Transversality and Alternating Projections for Nonconvex Sets

D. Drusvyatskiy, A. D. Ioffe & A. S. Lewis

 Springer

Springer

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

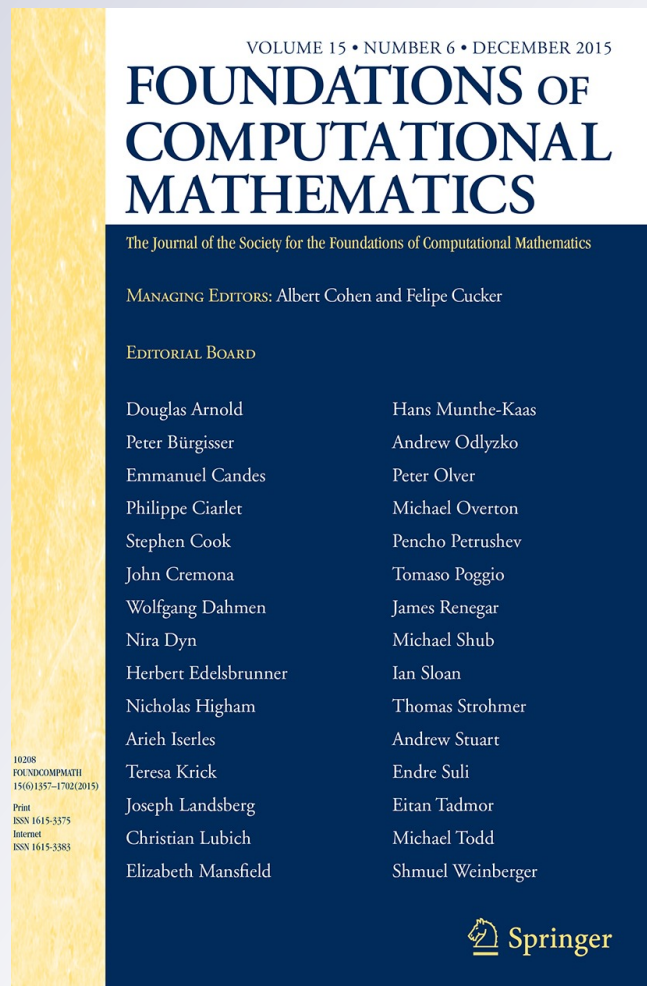The Journal of the Society for the Foundations of **Computational Mathematics**

CrossMark

Transversality and Alternating Projections for Nonconvex Sets

D. Drusvyatskiy[1] · A. D. Ioffe[2] · A. S. Lewis[3]

Abstract We consider the method of alternating projections for finding a point in the intersection of two closed sets, possibly nonconvex. Assuming only the standard transversality condition (or a weaker version thereof), we prove local linear convergence. When the two sets are semi-algebraic and bounded, but not necessarily transversal, we nonetheless prove subsequence convergence.

✉ A. S. Lewis
adrian.lewis@cornell.edu
http://people.orie.cornell.edu/aslewis

D. Drusvyatskiy
http://www.math.washington.edu/~ddrusv

A. D. Ioffe
ioffe@math.technion.ac.il

[1] Department of Mathematics, University of Washington, Seattle, WA 98195, USA

[2] Mathematics Department, Technion-Israel Institute of Technology, 32000 Haifa, Israel

[3] ORIE, Cornell University, Ithaca, NY 14853, USA

✦ Springer

1 Tranversality

The classical idea of *transversality* at an intersection point of two smooth manifolds in a Euclidean space $\mathbf{E}$ means that the two tangent spaces at the intersection point sum to $\mathbf{E}$: equivalently, their two orthogonal complements (the *normal* spaces) intersect trivially. Transversality assumptions are ubiquitous throughout analysis, guaranteeing in particular some stability for intersections under small perturbations. For two convex sets, for example, this stability requires that the intersection point lies on no separating hyperplane, a property also easy to express using normal vectors.

For more general sets that may be neither convex nor smooth, the correct generalization of transversality involves normals rather than tangents. Finite-dimensional variational analysis allows a unified treatment through a single, fundamental geometric construction: the normal cone (the monographs [8,10,30,35] are good references). In the space $\mathbf{E}$, we denote the inner product $\langle \cdot, \cdot \rangle$ and the norm $|\cdot|$. The set of points in a set $X \subset \mathbf{E}$ at minimum distance from a point $y \in \mathbf{E}$ is the *projection* $P_X(y)$, and that minimum distance is $d(y, X)$. For any point $x \in X$, vectors in the cone

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

are called *proximal normals* to X at x. Limits of proximal normals to X at points $x_n \in X$ approaching x are called *normals* and comprise the *normal cone* $N_X(x)$. When X is a smooth manifold, $N_X(x)$ is the classical normal space; when X is closed and convex, on the other hand, $N_X(x)$ consists of those linear functionals maximized over X by x.

By analogy with the classical case of manifolds, we say that two closed sets $X, Y \subset \mathbf{E}$ intersect *transversally* at a point $z \in X \cap Y$ when

$$N_X(z) \cap -N_Y(z) = \{0\}. \tag{1.1}$$

Central to many developments of variational analysis, this property is the pre-eminent theme, in particular, in [30,31]. The terminology, highlighting a fresh understanding of the classical analogy, is more recent: an early example is [10, p. 99].

In our entirely finite-dimensional setting, this definition of transversality is equivalent to several *metric regularity* properties [35], with profound consequences for stability, sensitivity, and robustness (and for algorithms [27]). Metric regularity of set-valued mappings is not our focus, but standard "coderivative" computations [23,28] prove the equivalence of property (1.1) with each of the following properties.

- The mapping $\Phi \colon \mathbf{E}^2 \rightrightarrows \mathbf{E}$ defined by $\Phi(x, y) = x - y$ for $x \in X$ and $y \in Y$ (and empty otherwise) is metrically regular at (z, z) for 0.
- The mapping $\Psi \colon \mathbf{E} \rightrightarrows \mathbf{E}^2$ defined by $\Psi(w) = (X - w) \times (Y - w)$ is metrically regular at z for $(0, 0)$.
- There exists a constant $\tau > 0$ such that

$$d(w, X' \cap Y') \ \le \ \tau\big(d(w, X') + d(w, Y')\big) \quad \text{for all } w \in \mathbf{E} \text{ near } z \tag{1.2}$$

and all small translations X', Y' of X, Y.

Transversality implies in particular the existence of a constant $\tau > 0$ such that (1.2) holds with $X' = X$ and $Y' = Y$. We call this weaker property *subtransversality*, because it corresponds to metric *sub*regularity [35] for the mappings Φ and Ψ. For an extended discussion of metric regularity and transversality, see [23].

Transversality is not just geometrically fundamental; it is also broad in applicability. As we discuss below, the property is strikingly unrestrictive in the class of sets to which it potentially applies, and, in concrete settings, it holds generically.

First, note that the transversality condition (1.1) implies little about the local structure of either of the two sets X and Y *individually*. In particular, it may hold without either set, considered alone, being "regular" in the standard single-set senses in the literature: for example, local convexity or smoothness, prox- or Clarke regularity [8,10,30,35], or super-regularity [28].[1] The epigraph

$$\{(x, \tau) \in \mathbf{E} \times \mathbf{R} : \tau \geq f(x)\}$$

of any locally Lipschitz function f intersects itself transversally at the point $(0, f(0))$, so in particular transversality is potentially applicable to any "epi-Lipschitz" set. In Sect. 3, we extend the reach of transversality further, to sets with empty interior.

Secondly, consider the broad but concrete class of *semi-algebraic* sets in the space $\mathbf{R}^n$. Such sets are finite unions of *basic* sets, each of which is a finite intersection of polynomial level sets $\{x : p(x) \leq 0\}$ and their complements [6,12]. Given any two closed semi-algebraic sets $X, Y \subset \mathbf{R}^n$, as we later prove, for almost all vectors $e \in \mathbf{E}$, transversality holds at every point in the intersection of the sets X and $Y - e$.

To summarize, transversality is a concise, fundamental, and widely studied geometric property, elegantly unifying classical notions, applicable to broad classes of sets, and commonly holding in concrete settings. Among its many powerful consequences, our aim here is to describe the first proof that transversality alone also guarantees local linear convergence for the method of alternating projections.

2 Alternating Projections

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 $x_n \in X$,

$$\text{choose} \quad y_n \in P_Y(x_n)$$
$$\text{choose} \quad x_{n+1} \in P_X(y_n).$$

This conceptually simple and widely used method for solving feasibility problems has been discovered and rediscovered by a number of authors, notably von Neumann [36]. We are particularly interested in *linear convergence* of the sequence of iterates,

---

[1] [34, Remark 10] states that "intrinsic transversality [the property, weaker than transversality, that we actually use to guarantee linear convergence] amalgamates transversality and regularity aspects". However, that statement concerns "regularity" aspects that involve *both* sets, like the idea of "linear regularity" in [3].

by which we mean that the sequence of distances from the iterates to the limit point is bounded above by a geometric sequence.

Convergence of the method for two intersecting closed *convex* sets is always guaranteed [9]. Linear convergence holds whenever the relative interiors of X and Y intersect [16], and more generally assuming just the global version of subtransversality [3].

Even for nonconvex sets, the method is popular: the first linear convergence guarantees in this setting appeared in [28,29], with extensions in [4,5,17] and another discussion in [1]. In particular, the main contribution in [28] was a proof that transversality at the point z suffices to guarantee linear convergence, starting sufficiently close to z, providing at least one of the sets X and Y is "super-regular" at z. Super-regularity is a nontrivial assumption, requiring the set to be not too nonconvex, and in particular implying Clarke regularity: it opens the door to a simple geometric proof. Super-regularity is, however, superfluous. Using a very different proof technique—the main topic of this paper—the authors derived the following striking theorem. We first discuss its provenance.

Theorem 2.1 *If two closed sets in a Euclidean space intersect transversally at a point, then the method of alternating projections, started nearby, converges linearly to a point in the intersection.*

The August 2013 PhD thesis of Drusvyatskiy seems to be the first reference for this result and its proof [13, Theorem 3.2.3], as well as the definition of "intrinsic" transversality on which it relies [13, Definition 3.2.1]. The result (without proof) was published in August 2014, in Lewis's lecture for the International Congress of Mathematicians [27, Theorem 2.1]. The initial version of the current work [14] was made public and submitted in January 2014.

A new and general framework for analyzing the convergence of nonconvex alternating projections (using properties of short sequences of iterates) was submitted for publication and disseminated in December 2013—see [32]. That manuscript did not show that transversality implies linear convergence (Theorem 2.1 above). Although result statements in a later revision [33, Corollary 7] (and the final published version [34]) do subsume Theorem 2.1, they do so by explicitly relying on our core argument in [14].[2] It is that argument that we present here.

In general, for two intersecting sets, no matter how close an initial point is to the intersection, the iterates generated by the method of alternating projections may not tend to the intersection. However, such pathologies rarely appear in practice. To end, we show how our basic technique proves that, if the two sets X and Y are semi-algebraic 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. Further regularity guarantees convergence of the sequence of iterates [34]. An analogous but independent result for averaged projections appears in [2].

---

[2] The proof of [33, Proposition 8] states: "Following entirely the argument in [14, p.6]…".

3 Intrinsic Transversality

Theorem 2.1 is not the last word on the local linear convergence of the method of alternating projections. It does not cover classically well-understood cases like two distinct intersecting lines in $\mathbf{R}^3$, a gap discussed in some recent work [1,4,5,17,34].

The approach in [17] depends, among other assumptions, on subtransversality (called there "local linear regularity"). Using a very different proof technique, we rely only on the following property, which we show to be intermediate in strength between transversality and subtransversality.

Definition 3.1 Two closed sets $X, Y \subset \mathbf{E}$ are *intrinsically transversal* at a common point if there exists an angle $\alpha > 0$ such that, nearby, any two points $x \in X \cap Y^c$ and $y \in Y \cap X^c$ cannot have difference $x - y$ simultaneously making an angle strictly less than α with both the cones $N_Y(y)$ and $-N_X(x)$.

(The normal cones may equivalently be replaced by proximal normal cones [14].)

While intrinsic transversality is widely applicable, we make no attempt at a detailed comparison with other variants of transversality. Some relationships may be found in [14], and extensive discussions appear in [34] and [24,25]. Our focus here is rather a clear and complete presentation of the new and subtle geometric argument opened up by the idea of intrinsic transversality, proving linear convergence.

Intrinsic transversality, as described in Definition 3.1, is a geometric idea. However, for the purposes of our proof, a variational perspective is more useful.

Using the normalization map $\widehat{}\colon \mathbf{E} \setminus \{0\} \to \mathbf{E}$, defined by $\widehat{z} = \frac{z}{|z|}$, we can rephrase the definition as follows. Two sets $X, Y \subset \mathbf{E}$ are intrinsically transversal at a common point if there exists a constant $\kappa \in (0, 1]$ (called a *constant of transversality*) such that nearby, any two points $x \in X \cap Y^c$ and $y \in Y \cap X^c$ have normalized displacement $u = \widehat{x - y}$ satisfying

$$\max \left\{ d\big(u, N_Y(y)\big), d\big(u, -N_X(x)\big) \right\} \geq \kappa. \tag{3.1}$$

With this in mind, we have the following easy observation.

Proposition 3.2 *Transversality implies intrinsic transversality. Indeed, if two closed sets $X, Y \subset \mathbf{E}$ intersect transversally at a point $z \in X \cap Y$, and $\theta \in (0, \pi]$ is the minimal angle between vectors in the normal cones $N_Y(z)$ and $-N_X(z)$, then intrinsic transversality holds with any constant in the interval $(0, \sin \frac{\theta}{2})$.*

Proof Clearly we have

$$\min_{|u|=1} \max \left\{ d\big(u, N_Y(z)\big), d\big(u, -N_X(z)\big) \right\} = \sin \frac{\theta}{2}.$$

(We define $\theta = \pi$ in the trivial case when either normal cone is $\{0\}$). Transversality simply means $\theta > 0$, and in that case an easy limiting argument shows that inequality (3.1) holds providing that $0 < \kappa < \sin \frac{\theta}{2}$. □

Several other common special cases are also covered under the umbrella of intrinsic transversality. Examples include two closed convex sets with intersecting relative interiors, and two arbitrary intersecting sets that are polyhedra (or finite unions of polyhedra).

We mention one particular classical generalization of transversality in this context (see [1]). Two smooth manifolds $X, Y \subset \mathbf{E}$ intersect *cleanly* at a point $z \in X \cap Y$ (see [18]) if the intersection $X \cap Y$ is a manifold and the tangent spaces at z satisfy

$$T_{X \cap Y}(z) = T_X(z) \cap T_Y(z).$$

One consequence is local linear convergence of the method of alternating projections, since, in fact, clean intersection implies intrinsic transversality. To see this, we note that, assuming clean intersection at z, there exists a diffeomorphism mapping z to zero and the two manifolds onto two linear subspaces [18, Proposition C.3.1]. However, intrinsic transversality always holds at zero for subspaces, and an easy exercise shows that it is preserved under diffeomorphisms, so the result follows.

4 Slope and The Coupling Function

Intrinsic transversality has its roots in a fundamental variational-analytic idea: *slope*. Denoting the extended reals $[-\infty, +\infty]$ by $\overline{\mathbf{R}}$, if a function $f : \mathbf{E} \to \overline{\mathbf{R}}$ is finite at a point $x \in \mathbf{E}$, then its slope there is

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

unless x is a local minimizer, in which case we set $|\nabla f|(x) = 0$. The *limiting slope* at x is

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

and when f is lower semi-continuous, we have the following relationship:

$$|\nabla f|(x) \geq \overline{|\nabla f|}(x) = d(0, \partial f(x)), \tag{4.1}$$

where ∂f denotes the set of subgradients [35]. For a proof, see [19, Propositions 1 and 2, Chapter 3], or, for example, [15, Proposition 4.6].

Now consider a "coupling" function for the two nonempty closed sets $X, Y \subset \mathbf{E}$. We define a function $\phi : \mathbf{E}^2 \to \overline{\mathbf{R}}$ 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 $\phi_y : \mathbf{E} \to \overline{\mathbf{R}}$ (for any fixed point $y \in \mathbf{E}$) defined

by $\phi_y(x) = \phi(x, y)$, and the marginal function $\phi_x : \mathbf{E} \to \overline{\mathbf{R}}$ (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 normalized displacement $u = \widehat{x - y}$,

$$\partial \phi_y(x) = u + N_X(x) \quad \text{and} \quad \partial \phi_x(y) = -u + N_Y(y),$$

and furthermore

$$\partial \phi(x, y) = \partial \phi_y(x) \times \partial \phi_x(y).$$

Using the relationship (4.1), we arrive at the following expressions for the limiting slopes:

$$\overline{|\nabla \phi_y|}(x) = d\big(u, -N_X(x)\big) \quad \text{and} \quad \overline{|\nabla \phi_x|}(y) = d\big(u, N_Y(y)\big).$$

Hence, the sets X and Y are intrinsically transversal at a common point z exactly when there exists a constant $\kappa \in (0, 1]$ such that, near z, for any two points $x \in X \cap Y^c$ and $y \in Y \cap X^c$, the limiting slopes $\overline{|\nabla \phi_y|}(x)$ and $\overline{|\nabla \phi_x|}(y)$ cannot both be strictly less than κ. The following result is an interesting alternative view.

Proposition 4.1 *Two closed sets $X, Y \subset \mathbf{E}$ are intrinsically transversal at a point $z \in X \cap Y$ if and only if the coupling function has slope uniformly bounded away from zero for pairs (x, y) with $x \in X \cap Y^c$ and $y \in Y \cap X^c$, both near z.*

Proof We note the relationship

$$\big(\overline{|\nabla \phi|}(x, y)\big)^2 = \big(\overline{|\nabla \phi_y|}(x)\big)^2 + \big(\overline{|\nabla \phi_x|}(y)\big)^2, \tag{4.2}$$

from which the result follows immediately. □

5 Error Bounds and Level Sets

We use the following key tool [19, Basic Lemma, Chapter 1], giving a slope-based criterion for a lower semi-continuous function f to have a nonempty level set

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

We note in passing that the result (and the proof we present) holds more generally, with $\mathbf{E}$ replaced by an arbitrary complete metric space.

Theorem 5.1 (Error bound) *Consider any lower semi-continuous function $f : \mathbf{E} \to \overline{\mathbf{R}}$, finite at the point $x \in \mathbf{E}$, and constants $\delta > 0$ and $\alpha < f(x)$ satisfying*

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

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 present two proofs. The first, suitable in any complete metric space, applies the Ekeland variational principle to the function $g : \mathbf{E} \to \overline{\mathbf{R}}$, defined as the positive part of the function $f - \alpha$. We deduce, for any constant $\gamma \in \left(\frac{1}{\delta}(f(x) - \alpha), K\right)$, 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.

A second, more elementary proof dispenses with the Ekeland principle when $\mathbf{E}$ is a Euclidean space. Assuming as before $\frac{1}{\delta}(f(x) - \alpha) < \gamma < K$, define a set

$$W = \{w \in \mathbf{E} : f(w) \leq f(x) - \gamma |w - x|\}.$$

The value $\beta = \inf_W f$ satisfies $\beta \leq f(x)$. Any sequence (w_r) in W with $f(w_r) \to \beta$ must be bounded, and any limit point $\bar{w}$ satisfies $\beta = f(\bar{w}) \leq f(x) - \gamma |\bar{w} - x|$.

We claim $\beta \leq \alpha$. If not, then $\alpha < f(\bar{w}) \leq f(x)$ and

$$|\bar{w} - x| \leq \frac{f(x) - f(\bar{w})}{\gamma} < \frac{f(x) - \alpha}{\gamma} < \delta,$$

so by definition $K \leq |\nabla f|(\bar{w})$. Hence, by definition of the slope, there is a point $w \in \mathbf{E}$ satisfying

$$f(w) < f(\bar{w}) - \gamma |\bar{w} - w| < \beta.$$

Furthermore, the central expression is bounded above by $f(x) - \gamma(|\bar{w} - x| + |\bar{w} - w|)$, and hence by $f(x) - \gamma |w - x|$, so $w \in W$, contradicting the definition of β.

We conclude $\beta \leq \alpha$, so we have constructed a point $\bar{w}$ in the level set $[f \leq \alpha]$, as required. To see the distance estimate, notice (as in the first proof) that we can assume $f \geq \alpha$, by replacing f by $\max\{f, \alpha\}$. In that case, we have in fact proved $\beta = \alpha$ and hence $|\bar{w} - x| \leq \frac{1}{\gamma}(f(x) - \alpha)$. Now let γ approach K as before. □

Our main result follows quickly from the following appealing geometric consequence of Theorem 5.1. 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 $|x - y|$ 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 5.2 (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*

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

then

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

Proof We apply Theorem 5.1 (Error bound) to the function $f = |\cdot - y| + \delta_X$. Inequality (4.1) and a repeat of our calculation of $\overline{|\nabla \phi_y|}$ in the previous section show

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

In the notation of Theorem 5.1, assuming $\alpha < f(x)$, we have

$$\mu \leq K = \inf_{w \in X} \big\{ |\nabla f|(w) : \alpha < |w - y| \leq |x - y|, |w - x| \leq \delta \big\}.$$

Hence, providing α satisfies $\mu > \frac{1}{\delta}(|x - y| - \alpha)$, we deduce the first conclusion of Theorem 5.1: the level set $[f \leq \alpha]$ is nonempty, or in other words $d(y, X) \leq \alpha$. The result now follows. $\qquad\square$

6 Main Result

The following is our main result.

Theorem 6.1 (Linear convergence) *If two closed sets $X, Y \subset \mathbf{E}$ are intrinsically transversal at a point $z \in X \cap Y$, with constant $\kappa > 0$, then, for any constant c in the interval $(0, \kappa)$, the method of alternating projections, initiated sufficiently near z, converges to a point in the intersection $X \cap Y$ with R-linear rate $1 - c^2$.*

Proof 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

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

Now consider any point $x \in X$, close to z, 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

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

since $x - y \in N_Y(y)$. To bound the right-hand side, note that any nonzero vectors $p, q \in \mathbf{E}$ satisfy the inequalities

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

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

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

and hence, using intrinsic transversality

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

We next apply Theorem 5.2 with $\delta = c\rho$. In the notation of that result, we have $\mu \geq \kappa > c$, so we deduce

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

as required. The result now follows. $\qquad\square$

Theorem 2.1 now follows, by applying Proposition 3.2. Indeed, if two closed sets $X, Y \subset \mathbf{E}$ intersect transversally at a point $z \in X \cap Y$, and the minimal angle between vectors in the closed cones $N_Y(z)$ and $-N_X(z)$ is $\theta \in (0, \pi]$, then for any constant $r \in (\cos^2 \frac{\theta}{2}, 1)$, the method of alternating projections, initiated sufficiently near z, must converge to a point in $X \cap Y$ with R-linear rate r.

Another consequence of the proof technique is the following result.

Theorem 6.2 *Intrinsic transversality implies subtransversality.*

Proof Under the assumptions of Theorem 6.1, consider any point $x \in X$ near the intersection point z, and any point $y \in P_Y(x)$. We proved $d(y, X) \leq (1 - c^2)d(x, Y)$, so by induction, alternating projections converges to a limit in the intersection $X \cap Y$ whose distance from x is no more than

$$d(x, Y) \sum_{n=0}^{\infty} (1 - c^2)^n.$$

Hence, $d(x, X \cap Y) \leq c^{-2} d(x, Y)$.

Now consider any point $w \in \mathbf{E}$ near the intersection point z, and any point $x \in P_X(w)$. The triangle inequality implies

$$d(x, Y) \leq |x - w| + d(w, Y) = d(w, X) + d(w, Y).$$

and hence

$$\begin{aligned}
d(w, X \cap Y) &\leq |w - x| + d(x, X \cap Y) \\
&\leq d(w, X) + c^{-2} d(x, Y) \\
&\leq (1 + c^{-2})\big(d(w, X) + d(w, Y)\big),
\end{aligned}$$

which proves subtransversality. $\qquad\square$

Transversality is a strictly stronger condition than intrinsic transversality, as shown by two intersecting lines in $\mathbf{R}^3$. In turn, intrinsic transversality is strictly stronger than subtransversality. Indeed, subtransversality alone does not even guarantee local convergence of alternating projections, as the following example shows.

Consider the two closed sets in $\mathbf{R}^2$

$$X = \big\{(x, 0) : x \geq 0\big\}$$

$$Y = \big\{(0, 0)\big\} \cup \bigcup_{n=-\infty}^{\infty} \Big\{\big(x, \tfrac{1}{2}(3^{n+1} - x)(3^{-n}x - 1)\big) : 3^n \leq x \leq 3^{n+1}\Big\}.$$

Geometrically, X is just the nonnegative x-axis, and Y consists of sequence of parabolic arches based on the x-axis and scaling linearly down toward the origin. Subtransversality at the origin is routine to check. However, intrinsic transversality fails: for any integer n, the point $(2 \cdot 3^n, 0) \in X$ projects onto the point $(2 \cdot 3^n, \tfrac{1}{2}3^n) \in Y$ and vice versa, so alternating projections cycles at points arbitrarily close to the origin.

7 Semi-algebraic Intersections

In this last section, we turn to the case where the two sets are semi-algebraic. Alternating projections on algebraic varieties, in particular, was considered in [1].

Classically (and reassuringly) two smooth manifolds "typically" intersect transversally: almost all translations to the manifolds satisfy transversality throughout the intersection. We first show the same result for semi-algebraic sets. See [22] for the role of semi-algebraic ideas in nonsmooth optimization, and for the more general "tame" setting, to which all results of this section extend.

Theorem 7.1 (Generic transversality) *Consider two closed semi-algebraic sets $X, Y \subset \mathbf{E}$. Then, for almost every vector e in $\mathbf{E}$, transversality holds at every point in the (possibly empty) 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 [20,21], we define a value of this mapping $(a, b) \in \mathbf{E}^2$ as *critical* when F is not metrically regular for (a, b) at some point

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

By a simple coderivative calculation (cf. [28, p. 491]), this is equivalent to the sets $X - a$ and $Y - b$ not 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. □

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. As the example in the previous section shows, 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 [7, Theorem 14], a result growing out of the work of [26] extending the classical Łojasiewicz inequality for analytic functions to a nonsmooth semi-algebraic setting. See [22, Corollary 6.2] for a related version.

Theorem 7.2 (Kurdyka–Łojasiewicz inequality) *Given a semi-algebraic lower semi-continuous function* $f : \mathbf{E} \rightarrow \overline{\mathbf{R}}$, *consider a bounded open set* $U \subset \mathbf{E}$. *Then, for any value* $\tau \in \mathbf{R}$, *there exists a constant* $\rho > 0$ *and a continuous strictly increasing semi-algebraic function* $\theta : (\tau, \tau + \rho) \rightarrow (0, +\infty)$, *such that the inequality*

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

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 sets* $X, Y \subset \mathbf{E}$, *with* X *bounded. If the method of alternating projections starts in* Y *and near* X, *then the distance of the iterates to the intersection* $X \cap Y$ *converges to zero, and hence every limit point lies in* $X \cap Y$.

Proof Define a bounded open set

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

The coupling function ϕ from Sect. 4 is semi-algebraic, so 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

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

From Eq. (4.2), we deduce

$$\max\left\{\overline{|\nabla \phi_x|}(y), \overline{|\nabla \phi_y|}(x)\right\} \geq h(|x - y|). \tag{7.1}$$

Suppose the initial point $y_0 \in Y$ satisfies $d(y_0, X) < \rho$. The distance between successive iterates $|x_n - y_n|$ is decreasing: our aim is to prove its limit is zero. By way of 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"

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

where $\mathbf{R}_+ v$ is the ray generated by v. For each iteration $n = 1, 2, 3, \ldots$, 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

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

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

$$\left| \widehat{x_n' - y_n} - \widehat{v_n} \right| \geq d\left(\widehat{x_n' - y_n}, \mathbf{R}_+ v_n \right) \geq h\left(|x_n' - y_n| \right) \to h(\alpha) > 0.$$

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

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

and hence, using inequality (7.1),

$$\overline{|\nabla \phi_{y_n}|}(x) \geq h\left(|x - y_n| \right). \tag{7.2}$$

Denote by β the minimum value of the strictly positive continuous function h on the interval $\left[\alpha, d(y_0, X) \right]$, so $\beta > 0$. Inequality (7.2) implies

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

and hence, using Theorem 5.2 (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. $\qquad\square$

We remark that the technique we present here applies to much broader classes of sets—those "definable in an o-minimal structure" [11]. Under a further regularity condition, [34, Corollary 9] proves convergence of the full alternating projection sequence in the semi-algebraic (or, more broadly, subanalytic) setting.

References

1. F. Andersson and M. Carlsson. Alternating projections on nontangential manifolds. *Constructive Approximation*, 38:489–525, 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. *Mathematics of Operations Research*, 35:438–457, 2010.

3. H.H. Bauschke and J.M. Borwein. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1:185–212, 1993.

4. 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 Analysis*, 21:475–501, 2013.

5. 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 Analysis*, 21:431–473, 2013.

6. J. Bochnak, M. Coste, and M.-F. Roy. *Real Algebraic Geometry*. Springer, Berlin, 1998.

7. J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18:556–572, 2007.

8. J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis*. Springer, New York, 2005.

9. L.M. Bregman. The method of successive projection for finding a common point of convex sets. *Soviet Mathematics Doklady*, 6:688–692, 1965.

10. F.H. Clarke, Yu. Ledyaev, R.I. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer, New York, 1998.

11. M. Coste. *An Introduction to O-Minimal Geometry*. RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.

12. M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

13. D. Drusvyatskiy. *Slope and Geometry in Variational Mathematics*. PhD thesis, School of Operations Research and Information Engineering, Cornell University, August 2013.

14. D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Alternating projections and coupling slope. arXiv:1401.7569v1, January 2014.

15. D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Curves of descent. *SIAM Journal on Control and Optimization*, 53:114–138, 2015.

16. L.G. Gubin, Polyak B.T., and Raik E.V. The method of projections for finding the common point of convex sets. *USSR Computational Mathematics and Mathematical Physics*, 7:1–24, 1967.

17. R. Hesse and D.R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization*, 23:2397–2419, 2013.

18. L. Hörmander. *The Analysis of Linear Partial Differential Operators*, volume III. Springer, Berlin, 1985.

19. A.D. Ioffe. Metric regularity and subdifferential calculus. *Uspekhi Matematicheskikh Nauk*, 55:103–162, 2000.

20. A.D. Ioffe. A Sard theorem for tame set-valued mappings. *Journal of Mathematical Analysis and Applications*, 335:882–901, 2007.

21. A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems. *Proceedings of the American Mathematical Society*, 136:3111–3119, 2008.

22. A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19:1894–1917, 2009.

23. A.D. Ioffe. Metric regularity. theory and applications—a survey. arXiv:1505.07920, May 2015.

24. A.Y. Kruger and N.H. Thao. Quantitative characterizations of regularity properties of collections of sets. *Journal of Optimization Theory and Applications*, 164:41–67, 2015.

25. A.Y. Kruger and N.H. Thao. Regularity of collections of sets and convergence of inexact alternating projections. arXiv:1501.04191, 2015.

26. K. Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier (Grenoble)*, 48:769–783, 1998.

27. A.S. Lewis. Nonsmooth optimization: conditioning, convergence and semi-algebraic models. In S.Y. Jang, Y.R. Kim, D.-W. Lee, and I. Yie, editors, *Proceedings of the International Congress of Mathematicians, Seoul*, volume IV: Invited Lectures, pages 872–895, Seoul, Korea, 2014. Kyung Moon Sa.

28. A.S. Lewis, D.R. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Foundations of Computational Mathematics*, 9:485–513, 2009.

29. A.S. Lewis and J. Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33:216–234, 2008.

30. B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Springer, Berlin, 2006.

31. B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications*. Springer, Berlin, 2006.

32. D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. arXiv:1312.5681v1, December 2013.
33. D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. arXiv:1312.5681v2, September 2014.
34. D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. *Foundations of Computational Mathematics*, 2015. doi:10.1007/s10208-015-9253-0.
35. R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.
36. J. von Neumann. *Functional Operators. II. The Geometry of Orthogonal Spaces*. Annals of Mathematics Studies, 22. Princeton University Press, Princeton, NJ, 1950.