FINDING BEST APPROXIMATION PAIRS RELATIVE TO A CONVEX AND A PROX-REGULAR SET IN A HILBERT SPACE

D. RUSSELL LUKE*

Abstract. We study the convergence of an iterative projection/reflection algorithm originally proposed for solving what are known as phase retrieval problems in optics. There are two features that frustrate any analysis of iterative methods for solving the phase retrieval problem: nonconvexity and infeasibility. The algorithm that we developed, called Relaxed Averaged Alternating Reflections (RAAR), was designed primarily to address infeasibility, though our strategy has advantages for nonconvex problems as well. In the present work we investigate the asymptotic behavior of the RAAR algorithm for the general problem of finding points that achieve the minimum distance between two closed convex sets in a Hilbert space with empty intersection, and for the problem of finding points that achieve a local minimum distance between one closed convex set and a closed *prox-regular set*, also possibly nonintersecting. The nonconvex theory includes and expands prior results limited to convex sets with nonempty intersection. To place the RAAR algorithm in context, we develop parallel statements about the standard alternating projections algorithm and gradient descent. All the various algorithms are unified as instances of iterated averaged alternating proximal reflectors applied to a sum of regularized maximal monotone mappings.

Keywords: Best approximation pair, convex set, prox-regular, inconsistent feasibility problems, projection, relaxed averaged alternating reflections, fixed point, resolvent, maximal monotone mappings.

AMS subject classifications. 90C26, 49M27, 49M20, 49J53, 65K05

PII. ABC-123

1. Introduction. Projection algorithms are simple yet powerful iterative techniques for finding the intersections of sets. Perhaps the most prevalent example of a projection algorithm is the alternating Projections Onto Convex Sets (POCS) dating back to von Neumann [54]. This and algorithms like it have been applied in image processing [19], medical imaging and economics [14], and optimal control [27] to name a few. For a review and historical background see [4]. The theory for these algorithms is limited mainly to convex setting and to *consistent feasibility problems*, that is problems where the set intersection is nonempty; if the intersection is empty the problem is referred to as an *inconsistent* feasibility problem. Examples abound of practitioners using these methods for nonconvex and/or inconsistent problems. We have in recent years been particularly interested in projection algorithms in crystallography and astronomy [6, 39], and more recently in inverse scattering [34, 33, 15, 12, 11]. Until now, we have been forced to rely on convex heuristics to justify certain strategies [7, 38] for want of an adequate nonconvex theory. In the absence of a nonconvex theory, practitioners resort to ad hoc stopping criteria and other strategies to get their algorithms to work according to user defined criteria. Depending on the algorithms, iterates tend to either stagnate at an undesirable point, or "blow up". We are particularly interested in those algorithms that appear to be unstable. Using convex heuristics we were able to provide plausible explanations [8] and remedies [38] for algorithmic instabilities, however a general theory was not pursued.

Our goal in this paper is two-fold: first to prove the convergence in the convex setting of an algorithm that we have proposed to solve inconsistent feasibility problems [38], and second to modify the theory to accommodate nonconvexity. Our algorithm, called relaxed averaged alternating reflections (RAAR), can be viewed as a relaxation of a fixed point mapping used by Lions and Mercier to solve generalized equations

* Department of Mathematical Sciences, University of Delaware, Newark, DE 19716 (rluke@math.udel.edu). The author's work was supported by NSF grant DMS-0712796.

involving the sum of maximal monotone mappings [37] and which is an extension of an implicit iterative algorithm by Douglas and Rachford [24] for solving linear partial differential equations.

In Section 2 we analyze the RAAR algorithm in the convex setting. Our task here is to characterize the fixed point set of the mapping, as well as to verify the assumptions of classical theorems. Our main result for this section is Theorem 2.7 which establishes convergence of the RAAR algorithm with approximate evaluation of the fixed point operator and variable relaxation parameter. The novelty of our mapping is that it addresses the crucial instance of inconsistent feasibility problems. Inconsistency is a source of instability for more conventional strategies. To place our new algorithm in the context of better-known strategies, we show in Proposition 2.5 that RAAR, alternating projections and gradient descent are all instances of iterated alternating averaged proximal reflectors – the Lions-Mercier algorithm – applied to the problem of minimizing the sum of two regularized maximal monotone mappings.

In Section 3 we expand our convex results to accommodate nonconvexity. In addition to characterizing the fixed point set, we formulate local, nonconvex versions of convex theorems, in particular formulations where *prox-regularity* is central. Our main result in this section is Theorem 3.12 which establishes local convergence of nonconvex applications of the RAAR slgorithm. While the nonconvex theory includes the convex case, we present both separately to highlight the places where nonconvexity requires extra care, and to make the nonconvex theory more transparent. The nature of nonconvexity requires focused attention to specific mappings, however we generalize wherever possible. Failing that, we detail parallel statements about the more conventional alternating projection algorithm; this also allows comparison of our algorithm with gradient descent methods for solving nonlinear least squares problems.

Our analysis complements other results on the convergence of projection algorithms for consistent nonconvex problems [22, 36, 35]. In particular we point out that the key assumption that we rely upon for convergence, namely a type of local nonexpansiveness of the fixed point mapping, does not appear to yield *rates* of convergence as are achieved in [36, 35] using notions of *regularity* of the intersection. On the other hand, regularity, in addition to assuming the intersection is nonempty, is a strong condition on the intersection that, in particular, is not satisfied for ill-posed inverse problems, our principal motivation.

To close this subsection we would like to clarify the relationship of the present work to previous work on the phase retrieval problem in crystallography and astronomy that has been a major motivation for these investigations. The results developed in this work apply in principle to the finite-dimensional phase retrieval problem (that is, discrete bandlimited images), so long as certain regularity of the fixed point mapping, namely local firm nonexpansiveness, can be determined. Such an investigation is beyond the scope of this work. The infinite dimensional phase retrieval problem, on the other hand, as studied in [13] does *not* fall within the theory developed here because the sets generated by the magnitude constraints are not weakly closed [39, Property 4.1], hence not prox-regular.

1.1. Basic Tools and Results. We begin with the central tools and basic results that we will use in our analysis. Throughout this paper $\mathcal{H}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. For $A, B \subset \mathcal{H}$ closed, the underlying problem is to

$$\text{find } \; x \in A \cap B. \tag{1.1}$$

Note that it could happen that $A \cap B = \varnothing$, in which case one might naturally formulate the problem as a nonlinear least squares problem

$$\underset{x}{\text{minimize}}\, J(x) = \frac{1}{2}\left(\tfrac{1}{2}\operatorname{dist}^2_A(x) + \tfrac{1}{2}\operatorname{dist}^2_B(x)\right) \tag{1.2}$$

where dist $_C(x)$ is the distance of the point x to a set C:

$$\text{dist}_C(x) := \inf_{c\in C} |x-c|.$$

If x_* is a locally optimal point then $0 \in \partial J(x_*)$ where ∂ denotes the subdifferential [17, 51, 18, 30, 31, 32, 43]. Another characterization of a locally optimal point x_* is to associate it with a *best approximation* pair (a, b) satisfying $b \in P_B x_* \subset P_B a$ and $a \in P_A x_* \subset P_A b$ where P_C is the projection defined by:

$$P_C x := \text{argmin}_{c\in C} |x-c| = \{y \in C \subset \mathcal{H} \mid |x-y| = \text{dist}_C(x)\}.$$

If C is convex then the projection is single-valued. If in addition C is closed and nonempty, then $P_C x$ is characterized by [23, Theorem 4.1]

$$P_C x \in C \quad \text{and} \quad \langle c - P_C x, x - P_C x\rangle \le 0 \quad \text{for all } c \in C. \tag{1.3}$$

If C is not convex, then, the projection, if it exists, is a set. If the projection exists and is single-valued near all points in C, then C is said to be *prox-regular* [49]. The relationship between the subdifferential of the squared distance function to a prox-regular set C and the projection is shown in [49, Proposition 3.1] to be

$$\partial\left(\text{dist}^2_C(x)\right) = 2(x - P_C x) \tag{1.4}$$

for x in a neighborhood of C. If C is convex this relationship holds globally. In particular, for A and B prox-regular and x in a proximal neighborhood of both sets, we have

$$\partial J(u) = \tfrac{1}{2}\left((x - P_A x) + (x - P_B x)\right).$$

Example 1.1 (gradient descent and averaged projections). *Consider the steepest descent iteration with step length λ_n where P_A and P_B are single-valued:*

$$\begin{aligned} x_{n+1} &= x_n - \lambda_n \partial J(x_n) \\ &= (1-\lambda_n)x_n + \lambda_n \tfrac{1}{2}\left(P_A x_n + P_B x_n\right). \end{aligned} \tag{1.5}$$

In other words, gradient descent for least squares minimization is a relaxed averaged projection algorithm. *We will come back to this particular algorithm below.* □

Projection algorithms seek to find an element in $A \cap B$, or best approximation thereof, by iterated projections, possibly with some relaxation strategy, onto A and B separately. The example above interprets the steepest descent algorithm as a relaxed averaged projection algorithm. Another elementary projection algorithm is the well-known alternating projections algorithm: Given $x_0 \in \mathcal{H}$ generate the sequence $\{x_n\}_{n\in\mathbb{N}}$ by

$$x_n = (P_A P_B)\, x_{n-1}. \tag{1.6}$$

Example 1.2 (averaged projections and alternating projections). *A standard formulation in the product space [46] identifies the averaged projections (and hence steepest descent) given by (1.5) with alternating projections. To see this, consider the product space $\mathcal{H} \times \mathcal{H}$ with inner product $\langle (x_1, x_2), (y_1, y_2)\rangle := \frac{1}{2}\left((x_1, y_1) + (x_2, y_2)\right)$. Let $C = \{(x, y) \in \mathcal{H} \times \mathcal{H} \mid x = y\}$ and $D = \{(x, y) \in \mathcal{H} \times \mathcal{H} \mid x \in A, y \in B\}$, then*

$$P_C P_D(x, x) = \left(\tfrac{1}{2}(P_A + P_B)x, \tfrac{1}{2}(P_A + P_B)x\right).$$

□

Our focus in this paper is on the convergence of projection algorithms, but the above example serves to emphasize that convergence results about such algorithms can be very broadly applied.

When $A \cap B = \varnothing$ we say that the feasibility problem (1.1) is inconsistent. The distinction between inconsistent and consistent feasibility problems has profound implications for the stability and convergence of numerical algorithms. It is convenient to define *difference set* $B - A$. The *gap* between the sets A and B is the point in $B - A$ closest to the origin. Specifically,

$$G := \overline{B - A}, \qquad g := P_G 0, \qquad E := A \cap (B - g), \quad \text{and} \quad F := (A + g) \cap B, \tag{1.7}$$

where $\overline{B - A}$ denotes the closure of $B - A$. Note that these definitions only make sense when A and B are convex. We will generalize these sets in Section 3. Basic characterizations are given in [3, 8]. We note that if $A \cap B \neq \varnothing$, then $E = F = A \cap B$. Even in the case where $A \cap B = \varnothing$, the gap vector g is unique and always well-defined. A useful characterization of the gap vector g is via the *normal cone mapping of* G: for G convex $-g \in N_G(g)$ where $N_G(g)$ is the defined by

$$N_G \colon g \mapsto \begin{cases} \{y \in \mathcal{H} \mid \langle c - g, y\rangle \leq 0 \quad \text{for all } c \in G\}, & \text{if } g \in G; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{1.8}$$

Example 1.3 (projections and normal cone mappings for convex sets). *If C is convex, then the normal cone mapping is the subdifferential of the (infinite) indicator function, ι_C, of the set C:*

$$\iota_C(x) := \begin{cases} 0 & \textit{for } x \in C \\ \infty & \textit{else} \end{cases} \tag{1.9}$$

$$\partial \iota_C(x) = N_C(x) \tag{1.10}$$

$$(I + \rho N_C)^{-1} x = P_C x \quad \textit{for all } \rho > 0. \tag{1.11}$$

The mapping $(I + N_C)^{-1}$ is called the resolvent *of the normal cone mapping, or equivalently in this case, the resolvent of $\partial \iota_C$. Specializing to $C = A \cap B$, the indicator function of the intersection is the sum of the indicator functions of the individual sets,*

$$\iota_{A+B} = \iota_A + \iota_B$$

and, for A and B convex, the resolvent $(I + (N_A + N_B))^{-1}$ is the projection onto $A \cap B$ supposing this is nonempty. In other words, an element of $A \cap B$ is a zero of $\partial \iota_{A \cap B}$. The parameter ρ in (1.11) is interpreted as a step size *consistent with backward-stepping descent algorithms (see [25]).* □

We compare throughout this work the asymptotic properties of alternating projections to more recent projection strategies. A common framework in the convex setting that provides an elegant synthesis of these algorithms is through operator splitting strategies for solving

$$\underset{x \in \mathcal{H}}{\text{minimize}} \; f_1(x) + f_2(x) \tag{1.12}$$

where f_1 and f_2 are proper, lower semi-continuous (l.s.c.) convex functions from $\mathcal{H}$ to $\mathbb{R} \cup \{\infty\}$. The subdifferentials ∂f_j are then *maximal monotone* [44, Proposition 12.b.], that is, $\operatorname{gph} \partial f_j$ cannot be enlarged in $\mathcal{H} \times \mathcal{H}$ without destroying monotonicity of ∂f_j defined by

$$\langle v_2 - v_1, x_2 - x_1 \rangle \geq 0 \quad \text{whenever} \quad v_1 \in \partial f_j(x_1),\ v_2 \in \partial f_j(x_2).$$

We then seek points that satisfy the inclusion for the sum of two maximal monotone mappings:

$$0 \in \partial f_1(x) + \partial f_2(x). \tag{1.13}$$

Iterative techniques for solving (1.13) are built on combinations of forward- and backward-stepping mappings of the form $(I - \lambda\partial f_j)$ and $(I + \lambda\partial f_j)^{-1}$ respectively. For proper, l.s.c. convex functions f_j Moreau [44] showed the correspondence between the resolvent $(I + \lambda\partial f_j)^{-1}$ and the argmin of the regularized mapping f_j centered on x. In particular, define the Moreau envelope, $e_{\lambda f}$, and the proximal mapping, $\operatorname{prox}_{\lambda,f}$, of a function f by

$$e_{\lambda,f}x := \inf_w \left\{ f(w) + \frac{1}{2\lambda}|w - x|^2 \right\} \quad \text{and} \quad \operatorname{prox}_{\lambda,f} x := \operatorname{argmin}_w \left\{ f(w) + \frac{1}{2\lambda}|w - x|^2 \right\}.$$

Then by [44, Proposition 6.a] we have

$$\operatorname{prox}_{1,f_j} x = J_{\partial f_j} x := (x + \partial f_j(x))^{-1} \tag{1.14}$$

where $J_{\partial f_j}$ is the resolvent of ∂f_j. The Moreau envelope at zero, $e_{\lambda,f}0$, is perhaps better known as Tikhonov regularization [53, 52].

Maximal monotonicity of ∂f_j is equivalent to *firm nonexpansiveness* of the resolvent $J_{\partial f_j}$ whose domain is all of $\mathcal{H}$ [42]. A mapping $T : \operatorname{dom} T = \mathbb{X} \to \mathbb{X}$ is *nonexpansive* on the closed convex subset $\mathbb{X} \subset \mathcal{H}$ if

$$|Tx - Ty| \le |x - y| \quad \text{for all} \quad x, y \in \mathbb{X}; \tag{1.15}$$

we say that T is firmly nonexpansive on $\mathbb{X}$ when

$$|Tx - Ty|^2 \le \langle x - y, Tx - Ty \rangle \quad \text{for all} \quad x, y \in \mathbb{X}. \tag{1.16}$$

Firmly nonexpansive mappings also satisfy the following convenient relation:

$$|Tx - Ty|^2 + |(I - T)x - (I - T)y|^2 \le |x - y|^2 \quad \text{for all} \quad x, y \in \mathbb{X}. \tag{1.17}$$

For more background see [28, Theorem 12.1] and [23, Theorem 5.5].

Example 1.4 (projections onto and reflections across convex sets). *Let C be a nonempty closed convex set in $\mathcal{H}$. The projection onto C is firmly nonexpansive on $\mathcal{H}$ [23, Theorem 5.5] and the corresponding reflection, defined by $R_C := 2P_C - I$ is nonexpansive.* □

The following central result upon which we build concerns the convergence of iterated nonexpansive mappings allowing for *approximate evaluation* of *dynamically relaxed* mappings with variable *step sizes*. Our formulation follows [20] which is a generalization of an analogous result in [25]. Both [25] and [20] synthesize previous work of Rockafellar [50], Gol'stein and Tret'yakov [29], and are also related to work of Martinet [40, 41] and Brezis and Lions [10] concerning resolvents of maximally monotone mappings. The theorem is formulated for a common relaxation of the fixed point mapping T. For any arbitrary nonexpansive mapping T, the standard relaxation of the iteration $x_{n+1} = Tx_n$ is to a *Krasnoselski-Mann* iteration [9] given by

$$x_{n+1} = U(T, \lambda_n)x_n := \lambda_n Tx_n + (1 - \lambda_n)x_n, \quad 0 < \lambda_n < 2. \tag{1.18}$$

By Example 1.1, gradient descent for the squared distance objective (1.2) with step length λ_n is equivalent to a Krasnoselski-Mann relaxation of the averaged projection mapping $T := \frac{1}{2}(P_A + P_B)$. In general, the Krasnoselski-Mann relaxation does not change the set of *fixed points* of T denoted $\mathsf{Fix}\, T$.

Lemma 1.5. *Let $T = (I + \rho S)^{-1}$ ($\rho > 0$) be firmly nonexpansive with* $\operatorname{dom} T = \mathcal{H}$. *Then* $\mathsf{Fix}\, T = \varnothing$ *if and only if there is no solution to* $0 \in Sx$.

Proof. T with $\operatorname{dom} T = \mathcal{H}$ is *firmly nonexpansive* if and only if it is the resolvent of a maximally monotone mapping $F : \mathcal{H} \to 2^{\mathcal{H}}$ [42]. Direct calculation then shows that $\mathsf{Fix}\, T = \varnothing$ is equivalent to $\{x \in \mathcal{H} \mid Fx = 0\} = \varnothing$. □

THEOREM 1.6 (inexact evaluation of firmly nonexpansive mappings). *Let T be a firmly nonexpansive mapping on $\mathcal{H}$ with* $\operatorname{dom} T = \mathcal{H}$. *Given any $x_0 \in \mathcal{H}$, let the sequence $\{x_n\}_{n\in\mathbb{N}}$ be generated by*

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n\left(Tx_n + \epsilon_n\right) \quad \forall n \geq 0 \tag{1.19}$$

where $\{\lambda_n\}_{n\in\mathbb{N}} \subset \left]0,2\right[$ and $\{\epsilon_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$ are sequences with

$$\sum_{n=0}^{\infty} |\epsilon_n| < \infty, \quad \lambda_- = \inf_{n\geq 0} \lambda_n > 0, \quad \lambda_+ = \sup_{n\geq 0} \lambda_n < 2. \tag{1.20}$$

Then if T possesses a fixed point, x_n converges weakly to a fixed point of T. Convergence is strong if any one of the following hold:

- $\underline{\lim}\, \operatorname{dist}_{\mathsf{Fix}\, T}(x_n) = 0$;
- $\operatorname{int} \mathsf{Fix}\, T \neq \varnothing$
- *T is* demicompact at 0*: that is, for every bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ with $Tx_n - x_n$ converging strongly to y, the set of strong cluster points of $\{x_n\}_{n\in\mathbb{N}}$ is nonempty.*

If T is firmly nonexpansive with $\operatorname{dom} T = \mathcal{H}$ *and* $\mathsf{Fix}\, T = \varnothing$, *then $\{x_n\}_{n\in\mathbb{N}}$ is unbounded.*

Proof. All but the last statement is the content of Theorem [20, Theorem 5.5]. To show that x_n is unbounded if T does not have a fixed point for T firmly nonexpansive with $\operatorname{dom} T = \mathcal{H}$, we note that by Lemma 1.5 $\mathsf{Fix}\, T = \varnothing$ if and only if there is no solution to $0 \in Fx$ where T is the resolvent of the maximally monotone mapping F. The result now follows from [25, Theorem 3]. □

For the remainder of this paper we will be concerned with applying the above results to particular instances of the mapping T for convex and nonconvex settings. Our principal task, therefore, is to characterize $\mathsf{Fix}\,(T)$ and to modify the above theory to accommodate nonconvexity. To account for realistic limitations in computing accuracy we consider fixed point iterations where T is only approximately evaluated. With this in mind, and in the context of (1.12), we compare the following approximate algorithms:

ALGORITHM 1.7 (approximate alternating proximal mappings). *Choose $x_0 \in \mathcal{H}$. For $n \in \mathbb{N}$ set*

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n\left(\operatorname{prox}_{1,f_1}\left(\operatorname{prox}_{1,f_2} x_n\right) + \epsilon_n\right). \tag{1.21}$$

ALGORITHM 1.8 (approximate averaged alternating proximal reflections). *Choose $x_0 \in \mathcal{H}$. For $n \in \mathbb{N}$ set*

$$x_{n+1} = (1-\lambda_n)x_n + \frac{\lambda_n}{2}\left(R_{f_1}\left(R_{f_2}x_n + \epsilon_n\right) + \rho_n + x_n\right). \tag{1.22}$$

where $R_{f_j}x := 2\operatorname{prox}_{1,f_j} x - x$.

The parameter λ_n is the Krasnoselski-Mann relaxation parameter, and the terms ρ_n and ϵ_n account for the error made in the calculation of each of the resolvents separately.

The exact version of Algorithm 1.8 was proposed by Lions and Mercier [37] who adapted the Douglas-Rachford [24] algorithm to solving $0 \in F+G$ for general maximal monotone mappings F and G. Convergence results for the application of this algorithm hinge on the following assumption:

ASSUMPTION 1.9. *There exist $x \in \mathcal{H}$, $a \in \partial f_1(x)$, and $b \in \partial f_2(x)$ such that $a + b = 0$.*

The key result of Lions and Mercier adapted to our setting is that, if Assumption 1.9 holds, then the sequence of iterates $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm 1.8 with $\epsilon_n = \rho_n = 0$ for all n converges weakly to $\overline{x} \in \mathcal{H}$ as $n \to \infty$ such that $x_* = J_{\partial f_2}\overline{x}$ solves (1.12) [37, Theorem 1].

Example 1.10 (specialization of (1.12) to convex feasibility). *Let $f_1 = \iota_A$ and $f_2 = \iota_B$ in (1.12) where A and B are convex. Then following Example 1.3 we have*

$$\operatorname{prox}_{1,f_1}\operatorname{prox}_{1,f_2} = P_A P_B \tag{1.23}$$

$$\tfrac{1}{2}(R_{f_1}R_{f_2} + I) = \tfrac{1}{2}(R_A R_B + I) \tag{1.24}$$

Specialization of Algorithm 1.7 to this setting yields the classical alternating projection algorithm. Convergence of the exact algorithm was obtained in [16, Theorem 4] under the assumption that either (a) one of A or B is compact, or (b) one of A or B is finite dimensional and the distance between the sets is attained. In other words, $A \cap B$ can be empty. Rates *of convergence, however, appear to require a certain regularity of the intersection [35].*

Specialization of Algorithm 1.8 yields the Averaged Alternating Reflection algorithm studied in [8]. It follows immediately from Example 1.4 that $\frac{1}{2}(R_A R_B + I)$ is firmly nonexpansive (see also [8, Proposition 3.1]). Assumption (1.9) reduces to $A \cap B \neq \varnothing$. If $A \cap B = \varnothing$ then by Theorem 1.6 we have $|\frac{1}{2}(R_A R_B + I)x_n| \to \infty$ as $n \to \infty$. Nevertheless, as long as there exist nearest points in B to A, then the sequences $\{P_B x_n\}_{n\in\mathbb{N}}$ and $\{P_A P_B x_n\}_{n\in\mathbb{N}}$ are bounded with weak cluster points belonging to the sets F and E defined by (1.7) [8, Theorem 3.13]. Indeed, regardless of whether or not $A \cap B = \varnothing$, the set $\mathsf{Fix}\,(T_{AAR} + g)$ is closed and convex and [8, Theorem 3.5]

$$F + N_G(g) \subset \mathsf{Fix}\,(T_{AAR} + g) \subset g + F + N_G(g). \tag{1.25}$$

In other words, if $A \cap B = \varnothing$ then T_{AAR} does not have fixed points, but rather has fixed directions *or* velocities. *Examples 3.7 and 3.8 of [8] show that the upper and lower bounds on this fixed point set are tight, consistent with the case $A \cap B \neq \varnothing$. The salient point here is that convergence of this algorithm is contingent on the consistency of the feasibility problem.* □

Generalizations of Lions and Mercier's results to approximate evaluation of the resolvents of maximally monotone mappings have been investigated in [25, 20, 21]. The following theorem, adapted from [21], is a specialization of Theorem 1.6 to Algorithms 1.7 and 1.8.

Corollary 1.11 (specialization to Algorithms 1.7 and 1.8). *Let f_1 and f_2 be proper, l.s.c. convex functions from $\mathcal{H}$ to $\mathbb{R} \cup \{\infty\}$, let $\{\rho_n\}_{n\in\mathbb{N}}$ and $\{\epsilon_n\}_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$, and let $\{\lambda_n\}_{n\in\mathbb{N}}$ be a sequence in $]0,2[$.*

(i) *Let $E := \mathsf{Fix}\operatorname{prox}_{1,f_1}\operatorname{prox}_{1,f_2} \neq \varnothing$ and $\{\epsilon_n\}_{n\in\mathbb{N}}$ and $\{\lambda_n\}_{n\in\mathbb{N}}$ satisfy (1.20). Then every sequence $\{x_n\}_{n\in\mathbb{N}}$ of Algorithm 1.7 converges weakly to a point in E. If* $\operatorname{int} E \neq \varnothing$ *then convergence is strong.*

(ii) *If Assumption 1.9 holds and $\{\lambda_n\}_{n\in\mathbb{N}} \subset]0,2[$ with $\sum_{n\in\mathbb{N}}\lambda_n(2-\lambda_n) = \infty$ and $\sum_{n\in\mathbb{N}}\lambda_n\left(\|\rho_n\| + \|\epsilon_n\|\right) < \infty$, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm 1.8 converges weakly to $\overline{x} \in \mathcal{H}$ as $n \to \infty$ such that $x_* = J_{\partial f_2}\overline{x}$ solves (1.12). If Assumption 1.9 does not hold, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm 1.8 is unbounded.*

Proof. (i) is an immediate specialization of Theorem 1.6. For (ii), all but the last statement is [21, Corollary 5.2] with $\gamma = 1$. The last statement of (ii) follows from Theorem 1.6 since $\frac{1}{2}(R_{f_1}R_{f_2} + I)$ is firmly nonexpansive. □

2. Convex Analysis. For this section we will assume that the sets A and B are closed and convex. Denote

$$T_{AP} := P_A P_B \quad \text{and} \quad T_{AAR} := \tfrac{1}{2}(R_A R_B + I) \tag{2.1}$$

discussed in Example 1.10. As discussed in Example 1.10 the existence of fixed points of T_{AP} is independent of whether or not the feasibility problem is consistent. Indeed, it is easy to see that $\mathsf{Fix}\, T_{AP} = E$ for E defined by (1.7). This is not the case for T_{AAR}. We will argue in the nonconvex setting that the fact that T_{AAR} has no fixed points if $A \cap B = \varnothing$ has tremendous algorithmic potential since it means that averaged alternating reflections will not get stuck in a local minimum. Other algorithms for solving feasibility problems do not suffer from such instabilities with inconsistent problems (alternating projections for instance), but for nonconvex problems, this stability is at the cost of getting caught in local minima. It is this resilience of the AAR algorithm in nonconvex applications that first attracted our attention and, we believe, warrants a closer look. In the next section we compare the behavior of these algorithms in the convex setting.

2.1. Relaxations/Regularizations. In this subsection we consider relaxations of T_{AAR} whose associated mappings have fixed points independent of whether or not $A \cap B = \varnothing$. The common relaxation that we have already discussed is of the form

$$U(T, \lambda) := \lambda T + (1 - \lambda) I, \quad 0 < \lambda < 2 \tag{2.2}$$

for the generic mapping T. If the mapping T is firmly nonexpansive (for instance T_{AP} or T_{AAR}) then this property is preserved under the relaxation $U(T, \beta)$ for $\beta \in]0, 1[$. Krasnoselski-Mann iterations have been extensively studied in Hilbert spaces and more general normed spaces [9] so there is ample theory to draw from for the study of the relaxation $U(T, \beta)$.

An advantage and disadvantage of this relaxation is that the fixed points of $U(T, \lambda)$ are the same as those of T. In particular, since T_{AAR} has a fixed point if and only if $A \cap B \neq \varnothing$, it follows immediately that the same holds for $U(T_{AAR}, \lambda)$: for inconsistent problems neither mapping has a fixed point. To remedy this we consider the following alternative relaxation:

$$V(T, \beta) := \beta T + (1 - \beta) P_B, \quad 0 < \beta < 1. \tag{2.3}$$

Like the Krasnoselski-Mann relaxation, for A and B convex and T firmly nonexpansive, then $V(T, \beta)$ is also firmly nonexpansive since it is the convex combination of firmly nonexpansive mappings. Hence if $\mathsf{Fix}\, V(T_{AAR}, \beta)$ is nonempty, then the associated approximate fixed point iteration converges to the fixed point set according to Theorem 1.6. One of the principal advantages of this relaxation is that, as we show in Lemma 2.1, $\mathsf{Fix}\, V(T_{AAR}, \beta)$ is independent of whether or not the associated problem (1.1) is feasible. Moreover, the relaxation parameter β can be used to exert some control on the iterates (see Subsection 2.2).

In characterizing the fixed points we note that the relaxation $V(T_{AAR}, \beta)$ is fundamentally different than the standard relaxation $U(T_{AAR}, \lambda)$ which has little qualitative effect on the set of fixed points. The two are independent and may be used together without any redundancy of effect. There can, however, be diminishing returns to the addition of parameters to algorithms of this sort. For our application we have found no significant advantage to employing relaxations of the form (2.2). Nevertheless, by Example 1.1, for cases where the relaxation is related to a step length in a gradient descent algorithm, then optimization of λ_n in (2.2) can clearly lead to improved performance. We therefore retain this relaxation and, for the sake of generality, consider nested relaxations of the form

$$U\left(V\left(U(T_{AAR}, \lambda_1), \beta\right), \lambda_2\right) = \lambda_2 V\left(U(T_{AAR}, \lambda_1), \beta\right) + (1 - \lambda_2) I, \quad \text{where } \lambda_2 \in]0, 2[\,. \tag{2.4}$$

The next theorem is a generalization of [38, Theorem 2.2] where we determined the fixed points of the mapping $V(T_{AAR}, \beta)$ alone. The following analysis of the nested relaxations demonstrates the relative

importance of the relaxation strategies. This is discussed in greater detail following the proof of the next observation.

LEMMA 2.1 (characterization of fixed points). *Let $\beta \in \,]0,1[$ and $\lambda_1, \lambda_2 \in \,]0,2[$. Then*

$$\mathsf{Fix}\, U\left(V\left(U(T_{AAR},\lambda_1),\beta\right),\lambda_2\right) = F - \frac{\beta\lambda_1}{1-\beta}g, \tag{2.5a}$$

where F and g are defined by (1.7). Moreover, $\mathsf{Fix}\, U\left(V\left(U(T_{AAR},\lambda_1),\beta\right),\lambda_2\right)$, is closed and convex and, for every $x \in \mathsf{Fix}\, U\left(V\left(U(T_{AAR},\lambda_1),\beta\right),\lambda_2\right)$, we have the following:

$$x = P_B x - \frac{\beta\lambda_1}{1-\beta}g; \tag{2.5b}$$

$$P_B x - P_A R_B x = g; \tag{2.5c}$$

$$P_B x \in F \quad \text{and} \quad P_A P_B x \in E. \tag{2.5d}$$

In the special case where $\beta = 1$, we have

$$F + N_G(g) \subset \mathsf{Fix}\left(U(T_{AAR},\lambda) + \lambda g\right) \subset g + F + N_G(g). \tag{2.5e}$$

By comparison,

$$\mathsf{Fix}\, U(V(U(T_{AP},\lambda_1),\beta),\lambda_2) = F - \beta g. \tag{2.6}$$

Proof. For all $\beta \in [0,1[$, since $\mathsf{Fix}\,(\lambda T + (1-\lambda)I) = \mathsf{Fix}\, T$ for *any* mapping T , the fixed point set is invariant with respect to the outer relaxation $(\lambda_2 V(U(T,\lambda_1)) + (1-\lambda_2)I)$, so without loss of generality we ignore this relaxation.

Equation (2.6) follows immediately from $\mathsf{Fix}\, T_{AP} = E$.

What remains, then, is to show (a) that $F - \frac{\beta\lambda_1}{(1-\beta)}g \subset \mathsf{Fix}\, V\left(U(T_{AAR},\lambda_1),\beta\right)$ and, conversely, (b) that $\mathsf{Fix}\, V\left(U(T_{AAR},\lambda_1),\beta\right) \subset F - \frac{\beta\lambda_1}{(1-\beta)}g$.

We first establish the inclusion $F - \frac{\beta\lambda_1}{1-\beta}g \subset \mathsf{Fix}\, V\left(U(T_{AAR},\lambda_1),\beta\right)$. Pick $f \in F$ and let $x = f - \frac{\beta\lambda_1}{1-\beta}g$ and define $e := f - g$. Now, since $f \in F$ and $g \in P_G 0$, then $e \in E$ and $-\gamma g \in N_B(f)$ and $\gamma g \in N_A(e)$ for all $\gamma > 0$. Hence $P_B x = f$ and $P_A(e + \gamma g) = e$, thus

$$R_B x = 2P_B x - x = f + \frac{\beta\lambda_1}{1-\beta}g,$$

and

$$P_A R_B x = P_A\left(f + \frac{\beta\lambda_1}{1-\beta}g\right) = P_A\left(e + \frac{1+\beta(\lambda_1-1)}{1-\beta}g\right) = e = f - g.$$

Hence $P_B x - P_A R_B x = g$. This together with the observation that

$$x - T_{AAR}x = P_B x - P_A R_B x \quad \text{for all } x \in \mathcal{H} \tag{2.7}$$

implies $x - \beta U(T_{AAR},\lambda_1)x - (1-\beta)P_B x = \beta\lambda_1(x - T_{AAR}x) + (1-\beta)(x - P_B x) = \beta\lambda_1 g + (1-\beta)(x-f) = 0$. Thus, as claimed, $F - \frac{\beta\lambda_1}{1-\beta}g \subset \mathsf{Fix}\left(\beta U(T_{AAR},\lambda_1) + (1-\beta)P_B\right)$.

We show next that $\mathsf{Fix}\,(\beta U(T_{AAR},\lambda_1)+(1-\beta)P_B)\subset F-\frac{\beta\lambda_1}{1-\beta}g$. To see this, pick any $x\in\mathsf{Fix}\,(\beta U(T_{AAR},\lambda_1)+(1-\beta)P_B)$. Let $f=P_Bx$ and $y=x-f$. Recall that

$$P_A(2f-x)=P_A(2P_Bx-x)=P_AR_Bx. \tag{2.8}$$

This, together with the identity (2.7) yields

$$P_A(2f-x)=f+T_{AAR}x-x. \tag{2.9}$$

For our choice of x we have $\beta U(T_{AAR},\lambda_1)x+(1-\beta)P_Bx=\beta\lambda_1T_{AAR}x+\beta x-\beta\lambda_1x+(1-\beta)P_Bx=x$, which yields

$$T_{AAR}x-x=\frac{1-\beta}{\beta\lambda_1}(x-P_Bx). \tag{2.10}$$

Then (2.9) and (2.10) give

$$P_A(2f-x)=f+\frac{1-\beta}{\beta\lambda_1}(x-f)=f+\frac{1-\beta}{\beta\lambda_1}y. \tag{2.11}$$

Now, for any $a\in A$, since A is nonempty, closed and convex, we have

$$\langle a-P_A(2f-u),(2f-u)-P_A(2f-u)\rangle\le 0, \tag{2.12}$$

and hence

$$\begin{aligned}0&\ge\left\langle a-\left(f+\frac{1-\beta}{\beta\lambda_1}y\right),\ (2f-x)-\left(f+\frac{1-\beta}{\beta\lambda_1}y\right)\right\rangle\\&=\left\langle a-\left(f+\frac{1-\beta}{\beta\lambda_1}y\right),\ -y-\frac{1-\beta}{\beta\lambda_1}y\right\rangle\\&=\frac{\beta(\lambda_1-1)+1}{\beta\lambda_1}\langle -a+f,\ y\rangle+\frac{(1-\beta)(\beta(\lambda_1-1)+1)}{(\beta\lambda_1)^2}|y|^2.\end{aligned} \tag{2.13}$$

Here we have used (2.12), (2.11) and the fact that $y=x-f$. On the other hand, for any $b\in B$, since B is a nonempty closed convex set and $f=P_Bx$, we have

$$\langle b-P_Bx,x-f\rangle\le 0, \tag{2.14}$$

which yields

$$\langle b-f,\ y\rangle=\langle b-f,\ x-f\rangle\le 0. \tag{2.15}$$

Note that for $\beta\in\,]0,1[$ and $\lambda_1\in\,]0,2[$ the numerator $\beta(\lambda_1-1)+1>0$, thus (2.13) and (2.15) yield

$$\langle b-a,\ y\rangle\le-\frac{1-\beta}{\beta\lambda_1}|y|^2\le 0. \tag{2.16}$$

Now take a sequence $\{a_n\}_{n\in\mathbb{N}}$ in A and a sequence $\{b_n\}_{n\in\mathbb{N}}$ in B such that $g_n=b_n-a_n\to g$. Then

$$\langle g_n,\ y\rangle\le-\frac{1-\beta}{\beta\lambda_1}|y|^2\le 0\quad\text{for all } n\in\mathbb{N}. \tag{2.17}$$

Taking the limit and using the Cauchy-Schwarz inequality yields

$$|y|\le\frac{\beta\lambda_1}{1-\beta}|g|. \tag{2.18}$$

Conversely, $x - (\beta U(T_{AAR}, \lambda_1)x + (1-\beta)P_B x) = \beta\lambda_1 (f - P_A(2f - x)) + (1-\beta)y = 0$ gives

$$|y| = \frac{\beta\lambda_1}{1-\beta}\Big|f - P_A(2f - x)\Big| \geq \frac{\beta\lambda_1}{1-\beta}|g|. \tag{2.19}$$

Hence $|y| = \frac{\beta\lambda_1}{1-\beta}|g|$ and, taking the limit in (2.17), $y = -\frac{\beta\lambda_1}{1-\beta}g$, which confirms the identity (2.5b). From (2.8) and (2.11) with $y = -\frac{\beta\lambda_1}{1-\beta}g$ it follows that $f - P_A R_B x = g$ which proves (2.5c) and, by definition, implies that $P_B x = f \in F$ and $P_A P_B x \in E$. This yields identity (2.5d) and proves (2.5a). The closedness and convexity of the fixed point set then follows from the fact that F is closed and convex. (More generally, the fixed point set of any nonexpansive map defined everywhere in a Hilbert space is closed convex; see [28, Lemma 3.4]).

For the special case where $\beta = 1$, a straightforward calculation shows that $\mathsf{Fix}\,(T_{AAR} + g) = \mathsf{Fix}\,(U(T_{AAR}, \lambda) + \lambda g)$. Since, by [8, Theorem 3.5], we have

$$F + N_G(g) \subset \mathsf{Fix}\,(T_{AAR} + g) \subset g + F + N_G(g), \tag{2.20}$$

the result follows immediately, which completes the proof. □

Remark 2.2. *Lemma 2.1 shows that the inner relaxation parameter λ_1 has only a marginal effect on the set of fixed points of T_{AAR} compared to the β relaxation, which, provided $g \neq 0$, is unbounded as $\beta \to 1$; it has no effect on the set of fixed points of T_{AP}. The outer relaxation parameter λ_2 has no effect on either mapping. In stark contrast to these, the relaxation parameter β in the relaxation $V(T, \beta)$ has a profound impact on the set of fixed points of T_{AAR} and marginal impact on the fixed points of T_{AP}. Indeed, from (2.20) and (2.5a) it is clear that, for all $0 < \beta < 1$, $\mathsf{Fix}\, V(T_{AAR}, \beta) \subset \mathsf{Fix}\,(T_{AAR} + g)$, thus, by definition, $x_\beta - T_{AAR}x_\beta = g$ where $x_\beta \in \mathsf{Fix}\, V(T_{AAR}, \beta)$. More interestingly, however, the fixed point set becomes vastly larger at $\beta = 1$. Similarly, at $\beta = 0$ the fixed point set becomes all of B.*

Having characterized the fixed points of $V(T_{AAR}, \beta)$ we turn our attention to *inexact Relaxed Averaged Alternating Reflections (RAAR)* iterations:

Algorithm 2.3 (inexact RAAR algorithm). *Choose $x_0 \in \mathcal{H}$ and the sequence $\{\beta_n\}_{n\in\mathbb{N}} \subset]0, 1[$. For $n \in \mathbb{N}$ set*

$$x_{n+1} = \frac{\beta_n}{2}\left(R_A\left(R_B x_n + \epsilon_n\right) + \rho_n + x_n\right) + (1-\beta_n)\left(P_B x_n + \frac{\epsilon_n}{2}\right). \tag{2.21}$$

The analogous algorithm to this for inexact alternating projections is the following.

Algorithm 2.4 (inexact alternating projection algorithm). *Choose $x_0 \in \mathcal{H}$ and the sequence $\{\eta_n\}_{n\in\mathbb{N}} \subset]0, 1[$. For $n \in \mathbb{N}$ set*

$$x_{n+1} = (1-\eta_n)x_n + \eta_n\left(P_A\left(P_B x_n + \epsilon_n\right) + \rho_n\right). \tag{2.22}$$

For fixed relaxation parameter β, additional insight into the relaxation $V(T_{AAR}, \beta)$ and the Krasnoselski-Mann-relaxed alternating projection algorithm is gained by considering regularizations of iterated proximal mappings applied to (1.12).

Proposition 2.5 (unification of algorithms).

(i) *Algorithm 1.8 applied to (1.12) with*

$$f_1(x) = \frac{\beta}{2(1-\beta)}\operatorname{dist}^2_A(x) \quad \text{and} \quad f_2(x) = \iota_B(x) \tag{2.23}$$

and $\lambda_n = 1$ for all n is equivalent to Algorithm 2.3 with $\beta_n = \beta$ for all n.

(ii) *Algorithm 1.8 applied to (1.12) with*

$$f_1(x) = \tfrac{1}{2}\operatorname{dist}^2_A(x), \quad \text{and} \quad f_2(x) = \tfrac{1}{2}\operatorname{dist}^2_B(x) \tag{2.24}$$

and relaxation parameter λ_n is equivalent to Algorithm 2.4 with $\eta_n = \lambda_n/2$.

(iii) *Algorithm 1.8 applied to (1.12) on the product space with f_1 and f_2 defined by*

$$f_1(x,y) = \tfrac{1}{2}\operatorname{dist}^2_C(x,y), \quad \text{and} \quad f_2(x,y) = \tfrac{1}{2}\operatorname{dist}^2_D(x,y) \tag{2.25}$$

for $C = \{(x,y) \in \mathcal{H}\times\mathcal{H} \mid x = y\}$, $D = \{(x,y) \in \mathcal{H}\times\mathcal{H} \mid x \in A, y \in B\}$, $\epsilon_n = \rho_n = 0$ for all n and relaxation parameter λ_n is equivalent to gradient descent with step length $\lambda_n/2$ applied to the nonlinear least squares problem (1.2).

Proof. (i) Let

$$f_1(x) = \frac{\beta}{2(1-\beta)}\operatorname{dist}^2_A(x), \quad \text{and} \quad f_2(x) = \iota_B(x)$$

then $\operatorname{prox}_{1,f_2}(x) = P_B x$ and a short calculation yields $\operatorname{prox}_{1,f_1}(x) = x + \beta(P_A x - x)$. The result then follows upon substituting these expressions into (1.22) with $\lambda_n = 1$ for all n and ρ_n of Algorithm 1.8 replaced by ρ_n of (2.21) scaled by β.

(ii) A similar calculation shows that when $f_1(x) = \frac{1-\beta}{2\beta}\operatorname{dist}^2_A(x)$, and $f_2(x) = \frac{\beta}{2(1-\beta)}\operatorname{dist}^2_B(x)$ the recursion (1.22) is equivalent to

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n\big((1-\beta)P_A y_n + \beta(2\beta-1)P_B x_n + 2(\beta-\beta^2)x_n + (\beta-\tfrac{1}{2})\epsilon_n + \tfrac{1}{2}\rho_n\big) \tag{2.26}$$

where $y_n = (1-2\beta)x_n + 2\beta P_B x_n + \epsilon_n$. In particular, when $\beta = 1/2$ we have $x_{n+1} = (1-\frac{\lambda_n}{2})x_n + \frac{\lambda_n}{2}(P_A(P_B x_n + \epsilon_n) + \rho_n)$, the Krasnoselski-Mann relaxation of approximate alternating projections given by (2.22) with relaxation parameter $\eta_n = \lambda_n/2$ for all n.

(iii) We use the product space formulation as in Example 1.2. By (ii) of this theorem, Algorithm 1.8, with relaxation parameter λ_n applied to (1.12) where f_1 and f_2 are defined by (2.25), is equivalent to alternating projections on the product space – Algorithm 2.4 with $x_{n+1} = (1-\lambda_n/2)x_n + \lambda_n/2 P_C P_D x_n$. But by Example 1.2 alternating projections is equivalent to Krasnoselski-Mann-relaxed averaged projections on the product space with relaxation parameter $\lambda_n/2$. To complete the proof we note that, by Example 1.1, Krasnoselski-Mann-relaxed averaged projections on the product space with relaxation parameter $\lambda_n/2$ is equivalent to gradient descent with step length $\lambda_n/2$ applied to the nonlinear least squares problem (1.2). □

In other words, $V(T_{AAR}, \beta)$ is not a relaxation of T_{AAR} but rather the exact instance of Algorithm 1.8 applied to (1.12) with f_1 and f_2 defined by (2.23). Similarly, alternating projections are also an instance of Algorithm 1.8, which, in turn, yields the equivalence of this algorithm to gradient descent.

Proposition 2.5 yields a proof of the next theorem by direct application of Corollary 1.11 with $\lambda_n = 1$ for all n.

THEOREM 2.6 (the inexact RAAR algorithm with fixed β). *Let $\{\rho_n\}_{n\in\mathbb{N}}$ and $\{\epsilon_n\}_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|\rho_n\| + \|\epsilon_n\| < \infty$, and fix $\beta \in]0,1[$, $x_0 \in \mathcal{H}$, and $\lambda_n = 1$ for all n. If F defined by (1.7) is nonempty, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm 2.3 with $\beta_n = \beta$ for all n converges weakly to $\overline{x} \in \mathcal{H}$ as $n \to \infty$ such that $x_* = P_B\overline{x}$ solves*

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \frac{\beta}{2(1-\beta)}\operatorname{dist}^2_A(x) + \iota_B(x). \tag{2.27}$$

If $F = \varnothing$, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (2.21) is unbounded.

Proof. The result follows from Corollary 1.11(ii) once the equivalence of the condition $F \neq \varnothing$ to Assumption (1.9) is established. To see this, note that by Proposition 2.5 the recursion (2.21) is equivalent to (1.22) with $\lambda_n = 1$ for all n applied to (2.27). Moreover, for $f_1 = \frac{\beta}{2(1-\beta)} \operatorname{dist}^2_A(x)$ and $f_2 = \iota_B$, $\partial f_1(x) = \frac{\beta}{1-\beta}(x - P_A x)$ and $\partial f_2(x) = N_B(x)$ (see (1.10) and (1.4)), so the existence of points $x \in F$ implies that $\frac{\beta}{1-\beta}(x - P_A x) = \frac{\beta}{1-\beta} g \in -N_B(x)$, hence Assumption (1.9) holds. Conversely, the existence of points $x \in \mathcal{H}$ and $a \in \partial f_1(x)$ and $b \in \partial f_2(x)$ such that $a + b = 0$ implies that, for such x, $\frac{\beta}{1-\beta}(x - P_A x) \in -N_B(x)$, hence $x \in F$, which completes the proof. □

While the above theorem takes advantage of regularizations to reinterpret the relaxation (2.3), it does not easily allow us to verify the effect of variable β. To account for variable β_n we take a different approach.

THEOREM 2.7 (the inexact RAAR algorithm with variable β). *Fix $\beta \in]0,1[$ and $x_0 \in \mathcal{H}$. Let $\{\beta_n\}_{n\in\mathbb{N}}$ be a sequence in $]0,1[$, and $\{x_n\}_{n\in\mathbb{N}} \in \mathcal{H}$ be generated by Algorithm 2.3 with corresponding errors $\{\epsilon_n\}_{n\in\mathbb{N}}, \{\rho_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$. Define*

$$\nu_n = 2|\beta_n - \beta|\,|(P_A - I)R_B x_n|\,. \tag{2.28}$$

If $F \neq \varnothing$ and

$$\sum_{n\in\mathbb{N}} |\epsilon_n| + |\rho_n| + \nu_n < +\infty, \tag{2.29}$$

then $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point $x_* \in F - \beta g/(1-\beta)$.

Proof. The mapping $V(T_{AAR}, \beta) = \beta T_{AAR} + (1-\beta)P_B$. Then $V(T_{AAR}, \beta)$ is firmly nonexpansive as a convex combination of the two firmly nonexpansive mappings T_{AAR} and P_B. Accordingly, the mapping $R = 2V(T_{AAR}, \beta) - I$ is nonexpansive since $V(T_{AAR}, \beta)$ is firmly nonexpansive if and only if $2V(T_{AAR}, \beta) - I$ is nonexpansive [28, Theorem 12.1]. Moreover, it follows from Lemma 2.1 that $\mathsf{Fix}\, R = \mathsf{Fix}\, V(T_{AAR}, \beta) = F - \beta g/(1-\beta) \neq \varnothing$. Setting $r_n = 2x_{n+1} - x_n$, an elementary calculation shows that

$$|r_n - Rx_n| \leq \beta_n |R_A\left(R_B x_n + \epsilon_n\right) - R_A R_B x_n| + \beta_n|\rho_n| + (1-\beta_n)|\epsilon_n| + 2|\beta_n - \beta|\,|(P_A - I)R_B x_n|\,. \tag{2.30}$$

Now, since R_A is nonexpansive

$$|R_A\left(R_B x_n + \epsilon_n\right) - R_A R_B x_n| \leq |\epsilon_n|$$

and from (2.28)

$$|r_n - Rx_n| \leq |\rho_n| + |\epsilon_n| + \nu_n. \tag{2.31}$$

The recursion (2.21) can thus be rewritten as

$$x_0 \in \mathcal{H} \quad \text{and} \quad x_{n+1} = \tfrac{1}{2}x_n + \tfrac{1}{2}r_n \quad \text{for all } n \in \mathbb{N}, \tag{2.32}$$

where, from (2.29) and (2.31),

$$\sum_{n\in\mathbb{N}} |r_n - Rx_n| < +\infty. \tag{2.33}$$

It then follows from [20, Theorem 5.5(i)] that $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a fixed point of R, which proves the result. □

Analogous results for the important case of the RAAR algorithm with $\beta = 1$, that is, the AAR algorithm, have been treated extensively in [8]. Surprisingly, the proof techniques for these two cases are distinct and it appears that a unification is not readily available.

A more detailed picture of the behavior of iterates of the exact RAAR algorithm can be obtained in the following restricted setting.

Corollary 2.8 (exact RAAR algorithm in Euclidean space). *Let $\mathcal{H}$ be Euclidean space. Fix $\beta \in]0,1[$ and let*

$$x_0 \in \mathbb{R}^n \quad \text{and} \quad x_{n+1} = V(T_{AAR}, \beta)x_n \quad \text{for all } n \in \mathbb{N}.$$

Suppose that $F \neq \varnothing$. Then $\{x_n\}_{n\in\mathbb{N}}$ converges to some point $x \in F - \beta g/(1-\beta)$ and furthermore,

(i) $P_B x_n - P_A P_B x_n \to g$;
(ii) $P_B x_n \to P_B x$ *and* $P_A P_B x_n \to P_A P_B x$;
(iii) $P_B x - P_A P_B x = g$, *hence* $P_B x \in F$ *and* $P_A P_B x \in E$.

Proof. The convergence of $\{x_n\}_{n\in\mathbb{N}}$ follows from Theorem 2.7 (with $\epsilon_n = \rho_n = \nu_n := 0$); denote the limit by x. From (1.7) we can write $x \in F - \beta g/(1-\beta)$ as $x = f - \beta g/(1-\beta)$, where $f = P_B x \in F$ (see also [8, Proposition 2.4.(ii)]). Since the mappings P_A, P_B, R_A, R_B are continuous, (ii) follows. Next, using (2.5c), we have

$$P_B x_n - P_A R_B x_n \to P_B x - P_A R_B x = g. \tag{2.34}$$

Hence

$$|g| \leq |P_B x_n - P_A P_B x_n| \leq |P_B x_n - P_A R_B x_n| \to |g|, \tag{2.35}$$

and thus $|P_B x_n - P_A P_B x_n| \to |g|$. Now (i) follows from [8, Proposition 2.5]. Taking the limit in (i) yields (iii). ∎

We would like to note in closing this subsection that the duality theory for (1.12) with f_1 and f_2 given by (2.23) has been detailed in [2, Section 2]. The connection between algorithms (1.8) and (1.7) and (2.3) allows for an attractive synthesis in the convex setting. However, at this time the nonconvex theory is much less developed than the convex theory. A notable exception is the recent work of Moudafi [45], who studies the convergence of the prox-gradient method in a prox-regular setting. Nevertheless, the view of the parameter β as a weight in a regularized objective does not, in our opinion, lead to a natural justification for dynamic β_n as does the interpretation of this parameter as a relaxation. This is discussed in greater detail in the next section.

2.2. Controlling the iterates. The implementation of the RAAR algorithm that we studied in [38] was motivated by the following observation that indicates that the relaxation parameter β might be used to *steer* the iterates.

Proposition 2.9. *Let $x \in \mathcal{H}$ and suppose that $F \neq \varnothing$.*

(i) $\operatorname{dist}\big(x, \operatorname{Fix} V(T_{AP}, \beta)\big) = \operatorname{dist}\big(x, F + \beta g\big)$ *for all* $\beta \in (0,1]$.
(ii) *If* $A \cap B \neq \varnothing$, *then* $\operatorname{dist}\big(x, \operatorname{Fix} V(T_{AAR}, \beta)\big) = \operatorname{dist}\big(x, A \cap B\big)$ *for all* $\beta \in]0,1[$; *otherwise,* $\lim_{\beta \nearrow 1} \operatorname{dist}\big(x, \operatorname{Fix} V(T_{AAR}, \beta)\big) = +\infty$.

Proof. The proof of (i) follows immediately from (2.6). To see (ii), note that if $A \cap B \neq \varnothing$, then $g = 0$ and $\operatorname{Fix} V(T_{AAR}, \beta) = A \cap B$, which proves the first part of the statement. Now assume $A \cap B = \varnothing$ and

fix $f_0 \in F$. Then $g \neq 0$ and F is contained in the hyperplane $\{x \in \mathcal{H} \mid \langle x - f_0,\ g\rangle = 0\}$ [3, Lemma 2.2(v)]. Hence, it follows from Lemma 2.1 that

$$\mathsf{Fix}\ V(T_{AAR}, \beta) = F - \frac{\beta}{1-\beta} g \subset \left\{ x \in \mathcal{H} \;\middle|\; \left\langle x + \frac{\beta}{1-\beta} g - f_0,\ g \right\rangle = 0 \right\} = H_\beta. \tag{2.36}$$

Accordingly,

$$\text{dist}\ \big(x, \mathsf{Fix}\ V(T_{AAR}, \beta)\big) \geq \text{dist}\ (x, H_\beta) = \frac{\left|\left\langle x + \frac{\beta}{1-\beta} g - f_0,\ g \right\rangle\right|}{|g|} \geq \frac{\beta}{1-\beta}|g| - \frac{|\langle x - f_0,\ g\rangle|}{|g|} \tag{2.37}$$

which proves the second assertion of part (ii). □

By Proposition 2.9, for any estimate x_n "close" to $\mathsf{Fix}\ V(T_{AAR}, \beta_n)$, there is a β_{n+1} such that x_n is comparatively distant to $\mathsf{Fix}\ V(T_{AAR}, \beta_{n+1})$. It will become clear in the next section that it is the proximity to the set F, rather than $\mathsf{Fix}\ V(T_{AAR}, \beta_n)$, that is critical to the quality of an iterate x_n. We therefore use the relaxation parameter β_n to control the *step size* of an iterate toward the set F. By comparison, the relaxation parameter β has very little effect on the iterates x_n of the alternating projection algorithm. The next proposition shows that by varying β the step size can be regulated in the direction of the gap vector g.

PROPOSITION 2.10. *Let $x \in \mathcal{H}$ satisfy $|x - x_{\beta_1}| < \delta$ where $x_{\beta_1} \in \mathsf{Fix}\ V(T_{AAR}, \beta_1)$, $\delta > 0$ and $\beta_1 \in\ ]0, 1[$. Then, for all $\beta_2 \in\ ]0, 1[$, we have*

$$\left| V(T_{AAR}, \beta_2)x - \left(f_{\beta_1} - \frac{\beta_2}{1-\beta_1} g \right) \right| < \delta, \tag{2.38}$$

where $f_{\beta_1} = P_B x_{\beta_1} \in F$.

Proof. This was proved in [38, Proposition 2.3] □

This ability to control the step lengths with the relaxation parameter stands out next to other relaxed projection algorithms. For this reason descent algorithms are often preferred since there is ample theory for determining optimal step sizes.

3. Nonconvex analysis.

3.1. Prox-regular Sets. In this section A is still convex, but we allow the set B to be nonconvex. Such a situation is encountered in the numerical solution to the phase retrieval problem in inverse scattering [39, 6, 38], and is therefore of great practical interest. Indeed, our results form the basis for proving local convergence of some phase retrieval algorithms for inconsistent (noisy) problems which, to our knowledge, would be the first such results. The central notion for getting a handle on this situation is *prox-regularity* as developed by Poliquin and Rockafellar[48, 47]. Prox-regular sets were, to our knowledge, first introduced by Federer [26] though he called them sets of *positive reach*, and are characterized as those sets C for which the projection is locally single-valued and continuous from the strong topology in the domain to the weak topology in the range [49, Theorem 3.1]. The main difficulty for our analysis is that prox-regularity is a local property relative to elements of the set B while the fixed points of the mapping $V(T, \beta)$ lie somewhere in the normal cone to B at the local best approximation points of this set. A localized normal cone mapping is obtained through the *truncated normal cone mapping*

$$N_C^r(x) := \begin{cases} N_C(x) \cap \text{int}\ \mathbb{B}(0, r) & x \in C \\ \varnothing & x \notin C \end{cases} \tag{3.1}$$

where $\mathbb{B}(0,r)$ is the closed ball of radius r centered on the origin. The principal result we draw from can be found in [49], Corollary 2.2 and Proposition 3.1.

LEMMA 3.1 (properties of prox-regular sets). *Let $C \subset \mathcal{H}$ be is prox-regular at $\overline{x}$. Then for some $r > 0$ and a neighborhood of $\overline{x}$, denoted $\mathcal{N}(\overline{x})$, the truncated normal cone mapping N_C^r is hypomonotone on $\mathcal{N}(\overline{x})$, that is, there is a $\sigma > 0$ such that*

$$\langle y_1 - y_2,\ x_1 - x_2\rangle \geq -\sigma|x_1 - x_2|^2 \quad \textit{whenever } y_i \in N_C^r(x_i) \textit{ and } x_i \in \mathcal{N}(\overline{x}).$$

As suggested by Proposition 2.9 we can control to some extent the location of the fixed points of $V(T,\beta)$ by adjusting the parameter β. In particular, note that for $\beta = 0$ we have $V(T,0) = P_B$ hence we can adjust β so that the fixed points remain in prox-neighborhoods of the best approximation points in B.

The next result is a prox-regular analog of (1.3).

LEMMA 3.2. *For C prox-regular at $\overline{x}$ there exist $\epsilon > 0$ and $\sigma > 0$ such that whenever $x \in C$ and $v \in N_C(x)$ with $|x - \overline{x}| < \epsilon$ and $|v| < \epsilon$ one has*

$$\langle x' - x, v\rangle \leq \sigma|x' - x|^2 \quad \textit{for all } x' \in C \quad \textit{with } |x' - x| < \epsilon. \tag{3.2}$$

Proof. Since C is prox-regular at $\overline{x}$, by Lemma 3.1 the truncated normal cone mapping $N_C^r(x)$ is hypomonotone on a neighborhood $\mathcal{N}(\overline{x})$, that is, there are $\sigma > 0$ and $\epsilon > 0$ such that

$$\langle x' - x, v\rangle \leq \sigma|x' - x|^2,$$

whenever $v \in N_C(x)$, and $0 \in N_C(x')$ with $|v| < \epsilon$ and $|x' - x| < \epsilon$. □

A stronger version (with different proof) of the above proposition can be found in [49, Proposition 1.2]

For this prox-regular setting we must define local versions of the sets G, E, and F defined in (1.7).

DEFINITION 3.3 (local best approximation points). *For A convex and B nonconvex, a point $f \in B$ is a* local best approximation point *if there exists a neighborhood $\mathcal{N}(f)$ on which $|f - P_A f| \leq |b - P_A b|$ for all $b \in B \cap \mathcal{N}(f)$. For such a point, we define*

$$\begin{aligned} &G_{\mathcal{N}(f)} := \overline{(B \cap \mathcal{N}(f)) - A} \quad \textit{and for } g := f - P_A f \in P_{G_{\mathcal{N}(f)}}0 \\ &E_{\mathcal{N}(f)}(g) := A \cap ((B \cap \mathcal{N}(f))) - g), \quad F_{\mathcal{N}(f)}(g) := (A + g) \cap (B \cap \mathcal{N}(f)). \end{aligned} \tag{3.3a}$$

If $|f - P_A f| \leq |b - P_A b|$ for all $b \in B$ then $f \in B$ is a global best approximation point. *Whether or not such a point exists, the following sets are well defined*

$$\begin{aligned} &G := \overline{B - A} \quad \textit{and for } g \in P_G 0 \\ &E(g) := A \cap (B - g), \quad F(g) := (A + g) \cap B. \end{aligned} \tag{3.3b}$$

From the above definition it is immediate that any global best approximation point is also a local best approximation point. Note that $P_{G_{\mathcal{N}(f)}}0$ and $P_G 0$ are now possibly *sets* of gap vectors since, for A convex and B prox-regular, $G = \overline{B - A}$ is not in general convex. For any $g_1, g_2 \in P_{G_{\mathcal{N}(f)}}0$, although it may happen that $g_1 \neq g_2$, it is still the case that $|g_1| = |g_2|$, hence the (local) gap between the sets A and B in the nonconvex setting is still well defined.

We will assume the following throughout the rest of this work.

Assumption 3.4 (prox-regularity of G). *The set G is prox-regular at all $g \in P_G 0$ and for every local best approximation point $f \in B$, the set $G_{\mathcal{N}(f)}$ is prox-regular at the corresponding point $g_f := f - P_A f \in P_{G_{\mathcal{N}(f)}} 0$.*

Example 3.5. *Consider the example in $\mathbb{R}^2$ where $A = \{(0, x_2)$ for $x_2 \in [-2, \epsilon]\}$ for $\epsilon \geq 0$ and*

$$B = \left\{ (x_1, x_2) \in \mathbb{R}^2 \;\middle|\; \begin{array}{cc} x_1 = \pm\sqrt{1 - x_2^2} & \text{for } x_2 \geq 0, \\ x_1 = -1 & \text{for } x_2 \in [0, -1], \\ x_1 = -\sqrt{1 - (x_2 + 1)^2} & \text{for } x_2 \in [-2, -1] \end{array} \right\}.$$

The corresponding set G is not prox-regular everywhere. In particular, if $\epsilon = 0$ it is not prox-regular at the point $(-\sqrt{3}/2, 1/2)$ since the projection onto G is multivalued along the line segment $(-\sqrt{3}/2, 1/2) + \tau(1, 0)$ for all $\tau \in [0, \sqrt{3}/2]$. For this example, however, this is the only point in G where prox-regularity fails. Since A and B intersect at the point $(0, -2)$, the global gap vector is $(0, 0)$, and $F = E = \{(0, -2)\}$. Each of the vectors in the set $\left\{ (x_1, x_2) \in \mathbb{R}^2 \middle| x_1 = \pm\sqrt{1 - x_2^2} \quad \text{for } x_2 \geq 0, \right\}$, on the other hand, is a local gap vector relative to some neighborhood of points in B. At the point $f = (0, 1)$, for example, all of these vectors are local gap vectors of the set $G_{\mathcal{N}(f)}$ where $\mathcal{N}(f)$ is a disk of radius $\sqrt{2}$. The corresponding local best approximation points in B are $F_{\mathcal{N}(f)}(g) = \{g\}$ while the local best approximation points in A, $E_{\mathcal{N}(f)}(g) = \{(0, 0)\}$ for all $g \in \left\{ (x_1, x_2) \in \mathbb{R}^2 \;\middle|\; x_1^2 + x_2^2 = 1 \text{ for } x_2 \geq 0 \right\}$. We call this collection of best approximation points a fan, *characterized by the nonuniqueness of the gap vector. We shall disallow such regions in what follows. Indeed, if in the definition of A we set $0 < \epsilon \ll 1$, then there is no such fan region and the unique best approximation point in B corresponding to $(0, \epsilon) \in A$ is $f = (0, 1)$ with corresponding unique gap vector $g = (0, 1 - \epsilon)$.* □

The next fact is an adjustment of [8, Proposition 2.5] for B prox regular.

Proposition 3.6. *Let A be closed convex and B prox-regular subsets of $\mathcal{H}$. Suppose that $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are sequences in A and $B \cap \mathcal{N}(f)$, respectively with f a local best approximation point, $\mathcal{N}(f)$ a suitable neighborhood and $b_n - a_n \to g = f - P_A f \in P_{G_{\mathcal{N}(f)}} 0$. Then the following hold.*

(i) *$b_n - P_A b_n \to g$ while $P_B a_n - a_n \to \widetilde{G}_{\mathcal{N}(f)} \subset P_{G_{\mathcal{N}(f)}} 0$.*
(ii) *The cluster points of $\{a_n\}_{n\in\mathbb{N}}$ and $\{P_A b_n\}_{n\in\mathbb{N}}$ belong to $E_{\mathcal{N}(f)}(g)$. The cluster points of $\{b_n\}_{n\in\mathbb{N}}$ belong to $F_{\mathcal{N}(f)}(g)$. Consequently, the cluster points of the sequences*
$$\big\{(a_n, b_n)\big\}_{n\in\mathbb{N}},\ \big\{(P_A b_n, b_n)\big\}_{n\in\mathbb{N}}$$
are local best approximation pairs relative to (A, B),
(iii) *If g is the unique gap vector on $\mathcal{N}(f)$, that is if $P_{G_{\mathcal{N}(f)}} 0 = g$, then*
$$b_n - a_n \to g \quad \Longleftrightarrow \quad |b_n - a_n| \to |g|.$$

Proof. Since

$$\begin{aligned} |b_n - a_n| &\geq \max\big\{|b_n - P_A b_n|, |P_B a_n - a_n|\big\} \\ &\geq \min\big\{|b_n - P_A b_n|, |P_B a_n - a_n|\big\} \\ &\geq |g|, \end{aligned}$$

we conclude that $\big\{|b_n - P_A b_n|\big\}_{n\in\mathbb{N}}$ and $\big\{|P_B a_n - a_n|\big\}_{n\in\mathbb{N}}$ both converge to $|\widetilde{g}|$ for any $\widetilde{g} \in P_{G_{\mathcal{N}(f)}} 0$. Since A is convex, $P_A b_n$ is single-valued and continuous, hence $b_n - P_A b_n \to g$. Since B is prox-regular P_B is possibly set-valued and $\lim_{n\to\infty} P_B a_n - a_n = \widetilde{G}_{\mathcal{N}(f)}$ a subset of $P_{G_{\mathcal{N}(f)}} 0$. Hence (i) holds. Let $a \in A$ be a cluster point of $\{a_n\}_{n\in\mathbb{N}}$, say $a_{n_k} \to a$. Then $b_{n_k} \to g + a \in B \cap \mathcal{N}(f) \cap (g + A) = F_{\mathcal{N}(f)}(g)$, hence $a \in A \cap (B \cap \mathcal{N}(f) - g) = E_{\mathcal{N}(f)}(g)$. The arguments for $\{b_n\}_{n\in\mathbb{N}}$ and $\{P_A b_n\}_{n\in\mathbb{N}}$ are similar. Finally, (iii) follows from (i) and the fact that g is the unique gap vector. □

Remark 3.7. *Convergence of* $|b_n - a_n| \to |g|$ *does not in general imply that* $b_n - a_n \to g$. *To see this, consider* B *and* A *in Example 3.5 with* $\epsilon = 0$. *Construct the sequences* $a_n := (0, -1/n)$ *and* $b_n := (-1, -1/n)$ *for all* n *and let* $g = (0, 1)$. *Now,* $a_n \to (0, 0)$, $b_n \to (-1, 0)$ *and* $b_n - a_n \to (-1, 0) = \widetilde{g} \neq g$, *even though both belong to* $P_{G_{\mathcal{N}(f)}} 0$ *and* $|b_n - a_n| \to |g|$ *when* $f = (0, 1)$ *and* $\mathcal{N}(f)$ *a disk of radius* $\sqrt{2}$. *Note also that* $P_B a_n \to (-1, 0)$ *while* $P_B a = \left\{(x_1, x_2) \in \mathbb{R}^2 \;\middle|\; x_1^2 + x_2^2 = 1 \quad \text{and } x_2 \geq 0\right\}$.

3.2. Relaxed Averaged Alternating Reflections: prox regular sets. The difference between T_{AP} and T_{AAR} is in the control exerted on the fixed points of the respective mappings by the relaxation strategy $V(T, \beta)$. As shown in (2.6) in the case of T_{AP}, the relaxation $\beta \in (0, 1]$ simply shifts the set of fixed points from best approximation points in A to their corresponding points in B. In the nonconvex setting this shift property is restricted to local best approximation points. Hence, the relaxation parameter does not change in any significant way the set of fixed points and, in particular, it does not change the correspondence between the set of local best approximation points and the fixed points of $V(T_{AP}, \beta)$. For T_{AAR} the relaxation parameter in (2.21) the nonconvex situation is quite different. Indeed, as we show in Lemma 3.8, the relaxation parameter β can be chosen to *limit* the set of local best approximation points that correspond to fixed points of $V(T_{AAR}, \beta)$, thus eliminating *bad* local best approximation points.

Before proceeding with the prox-regular versions of Theorem 2.7 and Corollary 2.8, we need to define what we mean by $V(T, \beta)x$ in the case when $P_B x$ is multi-valued. We shall define this as

$$V(T_{AP}, \beta)x := \{v = \beta P_A b + (1 - \beta)b \quad | \quad b \in P_B x\}. \tag{3.4a}$$

$$V(T_{AAR}, \beta)x := \left\{v = \frac{\beta}{2}\left(R_A(2b - x) + x\right) + (1 - \beta)b \quad | \quad b \in P_B x\right\}. \tag{3.4b}$$

Lemma 3.8 (characterization of fixed points). *For* A *closed and convex and* B *prox-regular, suppose that Assumption 3.4 holds and define* $V(T_{AAR}, \beta)$ *by (3.4b) for* $\beta \in]0, 1[$ *fixed. Then*

$$\mathsf{Fix}\, V(T_{AAR}, \beta) \subset \left\{f - \frac{\beta}{1 - \beta}(f - P_A f) \mid f \in B \text{ is a local best approximation point}\right\} \tag{3.5a}$$

and the two sets are equal for all $\beta \leq 1/2$. *Moreover, for every* $x \in \mathsf{Fix}\, V(T_{AAR}, \beta)$, *there is a local best approximation point* f *and corresponding gap vector* g_f *with*

$$x \in P_B x - \frac{\beta}{1 - \beta} g_f, \tag{3.5b}$$

$$g_f \in P_B x - P_A R_B x. \tag{3.5c}$$

By comparison, for all $\beta \in]0, 1[$ *fixed,*

$$\mathsf{Fix}\, V(T_{AP}, \beta) = \{f - \beta(f - P_A f) \mid f \in B \text{ is a local best approximation point}\}. \tag{3.6}$$

Proof. The proof is almost identical to that of Lemma 2.1. We skip most of the details and point only to the main differences.

To prove (3.5a)-(3.5c) we must take account of two issues: first, that P_B might not be single-valued at all $x \in \mathsf{Fix}\, V(T_{AAR}, \beta)$ and the relation (2.14) does not hold for B prox-regular. The possible multi-valuedness of $P_B x$ is handled by choosing $f \in P_B x$ for a given $x \in \mathsf{Fix}\, V(T_{AAR}, \beta)$ and setting $y = x - f$.

The corresponding gap vector is uniquely determined by $g_f = f - P_A(f)$. This changes (2.8) and (2.9) to *inclusions*

$$P_A(2f - x) \in P_A R_B x \quad \text{and} \quad P_A R_B x - P_B x = T_{AAR} x - x \tag{3.7}$$

by (2.7). The second equation is actually an expression of set equality. When T_{AAR} is restricted to the selection $f \in P_B x$, which we write as $T_{AAR}|_f$, this yields

$$P_A(2f - x) - f = T_{AAR}|_f x - x. \tag{3.8}$$

For $x \in \mathsf{Fix}\,(V(T_{AAR}, \beta))$ equations (3.7) and (3.8) give

$$(1 - \beta)(x - f) = \beta(T_{AAR}|_f - x) = \beta(P_A(2f - x) - f) \tag{3.9}$$

hence, with $y = x - f$,

$$f + \frac{1 - \beta}{\beta} y = P_A(2f - x). \tag{3.10}$$

This is the same result as (2.11) for the selection $f \in P_B x$. As with (2.13), using (2.12) and (3.10) we have, for any $a \in A$ nonempty, closed and convex,

$$\frac{1}{\beta} \langle -a + f,\ y \rangle + \frac{1 - \beta}{\beta^2} |y|^2 \leq 0. \tag{3.11}$$

On the other hand, since B is nonempty prox-regular and $f \in P_B x$, by Lemma 3.2 we have

$$\langle b - f, x - f \rangle \leq \sigma |f - b|^2 \tag{3.12}$$

where $x - f \in N_B(f)$, and $0 \in N_B(b)$ for b close enough to f. This yields (compare to (2.15))

$$\langle b - f,\ y \rangle = \langle b - f,\ x - f \rangle \leq \sigma |b - f|^2. \tag{3.13}$$

Now, (3.11) and (3.13) yield

$$\langle b - a,\ y \rangle \leq \langle b - f,\ y \rangle - \frac{1 - \beta}{\beta} |y|^2 \leq \sigma\, |f - b|^2 - \frac{1 - \beta}{\beta} |y|^2. \tag{3.14}$$

The right hand side is nonpositive for all b close enough to f. The rest of the proof follows the proof of Lemma 2.1 with the caveat that the sequence $b_n \to f$ be chosen close enough to f that

$$\sigma\, |f - b_n|^2 - \frac{1 - \beta}{\beta} |y|^2 \leq 0 \quad \text{for all } n.$$

The identities (3.5b)-(3.5c) follow immediately since $f \in P_B x$ is a local best approximation point.

To prove that the set inequality in (3.5a) is not, in general, tight we show that, given a local best approximation point $f \in B$ and corresponding gap vector g_f,

$$f - \frac{\beta}{1 - \beta} g_f \in \mathsf{Fix}\ V(T_{AAR}, \beta) \quad \text{if and only if} \quad f \in P_B \left(f - \frac{\beta}{1 - \beta} g_f \right). \tag{3.15}$$

The "easy" implication is that the left hand side of (3.15) implies the right hand side (expand $0 \in (I - V(T_{AAR}, \beta)) \left(f - \frac{\beta}{1-\beta} g_f \right)$ in terms of P_B and P_A and "solve" for $P_B \left(f - \frac{\beta}{1-\beta} g_f \right)$). The other implication follows exactly as in Lemma 2.1 with the generalization to inclusions since the projection onto B need not be single-valued.

Finally, to show that set equality holds in (3.5a) for all $\beta \leq 1/2$ note that, for any local best approximation point $f \in B$ with B prox-regular, $f \in P_B(f - \frac{\beta}{1-\beta} g_f)$ for all $\beta \in [0, 1/2]$ where g_f is the corresponding gap vector. The result then follows from (3.15).

For T_{AP}, since the parameter β simply shifts the fixed point within the gap of a local best approximation pair as β ranges from 0 to 1, the fixed points of $V(T_{AP}, \beta)$ coincide precisely with the local best approximation points of A and B, whence (3.6). □

Corollary 3.9. *$\mathsf{Fix}\, V(T_{AAR}, \beta) = \varnothing$ for all $\beta > 0$ if and only if B does not contain a local best approximation point.*

Proof. This follows immediately from (3.5). □

Remark 3.10. *Note that, while in the case of T_{AAR} all local best approximation points correspond to fixed points of $V(T_{AAR}, \beta)$ for $\beta \in [0, 1/2]$, this is not the case for $\beta > 1/2$. Indeed, by (3.15) of Lemma 3.8, the fixed points of $V(T_{AAR}, \beta)$ consist only of local best approximation points for which $f - \frac{\beta}{1-\beta} g_f$ is in a proximal neighborhood of B. This certainly will not hold for all $\beta \in [1/2, 1[$. One might envision an algorithmic strategy for* filtering out *certain local best approximation points by choosing β large enough. Of course, how such a filtering might work in practice depends entirely on local proximal properties the set B. The point is that by simply increasing β, one can avoid local minima. This is a potentially powerful algorithmic tool for global projection algorithms for nonconvex problems.*

We finish this section with nonconvex versions of Theorem 2.7 and Corollary 2.8. The convex results exploited the firm nonexpansiveness of the fixed point mapping, or equivalently maximal monotonicity. We show that, for the nonconvex problem, if this property holds locally, then local versions of the results of Section 2.1 follow. This is not an empty assumption as Example 3.5 illustrates. Indeed, for the sets defined there with $\epsilon > 0$, an elementary calculation of $V(T_{AAR}, 1/2)x$ for points x on convex neighborhoods of $(0, \epsilon) \in \mathsf{Fix}\, V(T_{AAR}, 1/2)$ (not even very small neighborhoods) shows that $2V(T_{AAR}, 1/2) - I$ is nonexpansive, hence $V(T_{AAR}, 1/2)$ is locally firmly nonexpansive.

One consequence of such an assumption is the following.

Proposition 3.11. *For either $T = T_{AAR}$ or $T = T_{AP}$, if $V(T, \beta)$ is firmly nonexpansive on a neighborhood $\mathcal{N}(x_0)$ of $x_0 \in \mathsf{Fix}\, V(T, \beta)$ then $g_0 = P_B x_0 - P_A P_B x_0$ is the unique gap vector on $\mathcal{N}(x_0)$, that is, for all $x \in \mathsf{Fix}\, V(T, \beta) \cap \mathcal{N}(x_0)$ one has $P_B x - P_A P_B x = P_B x_0 - P_A P_B x_0$. Moreover, P_B is single-valued on $\mathcal{N}(x_0)$.*

Proof. We prove the statement for $T = T_{AAR}$ as the proof for $T = T_{AP}$ is almost identical. Let x_1 be any fixed point on $\mathcal{N}(x_0)$ with corresponding gap vectors $g_1 \in P_B x_1 - P_A P_B x_1$ and let $b_j \in P_B x_j$, for $j = 0, 1$. Then by Lemma 3.8 $x_j = b_j - \frac{1-\beta}{\beta} g_j$, $j = 0, 1$. Since P_A is nonexpansive we have

$$|g_1 - g_0|^2 = |b_1 - b_0|^2 + |P_A b_1 - P_A b_0|^2 - 2\langle b_1 - b_0,\ P_A b_1 - P_A b_0\rangle \leq |b_1 - b_0|^2 - |P_A b_1 - P_A b_0|^2 \quad (3.16)$$

and

$$|P_A b_1 - P_A b_0| \leq |b_1 - b_0|. \quad (3.17)$$

If $|P_A b_1 - P_A b_0| = |b_1 - b_0|$ then by (3.16) $g_1 = g_0$. If, on the other hand, $|P_A b_1 - P_A b_0| < |b_1 - b_0|$ then $|y_1 - x_0| < |x_1 - x_0|$ where $y_1 := b_1 - \left(\frac{1-\beta}{\beta} + \epsilon\right) g_1$ for some $\epsilon > 0$ small enough. A straightforward calculation shows that

$$Ry_1 = b_1 - \left(\frac{1-\beta}{\beta} + (2\beta - 1)\epsilon\right) g_1 \quad \text{and} \quad Rx_0 = x_0$$

where $R := 2V(T_{AAR}, \beta) - I$. So for $\beta < 1$, $|y_1 - x_0| < |Ry_1 - Rx_0|$ which contradicts the assumption that $V(T_{AAR}, \beta)$ is firmly nonexpansive. It must hold, then, that $|P_A b_1 - P_A b_0| = |b_1 - b_0|$ hence $g_1 = g_0$.

To see that the projection P_B is single-valued on $\mathcal{N}(x_0)$, consider any $x \in \mathcal{N}(x_0)$ and the corresponding vectors $y_1 = b_1 - x$ and $y_2 = b_2 - x$ with $b_j \in P_B x$ $(j = 1, 2)$. The same argument as above above applies here with A replaced by $\{x\}$ to show that $y_1 = y_2$ since $V(T_{AAR}, \beta)$ is firmly nonexpansive, hence $b_1 = b_2$. Since x was arbitrarily chosen, this completes the proof. □

The inexact RAAR Algorithm 2.3 is modified in the obvious way to inclusions for B prox-regular. For variable relaxation parameters, we then have the following generalization of Theorem 2.7.

THEOREM 3.12 (inexact prox-regular RAAR algorithm, variable β). *For A closed and convex and B prox-regular, suppose that Assumption 3.4 holds. Let $\beta \in]0, 1[$ be small enough that $\mathsf{Fix}\, V(T_{AAR}, \beta) \neq \varnothing$. Suppose that $V(T_{AAR}, \beta)$ is firmly nonexpansive on a convex neighborhood $\mathcal{N}(\overline{x})$ of $\overline{x} \in \mathsf{Fix}\, V(T_{AAR}, \beta)$ with $\operatorname{dom} V(T_{AAR}, \beta) = \mathcal{H}$. Choose $x_0 \in \mathcal{N}(\overline{x})$, let $\{\beta_n\}_{n\in\mathbb{N}}$ be a sequence in $]0, 1[$, and $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$ be generated by Algorithm 2.3 with corresponding errors $\{\epsilon_n\}_{n\in\mathbb{N}}, \{\rho_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$. Define*

$$\nu_n = 2|\beta_n - \beta|\,|(P_A - I)R_B x_n| \tag{3.18}$$

and suppose that

$$\sum_{n\in\mathbb{N}} |\epsilon_n| + |\rho_n| + \nu_n = M \tag{3.19}$$

for $M \in \mathbb{R}$ small enough that $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{N}(\overline{x})$. Then $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in

$$\mathsf{Fix}\, V(T_{AAR}, \beta) \cap \mathcal{N}(\overline{x}) \subset \left\{ f - \frac{\beta}{1-\beta}(f - P_A f) \;\middle|\; f \in F_{P_B\mathcal{N}(\overline{x})} \right\},$$

where $F_{P_B\mathcal{N}(\overline{x})}$ denotes the set of best approximation points in B corresponding to the projected neighborhood $P_B\mathcal{N}(\overline{x})$. Convergence is strong if any one of the following hold:

- $\underline{\lim}\, \operatorname{dist}_{\mathsf{Fix}\, V(T_{AAR},\beta)}(x_n) = 0$;
- $\operatorname{int} \mathsf{Fix}\, V(T_{AAR}, \beta) \neq \varnothing$
- $V(T_{AAR}, \beta)$ *is* demicompact at 0.

Proof. Let $R = 2V(T_{AAR}, \beta) - I$ and note that $\mathsf{Fix}\, R = \mathsf{Fix}\, V(T_{AAR}, \beta)$, which, by assumption, is nonempty. Moreover, it follows from Lemma 3.8 that

$$\mathsf{Fix}\, R \cap \mathcal{N}(\overline{x}) = \mathsf{Fix}\, V(T_{AAR}, \beta) \cap \mathcal{N}(\overline{x}) \subset \left\{ f - \frac{\beta}{1-\beta}(f - P_A f) \;\middle|\; f \in F_{P_B\mathcal{N}(\overline{x})} \right\}.$$

Following the proof of Theorem 2.7, let $r_n = 2x_{n+1} - x_n$ and rewrite the recursion (2.21) as

$$x_0 \in \mathcal{H} \quad \text{and} \quad x_{n+1} = \tfrac{1}{2}x_n + \tfrac{1}{2}r_n \quad \text{for all } n \in \mathbb{N}, \tag{3.20}$$

where, from (3.19) and

$$|r_n - Rx_n| \leq |\rho_n| + |\epsilon_n| + \nu_n, \tag{3.21}$$

it holds that

$$\sum_{n\in\mathbb{N}} |r_n - Rx_n| < M \tag{3.22}$$

for M small enough that $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{N}(\overline{x})$. Since $V(T_{AAR},\beta)$ is firmly nonexpansive on this neighborhood with dom $V(T_{AAR},\beta) = \mathcal{H}$ then R is nonexpansive on the same neighborhood with dom $R = \mathcal{H}$. The result then follows from [20, Theorem 5.5]. □

A more detailed picture of the behavior of iterates of the exact RAAR algorithm can be obtained in the following restricted setting.

Corollary 3.13 (exact prox-regular RAAR algorithm in Euclidean space). *Let $\mathcal{H}$ be Euclidean space. For the assumptions of Theorem 3.12, fix $\beta \in]0,1[$ small enough that* $\mathsf{Fix}\, V(T_{AAR},\beta) \neq \varnothing$ *and let*

$$x_0 \in \mathcal{N}(\overline{x}) \quad \text{and} \quad x_{n+1} = V(T_{AAR},\beta)x_n \quad \text{for all } n \in \mathbb{N}$$

where $\overline{x} \in$ $\mathsf{Fix}\, V(T_{AAR},\beta)$ *with corresponding local best approximation point $f = P_B\overline{x}$ and gap vector $g_f = f - P_A f$. Then $\{x_n\}_{n\in\mathbb{N}}$ converges to a point $x \in F_{\mathcal{N}(f)}(g_f) - \beta g_f/(1-\beta)$ and*

(i) $P_B x_n - P_A P_B x_n \to g_f$;
(ii) $P_B x_n \to P_B x$ *and* $P_A P_B x_n \to P_A P_B x$;
(iii) $P_B x - P_A P_B x = g_f$, *hence* $P_B x \in F_{\mathcal{N}(f)}(g_f)$ *and* $P_A P_B x \in E_{\mathcal{N}(f)}(g_f)$.

Proof. The convergence of $\{x_n\}_{n\in\mathbb{N}}$ follows from Theorem 3.12 (with $\mu_n := \nu_n := 0$); denote the limit by x. From (3.3a) we can write $x \in F_{\mathcal{N}(f)}(g_f) - \beta g_f/(1-\beta)$ as $x = f - \beta g_f/(1-\beta)$, where $f = P_B x \in F_{\mathcal{N}(f)}(g_f)$ (see also [8, Proposition 2.4.(ii)]). Since the mappings P_A, P_B, R_A, R_B are continuous, (ii) follows. Next, using (3.5c), we have

$$P_B x_n - P_A R_B x_n \to P_B x - P_A R_B x = g_f. \tag{3.23}$$

Hence

$$|g_f| \leq |P_B x_n - P_A P_B x_n| \leq |P_B x_n - P_A R_B x_n| \to |g_f|, \tag{3.24}$$

and thus $|P_B x_n - P_A P_B x_n| \to |g_f|$. Now (i) follows from Proposition 3.6 and Proposition 3.11. Taking the limit in (i) yields (iii). □

4. Conclusion and Open Problems. In this work we have laid some groundwork for a comprehensive theory of the asymptotic behavior of projection algorithms in prox-regular settings, with particular focus on the RAAR algorithm. The RAAR algorithm has many attractive features, namely that it is robust for consistent and inconsistent problems, the relaxation parameter can be interpreted as a *step length* and thus can be optimized, and moreover, the relaxation parameter can be used to avoid "bad" local minima. In the convex setting the RAAR algorithm, together with the classical alternating projections and averaged projections algorithms, can be viewed as instances of the classical Lions-Mercier/Douglas-Rachford algorithm applied to the problem of minimizing the sum of two maximal monotone mappings, hence the analysis of the RAAR algorithm can be broadly applied. We conjecture that these correspondences carry over to the nonconvex setting, however the details of this correspondence are beyond the scope of the present study.

The analytical tools that we use derive from analogs in convex theory. As one would expect from nonconvex problems, our most general results are local in nature. Our hope is that this analysis can serve as a guide to the analysis of similar algorithms. For the purpose of proving the convergence of the RAAR algorithm for the phase retrieval problem in crystallography, it remains to be shown that the fixed point mapping $V(T_{AAR},\beta)$ is firmly nonexpansive on a neighborhood of a fixed point. This question would be quickly resolved by sufficient conditions under which firm nonexpansiveness holds in nonconvex settings. Less restrictive notions of firm nonexpansiveness, namely *quasi-1/2-averaged* mappings as studied in [5, 20], could also be quite fruitful here. Another key assumption for our results was the prox-regularity of the set $G_{\mathcal{N}(f)} = \overline{B \cap \mathcal{N}(f) - A}$

at local best approximation points f. We conjecture that this assumption as well as the assumption of local firm nonexpansiveness of the corresponding reflection can be removed on neighborhoods of local best approximation points. The assumption that one of the sets is convex, not just prox-regular, was useful for our proofs, but is probably not necessary in general. With the exception of what we called fan regions, local best approximation points will by definition be in the proximal neighborhood of the other set. We therefore conjecture that these results can be extended to two prox-regular sets. Finally we note that our use of hypomonotonicity defined in Lemma 3.1 might be relaxed to approximate monotonicity of the closed set C at a point $\overline{x} \in C$ defined by

$$\langle y_1 - y_2,\ x_1 - x_2 \rangle \geq -\sigma |x_1 - x_2| \quad \text{whenever } y_i \in N_C^r(x_i) \text{ and } x_i \in \mathcal{N}(\overline{x}).$$

for all $\sigma > 0$. By [35, Corollary4.11] this is equivalent to the super-regularity of C [35, Definition 4.4], a weaker condition than prox-regularity. This generalization would immediately extend our results to sets with other types of regularity such as subsmooth sets [1].

Acknowledgments The author would like to thank the anonymous reviewers for their thoughtful comments, particularly with regard to the interpretation of RAAR as a regularization rather than relaxation.

REFERENCES

[1] D. Aussel, A. Daniilidis, and L. Thibault, *Subsmooth sets: functional characterizations and related concepts*, Trans. Amer. Math. Soc., 357 (2004), pp. 1275–1301.
[2] H. H. Bauschke and J. M. Borwein, *On the convergence of von Neumann's alternating projection algorithm for two sets*, Set-Valued Anal., 1 (1993), pp. 185–212.
[3] ———, *Dykstra's alternating projection algorithm for two sets*, J. Approx. Theory, 79 (1994), pp. 418–443.
[4] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.
[5] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert space*, Math. Oper. Res., 26 (2001), pp. 248–264.
[6] 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 (2002), pp. 1334–45.
[7] ———, *A hybrid projection reflection method for phase retrieval*, J. Opt. Soc. Amer. A., 20 (2003), pp. 1025–34.
[8] ———, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*, J. Approx. Theory, 127 (2004), pp. 178–92.
[9] J. M. Borwein, S. Reich, and I. Shafrir, *Krasnoselski-Mann iterations in normed spaces*, Canad.Math.Bull., 35 (1992), pp. 21–28.
[10] H. Brézis and P.-L. Lions, *Produits infinis de resolvantes*, Isreal J. Math., 29 (1978), pp. 329–345.
[11] M. Brigonne, J. Coyle, and M. Piana, *Improving Born approximation by means of prior constraints.* Manuscript, 2005.
[12] M. Brigonne and M. Piana, *The use of constraints for solving inverse scattering problems: physical optics and the linear sampling method*, Inverse Problems, 21 (2005), pp. 207–222.
[13] J. V. Burke and D. R. Luke, *Variational analysis applied to the problem of optical phase retrieval*, SIAM J. Contr. Optim., 42 (2003), pp. 576–595.
[14] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory Algorithms and Applications*, Oxford University Press, 1997.
[15] R. Chapko and R. Kress, *A hybrid method for inverse boundary value problems in potential theory*, J. Inv. Ill-Posed Prob., 13 (2005), pp. 27–40.
[16] W. Cheney and A. A. Goldstein, *Proximity maps for convex sets*, Proc. Amer. Math. Soc., 10 (1959), pp. 448–450.
[17] F. H. Clarke, *Optimization and Nonsmooth Analysis*, vol. 5 of Classics in Applied Mathematics, SIAM, 1990.
[18] F. H. Clarke, R. J. Stern, Y. S. Ledyaev, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer Verlag, 1998.
[19] P. L. Combettes, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, P. W. Hawkes, ed., vol. 95, Academic Press, New York, 1996, pp. 155–270.
[20] ———, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., New York, 2001, Elsevier, pp. 115–152.
[21] ———, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504.

[22] P. L. Combettes and H. J. Trussell, *Method of successive projections for finding a common point of sets in metric spaces*, J. Opt. Theory and Appl., 67 (1990), pp. 487–507.

[23] F. Deutsch, *Best Approximation in Inner Product Spaces*, CMS Books in Mathematics, Springer, New York, 2001.

[24] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two or three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[25] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Prog., 55 (1992), pp. 293–318.

[26] H. Federer, *Curvature measures*, Trans. Amer. Math. Soc., 93 (1959), pp. 418–491.

[27] A. E. Frazho, K. M. Grigoriadis, and S. M. Kherat, *Alternating projection methods for mixed H^2 and H^∞ Nehari problems*, IEEE Trans. Automatic Control, 40 (1995), pp. 2127–2131.

[28] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, 1990.

[29] E. G. Gol'shtein and N. Tret'yakov, *Modified Lagrangians in convex programming and their generalizations*, Math. Prog. Study, 10 (1979), pp. 86–97.

[30] A. D. Ioffe, *Approximate subdifferentials and applications: II*, Mathematika, 33 (1986), pp. 111–128.

[31] ———, *Approximate subdifferentials and applications: III*, Mathematika, 36 (1989), pp. 1–38.

[32] ———, *Proximal analysis and approximate subdifferentials*, J. London Math Soc., 41 (1990), pp. 175–192.

[33] R. Kress and W. Rundell, *Nonlinear integral equations and the iterative solution for an inverse boundary value problem*, Inverse Problems, 21 (2005), pp. 1207–1223.

[34] R. Kress and P. Serranho, *A hybrid method for two-dimensional crack reconstruction*, Inverse Problems, 21 (2005), pp. 773–784.

[35] A. S. Lewis, D. R. Luke, and J. Malick, *Local linear convergence of alternating and averaged projections.* Preprint, July 2007.

[36] A. S. Lewis and J. Malick, *Alternating projections on manifolds.* Preprint, July 2006.

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

[38] D. R. Luke, *Relaxed averaged alternating reflections for diffraction imaging*, Inverse Problems, 21 (2005), pp. 37–50.

[39] D. R. Luke, J. V. Burke, and R. G. Lyon, *Optical wavefront reconstruction: Theory and numerical methods*, SIAM Rev., 44 (2002), pp. 169–224.

[40] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), Sér. R-3, pp. 154–158.

[41] ———, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox*, C. R. Acad. Sci. Paris Sér. A-B, 274 (1972), pp. A163–A165.

[42] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke. Math. J., 29 (1962), pp. 341–346.

[43] B. S. Mordukhovich and Y. Shao, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc., 328 (1996), pp. 1235–1280.

[44] J. J. Moreau, *Proximité et dualité dans un espace Hilbertian*, Bull. de la Soc. math. de France, 93 (1965), pp. 273–299.

[45] A. Moudafi, *An algorithmic approach to prox-regular variational inequalities*, Appl. Math. Comput., 155 (2004), pp. 845–852.

[46] G. Pierra, *Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique*, Lecture Notes in Computer Science, 41 (1976), pp. 200–218.

[47] R. A. Poliquin and R. T. Rockafellar, *Generalized Hessian properties of regularized nonsmooth functions*, SIAM J. Optim., 6 (1996), pp. 1121–1137.

[48] ———, *Prox-regular functions in variational analysis*, Trans. Amer. Math. Soc., 348 (1996), pp. 1805–1838.

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

[50] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Contr. Optim., 14 (1976), pp. 87–898.

[51] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[52] A. N. Tihonov, *On the regularization of ill-posed problems. (russian)*, Dokl. Akad. Nauk SSSR, 153 (1963), pp. 49–52.

[53] ———, *On the solution of ill-posed problems and the method of regularization (Russian)*, Dokl. Akad. Nauk SSSR, 151 (1963), pp. 501–504.

[54] J. von Neumann, *On rings of operators, reduction theory*, Ann. Math., 50 (1949), pp. 401–485.