Local Linear Convergence for Alternating and Averaged Nonconvex Projections

A.S. Lewis · D.R. Luke · J. Malick

Abstract The idea of a finite collection of closed sets having "linearly regular inter-section" at a point is crucial in variational analysis. This central theoretical condition also has striking algorithmic consequences: in the case of two sets, one of which satisfies a further regularity condition (convexity or smoothness, for example), we prove that von Neumann's method of "alternating projections" converges locally to a point in the intersection, at a linear rate associated with a modulus of regularity. As a consequence, in the case of several arbitrary closed sets having linearly regular intersection at some point, the method of "averaged projections" converges locally at a linear rate to a point in the intersection. Inexact versions of both algorithms also converge linearly.

A.S. Lewis (✉)
ORIE, Cornell University, Ithaca, NY 14853, USA
e-mail: aslewis@orie.cornell.edu
url: people.orie.cornell.edu/~aslewis

D.R. Luke
Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, USA
e-mail: rluke@math.udel.edu

J. Malick
INRIA Rhone-Alpes, 655 avenue de l'Europe, 38334 St Ismier Cedex, France
e-mail: jerome.malick@inria.fr

1 Introduction

An important theme in computational mathematics is the relationship between "conditioning" of a problem instance and speed of convergence of iterative solution algorithms on that instance. A classical example is the method of conjugate gradients for a positive definite system of linear equations: the relative condition number of the associated matrix gives a bound on the linear convergence rate. More generally, Renegar [41–43] showed that the rate of convergence of interior-point methods for conic convex programming can be bounded in terms of the "distance to ill-posedness" of the program.

In studying the convergence of iterative algorithms for nonconvex minimization problems or nonmonotone variational inequalities, we must content ourselves with a local theory. A suitable analogue of the distance to ill-posedness is then the notion of "metric regularity," fundamental in variational analysis. Loosely speaking, a constraint system, such as a system of inequalities, for example, is metrically regular when, locally, we can bound the distance from a trial solution to an exact solution by a constant multiple of the error in the equation generated by the trial solution. The constant needed is called the "regularity modulus," and its reciprocal has a natural interpretation as a distance to ill-posedness [19]. While not appropriate as a universal condition on general variational systems [34], metric regularity is often a reasonable assumption for constraint systems.

This philosophy suggests understanding the speed of convergence of algorithms for solving constraint systems in terms of the regularity modulus at a solution. Recent literature focuses in particular on the proximal point algorithm (see, for example [1, 13, 26, 37]). After the initial version [29] of this article, an independent but related, proximal-type development was announced in [2]. A unified approach to the relationship between metric regularity and the linear convergence of a family of conceptual algorithms appears in [27].

We here study a very basic algorithm for a very basic problem. We consider the problem of finding a point in the intersection of several closed sets, using the *method of averaged projections*: at each step, we project the current iterate onto each set, and average the results to obtain the next iterate. Global convergence of this method for convex sets was proved in 1969 in [3]. Here we show, in complete generality, that this method converges locally to a point in the intersection of the sets, at a linear rate governed by an associated regularity modulus. Our linear convergence proof is elementary: although we use the idea of the normal cone, we apply only the definition, and we discuss metric regularity only to illuminate the rate of convergence.

Finding a point in the intersection of several sets is a problem of fundamental computational significance. In the case of closed halfspaces, for example, the problem is equivalent to linear programming. We mention some nonconvex examples below.

Our approach to the convergence of the method of averaged projections is standard [5, 38, 39]: we identify the method with von Neumann's *alternating* projections algorithm [49] on two closed sets (one of which is a linear subspace) in a suitable product space. A nice development of the classical method of alternating projections in the convex case may be found in [15]. The convergence of the method for two intersecting closed convex sets was proved in [8], and *linear* convergence under a regular intersection assumption was proved in [5], strengthening a classical result of [25]. Our algorithmic contribution is to show that assuming linear regularity, *local* linear convergence does not depend on convexity of both sets, but rather on a good geometric property (such as convexity, smoothness, or more generally, "amenability" or "prox-regularity") of just *one* of the two.

One consequence of our convergence proof is an algorithmic demonstration for the "exact extremal principle" of [31] (see also [33, Theorem 2.8]). This result, a unifying theme in [33], asserts that if several sets have linearly regular intersection at a point, then that point is not "locally extremal": that is, translating the sets by sufficiently small vectors cannot render the intersection empty locally. To prove this result, we simply apply the method of averaged projections, starting from the point of regular intersection. In a further section, we show that inexact versions of the method of averaged projections, closer to practical implementations, also converge linearly.

The method of averaged projections is a conceptual algorithm that might appear hard to implement on concrete nonconvex problems. However, the projection problem for some nonconvex sets is relatively easy. A good example is the set of matrices of some fixed rank: given a singular value decomposition of a matrix, projecting it onto this set is immediate. Furthermore, nonconvex alternating projection algorithms and analogous heuristics are quite popular in practice, in areas such as inverse eigenvalue problems [10, 11], pole placement [35, 51], information theory [48], low-order control design [23, 24, 36], and image processing [7, 50]. Previous convergence results on nonconvex alternating projection algorithms have been uncommon, and have either focussed on a very special case (see, for example [10, 30]), or have been much weaker than for the convex case [14, 48]. For more discussion, see [30].

Our results primarily concern R-linear convergence: we show that our sequences of iterates converge, with error bounded by a geometric sequence. In a final section, we employ a completely different approach to show that the method of averaged projections, for prox-regular sets with regular intersection, has a Q-linear convergence property: each iteration guarantees a fixed rate of improvement. In a final section, we illustrate these theoretical results with an elementary numerical example coming from signal processing.

Our interest here is not in the development of practical numerical methods. Notwithstanding linear convergence proofs, basic alternating and averaged projection schemes may be slow in practice. Rather we aim to study the interplay between a simple, popular, fundamental algorithm and a variety of central ideas from variational analysis. Whether such an approach can help in the design and analysis of more practical algorithms remains to be seen.

2 Notation and Definitions

We fix some notation and definitions. Our underlying setting throughout this work is a Euclidean space $\mathbf{E}$ with corresponding closed unit ball B. For any point $x \in \mathbf{E}$ and radius $\rho > 0$, we write $B_\rho(x)$ for the set $x + \rho B$.

Consider first two sets $F, G \subset \mathbf{E}$. A point $\bar{x} \in F \cap G$ is *locally extremal* [33] for this pair of sets if there exists a constant $\rho > 0$ and a sequence of vectors $z_r \to 0$ in $\mathbf{E}$ such that $(F + z_r) \cap G \cap B_\rho(\bar{x}) = \emptyset$ for all $r = 1, 2, \ldots$. In other words, restricting to a neighborhood of $\bar{x}$ and then translating the sets by arbitrarily small distances can render their intersection empty. Clearly $\bar{x}$ is not locally extremal if and only if

$$0 \in \text{int} \left(\left((F - \bar{x}) \cap \rho B \right) - \left((G - \bar{x}) \cap \rho B \right) \right) \quad \text{for all } \rho > 0.$$

For recognition purposes, it is easier to study a weaker property than local extremality. We say that two sets $F, G \subset \mathbf{E}$ have *linearly regular intersection* at the point $\bar{x} \in F \cap G$ if there exist constants $\alpha, \delta > 0$ such that for all points $x \in F \cap B_\delta(\bar{x})$ and $z \in G \cap B_\delta(\bar{x})$, and all $\rho \in (0, \delta]$, we have

$$\alpha \rho B \subset \left((F - x) \cap \rho B \right) - \left((G - z) \cap \rho B \right).$$

(In [28], this property is called "strong regularity.") By considering the case $x = z = \bar{x}$, we see that linear regularity implies that $\bar{x}$ is not locally extremal. This "primal" definition of linear regularity is often not the most convenient way to handle linear regularity, either conceptually or theoretically. By contrast, a "dual" approach, using normal cones, is very helpful.

Given a set $F \subset \mathbf{E}$, we define the *distance function* and (multi-valued) *projection* for F by

$$d_F(x) = d(x, F) = \inf \{ \|z - x\| : z \in F \},$$
$$P_F(x) = \text{argmin} \{ \|z - x\| : z \in F \}.$$

The *normal cone* to a closed set $F \subset \mathbf{E}$ at a point $\bar{x} \in F$ is

$$N_F(\bar{x}) = \left\{ \lim_i t_i (x_i - z_i) : t_i \geq 0, x_i \to \bar{x}, z_i \in P_F(x_i) \right\}.$$

The centrality of this idea in variational analysis is described at length in [12, 33, 44]. This construction dates back to [31]: see [44, Chap. 6 Commentary] and [33, Chap. 1 Commentary] for a discussion of the equivalence between this definition and that of [44, p. 199]. Notice two properties in particular. First,

$$z \in P_F(x) \quad \Longrightarrow \quad x - z \in N_F(z). \tag{2.1}$$

Secondly, the normal cone is a "closed" multifunction: for any sequence of points $x_r \to \bar{x}$ in F, any limit of a sequence of normals $y_r \in N_F(x_r)$ must lie in $N_F(\bar{x})$. Indeed, the normal cone is the smallest cone satisfying the two properties. A noteworthy equivalence is $N_F(x) = \{0\} \Leftrightarrow x \in \text{int } F$.

Normal cones provide an elegant alternative approach to defining linear regularity. A family of closed sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ has *linearly regular intersection* at a point $\bar{x} \in \bigcap_i F_i$, when the only solution to the system

$$\sum_{i=1}^{m} y_i = 0, \quad \text{with } y_i \in N_{F_i}(\bar{x}) \ (i = 1, 2, \ldots, m)$$

is $y_i = 0$ for $i = 1, 2, \ldots, m$ (cf. the "exact extremal principle" of [33, Theorem 2.8]). In the case $m = 2$, this condition can be written

$$N_{F_1}(\bar{x}) \cap -N_{F_2}(\bar{x}) = \{0\}, \tag{2.2}$$

and it is equivalent to our previous definition (see [28, Corollary 2], for example). We also note that this condition appears throughout variational-analytic theory. For example, it guarantees the crucial inclusion (see [32, Theorem 1] and also [44, Theorem 6.42])

$$N_{F_1 \cap \cdots \cap F_m}(\bar{x}) \subset N_{F_1}(\bar{x}) + \cdots + N_{F_m}(\bar{x}).$$

For convex F_1 and F_2, condition (2.2) asserts the nonexistence of a separating hyperplane. More generally, linear regularity was introduced in [32] as the "generalized nonseparation property." The notion of a "linear regular" family of convex sets [6] is also related, though the definition we use here is local.

We will find it helpful to quantify the notion of linear regularity (cf. [28]). A straightforward compactness argument shows the following result.

Proposition 2.3 (Quantifying linear regularity) *A collection of closed sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ have linearly regular intersection at a point $\bar{x} \in \cap F_i$ if and only if there exists a constant $k > 0$ such that the following condition holds*:

$$y_i \in N_{F_i}(\bar{x}) \quad (i = 1, 2, \ldots, m) \quad \Longrightarrow \quad \sqrt{\sum_i \|y_i\|^2} \le k \left\| \sum_i y_i \right\|. \tag{2.3}$$

We define the *condition modulus* $\mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x})$ to be the infimum of all constants $k > 0$ such that property (2.3) holds. Since $\| \cdot \|^2$ is convex, we notice that vectors $y_1, y_2, \ldots, y_m \in \mathbf{E}$ always satisfy the inequality

$$\sum_i \|y_i\|^2 \ge \frac{1}{m} \left\| \sum_i y_i \right\|^2, \tag{2.4}$$

which yields

$$\mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x}) \ge \frac{1}{\sqrt{m}}, \tag{2.5}$$

except in the special case when $N_{F_i}(\bar{x}) = \{0\}$ (or equivalently $\bar{x} \in \mathrm{int}\, F_i$) for all $i = 1, 2, \ldots, m$; in this case the condition modulus is zero.

One goal of this paper is to show that far from being of purely analytic significance, linear regularity has central algorithmic consequences, specifically for the

method of averaged projections for finding a point in the intersection $\bigcap_i F_i$. Given any initial point $x_0 \in \mathbf{E}$, the algorithm proceeds iteratively as follows:

$$z_n^i \in P_{F_i}(x_n) \quad (i = 1, 2, \ldots, m),$$

$$x_{n+1} = \frac{1}{m} \left(z_n^1 + z_n^2 + \cdots + z_n^m \right).$$

Our main result shows, assuming only linear regularity that providing the initial point x_0 is sufficiently near $\bar{x}$, any sequence $x_1, x_2, x_3, \ldots$ generated by the method of averaged projections converges linearly to a point in the intersection $\bigcap_i F_i$, at a rate governed by the condition modulus.

3 Linear and Metric Regularity

The notion of linear regularity is well known to be closely related to another central idea in variational analysis: "metric regularity." A concise summary of the relationships between a variety of regular intersection properties and metric regularity appears in [28]. We summarize the relevant ideas here.

Consider a set-valued mapping $\Phi \colon \mathbf{E} \rightrightarrows \mathbf{Y}$, where $\mathbf{Y}$ is a second Euclidean space. The inverse mapping $\Phi^{-1} \colon \mathbf{Y} \rightrightarrows \mathbf{E}$ is defined by $x \in \Phi^{-1}(y) \Leftrightarrow y \in \Phi(x)$, for $x \in \mathbf{E}$ and $y \in \mathbf{Y}$. For vectors $\bar{x} \in \mathbf{E}$ and $\bar{y} \in \Phi(\bar{x})$, we say Φ is *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exists a constant $\kappa > 0$ such that all pairs $(x, y) \in \mathbf{E} \times \mathbf{Y}$ sufficiently near $(\bar{x}, \bar{y})$ satisfy the inequality

$$d\big(x, \Phi^{-1}(y)\big) \le \kappa d\big(y, \Phi(x)\big).$$

The infimum of all such constants κ is called the *modulus of metric regularity* of Φ at $\bar{x}$ for $\bar{y}$, denoted $\operatorname{reg}\Phi(\bar{x}|\bar{y})$. See [44, Chap. 9G] for a discussion.

Intuitively, metric regularity gives a local linear bound for the distance to a solution of the constraint system $y \in \Phi(x)$ (where the vector y is given and we seek the unknown vector x), in terms of the distance from y to the set $\Phi(x)$. The modulus is a measure of the sensitivity or "conditioning" of the constraint system $y \in \Phi(x)$. To take one simple example, if Φ is a single-valued linear map, the modulus of regularity is the reciprocal of its smallest singular value. In general, variational analysis provides a powerful calculus for computing the regularity modulus. In particular, we have the following formula (see [32, Theorem 8] and [44, Theorem 9.43]):

$$\frac{1}{\operatorname{reg}\Phi(\bar{x}|\bar{y})} = \min\big\{ d\big(0, D^*\Phi(\bar{x}|\bar{y})(w)\big) : w \in \mathbf{Y}, \|w\| = 1 \big\}, \qquad (3.1)$$

where D^* denotes the "coderivative." We now study these ideas for a particular mapping, highlighting the connections between metric and linear regularity. As in the previous section, consider closed sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ and a point $\bar{x} \in \bigcap_i F_i$. We endow the space $\mathbf{E}^m$ with the inner product

$$\big\langle (x_1, x_2, \ldots, x_m), (y_1, y_2, \ldots, y_m) \big\rangle = \sum_i \langle x_i, y_i \rangle,$$

and define a set-valued mapping $\Phi \colon \mathbf{E} \rightrightarrows \mathbf{E}^m$ by

$$\Phi(x) = (F_1 - x) \times (F_2 - x) \times \cdots \times (F_m - x).$$

Then the inverse mapping is given by $\Phi^{-1}(y) = \bigcap_i (F_i - y_i)$, for $y \in \mathbf{E}^m$, and finding a point in the intersection $\bigcap_i F_i$ is equivalent to finding a solution of the constraint system $0 \in \Phi(x)$. By definition, the mapping Φ is metrically regular at $\bar{x}$ for 0 if and only if there is a constant $\kappa > 0$ such that the following *strong metric inequality* holds:

$$d\left(x, \bigcap_i (F_i - z_i)\right) \leq \kappa \sqrt{\sum_i d^2(x, F_i - z_i)} \quad \text{for all } (x, z) \text{ near } (\bar{x}, 0). \qquad (3.2)$$

Furthermore, the regularity modulus $\operatorname{reg} \Phi(\bar{x}|0)$ is just the infimum of those constants $\kappa > 0$ such that inequality (3.2) holds.

To compute the coderivative $D^*\Phi(\bar{x}|0)$, we decompose the mapping Φ as $\Psi - A$, where, for points $x \in \mathbf{E}$, we define $\Psi(x) = F_1 \times F_2 \times \cdots \times F_m$ and $Ax = (x, x, \ldots, x)$. The calculus rule [44, 10.43] yields $D^*\Phi(\bar{x}|0) = D^*\Psi(\bar{x}|A\bar{x}) - A^*$. Then by definition,

$$v \in D^*\Psi(\bar{x}|A\bar{x})(w) \quad \Longleftrightarrow \quad (v, -w) \in N_{\operatorname{gph}\Psi}(\bar{x}, A\bar{x}),$$

and since $\operatorname{gph}\Psi = \mathbf{E} \times F_1 \times F_2 \times \cdots \times F_m$, we deduce

$$D^*\Psi(\bar{x}|A\bar{x})(w) = \begin{cases} \{0\} & \text{if } w_i \in -N_{F_i}(\bar{x}) \ \forall i, \\ \emptyset & \text{otherwise} \end{cases}$$

and hence

$$D^*\Phi(\bar{x}|0)(w) = \begin{cases} -\sum_i w_i & \text{if } w_i \in -N_{F_i}(\bar{x}) \ \forall i, \\ \emptyset & \text{otherwise.} \end{cases}$$

From the coderivative formula (3.1), we now obtain

$$\frac{1}{\operatorname{reg}\Phi(\bar{x}|0)} = \min\left\{ \left\| \sum_i y_i \right\| : \sum_i \|y_i\|^2 = 1, \ y_i \in N_{F_i}(\bar{x}) \right\}, \qquad (3.3)$$

where, as usual, we interpret the right-hand side as $+\infty$ if $N_{F_i}(\bar{x}) = \{0\}$ (or equivalently $\bar{x} \in \operatorname{int} F_i$) for all $i = 1, 2, \ldots, m$. Thus, the regularity modulus agrees exactly with the condition modulus that we defined in the previous section: $\operatorname{reg}\Phi(\bar{x}|0) = \operatorname{cond}(F_1, F_2, \ldots, F_m|\bar{x})$. It is well known [28] that linear regularity is equivalent to the strong metric inequality (3.2).

4 Clarke Regularity and Refinements

"Clarke regularity" is a basic variational-geometric property of sets, shared in particular by closed convex sets and smooth manifolds. We next study a slight refinement, crucial for our development. In the interest of maintaining as elementary approach as possible, we use the following definition of Clarke regularity, easy to interpret geometrically in terms of certain angles.

Definition 4.1 (Clarke regularity) A closed set $C \subset \mathbf{R}^n$ is *Clarke regular* at a point $\bar{x} \in C$ if, for all $\delta > 0$, any two points u, z sufficiently near $\bar{x}$ with $z \in C$, and any point $y \in P_C(u)$, satisfy $\langle z - \bar{x}, u - y \rangle \leq \delta \|z - \bar{x}\| \cdot \|u - y\|$.

Remark 4.2 This property is equivalent to the standard notion of Clarke regularity [44, Definition 6.4]. To see this, suppose the property in the definition holds. Consider any unit vector $v \in N_C(\bar{x})$, and any unit "tangent direction" w to C at $\bar{x}$. By definition, there exists a sequences $u_r \to \bar{x}$, $y_r \in P_C(u_r)$, and $z_r \to \bar{x}$ with $z_r \in C$, such that

$$v_r = \frac{u_r - y_r}{\|u_r - y_r\|} \to v,$$

$$w_r = \frac{z_r - \bar{x}}{\|z_r - \bar{x}\|} \to w.$$

By assumption, given any $\delta > 0$, for all sufficiently large r, we have $\langle v_r, w_r \rangle \leq \delta$, and hence $\langle v, w \rangle \leq \delta$. Thus, $\langle v, w \rangle \leq 0$, so Clarke regularity follows, by [44, Corollary 6.29]. Conversely, if the property described in the definition fails, then for some $\delta > 0$ and some sequences $u_r \to \bar{x}$, $y_r \in P_C(u_r)$, and $z_r \to \bar{x}$ with $z_r \in C$, we have

$$\left\langle \frac{u_r - y_r}{\|u_r - y_r\|}, \frac{z_r - \bar{x}}{\|z_r - \bar{x}\|} \right\rangle \geq \delta \quad \text{for all } r.$$

Then any cluster points v and w of the two sequences of unit vectors defining the above inner product are respectively an element of $N_C(\bar{x})$ and a tangent direction to C at $\bar{x}$, and satisfy $\langle v, w \rangle > 0$, contradicting Clarke regularity.

The property we need for our development is an apparently-slight modification of Clarke regularity, again easy to interpret geometrically.

Definition 4.3 (Super-regularity) A closed set $C \subset \mathbf{R}^n$ is *super-regular* at a point $\bar{x} \in C$ if, for all $\delta > 0$, any two points u, z sufficiently near $\bar{x}$ with $z \in C$, and any point $y \in P_C(u)$, satisfy $\langle z - y, u - y \rangle \leq \delta \|z - y\| \cdot \|u - y\|$.

An equivalent statement involves the normal cone.

Proposition 4.4 (Super-regularity and normal angles) *A closed set $C \subset \mathbf{R}^n$ is super-regular at a point $\bar{x} \in C$ if and only if, for all $\delta > 0$, the inequality $\langle v, z - y \rangle \leq \delta \|v\| \cdot \|z - y\|$ holds for all points $y, z \in C$ sufficiently near $\bar{x}$ and all vectors $v \in N_C(y)$.*

Proof Super-regularity follows immediately from the normal cone property described in the proposition, by property (2.1). Conversely, suppose the normal cone property fails, so for some $\delta > 0$ and sequences of distinct points $y_r, z_r \in C$ approaching $\bar{x}$ and unit normal vectors $v_r \in N_C(y_r)$, we have, for all $r = 1, 2, \ldots,$

$$\left\langle v_r, \frac{z_r - y_r}{\|z_r - y_r\|} \right\rangle > \delta.$$

Fix an index r. By definition of the normal cone, there exist sequences of distinct points $u_r^j \to y_r$ and $y_r^j \in P_C(u_r^j)$ such that

$$\lim_{j\to\infty} \frac{u_r^j - y_r^j}{\|u_r^j - y_r^j\|} = v_r.$$

Since $\lim_j y_r^j = y_r$, we must have, for all sufficiently large j,

$$\left\langle \frac{u_r^j - y_r^j}{\|u_r^j - y_r^j\|}, \frac{z_r - y_r^j}{\|z_r - y_r^j\|} \right\rangle > \delta.$$

Choose j sufficiently large to ensure both the above inequality and the inequality $\|u_r^j - y_r\| < \frac{1}{r}$, and then define points $u_r' = u_r^j$ and $y_r' = y_r^j$.

We now have sequences of points u_r', z_r approaching $\bar{x}$ with $z_r \in C$, and $y_r' \in P_C(u_r')$, and satisfying

$$\left\langle \frac{u_r' - y_r'}{\|u_r' - y_r'\|}, \frac{z_r - y_r'}{\|z_r - y_r'\|} \right\rangle > \delta.$$

Hence, C is not super-regular at $\bar{x}$. $\square$

Super-regularity is a strictly stronger property than Clarke regularity, as the following result and example make clear.

Corollary 4.5 (Super-regularity implies Clarke regularity) *At any point in a closed set $C \subset \mathbf{R}^n$, super regularity implies Clarke regularity.*

Proof Suppose the point in question is $\bar{x}$. Fix any $\delta > 0$, and set $y = \bar{x}$ in Proposition 4.4. Then clearly any unit tangent direction d to C at $\bar{x}$ and any unit normal vector $v \in N_C(\bar{x})$ satisfy $\langle v, d \rangle \le \delta$. Since δ was arbitrary, in fact $\langle v, d \rangle \le 0$, so Clarke regularity follows by [44, Corollary 6.29]. $\square$

Example 4.6 Consider the following function $f : \mathbf{R} \to (-\infty, +\infty]$, taken from an example in [46]:

$$f(t) = \begin{cases} 2^r(t - 2^r) & (2^r \le t < 2^{r+1}, r \in \mathbf{Z}), \\ 0 & (t = 0), \\ +\infty & (t < 0). \end{cases}$$

This function has Clarke-regular epigraph at $(0, 0)$, but an exercise shows it is not super-regular there. Indeed, a minor refinement of this example (smoothing the set slightly close to the nonsmooth points $(2^r, 0)$ and $(2^r, 4^{r-1})$) shows that a set can be everywhere Clarke regular, and yet not super-regular.

Super-regularity is a common property: indeed, it is implied by two well-known properties that we discuss next. Following [44], we say that a set $C \subset \mathbf{R}^n$ is *amenable* at a point $\bar{x} \in C$ when there exists a neighborhood U of $\bar{x}$, a C^1 mapping G :

$U \to \mathbf{R}^\ell$, and a closed convex set $D \subset \mathbf{R}^\ell$ containing $G(\bar{x})$, and satisfying the constraint qualification

$$N_D\big(G(\bar{x})\big) \cap \ker\big(\nabla G(\bar{x})^*\big) = \{0\}, \tag{4.1}$$

such that

$$C \cap U = \big\{x \in U : G(x) \in D\big\}.$$

In particular, if C is defined by C^1 equality and inequality constraints and the Mangasarian–Fromovitz constraint qualification holds, then C is amenable.

Proposition 4.8 (Amenable implies super-regular) *If a closed set $C \subset \mathbf{R}^n$ is amenable at a point in C, then it is super-regular there.*

Proof Suppose the result fails at some point $\bar{x} \in C$. Assume as in the definition of amenability that in a neighborhood of $\bar{x}$, the set C is identical with the inverse image $G^{-1}(D)$, where the C^1 map G and the closed convex set D satisfy the condition (4.1). Then by definition, for some $\delta > 0$, there are sequences of points $y_r, z_r \in C$ converging to $\bar{x}$ and unit normal vectors $v_r \in N_C(y_r)$ satisfying $\langle v_r, z_r - y_r \rangle > \delta \|z_r - y_r\|$ for all $r = 1, 2, \ldots$. Since the normal cone mapping N_D is outer semicontinuous relative to D [44, Proposition 6.6], it is easy to check the condition

$$N_D\big(G(y_r)\big) \cap \ker\big(\nabla G(y_r)^*\big) = \{0\},$$

for all sufficiently large r, since otherwise we contradict assumption (4.1). Consequently, using the standard chain rule [44, Exercise 10.26(d)], we deduce $N_C(y_r) = \nabla G(y_r)^* N_D(G(y_r))$, so there are vectors $u_r \in N_D(G(y_r))$ such that $\nabla G(y_r)^* u_r = v_r$. The sequence (u_r) must be bounded, since otherwise, by taking a subsequence, we could suppose $\|u_r\| \to \infty$ and $\|u_r\|^{-1} u_r$ approaches some unit vector $\hat{u}$, leading to the contradiction

$$\hat{u} \in N_D\big(G(\bar{x})\big) \cap \ker\big(\nabla G(\bar{x})^*\big) = \{0\}.$$

For all sufficiently large r, we now have $\langle \nabla G(y_r)^* u_r, z_r - y_r \rangle > \delta \|z_r - y_r\|$, and by convexity of D, since $u_r \in N_D(G(y_r))$, we have $\langle u_r, G(z_r) - G(y_r) \rangle \leq 0$. Adding these two inequalities gives

$$\big\langle u_r, G(z_r) - G(y_r) - \nabla G(y_r)(z_r - y_r) \big\rangle < -\delta \|z_r - y_r\|.$$

But as $r \to \infty$, the left-hand side is $o(\|z_r - y_r\|)$, since the sequence (u_r) is bounded and G is C^1. This contradiction completes the proof.   □

A rather different refinement of Clarke regularity is the notion of "prox-regularity." Following [40, Theorem 1.3], we call a set $C \subset \mathbf{E}$ is *prox-regular* at a point $\bar{x} \in C$ if the projection mapping P_C is single-valued around $\bar{x}$. (In this case, clearly C must be locally closed around $\bar{x}$.) For example, if in the definition of an amenable set that we gave earlier, we strengthen our assumption on the map G to be C^2 rather than just C^1, the resulting set must be prox-regular. Without this strengthening, however, notice the

set $\{(s, t) \in \mathbf{R}^2 : t = |s|^{3/2}\}$ is amenable at the point $(0, 0)$ (and hence super-regular there), but is not prox-regular there.

Proposition 4.9 (Prox-regular implies super-regular) *If a closed set $C \subset \mathbf{R}^n$ is prox-regular at a point in C, then it is super-regular there.*

Proof If the results fails at $\bar{x} \in C$, then for some $\delta > 0$, there exist sequences of points $y_r, z_r \in C$ converging to the point $\bar{x}$, and a sequence of normal vectors $v_r \in N_C(y_r)$ satisfying the inequality $\langle v_r, z_r - y_r \rangle > \delta \| v_r \| \cdot \| z_r - y_r \|$. By [40, Proposition 1.2], there exist constants $\epsilon, \rho > 0$ such that

$$\left\langle \frac{\epsilon}{2\|v_r\|} v_r, z_r - y_r \right\rangle \le \frac{\rho}{2} \| z_r - y_r \|^2$$

for all large r. This gives a contradiction, since $\| z_r - y_r \| \le \frac{\delta \epsilon}{\rho}$ eventually. □

We digress briefly to discuss relationships between super-regularity and other notions in the literature. First note the following equivalent definition, which is an immediate consequence of Proposition 4.4, and which gives an alternate proof of Proposition 4.9 via "hypomonotonicity" of the truncated normal cone mapping $x \mapsto N_C(x) \cap B$ for prox-regular sets C [40, Theorem 1.3].

Corollary 4.10 (Approximate monotonicity) *A closed set $C \subset \mathbf{R}^n$ is super-regular at a point $\bar{x} \in C$ if and only if, for all $\delta > 0$, the inequality $\langle v - w, y - z \rangle \ge -\delta \| y - z \|$ holds for all points $y, z \in C$ sufficiently near $\bar{x}$ and all normal vectors $v \in N_C(y) \cap B$ and $w \in N_C(z) \cap B$.*

If we replace the normal cone N_C in the property described in the result above by its convex hull, the "Clarke normal cone," we obtain a stronger property, called "sub-smoothness" in [4]. Similar proofs to those above show that like super-regularity, sub-smoothness is a consequence of either amenability or prox-regularity. However, sub-smoothness is strictly stronger than super-regularity. To see this, consider the graph of the function $f : \mathbf{R} \to \mathbf{R}$ defined by the following properties: $f(0) = 0$, $f(2^r) = 4^r$ for all integers r, f is linear on each interval $[2^r, 2^{r+1}]$, and $f(t) = f(-t)$ for all $t \in \mathbf{R}$. The graph of f is super-regular at $(0, 0)$, but is not subsmooth there.

In a certain sense, however, the distinction between subsmoothness and super-regularity is slight. Suppose the set F is super-regular at every point in $F \cap U$, for some open set $U \subset \mathbf{R}^n$. Since super-regularity implies Clarke regularity, the normal cone and Clarke normal cone coincide throughout $F \cap U$, and hence F is also sub-smooth throughout $F \cap U$. In other words, "local" super regularity coincides with "local" subsmoothness, which in turn, by [4, Theorem 3.16] coincides with the "first order Shapiro property" [45] (also called "near convexity" in [47]) holding locally.

5 Alternating Projections with Nonconvexity

Having reviewed or developed over the last few sections the key variational-analytic properties that we need, we now turn to projection algorithms. In this section, we

develop our convergence analysis of the method of alternating projections. The following result is our basic tool, guaranteeing conditions under which the method of alternating projections converges linearly. For flexibility, we state it in a rather technical manner. For clarity, we point out afterward that the two main conditions, (5.3) and (5.4), are guaranteed in applications via assumptions of linear regularity and super-regularity (or in particular, amenability or prox-regularity), respectively.

Given any sets $F, C \subset \mathbf{E}$, an *alternating projection sequence* is any sequence of points $\{x_j\}$ in $\mathbf{E}$ satisfying the condition

$$x_{2n+1} \in P_F(x_{2n}) \quad \text{and} \quad x_{2n+2} \in P_C(x_{2n+1}) \quad (n = 0, 1, 2, \ldots), \qquad (5.1)$$

or the same property with F and C interchanged.

Theorem 5.2 (Linear convergence of alternating projections) *Consider the closed sets $F, C \subset \mathbf{E}$, and a point $\bar{x} \in F$. Fix any constant $\epsilon > 0$. Suppose for some constant $c' \in (0, 1)$, the following condition holds*:

$$\left. \begin{array}{ll} x \in F \cap (\bar{x} + \epsilon B), & u \in -N_F(x) \cap B \\ y \in C \cap (\bar{x} + \epsilon B), & v \in N_C(y) \cap B \end{array} \right\} \implies \langle u, v \rangle \le c'. \qquad (5.3)$$

Suppose for some constant $\delta \in [0, \frac{1-c'}{2})$ the following condition holds:

$$\left. \begin{array}{c} y, z \in C \cap (\bar{x} + \epsilon B) \\ v \in N_C(y) \cap B \end{array} \right\} \implies \langle v, z - y \rangle \le \delta \|z - y\|. \qquad (5.4)$$

Define a constant $c = c' + 2\delta < 1$. Then for any initial point $x_0 \in C$ satisfying $\|x_0 - \bar{x}\| \le \frac{1-c}{4}\epsilon$, any alternating projection sequence $\{x_j\}$ for the sets F and C must converge with R-linear rate $\sqrt{c}$ to a point $\hat{x} \in F \cap C$ satisfying the inequality $\|\hat{x} - x_0\| \le \frac{1+c}{1-c}\|x_0 - \bar{x}\|$.

Proof Assume property (5.1). By the definition of the projections, we have

$$\|x_{2n+3} - x_{2n+2}\| \le \|x_{2n+2} - x_{2n+1}\| \le \|x_{2n+1} - x_{2n}\|. \qquad (5.5)$$

Clearly, we therefore have

$$\|x_{2n+2} - x_{2n}\| \le 2\|x_{2n+1} - x_{2n}\|. \qquad (5.6)$$

We next claim

$$\left. \begin{array}{ll} \|x_{2n+1} - \bar{x}\| \le \frac{\epsilon}{2} & \text{and} \\ \|x_{2n+1} - x_{2n}\| \le \frac{\epsilon}{2} \end{array} \right\} \implies \|x_{2n+2} - x_{2n+1}\| \le c\|x_{2n+1} - x_{2n}\|. \qquad (5.7)$$

To see this, note that if $x_{2n+2} = x_{2n+1}$, the result is trivial, and if $x_{2n+1} = x_{2n}$ then $x_{2n+2} = x_{2n+1}$ so again the result is trivial. Otherwise, we have

$$\frac{x_{2n} - x_{2n+1}}{\|x_{2n} - x_{2n+1}\|} \in N_F(x_{2n+1}) \cap B$$

while

$$\frac{x_{2n+2} - x_{2n+1}}{\|x_{2n+2} - x_{2n+1}\|} \in -N_C(x_{2n+2}) \cap B.$$

Furthermore, using inequality (5.5), the left-hand side of the implication (5.7) ensures

$$\|x_{2n+2} - \bar{x}\| \le \|x_{2n+2} - x_{2n+1}\| + \|x_{2n+1} - \bar{x}\|$$
$$\le \|x_{2n+1} - x_{2n}\| + \|x_{2n+1} - \bar{x}\| \le \epsilon.$$

Hence, by assumption (5.3) we deduce

$$\left\langle \frac{x_{2n} - x_{2n+1}}{\|x_{2n} - x_{2n+1}\|}, \frac{x_{2n+2} - x_{2n+1}}{\|x_{2n+2} - x_{2n+1}\|} \right\rangle \le c',$$

so

$$\langle x_{2n} - x_{2n+1}, x_{2n+2} - x_{2n+1}\rangle \le c'\|x_{2n} - x_{2n+1}\| \cdot \|x_{2n+2} - x_{2n+1}\|.$$

On the other hand, by assumption (5.4), we know

$$\langle x_{2n} - x_{2n+2}, x_{2n+1} - x_{2n+2}\rangle \le \delta\|x_{2n} - x_{2n+2}\| \cdot \|x_{2n+1} - x_{2n+2}\|$$
$$\le 2\delta\|x_{2n} - x_{2n+1}\| \cdot \|x_{2n+2} - x_{2n+1}\|,$$

using inequality (5.6). Adding this inequality to the previous inequality then gives the right-hand side of (5.7), as desired.

Now let $\alpha = \|x_0 - \bar{x}\|$. We will show by induction the inequalities

$$\|x_{2n+1} - \bar{x}\| \le 2\alpha \frac{1 - c^{n+1}}{1 - c} < \frac{\epsilon}{2}, \tag{5.8}$$

$$\|x_{2n+1} - x_{2n}\| \le \alpha c^n < \frac{\epsilon}{2}, \tag{5.9}$$

$$\|x_{2n+2} - x_{2n+1}\| \le \alpha c^{n+1}. \tag{5.10}$$

Consider first the case $n = 0$. Since $x_1 \in P_F(x_0)$ and $\bar{x} \in F$, we deduce $\|x_1 - x_0\| \le \|\bar{x} - x_0\| = \alpha < \epsilon/2$, which is inequality (5.9). Furthermore,

$$\|x_1 - \bar{x}\| \le \|x_1 - x_0\| + \|x_0 - \bar{x}\| \le 2\alpha < \frac{\epsilon}{2},$$

which shows inequality (5.8). Finally, since $\|x_1 - x_0\| < \epsilon/2$ and $\|x_1 - \bar{x}\| < \epsilon/2$, the implication (5.7) shows

$$\|x_2 - x_1\| \le c\|x_1 - x_0\| \le c\|\bar{x} - x_0\| = c\alpha,$$

which is inequality (5.10).

For the induction step, suppose inequalities (5.8), (5.9), and (5.10) all hold for some n. Inequalities (5.5) and (5.10) imply

$$\|x_{2n+3} - x_{2n+2}\| \le \alpha c^{n+1} < \frac{\epsilon}{2}. \tag{5.11}$$

We also have, using inequalities (5.11), (5.10), and (5.8)

$$\|x_{2n+3} - \bar{x}\| \leq \|x_{2n+3} - x_{2n+2}\| + \|x_{2n+2} - x_{2n+1}\| + \|x_{2n+1} - \bar{x}\|$$

$$\leq \alpha c^{n+1} + \alpha c^{n+1} + 2\alpha \frac{1 - c^{n+1}}{1 - c},$$

so

$$\|x_{2n+3} - \bar{x}\| \leq 2\alpha \frac{1 - c^{n+2}}{1 - c} < \frac{\epsilon}{2}. \tag{5.12}$$

Now implication (5.7) with n replaced by $n+1$ implies $\|x_{2n+4} - x_{2n+3}\| \leq c\|x_{2n+3} - x_{2n+2}\|$, and using inequality (5.11) we deduce

$$\|x_{2n+4} - x_{2n+3}\| \leq \alpha c^{n+2}. \tag{5.13}$$

Since inequalities (5.12), (5.11), and (5.13) are exactly inequalities (5.8), (5.9), and (5.10) with n replaced by $n+1$, the induction step is complete and our claim follows.

We can now easily check that the sequence (x_k) is Cauchy and, therefore, converges. To see this, note for any integer $n = 0, 1, 2, \ldots$ and any integer $k > 2n$, we have

$$\|x_k - x_{2n}\| \leq \sum_{j=2n}^{k-1} \|x_{j+1} - x_j\|$$

$$\leq \alpha\left(c^n + c^{n+1} + c^{n+1} + c^{n+2} + c^{n+2} + \cdots\right)$$

so

$$\|x_k - x_{2n}\| \leq \alpha c^n \frac{1 + c}{1 - c},$$

and a similar argument shows

$$\|x_{k+1} - x_{2n+1}\| \leq \frac{2\alpha c^{n+1}}{1 - c}. \tag{5.14}$$

Hence, x_k converges to some point $\hat{x} \in \mathbf{E}$, and for all $n = 0, 1, 2, \ldots$ we have

$$\|\hat{x} - x_{2n}\| \leq \alpha c^n \frac{1 + c}{1 - c} \quad \text{and} \quad \|\hat{x} - x_{2n+1}\| \leq \frac{2\alpha c^{n+1}}{1 - c}. \tag{5.15}$$

We deduce that the limit $\hat{x}$ lies in the intersection $F \cap C$ and satisfies the inequality $\|\hat{x} - x_0\| \leq \alpha \frac{1+c}{1-c}$, and furthermore that the inequality

$$\|\hat{x} - x_r\| \leq \alpha\left(\sqrt{c}\right)^r \frac{1 + c}{1 - c}$$

holds for all $r = 0, 1, 2, \ldots$, so the convergence is R-linear with rate $\sqrt{c}$.   □

We can now prove our key result. To apply Theorem 5.2 to alternating projections between a closed and a super-regular set, we make use of the key geometric property of super-regular sets (Proposition 4.4).

Theorem 5.16 (Alternating projections with a super-regular set) *Consider closed sets $F, C \subset \mathbf{E}$ and a point $\bar{x} \in F \cap C$. Suppose C is super-regular at $\bar{x}$ (as holds, for example, if it is amenable or prox-regular there). Suppose furthermore that F and C have linearly regular intersection at $\bar{x}$: that is, $N_F(\bar{x}) \cap -N_C(\bar{x}) = \{0\}$, or equivalently, the constant*

$$\bar{c} = \max\big\{\langle u, v\rangle : u \in N_F(\bar{x}) \cap B, v \in -N_C(\bar{x}) \cap B\big\} \qquad (5.17)$$

is strictly less than one. Fix any constant $c \in (\bar{c}, 1)$. Then any alternating projection sequence with initial point sufficiently near $\bar{x}$ must converge to a point in $F \cap C$ with R-linear rate $\sqrt{c}$.

Proof Let us show first the equivalence between $\bar{c} < 1$ and linear regularity. The compactness of the intersections between normal cones and the unit ball guarantees the existence of u and v achieving the maximum in (5.17). Observe then that $\langle u, v\rangle \leq \|u\| \|v\| \leq 1$. The cases of equality in the Cauchy–Schwarz inequality permits to write

$$\bar{c} = 1 \quad \Longleftrightarrow \quad u \text{ and } v \text{ are colinear} \quad \Longleftrightarrow \quad N_F(\bar{x}) \cap -N_C(\bar{x}) \neq \{0\},$$

which corresponds to the desired equivalence.

Denote the alternating sequence $\{x_j\}$. We can suppose $x_0 \in C$. Fix any constant $c' \in (\bar{c}, c)$ and define $\delta = \frac{c-c'}{2}$. To apply Theorem 5.2, we just need to check the existence of a constant $\epsilon > 0$ such that conditions (5.3) and (5.4) hold. Condition (5.4) holds for all sufficiently small $\epsilon > 0$, by Proposition 4.4. On the other hand, if condition (5.3) fails for all sufficiently small $\epsilon > 0$, then there exist sequences of points $x_r \to \bar{x}$ in the set F and $y_r \to \bar{x}$ in the set C, and sequences of vectors $u_r \in -N_F(x_r) \cap B$ and $v_r \in N_C(y_r) \cap B$, satisfying $\langle u_r, v_r \rangle > c'$. After taking subsequences, we can suppose u_r approaches some vector $u \in -N_F(\bar{x}) \cap B$ and v_r approaches some vector $v \in N_C(\bar{x}) \cap B$, and then $\langle u, v\rangle \geq c' > \bar{c}$, contradicting the definition of the constant $\bar{c}$. $\qquad\square$

Corollary 5.18 (Improved convergence rate) *With the assumptions of Theorem 5.16, suppose the set F is also super-regular at $\bar{x}$. Then the alternating projection sequence converges with R-linear rate c.*

Proof Inequality (5.7) and its analog when the roles of F and C are interchanged, together show $\|x_{k+1} - x_k\| \leq c\|x_k - x_{k-1}\|$ for all sufficiently large k, and the result then follows easily, using an argument analogous to that at the end of the proof of Theorem 5.2. $\qquad\square$

In the light of our discussion in the previous section, the linear regularity assumption of Theorem 5.16 is equivalent to the metric regularity at $\bar{x}$ for 0 of the set-valued

mapping $\Psi : \mathbf{E} \rightrightarrows \mathbf{E}^2$ defined by $\Psi(x) = (F - x) \times (C - x)$, for $x \in \mathbf{E}$. Using (3.3), the regularity modulus is determined by

$$\frac{1}{\operatorname{reg} \Psi(\bar{x}|0)} = \min\{\|u + v\| : u \in N_F(\bar{x}), v \in N_C(\bar{x}), \|u\|^2 + \|v\|^2 = 1\},$$

and a short calculation then shows

$$\operatorname{reg} \Psi(\bar{x}|0) = \frac{1}{\sqrt{1 - \bar{c}}}. \tag{5.19}$$

The closer the constant $\bar{c}$ is to one, the larger the regularity modulus. We have shown that $\bar{c}$ also controls the speed of linear convergence for the method of alternating projections applied to the sets F and C.

Inevitably, Theorem 5.16 concerns *local* convergence: it relies on finding an initial point x_0 sufficiently close to a point of linearly regular intersection. How might we find such a point? One natural context in which to pose this question is that of sensitivity analysis. Suppose we already know a point of linearly regular intersection of two closed sets, but now want to find a point in the intersection of two slight translations of these sets. The following result shows that starting from the original point of intersection, the method of alternating projections will converge linearly to the new intersection.

Theorem 5.20 (Perturbed intersection) *Given any closed sets $F, C \subset \mathbf{E}$ and any point $\bar{x} \in F \cap C$, suppose the assumptions of Theorem 5.16 hold. Then for any sufficiently small vector $d \in \mathbf{E}$, any alternating projection sequence for the sets $d + F$ and C, with the initial point $\bar{x}$, must converge with R-linear rate $\sqrt{c}$ to a point in the set $(d + F) \cap C \cap B_\rho(\bar{x})$, where $\rho = \frac{1+c}{1-c}\|d\|$.*

Proof As in the proof of Theorem 5.16, if we fix any constant $c' \in (\bar{c}, c)$ and define $\delta = \frac{c - c'}{2}$, then there exists a constant $\epsilon > 0$ such that conditions (5.3) and (5.4) hold. Suppose the vector d satisfies

$$\|d\| \leq \frac{(1 - c)\epsilon}{8} < \frac{\epsilon}{2}.$$

Since

$$y \in (C - d) \cap \left(\bar{x} + \frac{\epsilon}{2}B\right) \quad \text{and} \quad v \in N_{C-d}(y)$$

$$\implies \quad y + d \in C \cap (\bar{x} + \epsilon B) \quad \text{and} \quad v \in N_C(y + d),$$

we deduce from condition (5.3) the implication

$$\left.\begin{array}{ll} x \in F \cap \left(\bar{x} + \dfrac{\epsilon}{2}B\right), & u \in -N_F(x) \cap B \\[2mm] y \in (C - d) \cap \left(\bar{x} + \dfrac{\epsilon}{2}B\right), & v \in N_{C-d}(y) \cap B \end{array}\right\} \implies \langle u, v \rangle \leq c'.$$

Furthermore, using condition (5.4), we deduce the implication

$$y, z \in (C - d) \cap \left(\bar{x} + \frac{\epsilon}{2} B\right) \quad \text{and} \quad v \in N_{C-d}(y) \cap B$$

$$\implies \quad y + d, z + d \in C \cap (\bar{x} + \epsilon B) \quad \text{and} \quad v \in N_C(y + d) \cap B$$

$$\implies \quad \langle v, z - y \rangle \leq \delta \|z - y\|.$$

We now apply Theorem 5.2 with the set C replaced by $C - d$ and ϵ replaced by $\frac{\epsilon}{2}$. We deduce that any alternating projection sequence for the sets F and $C - d$, starting at the point $x_0 = \bar{x} - d \in C - d$, converges with R-linear rate $\sqrt{c}$ to a point $\hat{x} \in F \cap (C - d)$ satisfying the inequality $\|\hat{x} - x_0\| \leq \frac{1+c}{1-c} \|x_0 - \bar{x}\|$. The theorem statement then follows by translation. □

Lack of convexity notwithstanding, more structure sometimes implies that the method of alternating projections converges Q-linearly, rather than just R-linearly, on a neighborhood of point of linearly regular intersection of two closed sets. One example is the case of two manifolds [30].

6 Inexact Alternating Projections

Our basic tool, the method of alternating projections for a super-regular set C and an arbitrary closed set F, is a conceptual algorithm that may be challenging to realize in practice. We might reasonably consider the case of exact projection on the super-regular set C: for example, in the next section, for the method of averaged projections, C is a subspace and computing projections is trivial. However, projecting onto the set F may be much harder, so a more realistic analysis allows relaxed projections.

We sketch one approach. Given two iterates $x_{2n-1} \in F$ and $x_{2n} \in C$, a necessary condition for the new iterate x_{2n+1} to be an exact projection on F, that is $x_{2n+1} \in P_F(x_{2n})$, is

$$\|x_{2n+1} - x_{2n}\| \leq \|x_{2n} - x_{2n-1}\| \quad \text{and} \quad x_{2n} - x_{2n+1} \in N_F(x_{2n+1}).$$

In the following result, we assume only that we choose the iterate x_{2n+1} to satisfy a relaxed version of this condition, where we replace the second part by the assumption that the distance

$$d_{N_F(x_{2n+1})} \left(\frac{x_{2n} - x_{2n+1}}{\|x_{2n} - x_{2n+1}\|} \right)$$

from the normal cone at the iterate to the normalized direction of the last step is *sufficiently small*.

Theorem 6.1 (Inexact alternating projections) *With the assumptions of Theorem 5.16, fix any constant $\gamma < \sqrt{1 - c^2}$, and consider the following inexact alternating projection iteration. Given any initial points $x_0 \in C$ and $x_1 \in F$, for*

$n = 1, 2, 3, \ldots$ *suppose* $x_{2n} \in P_C(x_{2n-1})$ *and* $x_{2n+1} \in F$ *satisfies*

$$\|x_{2n+1} - x_{2n}\| \le \|x_{2n} - x_{2n-1}\| \quad \text{and} \quad d_{N_F(x_{2n+1})}\left(\frac{x_{2n} - x_{2n+1}}{\|x_{2n} - x_{2n+1}\|} \right) \le \gamma.$$

Then providing x_0 *and* x_1 *are sufficiently close to* $\bar{x}$*, the iterates converge to a point in* $F \cap C$ *with R-linear rate*

$$\sqrt{c\sqrt{1-\gamma^2} + \gamma\sqrt{1-c^2}} < 1.$$

Sketch proof Once again as in the proof of Theorem 5.16, we fix any constant $c' \in (\bar{c}, c)$ and define $\delta = \frac{c-c'}{2}$, so there exists a constant $\epsilon > 0$ such that conditions (5.3) and (5.4) hold. Define a vector

$$z = \frac{x_{2n} - x_{2n+1}}{\|x_{2n} - x_{2n+1}\|}.$$

By assumption, there exists a vector $w \in N_F(x_{2n+1})$ satisfying $\|w - z\| \le \gamma$. Easy manipulation then shows that the unit vector $\hat{w} = \|w\|^{-1}w$ satisfies $\langle \hat{w}, z \rangle \ge \sqrt{1-\gamma^2}$. As in the proof of Theorem 5.2, assuming inductively that x_{2n+1} is sufficiently close to both $\bar{x}$ and x_{2n}, since $\hat{w} \in N_F(x_{2n+1})$, and

$$u = \frac{x_{2n+2} - x_{2n+1}}{\|x_{2n+2} - x_{2n+1}\|} \in -N_C(x_{2n+2}) \cap B,$$

we deduce $\langle \hat{w}, u \rangle \le c'$.

We now see that on the unit sphere, the arc distance between the unit vectors $\hat{w}$ and z is no more than $\arccos(\sqrt{1-\gamma^2})$, whereas the arc distance between $\hat{w}$ and the unit vector u is at least $\arccos c'$. Hence, by the triangle inequality, the arc distance between z and u is at least

$$\arccos c' - \arccos(\sqrt{1-\gamma^2}),$$

so

$$\langle z, u \rangle \le \cos\left(\arccos c' - \arccos(\sqrt{1-\gamma^2}) \right) = c'\sqrt{1-\gamma^2} + \gamma\sqrt{1-c'^2}.$$

Some elementary calculus shows that the quantity on the right-hand side is strictly less than one. Again as in the proof of Theorem 5.2, this inequality shows, providing x_0 is sufficiently close to $\bar{x}$, the inequality

$$\|x_{2n+2} - x_{2n+1}\| \le \left(c\sqrt{1-\gamma^2} + \gamma\sqrt{1-c^2} \right)\|x_{2n+1} - x_{2n}\|,$$

and in conjunction with the inequality $\|x_{2n+1} - x_{2n}\| \le \|x_{2n} - x_{2n-1}\|$, this suffices to complete the proof by induction. □

7 Local Convergence for Averaged Projections

We return to the problem of finding a point in the intersection of several closed sets via averaged projections. Given sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$, an *averaged projection sequence* is any sequence of points $\{x_j\}$ in $\mathbf{E}$ satisfying

$$x_{j+1} \in \frac{1}{m} \sum_{i=1}^{m} P_{F_i}(x_j) \quad (j = 0, 1, 2, \ldots).$$

We apply our previous results to the method of averaged projections via the well-known reformulation of the algorithm as alternating projections in a product space. This leads to the main result of this section, Theorem 7.3, which shows linear convergence in a neighborhood of any point of linearly regular intersection, at a rate governed by the associated regularity modulus.

We begin with a characterization of linearly regular intersection, relating the condition modulus with a generalized notion of angle for several sets. Such notions, for collections of convex sets, have also been studied recently in the context of projection algorithms in [16, 17].

Proposition 7.1 (Variational characterization of linear regularity) *Closed sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ have linearly regular intersection at a point $\bar{x} \in \bigcap_i F_i$ if and only if the optimal value $\bar{c}$ of the optimization problem*

$$\text{maximize} \quad \sum_i \langle u_i, v_i \rangle$$

$$\text{subject to} \quad \sum_i \|u_i\|^2 \leq 1,$$

$$\sum_i \|v_i\|^2 \leq 1,$$

$$\sum_i u_i = 0,$$

$$u_i \in \mathbf{E}, \quad v_i \in N_{F_i}(\bar{x}) \quad (i = 1, 2, \ldots, m)$$

is strictly less than one. Indeed, we have

$$\bar{c}^2 = \begin{cases} 0 & (\bar{x} \in \bigcap_i \text{int } F_i), \\ 1 - \frac{1}{m \cdot \text{cond}^2(F_1, F_2, \ldots, F_m|\bar{x})} & (otherwise). \end{cases} \quad (7.2)$$

Proof When $\bar{x} \in \bigcap_i \text{int } F_i$, the result follows by definition. Henceforth, we therefore rule out that case.

For any vectors $u_i, v_i \in \mathbf{E}$ $(i = 1, 2, \ldots, m)$, by Lagrangian duality and differentiation, we obtain

$$\max_{u_i} \left\{ \sum_i \langle u_i, v_i \rangle : \sum_i \|u_i\|^2 \leq 1, \sum_i u_i = 0 \right\}$$

$$
= \min_{\lambda \in \mathbf{R}_+, z \in \mathbf{E}} \max_{u_i} \left\{ \sum_i \langle u_i, v_i \rangle + \frac{\lambda}{2} \left(1 - \sum_i \|u_i\|^2 \right) + \langle z, \textstyle\sum_i u_i \rangle \right\}
$$

$$
= \min_{\lambda \in \mathbf{R}_+, z \in \mathbf{E}} \left\{ \frac{\lambda}{2} + \sum_i \max_{u_i} \left\{ \langle u_i, v_i + z \rangle - \frac{\lambda}{2} \|u_i\|^2 \right\} \right\}
$$

$$
= \min_{\lambda > 0, z \in \mathbf{E}} \left\{ \frac{\lambda}{2} + \frac{1}{2\lambda} \sum_i \|v_i + z\|^2 \right\} = \min_{z \in \mathbf{E}} \sqrt{\sum_i \|v_i + z\|^2}
$$

$$
= \sqrt{\sum_{i=1}^m \left\| v_i - \frac{1}{m} \sum_j v_j \right\|^2} = \sqrt{\sum_i \|v_i\|^2 - \frac{1}{m} \left\| \sum_i v_i \right\|^2}.
$$

Consequently, $\bar{c}^2$ is the optimal value of the optimization problem

$$
\text{maximize} \quad \sum_i \|v_i\|^2 - \frac{1}{m} \left\| \sum_i v_i \right\|^2
$$

$$
\text{subject to} \quad \sum_i \|v_i\|^2 \le 1,
$$

$$
v_i \in N_{F_i}(\bar{x}) \quad (i = 1, 2, \ldots, m).
$$

By homogeneity, the optimal solution must occur when the inequality constraint is active, so we obtain an equivalent problem by replacing that constraint by the corresponding equation. By (3.3) and the definition of the condition modulus, the optimal value of this new problem is

$$
1 - \frac{1}{m \cdot \text{cond}^2(F_1, F_2, \ldots, F_m | \bar{x})}
$$

as required.   □

Theorem 7.3 (Linear convergence of averaged projections) *Suppose closed sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ have linearly regular intersection at a point $\bar{x} \in \bigcap_i F_i$. Define a constant $\bar{c} \in [0, 1)$ by (7.2), and fix any constant $c \in (\bar{c}, 1)$. Then any averaged projection sequence with initial point sufficiently near $\bar{x}$ converges to a point in the intersection $\bigcap_i F_i$, with R-linear rate c (and if each set F_i is super-regular at $\bar{x}$, or in particular, prox-regular or amenable there, then the convergence rate is c^2). Furthermore, for any sufficiently small perturbations $d_i \in \mathbf{E}$ for $i = 1, 2, \ldots, m$, any averaged projection sequence for the sets $d_i + F_i$ with the initial point $\bar{x}$ converges linearly to a nearby point in the intersection, with R-linear rate c.*

Proof In the product space $\mathbf{E}^m$ with the inner product

$$
\langle (u_1, u_2, \ldots, u_m), (v_1, v_2, \ldots, v_m) \rangle = \sum_i \langle u_i, v_i \rangle,
$$

we consider the closed set $F = \prod_i F_i$ and the subspace $L = \{Ax : x \in \mathbf{E}\}$, where the linear map $A \colon \mathbf{E} \to \mathbf{E}^m$ is defined by $Ax = (x, x, \ldots, x)$. Notice $A\bar{x} \in F \cap L$, and it is easy to check $N_F(A\bar{x}) = \prod_i N_{F_i}(\bar{x})$ and

$$L^\perp = \left\{ (u_1, u_2, \ldots, u_m) : \sum_i u_i = 0 \right\}.$$

Hence, $F_1, F_2, \ldots, F_m$ have linearly regular intersection at $\bar{x}$ if and only if F and L have linearly regular intersection at the point $A\bar{x}$. This latter property is equivalent to the constant $\bar{c}$ in Theorem 5.16 (with $C = L$) being strictly less than one. But that constant agrees exactly with that defined by (7.2), so we show next that we can apply Theorems 5.16 and 5.20.

To see this note that for any point $x \in \mathbf{E}$, we have the equivalence

$$(z_1, z_2, \ldots, z_m) \in P_F(Ax) \quad \Longleftrightarrow \quad z_i \in P_{F_i}(x) \quad (i = 1, 2, \ldots, m).$$

Furthermore a quick calculation shows, for any $z_1, z_2, \ldots, z_m \in \mathbf{E}$,

$$P_L(z_1, z_2, \ldots, z_m) = \frac{1}{m}(z_1 + z_2 + \cdots + z_m).$$

Hence, in fact the method of averaged projections for the sets $F_1, F_2, \ldots, F_m$, starting at an initial point x_0, is essentially identical with the method of alternating projections for the sets F and L, starting at the initial point Ax_0. If $x_0, x_1, x_2, \ldots$ is a possible sequence of iterates for the former method, then a possible sequence of even iterates for the latter method is $Ax_0, Ax_1, Ax_2, \ldots$. For x_0 sufficiently close to $\bar{x}$, this latter sequence must converge to a point $A\hat{x} \in F \cap L$ with R-linear rate c, by Theorem 5.16 and its corollary. Thus, the sequence $x_0, x_1, x_2, \ldots$ converges to $\hat{x} \in \bigcap_i F_i$ at the same linear rate. When each of the sets F_i is super-regular at $\bar{x}$, it is easy to check that the Cartesian product F is super-regular at $A\bar{x}$, so the rate is c^2. The last part of the theorem follows from Theorem 5.20.   □

Applying Theorem 6.1 to the product-space formulation of averaged projections shows in a similar fashion that an inexact variant of the method of averaged projections will also converge linearly.

Remark 7.4 (Linear regularity and local extremality) In the language of [33], we have proved algorithmically that if closed sets have linearly regular intersection at a point, then that point is not "locally extremal".

Remark 7.5 (Alternating versus averaged projections) For a feasibility problem for two super-regular sets F_1 and F_2, assume that linear regularity holds at $\bar{x} \in F_1 \cap F_2$ and set $\kappa = \mathrm{cond}(F_1, F_2 | \bar{x})$. Theorem 7.3 gives a bound on the rate of convergence of the method of averaged projections as

$$r_{\mathrm{av}} \leq 1 - \frac{1}{2\kappa^2}.$$

Notice that each iteration involves two projections: one onto each of the sets F_1 and F_2. On the other hand, Corollary 5.18 and (5.19) give a bound on the rate of convergence of the method of alternating projections as

$$r_{\text{alt}} \le 1 - \frac{1}{\kappa^2},$$

and each iteration involves just one projection. Thus, we note that our bound on the rate of alternating projections r_{alt} is always better than the bound on the rate of averaged projections r_{av}. From the perspective of this analysis, averaged projections seems to have no advantage over alternating projections, although our proof of linear convergence for alternating projections needs a super-regularity assumption not necessary in the case of averaged projections.

8 Prox-Regularity and Averaged Projections

If we assume that the sets $F_1, F_2, \ldots, F_m$ are prox-regular, then we can refine our understanding of local convergence for the method of averaged projections using a completely different approach, explored in this section.

Proposition 8.1 *Around any point $\bar{x}$ at which the set $F \subset \mathbf{E}$ is prox-regular, the squared distance to F is continuously differentiable, and its gradient $\nabla d_F^2 = 2(I - P_F)$ has Lipschitz constant 2.*

Proof This result corresponds essentially to [40, Proposition 3.1], which yields the smoothness of d_F^2 together with the gradient formula. This proof of this proposition also shows that for any sufficiently small $\delta > 0$, all points $x_1, x_2 \in \mathbf{E}$ near $\bar{x}$ satisfy the inequality

$$\langle x_1 - x_2, P_F(x_1) - P_F(x_2) \rangle \ge (1 - \delta) \| P_F(x_1) - P_F(x_2) \|^2$$

(see "Claim" in [40, p. 5,239]). Consequently, we have

$$\begin{aligned}
&\left\| (I - P_F)(x_1) - (I - P_F)(x_2) \right\|^2 - \|x_1 - x_2\|^2 \\
&\quad = \left\| (x_1 - x_2) - \left(P_F(x_1) - P_F(x_2) \right) \right\|^2 - \|x_1 - x_2\|^2 \\
&\quad = -2 \langle x_1 - x_2, P_F(x_1) - P_F(x_2) \rangle + \left\| P_F(x_1) - P_F(x_2) \right\|^2 \\
&\quad \le (2\delta - 1) \left\| P_F(x_1) - P_F(x_2) \right\|^2 \le 0,
\end{aligned}$$

provided we choose $\delta \le 1/2$.   □

As before, consider sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ and a point $\bar{x} \in \bigcap_i F_i$, but now let us suppose moreover that each set F_i is prox-regular at $\bar{x}$. Define a function $f : \mathbf{E} \to \mathbf{R}$ by

$$f = \frac{1}{2m} \sum_{i=1}^{m} d_{F_i}^2. \tag{8.2}$$

This function is half the *mean-squared-distance* from the point x to the set system $\{F_i\}$. According to the preceding result, f is continuously differentiable around $\bar{x}$, and its gradient

$$\nabla f = \frac{1}{m} \sum_{i=1}^{m} (I - P_{F_i}) = I - \frac{1}{m} \sum_{i=1}^{m} P_{F_i} \tag{8.3}$$

is Lipschitz continuous with constant 1 on a neighborhood of $\bar{x}$. The method of averaged projections constructs the new iterate $x_+ \in \mathbf{E}$ from the old iterate $x \in \mathbf{E}$ via the update

$$x_+ = \frac{1}{m} \sum_{i=1}^{m} P_{F_i}(x) = x - \nabla f(x), \tag{8.4}$$

so we can interpret it as the method of steepest descent with a step size of one when the sets F_i are all prox-regular. To understand its convergence, we return to our linear regularity assumption.

The condition modulus controls the behavior of normal vectors not just at the point $\bar{x}$ but also at nearby points.

Proposition 8.5 (Local effect of condition modulus) *Consider closed sets $F_1, F_2, \ldots,$ $F_m \subset \mathbf{E}$ having linearly regular intersection at a point $\bar{x} \in \cap F_i$, and any constant $k > \mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x})$. Then for any points $x_i \in F_i$ sufficiently near $\bar{x}$, any vectors $y_i \in N_{F_i}(x_i)$ (for $i = 1, 2, \ldots, m$) satisfy the inequality*

$$\sqrt{\sum_i \|y_i\|^2} \le k \left\| \sum_i y_i \right\|.$$

Proof If the result fails, then we can find sequences of points $x_i^r \to \bar{x}$ in F_i and sequences of vectors $y_i^r \in N_{F_i}(x_i^r)$ (for $i = 1, 2, \ldots, m$) satisfying

$$\sqrt{\sum_i \|y_i^r\|^2} > k \left\| \sum_i y_i^r \right\|$$

for all $r = 1, 2, \ldots$. Define new vectors

$$u_i^r = \frac{1}{\sqrt{\sum_j \|y_j^r\|^2}} y_i^r \in N_{F_i}(x_i^r)$$

for each index $j = 1, 2, \ldots, m$ and r. Notice $\sum_i \|u_i^r\|^2 = 1$ and $\left\| \sum_i u_i^r \right\| < \frac{1}{k}$. For each $i = 1, 2, \ldots$, the sequence $u_i^1, u_i^2, \ldots$ is bounded, so after taking subsequences we can suppose it converges to some vector $u_i \in \mathbf{E}$, and since the normal cone N_{F_i} is closed as a set-valued mapping from F_i to $\mathbf{E}$, we deduce $u_i \in N_{F_i}(\bar{x})$. But then we have $\sum_i \|u_i\|^2 = 1$ and $\left\| \sum_i u_i \right\| \le \frac{1}{k}$, contradicting the definition of the modulus $\mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x})$. □

The size of the gradient of the mean-squared-distance function f, defined by (8.2), is closely related to the value of the function near a point of linearly regular intersection. To be precise, we have the following result.

Proposition 8.6 (Gradient of mean-squared-distance) *Consider prox-regular sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ having linearly regular intersection at a point $\bar{x} \in \cap F_i$, and any constant $k > \mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x})$. Then on a neighborhood of $\bar{x}$, the mean-squared-distance function*

$$f = \frac{1}{2m} \sum_{i=1}^{m} d_{F_i}^2$$

satisfies the inequalities

$$\frac{1}{2} \|\nabla f\|^2 \le f \le \frac{k^2 m}{2} \|\nabla f\|^2. \tag{8.7}$$

Proof Consider any point $x \in \mathbf{E}$ sufficiently near $\bar{x}$. Equation (8.3) implies $\nabla f(x) = \frac{1}{m} \sum_i y_i$, where $y_i = x - P_{F_i}(x) \in N_{F_i}(P_{F_i}(x))$ for each $i = 1, 2, \ldots, m$. By definition, we have $f(x) = \frac{1}{2m} \sum_i \|y_i\|^2$. Using inequality (2.4), we obtain

$$m^2 \|\nabla f(x)\|^2 = \left\| \sum_{i=1}^{m} y_i \right\|^2 \le m \sum_{i=1}^{m} \|y_i\|^2 = 2m^2 f(x).$$

But since x is sufficiently near $\bar{x}$, so are the projections $P_{F_i}(x)$, so

$$2m f(x) = \sum_i \|y_i\|^2 \le k^2 \left\| \sum_i y_i \right\|^2 = k^2 m^2 \|\nabla f(x)\|^2$$

by Proposition 8.5. The result now follows.   □

A standard argument now gives the main result of this section.

Theorem 8.8 (Q-linear convergence for averaged projections) *Consider prox-regular sets $F_1, F_2, \ldots, F_m \subset \mathbf{E}$ having linearly regular intersection at a point $\bar{x} \in \bigcap F_i$, and any constant $k > \mathrm{cond}(F_1, F_2, \ldots, F_m | \bar{x})$. Then for any averaged projection sequence $\{x_j\}$ with initial point x_0 sufficiently near $\bar{x}$, the mean-squared-distance*

$$f = \frac{1}{2m} \sum_{i=1}^{m} d_{F_i}^2$$

is reduced by at least a constant factor at each iteration:

$$f(x_{j+1}) \le \left(1 - \frac{1}{k^2 m}\right) f(x_j) \quad (j = 0, 1, 2, \ldots).$$

Proof Consider any point $x \in \mathbf{E}$ near $\bar{x}$. The function f is continuously differentiable around the minimizer $\bar{x}$, so the gradient $\nabla f(x)$ must be small, and hence the new iterate $x_+ = x - \nabla f(x)$ must also be near $\bar{x}$. Hence, as we observed after (8.3), the gradient ∇f has Lipschitz constant one on a neighborhood of the line segment $[x, x_+]$. Consequently,

$$
\begin{aligned}
&f(x_+) - f(x) \\
&= \int_0^1 \frac{d}{dt} f\big(x - t\nabla f(x)\big)\, dt = \int_0^1 \big\langle -\nabla f(x), \nabla f\big(x - t\nabla f(x)\big)\big\rangle dt \\
&= \int_0^1 \big(-\big\|\nabla f(x)\big\|^2 + \big\langle \nabla f(x), \nabla f(x) - \nabla f\big(x - t\nabla f(x)\big)\big\rangle \big)\, dt \\
&\leq -\big\|\nabla f(x)\big\|^2 + \int_0^1 \big\|\nabla f(x)\big\| \cdot \big\|\nabla f(x) - \nabla f\big(x - t\nabla f(x)\big)\big\|\, dt \\
&\leq -\big\|\nabla f(x)\big\|^2 + \int_0^1 \big\|\nabla f(x)\big\|^2 t\, dt = -\frac{1}{2}\big\|\nabla f(x)\big\|^2 \leq -\frac{1}{k^2 m} f(x),
\end{aligned}
$$

using Proposition 8.6. $\qquad\square$

A simple induction argument now gives an independent proof in the prox-regular case that the method of averaged projections converges linearly to a point in the intersection of the given sets. Specifically, the result above shows that mean-squared-distance $f(x_k)$ decreases by at least a constant factor at each iteration, and Proposition 8.6 shows that the size of the step $\|\nabla f(x_k)\|$ also decreases by a constant factor. Hence, the sequence (x_k) must converge R-linearly to a point in the intersection.

Comparing this result to Theorem 7.3 (linear convergence of averaged projections), we see that the predicted rates of linear convergence are the same. Theorem 7.3 guarantees that the squared distance to the intersection converges to zero with R-linear rate c^2 (for any constant $c \in (\bar{c}, 1)$). The argument gives no guarantee about improvements in a particular iteration: it only describes the asymptotic behavior of the iterates. By contrast, the argument of Theorem 8.8, with the added assumption of prox-regularity, guarantees the same behavior but with the stronger information that the mean-squared-distance decreases monotonically to zero with Q-linear rate c^2. In particular, each iteration must decrease the mean-squared-distance.

9 A Numerical Example

In this final section, we give a numerical illustration showing the linear convergence of alternating and averaged projections algorithms. Some major problems in signal or image processing come down to reconstructing an object from as few linear measurements as possible. Several recovery procedures from randomly sampled signals have been proved to be effective when combined with sparsity constraints (see, for instance, the recent developments of compressed sensing [18, 20]). These optimization problems can be cast as linear programs. However, for extremely large and/or

nonlinear problems, projection methods become attractive alternatives. In the spirit of compressive sampling, we use projection algorithms to optimize the compression matrix. This speculative example is meant simply to illustrate the theory rather than make any claim on real applications.

We consider the decomposition of images $x \in R^n$ as $x = Wz$ where $W \in \mathbf{R}^{n \times m}$ ($n < m$) is a "dictionary" (that is, a redundant collection of basis vectors). Compressed sensing consists in linearly reducing x to $y = Px = PWz$ with the help of a compression matrix $P \in \mathbf{R}^{d \times n}$ (with $d \ll n$); the inverse operation is to recover x (or z) from y. Compressed sensing theory gives sparsity conditions on z to ensure exact recovery [18, 20]. Reference [20], in fact, proposes a recovery algorithm based on alternating projections (on two convex sets). In general, we might want to design a specific sensing matrix P adapted to W, to ease this recovery process. An initial investigation on this question is [21]; we suggest here another direction, inspired by [9] and [22], where averaged projections naturally appear.

Candes and Romberg [9] showed that under orthogonality conditions, sparse recovery is more efficient when the entries $|(PW)_{ij}|$ are small. One could thus use the componentwise ℓ_∞ norm of PW as a measure of quality of P. This leads to the following feasibility problem: to find $U = PW$ such that $UU^\top = I$ and with the infinity norm constraint $\|U\|_\infty \le \alpha$ (for a fixed tolerance α). The sets corresponding to these constraints are given by

$$L = \left\{ U \in \mathbf{R}^{d \times m} : U = PW \right\},$$

$$M = \left\{ U \in \mathbf{R}^{d \times m} : UU^\top = I \right\},$$

$$C = \left\{ U \in \mathbf{R}^{d \times m} : \|U\|_\infty \le \alpha \right\}.$$

The first set L is a subspace, the second set M is a smooth manifold while the third C is convex; hence the three are prox-regular. Moreover, we can easily compute the projections. The projection onto the linear subspace L can be computed with a pseudo-inverse. The manifold M corresponds to the set of matrices U whose singular values are all ones; it turns out that naturally the projection onto M is obtained by computing the singular value decomposition of U, and setting singular values to 1 (apply, for example, Theorem 27 of [30]). Finally, the projection onto C comes by shrinking entries of U (specifically, we operate $\min\{\max\{u_{ij}, -\alpha\}, \alpha\}$ for each entry u_{ij}). This feasibility problem can thus be treated by projection algorithms, and hopefully a matrix $U \in L \cap M \cap C$ will correspond to a good compression matrix P.

To illustrate, we generate random entries (normally distributed) of the dictionary W (size 128×512, redundancy factor 4) and of an initial iterate $U_0 \in L$. (In practice, since the theory only guarantees *local convergence*, we would need a heuristic to find an initial iterate.) We fix $\alpha = 0.1$ and run the averaged projection algorithm, thereby computing a sequence of U_k that appear to be converging, as hoped, to a feasible solution to our problem. Furthermore, the convergence appears linear: Figure 1 shows

$$10 \log_{10} f(U_k) \quad \text{with } f(U) = d_L^2(U) + d_M^2(U) + d_C^2(U)$$

for each iteration k. We observe $f(U_{k+1})/f(U_k) < 0.9627$ for all k, suggesting the expected local Q-linear convergence. Random examples are interesting for our simple

Fig. 1 Convergence of
averaged projection algorithm
for designing compression
matrix in compressed sensing
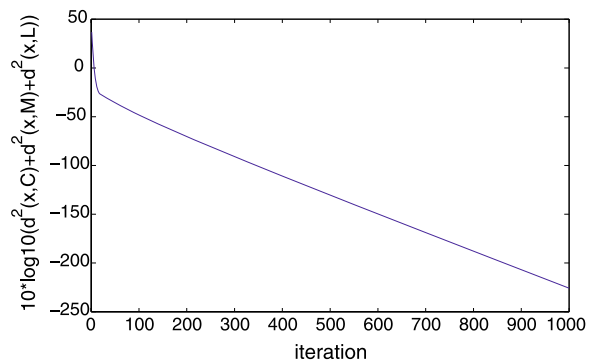
test of averaged projections: the challenging question of checking a priori the linear
regularity of the intersection of the three sets is open, but randomness seems to pre-
vent irregular solutions, providing α is not too small. So in this situation, we would
hope that the algorithm will converge locally linearly; this is indeed what the numer-
ical results in Fig. 1 suggest. We note furthermore that we tested iterated projections
on this problem (involving three sets, so not explicitly covered by Theorem 5.16).
We observed that the method still appears locally linearly convergent in practice, and
again, that the rate is better than for averaged projections.

This example illustrates how the projection algorithm behaves on random feasibil-
ity problems of this type. However, the potential benefits of using optimized compres-
sion matrix versus random compression matrix in practice are still unclear. Further
study and more complete testing have to be done for these questions; this is beyond
the scope of this paper.

References

1. F.J. Aragón Artacho, A.L. Dontchev, M.H. Geoffroy, Convergence of the proximal point method for
   metrically regular mappings, *ESAIM Proc.* **17**, 1–8 (2007).
2. H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Alternating minimization and projection methods for
   nonconvex problems. arXiv:0801.1780v1, 11 Jan. 2008.
3. A. Auslender, *Méthodes Numériques pour la Résolution des Problèmes d'Optimisation avec Con-
   traintes*. PhD thesis, Uni. Grenoble, 1969.
4. D. Aussel, A. Daniilidis, L. Thibault, Subsmooth sets: functional characterizations and related con-
   cepts, *Trans. Am. Math. Soc.* **357**, 1275–1301 (2004).
5. H.H. Bauschke, J.M. Borwein, On the convergence of von Neumann's alternating projection algorithm
   for two sets, *Set-Valued Anal.* **1**, 185–212 (1993).
6. H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems,
   SIAM Rev. **38**, 367–426 (1996).
7. H.H. Bauschke, P.L. Combettes, D.R. Luke, Phase retrieval, error reduction algorithm, and Fienup
   variants: A view from convex optimization, *J. Opt. Soc. Am.* **19**(7), 1334–1345 (2002).
8. L.M. Bregman, The method of successive projection for finding a common point of convex sets, *Sov.
   Math. Dokl.* **6**, 688–692 (1965).
9. E.J. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Probl.* **23**(3),
   969–986 (2007).
10. X. Chen, M.T. Chu, On the least squares solution of inverse eigenvalue problems, *SIAM J. Numer.
    Anal.* **33**, 2417–2430 (1996).
11. M.T. Chu, Constructing a Hermitian matrix from its diagonal entries and eigenvalues, *SIAM J. Matrix
    Anal.* **16**, 207–217 (1995).

12. F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, P.R. Wolenski, *Nonsmooth Analysis and Control Theory* (Springer, New York, 1998).
13. P.L. Combettes, T. Pennanen, Proximal methods for cohypomonotone operators, *SIAM J. Control Opt.* **43**, 731–742 (2004).
14. P.L. Combettes, H.J. Trussell, Method of successive projections for finding a common point of sets in metric spaces, *J. Optim. Theory Appl.* **67**(3), 487–507 (1990).
15. F. Deutsch, *Best Approximation in Inner Product Spaces* (Springer, New York, 2001).
16. F. Deutsch, H. Hundal, The rate of convergence for the cyclic projections algorithm I: angles between convex sets, *J. Approx. Theory* **142**, 36–55 (2006).
17. F. Deutsch, H. Hundal, The rate of convergence for the cyclic projections algorithm II: norms of nonlinear operators, *J. Approx. Theory* **142**, 56–82 (2006).
18. D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
19. A.L. Dontchev, A.S. Lewis, R.T. Rockafellar, The radius of metric regularity, *AMS Trans.* **355**, 493–517 (2003).
20. J. Romberg, E.J. Candès, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* **52**, 489–509 (2005).
21. M. Elad, Optimized projections for compressed-sensing, *IEEE Trans. Signal Process.* **55**, 5695–5702 (2007).
22. J. Fadili, G. Peyré, Personal communication, 2007.
23. K.M. Grigoriadis, E. Beran, Alternating projection algorithm for linear matrix inequalities problems with rank constraints, in *Advances in Linear Matrix Inequality Methods in Control* (SIAM, Philadelphia, 2000).
24. K.M. Grigoriadis, R.E. Skelton, Low-order control design for LMI problems using alternating projection methods, *Automatica* **32**, 1117–1125 (1996).
25. L.G. Gubin, B.T. Polyak, E.V. Raik, The method of projections for finding the common point of convex sets, *USSR Comput. Math. Math. Phys.* **7**, 1–24 (1967).
26. A.N. Iusem, T. Pennanen, B.F. Svaiter, Inexact versions of the proximal point algorithm without monotonicity, *SIAM J. Optim.* **13**, 1080–1097 (2003).
27. D. Klatte, B. Kummer, Optimization methods and stability of inclusions in Banach spaces, *Math. Program.* **17**, 305–330 (2009).
28. A.Y. Kruger, About regularity of collections of sets, *Set-Valued Anal.* **14**, 187–206 (2006).
29. A.S. Lewis, D.R. Luke, J. Malick, Local convergence for alternating and averaged nonconvex projections. arXiv:0709.0109v1, 2 Sep. 2007.
30. A.S. Lewis, J. Malick, Alternating projections on manifolds, *Math. Oper. Res.* **33**, 216–234 (2008).
31. B.S. Mordukhovich, Maximum principle in the problem of time optimal response with nonsmooth constraints, *J. Appl. Math. Mech.* **40**, 960–969 (1976).
32. B.S. Mordukhovich, Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings, *Dokl. Akad. Nauk BSSR* **28**, 976–979 (1984).
33. B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications* (Springer, New York, 2006).
34. B.S. Mordukhovich, Failure of metric regularity for major classes of variational systems, *Nonlinear Anal.* **69**, 918–924 (2008).
35. R. Orsi, Numerical methods for solving inverse eigenvalue problems for nonnegative matrices, *SIAM J. Matrix Anal.* **28**, 190–212 (2006).
36. R. Orsi, U. Helmke, J. Moore, A Newton-like method for solving rank constrained linear matrix inequalities, *Automatica* **42**, 1875–1882 (2006).
37. T. Pennanen, Local convergence of the proximal point algorithm and multiplier methods without monotonicity, *Math. Oper. Res.* **27**, 170–191 (2002).
38. G. Pierra, Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique, *Lect. Notes Comput. Sci.* **41**, 200–218 (1976).
39. G. Pierra, Decomposition through formalization in a product space, *Math. Program.* **28**, 96–115 (1984).
40. R.A. Poliquin, R.T. Rockafellar, L. Thibault, Local differentiability of distance functions, *AMS Trans.* **352**, 5231–5249 (2000).
41. J. Renegar, Incorporating condition measures into the complexity theory of linear programming, *SIAM J. Optim.* **5**, 506–524 (1995).
42. J. Renegar, Linear programming, complexity theory and elementary functional analysis, *Math. Program.* **70**, 279–351 (1995).

43. J. Renegar, Condition numbers, the barrier method, and the conjugate gradient method, *SIAM J. Optim.* **6**, 879–912 (1996).
44. R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis* (Springer, Berlin, 1998).
45. A. Shapiro, Existence and differentiability of metric projections in Hilbert space, *SIAM J. Optim.* **4**, 130–141 (1994).
46. A. Shapiro, On the asymptotics of constrained local M-estimation, *Ann. Stat.* **28**, 948–960 (2000).
47. A. Shapiro, F. Al-Khayyal, First-order conditions for isolated locally optimal solutions, *J. Optim. Theory Appl.* **77**, 189–196 (1993).
48. J.A. Tropp, I.S. Dhillon, R.W. Heath, T. Strohmer, Designing structured tight frames via an alternating projection method, *IEEE Trans. Inf. Theory* **51**, 188–209 (2005).
49. J. von Neumann, *Functional Operators*, vol. II (Princeton University Press, Princeton, 1950). Reprint of notes distributed in 1933.
50. C.A. Weber, J.P. Allebach, Reconstruction of frequency-offset Fourier data by alternating projection on constraint sets, in *24th Allerton Conference Proc.*, pp. 194–201. Urbana-Champaign, IL, 1986.
51. K. Yang, R. Orsi, Generalized pole placement via static output feedback: a methodology based on projections, *Automatica* **42**, 2143–2150 (2006).