Local linear convergence for inexact alternating projections on nonconvex sets

D. Drusvyatskiy [*]   A.S. Lewis [†]

March 22, 2019

*Dedicated to our friend, colleague, and inspiration, Alex Ioffe,
on the occasion of his 80th birthday.*

Abstract

Given two arbitrary closed sets in Euclidean space, to guarantee that the method of alternating projections converges locally at linear rate to a point in the intersection, a simple transversality assumption suffices. Exact projection onto nonconvex sets is typically intractable, but we show that computationally-cheap inexact projections may suffice instead. In particular, if one set is defined by sufficiently regular smooth constraints, then projecting onto the approximation obtained by linearizing those constraints around the current iterate suffices. On the other hand, if one set is a smooth manifold represented through local coordinates, then the approximate projection resulting from linearizing the coordinate system around the preceding iterate on the manifold also suffices.

Key words: Alternating projections, Linear convergence, Variational analysis, Transversality, Approximate projection

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

1 Nonconvex alternating projections

The method of alternating projections for finding a point in the intersection of two closed convex sets enjoys an enduring popularity. A good exposition is [1]. A suitable regularity or transversality condition ensures local linear convergence, even for nonconvex sets. With a few exceptions (notably including manifolds of fixed-rank matrices), exact projection—finding exact nearest points—for nonconvex sets is practically unrealistic. Nonetheless, the alternating projection philosophy remains appealing because we can often approximate the projection cheaply without losing linear convergence.

We present two examples, both feasibility problems involving two closed sets with distinct roles: an "easier" set Q, and a second set M. For simplicity, we assume that we can compute exact nearest points in Q, which might, for example, be convex and semidefinite-representable. On the other hand, we rely only on relaxed projections onto M, projecting instead onto approximations of M using information from previous iterates. Earlier general results on inexact alternating projection schemes, including [5, Thm 6.1] and [4], typically emerge naturally from the convergence theory for the exact algorithm, but often do not apply immediately to common feasibility problems. Here, we therefore pursue a more transparent approach, motivated directly by the models.

The first example concerns the constraint system

$$\begin{cases} G(x) & \leq & 0 \\ H(x) & = & 0 \\ x & \in & Q, \end{cases}$$

for C^2-smooth maps G (into $\mathbf{R}^n$) and H. For this problem, we consider the following algorithm, which involves just projections onto Q and various polyhedra.

Algorithm 1.1
   while $x \in Q$ not feasible **do**
      choose minimal norm s satisfying
         $G(x) + \nabla G(x)s \leq 0$ and $H(x) + \nabla H(x)s = 0$;
      choose $x_+ \in Q$ minimizing $|x + s - x_+|$;
      $x = x_+$;
   end while

The second example concerns the system

$$F(x) \in Q,$$

for a C^2-smooth map F. Now we consider the following algorithm.

Algorithm 1.2
  while x not feasible **do**
    choose $y \in Q$ minimizing $|y - F(x)|$;
    choose s minimizing $|F(x) + \nabla F(x)s - y|$;
    $x = x + s$;
  end while

By interpreting each algorithm as an approximate alternating projection scheme, we derive reasonable conditions for local linear convergence to a feasible solution. We consider two distinct notions of inexact or approximate projections, attuned to the two distinct model problems.

2   The geometry of alternating projections

We now describe our setting more formally. Consider a Euclidean space $\mathbf{X}$ with norm $|\cdot|$. Given a nonempty closed set $Q \subset \mathbf{X}$, we define the *distance function* $d_Q \colon \mathbf{X} \to \mathbf{R}$ by

$$d_Q(z) \;=\; \min \big\{ |x - z| : x \in Q \big\},$$

for any point $z \in \mathbf{X}$. The *projection* $P_Q(z)$ consists of the attaining (or *nearest*) points x: it is a nonempty closed set, but may not be a singleton if Q is nonconvex. Given a second nonempty closed set $M \subset \mathbf{X}$, the *method of alternating projections* is the following iteration.

Algorithm 2.1 (Alternating projections)
  choose $z \in Q$;
  while $z \notin M$ **do**
    choose $x \in P_M(z)$;
    choose $z' \in P_Q(x)$;
    $z = z'$;
  end while

Consider any point of intersection $\bar{x} \in Q \cap M$ satisfying the transversality condition

$$(2.2) \qquad\qquad N_Q(\bar{x}) \cap -N_M(\bar{x}) \;=\; \{0\},$$

where the notation $N_Q(\cdot)$ denotes the (limiting) normal cone to Q. (We refer the reader to [10] for standard terminology in variational analysis.) Then, as shown in [2], providing the initial point $z^0 \in Q$ is sufficiently close to $\bar{x}$, the generated sequence of points $z^0, z^1, z^2, \ldots$ in Q converges linearly to a point $\hat{z} \in Q \cap M$: there exist positive constants $c < 1$ (a "rate" determined by a quantitative version of the transversality condition (2.2)) and $\rho > 0$ such that $|z^k - \hat{z}| < \rho c^k$ for all iterations $k = 0, 1, 2, \ldots$. In fact just a weaker notion of *intrinsic transversality* [2] suffices.

3

This convergence result is striking in its generality, assuming nothing about the closed sets Q and M beyond their transversal intersection. This underlying assumption is subtle, but in fact a more intuitive assumption suffices, under a modest geometric assumption about one of the sets.

Specifically, consider the following idea from [5].

Definition 2.3 A set $Q \subset \mathbf{X}$ is *super-regular* at a point $\bar{x}$ if, given any angle $\gamma > 0$, for any points $z \in Q$ and $x \notin Q$ sufficiently close to $\bar{x}$, and any point $z' \in P_Q(x)$ with $z' \neq z$, the angle between the vectors $z - z'$ and $x - z'$ is at least $\frac{\pi}{2} - \gamma$.

In particular, closed convex sets are everywhere super-regular, as are manifolds defined by C^1-smooth equations with linearly independent gradients.

Somewhat stronger but more standard than super-regularity is the notion of prox-regularity: the set Q is *prox-regular* at a point $\bar{x} \in Q$ if the projection P_Q is single-valued near $\bar{x}$ (see [9]). In that case P_Q is also Lipschitz near $\bar{x}$, and satisfies the following characterization:

$$(2.4) \qquad P_Q(z) = x \quad \Leftrightarrow \quad z - x \in N_Q(x)$$

for all points $z, x \in \mathbf{X}$ near $\bar{x}$. Furthermore, Q is then super-regular at $\bar{x}$, as shown in [5].

Now consider the following simple geometric condition (called "0-separability" in [6]), specifically designed with alternating projections in mind.

Definition 2.5 A set $Q \subset \mathbf{X}$ intersects a set $M \subset \mathbf{X}$ *separably* at a point $\bar{x} \in M \cap Q$ if there exists an angle $\alpha > 0$ such that, for any point $z \in Q \setminus M$ sufficiently close to $\bar{x}$, and any points $x \in P_M(z) \setminus Q$ and $z' \in P_Q(x)$, the angle between the vectors $z - x$ and $z' - x$ is at least α.

It is easy to check that this property follows from intrinsic transversality, and hence in particular from transversality. Notice that, unlike those conditions, the separable intersection property is not symmetric in the sets M and Q.

These properties together imply linear convergence of alternating projections, beginning close to the intersection. To see this, suppose that both Definition 2.3 and Definition 2.5 hold: the set Q intersects the set M separably at the point $\bar{x}$ and is super-regular there. Now recall the following simple geometric argument, identical to the approach in [5], and extended further in [6]. Consider one iteration of the algorithm, starting from a point z, and generating the points x and z' in turn. Given any angle $\gamma > 0$, the triangle near the intersection point $\bar{x}$ defined by the three points z, x and z' has angle at least α at x (by separability) and at least $\frac{\pi}{2} - \gamma$ at z' (by super-regularity), and hence at most $\frac{\pi}{2} + \gamma - \alpha$ at z. The Sine Rule shows

$$\frac{|z' - x|}{\sin(\frac{\pi}{2} + \gamma - \alpha)} \leq \frac{|z - x|}{\sin(\frac{\pi}{2} - \gamma)}.$$

4

Now fix any constant τ in the interval $(\cos\alpha, 1)$. By fixing γ sufficiently small, we deduce that the next point generated by the algorithm, $x' \in P_M(z')$, satisfies

$$(2.6) \qquad |z' - x'| \ \leq \ |z' - x| \ \leq \ \tau|z - x|.$$

An easy induction argument now shows linear convergence of the iterate sequence $z_0, z_1, z_2, \dots$ to a point in the intersection $Q \cap M$. The rate is $\sqrt{\tau}$, which we can make arbitrarily close to $\sqrt{\cos\alpha}$.

3 Inexact projections

In this work we consider to what extent we can relax the projection operations while still retaining the geometry underlying the simple linear convergence argument. We first consider the following generalization of alternating projections, which allows small errors in the computation of the projection $P_M(z)$.

Algorithm 3.1 (Inexact alternating projections)
    choose $z \in Q$ and $\epsilon \geq 0$;
    while $z \notin M$ **do**
        choose $x \in \mathbf{X}$ with $d_{P_M(z)}(x) \leq \epsilon d_M(z)$;
        choose $z_+ \in P_Q(x)$;
        $z = z_+$;
    end while

When the tolerance ϵ is zero, we recover the exact alternating projections scheme. When $\epsilon > 0$, however, we no longer even require the point x to lie in the set M.

In what follows we suppose that the set Q intersects the set M separably at a point $\bar{x}$, as holds in particular under the usual transversality assumption (2.2). We suppose furthermore that Q is prox-regular (and hence super-regular) at $\bar{x}$. We make the following definition.

Definition 3.2 A map $\Phi\colon \mathbf{X} \to \mathbf{X}$ is an *inexact projection* onto a closed set $M \subset \mathbf{X}$ around a point $\bar{x} \in M$ if
$$d_{P_M(z)}\big(\Phi(z)\big) = o\big(d_M(z)\big)$$
as $z \to \bar{x}$.

We then have the following result.

Theorem 3.3 *Suppose that a closed set Q intersects a closed set M separably at a point $\bar{x}$, and is prox-regular there. If Φ is an inexact projection onto M around $\bar{x}$, then starting from any nearby point $z \in Q$, the inexact alternating projection iteration*
$$z \leftarrow P_Q\big(\Phi(z)\big)$$
converges linearly to a point in the intersection $M \cap Q$.

Proof We claim in fact the following: providing that the tolerance ϵ is sufficiently small, starting Algorithm 3.1 at any point $z \in Q$ near $\bar{x}$ results in linear convergence to a point in the intersection $Q \cap M$. To see this, notice that at each iteration there exists a point $x' \in P_M(z)$ satisfying $|x - x'| \leq \epsilon|z - x'|$. Denote by $x'_+ \in P_M(z_+)$ a corresponding such point at the next iteration, and let $z' = P_Q(x')$. As before, after initially fixing any constant τ in the interval $(\cos\alpha, 1)$, we deduce $|z' - x'| \leq \tau|z - x'|$. Furthermore, if we fix a local Lipschitz constant $\kappa > 0$ for the projection P_Q, then we have

$$|z_+ - z'| = |P_Q(x) - P_Q(x')| \leq \kappa|x - x'| \leq \kappa\epsilon|z - x'|.$$

An immediate consequence is

$$|z_+ - x'_+| \leq |z_+ - x'| \leq |z_+ - z'| + |z' - x'| \leq (\kappa\epsilon + \tau)|z - x'|.$$

Again, providing $\epsilon < \frac{1-\tau}{\kappa}$, we deduce linear convergence of the iterates z to a point in the intersection $Q \cap M$. $\qquad\square$

In this proof, notice that, as we reduce the tolerance ϵ to zero, the convergence rate of the approximate algorithm grows arbitrarily close to that of the exact alternating projection scheme.

4   Inexact projection onto constraint sets

In this section we show how to calculate computationally-cheap inexact projections for constraint sets of the form

(4.1) $$M = \{x \in \mathbf{X} : G(x) \leq 0,\ P(x) \leq 0,\ H(x) = 0\},$$

for maps $G \colon \mathbf{X} \to \mathbf{R}^n$ and $H \colon \mathbf{X} \to \mathbf{Y}$, where $\mathbf{Y}$ is a second Euclidean space. We focus on a neighborhood of a point $\bar{x} \in M$ where $P(\bar{x}) < 0$: nearby, the map P models the inactive constraints. The assumption that the maps G, P, and H are everywhere defined is just for notational convenience. We make the following classical assumption about the derivative maps $\nabla G(\bar{x}) \colon \mathbf{X} \to \mathbf{R}^n$ and $\nabla H(\bar{x}) \colon \mathbf{X} \to \mathbf{Y}$ and their adjoints $\nabla G(\bar{x})^* \colon \mathbf{R}^n \to \mathbf{X}$ and $\nabla H(\bar{x})^* \colon \mathbf{Y} \to \mathbf{X}$.

Assumption 4.2 (Linear independence constraint qualification)
The maps G, P and H are C^2-smooth around a point $\bar{x} \in \mathbf{X}$ satisfying $G(\bar{x}) = 0$, $P(\bar{x}) < 0$, and $H(\bar{x}) = 0$, and they satisfy the condition

$$\nabla G(\bar{x})^*w + \nabla H(\bar{x})^*y = 0 \quad \Rightarrow \quad w = 0 \text{ and } y = 0.$$

Under these assumptions, the set M is strongly amenable at $\bar{x}$, in the sense of [7], and hence in particular prox-regular there [8, Proposition 2.5].

We can construct an inexact projection for the constraint set M as follows. For any point z in $\mathbf{X}$ near $\bar{x}$, we consider the projection of z onto the polyhedral approximation to the set M obtained by linearizing its defining constraints at z: in other words, we solve the convex quadratic program

(4.3)
$$\begin{cases} \text{minimize} & \frac{1}{2}|x - z|^2 \\ \text{subject to} & G(z) + \nabla G(z)(x - z) \leq 0 \\ & P(z) + \nabla P(z)(x - z) \leq 0 \\ & H(z) + \nabla H(z)(x - z) = 0 \\ & x \in \mathbf{X}. \end{cases}$$

As a consequence we deduce, under reasonable conditions, local linear convergence of the first algorithm we presented in the introduction.

Our aim, then, is to prove the following result.

Theorem 4.4 *Consider the constraint set M defined by equation (4.1). Suppose the linear independence constraint qualification (Assumption 4.2) holds. For points $z \in \mathbf{X}$ near the point $\bar{x}$, denote by $\Phi(z)$ the projection onto the linearized constraints, so $\Phi(z)$ is the unique optimal solution of the quadratic program (4.3). Then, the map Φ is an inexact projection onto M around $\bar{x}$:*

$$|\Phi(z) - P_M(z)| = O(d_M(z)^2) \quad as \ z \to \bar{x}.$$

Proof Near $\bar{x}$, the constraint involving the continuous map P has no impact on the set M, so for the time being we ignore it. Denote the nonnegative orthant in $\mathbf{R}^n$ by $\mathbf{R}^n_+$, and the set of complementary pairs of vectors by

$$\Omega = \{(w, s) \in \mathbf{R}^n_+ \times \mathbf{R}^n_+ : \langle w, s \rangle = 0\}.$$

Then any point $x \in \mathbf{X}$ near $\bar{x}$ and vector $p \in \mathbf{X}$ satisfies $x \in M$ and $p \in N_M(x)$ if and only if there exists a pair $(w, s) \in \Omega$ and a vector $y \in \mathbf{Y}$ satisfying $G(x) + s = 0$ and $p = \nabla G(x)^* w + \nabla H(x)^* y$.

For any nearby point $z \in \mathbf{X}$, we can now characterize the unique projection $P_M(z)$ using the property (2.4). Consider the following set-valued mapping

$$\Psi \colon \mathbf{X} \times \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{Y} \rightrightarrows \mathbf{R}^n \times \mathbf{Y} \times \mathbf{X}$$

defined by setting $\Psi(x, w, s, y)$ to be the single point

$$\begin{pmatrix} G(x) + s \\ H(x) \\ x + \nabla G(x)^* w + \nabla H(x)^* y \end{pmatrix}^T$$

7

if $(w, s) \in \Omega$, and to be empty otherwise. Then, for all points z and x near the point $\bar{x}$, the condition

$$
\begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}^T \in \Psi(x, w, s, y) \quad \text{for some } w \in \mathbf{R}^n, \; s \in \mathbf{R}^n, \; y \in \mathbf{Y}
$$

is necessary and sufficient for $x = P_M(z)$.

We claim that the map Ψ is metrically regular at $(\bar{x}, 0, 0, 0)$ for $(0, 0, \bar{x})$. To see this, we linearize around the point $\bar{x}$ to obtain the following new set-valued mapping $\hat{\Psi} \colon \mathbf{X} \times \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{Y} \rightrightarrows \mathbf{R}^n \times \mathbf{Y} \times \mathbf{X}$. Define the set $\hat{\Psi}(x, w, s, y)$ to be the single point

$$
\begin{pmatrix} \nabla G(\bar{x})(x - \bar{x}) + s \\ \nabla H(\bar{x})(x - \bar{x}) \\ x + \nabla G(\bar{x})^* w + \nabla H(\bar{x})^* y \end{pmatrix}^T
$$

if $(w, s) \in \Omega$, and to be empty otherwise. For $(x, w, s, y) \in \mathbf{X} \times \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{Y}$, we have

$$
(\Psi - \hat{\Psi})(x, w, s, y) \;=\; \begin{pmatrix} G(x) - \nabla G(\bar{x})(x - \bar{x}) \\ H(x) - \nabla H(\bar{x})(x - \bar{x}) \\ \big(\nabla G(x) - \nabla G(\bar{x})\big)^* w + \big(\nabla H(x) - \nabla H(\bar{x})\big)^* y \end{pmatrix}^T .
$$

The map $\Psi - \hat{\Psi}$ is therefore continuously differentiable around the point $(\bar{x}, 0, 0, 0)$, with zero derivative. Hence, using the coderivative criterion [10], metric regularity of the maps Ψ and $\hat{\Psi}$ at $(\bar{x}, 0, 0, 0)$ for $(0, 0, \bar{x})$ are equivalent.

For all $(u, v, z) \in \mathbf{R}^n \times \mathbf{Y} \times \mathbf{X}$ and $x \in \mathbf{X}$, the condition

$$
\begin{pmatrix} u \\ v \\ z \end{pmatrix}^T \in \hat{\Psi}(x, w, s, y) \quad \text{for some } w \in \mathbf{R}^n, \; s \in \mathbf{R}^n, \; y \in \mathbf{Y}
$$

is just the necessary and sufficient first-order condition for x to be an optimal solution for the convex quadratic program

$$
\begin{cases} \text{minimize} & \frac{1}{2}|x - z|^2 \\ \text{subject to} & \nabla G(\bar{x})x \;\leq\; \nabla G(\bar{x})\bar{x} + u \\ & \nabla H(\bar{x})x \;=\; \nabla H(\bar{x})\bar{x} + v \\ & x \;\in\; \mathbf{X}. \end{cases}
$$

This problem is feasible, by the linear independence constraint qualification, and hence has a unique optimal solution $\xi(u, v, z)$, which satisfies the first-order condition. Hence the set $\hat{\Psi}^{-1}(u, v, z)$ is nonempty, and every element $(x, w, s, y) \in$

$\hat{\Psi}^{-1}(u, v, z)$ must satisfy

$$
\begin{aligned}
x &= \xi(u, v, z) \\
s &= u - \nabla G(\bar{x})(x - \bar{x}) \\
\nabla G(\bar{x})^* w + \nabla H(\bar{x})^* y &= z - x.
\end{aligned}
$$

Thus the point $z - \xi(u, v, z)$ must lie in the range $\mathbf{S} = \nabla G(\bar{x})^* \mathbf{R}^n + \nabla H(\bar{x})^* \mathbf{Y}$. By the linear independence constraint qualification, the linear map

$$
(w, y) \mapsto \nabla G(\bar{x})^* w + \nabla H(\bar{x})^* y
$$

has a linear inverse $K \colon \mathbf{S} \to \mathbf{R}^n \times \mathbf{Y}$. We deduce that $\hat{\Psi}^{-1}(u, v, z)$ is a singleton $\{(x, w, s, y)\}$, where

$$
\begin{aligned}
x &= \xi(u, v, z) \\
s &= u - \nabla G(\bar{x})(\xi(u, v, z) - \bar{x}) \\
(w, y) &= K(z - \xi(u, v, z)).
\end{aligned}
$$

For any fixed point $z \in X$, the dependence of the optimal solution $\xi(u, v, z)$ on the parameters $(u, v) \in \mathbf{R}^n \times \mathbf{Y}$ is globally Lipschitz, with constant independent of z (see [11], and also the discussion in [3, Section 4.7], and on the other hand, for fixed (u, v), as a function of the point z, it is globally 1-Lipschitz (being a projection). Hence the function ξ is globally Lipschitz. Consequently the map $\hat{\Psi}^{-1}$ is single-valued and Lipschitz, so the map $\hat{\Psi}$ is metrically regular, and hence so is the map Ψ at $(\bar{x}, 0, 0, 0)$ for for $(0, 0, \bar{x})$, as we claimed.

Equipped with this metric regularity property, we can now prove the desired inexact projection property for the constraint set M defined by equation (4.1). For any point z in $\mathbf{X}$ near $\bar{x}$, we solve the convex quadratic program (4.3). To see that this calculation results in an inexact projection, in the sense of Definition 3.2, we can ignore the inactive constraints modeled by the map P, and consider the unique optimal solution x_z of the convex quadratic program

$$
\begin{cases}
\text{minimize} & \frac{1}{2}|x - z|^2 \\
\text{subject to} & G(z) + \nabla G(z)(x - z) \leq 0 \\
& H(z) + \nabla H(z)(x - z) = 0 \\
& x \in \mathbf{X}.
\end{cases}
$$

Define a C^2-smooth map $A \colon \mathbf{X} \to \mathbf{R}^n \times \mathbf{Y}$ by $A(z) = \big(G(z), H(z)\big)$, for points $z \in \mathbf{X}$. By the linear independence constraint qualification, the linear map $\nabla A(z) \colon \mathbf{X} \to \mathbf{R}^n \times \mathbf{Y}$ is one-to-one for all z near $\bar{x}$, and hence the composition

$$
\nabla A(z) \nabla A(z)^* \colon \mathbf{R}^n \times \mathbf{Y} \to \mathbf{R}^n \times \mathbf{Y}
$$

is positive definite. The point

$$x = z - \nabla A(z)^* \big(\nabla A(z) \nabla A(z)^* \big)^{-1} A(z)$$

is feasible for the quadratic program. Hence we deduce

$$|x_z - z| \leq \left| \nabla A(z)^* \big(\nabla A(z) \nabla A(z)^* \big)^{-1} A(z) \right| \;\to\; 0$$

as $z \to \bar{x}$, so $x_z \to \bar{x}$. Furthermore, for all z near $\bar{x}$, the optimal solution x_z is also feasible for the the linearized problem (4.3), so it is also the unique optimal solution of that problem.

Using the first-order optimality conditions, there exist a unique slack vector $s_z \in \mathbf{R}^n$ and multiplier vectors $w_z \in \mathbf{R}^n$ and $y_z \in \mathbf{Y}$ such that $(w_z, s_z) \in \Omega$ and

$$
\begin{aligned}
G(z) + \nabla G(z)(x_z - z) + s_z &= 0 \\
H(z) + \nabla H(z)(x_z - z) &= 0 \\
x_z + \nabla G(z)^* w_z + \nabla H(z)^* y_z &= z.
\end{aligned}
$$

In fact w_z and y_z are also unique, by the linear independence constraint qualification, which furthermore implies

(4.5) $$|w_z| + |y_z| = O(|x_z - z|)$$

as $z \to \bar{x}$. The C^2-smoothness of G and H implies

$$
\begin{aligned}
G(x_z) + s_z &= O(|x_z - z|^2) \\
H(x_z) &= O(|x_z - z|^2).
\end{aligned}
$$

Furthermore we have

$$
\begin{aligned}
x_z &+ \nabla G(x_z)^* w_z + \nabla H(x_z)^* y_z - z \\
&= (\nabla G(x_z) - \nabla G(z))^* w_z + (\nabla H(x_z) - \nabla H(z))^* y_z \\
&= O(|x_z - z|)|w_z| + O(|x_z - z|)|y_z| \\
&= O(|x_z - z|^2)
\end{aligned}
$$

by equation (4.5). By the metric regularity of the map Ψ, we deduce

$$|x_z - P_M(z)| = O(|x_z - z|^2) \quad \text{as } z \to \bar{x}.$$

Notice the inequalities

$$
\begin{aligned}
|x_z - z| &\leq |x_z - P_M(z)| + |P_M(z) - z| \\
&= O(|x_z - z|^2) + d_M(z) \leq \frac{1}{2}|x_z - z| + d_M(z)
\end{aligned}
$$

10

for all z near $\bar{x}$, which imply $|x_z - z| \leq 2d_M(z)$, and hence

$$|x_z - P_M(z)| = O(d_M(z)^2) \quad \text{as } z \to \bar{x}.$$

This completes the proof that the map $z \mapsto x_z$ is an inexact projection. $\qquad\square$

When the set M is defined only by equalities, the assumptions imply that it is a manifold around the point $\bar{x}$. As we note in the next section, manifolds also have dual representations, in terms of local coordinates. Given such a representation, as we see next, another approach to approximate alternating projections is available, which in particular allows us to dispense with the assumption that the set Q is prox-regular.

5 Approximate projections: a second approach

We return to the basic alternating projections convergence scheme that we outlined in Section 2. In contrast with the previous section, we consider a different approach to errors in the computation of the projection $P_M(y)$. Shifting notation slightly, we again consider alternating projections for two closed sets M and Q, this time in a Euclidean space $\mathbf{Y}$. We suppose, at a common point $\bar{y}$, that M intersects Q separably (as opposed to vice versa, as before) and that M is super-regular there (as opposed to prox-regularity of Q, as before).

Consider the following algorithm.

Algorithm 5.1 (Approximate alternating projections)
   choose $z \in M$ and $\epsilon \geq 0$;
   while $z \notin Q$ **do**
     choose $y \in P_Q(z)$;
     choose $z' \in M$ with $d_{P_M(y)}(z') \leq \epsilon|y - z|$;
     $z = z'$;
   end while

Notice that, unlike in the previous inexact alternating projection scheme (Algorithm 3.1), we now require the approximate projection z' to lie in the set M. As we shall see, our assumptions again imply linear convergence, beginning close to the intersection point $\bar{y}$, providing that the tolerance ϵ is sufficiently small.

Aiming at this convergence result, we make the following definition, capturing the idea of approximating the projection of a point $y \in \mathbf{Y}$ onto a set $M \subset Y$ by an inexact estimate $\Phi(z, y)$ that depends on a base point $z \in M$.

Definition 5.2 A *faithful approximation* of the projection onto a set $M \subset \mathbf{Y}$ around a point $\bar{y} \in M$ is a map $\Phi \colon V \times \mathbf{Y} \to M$, where V is a neighborhood of $\bar{y}$ in M, with

the following property: given any constant $\epsilon > 0$ and angle $\alpha > 0$, if points $z \in M$ and $y \notin M$ are sufficiently close to $\bar{y}$, then any point $\hat{z} \in P_M(y)$ such that the angle between the vectors $z - y$ and $\hat{z} - y$ is at least α must satisfy $|\hat{z} - \Phi(z, y)| \le \epsilon|y - z|$.

We then have the following result.

Theorem 5.3 *Suppose a closed set M intersects a closed set Q separably at a point $\bar{y}$, and is super-regular there. If Φ is a faithful approximation of the projection onto M around $\bar{y}$, then starting from any nearby point $z \in M$, the approximate alternating projection iteration*

$$z \leftarrow \Phi(z, y) \quad \text{for any } y \in P_Q(z)$$

converges linearly to a point in the intersection $M \cap Q$.

Proof Consider one iteration of Algorithm 5.1, starting from a point z, and generating the points y and z' in turn, choose $z'' \in P_{P_M(y)}(z')$ and let $y' \in P_Q(z')$ denote the next point generated. For any constant τ in the interval $(\cos\alpha, 1)$, the analog of inequalities (2.6) still hold,

$$|z'' - y| \le \tau|z - y|,$$

and so we have

$$|z' - y'| \le |z' - y| \le |z' - z''| + |z'' - y| \le (\epsilon + \tau)|z - y|.$$

For any nonnegative $\epsilon < 1 - \tau$, we deduce linear convergence of the iterate sequence of iterates z to a point in the intersection $Q \cap M$. $\qquad\square$

Notice that this proof shows the linear convergence rate $\sqrt{\tau + \epsilon}$, which again we can make arbitrarily close to $\sqrt{\cos\alpha}$.

6   Smooth inclusions

We apply the approximate alternating projections framework of the previous section to the following problem. Given two Euclidean spaces $\mathbf{X}$ and $\mathbf{Y}$, a continuously differentiable map $F \colon \mathbf{X} \to \mathbf{Y}$, and a closed set $Q \subset \mathbf{Y}$, we consider the problem of finding a point $x \in \mathbf{X}$ satisfying

$$F(x) \in Q.$$

We can then consider the following simple splitting algorithm, which alternates between projections onto the set Q and linear least squares problems.

Algorithm 6.1 (Linearized alternating projections)
   choose $x \in \mathbf{X}$;
   while $F(x) \notin Q$ **do**
     choose $y \in P_Q(F(x))$;
     choose $s \in \mathbf{X}$ minimizing $|F(x) + \nabla F(x)s - y|$;
     $x = x + s$;
   end while

In the case $Q = \{0\}$, this is simply the Gauss-Newton method for the nonlinear least squares problem of minimizing $|F(\cdot)|^2$. On the other hand, if we replace the linear approximation $F(x) + \nabla F(x)s$ in the algorithm by the exact value $F(x + s)$, then, under reasonable conditions, the method becomes the exact alternating projection algorithm of Section 2 for the intersection of Q with a certain smooth manifold, namely the image under the map F of an open set containing the iterates x. We can then deduce linear convergence close to well-behaved solutions. Here, we parallel that development to prove local linear convergence of Algorithm 6.1 by interpreting it as a version of the approximate projection scheme in the previous section: Algorithm 5.1.

   More formally, we summarize the assumptions we need below.

Assumption 6.2

(i) The map F is continuously differentiable around the point $\bar{x}$, with one-to-one derivative $\nabla F(\bar{x})$, and satisfies $F(\bar{x}) \in Q$.

(ii) There exists an angle $\alpha > 0$ such that whenever two vectors $x, x' \in \mathbf{X}$ near $\bar{x}$ satisfy

$$\nabla F(x')^*(y - F(x')) = 0 \quad \text{and} \quad F(x) \neq y \neq F(x') \quad \text{for some} \quad y \in P_Q(F(x)),$$

then the angle between the vectors $F(x) - y$ and $F(x') - y$ is at least α.

Under this assumption, we study Algorithm 6.1 in a neighborhood of a particular solution $\bar{x}$ of our problem. The one-to-one assumption on the derivative in particular allows us to work with the following basic terminology for manifolds.

Definition 6.3 A set $M \subset \mathbf{Y}$ is a *manifold* around a point $\bar{y} \in M$ when there exists a corresponding *local coordinate system*: a continuously differentiable map F from another Euclidean space $\mathbf{X}$ into $\mathbf{Y}$ whose restriction to some open set $U \subset \mathbf{X}$ is a diffeomorphism onto an open neighborhood V of $\bar{y}$ in M.

Assumption 6.2 hence guarantees that, for any sufficiently small neighborhood U of $\bar{x}$, the image $M = F(U)$ is a manifold around the point $\bar{y} = F(\bar{x})$, and the map F is a local coordinate system. Furthermore, the normal space to the manifold,

13

$N_M\big(F(x)\big)$, is given by the null space $\mathrm{Null}\big(\nabla F(x)^*\big)$, for all x near $\bar{x}$. Thus the angle condition is exactly the property that M intersects the set Q separably at the point $\bar{y}$. It holds in particular under the usual transversality condition (2.2), which here becomes

$$N_Q\big(F(\bar{x})\big) \cap \mathrm{Null}\big(\nabla F(\bar{x})^*\big) = \{0\}.$$

Our aim is to prove the following result.

Theorem 6.4 (Local linear convergence) *If Assumption 6.2 holds, then, starting from any point $x \in \mathbf{X}$ sufficiently close to the solution $\bar{x}$, the linearized alternating projections method, Algorithm 6.1, converges linearly to a point in the solution set $F^{-1}(Q)$.*

7 Faithful approximate projections on manifolds

As in Definition 6.3, consider a local coordinate system F for a manifold M around a point $\bar{y} \in M$. We can calculate approximate projections onto the manifold M via a map $\Phi \colon V \times \mathbf{Y} \to \mathbf{Y}$ defined as follows. For coordinate vectors $x \in U$ and points $y \in \mathbf{Y}$, define

$$(7.1) \quad \Phi(F(x), y) = F(x + s) \quad \text{where } s \in \mathbf{Y} \text{ minimizes} \quad |F(x) + \nabla F(x)s - y|.$$

Notice that the vector s is unique, since the derivative map $\nabla F(x)$ is one-to-one. Calculating s is a simple least squares problem. If y is close to the point $F(x)$, then the vector s is small, and hence the point $F(x + s)$ lies on the manifold M, and is close to the point $F(x) + \nabla F(x)s$, which is the projection of y onto the tangent space $T_M(x)$. In this sense, the map $\Phi(F(x), \cdot)$ approximates the projection mapping P_M near the point $F(x)$. We can make this statement more precise as follows.

Theorem 7.2 *If Assumption 6.2(i) holds, then the map Φ defined by equation (7.1) is a faithful approximation of the projection for the manifold M around the point $\bar{y}$.*

Proof Suppose the result fails, so there exists an angle $\alpha > 0$ and a constant $\epsilon > 0$, and sequences $z_k \in M$, $y_k \in \mathbf{Y}$, and $z_k' \in P_M(y_k)$, all converging to $\bar{y}$, such that, for all $k = 1, 2, 3, \ldots$, the angle between the vectors $z_k - y_k$ and $z_k' - y_k$ is at least α, or in other words

$$\gamma_k = |z_k' - z_k| \geq |y_k - z_k| \sin \alpha$$

and furthermore

$$|z_k' - \Phi(z_k, y_k)| > \epsilon |y_k - z_k|.$$

The inequality $|z_k' - y_k| \leq |z_k - y_k|$ follows from the definition of z_k', so we deduce

$$(7.3) \qquad \gamma_k \leq |z_k' - y_k| + |y_k - z_k| \leq 2|y_k - z_k|.$$

Denote the coordinates of z_k, z_k' by $x_k, x_k' \in U$ respectively (clearly distinct), so $F(x_k) = z_k$ and $F(x_k') = z_k'$. Since F is a diffeomorphism between U and V, the sequences (x_k) and (x_k') both converge to the common limit $\bar{x} \in U$, the coordinates of $\bar{y}$. Define $\delta_k = |x_k' - x_k|$: the diffeomorphism property ensures $\gamma_k = O(\delta_k)$ and $\delta_k = O(\gamma_k)$. After taking a subsequence, we can suppose the unit vectors $\frac{1}{\delta_k}(x_k' - x_k)$ converge to a unit vector u.

Now define vectors $w_k = y_k - z_k' \in N_M(z_k')$. Notice $\nabla F(x_k')^* w_k = 0$, and

$$|w_k| \leq |y_k - z_k| + |z_k - z_k'| = O(\gamma_k) + \gamma_k = O(\gamma_k) = O(\delta_k).$$

Taking a further subsequence, we can then suppose that the vectors $\frac{1}{\delta_k} w_k$ converge to some vector in the null space of $\nabla F(\bar{x})^*$.

By definition, $\Phi(z_k, y_k) = F(x_k + s_k)$, where the vector $s_k \in \mathbf{X}$ minimizes the function

$$s \mapsto |F(x_k) + \nabla F(x_k)s - y_k|.$$

We deduce

$$\nabla F(x_k)^* (F(x_k) + \nabla F(x_k)s_k - F(x_k') - w_k) = 0,$$

so

$$s_k = \big(\nabla F(x_k)^* \nabla F(x_k)\big)^{-1} \nabla F(x_k)^* \big(F(x_k') - F(x_k) + w_k\big).$$

We deduce $\frac{1}{\delta_k} s_k \to u$. Hence

$$(x_k + s_k) - x_k' = o(\delta_k)$$

so inequality (7.3) implies

$$\Phi(z_k, y_k) - z_k' = F(x_k + s_k) - F(x_k') = o(\delta_k) = o(\gamma_k) = o(|y_k - z_k|)$$

which is a contradiction. $\qquad\square$

Theorem 6.4 (Local linear convergence) now follows directly from Theorem 5.3.

References

[1] 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.

[2] D. Drusvyatskiy, A.D. Ioffe, and A.S. Lewis. Transversality and alternating projections for nonconvex sets. *Found. Comput. Math.*, 15(6):1637–1651, 2015.

[3] F. Facchinei and J. Pang. *Finite Dimensional Variational Inequalities and Complementarity Problems.* Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[4] A.Y. Kruger and N.H. Thao. Regularity of collections of sets and convergence of inexact alternating projections. *J. Convex Anal.*, 23(3):823–847, 2016.

[5] 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.

[6] D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. *Found. Comput. Math.*, 16:425–455, 2016.

[7] R.A. Poliquin and R.T. Rockafellar. Amenable functions in optimization. In *Nonsmooth Optimization: Methods and Applications (Erice, 1991)*, pages 338–353. Gordon and Breach, Montreux, 1992.

[8] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.

[9] R.A. Poliquin, R.T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

[10] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis.* Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.

[11] N. D. Yen. Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint. *Mathematics of Operations Research*, 20:695–708, 1995.