Nonlinear Analysis 75 (2012) 1531–1546

Contents lists available at SciVerse ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

Local linear convergence of approximate projections onto regularized sets

D. Russell Luke *

Institute for Numerical and Applied Mathematics, Universität Göttingen, Germany

ARTICLE INFO

Communicated by Ravi Agarwal

MSC:
49M20
65K10
90C30

Keywords:
Alternating projections
Linear convergence
Ill-posed
Regularization
Metric regularity
Distance to ill-posedness
Variational analysis
Nonconvex
Extremal principle
Prox-regular

ABSTRACT

The numerical properties of algorithms for finding the intersection of sets depend to some extent on the regularity of the sets, but even more importantly on the regularity of the intersection. The alternating projection algorithm of von Neumann has been shown to converge locally at a linear rate dependent on the regularity modulus of the intersection. In many applications, however, the sets in question come from inexact measurements that are matched to idealized models. It is unlikely that any such problems in applications will enjoy metrically regular intersection, let alone set intersection. We explore a regularization strategy that generates an intersection with the desired regularity properties. The regularization, however, can lead to a significant increase in computational complexity. In a further refinement, we investigate and prove linear convergence of an approximate alternating projection algorithm. The analysis provides a regularization strategy that fits naturally with many ill-posed inverse problems, and a mathematically sound stopping criterion for extrapolated, approximate algorithms. The theory is demonstrated on the phase retrieval problem with experimental data. The conventional early termination applied in practice to unregularized, consistent problems in diffraction imaging can be justified fully in the framework of this analysis providing, for the first time, proof of convergence of alternating approximate projections for finite dimensional, consistent phase retrieval problems.

1. Introduction

The role of local regularity for nonconvex minimization problems or nonmonotone variational inequalities is well-established. In broad terms, a generalized equation is said to be ''regular'' (or ''metrically regular'') if the distance from a proposed solution to an exact solution can be bounded by a constant multiple of the model error of the proposed solution. A particular focus has been the proximal point algorithm and alternating projections [1–4].

It is often the case, however, that the problems in question are *ill-posed*; in other words, there is no constant of proportionality between the model error and the distance of an approximate solution to the true solution. For some algorithms, such an ill-posedness would not prevent the iterates from converging to a *best approximate solution*, but numerical performance will suffer. An example of such behavior can be observed with the classical alternating projection algorithm of von Neumann [5] applied to a general *feasibility problem*: that is, the problem of finding the intersection of sets. Ill-posedness for feasibility problems can be characterized by problem *inconsistency*, that is, the nonexistence of an intersection of the sets in question. More generally, the feasibility problem will be ill-posed if the intersection vanishes under arbitrarily small perturbations of the sets.

* Tel.: +49 5513910324.
E-mail address: r.luke@math.uni-goettingen.de.

0362-546X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2011.08.027

Local Linear Convergence of Approximate Projections onto Regularized Sets

D. Russell Luke *

August 10, 2011

Key words: alternating projections, linear convergence, ill-posed, regularization, metric regularity, distance to ill-posedness, variational analysis, nonconvex, extremal principle, prox-regular

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

Abstract

The numerical properties of algorithms for finding the intersection of sets depend to some extent on the regularity of the sets, but even more importantly on the regularity of the intersection. The alternating projection algorithm of von Neumann has been shown to converge locally at a linear rate dependent on the regularity modulus of the intersection. In many applications, however, the sets in question come from inexact measurements that are matched to idealized models. It is unlikely that any such problems in applications will enjoy metrically regular intersection, let alone set intersection. We explore a regularization strategy that generates an intersection with the desired regularity properties. The regularization, however, can lead to a significant increase in computational complexity. In a further refinement, we investigate and prove linear convergence of an approximate alternating projection algorithm. Our results provide for the first time a mathematically sound stopping criterion for alternating projections applied to consistent phase retrieval problems. The theory is demonstrated on a problem from a laser diffraction experiment.

1 Introduction

The role of local regularity for nonconvex minimization problems or nonmonotone variational inequalities is well-established. In broad terms, a generalized equation is said to be "regular" (or "metrically regular") if the distance from a proposed solution to an exact solution can be bounded by a constant multiple of the model error of the proposed solution. A particular focus has been the proximal point algorithm and alternating projections [1, 9, 11, 18].

It is often the case, however, that the problems in question are *ill-posed*; in other words, there is no constant of proportionality between the model error and the distance of an approximate solution to the true solution. For some algorithms such an ill-posedness would not prevent the iterates from converging to

*Institute for Numerical and Applied Mathematics, Universität Göttingen, Germany. The author was supported by the German DFG grant SFB755-C2.

a *best approximate solution*, but numerical performance will suffer. An example of such behavior can be observed with the classical alternating projection algorithm of von Neumann [22] applied to a general *feasibility problem:* that is, the problem of finding the intersection of sets. Ill-posedness for feasibility problems can be characterized by problem *inconsistency*, that is, the nonexistence of an intersection of the sets in question. More generally, the feasibility problem will be ill-posed if the intersection vanishes under arbitrarily small perturbations of the sets.

For the applications we have in mind, at least one of the sets in question comes from a finite precision measurement or calculation. It is quite reasonable to expect an inconsistency between the idealized model and the measured data, which can be represented as a perturbation of the idealized data set. When only two convex sets are involved, alternating projections can be shown to converge to *nearest points* [6, Theorem 4], however the *rate* of convergence will in general be arbitrarily slow. For other algorithms ill-posedness leads to instability in the sense that the iterates do not converge to a fixed point. The Douglas Rachford algorithm, for example, applied to inconsistent feasibility problems has no fixed points [4, 12, 14].

Insofar as ill-posed problems can be *regularized*, the theory cited above can be applied to numerical methods for the regularized problems. Our focus here is on a particular regularization for ill-posed feasibility problems and efficient *approximate* projection algorithms. The problem of nonconvex best approximation was considered in [13, 14] where the focus was on instability of the Douglas Rachford algorithm resulting from problem inconsistency. A relaxation of this algorithm was proposed that has fixed points for inconsistent problems and has been successful in practice [15]. As is often the case for relaxed projection algorithms, there is no systematic rule for choosing the relaxation parameter. It was shown in [14] that the size of the relaxation parameter at the solution is related to the optimal gap distance between the sets. This observation suggests a different approach to algorithmic design that is based on regularization of the underlying problem rather than stabilization of the algorithm as was the focus in [13].

We further develop this viewpoint here, where we study *local* regularization of the underlying problem while retaining the character of the original problem. In particular, we expand one of the sets in order to create an intersection with all the desired regularity properties described in [11]. The strategy is a local regularization in the sense that indicator functions are still used as the central penalty function, in contrast to [14] where the indicator function was relaxed to a distance function. One then can apply any number of algorithms for finding the intersection of regularized sets. We are particularly interested in projection algorithms and specifically the classical alternating projection algorithm. We show in section 3 that, for the problems of interest to us, such a regularization of the sets results in a significant increase in the complexity of computing the corresponding projections. To address computational complexity of the regularized problem we consider approximate alternating projections based on the projection operators of the original, unregularized problem. An approximate algorithm is stated in section 4. We prove local linear convergence of this algorithm to a solution of the regularized problem under regularity assumptions that are natural for regularized problems. In section 5 we apply specific approximation motivated in section 3 to the approximate projection algorithm. We demonstrate the effectiveness of this approximation in section 6 with an example from diffraction imaging with real experimental data. We do not claim that the approximate alternating projection algorithm is the best, or even a very good strategy for solving this particular problem. However to our knowledge, our analysis yields the first mathematically sound stopping criteria for alternating projections applied to the phase retrieval problem. Our goal is to demonstrate the theory and to motivate the adaptation of our proposed regularization and approximation to more sophisticated projection algorithms.

2 Notation, Definitions and Basic Theory

We begin with basic theory and notation. For the most part, we present only the results with pointers to the literature for interested readers. The setting we consider is finite dimensional Euclidean space $\mathbb{E}$. The closed unit ball centered at x is denoted by $\mathbb{B}(x)$; when it is centered at the origin, we simply write $\mathbb{B}$. We denote the open interval from a to b by (a, b); the closed interval is denoted as usual by $[a, b]$.

Given a set $C \subset \mathbb{E}$, we define the *distance function* and (multivalued) *projection* for C by

$$\begin{aligned} d_C(x) &= d(x, C) = \inf\{\|z - x\| : z \in C\} \\ P_C(x) &= \operatorname{argmin}\{\|z - x\| : z \in C\}. \end{aligned}$$

If C is closed, then the projection is nonempty. Following [17, Definition 1.6] we define the *normal cone* to a closed set $C \subset \mathbb{E}$ as follows:

Definition 2.1 (normal cone) *A vector v is* normal *to a closed set $C \subset \mathbb{E}$ at $\overline{x}$, written $v \in N_C(\overline{x})$ if there are sequences $x^k \to \overline{x}$ and $v^k \to v$ with*

$$v^k \in \left\{ t(x^k - z) \;\middle|\; t \geq 0,\ z \in P_C(x^k) \right\} \qquad \textit{for all } k \in \mathbb{N}.$$

The vectors v^k are proximal normals *to C at $z \in P_C(x^k)$ and the cone of proximal normals at z is denoted $N_C^P(z)$.*

It follows immediately from the definition that the normal cone is a closed multifunction: for any sequence of points $x^k \to \overline{x}$ in C, any limit of a sequence of normals $v^k \in N_C(x^k)$ must lie in $N_C(\overline{x})$. The relation of the projection to the normal cone is also evident from the definition:

$$z \in P_C(x) \quad \Rightarrow \quad x - z \in N_C(z). \tag{2.1}$$

Notice too that $N_C(x) = \{0\} \iff x \in \operatorname{int} C$.

Definition 2.2 (basic set intersection qualification) *A family of closed sets $C_1, C_2, \ldots C_m \subset \mathbb{E}$ satisfies the basic set intersection qualification at a point $\overline{x} \in \cap_i C_i$, if the only solution to*

$$\sum_{i=1}^{m} y_i = 0, \quad y_i \in N_{C_i}(\overline{x}) \quad (i = 1, 2, \ldots, m)$$

is $y_i = 0$ for $i = 1, 2, \ldots, m$. We say that the intersection is strongly regular *at $\overline{x}$ if the basic set constraint qualification is satisfied there.*

In the case $m = 2$, this condition can be written

$$N_{C_1}(\bar{x}) \cap -N_{C_2}(\bar{x}) = \{0\}.$$

The two set case is is called the *basic constraint qualification for sets* in [17, Definition 3.2] and has its origins in the the *generalized property of nonseparability* [16] which is the n-set case. It was later recovered as a dual characterization of what is called *strong regularity* of the intersection in [10, Proposition 2]. This property was called *linear regularity* in [11]. The case of two sets also yields the following simple quantitative characterization of strong regularity.

Proposition 2.3 (Theorem 5.16 of [11]) *Suppose that C_1 and C_2 are closed subsets of $\mathbb{E}$. The intersection $C_1 \cap C_2$ satisfies the basic set intersection qualification at $\overline{x}$ if and only if the constant*

$$\overline{c} \;:=\; \max\{\langle u,\; v\rangle \mid u \in N_{C_1}(\overline{x}) \cap \mathbb{B},\; v \in -N_{C_2}(\overline{x}) \cap \mathbb{B}\} < 1. \tag{2.2}$$

Definition 2.4 (angle of regular intersections) *We say that the intersection $C_1 \cap C_2$ is* strongly regular at $\overline{x}$ with angle $\overline{\theta} := \cos^{-1}(\overline{c}) > 0$ *where $\overline{c}$ is given by* (2.2).

In order to achieve linear rates of convergence of alternating projections to the intersection of sets, we require pointwise strong regularity of the intersection [11]. In the absence of this property the above definitions suggest a general regularization philosophy: *promote strong regularity.* This is most obviously achieved by augmenting at least one of the sets by some ϵ ball: $C_1(\epsilon) = C_1 + \epsilon\mathbb{B}$, for instance. Similar ideas been used extensively in the development of proximally smooth sets by Clarke, Stern and Wolenski [7]. We pursue this idea in section 3 with the generalization that the ball, or "tube" around the set of interest is with respect to a generic metric in the image space of a continuous mapping, the tube having no relation to the native space in which the projectors onto the sets are defined.

Somewhat stronger results are possible when the sets have additional regularity. We call a set $C \subset \mathbb{E}$ is *prox-regular* at a point $\overline{x} \in C$ if the projection mapping P_C is single-valued around $\overline{x}$ [19]. Convex sets, in particular, are prox-regular. More generally, any set defined by C^2 equations and inequalities is prox-regular at any point satisfying the Mangasarian-Fromovitz constraint qualification, for instance.

Proposition 2.5 (angle of normals of prox-regular set) *Suppose the set $C \subset \mathbb{E}$ is prox-regular at the point $\overline{x} \in C$. Then for any constant $\delta > 0$, any points $y, z \in C$ near $\overline{x}$ and any normal vector $v \in N_C(y)$ satisfy the inequality*

$$\langle v,\; z - y\rangle \le \delta\|v\| \cdot \|z - y\|.$$

Proof. This is a special case of the same property for *super regular sets* ([11, Definition 4.3] and [11, Proposition 4.4]) since by [11, Proposition 4.9] prox-regularity implies super regularity.

Alternatively, for prox-regular sets we can proceed directly from [19, Proposition 1.2] which shows that, for any sequences of points $y^k, z^k \in C$ converging to $\overline{x}$ and any corresponding sequence of normal vectors $v^k \in N_C(y^k)$, there exist constants $\epsilon, \rho > 0$ such that

$$\left\langle \frac{\epsilon}{2\|v^k\|} v^k,\ z^k - y^k \right\rangle \le \frac{\rho}{2}\|z^k - y^k\|^2$$

for all large k. Since for any fixed $\delta > 0$ we will eventually have $\|z^k - y^k\| \le \frac{\delta\epsilon}{\rho}$, it follows that

$$\langle v^k,\ z^k - y^k\rangle \le \delta \|v^k\| \cdot \|z^k - y^k\|$$

for k large enough. □

The next result provides bounds on the angle between sets in the neighborhood of a point in a strongly regular intersection of a closed and a prox-regular set. In [11, Theorem 5.2] implications (2.3) and (2.4) are used to characterize sets for which linear convergence of the alternating projections algorithm holds. We do not seek such generality here and are content with identifying classes of sets which satisfy these conditions, namely prox-regular sets. The proof of the following assertion can be found in the proof of Theorem 5.16 of [11].

Proposition 2.6 *Let $M, C \subset \mathbb{E}$ be closed. Suppose that C is prox-regular at a point $\overline{x} \in M \cap C$ and that M and C have strongly regular intersection at $\overline{x}$ with angle $\overline{\theta}$. Define $\overline{c} := \cos(\overline{\theta})$ and fix the constant c' with $\overline{c} < c' < 1$. There exists a constant $\epsilon > 0$ such that*

$$\left.\begin{array}{ll} x \in M \cap (\overline{x} + \epsilon\mathbb{B}), & u \in -N_M(x) \cap \mathbb{B} \\ y \in C \cap (\overline{x} + \epsilon\mathbb{B}), & v \in N_C(y) \cap \mathbb{B} \end{array}\right\} \implies \langle u,\ v\rangle \le c', \tag{2.3}$$

and, for some constant $\delta \in [0, \frac{1-c'}{2})$,

$$\left.\begin{array}{rcl} y, z & \in & C \cap (\overline{x} + \epsilon\mathbb{B}) \\ v & \in & N_C(y) \cap \mathbb{B} \end{array}\right\} \implies \langle v,\ z - y\rangle \le \delta\|z - y\|. \tag{2.4}$$

In what follows, we define an approximate alternating projection algorithm in terms of the distance of the normal cone associated with the approximate projection to the "true" normal cone. In order to guarantee that for our proposed approximation we can get arbitrarily close to the true projection, we need the notion of convergence of the associated normal cone mappings. Let $S : \mathbb{E} \rightrightarrows \mathbb{Y}$ denote a set-valued mapping where $\mathbb{Y}$ is another Euclidean space. We define the domain of S to be the set of points whose image is not empty, that is

$$\text{dom } S := \{x \mid S(x) \neq \emptyset\}.$$

Following [20, Definition 4.1] we define continuous set-valued mappings relative to some subset D as those which are both outer and inner semicontinuous relative to D.

Definition 2.7 (continuity of set-valued mappings) *A set-valued mapping $S : \mathbb{E} \rightrightarrows \mathbb{Y}$ is continuous at a point $\overline{x} \in D$ relative to $D \subset \mathbb{E}$ if*

$$S(\overline{x}) \subset \left\{ y \,\middle|\, \forall\ x^k \underset{D}{\to} \overline{x},\ \exists\ K > 0 \text{ such that for } k > K,\ y^k \to y \text{ with } y^k \in S(x^k) \right\}$$

and

$$\{y \mid \exists\, x^k \underset{D}{\to} \overline{x},\ \exists\, y^k \to y \ \ \text{with } y^k \in S(x^k)\} \subset S(\overline{x})$$

where $\underset{D}{\to}$ *indicates that the sequence lies within* D. *We denote this as* $S(x) \to S(\overline{x})$ *for all sequences* $x \underset{D}{\to} \overline{x}$.

3 The problem

In this section we formulate our abstract problem and motivate the regularization and approximation strategies that we propose. Our initial, naive problem formulation involves finding points $x \in C \subset \mathbb{E}$, a Euclidean space, that explain some measurement $b \in \mathbb{Y}$ modeled as the image of the continuous mapping $g : \mathbb{E} \to \mathbb{Y}$, that is

$$\text{Find } x \in C \cap M_0$$

for

$$M_0 := \{x \in \mathbb{E} \mid g(x) = b\}.$$

The set C usually captures a qualitative feature of solutions, such as nonnegativity, or a prescribed support. If b is a physical/empirical measurement, it is likely that the intersection is empty, or that the solution consists only of extremal points. In the case of measurements with discrepancies modeled by statistical noise, the noise could be Gaussian or Poisson distributed (among still other possibilities). To accommodate a variety of instances we consider the following regularizations of the set M_0:

$$M_\epsilon := \{x \in \mathbb{E} \mid d_\phi(g(x), b) \le \epsilon\} \tag{3.1}$$

where $\epsilon \ge 0$ and d_ϕ is a *Bregman distance* defined by

$$d_\phi(z, y) := \phi(z) - \phi(y) - \phi'(y)(z - y)$$

for $\phi : \mathbb{Y} \to (-\infty, +\infty]$ strictly convex and differentiable on $\operatorname{int}(\operatorname{dom}\phi)$. The Bregman distance with $\phi := \frac{1}{2}\|\cdot\|^2$ corresponds to the Euclidean norm which is appropriate for Gaussian noise. If $\mathbb{Y} = \mathbb{R}^m$ and

$$\phi(y) = \sum_{j=1}^{m} h(y_j) \quad \text{for } h(t) := \begin{cases} t\log t - t & \text{for } t > 0 \\ 0 & \text{for } t = 0 \\ +\infty & \text{for } t < 0 \end{cases}$$

then the Bregman distance leads to the *Kullback-Leibler divergence*,

$$d_\phi(z, y) = KL(x, y) := \sum_{j=1}^{m} z_j \log \frac{z_j}{y_j} + y_j - z_j. \tag{3.2}$$

The Kullback-Leibler divergence is appropriate for Poisson noise.

Remark 3.1 The regularization (3.1) bears some resemblance to closed neighborhoods of the type $X(\epsilon) := \{x \mid d(x, X) \le \epsilon\}$ considered by Clarke, Stern and Wolenski [7] in their development of proximally smooth sets, except that the neighborhood around the set of interest is with respect to a generic metric in the image space of a continuous mapping, the neighborhood having no relation to the metric upon which the projectors onto the sets are defined. Still, we will rely on prox-regularity of the regularized set for the approximation strategy discussed in Section 5. □

Regardless of the metric, the first algorithm we consider for finding this intersection is the classical alternating projection algorithm.

Algorithm 3.2 (exact alternating projections) Choose $x^0 \in C$. For $k = 1, 2, 3, \dots$ generate the sequence $\{x^{2k}\} \subset C$ with $x^{2k} \in P_C(x^{2k-1})$ where the sequence $\{x^{2k+1}\}$ consists of points $x^{2k+1} \in P_{M_\epsilon}(x^{2k})$.

We show next that the projection onto the fattened set M_ϵ could be considerably more costly to calculate than for the unregularized set M_0. This motivates the approximate projection algorithm studied in section 4

We want to compute

$$x^* \in P_{M_\epsilon}(\widehat{x}) := \operatorname{argmin}_{x \in M_\epsilon} \tfrac{1}{2}\|x - \widehat{x}\|^2.$$

Assume $d_\phi(g(\widehat{x}), b) > \epsilon$, then we seek a solution on the ϵ-sphere around b with respect to d_ϕ. This is an instance of a *trust region* problem.

Suppose that $\overline{x} \in P_{M_\epsilon}(\widehat{x})$ and that the standard constraint qualification holds, that is

$$-\nabla d_\phi(g(\overline{x}, b))^* \eta = 0, \quad \eta \geq 0 \quad \Longrightarrow \quad \eta = 0. \tag{3.3}$$

Then

$$\begin{aligned} (\overline{x} - \widehat{x}) + \nabla d_\phi(g(\overline{x}), b)^* \overline{\eta} &= 0 \qquad (\overline{\eta} \geq 0) && (3.4) \\ d_\phi(g(\overline{x}), b) - \epsilon &= 0. && (3.5) \end{aligned}$$

These are the standard KKT conditions (see, for example [20, Theorem 10.6]). Numerical methods for computing the projection $P_{M_\epsilon}(\widehat{x})$ involve solving a possibly large-scale nonlinear system of equations with respect to x and η; this could well be as difficult to solve as the original problem.

Example 3.3 (affine subspaces) Let $\mathbb{E} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^m$ with $m < n$. Take g to be the linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ and $d_\phi(x, y) = \frac{1}{2}\|x - y\|^2$ (that is, $\phi(x) = \frac{1}{2}\|x\|^2$). The projection can then be written as the solution to a quadratically constrained quadratic program:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\|x - z\|^2 \\ \text{subject to} & \frac{1}{2}\|Ax - b\|^2 \leq \epsilon. \end{array}$$

For small problem sizes this can be efficiently solved via interior point methods. Still, even the most efficient numerical methods cannot compare to computing the projection onto the affine space $M_0 := \{x \in \mathbb{E} \mid Ax = b\}$ which has the trivial closed form

$$P_{M_0}(z) = (I - A^T(AA^T)^{-1}A)z + A^T(AA^T)^{-1}b.$$

This suggests an alternative strategy for computing the projection onto the "fattened" set.

Indeed, we can efficiently compute the projection $P_{M_\epsilon}(z)$ as a convex combination of the points $y = P_{M_0}(z)$ and z

$$x^* = \lambda_\epsilon z + (1 - \lambda_\epsilon)y$$

where $\lambda_\epsilon \in [0, 1)$ solves $\frac{1}{2}(1 - \lambda)^2\|z - y\|^2 = \epsilon$. This also has a closed form: the quadratic formula. For general Bregman distances such shortcuts are not available, but this forms the basis for our approximations. □

Example 3.4 (boxes) Let $\mathbb{E} = \mathbb{Y} = \mathbb{R}^n$. Define $g : \mathbb{R}^n \to \mathbb{R}^n$ by

$$g(x) = \left(|x_1|^2, \dots, |x_n|^2\right)^T$$

and, again, let the distance d_ϕ be the standard normalized squared Euclidean distance to some point $b \in \mathbb{R}^n_+$. The projection can then be written as the solution to the nonconvex optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\|x - \widehat{x}\|^2 \\ \text{subject to} & \frac{1}{2}\sum_{j=1}^n (|x_j|^2 - b_j)^2 = \epsilon. \end{array}$$

Notice that the corresponding set M_ϵ is not convex: the origin is projected in the positive and negative direction in each component. Generally, nonconvex problems are hard to solve. On the other hand, the projection onto the box with length $2b$, $y = (y_1, \dots, y_n)^T \in P_{M_0}(\widehat{x})$, is trivial and has the form

$$y_j \begin{cases} = b_j \frac{\widehat{x}_j}{|\widehat{x}_j|} & \widehat{x}_j \neq 0 \\ \in \{-b_j, b_j\} & \widehat{x}_j = 0. \end{cases}$$

See [5] for analysis of this projection in higher dimensional product spaces.

For this example there is no shortcut to computing the projection P_{M_ϵ} for $\epsilon > 0$, but we show below that the convex combination of the projection of $\widehat{x}$ onto M_0 and $\widehat{x}$ is a effective approximation that still yields linear rates of convergence for the method of alternating projections for finding the intersection of $M_\epsilon \cap C$. □

Remark 3.5 We note that in both of the above examples the constraint qualification (3.3) is no longer satisfied in the limit $\epsilon = 0$ for the set M_ϵ. This obviously does not prevent us from calculating the projection onto the set M_0. Indeed, as we showed, the projection sometimes even has an explicit representation. □

4 Inexact alternating projections

There is more than one way to formulate inexact algorithms. One template for this is to add summable error terms to the operators involved in the exact algorithm. Another approach – the one we take here – is less general but has a more geometric appeal. More to the point, it is appropriate for our intended application.

Given two iterates $x^{2k-1} \in M$ and $x^{2k} \in C$, a necessary condition for the new iterate x^{2k+1} to be an exact projection on M, that is $x^{2k+1} \in P_M(x^{2k})$, is

$$\|x^{2k+1} - x^{2k}\| \leq \|x^{2k} - x^{2k-1}\| \quad \text{and} \quad x^{2k} - x^{2k+1} \in N_M(x^{2k+1}).$$

In a modification of [11] we assume only that we choose the odd iterates x^{2k+1} to satisfy a relaxed version of this condition, where we replace the second part by the assumption that the distance of the normalized direction of the current step to the normal cone to M at the intersection of the boundary of M with the line segment between x^{2k+1} and x^{2k} is small.

Consider the following inexact alternating projection iteration for finding the intersection of two sets $M, C \subset \mathbb{E}$.

Algorithm 4.1 (inexact alternating projections) Fix $\gamma > 0$ and choose $x^0 \in C$ and $x^1 \in M$. For $k = 1, 2, 3, \ldots$ generate the sequence $\{x^{2k}\} \subset C$ with $x^{2k} \in P_C(x^{2k-1})$ where the sequence $\{x^{2k+1}\} \subset M$ satisfies

$$\|x^{2k+1} - x^{2k}\| \le \|x^{2k} - x^{2k-1}\|, \tag{4.1a}$$

$$x^{2k+1} = x^{2k} \quad \text{if } x_*^{2k+1} = x^{2k}, \tag{4.1b}$$

and

$$d_{N_M(x_*^{2k+1})}(\widehat{z}^k) \le \gamma \tag{4.1c}$$

for

$$x_*^{2k+1} = P_{M \cap \{x^{2k} - \tau \widehat{z},\ \tau \ge 0\}}(x^{2k})$$

and

$$\widehat{z}^k := \begin{cases} \frac{x^{2k} - x^{2k+1}}{\|x^{2k} - x^{2k+1}\|} & \text{if } x_*^{2k+1} \ne x^{2k} \\ 0 & \text{if } x_*^{2k+1} = x^{2k}. \end{cases}$$

Note that the odd iterates x^{2k+1} can lie on the interior of M. This is the major difference between Algorithm 4.1 and the one specified in [11] where all of the iterates are assumed to lie on the boundary of M. We include this feature to allow for *extrapolated* iterates in the case where M has interior. Extrapolation, or over relaxation, is a common technique for accelerating algorithms, though its basis is rather heuristic. Empirical experience reported in the literature shows that extrapolation can be quite effective (see [8, 21]). The algorithm given in Theorem 5.1 below explicitly includes extrapolation. Our numerical results at the end of this paper do not contradict the conventional experience with extrapolation.

Lemma 4.2 and Theorem 4.4 below were sketched in [11, Theorem 6.1] for the variation of Algorithm 4.1 just described.

Lemma 4.2 *Let $M, C \subset \mathbb{E}$ be closed. Suppose that C is prox-regular at a point $\overline{x} \in M \cap C$ and that M and C have strongly regular intersection at $\overline{x}$ with angle $\overline{\theta}$. Define $\overline{c} := \cos(\overline{\theta})$ and fix the constants c with $\overline{c} < c < 1$ and $\gamma < \sqrt{1 - c^2}$. Then there is an $\epsilon > 0$ such that the iterates of Algorithm 4.1 satisfy*

$$\left.\begin{array}{ll} \|x^{2k+1} - \overline{x}\| & \le \frac{\epsilon}{2} \\ \|x^{2k+1} - x^{2k}\| & \le \frac{\epsilon}{2} \end{array}\right\} \implies \|x^{2k+2} - x^{2k+1}\| \le \eta \|x^{2k+1} - x^{2k}\| \tag{4.2}$$

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

Proof. Fix c' with $\overline{c} < c' < c < 1$ and define $\delta = \frac{1}{2}(\eta - \eta')$ where $\eta' = c'\sqrt{1 - \gamma^2} + \gamma\sqrt{1 - c'^2}$. ($\delta > 0$ since, as is easily verified, $c\sqrt{1 - \gamma^2} + \gamma\sqrt{1 - c^2}$ increases monotonically with respect to c on $[0, 1]$.) Since C is prox-regular at $\overline{x}$ and the intersection is strongly regular, by Proposition 2.6 for this δ there is an $\epsilon > 0$ such that implications (2.3) and (2.4) hold. We apply this result here.

The assumptions and the triangle inequality yield

$$\|x^{2k} - \overline{x}\| \le \|x^{2k} - x^{2k+1}\| + \|\overline{x} - x^{2k+1}\| \le \epsilon. \tag{4.3}$$

By the definition of x_*^{2k+1} we have $x_*^{2k+1} = (1-\lambda)x^{2k} + \lambda x^{2k+1}$ for some $\lambda \in [0, 1]$ so that

$$\begin{aligned}
\|x_*^{2k+1} - \overline{x}\| &= \|\lambda(x^{2k+1} - \overline{x}) + (1-\lambda)(x^{2k} - \overline{x})\| \\
&\le \lambda\|x^{2k+1} - \overline{x}\| + (1-\lambda)\|x^{2k} - \overline{x}\| \\
&\le \lambda\frac{\epsilon}{2} + (1-\lambda)\epsilon \le \epsilon \quad (\lambda \in [0, 1])
\end{aligned} \tag{4.4}$$

where the last inequality combines the left hand side of(4.2) and (4.3). Next, by the triangle inequality and the definition of the projection

$$\begin{aligned}
\|x^{2k+2} - \overline{x}\| &\le \|x^{2k+2} - x^{2k+1}\| + \|\overline{x} - x^{2k+1}\| \\
&\le \|x^{2k} - x^{2k+1}\| + \|\overline{x} - x^{2k+1}\| \le \epsilon.
\end{aligned} \tag{4.5}$$

If $x^{2k+1} = x^{2k}$ then this is a fixed point of the algorithm and the result is trivial. Similarly, if $x^{2k+1} = x^{2k+2}$, then $x^{2k+1} \in C \cap M$ and by the first condition in (4.1) this is a fixed point of the algorithm. So we assume that $x^{2k+1} \ne x^{2k}$ and define $\widehat{w} \in N_M(x_*^{2k+1})$ with $\|\widehat{w}\| = 1$ and $\widehat{u} := \frac{x^{2k+2} - x^{2k+1}}{\|x^{2k+2} - x^{2k+1}\|}$. Now applying Proposition 2.6 to x_*^{2k+1} satisfying (4.4) with $-\widehat{w} \in -N_M(x_*^{2k+1}) \cap \mathbb{B}$ and to x^{2k+2} satisfying (4.5) with $-\widehat{u} \in N_C(x^{2k+2}) \cap \mathbb{B}$ we have that for ϵ small enough

$$\langle \widehat{w},\ \widehat{u}\rangle = \langle -\widehat{w},\ -\widehat{u}\rangle \le c'. \tag{4.6}$$

In other words the angular separation between the unit vectors $\widehat{w}$ and $\widehat{u}$ is bounded below by $\arccos c'$.

On the other hand, define

$$\widehat{z} := \frac{x^{2k} - x^{2k+1}}{\|x^{2k} - x^{2k+1}\|}.$$

Our goal is to obtain a lower bound the angle between $\widehat{z}$ and $\widehat{u}$. If it were the case that $x^{2k+1} \in P_M(x^{2k})$ then $\widehat{z} = \widehat{w}$ and c' would already be our bound. But since x^{2k+1} only approximates the projection, we must work a little harder. Since the iterates satisfy (4.1), for some $w \in N_M(x_*^{2k+1})$ we have $\|w - \widehat{z}\| \le \gamma$.

There are two cases to consider. If $\widehat{z} = 0$, then we are done. Otherwise $\widehat{z}$ has length one, and

$$\frac{\|w\|^2 + 1 - \gamma^2}{2\|w\|} \le \left\langle \frac{w}{\|w\|},\ \widehat{z} \right\rangle.$$

Maximizing the left hand side as a function of $\|w\| \in [1-\gamma, 1+\gamma]$ yields the largest possible angular separation from $\widehat{z}$, that is

$$\langle \widehat{w},\ \widehat{z}\rangle \ge \sqrt{1-\gamma^2} \tag{4.7}$$

where $\widehat{w} = \frac{w}{\|w\|}$.

Note that $\gamma < \sqrt{1-c^2} < \sqrt{1-c'^2}$ for $c' < c$ so that $c' < \sqrt{1-\gamma^2}$. Thus, combining (4.6) and (4.7), we have

$$\begin{array}{ccc}
\langle \widehat{w},\ \widehat{z}\rangle \ge \sqrt{1-\gamma^2} & > & c' \ge \langle \widehat{w},\ \widehat{u}\rangle \\
& \Longleftrightarrow & \\
\arccos\langle \widehat{w},\ \widehat{z}\rangle \le \arccos(\sqrt{1-\gamma^2}) & < & \arccos c' \le \arccos\langle \widehat{w},\ \widehat{u}\rangle\,.
\end{array}$$

It follows immediately, then, that

$$\begin{aligned} 0 &< \arccos c' - \arccos(\sqrt{1-\gamma^2}) \\ &< \arccos\langle \widehat{w},\ \widehat{u}\rangle - \arccos\langle \widehat{w},\ \widehat{z}\rangle \le \arccos\langle \widehat{u},\ \widehat{z}\rangle \end{aligned}$$

which is equivalent to

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

Letting $\eta' = c'\sqrt{1-\gamma^2} + \gamma\sqrt{1-c'^2}$ and removing the normalization yields

$$\left\langle x^{2k} - x^{2k+1},\ x^{2k+2} - x^{2k+1}\right\rangle \le \eta' \|x^{2k} - x^{2k+1}\| \|x^{2k+2} - x^{2k+1}\|. \tag{4.8}$$

Now by our choice of ϵ, implication (2.4) holds for x^{2k} and $x^{2k+2} \in C \cap \{\overline{x} + \epsilon\mathbb{B}\}$ with $-\widehat{u} \in N_C(x^{2k+2}) \cap \mathbb{B}$, namely

$$\left\langle -\widehat{u},\ x^{2k} - x^{2k+2}\right\rangle \le \delta\|x^{2k+2} - x^{2k}\|$$

which is equivalent to

$$\left\langle x^{2k+2} - x^{2k+1},\ x^{2k+2} - x^{2k}\right\rangle \le \delta\|x^{2k+2} - x^{2k}\|\|x^{2k+2} - x^{2k+1}\|.$$

By the triangle inequality and the definition of the projection

$$\|x^{2k+2} - x^{2k}\| \le \|x^{2k+2} - x^{2k+1}\| + \|x^{2k+1} - x^{2k}\| \le 2\|x^{2k+1} - x^{2k}\|$$

so that

$$\left\langle x^{2k+2} - x^{2k+1},\ x^{2k+2} - x^{2k}\right\rangle \le 2\delta\|x^{2k+1} - x^{2k}\|\|x^{2k+2} - x^{2k+1}\|. \tag{4.9}$$

Adding (4.8) and (4.9) yields

$$\|x^{2k+2} - x^{2k+1}\|^2 \le (2\delta + \eta')\,\|x^{2k+1} - x^{2k}\|\|x^{2k+2} - x^{2k+1}\|,$$

which by our construction of δ yields

$$\|x^{2k+2} - x^{2k+1}\| \le \eta\|x^{2k+1} - x^{2k}\|$$

as claimed. □

Lemma 4.3 *With the same assumptions as Lemma 4.2, choose* x^0 *and* x^1 *so that*

$$\|x^1 - \overline{x}\| \le \|x^0 - \overline{x}\| = \beta < \frac{1-\eta}{4}\epsilon \tag{4.10}$$

where ϵ *is chosen to satisfy* (4.2)*. Let* $\eta = c\sqrt{1-\gamma^2} + \gamma\sqrt{1-c^2}$*. Then for all* $k \ge 0$

$$\|x^{2k+1} - \overline{x}\| \ \le\ 2\beta\frac{1-\eta^{k+1}}{1-\eta} < \frac{\epsilon}{2}, \tag{4.11a}$$

$$\|x^{2k+1} - x^{2k}\| \ \le\ \beta\eta^k < \frac{\epsilon}{2} \quad \textit{and} \tag{4.11b}$$

$$\|x^{2k+2} - x^{2k+1}\| \ \le\ \beta\eta^{k+1}. \tag{4.11c}$$

If in addition M is prox regular at $\overline{x}$, then for all $k \geq 0$

$$\begin{aligned} \|x^{k+1} - \overline{x}\| &\leq 2\beta\frac{1-\eta^{k+1}}{1-\eta} < \frac{\epsilon}{2}, && (4.12a)\\ \|x^{k+1} - x^{k}\| &\leq \beta\eta^{k} < \frac{\epsilon}{2} \quad \textit{and} && (4.12b)\\ \|x^{k+2} - x^{k+1}\| &\leq \beta\eta^{k+1}. && (4.12c) \end{aligned}$$

Proof. The proof is by induction. For the case $k = 0$ inequality (4.11a) holds trivially. Inequality (4.11b) follows from the triangle inequality and (4.10). Inequality (4.11c) then follows from (4.11a), (4.11b) and Lemma 4.2. Since for the case $k = 0$ inequalities (4.11a)-(4.11c) are equivalent to (4.12a)-(4.12c) this case is true whether M is prox-regular or not.

To show that these relations hold for $k + 1$ with M *not* prox regular, note that by (4.1) $\|x^{2k+3} - x^{2k+2}\| \leq \|x^{2k+2} - x^{2k+1}\|$. In light of (4.11c) this implies

$$\|x^{2k+3} - x^{2k+2}\| \leq \beta\eta^{k+1} < \frac{\epsilon}{2}. \tag{4.13}$$

This together with (4.11a) and (4.11c), yields

$$\begin{aligned} \|x^{2k+3} - \overline{x}\| &\leq \|x^{2k+3} - x^{2k+2}\| + \|x^{2k+2} - x^{2k+1}\| + \|x^{2k+1} - \overline{x}\|\\ &\leq \beta\eta^{k+1} + \beta\eta^{k+1} + 2\beta\frac{1-\eta^{k+1}}{1-\eta}\\ &\leq 2\beta\eta^{k+1} + 2\beta\frac{1-\eta^{k+1}}{1-\eta} = 2\beta\frac{1-\eta^{k+2}}{1-\eta} < \frac{\epsilon}{2}. && (4.14) \end{aligned}$$

Now, Lemma 4.2 applied to (4.13) and (4.14) yields

$$\|x^{2k+4} - x^{2k+3}\| \leq \eta\|x^{2k+3} - x^{2k+2}\| \leq \beta\eta^{k+2}. \tag{4.15}$$

As (4.13)-(4.15) are just (4.11a)-(4.11c) with k replaced by $k+1$, this completes the induction and the proof for the case where M is not prox regular.

If we assume, in addition, that M is prox regular, then by (4.12c)

$$\|x^{k+2} - x^{k+2}\| \leq \beta\eta^{k+1} < \frac{\epsilon}{2}. \tag{4.16}$$

This together with (4.12a) yields

$$\begin{aligned} \|x^{k+2} - \overline{x}\| \leq \|x^{k+2} - x^{k+1}\| + \|x^{k+1} - \overline{x}\| &\\ &\leq \beta\eta^{k+1} + \beta\frac{1-\eta^{k+1}}{1-\eta}\\ &= \beta\frac{1-\eta^{k+2}}{1-\eta} \leq \frac{\eta}{2} && (4.17) \end{aligned}$$

Now Lemma 4.2 with the rolls of C and M reversed, together with (4.12c) yields

$$\|x^{k+3} - x^{k+2}\| \leq \eta\|x^{k+2} - x^{k+1}\| \leq \beta\eta^{k+2} \tag{4.18}$$

Again, since (4.16)-(4.18) are just (4.12a)-(4.12c) with k replaced by $k+1$, this completes the induction and the proof. □

Theorem 4.4 (convergence of inexact alternating projections) *Let $M, C \subset \mathbb{E}$ and suppose C is prox-regular at a point $\overline{x} \in M \cap C$. Suppose furthermore that M and C have strongly regular intersection at $\overline{x}$ with angle $\overline{\theta}$. Define $\overline{c} := \cos(\overline{\theta}) < 1$ and fix the constants $c \in (\overline{c}, 1)$ and $\gamma < \sqrt{1-c^2}$. For x^0 and x^1 close enough to $\overline{x}$, the iterates in Algorithm 4.1 converge to a point in $M \cap C$ with R-linear rate*

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

If, in addition, M is prox-regular at $\overline{x}$, then the iterates converge with R-linear rate

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

Proof. We prove in detail the case where M is *not* assumed to be prox-regular. Choose x^0 and x^1 so that (4.10) holds with ϵ is chosen as in Lemma 4.2. Let $\eta = c\sqrt{1-\gamma^2} + \gamma\sqrt{1-c^2}$. To establish convergence of the sequence we check that the iterates form a Cauchy sequence. To see this, note that for any integer $k = 0, 1, 2, \ldots$ and any integer $j > 2k$, by (4.11b) and (4.11c) of Lemma 4.3 we have

$$\begin{aligned}
\|x^j - x^{2k}\| &\le \sum_{i=2k}^{j-1} \|x^{i+1} - x^i\| \\
&\le (\alpha+\beta)\left(\eta^k + 2\eta^{k+1} + 2\eta^{k+2} + \ldots\right) \\
&\le (\alpha+\beta)\frac{1+\eta}{1-\eta}\eta^k \le 2\beta\frac{1+\eta}{1-\eta}\eta^k.
\end{aligned}$$

Similarly, it can be shown that

$$\|x^{j+1} - x^{2k+1}\| \le (\alpha+\beta)\frac{\eta^{k+1}}{1-\eta} \le 2\beta\frac{\eta^{k+1}}{1-\eta}.$$

So the sequence is a Cauchy sequence and converges to some $\widehat{x} \in \mathbb{E}$. The fixed point of the sequence must belong to $M \cap C$ and satisfies

$$\|\widehat{x} - x^0\| \le 2\beta\frac{1+\eta}{1-\eta}.$$

Moreover, for all $j = 0, 1, 2, \ldots$

$$\|\widehat{x} - x^j\| \le \beta\eta^{j/2}\frac{1+\eta}{1-\eta}.$$

We conclude that convergence is R-linear with rate $\sqrt{\eta}$ as claimed.

The proof for the case where M is also prox-regular at $\overline{x}$ proceeds analogously using inequalities (4.12a)-(4.12c) of Lemma 4.2 instead. □

Note that the worse the approximation to the projection, the slower the convergence. As we showed in the previous section, the projection onto the unfattened set can be easier (sometimes *much* easier) to compute than the projection onto the fattened set, so although the rate of convergence suffers from taking only an approximate projection, we gain in the per-iteration complexity of calculating the projections.

5 Approximate alternating projections onto fattened sets

We now apply Algorithm 4.1 to finding the intersection of a prox-regular set C and a fattened set M_ϵ of the form (3.1). Motivated by the observation in section 3 that the projection onto the unregularized set M_0 can be easier to compute than the projection onto M_ϵ, we use P_{M_0} to approximate P_{M_ϵ}.

We examine when (4.1) holds if the odd iterates x^{2k+1} are chosen to lie on the line segment between the point x^{2k} and the projected point $P_{M_0}(x^{2k})$. Given an ϵ for which the conditions of Theorem 4.4 hold, for starting points close enough to a point of strong regularity $\overline{x}$, the iterates of Algorithm 4.1 will converge at an R-linear rate governed by the angle of intersection $C \cap M_\epsilon$ at $\overline{x}$ and by the accuracy of the approximate projection. In particular, we will prove

Theorem 5.1 *For M_ϵ defined by* (3.1) *and $C \subset \mathbb{E}$ closed, suppose that $M_\epsilon \cap C \neq \emptyset$ for all $\epsilon > 0$. Suppose that at a point $\overline{x} \in M_\epsilon \cap C$ the sets M_ϵ and C have strongly regular intersection with fixed angle $\overline{\theta}$ and that C is prox-regular there. In addition to the assumptions of Lemma 5.2, suppose that M_ϵ is prox-regular at all points $x \in (\overline{x} + \lambda\mathbb{B}) \cap M_\epsilon$ with nonzero proximal normals at points $x \in [(x_0 + \lambda\mathbb{B}) \cap M_\epsilon] \setminus \operatorname{int}(M_\epsilon)$.*

Define $\overline{c} := \cos(\overline{\theta}) < 1$ and fix the constants $c \in (\overline{c}, 1)$ and $\gamma < \sqrt{1 - c^2}$. Compute the sequence $\{x^{2k+1}\}$ of Algorithm 4.1 by

$$x^{2k+1} = (1 - \lambda_k)x^{2k} + \lambda_k x_0^{2k+1} \tag{5.1}$$

for $x_0^{2k+1} \in P_{M_0}(x^{2k})$ and $\lambda_k > 0$ chosen so that $x^{2k+1} \in M_\epsilon$.

There exist $\lambda_k > 0$ and $\epsilon > 0$ such that the iterates x^{2k+1} belong to M_ϵ and satisfy (4.1a) *and* (4.1c) *for all $k \in \mathbb{N}$. For such parameter values $\{\lambda_k\}_{k \in \mathbb{N}}$, ϵ and for x^0 and x^1 close enough to $\overline{x}$, the iterates of Algorithm 4.1 converge to a point in $M_\epsilon \cap C$ with R-linear rate*

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

The odd iterates of the proposed algorithm do not necessarily lie on the surface of the regularized set M_ϵ, but could be on the interior of this set. Were we computing true projections, all the odd iterates would lie on the boundary of M_ϵ – instead we take *larger* steps than the projections would indicate. In this sense, the algorithm defined in Theorem 5.1 is a regularized approximate alternating projection with *extrapolation*. The theorem does not tell us what such extrapolation buys us, but at least it says that we will not do any worse than without it.

We begin next developing the groundwork for the proof of Theorem 5.1.

Lemma 5.2 (level-boundedness) *Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces, and $\phi : \mathbb{Y} \to (-\infty, +\infty]$ be lsc, strictly convex and differentiable on* $\operatorname{int}(\operatorname{dom} \phi)$. *Define the function $f := d_\phi(g(\cdot), b)$ where $d_\phi(y, b)$ is the Bregman distance of y to the point $b \in \operatorname{dom} \phi$ and the function $g : \mathbb{E} \to \mathbb{Y}$ is continuous with* $\operatorname{range}(g) \subset \operatorname{dom} \phi$ *and satisfies*

$$\liminf_{|x| \to \infty} \frac{d_\phi(g(x), b)}{|x|} > 0. \tag{5.2}$$

Then the lower level sets of f, $\{x \in \mathbb{E} \mid f(x) \le \alpha\}$ for fixed $\alpha \in \mathbb{R}$, are compact. In particular, the set argmin f *is nonempty and compact and* $\inf f = \min f \ge 0$.

Proof. For easy reference we recall the definition of the Bregman distance:

$$f(x) := d_\phi(g(x), b) = \phi(g(x)) - \phi(b) - \langle \phi'(b),\ g(x) - b\rangle\,.$$

Since range$(g) \subset$ dom ϕ and $b \in$ dom ϕ there is an $x \in \mathbb{E}$ at which $f(x) < \infty$. Moreover, since ϕ is convex, the Bregman distance is bounded below by 0, hence $\inf f \ge 0$ and f is *proper* (that is, not everywhere equal to infinity, and does not take the value $-\infty$ on $\mathbb{E}$). Also f is lsc as the composition of the sum of a lsc function ϕ and a linear function $\langle \phi'(b),\ \cdot\rangle$ with a continuous function g. The lower level sets of f are therefore closed (see for instance [20, Theorem 1.6]). The coercivity condition (5.2) then implies that the lower level-sets are *bounded* [20, Corollary 3.27], thus the lower level sets are compact and argmin f is nonempty and compact. □

Theorem 5.3 (continuity of the level set mapping) *Let $f := d_\phi(g(\cdot), b)$ with ϕ, g, b and d_ϕ as in Lemma 5.2. The corresponding level-set mapping*

$$M(\alpha) := \{x \in \mathbb{E} \mid f(x) \le \alpha\} \tag{5.3}$$

is continuous on $[\overline{\epsilon}, \infty)$ where $\overline{\epsilon} := \min f$.

Proof. By Lemma 5.2 $M(\cdot)$ is compact and dom $M(\cdot) = [\overline{\epsilon}, \infty) \subset [0, \infty)$. Consequently the graph of $M(\cdot)$ is closed (in fact, closed-valued) in $\mathbb{E} \times \mathbb{R}$ and satisfies

$$\left\{y \;\middle|\; \exists\ \alpha^k \to \overline{\alpha},\ \exists\ y^k \to y \ \text{ with } y^k \in M(\alpha^k)\right\} \subset M(\overline{\alpha}) \quad \text{ for all } \overline{\alpha} \in \mathbb{R}. \tag{5.4}$$

On the other hand, the inverse of the level-set mapping (the epigraphical profile mapping)

$$M^{-1}(x) := \{\alpha \in \mathbb{R} \mid \alpha \ge f(x)\}$$

maps open sets to open sets relative to $[\overline{\epsilon}, \infty)$, that is $M^{-1}(O)$ is open relative to $[\overline{\epsilon}, \infty)$ for every open set $O \subset \mathbb{E}$. Thus by [20, Theorem 5.7] the level set mapping satisfies

$$\begin{aligned} &M(\overline{\alpha}) \subset \\ &\quad \left\{y \;\middle|\; \forall\ \alpha^k \underset{[\overline{\epsilon},\infty)}{\rightrightarrows} \overline{\alpha},\ \exists\ K > 0 \text{ such that for } k > K,\ y^k \to y \text{ with } y^k \in M(\alpha^k)\right\} \end{aligned} \tag{5.5}$$

for all $\overline{\alpha} \ge \overline{\epsilon}$. Since the right hand side of (5.5) is a subset of the left hand side of (5.4) we have equality of these limiting procedures, and thus continuity of $M(\cdot)$ on $[\overline{\epsilon}, \infty)$ according to Definition 2.7. □

Proposition 5.4 *Let $f := d_\phi(g(\cdot), b)$ with ϕ, g, b and d_ϕ be as in Lemma 5.2 and let $M(\alpha)$ be defined by (5.3). For $\{\alpha^k\} \subset [\overline{\epsilon}, \infty)$ with $\alpha^k \to \overline{\alpha}$ where $\overline{\epsilon} := \min f$, the corresponding sequence of projections onto $M(\alpha^k)$, $P_{M(\alpha^k)}$, converges graphically to $P_{M(\overline{\alpha})}$, that is*

$$\operatorname{gph} P_{M(\alpha^k)} \to \operatorname{gph} P_{M(\overline{\alpha})}.$$

Proof. Since $M(\alpha^k) \to M(\overline{\alpha})$ by Theorem 5.3, graphical convergence of the projection mapping follows from a minor extension of [20, Proposition 4.9] (see [20, Example 5.35]). □

In light of the discussion in section 3, our numerical strategy for approximating the projection to the regularized set M_ϵ defined by (3.1) will be to compute the intersection of the boundary of M_ϵ with line segment between the current iterate and the projection onto the unregularized set M_0. Specifically, for $x \notin M_\epsilon$ we define $x_0 = P_{M_0}(x)$ and calculate the point

$$x_\epsilon := (1-\tau_\epsilon)x + \tau_\epsilon x_0 \quad \text{where} \quad \tau_\epsilon := \min\{\tau > 0 \mid (1-\tau)x + \tau x_0 \in M_\epsilon\}. \tag{5.6}$$

The next proposition shows that this approximation can achieve any specified accuracy for sets with a certain regularity. This will then be used to guarantee that the approximation to the projection given by (5.6) satisfies (4.1c) on neighborhoods of a fixed point of Algorithm 4.1.

Proposition 5.5 (uniform normal cone approximation) *Let $\overline{\epsilon} > 0$ and M_ϵ ($\epsilon \in [0, \overline{\epsilon}]$) be defined by (3.1). Let $x_0 \in M_0$, and $(x_0 + \lambda\mathbb{B}) \cap (\mathbb{E} \setminus M_{\overline{\epsilon}}) \neq \emptyset$ for $\lambda > 0$ fixed. In addition to the assumptions of Lemma 5.2, suppose that M_ϵ ($\epsilon \in [0, \overline{\epsilon}]$) is prox-regular at all points $x \in (x_0 + \lambda\mathbb{B}) \cap M_\epsilon$ with nonzero proximal normals at points $x \in [(x_0 + \lambda\mathbb{B}) \cap M_\epsilon] \setminus \operatorname{int}(M_\epsilon)$. Then given any $\gamma > 0$ there exists an $\epsilon' \in (0, \overline{\epsilon}]$ such that for all $\epsilon \in (0, \epsilon']$*

$$d_{N_{M_\epsilon}(z_\epsilon)}\left(\frac{z - z_0}{\|z - z_0\|}\right) < \gamma. \tag{5.7}$$

holds where $z_0 = P_{M_0}(z)$, z_ϵ is given by (5.6) and z is any point near $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$.

Proof. Since for all $\epsilon \in [0, \overline{\epsilon}]$ the sets M_ϵ are prox-regular on $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$, all nonzero proximal normals to M_ϵ can be realized by an r-ball on open neighborhoods of points on $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$ for r small enough [19, Theorem 1.3.f]. There is thus a ball with radius $r_\epsilon > 0$ on which the nonzero proximal normals to M_ϵ can be realized uniformly on $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$. Also by assumption, the proximal normal cones to all points on the boundary of $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$ are nonzero. Thus, by Definition 2.1 the normal cone to M_ϵ at all points on the boundary of $(x_0 + \lambda\mathbb{B}) \cap M_\epsilon$ can be identified with the projection of points z in a r_ϵ-neighborhood of this boundary. The result then follows from Proposition 5.4, identifying the level set mapping $M(\epsilon)$ with the parameterized set M_ϵ. □

Remark 5.6 We conjecture that the assumption of prox-regularity and nontriviality of the proximal normal can be relaxed. The assumptions of Lemma 5.2 are used to guarantee graphical convergence of the projection mappings; the issue here is that the points on the boundary of the M_ϵ generated by (5.6) do not have to correspond to projections. Prox-regularity, and more restrictive still, the nontriviality of the proximal normals to M_ϵ on the boundary is used, in essence, locally to guarantee the reverse implication of (2.1). Definition 2.1 only relies on the *existence* of sequences of proximal normals whose limits constitute the normal cone. Our approximation scheme (5.6), in contrast, generates a specific sequence of points, which could conceivably correspond only to zero proximal normals without further assumptions on the regularity of M_ϵ, though we are unaware of a counterexample. That prox-regularity alone is not enough to assure that the proximal normal cone is nonzero is nicely illustrated by the set $M = \left\{ x \in \mathbb{R}^2 \;\middle|\; x_2 \geq x_1^{3/5} \right\}$ which is prox-regular at the origin, but has only a zero proximal normal cone there (see [20, Fig. 6-12.]). Obviously, such regularity will depend on the metric d_ϕ and the mapping g used in the construction of M_ϵ. □

Proof of Theorem 5.1. We show first that there are $\lambda_k > 0$ such that for any $\epsilon \geq 0$ the iterates x^{2k+1} lie in M_ϵ and satisfy (4.1a). Consider $\lambda_k = 1$ for all k. Then $x^{2k+1} = x_0^{2k+1} \in M_\epsilon$ for all k and all $\epsilon \geq 0$ and by the definition of the projection

$$\|x_0^{2k+1} - x^{2k}\| \leq \|x^{2k} - x_0^{2k-1}\|$$

which suffices to prove the claim.

Next, the existence of $\epsilon > 0$ such that (4.1c) is satisfied follows immediately from Proposition 5.5. Convergence then follows from Theorem 4.4. □

Remark 5.7 The theorem above guarantees convergence of Algorithm 4.1 with approximation strategy given by (5.1) for instances where the intersection of the unregularized problem need not be strongly regular. When the unregularized problem is inconsistent the strategy may fail. In particular, suppose that $M_0 \cap C = \emptyset$. Then for some $\overline{\epsilon}$ the intersection $M_\epsilon \cap C = \emptyset$ for all $\epsilon < \overline{\epsilon}$. If γ is such that (4.1c) is only satisfied for $\epsilon < \overline{\epsilon}$, then the proposed approximation will fail.

To the degree that the coupling between the regularization parameter ϵ and γ is *weak*, we can still obtain positive results. One instance where the coupling is very weak is if the fattened set has interior and $\overline{x}$ is some point in this interior. In this case $\overline{c} = 0$ in (2.2), γ can be arbitrarily close to 1 and the condition (4.1c) is almost trivial to satisfy. This is indeed the case for our intended application. Of course, the closer γ is to 1, that is, the worse our approximation of the true projection, the slower the convergence; so the trade off between efficient computations and rates of convergence must be balanced. The addition of extrapolation to the approximate algorithm is meant to mitigate any adverse effects of the approximation. The effectiveness of extrapolation is illustrated in the following section. □

6 An example from diffraction imaging

We present an application of the theory developed here to image reconstruction from laser diffraction experiments produced at the Institute for X-Ray Physics at the University of Göttingen. Shown in Figure 1 is the observed diffraction image produced by an object resembling a coffee cup that has been placed in the path of a helium-neon laser. The imaging model is

$$|Fx|^2 = b \tag{6.1}$$

where $b \in \mathbb{R}^n$ is the observed image intensity, F is a discrete Fourier transform, $|\cdot|^2$ is the componentwise (pixelwise) modulus-squared, and $x \in \mathbb{C}^n$ is the object

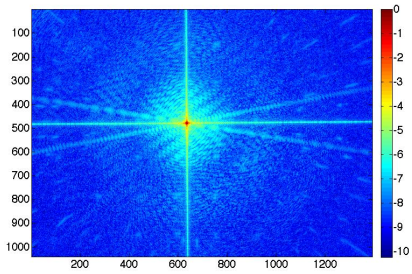

Figure 1: Diffraction image of real object.

to be found. The image is corrupted by noise modeled by a Poisson distribution. In the context of (3.1) the solution we seek lies in the fattened set

$$M_\epsilon := \left\{ x \,\middle|\, KL(|Fx|^2, b) \le \epsilon \right\} \tag{6.2}$$

for $KL(x, y)$ the Kullback-Leibler divergence given by (3.2). This set can be shown to be prox-regular everywhere with nonzero proximal normals at all points on the boundary. To this, we add the qualitative constraint that the object is nonnegative (that is, real) and lies within a specified support: for a given index set $\mathbb{J} \subset \{1, 2, \ldots, n\}$

$$C := \{x \in \mathbb{R}^n_+ \mid x_j = 0 \text{ for } j \in \mathbb{J}\}\,.$$

This set is not only prox-regular, but in fact convex.

Despite the good features of these sets, the problem is inconsistent/ill-posed. The set C is a set of real vectors, but the observation b is corrupted by noise. If b is not symmetric, as happens to be the case here, then the image cannot come from a real-valued object. Sometimes practitioners will "preprocess" the data by symmetrizing the raw data. If this is done, then the corresponding feasibility problem is provably consistent, and the results of Theorem 5.1 can be applied. In the numerical examples below, however, we choose to keep closer to the true nature of the experiment and demonstrate the success of Algorithm 4.1 as prescribed by Theorem 4.4 despite the absence of guarantees that the condition (4.1c) is satisfied.

The state of the art for iterative methods for solving this problem can be found in [15]. The main problem for these algorithms is the absence of a stopping criterion. Often what is done in practice is one algorithm (often the Douglas Rachford algorithm or variants [2, 3, 13]) is used to get close to a solution, and then alternating projections is used to refine the image according to the "eye-ball" norm. In the application literature alternating projections is often known as the "Error Reduction" algorithm. Different communities have different opinions as to what constitutes a stopping criteria, but in our reading of the application literature, none of the proposed criteria involve iterates approaching a numerical fixed point. Typical behavior of alternating projections onto the unregularized problem, together with the corresponding reconstruction are shown in Figure 2. The true object was a coffee cup, which can be seen in the upper left hand corner of the reconstruction in Figure 2, with the handle on the right hand side.

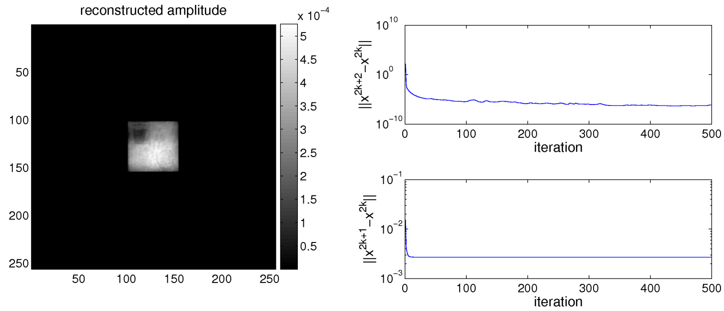

Figure 2: Reconstruction and behavior of odd and even iterates of unregularized (data set M_0 given by (6.2)) exact alternating projections applied to the diffraction imaging problem. Only 500 iterations are shown, but the iterates past 500 behave similarly.

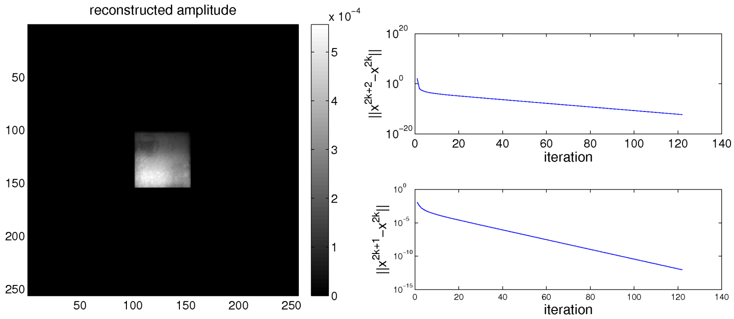

Figure 3: Reconstruction and behavior of odd and even iterates of regularized (data set M_ϵ given by (6.2) with $\epsilon = 1.9$) inexact alternating projections with λ_k chosen so that the iterates lie on the surface of the M_ϵ set.

Next we apply Algorithm 4.1 with the approximate projection computed as in Theorem 5.1 for different regularization parameters ϵ and different step-length strategies. Figure 3 shows the reconstruction and behavior of iterates for $\epsilon = 1.9$ and λ_k chosen so that the iterates remain on the surface of the M_ϵ set. Figure 4 shows the reconstruction and behavior of iterates for $\epsilon = 1.2$ and $\lambda_k = 1$ for all k. The reconstruction of the true object is only unique up to rotations, shifts and reflections. This is apparent with the "upside down" reconstruction of Figure 4. Figure 5 shows the *apparent* convergence rates for different values of the relaxation parameter ϵ in (6.2) and different settings for the step-length parameters λ_k. The black line shows again the change between the even iterates of the unregularized, exact alternating projection algorithm. The blue and green lines show the apparent rate of convergence of the regularized problems without extrapolation, that is, λ_k is computed so that the iterates lie on the surface of the set M_ϵ (to numerical precision). As expected, the lower the value of ϵ, the poorer the (asymptotic) rate of convergence since the sets are closer to ill-posedness for smaller regularization values. The red line shows what can be

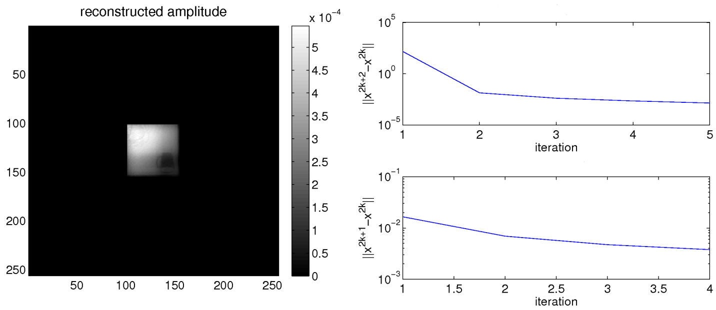

Figure 4: Reconstruction and behavior of odd and even iterates of regularized (data set M_ϵ given by (6.2) with $\epsilon = 1.2$) inexact extrapolated alternating projections with $\lambda_k = 1$ for all k. The algorithm terminates at the 5th iterate which achieves condition (4.1b) to numerical precision.

gained by extrapolation. Here the step-length parameter $\lambda_k = 1$ for all k and the algorithm proceeds with a convergence rate indicated by Theorem 4.4, but then terminates finitely as it finds a point on the interior of the intersection. Note that the only difference between this implementation and the unregularized exact alternating projections implementation (the black line) is early termination of the algorithm. This is what is usually done heuristically in practice. What this example shows is a mathematically sound explanation of this practice in terms of regularization, extrapolation and approximate alternating projections.

For this example it is not possible to compute an a priori rate of convergence as specified by Theorem 4.4 since the set M_ϵ has no analytic form and we are unable to compute the angle of the intersection. We observe a linear convergence rate, at least to the limit of machine precision. To a certain extent, this is beside the point. The value of the theory outlined above lies not with the computation of rates of convergence, but rather with the provision of regularization strategies and corresponding stopping rules. We can, however, verify numerically whether a point lies in the interior of the intersection of the regularized set with the qualitative constraint set. For the extrapolated example shown in Figure 4 it was verified that this point lies in the interior of the intersection by perturbing the point slightly and verifying that the perturbed point is still a numerical fixed point. Thus, even if the unregularized problem is not consistent as required by Theorem 5.1 to *guarantee* that the approximate projection achieves a sufficient accuracy for linear convergence, since the fixed point of the algorithm is an interior point, the required accuracy for the approximate projection is quite easy to satisfy as discussed in Remark 5.7.

Finally, note that while the rate of convergence for the more regularized problems is better, as indicated by comparing the reconstructions in Figures 3 and 4, the reconstruction can be poorer since this reconstruction is apparently further away from the ideal solution than the less regularized reconstructions.

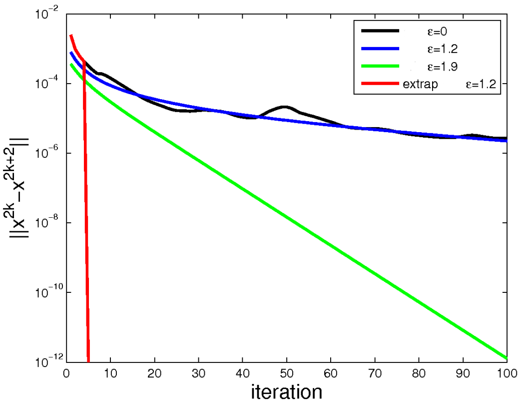

Figure 5: Comparison of implementations of Algorithm 4.1 with the approximate projection computed as in Theorem 5.1 for different parameters ϵ and step-length strategies (λ_k) for the fattened set M_ϵ given by (6.2). The black line is the unregularized alternating projection algorithm with exact projections. The blue and green lines are the regularized approximate alternating projection algorithms with step lengths λ_k computed so that the iterates lie on the surface of the M_ϵ set. The red line is the extrapolated approximate alternating projection algorithm with $\lambda_k = 1$ for all k.

7 Conclusion

The main achievement of this note is not our algorithm. Indeed, the regularized extrapolated ($\lambda_k = 1$ for all k) inexact projection algorithm specified in Theorem 5.1 in fact has been used successfully for decades in diffraction imaging with heuristic stopping criteria and *early termination* effectively serving as the regularization. What the analysis here provides, for the first time, is a regularization strategy that fits naturally with many ill-posed inverse problems, and a mathematically sound stopping criterion. The conventional early termination used in practice can be fully explained in the framework of this regularization strategy. While all of the regularity assumptions on the sets M_ϵ and C are satisfied for the finite dimensional phase problem discussed in Section 6, since the unregularized phase problem with noise is still inconsistent, Theorem 5.1 does not apply. If exact projections onto the regularized sets were computed, then Theorem 5.16 of [11] would suffice to prove convergence of exact alternating projections applied to the regularized problem. Proof of convergence of the inexact algorithm with extrapolation strategy $\lambda_k = 1$ for all k for the regularized phase retrieval problem (what has in fact been applied in the application literature for decades) hinges on verifying that condition 4.1c of Algorithm 4.1 is satisfied locally for all iterates. This is an open problem.

Acknowledgements

We thank Katharina Echternkamp, Ann-Kathrin Günther, Daja Herzog, Dong-Du Mai, Jelena Panke, Aike Ruhlandt and Jan Thiart at the Institut für

Röntgenphysik at the University of Göttingen for the diffraction data used in our numerical experiments.

References

[1] F.J. Aragón Artacho, A.L. Dontchev, and M.H. Geoffroy. Convergence of the proximal point method for metrically regular mappings. *ESAIM: COCV*, 2005.

[2] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm and Fienup variants: a view from convex feasibility. *J. Opt. Soc. Amer. A.*, 19(7):1334–45, 2002.

[3] H. H. Bauschke, P. L. Combettes, and D. R. Luke. A hybrid projection reflection method for phase retrieval. *J. Opt. Soc. Amer. A.*, 20(6):1025–34, 2003.

[4] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. *J. Approx. Theory*, 127:178–92, 2004.

[5] J. V. Burke and D. R. Luke. Variational analysis applied to the problem of optical phase retrieval. *SIAM J. Control Optim.*, 42(2):576–595, 2003.

[6] W. Cheney and A. A. Goldstein. Proximity maps for convex sets. *Proc. Amer. Math. Soc.*, 10(3):448–450, 1959.

[7] F.H. Clarke, R.J. Stern, and P.R. Wolenski. Proximal smoothness and the lower-C^2 property. *Journal of Convex Analysis*, 2:117–144, 1995.

[8] P. L. Combettes. Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. *IEEE Trans. Image Process.*, 6:493–506, 1997.

[9] A.N. Iusem, T. Pennanen, and B.F. Svaiter. Inexact versions of the proximal point algorithm without monotonicity. *SIAM J. Optim.*, 13:1080–1097, 2003.

[10] A.Y. Kruger. About regularity of collections of sets. *Set-Valued Analysis*, 14:187–206, 2006.

[11] A. S. Lewis, D. R. Luke, and J. Malick. Local linear convergence of alternating and averaged projections. *Found. Comput. Math.*, 9(4):485–513, 2009.

[12] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16:964–979, 1979.

[13] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21:37–50, 2005.

[14] D. R. Luke. Finding best approximation pairs relative to a convex and a prox-regular set in Hilbert space. *SIAM J. Optim.*, 19(2):714–739, 2008.

[15] S. M. Marchesini. A unified evaluation of iterative projection algorithms for phase retrieval. *Rev. Sci, Inst.*, 78:011301, 2007.

[16] B. S. Mordukhovich. Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings. *Dokl. Akad. Nauk BSSR*, 28:976–979, 1984. Russian.

[17] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 2006.

[18] T. Pennanen. Local convergence of the proximal point algorithm and multiplier methods without monotonicity. *Mathematics of Operations Research*, 27:170–191, 2002.

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

[20] R. T. Rockafellar and R. J. Wets. *Variational Analysis.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, 1998.

[21] J. E. Spingarn. A primal-dual projection method for solving systems of linear inequalities. *Linear Algebra Appl.*, 65:45–62, 1985.

[22] J. von Neumann. On rings of operators, reduction theory. *Ann. Math.*, 50:401–485, 1949.