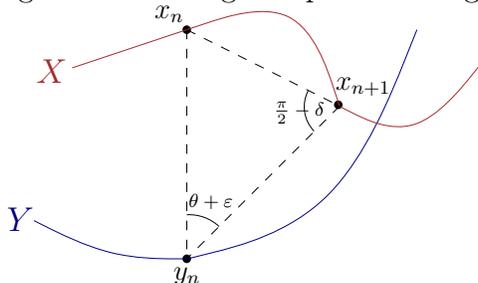
$$x_n - x_{n+1} \quad \text{and} \quad y_n - x_{n+1}$$

cannot be much less than $\pi/2$.

A consequence, via the elementary geometry of the triangle with vertices x_n , y_n and x_{n+1} , is

$$\frac{|y_n - x_{n+1}|}{|x_n - y_n|} < \cos \theta,$$

Figure 1: Convergence proof strategy.



and then a routine argument proves that the sequence of iterates converges to a point \bar{x} in the intersection with R-linear rate $\cos \theta$. In other words, the error $|x_n - \bar{x}|$ is uniformly bounded by a multiple of $(\cos \theta)^n$.

Classically, the second ingredient above consisted simply of the assumption that both sets X and Y were convex; see for example [4, 17]. Recent nonconvex work [1, 6, 7, 18, 24] showed that, to guarantee local linear convergence, a weaker property of the set X suffices, a property slightly stronger than the familiar variational-analytic notion of “Clarke regularity” [28, Definition 6.4].

Definition 2.1 (Super-regularity) A closed set $X \subset \mathbf{E}$ is *super-regular* at a point $z \in X$ when, for all $\delta > 0$, if distinct points $w, x \in X$ are sufficiently near z , then their difference $w - x$ makes an angle of at least $\frac{\pi}{2} - \delta$ with any nonzero normal $v \in N_X(x)$.

In particular, this property holds providing the projection operator P_X is single-valued on a neighborhood of z , a property known as “prox-regularity” [28, Exercise 13.38]. This property was even further weakened in [26] for subanalytic intersections.

In this work, we dispense entirely with this second ingredient: we make no assumptions whatsoever about regularity properties of either set X or Y . Instead we rely solely on transversality-like behavior of the intersection, as follows.

Definition 2.2 (Intrinsic transversality) Two sets $X, Y \subset \mathbf{E}$ are *intrinsically transversal* at a point $z \in X \cap Y$ if there exists $\kappa \in (0, 1]$ (a *constant of transversality*) and a neighborhood Z of z such that the difference $x - y$ between any points $x \in X \cap Y^c \cap Z$ and $y \in Y \cap X^c \cap Z$ cannot simultaneously make an angle strictly less than $\arcsin \kappa$ with both the cones $N_Y^p(y)$ and $-N_X^p(x)$

For notational convenience, we introduce the normalization map $\widehat{\cdot}: \mathbf{E} \setminus \{0\} \rightarrow \mathbf{E}$, defined by $\widehat{z} = \frac{z}{|z|}$. Then, more succinctly, this angle condition is:

$$\max \left\{ d(u, N_Y^p(y)), d(u, -N_X^p(x)) \right\} \geq \kappa \quad \text{for } u = \widehat{x - y}.$$

For future reference, if we replace proximal normal cones with normal cones in this definition, we call the resulting property *strong intrinsic transversality*. Clearly we have

$$\text{strong intrinsic transversality} \implies \text{intrinsic transversality},$$

with the same constant κ . (In a later section, we prove these properties are in fact equivalent.)

The following is our main result.

Theorem 2.3 (Linear convergence) *Suppose two closed sets $X, Y \subset \mathbf{E}$ are intrinsically transversal at a point $z \in X \cap Y$, with constant $\kappa > 0$. Then, given any constant c in the interval $(0, \kappa)$, the method of alternating projections, initiated sufficiently near z , must converge to a point in the intersection $X \cap Y$ with R -linear rate $1 - c^2$.*

As we discuss later, this theorem immediately implies local linear convergence of alternating projections near any point z where the two sets X and Y intersect *transversally*: $N_X(z) \cap -N_Y(z) = \{0\}$. The authors of [24] prove the same result, but under the additional assumption that at least one of the two sets is super-regular at z .

It is clear that without intrinsic transversality, local linear convergence is difficult to guarantee. However, one may ask whether local (perhaps sublinear) convergence is always assured. It is easy to concoct pathological examples where, without intrinsic transversality, no matter how close the initial point is to the intersection, the iterates generated by the method of alternating projections never tend to the intersection. On the other hand, such pathologies rarely appear in practice. In particular we will show that this does not happen when both of the sets are semi-algebraic (Theorem 7.3).

3 Proof of the basic result

Our main result, Theorem 2.3 (Linear convergence), follows quickly from the following appealing geometric tool. This tool concerns the distance from a point y to a set X . Given a trial point $x \in X$, it shows how to improve the upper bound on $|x - y|$ under the assumption of a uniform lower bound on the angle between normal vectors to X at points w near x and line segments from w to y .

Theorem 3.1 (Distance decrease) *Consider a closed set $X \subset \mathbf{E}$ and points $x \in X$ and $y \notin X$ with $\rho := |y - x|$. Given any constant $\delta > 0$, if*

$$\inf_{w \in X} \{d(\widehat{y-w}, N_X(w)) : w \in B_\rho(y) \cap B_\delta(x)\} = \mu > 0,$$

then

$$d(y, X) \leq |y - x| - \mu\delta.$$

This result is a consequence of the Ekeland Variational Principle. We prove it later, but we first observe how our basic result follows.

We consider two closed sets $X, Y \subset \mathbf{E}$ that are intrinsically transversal at a point $z \in X \cap Y$, with constant $\kappa \in (0, 1]$. As we shall see (Proposition 6.9 (Intrinsic transversality)), we can assume strong intrinsic transversality holds. Hence there exists a constant $\epsilon > 0$ such that for any distinct points $x \in X$ and $y \in Y$, both lying in the open ball $B_\epsilon(z)$, we have

$$\max \{d(u, N_Y(y)), d(u, -N_X(x))\} \geq \kappa \text{ for } u = \widehat{x - y}.$$

Now consider any point $x \in X$ and $y \in P_Y(x)$. We aim to deduce the inequality

$$d(y, X) \leq (1 - c^2)|y - x|.$$

The result will then follow by a routine induction. We can assume $y \notin X$, since otherwise there is nothing to prove.

Define $\rho = |x - y|$ and assume

$$x, y \in B_{\epsilon - \kappa\rho}(z),$$

as certainly holds if x is sufficiently close to z . Consider any point $w \in X \cap B_{\kappa\rho}(x)$. Since $|w - x| < \kappa|x - y| \leq d_Y(x)$, we see $w \notin Y$, and in particular $w \neq y$. We have

$$d(\widehat{w - y}, N_Y(y)) \leq d(\widehat{w - y}, \mathbf{R}_+(x - y)),$$

since $x - y \in N_Y(y)$. The following simple result is useful in bounding the right-hand side.

Lemma 3.2 *Any nonzero vectors $p, q \in \mathbf{E}$ satisfy*

$$d(\widehat{p}, \mathbf{R}_+q) \leq \frac{|p - q|}{|q|}.$$

Proof The following more precise inequality is simple to verify algebraically:

$$|\widehat{p} - \langle \widehat{p}, \widehat{q} \rangle \widehat{q}| \leq \frac{|p - q|}{|q|}.$$

The result follows. □

Applying this result with $p = w - y$ and $q = x - y$ shows, for any $w \in X \cap B_{\kappa\rho}(x)$, the inequality

$$d(\widehat{w - y}, N_Y(y)) < \kappa,$$

and hence, using intrinsic transversality

$$d(\widehat{y - w}, N_X(w)) \geq \kappa.$$

We next apply Theorem 3.1 with $\delta = c\rho$, and deduce

$$d(y, X) \leq (1 - c^2)|y - x|,$$

as required. The result now follows.

4 Transversality and related work

We next pause to study the idea of intrinsic transversality in the context of related notions. Once again, we consider two sets $X, Y \subset \mathbf{E}$ and a point z in their intersection $X \cap Y$.

The standard variational-analytic idea of *transversality* (in the finite dimensional case) is simply the property

$$N_X(z) \cap -N_Y(z) = \{0\}.$$

In that case, an easy limiting argument shows that the quantity

$$(4.1) \quad \min_{|u|=1} \max \left\{ d(u, N_Y(z)), d(u, -N_X(z)) \right\}$$

is strictly positive, and consequently that strong intrinsic transversality holds for any strictly smaller constant $\kappa > 0$. We immediately deduce the following corollary.

Corollary 4.2 (Transversality and linear convergence)

Suppose two closed sets $X, Y \subset \mathbf{E}$ intersect transversally at a point $z \in X \cap Y$. Then the method of alternating projections, initiated sufficiently near z , must converge R -linearly to a point in the intersection $X \cap Y$. Indeed, if the minimal angle between vectors in the closed cones $N_Y(z)$ and $-N_X(z)$ is $\theta \in (0, \pi]$, then for any positive constant $c < \cos \frac{\theta}{2}$, the method of alternating projections, initiated sufficiently near z , must converge to a point in $X \cap Y$ with R -linear rate $1 - c^2$.

The simple example of two intersecting lines in \mathbf{R}^3 shows that transversality is a strictly stronger property than intrinsic transversality. Indeed, if the affine span of the union $X \cup Y$ fails to be \mathbf{E} , it is easy to see that transversality must fail. On the other hand, such a deficit in the affine span has no direct bearing on the method of alternating projections, making it appealing to consider an appropriately modified version of transversality. We proceed as follows.

If the point zero lies in the intersection $X \cap Y$, we say that X and Y intersect *relatively transversally at zero* if, when considered as subsets of the Euclidean space $\text{span}(X \cup Y)$, they intersect transversally at zero. In general, we say that X and Y intersect *relatively transversally* at a point $z \in X \cap Y$ when the translates $X - z$ and $Y - z$ intersect relatively transversally at zero. Relative transversality is easy to recognize in the convex case, as the following result shows.

Proposition 4.3 (Relative transversality and relative interiors)

Two convex sets intersect relatively transversally at a point in their intersection if and only if their relative interiors intersect.

The proof is a straightforward exercise in convex analysis.

Essentially the same argument as before shows that relative transversality implies intrinsic transversality. We deduce a version of the linear convergence result Corollary 4.2 with the assumption of transversality weakened to relative transversality. In the convex case we retrieve classical convergence results in the case of intersecting relative interiors.

Summarizing, we have the following implications.

$$\begin{aligned}
 & \text{transversality} \\
 & \implies \text{relative transversality} \\
 & \implies \text{strong intrinsic transversality} \\
 & \implies \text{intrinsic transversality}
 \end{aligned}$$

Of these four properties, the first is strictly stronger than the second, as shown by the example of two intersecting lines in \mathbf{R}^3 . Equally simple convex examples show that the second is strictly stronger than the third: consider for example the space $\mathbf{E} = \mathbf{R}^2$ with sets $X = \mathbf{R} \times \{0\}$ and $Y = \{0\} \times \mathbf{R}_+$, at their point of intersection 0. We later show the equivalence of the third and fourth properties.

Bauschke, Luke, Phan and Wang [6, 7] weaken the assumption of transversality (and indeed relative transversality) in a different way, also attuned to the method of alternating projections. They introduce a notion of *restricted normal cone* that leads (though not in their precise language) to the following idea.

Definition 4.4 (Inherent transversality) Two sets $X, Y \subset \mathbf{E}$ are *inherently transversal* at a point $z \in X \cap Y$ if there exists a constant $\epsilon > 0$ and a neighborhood Z of z such that, for any points $x \in X \cap Y^c \cap Z$ and $y \in Y \cap X^c \cap Z$, and any corresponding nearest points $p \in P_Y(x)$ and $w \in P_X(y)$, the angle between the differences $x - p$ and $w - y$ is at least ϵ .

With this assumption, along with an additional super-regularity assumption, the authors of [6] prove linear convergence. In this context, the following result is notable.

Proposition 4.5 (Inherent versus intrinsic transversality)

Suppose two closed sets $X, Y \subset \mathbf{E}$ are inherently transversal at a point $z \in X \cap Y$. If both X and Y are super-regular at z , then they are also strongly intrinsically transversal at z .

Proof Suppose strong intrinsic transversality fails. Then there exists sequences of points (x_r) in $X \cap Y^c$ and (y_r) in $Y \cap X^c$, both approaching the point z , and corresponding nonzero normals $v_r \in N_X(x_r)$ and $q_r \in N_Y(y_r)$ (for $r = 1, 2, 3, \dots$), such that the angles θ_r and ψ_r between the difference $x_r - y_r$ and the vectors $-v_r$ and q_r both converge to zero.

Fix any constant $\delta > 0$. Choose a point $w_r \in P_X(y_r)$ (for $r = 1, 2, 3, \dots$). Since the set X is super-regular at the limit point z , the angle between the difference $w_r - x_r$ and the normal vector v_r is eventually always at least $\frac{\pi}{2} - \delta$. Similarly, the angle between the difference $x_r - w_r$ and the normal vector $y_r - w_r \in N_X(w_r)$ is eventually always at least $\frac{\pi}{2} - \delta$. It follows, geometrically, that the angle between the two vectors $w_r - y_r$ and $x_r - y_r$ is no larger than $\theta_r + 2\delta$, and hence converges to zero, since δ was arbitrary.

We now argue, symmetrically, that for any sequence of points $p_r \in P_Y(x_r)$ (for $r = 1, 2, 3, \dots$), the angle between the two vectors $p_r - x_r$ and $y_r - x_r$ is eventually no larger than $\psi_r + 2\delta$. Combining this with our previous observation shows that the angle between $x_r - p_r$ and $w_r - y_r$ is eventually no larger than $\theta_r + \psi_r + 4\delta$, and hence converges to zero, since the constant $\delta > 0$ was arbitrary. \square

5 Slope

The intrinsic transversality properties are closely related to a coupling function, which we discuss in the next section, and its “slope”, a notion to which we turn next. Denoting the extended reals $[-\infty, +\infty]$ by $\overline{\mathbf{R}}$, if a function $f: \mathbf{E} \rightarrow \overline{\mathbf{R}}$ is finite at a point $x \in \mathbf{E}$, then its *slope* there is

$$|\nabla f|(x) = \limsup_{\substack{w \rightarrow x \\ w \neq x}} \frac{f(x) - f(w)}{|x - w|},$$

unless x is a local minimizer, in which case we set $|\nabla f|(x) = 0$. A vector v lies in the *proximal subdifferential* $\partial^p f(x)$ whenever $(v, -1)$ is a proximal normal to the epigraph of f at $(\bar{x}, f(\bar{x}))$; the *limiting subdifferential* $\partial f(x)$ is defined analogously. We always have the inequality

$$(5.1) \quad |\nabla f|(x) \leq d(0, \partial^p f(x)),$$

relating the slope and the proximal subdifferential. In contrast, the *limiting slope*

$$\overline{|\nabla f|}(x) = \liminf_{\substack{w \rightarrow x \\ f(w) \rightarrow f(x)}} |\nabla f|(w),$$

has an exact relationship with the limiting subdifferential

$$(5.2) \quad \overline{|\nabla f|}(x) = d(0, \partial f(x)),$$

when f is lower semicontinuous. For a proof, see for example [14, Proposition 4.6].

Our approach to linear convergence of alternating projections relies on the following key tool from Ioffe [19, Basic Lemma, Chapter 1], giving a slope-based criterion for f to have a nonempty level set

$$[f \leq \alpha] = \{u \in \mathbf{E} : f(u) \leq \alpha\}.$$

We note in passing that the result holds more generally, with \mathbf{E} replaced by any complete metric space.

Theorem 5.3 (Error bound)

Suppose that the lower semicontinuous function $f: \mathbf{E} \rightarrow \overline{\mathbf{R}}$ is finite at the point $x \in \mathbf{E}$, and that the constants $\delta > 0$ and $\alpha < f(x)$ satisfy

$$\inf_{w \in \mathbf{E}} \{ |\nabla f|(w) : \alpha < f(w) \leq f(x), |w - x| \leq \delta \} > \frac{f(x) - \alpha}{\delta}.$$

Then the level set $[f \leq \alpha]$ is nonempty, and furthermore its distance from x is no more than $\frac{1}{K}(f(x) - \alpha)$, where K denotes the left-hand side of the inequality above.

Proof We apply the Ekeland variational principle to the function $g: \mathbf{E} \rightarrow \overline{\mathbf{R}}$, defined as the positive part of the function $f - \alpha$. We deduce, for any constant $\gamma \in (\frac{1}{\delta}(f(x) - \alpha), K)$, the existence of a point v minimizing the function $g + \gamma|\cdot - v|$ and satisfying the properties $g(v) \leq g(x)$ and $|v - x| \leq \frac{1}{\gamma}g(x) < \delta$. The minimizing property of v shows $|\nabla g|(v) \leq \gamma < K$. Hence $g(v) = 0$, since otherwise $|\nabla g|(v) = |\nabla f|(v) \geq K$. The result now follows by letting γ approach K . \square

Our main geometric tool, Theorem 3.1, is a special case. To see this, we apply the first part of Theorem 5.3 (Error bound) to the function $f = |\cdot - y| + \delta_X$, and use the inequality

$$|\nabla f|(w) \geq \overline{|\nabla f|}(w) = d(\widehat{y - w}, N_X(w)),$$

which follows from equation (5.2). For any constant $\alpha < |x - y|$ such that $\mu > \frac{1}{\delta}(|x - y| - \alpha)$, we deduce $d(y, X) \leq \alpha$. The result now follows.

6 The coupling function

Proving the equivalence of various intrinsic transversality notions relies on a careful analysis of a “coupling” function for the two nonempty closed sets $X, Y \subset \mathbf{E}$. We define a function $\phi: \mathbf{E}^2 \rightarrow \overline{\mathbf{R}}$ by

$$\phi(x, y) = \delta_X(x) + |x - y| + \delta_Y(y),$$

where δ_X and δ_Y denote the indicator functions of the sets X and Y respectively. We also consider the marginal function $\phi_y: \mathbf{E} \rightarrow \overline{\mathbf{R}}$ (for any fixed point $y \in \mathbf{E}$) defined by $\phi_y(x) = \phi(x, y)$, and the marginal function $\phi_x: \mathbf{E} \rightarrow \overline{\mathbf{R}}$ (for any fixed point $x \in \mathbf{E}$) defined by $\phi_x(y) = \phi(x, y)$. Standard subdifferential calculus applied

to the coupling function and its marginals shows, for distinct points $x \in X$ and $y \in Y$, with $u = \widehat{x - y}$,

$$(6.1) \quad \partial\phi_y(x) = u + N_X(x) \quad \text{and} \quad \partial^p\phi_y(x) = u + N_X^p(x)$$

$$(6.2) \quad \partial\phi_x(y) = -u + N_Y(y) \quad \text{and} \quad \partial^p\phi_x(y) = -u + N_Y^p(y)$$

$$(6.3) \quad \partial\phi(x, y) = \partial\phi_y(x) \times \partial\phi_x(y) \quad \text{and} \quad \partial^p\phi(x, y) = \partial^p\phi_y(x) \times \partial^p\phi_x(y)$$

Using the relationships (5.1) and (5.2), we arrive at the following expressions for the slopes and limiting slopes:

$$(6.4) \quad \overline{|\nabla\phi_y|}(x) = d(u, -N_X(x)) \quad \text{and} \quad |\nabla\phi_y|(x) \leq d(u, -N_X^p(x))$$

$$(6.5) \quad \overline{|\nabla\phi_x|}(y) = d(u, N_Y(y)) \quad \text{and} \quad |\nabla\phi_x|(y) \leq d(u, N_Y^p(y)).$$

We also note

$$(6.6) \quad \overline{|\nabla\phi|}(x, y) = \sqrt{(\overline{|\nabla\phi_y|}(x))^2 + (\overline{|\nabla\phi_x|}(y))^2}.$$

Thus equipped, we can rewrite strong intrinsic transversality at a point $z \in X \cap Y$ in an equivalent form.

Definition 6.7 Two sets $X, Y \subset \mathbf{E}$ are *strongly intrinsically transversal* at a point $z \in X \cap Y$ if there exists $\kappa > 0$ (a *constant*) and a neighborhood Z of z such that, for any points $x \in X \cap Y^c \cap Z$ and $y \in Y \cap X^c \cap Z$, the limiting slopes $\overline{|\nabla\phi_y|}(x)$ and $\overline{|\nabla\phi_x|}(y)$ cannot both be strictly less than κ .

Crucial to our development is the corresponding property with slopes replacing limiting slopes.

Definition 6.8 Two sets $X, Y \subset \mathbf{E}$ are *strictly intrinsically transversal* at a point $z \in X \cap Y$ if there exists $\kappa > 0$ (the *constant*) and a neighborhood Z of z such that, for any points $x \in X \cap Y^c \cap Z$ and $y \in Y \cap X^c \cap Z$, the slopes $|\nabla\phi_y|(x)$ and $|\nabla\phi_x|(y)$ cannot both be strictly less than κ :

$$\max \{ |\nabla\phi_y|(x), |\nabla\phi_x|(y) \} \geq \kappa.$$

Proposition 6.9 (Intrinsic transversality) *Consider a point z in the intersection of two closed sets $X, Y \subset \mathbf{E}$. For any constant $\kappa > 0$, the properties of intrinsic transversality, strong intrinsic transversality, and strict intrinsic transversality are all equivalent.*

Proof Clearly strong intrinsic transversality implies strict intrinsic transversality, because the limiting slope is never larger than the slope. On the other hand, strict intrinsic transversality implies intrinsic transversality, due to the inequalities

in (6.4) and (6.5). It remains to prove intrinsic transversality implies strong intrinsic transversality.

Suppose strong intrinsic transversality fails. Then, given any open neighborhood Z of z , there exist points $x \in X \cap Y^c \cap Z$ and $y \in Y \cap X^c \cap Z$ with both the limiting slopes $|\overline{\nabla\phi_y}|(x)$ and $|\overline{\nabla\phi_x}|(y)$ strictly less than κ . Equation (5.2) then ensures the existence of subgradients $p \in \partial\phi_y(x) \cap \kappa B$ and $q \in \partial\phi_x(y) \cap \kappa B$, where B denotes the open unit ball. The relationship (6.3) shows $(p, q) \in \partial\phi(x, y)$, so there exist points $x' \in X \cap Y^c \cap Z$ and $y' \in Y \cap X^c \cap Z$ and a proximal subgradient

$$(p', q') \in \partial^p\phi(x', y') \cap \kappa(B \times B) = (\partial^p\phi_{y'}(x') \cap \kappa B) \times (\partial^p\phi_{x'}(y') \cap \kappa B).$$

Applying the inequalities in (6.4) and (6.5), we deduce that both the slopes $|\nabla\phi_{y'}|(x')$ and $|\nabla\phi_{x'}|(y')$ are strictly less than κ . Since the neighborhood Z was arbitrary, intrinsic transversality also fails. \square

Notice that, according to equation (6.6), our intrinsic transversality properties amount to assuming that the *slope of the coupling function* $|\nabla\phi|(x, y)$ *is uniformly bounded away from zero as points* $x \in X \cap Y^c$ *and* $y \in Y \cap X^c$ *approach* z .

7 Semi-algebraic intersections

In our main result, Theorem 2.3 (Linear convergence), we showed that intrinsic transversality implies local linear convergence of alternating projections. Motivated by this result, we begin this section by asking: to what extent is intrinsic transversality a common property? Classically (and reassuringly) two smooth manifolds “typically” intersect transversally: that is, almost all linear perturbations to the manifolds yield a transverse intersection. The following theorem show that the analogous result holds much more generally for semi-algebraic sets — those sets that are finite unions of sets, each defined by finitely many polynomial inequalities. Semi-algebraic sets and functions (those whose graphs are semi-algebraic) nicely exemplify concrete optimization problems; see [22] for their role in nonsmooth optimization. We mention in passing that all results of this section extend from semi-algebraic to “tame” sets [22].

Theorem 7.1 (Generic transversality)

Consider closed semi-algebraic subsets X and Y of a Euclidean space \mathbf{E} . Then for almost every vector e in \mathbf{E} , transversality holds at every point in the intersection of the sets X and $Y - e$.

Proof Consider the set-valued mapping F from \mathbf{E} to \mathbf{E}^2 defined by

$$F(z) = (X - z) \times (Y - z).$$

In standard variational-analytic language, a value of this mapping $(a, b) \in \mathbf{E}^2$ is critical exactly when F is not metrically regular for (a, b) at some point

$$z \in F^{-1}(a, b) = (X - a) \cap (Y - b).$$

A simple coderivative calculation (cf. [24]) shows that this is equivalent to the sets $X - a$ and $Y - b$ intersecting transversally at z . The semi-algebraic Sard theorem [20, 21] now shows that the semi-algebraic set of critical values $(a, b) \in \mathbf{E}^2$ has dimension strictly less than $2 \dim \mathbf{E}$. The result now follows by Fubini's theorem. \square

It is interesting to ask whether convergence of alternating projections is guaranteed (albeit sublinear) even without intrinsic transversality. This is particularly important, since in practice to check whether transversality holds requires knowledge of a point in the intersection. Easy examples show that, without intrinsic transversality, even limit points of alternating projection iterates may not lie in the intersection. Nevertheless, such pathologies do not occur in the semi-algebraic setting. A key tool for establishing this result is the following theorem based on [9, Theorem 14], a result growing out of the work of [23] extending the classical Łojasiewicz inequality for analytic functions to a nonsmooth semi-algebraic setting. See [20, Corollary 6.2] for a related version.

Theorem 7.2 (Kurdyka-Łojasiewicz inequality)

Consider a lower-semicontinuous semi-algebraic function $f: \mathbf{E} \rightarrow \overline{\mathbf{R}}$ and let U be a bounded open subset of the Euclidean space \mathbf{E} . Then for any value $\tau \in \mathbf{R}$ there exist a constant $\rho > 0$ and a continuous semi-algebraic function $\theta: (\tau, \tau + \rho) \rightarrow (0, +\infty)$, such that the inequality

$$|\overline{\nabla f}|(x) \geq \theta(f(x))$$

holds for all $x \in U$ with $\tau < f(x) < \tau + \rho$.

We now arrive at the main result of this section.

Theorem 7.3 (Semi-algebraic intersections)

Consider two nonempty closed semi-algebraic subsets X and Y of the Euclidean space \mathbf{E} , and suppose that X is bounded. If the method of alternating projections starts at a point in Y sufficiently close to X , then the distance of the iterates to $X \cap Y$ converges to zero, and consequently every limit point of the sequence lies in the intersection $X \cap Y$.

Proof Define a bounded open set

$$U = \{z : d(z, X) < 1\}.$$

Observe that the coupling function

$$\phi(x, y) = \delta_X(x) + |x - y| + \delta_Y(y),$$

is semi-algebraic. The Kurdyka-Łojasiewicz inequality above implies that there exists a constant $\rho \in (0, 1)$ and a continuous function $h: (0, \rho) \rightarrow (0, +\infty)$ such that all points $x \in X$ and $y \in Y \cap U$ with $0 < |x - y| < \rho$ satisfy the inequality

$$|\overline{\nabla\phi}|(x, y) \geq \sqrt{2} \cdot h(|x - y|),$$

and hence, using equation (6.6),

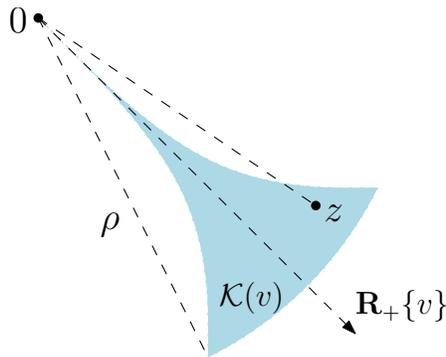
$$(7.7) \quad \max \{ |\overline{\nabla\phi_x}|(y), |\overline{\nabla\phi_y}|(x) \} \geq h(|x - y|).$$

Suppose the initial point $y_0 \in Y$ satisfies $d(y_0, X) < \rho$. The distance between successive alternating projection iterates $|x_n - y_n|$ is decreasing: we will prove its limit is 0. Arguing by contradiction, suppose in fact $|x_n - y_n| \downarrow \alpha \in (0, \rho)$.

Associated with the continuous function h , around any nonzero vector v , we define a radially symmetric open “cusp”

$$\mathcal{K}(v) := \{z \in \mathbf{E} : 0 < |z| < \rho \text{ and } d(\widehat{z}, \mathbf{R}_+v) < h(|z|)\},$$

where \mathbf{R}_+v is the ray generated by v . See the figure below for an illustration.



For each iteration $n = 1, 2, 3, \dots$, consider the normal vector $v_n = x_n - y_n \in N_Y(y_n)$. Note $|v_n| > \alpha$. Obviously the open set $\mathcal{K}(v_n)$ contains the point v_n : we next observe that in fact it also contains a uniform neighborhood. Specifically, we claim there exists a constant $\epsilon > 0$ such that $x - y_n \in \mathcal{K}(v_n)$ whenever $|x - x_n| \leq \epsilon$. Otherwise there would exist a sequence (x'_n) in \mathbf{E} satisfying

$$|x'_n - x_n| \rightarrow 0 \text{ and } x'_n - y_n \notin \mathcal{K}(v_n).$$

Notice that the distance between the vector $x'_n - y_n$ and the vector v_n converges to 0, so the same is true after we normalize them (normalization being a Lipschitz map on the set $\{v \in \mathbf{E} : |v| > \frac{\alpha}{2}\}$). But continuity of h now gives the contradiction

$$|\widehat{x'_n - y_n} - \widehat{v_n}| \geq d(\widehat{x'_n - y_n}, \mathbf{R}_+v_n) \geq h(|x'_n - y_n|) \rightarrow h(\alpha) > 0.$$

We deduce that any point $x \in X$ with $|x - y_n| < \rho$ and $|x - x_n| \leq \epsilon$ satisfies

$$\overline{|\nabla\phi_x|}(y_n) = d(\widehat{x - y_n}, N_Y(y_n)) \leq d(\widehat{x - y_n}, \mathbf{R}_+v_n) < h(|x - y_n|),$$

and hence, using inequality (7.7),

$$(7.8) \quad \overline{|\nabla\phi_{y_n}|}(x) \geq h(|x - y_n|).$$

Denote by β the minimum value of the strictly positive continuous function h on the interval $[\alpha, d(y_0, X)]$, so $\beta > 0$. Inequality (7.8) implies

$$d(\widehat{y_n - x}, N_X(x)) \geq \beta,$$

and hence, using Theorem 3.1 (Distance decrease),

$$|x_{n+1} - y_{n+1}| \leq |x_{n+1} - y_n| \leq |x_n - y_n| - \epsilon\beta,$$

giving a contradiction for large n . The result follows. \square

References

- [1] F. Andersson and M. Carlsson. Alternating projections on nontangential manifolds. *Constr. Approx.*, 38(3):489–525, 2013.
- [2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.
- [3] H.H. Bauschke and J.M. Borwein. On the convergence of von Neumann’s alternating projection algorithm for two sets. *Set-Valued Anal.*, 1(2):185–212, 1993.
- [4] H.H. Bauschke and J.M. Borwein. On the convergence of von Neumann’s alternating projection algorithm for two sets. *Set-Valued Anal.*, 1(2):185–212, 1993.
- [5] H.H. Bauschke, P.L. Combettes, and D.R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *J. Opt. Soc. Amer. A*, 19(7):1334–1345, 2002.
- [6] H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections: Applications. *Set-Valued and Variational Anal.*, pages 1–27, 2013.

- [7] H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections: Theory. *Set-Valued and Variational Anal.*, pages 1–43, 2013.
- [8] J. Bochnak, M. Coste, and M.-F. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1998. Translated from the 1987 French original, Revised by the authors.
- [9] J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optim.*, 18(2):556–572, 2007.
- [10] L.M. Bregman. The method of successive projection for finding a common point of convex sets. *Sov. Math. Dokl.*, 6:688–692, 1965.
- [11] X. Chen and M.T. Chu. On the least squares solution of inverse eigenvalue problems. *SIAM J. Numer. Anal.*, 33(6):2417–2430, 1996.
- [12] M.T. Chu. Constructing a Hermitian matrix from its diagonal entries and eigenvalues. *SIAM J. Matrix Anal. Appl.*, 16(1):207–217, 1995.
- [13] M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.
- [14] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Curves of descent. *Under review, arXiv:1212.1231 [math.OC]*, 2013.
- [15] K.M. Grigoriadis and E.B. Beran. Alternating projection algorithms for linear matrix inequalities problems with rank constraints. In *Advances in linear matrix inequality methods in control*, volume 2 of *Adv. Des. Control*, pages xx–xxi, 251–267. SIAM, Philadelphia, PA, 2000.
- [16] K.M. Grigoriadis and R.E. Skelton. Low-order control design for LMI problems using alternating projection methods. *Automatica J. IFAC*, 32(8):1117–1125, 1996.
- [17] L.G. Gubin, B.T. Polyak, and E.V. Raik. The method of projections for finding the common point of convex sets. *USSR Computational Mathematics and Mathematical Physics*, 7(6):1–24, 1967.
- [18] R. Hesse and D.R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *Under review, arXiv:1212.3349 [math.OC]*, 2012.
- [19] A.D. Ioffe. Metric regularity and subdifferential calculus. *Uspekhi Mat. Nauk*, 55(3(333)):103–162, 2000.

- [20] A.D. Ioffe. A Sard theorem for tame set-valued mappings. *J. Math. Anal. Appl.*, 335(2):882–901, 2007.
- [21] A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems. *Proc. Amer. Math. Soc.*, 136(9):3111–3119, 2008.
- [22] A.D. Ioffe. An invitation to tame optimization. *SIAM J. Optim.*, 19(4):1894–1917, 2009.
- [23] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)*, 48(3):769–783, 1998.
- [24] A.S. Lewis, D.R. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Found. Comput. Math.*, 9(4):485–513, 2009.
- [25] A.S. Lewis and J. Malick. Alternating projections on manifolds. *Math. Oper. Res.*, 33(1):216–234, 2008.
- [26] D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. *arXiv:1312.5681 [math.OA]*, 2013.
- [27] R. Orsi, U. Helmke, and J.B. Moore. A Newton-like method for solving rank constrained linear matrix inequalities. *Automatica J. IFAC*, 42(11):1875–1882, 2006.
- [28] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.
- [29] J.A. Tropp, I.S. Dhillon, R.W. Heath, Jr., and T. Strohmer. Designing structured tight frames via an alternating projection method. *IEEE Trans. Inform. Theory*, 51(1):188–209, 2005.
- [30] J. von Neumann. *Functional Operators. II. The Geometry of Orthogonal Spaces*. Annals of Mathematics Studies, no. 22. Princeton University Press, Princeton, N. J., 1950.
- [31] C.A. Weber and J.P. Allebach. Reconstruction of frequency-offset fourier data by alternating projectionso on constraint sets. *Proc. 24th Allerton Conf. Comm. Control and Comput.*, pages 194–203, 1986.
- [32] K. Yang and R. Orsi. Generalized pole placement via static output feedback: a methodology based on projections. *Automatica J. IFAC*, 42(12):2143–2150, 2006.