

ALGORITHMS FOR STRUCTURED NONCONVEX OPTIMIZATION: THEORY AND PRACTICE

Dissertation
for the award of degree
“Doctor of Philosophy”
Ph.D. Division of Mathematics and Natural Sciences
of the Georg-August-Universität-Göttingen

within the doctoral program mathematics
of the Georg-August University School of Science (GAUSS)

submitted by
Hieu Thao Nguyen

from Ca Mau, Vietnam

Göttingen, 2018

Contents

1	Introduction and preliminary results	7
1.1	Introduction	7
1.2	Notation and basic definitions	8
1.3	Theory of pointwise almost averaging operators	14
2	Regularity theory	22
2.1	Elemental regularity of sets	22
2.2	Metric (sub)regularity of set-valued mappings	27
2.2.1	Primal characterizations	27
2.2.2	Dual characterizations	31
2.3	(Sub)transversality of collections of sets	33
2.3.1	Primal characterizations	33
2.3.2	(Sub)transversality versus metric (sub)regularity	39
2.3.3	Dual characterizations	41
2.3.4	Special cases: convex sets, cones and manifolds	51
3	Convergence analysis	53
3.1	Abstract convergence of Picard iterations	53
3.2	Cyclic projections	61
3.3	Alternating projections	72
3.4	Forward–backward algorithms	79
3.5	Douglas–Rachford algorithm and its relaxations	83
3.6	ADMM algorithms	87
4	Necessary conditions for convergence	88
4.1	Existence of implicit error bounds	88
4.2	Necessary conditions for linear convergence of alternating projections	96
4.3	Further discussion on convex alternating projections	99

5	Applications	103
5.1	Source location problem	103
5.1.1	Cyclic and averaged projections	104
5.1.2	Forward–backward algorithm and variants of the DR method	105
5.1.3	ADMM algorithm	105
5.1.4	Numerical simulation	106
5.2	Phase retrieval problem	108
5.2.1	Cyclic and averaged projections	109
5.2.2	Forward–backward algorithm and variants of the DR method	109
5.2.3	ADMM algorithm	110
5.2.4	Numerical simulation	111
6	Conclusion	114

Acknowledgements

This Ph.D. project has been conducted thanks to the scholarship from the German Israeli Foundation Grant G-1253-304.6 lead by Prof. Dr. Russell Luke and Prof. Dr. Marc Teboulle. I would like to thank the *Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen* for providing sufficient and adequate working conditions and support during my Ph.D. candidature. I would like to express my deepest gratitude to Prof. Dr. Russell Luke for his being a very helpful and highly encouraging advisor. I have also benefited from many academic discussions and collaborations with many colleagues and senior researchers in the field. In particular, I would like to thank my second advisor Prof. Dr. Thorsten Hohage, Prof. Dr. Alexander Kruger, Prof. Dr. Marc Teboulle, Dr. Mathew Tam and Dr. Shoham Sabach. My special thanks go to my working group at the *Institut für Numerische und Angewandte Mathematik* for their comradery during my time in Göttingen.

Abstract

We first synthesize and unify notions of regularity, both of individual functions/sets and of families of functions/sets, as they appear in the convergence theory of fixed point iterations. Several new primal and dual characterizations of regularity notions are presented with the focus on convergence analysis of numerical methods. A theory of almost averaged mappings is developed with a specialization to the projectors and reflectors associated with elemental regular sets.

Based on the knowledge of regularity notions, we develop a framework for quantitative convergence analysis of Picard iterations of expansive set-valued fixed point mappings. As application of the theory, we provide a number of results showing local convergence of nonconvex cyclic projections for both inconsistent and consistent feasibility problems, local convergence of the forward–backward algorithm for structured optimization without convexity, and local convergence of the Douglas–Rachford algorithm for structured nonconvex minimization. In particular, we establish a unified and weakest criterion for linear convergence of consistent alternating projections. As preparation for subsequent applications, we also discuss convergence of several relaxed versions of Douglas–Rachford algorithm and the alternating direction method of multipliers (ADMM).

Our development of regularity theory also sheds light on the relations between seemingly different ideas and point to possible necessary conditions for local linear convergence of fixed point iterations. We show that metric subregularity is necessary for linear monotonicity of fixed point iterations. This is specialized to an intensive discussion on subtransversality and alternating projections. In particular, we show that subtransversality is not only sufficient but also necessary for linear convergence of convex consistent alternating projections. More general results on gauge metric subregularity as necessary conditions for convergence are also discussed.

The algorithms together with their convergence theory are illustrated and simulated for the source location and phase retrieval problems.

Chapter 1

Introduction and preliminary results

1.1 Introduction

Convergence analysis has been one of the central and very active applications of variational analysis and mathematical optimization. Examples of recent contributions to the theory of the field that have initiated efficient programs of analysis are [4, 5, 26, 27, 97, 103]. It is the common recipe emphasized in these and many other works that there are two key ingredients required in order to derive convergence of a numerical method 1) regularity of the individual functions or sets such as *convexity* and *averagedness*, and 2) regularity of families of functions or sets at their critical points such as *transversality*, *Kurdyka-Łojasiewicz property* and *metric regularity*. The question of convergence for a given method can therefore be reduced to checking regularity properties of the problem data. There have been a considerable number of works studying the two ingredients of convergence analysis in order to provide sharper tools in various circumstances, especially in nonconvex cases, e.g., [20, 51, 59, 84, 83, 88, 90, 103, 118, 125]. The current thesis on “*Algorithms for structured nonconvex optimization: theory and practice*” consists of an investigation on this important and currently active research topic with application to source location and phase retrieval problems.

In Chapter 1, following this introductory section is an explanation of notation and basic definitions that will be used in the thesis.

Chapter 2 is devoted to a study of regularity theory with the emphasis on convergence analysis of numerical methods. This chapter consists of recent developments on 1) regularity of individual functions and sets, 2) theory of almost averaged mappings, 3) regularity of set-valued mappings and collections of sets, and 4) relationships amongst a range of regularity notions. Several new primal and dual characterizations of regularity notions are presented.

Chapter 3 is devoted to study convergence analysis of numerical algorithms based on the knowledge of regularity notions developed in Chapter 2. An abstract analysis program of Picard iterations of expansive set-valued fixed point mappings is established. As ap-

plications, we provide a number of results showing local convergence of nonconvex cyclic projections for feasibility, forward–backward algorithm for structured optimization, and Douglas–Rachford algorithm for structured minimization. In particular, we establish a unified and weakest criterion for linear convergence of consistent alternating projections. For subsequent applications, we also discuss convergence of several relaxed versions of Douglas–Rachford algorithm and the alternating direction method of multipliers.

Chapter 4 is devoted to discuss necessary conditions for linear convergence of fixed point iterations based on the knowledge of Chapters 2 and 3. This chapter consists of results on metric subregularity/error bounds for general fixed point iterations and an intensive specialization to subtransversality and the alternating projections method. In particular, we show that subtransversality is not only sufficient but also necessary for linear convergence of convex consistent alternating projections. More general results on nonlinear model of metric subregularity as necessary conditions for convergence are also discussed in this chapter.

Chapter 5 is devoted to application and numerical simulation. The source location and the phase retrieval problems are analyzed and simulated for the methods discussed in Chapter 3. Regularity properties from the problem data are discussed in accordance with the available convergence theory for each of the algorithms.

Most of the main results of the thesis can be found in [84, 83, 103, 102, 101] which are the joint research papers of the author with his advisor and collaborators during his Ph.D. candidature.

1.2 Notation and basic definitions

The underlying spaces will be specified in each section of this thesis. We use notation X, Y for general normed linear spaces, \mathcal{H} for infinite dimensional Hilbert spaces and \mathbb{E} for finite dimensional Euclidean spaces. For a normed linear space X , its topological dual is denoted X^* while $\langle \cdot, \cdot \rangle$ denotes the bilinear form defining the pairing between the two spaces. For Hilbert space \mathcal{H} , \mathcal{H}^* is identified with \mathcal{H} while $\langle \cdot, \cdot \rangle$ denotes the scalar product. Finite dimensional spaces are assumed equipped with the Euclidean norm. The notation $\| \cdot \|$ denotes the norm in the current setting. The open unit ball and the unit sphere are respectively denoted \mathbb{B} and \mathbb{S} while \mathbb{B}^* stands for the *closed* unit ball of the dual space X^* . $\mathbb{B}_\delta(x)$ stands for the open ball with radius $\delta > 0$ and center x . We denote the extended reals by $(-\infty, +\infty] := \mathbb{R} \cup \{+\infty\}$. The domain of a function $f : U \rightarrow (-\infty, +\infty]$ is defined by $\text{dom } f = \{u \in \mathbb{E} \mid f(u) < +\infty\}$. The (*Fréchet*) *subdifferential* of f at $\bar{x} \in \text{dom } f$ is defined by

$$\partial f(\bar{x}) := \left\{ v \mid \exists v^k \rightarrow v \text{ and } x^k \xrightarrow{f} \bar{x} \text{ such that } f(x) \geq f(x^k) + \langle v^k, x - x^k \rangle + o(\|x - x^k\|) \right\}. \quad (1.1)$$

Here the notation $x^k \xrightarrow{f} \bar{x}$ means that $x^k \rightarrow \bar{x} \in \text{dom } f$ and $f(x^k) \rightarrow f(\bar{x})$. When f is convex, (1.1) reduces to the usual convex subdifferential given by

$$\partial f(\bar{x}) := \{v \in U \mid \langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}), \text{ for all } x \in U\}.$$

When $\bar{x} \notin \text{dom } f$ the subdifferential is defined to be empty. Elements of the subdifferential are called *subgradients*.

A set-valued mapping T from X to another space Y is denoted $T : X \rightrightarrows Y$ and its *inverse* is given by

$$T^{-1}(y) := \{x \in X \mid y \in T(x)\}.$$

In the Hilbert space setting, a self mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is said to be *monotone* on $A \subset \mathcal{H}$ if

$$\forall x, y \in A \quad \inf_{\substack{x^+ \in T(x) \\ y^+ \in T(y)}} \langle x^+ - y^+, x - y \rangle \geq 0.$$

T is called *strongly monotone* on A if there exists a $\tau > 0$ such that

$$\forall x, y \in A \quad \inf_{\substack{x^+ \in T(x) \\ y^+ \in T(y)}} \langle x^+ - y^+, x - y \rangle \geq \tau \|x - y\|^2.$$

A *maximally monotone* mapping is one whose graph cannot be augmented by any more points without violating monotonicity. The subdifferential of a proper, lower semicontinuous (l.s.c.), convex function, for example, is a maximally monotone set-valued mapping [129, Theorem 12.17]. We denote the *resolvent* of T by $J_T := (\text{Id} + T)^{-1}$ where Id denotes the identity mapping. The corresponding *reflector* is defined by $R_T := 2J_T - \text{Id}$. A basic and fundamental fact is that the resolvent of a monotone mapping is firmly nonexpansive and hence single-valued [33, 105]. Of particular interest are *polyhedral* (or *piecewise polyhedral* [129]) mappings, that is, mappings $T : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ whose graph is the union of finitely many sets that are polyhedral convex in $\mathcal{H}_1 \times \mathcal{H}_2$ [50].

Notions of *continuity* of set-valued mappings have been thoroughly developed over the last 40 years. Readers are referred to the monographs [8, 50, 129] for basic results. A mapping $T : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ is said to be *Lipschitz continuous* if it is closed-valued and there exists a $\tau \geq 0$ such that, for all $u, u' \in \mathcal{H}_1$,

$$T(u') \subset T(u) + \tau \|u' - u\| \mathbb{B}.$$

Lipschitz continuity is, however, too strong a notion for set-valued mappings. We will mostly only require *calmness*, which is a pointwise version of Lipschitz continuity. A mapping $T : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ is said to be *calm* at \bar{u} for \bar{v} if $(\bar{u}, \bar{v}) \in \text{gph } T$ and there is a constant κ together with neighborhoods $U \times V$ of (\bar{u}, \bar{v}) such that

$$T(u) \cap V \subset T(\bar{u}) + \kappa \|u - \bar{u}\| \mathbb{B} \quad \forall u \in U.$$

When T is single-valued, calmness is just pointwise Lipschitz continuity:

$$\|T(u) - T(\bar{u})\| \leq \kappa \|u - \bar{u}\| \quad \forall u \in U.$$

The *graphical derivative* of a mapping $T : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ at a point $(x, y) \in \text{gph} T$ is denoted $DT(x|y) : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ and defined as the mapping whose graph is the tangent cone to $\text{gph} T$ at (x, y) (see [7] where it is called the contingent derivative). That is,

$$v \in DT(x|y)(u) \iff (u, v) \in \mathcal{T}_{\text{gph} T}(x, y) \quad (1.2)$$

where \mathcal{T}_A is the tangent cone mapping associated with the set A defined by

$$\mathcal{T}_A(\bar{x}) := \left\{ w \mid \frac{(x^k - \bar{x})}{\tau} \rightarrow w \quad \text{for some } x^k \xrightarrow{A} \bar{x}, \tau \searrow 0 \right\}.$$

Here the notation $x^k \xrightarrow{A} \bar{x}$ means that the sequence of points $\{x^k\}$ approaches \bar{x} from within A .

The distance to a set $A \subset \mathcal{H}$ with respect to the bivariate function $\text{dist}(\cdot, \cdot)$ is defined by

$$\text{dist}(\cdot, A) : \mathcal{H} \rightarrow \mathbb{R} : x \mapsto \inf_{y \in A} \text{dist}(x, y).$$

We use the convention that the distance to the empty set is $+\infty$. We use the *excess* to characterize the distance between two sets A and B

$$\text{excess}(A, B) := \sup\{\text{dist}(x, B) : x \in A\}$$

This is finite whenever B is nonempty and A is bounded and nonempty.

The set-valued mapping

$$P_A : \mathcal{H} \rightrightarrows \mathcal{H} : x \mapsto \{y \in A \mid \text{dist}(x, A) = \text{dist}(x, y)\}$$

is the *projector* on A . An element $y \in P_A(x)$ is called a *projection*. A *projection* is a selection from the projector. This exists for any closed nonempty set $A \subset \mathcal{H}$, as can be deduced by the continuity and coercivity of the norm. Note that the projector is not, in general, single-valued, and indeed uniqueness of the projector defines a type of regularity of the set A : local uniqueness characterizes *prox-regularity* [127] while in finite dimensional settings global uniqueness characterizes convexity [34].

Closely related to the projector is the *prox* mapping [111]

$$\text{prox}_{\lambda, f}(x) := \operatorname{argmin}_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}.$$

When $f(x) = \iota_A$, then $\text{prox}_{\lambda, \iota_A} = P_A$ for all $\lambda > 0$. The value function corresponding to the prox mapping is known as the *Moreau* envelope, which we denote by $e_{\lambda, f}(x) :=$

$\inf_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$. When $\lambda = 1$ and $f = \iota_A$ the Moreau envelope is just one-half the squared distance to the set A : $e_{1, \iota_A}(x) = \frac{1}{2} \text{dist}^2(x, A)$. The *inverse* projector P_A^{-1} is defined by

$$P_A^{-1}(y) := \{x \in \mathcal{H} \mid P_A(x) \ni y\}.$$

In the finite dimensional Euclidean space setting, we will assume the distance corresponds to the Euclidean norm unless otherwise specified. When $\text{dist}(x, y) = \|x - y\|$ then one has the following variational characterization of the projector: $\bar{z} \in P_A^{-1}\bar{x}$ if and only if

$$\langle \bar{z} - \bar{x}, x - \bar{x} \rangle \leq \frac{1}{2} \|x - \bar{x}\|^2 \quad \forall x \in A.$$

The *Fréchet normal cone* to $A \subset X$ at $\bar{x} \in A$ is defined

$$\widehat{N}_A(\bar{x}) := \left\{ v \mid \limsup_{x \xrightarrow{A} \bar{x}, x \neq \bar{x}} \frac{\langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0 \right\}. \quad (1.3)$$

The (*limiting*) *normal cone* to A at $\bar{x} \in A$, denoted $N_A(\bar{x})$, is defined as the limsup of the Fréchet normal cones. That is, a vector $v \in N_A(\bar{x})$ if there are sequences $x^k \xrightarrow{A} \bar{x}$, $v^k \rightarrow v$ with $v^k \in \widehat{N}_A(x^k)$. The *proximal normal cone* to A at \bar{x} is the set

$$N_A^{\text{prox}}(\bar{x}) := \text{cone}(P_A^{-1}\bar{x} - \bar{x}).$$

If $\bar{x} \notin A$, then all normal cones are defined to be empty.

The proximal normal cone need not be closed. The limiting normal cone is, of course, closed by definition. See [109, Definition 1.1] or [129, Definition 6.3] (where this is called the regular normal cone) for an in-depth treatment as well as [109, page 141] for historical notes. All these three sets are clearly cones. Unlike the first two cones, the third one can be nonconvex. It is easy to verify that $N_A^{\text{prox}}(\bar{x}) \subseteq \widehat{N}_A(\bar{x})$.

If $\bar{x} \in \text{bd } A$, then $N_A(\bar{x}) \neq \{0\}$. If A is a convex set, then all three cones $\widehat{N}_A(\bar{x})$, $N_A(\bar{x})$ and $N_A^{\text{prox}}(\bar{x})$ coincide and reduce to the normal cone in the sense of convex analysis:

$$N_A(\bar{x}) := \{v \in X \mid \langle v, x - \bar{x} \rangle \leq 0 \quad \text{for all } x \in A\}. \quad (1.4)$$

In the finite dimensional setting, when the projection is with respect to the Euclidean norm, the limiting normal cone can be written as the limsup of proximal normal cones:

$$N_A(\bar{x}) = \text{Lim sup}_{x \xrightarrow{A} \bar{x}} N_A^{\text{prox}}(x). \quad (1.5)$$

In differential geometry it is more common to work with the tangent space, but for smooth manifolds the normal cone (1.3) (the same as (1.5)) is a subspace and dual to the tangent space. Following Rockafellar and Wets [129, Example 6.8], we say that a subset $A \subset \mathbb{E}$ is a k -dimensional ($0 < k < n := \dim \mathbb{E}$) smooth manifold around a point $\bar{x} \in A$ if

there are a neighborhood U of \bar{x} in \mathbb{E} and a smooth (i.e., of \mathcal{C}^1 class) mapping $F : U \rightarrow \mathbb{R}^m$ ($m := n - k$) with $\nabla F(\bar{x})$ of full rank m such that $A \cap U = \{x \in U \mid F(x) = 0\}$. The *tangent space* to A at \bar{x} is a linear approximation of A near \bar{x} and is given by

$$T_A(\bar{x}) := \{x \in \mathbb{E} \mid \nabla F(\bar{x})x = 0\}.$$

The *normal space* to A at \bar{x} is defined as the orthogonal complement of $T_A(\bar{x})$ and can be written as

$$N_A(\bar{x}) := \{\nabla F(\bar{x})^*y \mid y \in \mathbb{R}^m\}. \quad (1.6)$$

It is in a sense a dual space object. If A is a smooth manifold, then cones (1.3), (1.5) and (1.4) reduce to the normal space (1.6).

Normal cones are central to characterizations both of the regularity of individual sets as well as of the regularity (transversality) of collections of sets. For collections of sets, when dealing with projection algorithms, it is important to account for the *relation* of the sets to each other and so the classical definitions of the normal cones above are too blunt for a refined numerical analysis. A typical situation: two nonempty sets A and B such that the affine span of $A \cup B$ is not equal to the whole space (e.g., two distinct intersecting lines in \mathbb{R}^3). One would expect all projections to lie in this affine span and the convergence to depend only on the mutual arrangement of the sets within the span. However, the normals (of any kind) to this affine span are also normals to the sets. They make a nontrivial subspace and this causes problems for the regularity conditions on collections of sets discussed below. In the context of algorithms, the only regularity conditions that are relevant are those that apply to the space where the iterates lie. In the case of algorithms like alternating projections, this is often an affine subspace of dimension smaller than the space in which the problem is formulated, as the example above illustrates. The essence of what we call “dual regularity conditions” consists in computing appropriate normal cones (limiting, Fréchet, or proximal) to each of the sets at the reference point (or nearby) and ensuring that the cones do not contain oppositely directed nonzero vectors. Such conditions are important for many applications including convergence analysis of projection algorithms.

For a subspace V of a Euclidean space \mathbb{E} , $V^\perp := \{u \in \mathbb{E} \mid \langle u, v \rangle = 0 \text{ for all } v \in V\}$ is the orthogonal complement subspace of V . For a real number α , $[\alpha]_+$ denotes $\max\{\alpha, 0\}$.

To quantify convergence of sequences and fixed point iterations, we encounter various forms of linear convergence listed next.

DEFINITION 1.2.1 (R- and Q-linear convergence to points, Chapter 9 of [120]). *Let $(x_k)_{k \in \mathbb{N}}$ be a sequence in X .*

- (i) $(x_k)_{k \in \mathbb{N}}$ is said to converge R-linearly to \tilde{x} with rate $c \in [0, 1)$ if there is a constant $\gamma > 0$ such that

$$\|x_k - \tilde{x}\| \leq \gamma c^k \quad \forall k \in \mathbb{N}.$$

(ii) $(x_k)_{k \in \mathbb{N}}$ is said to converge Q-linearly to \tilde{x} with rate $c \in [0, 1)$ if

$$\|x_{k+1} - \tilde{x}\| \leq c \|x_k - \tilde{x}\| \quad \forall k \in \mathbb{N}.$$

By definition, Q-linear convergence implies R-linear convergence with the same rate. Elementary examples show that the inverse implication does not hold in general.

One of the central concepts in the convergence of sequences is *Fejér monotonicity* [16, Definition 5.1]: a sequence $(x_k)_{k \in \mathbb{N}}$ is *Fejér monotone* with respect to a nonempty convex set A if

$$\|x_{k+1} - x\| \leq \|x_k - x\|, \quad \forall x \in A, \forall k \in \mathbb{N}.$$

In the context of convergence analysis of fixed point iterations, the following generalization of Fejér monotonicity of sequences is central.

DEFINITION 1.2.2 (μ -monotonicity). [101, Definition 2.2] Let $(x_k)_{k \in \mathbb{N}}$ be a sequence in X , $A \subset X$ be nonempty and $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfy $\mu(0) = 0$ and

$$\mu(t_1) < \mu(t_2) \leq t_2 \quad \text{whenever } 0 \leq t_1 < t_2.$$

(i) $(x_k)_{k \in \mathbb{N}}$ is said to be μ -monotone with respect to A if

$$\text{dist}(x_{k+1}, A) \leq \mu(\text{dist}(x_k, A)) \quad \forall k \in \mathbb{N}. \quad (1.7)$$

(ii) $(x_k)_{k \in \mathbb{N}}$ is said to be linearly monotone with respect to A if (1.7) is satisfied for $\mu(t) = c \cdot t$ for all $t \in \mathbb{R}_+$ and some constant $c \in [0, 1]$.

The next result is clear.

PROPOSITION 1.2.3 (Fejér monotonicity implies μ -monotonicity). [101, Proposition 2.3] If the sequence $(x_k)_{k \in \mathbb{N}}$ is Fejér monotone with respect to $A \subset X$, then it is μ -monotone with respect to A with $\mu = \text{Id}$.

The converse is not true, as the next example shows.

EXAMPLE 1.2.4 (μ -monotonicity is not Fejér monotonicity). [101, Example 2.4] Let $A := \{(x, y) \in \mathbb{R}^2 \mid y \leq 0\}$ and consider the sequence $x_k := (1/2^k, 1/2^k)$ for all $k \in \mathbb{N}$. This sequence is linearly monotone with respect to A with constant $c = 1/2$, but not Fejér monotone since $\|x_{k+1} - (2, 0)\| > \|x_k - (2, 0)\|$ for all k .

The next definition will come into play in Sections 4.2 and 4.3. It provides a way to analyze fixed point iterations which, like the classical example of alternating projections, are compositions of mappings.

The subset Λ appearing in Definition 1.2.5 and throughout this thesis is always assumed to be closed and nonempty. We use this set to isolate specific elements of the fixed point set (most often restricted to affine subspaces). This is more than just a formal generalization since in some concrete situations the required assumptions do not hold on X but they do hold on relevant subsets.

DEFINITION 1.2.5 (linearly extendible sequences). [101, Definition 2.5] A sequence $(x_k)_{k \in \mathbb{N}}$ on $\Lambda \subset X$ is said to be linearly extendible on Λ with frequency $m \geq 1$ ($m \in \mathbb{N}$ is fixed) and rate $c \in [0, 1)$ if there is a sequence $(z_k)_{k \in \mathbb{N}}$ on Λ such that $x_k = z_{mk}$ for all $k \in \mathbb{N}$ and the following conditions are satisfied for all $k \in \mathbb{N}$:

$$\begin{aligned} \|z_{k+2} - z_{k+1}\| &\leq \|z_{k+1} - z_k\|, \\ \|z_{m(k+1)+1} - z_{m(k+1)}\| &\leq c\|z_{mk+1} - z_{mk}\|. \end{aligned}$$

When $\Lambda = X$, the quantifier “on Λ ” is dropped.

The requirement on the linear extension sequence $(z_k)_{k \in \mathbb{N}}$ means that the sequence of the distances between its two consecutive iterates is uniformly non-increasing and possesses a subsequence of type $(\|z_{mk+1} - z_{mk}\|)_{k \in \mathbb{N}}$ that converges Q-linearly with a global rate to zero.

The extension of sequences of fixed point iterations $(x_k)_{k \in \mathbb{N}}$ will most often be to the intermediate points generated by the composite mappings. In the case of alternating projections this is $z_{2k} := x_k \in P_A P_B x_{k-1}$, and $z_{2k+1} \in P_B z_{2k}$. This strategy of analyzing alternating projections by keeping track of the intermediate projections has been exploited to great effect in [20, 51, 90, 91, 118, 103]. From the Cauchy property of $(z_k)_{k \in \mathbb{N}}$, one can deduce R-linear convergence from linear extendability.

PROPOSITION 1.2.6 (linear extendability implies R-linear convergence). [101, Proposition 2.6] If the sequence $(x_k)_{k \in \mathbb{N}}$ on $\Lambda \subset X$ is linearly extendible on Λ with some frequency $m \geq 1$ and rate $c \in [0, 1)$, then $(x_k)_{k \in \mathbb{N}}$ converges R-linearly to a point $\tilde{x} \in \Lambda$ with rate c .

For ease of exposition, in most of discussion of collections of sets we consider the case of two nonempty subsets $A, B \subset X$ except in Subsection 3.2 where the most general convergence result for cyclic projections is of target. The analogous extension of most of the results to the case of any finite collection of n sets ($n > 2$) does not require much effort (cf. [78, 79, 80, 85, 87]).

Recall that a Banach space is *Asplund* if the dual of each its separable subspace is separable; see, e.g., [30, 109] for discussions and characterizations of Asplund spaces. All reflexive, in particular, all finite dimensional Banach spaces are Asplund.

A function $\mu : [0, \infty) \rightarrow [0, \infty)$ is a *gauge function* if μ is continuous and strictly increasing with $\mu(0) = 0$ and $\lim_{t \rightarrow \infty} \mu(t) = \infty$.

1.3 Theory of pointwise almost averaging operators

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} . The content of this section is taken from our joint work with Dr. Matthew K. Tam [103].

We first clarify what is meant by a fixed point of a set-valued mapping.

DEFINITION 1.3.1 (fixed points of set-valued mappings). [103, Definition 2.1] *The set of fixed points of a set-valued mapping $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is defined by*

$$\text{Fix } T := \{x \in \mathbb{E} \mid x \in T(x)\}.$$

In the set-valued setting, it is important to keep in mind a few things that can happen that cannot happen when the mapping is single-valued.

EXAMPLE 1.3.2 (inhomogeneous fixed point sets). [103, Example 2.1] *Let $T := P_A P_B$ where*

$$\begin{aligned} A &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq -2x_1 + 3\} \cap \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq 1\}, \\ B &= \mathbb{R}^2 \setminus (0, +\infty)^2. \end{aligned}$$

Here $P_B(1, 1) = \{(0, 1), (1, 0)\}$ and the point $(1, 1)$ is a fixed point of T since $(1, 1) \in P_A \{(0, 1), (1, 0)\}$. However, the point $P_A(0, 1)$ is also in $T(1, 1)$, and this is not a fixed point of T .

To help rule out inhomogeneous fixed point sets like the one in the previous example, we introduce the following strong calmness of fixed point mappings that is an extension of conventional nonexpansiveness and firm nonexpansiveness. What we call *almost nonexpansive mappings* below were called (S, ϵ) -nonexpansive mappings in [59, Definition 2.3], and almost averaged mappings are slight generalization of (S, ϵ) -firmly nonexpansive mappings also defined there.

DEFINITION 1.3.3 (almost nonexpansive/averaged mappings). [103, Definition 2.2] *Let D be a nonempty subset of \mathbb{E} and let T be a (set-valued) mapping from D to \mathbb{E} .*

- (i) *T is said to be pointwise almost nonexpansive on D at $y \in D$ if there exists a constant $\epsilon \in [0, 1)$ such that*

$$\begin{aligned} \|x^+ - y^+\| &\leq \sqrt{1 + \epsilon} \|x - y\| \\ \forall y^+ \in Ty \text{ and } \forall x^+ \in Tx \text{ whenever } x \in D. \end{aligned} \tag{1.8}$$

If (1.8) holds with $\epsilon = 0$ then T is called pointwise nonexpansive at y on D .

If T is pointwise (almost) nonexpansive at every point on a neighborhood of y (with the same violation constant ϵ) on D , then T is said to be (almost) nonexpansive at y (with violation ϵ) on D .

If T is pointwise (almost) nonexpansive on D at every point $y \in D$ (with the same violation constant ϵ), then T is said to be pointwise (almost) nonexpansive on D (with violation ϵ). If D is open and T is pointwise (almost) nonexpansive on D , then it is (almost) nonexpansive on D .

- (ii) T is called pointwise almost averaged on D at $y \in D$ if there is an averaging constant $\alpha \in (0, 1)$ and a violation constant $\varepsilon \in [0, 1)$ such that the mapping \tilde{T} defined by

$$T = (1 - \alpha)\text{Id} + \alpha\tilde{T}$$

is pointwise almost nonexpansive at y with violation ε/α on D .

Likewise if \tilde{T} is (pointwise) (almost) nonexpansive on D (at y) (with violation ε), then T is said to be (pointwise) (almost) averaged on D (at y) (with averaging constant α and violation $\alpha\varepsilon$).

If the averaging constant $\alpha = 1/2$, then T is said to be (pointwise) (almost) firmly nonexpansive on D (with violation ε) (at y).

Note that the mapping T need not be a self-mapping from D to itself. In the special case where T is (firmly) nonexpansive at all points $y \in \text{Fix } T$, mappings satisfying (1.8) are also called *quasi-(firmly)nonexpansive* [16].

The term ‘‘almost nonexpansive’’ has been used for different purposes by Nussbaum [119] and Rouhani [130]. Rouhani uses the term to indicate sequences, in the Hilbert space setting, that are asymptotically nonexpansive. Nussbaum’s definition is the closest in spirit and definition to ours, except that he defines f to be locally almost nonexpansive when $\|f(y) - f(x)\| \leq \|y - x\| + \varepsilon$. In this context, see also [128]. At the risk of some confusion, we re-purpose the term here. Our definition of pointwise almost nonexpansiveness of T at \bar{x} is stronger than *calmness* [129, Chapter 8.F] with constant $\lambda = \sqrt{1 + \varepsilon}$ since the inequality must hold for all pairs $x^+ \in Tx$ and $y^+ \in Ty$, while for calmness the inequality would hold only for points $x^+ \in Tx$ and their *projections* onto Ty . We have avoided the temptation to call this property ‘‘strong calmness’’ in order to make clearer the connection to the classical notions of (firm) nonexpansiveness. A theory based only on calm mappings, what one might call ‘‘weakly almost averaged/nonexpansive’’ operators is possible and would yield statements about the existence of convergent *selections* from sequences of iterated set-valued mappings. In light of the other requirement of the mapping T that we will explore in Section 2.2, namely metric subregularity, this would illuminate an aesthetically pleasing and fundamental symmetry between requirements on T and its inverse. We leave this avenue of investigation open. Our development of the properties of almost averaged operators parallels the treatment of averaged operators in [16].

PROPOSITION 1.3.4 (characterizations of almost averaged operators). [103, Proposition 2.1] Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$, $U \subset \mathbb{E}$ and let $\alpha \in (0, 1)$. The following are equivalent.

- (i) T is pointwise almost averaged at y on U with violation ε and averaging constant α .
- (ii) $(1 - \frac{1}{\alpha})\text{Id} + \frac{1}{\alpha}T$ is pointwise almost nonexpansive at y on $U \subset \mathbb{E}$ with violation ε/α .
- (iii) For all $x \in U$, $x^+ \in T(x)$ and $y^+ \in T(y)$ it holds that

$$\|x^+ - y^+\|^2 \leq (1 + \varepsilon)\|x - y\|^2 - \frac{1 - \alpha}{\alpha} \|(x - x^+) - (y - y^+)\|^2.$$

Consequently, if T is pointwise almost averaged at y on U with violation ε and averaging constant α then T is pointwise almost nonexpansive at y on U with violation at most ε .

Proposition 1.3.4 is a slight extension of [16, Proposition 4.25].

EXAMPLE 1.3.5 (alternating projections). [103, Example 2.2] Let $T := P_A P_B$ for the closed sets A and B defined below.

(i) If A and B are convex, then T is nonexpansive and averaged (i.e. pointwise everywhere, no violation).

(ii) Let

$$\begin{aligned} A &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, -1/2x_1 \leq x_2 \leq x_1, x_1 \geq 0\} \subset \mathbb{R}^2 \\ B &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, x_1 \leq |x_2|\} \subset \mathbb{R}^2, \\ \bar{x} &= (0, 0). \end{aligned}$$

The mapping T is not almost nonexpansive on any neighborhood for any finite violation at $y = (0, 0) \in \text{Fix } T$, but it is pointwise nonexpansive (no violation) at $y = (0, 0)$ and nonexpansive at all $y \in (A \cap B) \setminus \{(0, 0)\}$ on small enough neighborhoods of these points.

(iii) T is pointwise averaged at $(1, 1)$ when

$$\begin{aligned} A &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \leq 2x_1 - 1\} \cap \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq \frac{1}{2}x_1 + \frac{1}{2}\} \\ B &= \mathbb{R}^2 \setminus \mathbb{R}^2_{++}. \end{aligned}$$

This illustrates that whether or not A and B have points in common is not relevant to the property.

(iv) T is not pointwise almost averaged at $(1, 1)$ for any $\varepsilon > 0$ when

$$\begin{aligned} A &= \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq -2x_1 + 3\} \cap \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq 1\} \\ B &= \mathbb{R}^2 \setminus \mathbb{R}^2_{++} \end{aligned}$$

In light of Example 1.3.2, this shows that the pointwise almost averaged property is incompatible with inhomogeneous fixed points (see Proposition 1.3.6).

PROPOSITION 1.3.6 (pointwise single-valuedness). [103, Proposition 2.2] If $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is pointwise almost nonexpansive on $D \subseteq \mathbb{E}$ at $\bar{x} \in D$ with violation $\varepsilon \geq 0$, then T is single-valued at \bar{x} . In particular, if $\bar{x} \in \text{Fix } T$ (that is $\bar{x} \in T\bar{x}$) then $T\bar{x} = \{\bar{x}\}$.

EXAMPLE 1.3.7 (pointwise almost nonexpansive mappings not single-valued). [103, Example 2.3] Although a pointwise almost nonexpansive mapping is single-valued at the reference

point, it need not be single-valued on neighborhoods of the reference points. Consider, for example, the coordinate axes in \mathbb{R}^2 ,

$$A = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}.$$

The metric projector P_A is single-valued and even pointwise nonexpansive (no almost) at every point in A , but multivalued on $L := \{(x, y) \in \mathbb{R}^2 \setminus \{0\} \mid |x| = |y|\}$.

Almost firmly nonexpansive mappings have particularly convenient characterizations. In our development below and thereafter we use the set S to denote the collection of points at which the property holds. This is useful for distinguishing points where the regularity holds. In Section 3.1, the set S is used to isolate a subset of fixed points. The idea here is that the properties required for quantifying convergence need not hold on the space where a problem is formulated, but may only hold on a subset of this space where the iterates of a particular algorithm may be, naturally, confined. This is used in [4] to achieve linear convergence results for the alternating directions method of multipliers algorithm. Alternatively, S can also include points that are not fixed points of constituent operators in an algorithm, but are closely related to fixed points. One example of this is *local best approximation points*, that is, points in one set that are locally nearest to another. In Section 2.1 we will need to quantify the violation of the averaging property for a projector onto a nonconvex set A at points in another set, say B , that are locally *nearest points* to A . This will allow us to tackle *inconsistent feasibility* where the alternating projections iteration converges not to the intersection, but to local best approximation points.

PROPOSITION 1.3.8 (almost firmly nonexpansive mappings). [103, Proposition 2.3] Let $S \subset U \subset \mathbb{E}$ be nonempty and $T : U \rightrightarrows \mathbb{E}$. The following are equivalent.

(i) T is pointwise almost firmly nonexpansive on U at all $y \in S$ with violation ε .

(ii) The mapping $\tilde{T} : U \rightrightarrows \mathbb{E}$ given by

$$\tilde{T}x := (2Tx - x) \quad \forall x \in U$$

is pointwise almost nonexpansive on U at all $y \in S$ with violation 2ε .

(iii) $\|x^+ - y^+\|^2 \leq \frac{\varepsilon}{2} \|x - y\|^2 + \langle x^+ - y^+, x - y \rangle$ for all $x^+ \in Tx$, and all $y^+ \in Ty$ at each $y \in S$ whenever $x \in U$.

(iv) Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$ be a mapping whose resolvent is T , i.e., $T = (\text{Id} + F)^{-1}$. At each $x \in U$, for all $u \in Tx$, $y \in S$ and $v \in Ty$, the points (u, z) and (v, w) are in $\text{gph } F$ where $z = x - u$ and $w = y - v$, and satisfy

$$-\frac{\varepsilon}{2} \|(u + z) - (v + w)\|^2 \leq \langle z - w, u - v \rangle.$$

Property (iv) of Proposition 1.3.8 characterizes a type of nonmonotonicity of the mapping F on D with respect to S ; for lack of a better terminology we call this *Type-I nonmonotonicity*. It can be shown that, for small enough parameter values, this is a generalization of another well-established property known as *hypomonotonicity* [127]. In [42] the notion of *submonotonicity* proposed by Spingarn [132] in relation to *approximate convexity* [115] was studied. Their relation to the definition below is the topic of future research.

DEFINITION 1.3.9 (nonmonotone mappings). [103, Definition 2.3]

- (a) A mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is pointwise Type-I nonmonotone at \bar{v} if there is a constant τ together with a neighborhood U of \bar{v} such that

$$-\tau \|(u+z) - (\bar{v}+w)\|^2 \leq \langle z-w, u-\bar{v} \rangle \quad \forall z \in Fu, \forall u \in U, \forall w \in F\bar{v}. \quad (1.9)$$

The mapping F is said to be Type-I nonmonotone on U if (1.9) holds for all \bar{v} on U .

- (b) The mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is said to be pointwise hypomonotone at \bar{v} with constant τ on U if

$$-\tau \|u - \bar{v}\|^2 \leq \langle z - w, u - \bar{v} \rangle \quad \forall z \in Fu, \forall u \in U, \forall w \in F\bar{v}. \quad (1.10)$$

If (1.10) holds for all $\bar{v} \in U$ then F is said to be hypomonotone with constant τ on U .

In the event that T is in fact firmly nonexpansive (that is, $S = D$ and $\tau = 0$) then Proposition 1.3.8(iv) just establishes the well known equivalence between monotonicity of a mapping and firm nonexpansiveness of its resolvent [105]. Moreover, if a single-valued mapping $f : \mathbb{E} \rightarrow \mathbb{E}$ is calm at \bar{v} with calmness modulus L , then it is pointwise hypomonotone at \bar{v} with violation at most L . Indeed,

$$\langle u - \bar{v}, f(u) - f(\bar{v}) \rangle \geq -\|u - \bar{v}\| \|f(u) - f(\bar{v})\| \geq -L \|u - \bar{v}\|^2.$$

This also points to a relationship to *cohypomonotonicity* developed in [41]. More recently the notion of pointwise quadratically supportable functions was introduced [100, Definition 2.1]; for smooth functions, this class – which is not limited to convex functions – was shown to include functions whose gradients are pointwise strongly monotone (pointwise hypomonotone with constant $\tau < 0$) [100, Proposition 2.2].

The next result shows the inheritance of the averaging property under compositions and averages of averaged mappings.

PROPOSITION 1.3.10 (compositions and averages of relatively averaged operators). [103, Proposition 2.4] Let $T_j : \mathbb{E} \rightrightarrows \mathbb{E}$ for $j = 1, 2, \dots, m$ be pointwise almost averaged on U_j at all $y_j \in S_j \subset \mathbb{E}$ with violation ε_j and averaging constant $\alpha_j \in (0, 1)$ where $U_j \supset S_j$ for $j = 1, 2, \dots, m$.

- (i) If $U := U_1 = U_2 = \dots = U_m$ and $S := S_1 = S_2 = \dots = S_m$ then the weighted mapping $T := \sum_{j=1}^m w_j T_j$ with weights $w_j \in [0, 1]$, $\sum_{j=1}^m w_j = 1$, is pointwise almost averaged at all $y \in S$ with violation $\varepsilon = \sum_{j=1}^m w_j \varepsilon_j$ and averaging constant $\alpha = \max_{j=1,2,\dots,m} \{\alpha_j\}$ on U .
- (ii) If $T_j U_j \subseteq U_{j-1}$ and $T_j S_j \subseteq S_{j-1}$ for $j = 2, 3, \dots, m$, then the composite mapping $T := T_1 \circ T_2 \circ \dots \circ T_m$ is pointwise almost nonexpansive at all $y \in S_m$ on U_m with violation at most

$$\varepsilon = \prod_{j=1}^m (1 + \varepsilon_j) - 1. \quad (1.11)$$

- (iii) If $T_j U_j \subseteq U_{j-1}$ and $T_j S_j \subseteq S_{j-1}$ for $j = 2, 3, \dots, m$, then the composite mapping $T := T_1 \circ T_2 \circ \dots \circ T_m$ is pointwise almost averaged at all $y \in S_m$ on U_m with violation at most ε given by (1.11) and averaging constant at least

$$\alpha = \frac{m}{m - 1 + \frac{1}{\max_{j=1,2,\dots,m} \{\alpha_j\}}}.$$

REMARK 1.3.11. [103, Remark 2.1] We remark that Proposition 1.3.10(ii) holds in the case when T_j ($j = 1, 2, \dots, m$) are merely pointwise almost nonexpansive. The counterpart for T_j ($j = 1, \dots, m$) pointwise almost nonexpansive to Proposition 1.3.10(i) is given by allowing $\alpha = 0$.

COROLLARY 1.3.12 (Krasnoselski–Mann relaxations). [103, Corollary 2.1] Let $\lambda \in [0, 1]$ and define $T_\lambda := (1 - \lambda) \text{Id} + \lambda T$ for T pointwise almost averaged at y with violation ε and averaging constant α on U . Then T_λ is pointwise almost averaged at y with violation $\lambda \varepsilon$ and averaging constant α on U . In particular, when $\lambda = 1/2$ the mapping $T_{1/2}$ is pointwise almost firmly nonexpansive at y with violation $\varepsilon/2$ on U .

A particularly attractive consequence of Corollary 1.3.12 is that the violation of almost averaged mappings can be mitigated by taking smaller steps via Krasnoselski–Mann relaxation.

To conclude this section we prove the following lemma, a special case of which will be required in Section 3.2, which relates the fixed point set of the composition of pointwise almost averaged operators to the corresponding *difference vector*.

DEFINITION 1.3.13 (difference vectors of composite mappings). [103, Definition 2.4] For a collection of operators $T_j : \mathbb{E} \rightrightarrows \mathbb{E}$ ($j = 1, 2, \dots, m$) and $T := T_1 \circ T_2 \circ \dots \circ T_m$ the set of difference vectors of T at u is given by the mapping $\mathcal{Z} : \mathbb{E} \rightrightarrows \mathbb{E}^m$ defined by

$$\mathcal{Z}(u) := \{ \zeta := z - \Pi z \mid z \in W_0 \subset \mathbb{E}^m, z_1 = u \},$$

where

$$\Pi : (x_1, x_2, \dots, x_m) \mapsto (x_2, \dots, x_m, x_1) \quad \forall (x_1, x_2, \dots, x_m) \in \mathbb{E}^m \quad (1.12)$$

is the permutation mapping and

$$W_0 := \{x = (x_1, \dots, x_m) \in \mathbb{E}^m \mid x_m \in T_m x_1, x_j \in T_j(x_{j+1}), j = 1, 2, \dots, m-1\}.$$

LEMMA 1.3.14 (difference vectors of averaged compositions). [103, Lemma 2.1] Given a collection of operators $T_j : \mathbb{E} \rightrightarrows \mathbb{E}$ ($j = 1, 2, \dots, m$), set $T := T_1 \circ T_2 \circ \dots \circ T_m$. Let $S_0 \subset \text{Fix } T$, let U_0 be a neighborhood of S_0 and define $U := \{z = (z_1, z_2, \dots, z_m) \in W_0 \mid z_1 \in U_0\}$. Fix $\bar{u} \in S_0$ and the difference vector $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ with $\bar{\zeta} = \bar{z} - \Pi \bar{z}$ for the point $\bar{z} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_m) \in W_0$ having $\bar{z}_1 = \bar{u}$. Let T_j be pointwise almost averaged at \bar{z}_j with violation ε_j and averaging constant α_j on $U_j := p_j(U)$ where $p_j : \mathbb{E}^m \rightarrow \mathbb{E}$ denotes the j th coordinate projection operator ($j = 1, 2, \dots, m$). Then, for $u \in S_0$ and $\zeta \in \mathcal{Z}(u)$ with $\zeta = z - \Pi z$ for $z = (z_1, z_2, \dots, z_m) \in W_0$ having $z_1 = u$,

$$\frac{1 - \alpha}{\alpha} \|\bar{\zeta} - \zeta\|^2 \leq \sum_{j=1}^m \varepsilon_j \|\bar{z}_j - z_j\|^2 \text{ where } \alpha = \max_{j=1,2,\dots,m} \alpha_j.$$

If the mapping T_j is in fact pointwise averaged at \bar{z}_j on U_j ($j = 1, 2, \dots, m$), then the set of difference vectors of T is a singleton and independent of the initial point; that is, there exists $\bar{\zeta} \in \mathbb{E}^m$ such that $\mathcal{Z}(u) = \{\bar{\zeta}\}$ for all $u \in S_0$.

Chapter 2

Regularity theory

In the last decade there has been a great deal of interest in extending the classical notions of regularity to include nonconvex and nonsmooth sets, motivated to a large extent by nonsmooth and nonconvex optimization and attendant subdifferential and coderivative calculus, optimality and stationarity conditions and convergence analysis of algorithms.

On the one hand convergence analysis has clearly served as a main motivator for the regularity theory, but on the other hand these regularity properties, which are amongst the corner stones of variational analysis and mathematical optimization, are themselves of importance. In fact, investigations of these regularity properties have led to many fundamental ideas and important applications in variational analysis and optimization, e.g., [73, 79, 87].

2.1 Elemental regularity of sets

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} . The content of this section is taken from our joint papers with Prof. Alexander Y. Kruger [84] and Dr. Matthew K. Tam [103].

This section discusses a general framework for elemental regularity of sets that provides a common language for the many different definitions that have appeared to date. This new framework makes the cascade of implications between the different types of regularity more transparent, namely that convexity \implies prox-regularity \implies super-regularity \implies Clarke regularity \implies (ε, δ) -regularity \implies (ε, δ) -subregularity \implies σ -Hölder regularity see Theorem 2.1.4.

We first recall these widely known regularity notions of individual sets.

DEFINITION 2.1.1 (regularity notions of sets). *Let $A \subset \mathbb{E}$ be closed and nonempty.*

(i) *A is convex if it holds that $tx + (1 - t)y \in A$ for all $t \in [0, 1]$ whenever $x, y \in A$.*

- (ii) [127] A is prox-regular at $\bar{x} \in A$ if the projector P_A is single-valued around \bar{x} .
- (iii) [90, Definition 4.3] A is super-regular at $\bar{x} \in A$ if for every $\varepsilon > 0$, there exists a number $\delta > 0$ such that

$$\langle x - z, y - z \rangle \leq \varepsilon \|x - z\| \|y - z\| \quad \forall x \in A \cap \mathbb{B}_\delta(\bar{x}), y \in \mathbb{B}_\delta(\bar{x}), z \in P_A(y).$$

- (iv) [129, Definition 6.4] A is Clarke regular at $\bar{x} \in A$ if every (limiting) normal vector to A at \bar{x} is a Fréchet normal vector, i.e., $\widehat{N}_A(\bar{x}) = N_A(\bar{x})$.
- (v) [21, Definition 8.1] Let $\varepsilon, \delta > 0$. A is (ε, δ) -regular at $\bar{x} \in A$ if

$$\langle u, x - z \rangle \leq \varepsilon \|u\| \|x - z\| \quad \forall x, z \in A \cap \mathbb{B}_\delta(\bar{x}), u \in N_A^{\text{prox}}(z).$$

- (vi) [59, Definition 2.9] Let $B \subset \mathbb{E}$ and $\varepsilon, \delta > 0$. A is (ε, δ) -subregular at $\bar{x} \in A$ relative to B if

$$\langle u, x - z \rangle \leq \varepsilon \|u\| \|x - z\| \quad \forall z \in A \cap \mathbb{B}_\delta(\bar{x}), x \in B \cap \mathbb{B}_\delta(\bar{x}), u \in N_A^{\text{prox}}(z).$$

- (vii) [118, Definition 2] Let $B \subset \mathbb{E}$ and $\sigma \in [0, 1)$. A is σ -Hölder regular at $\bar{x} \in A$ relative to B with neighborhood U and constant $\gamma \in [0, 1)$ if for every $b \in B \cap U$ and $a^+ \in P_A(b) \cap U$, it holds that

$$A \cap \mathbb{B}_{(1+\gamma^2)\|b-a^+\|}(b) \cap \{a \in P_B^{-1}(b) : \langle b - a^+, a - a^+ \rangle > \gamma \|b - a^+\|^{\sigma+1} \|a - a^+\|\} = \emptyset.$$

The following concept of *elemental regularity* places under one schema the many different kinds of set regularity appearing in Definition 2.1.1.

DEFINITION 2.1.2 (elemental regularity of sets). [84, Definition 5] Let $A \subset \mathbb{E}$ be nonempty and let $(\bar{y}, \bar{v}) \in \text{gph}(N_A)$.

- (i) A is elementally subregular of order σ relative to Λ at \bar{x} for (\bar{y}, \bar{v}) with constant ε if there exists a neighborhood U of \bar{x} such that

$$\langle \bar{v} - (x - x^+), x^+ - \bar{y} \rangle \leq \varepsilon \|\bar{v} - (x - x^+)\|^{1+\sigma} \|x^+ - \bar{y}\|, \quad \forall x \in \Lambda \cap U, x^+ \in P_A(x). \quad (2.1)$$

- (ii) The set A is said to be uniformly elementally subregular of order σ relative to Λ at \bar{x} for (\bar{y}, \bar{v}) if for any $\varepsilon > 0$ there is a neighborhood U (depending on ε) of \bar{x} such that (2.1) holds.

- (iii) The set A is said to be elementally regular of order σ at \bar{x} for (\bar{y}, \bar{v}) with constant ε if it is elementally subregular of order σ relative to $\Lambda = A$ at \bar{x} for all (\bar{y}, \bar{v}) with constant ε where $v \in N_A(\bar{y}) \cap V$ for some neighborhood V of \bar{v} .

- (iv) The set A is said to be uniformly elementally regular of order σ at \bar{x} for (\bar{y}, \bar{v}) if it is uniformly elementally subregular of order σ relative to $\Lambda = A$ at \bar{x} for all (\bar{y}, v) where $v \in N_A(\bar{y}) \cap V$ for some neighborhood V of \bar{v} .

If $\Lambda = \{\bar{x}\}$ in (i) or (ii), then the respective qualifier “relative to” is dropped. If $\sigma = 0$, then the respective qualifier “of order” is dropped in the description of the properties. The modulus of elemental (sub)regularity is the infimum over all ε for which (2.1) holds.

In all properties in Definition 2.1.2, \bar{x} need not be in Λ and \bar{y} need not be in either U or Λ . In case of order $\sigma = 0$, the properties are trivial for any constant $\varepsilon \geq 1$. When saying a set is not elementally (sub)regular but without specifying a constant, it is meant for any constant $\varepsilon < 1$.

EXAMPLE 2.1.3. [84, Example 2]

- (a) (cross) Consider the set

$$A = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}.$$

This example is of particular interest for the study of sparsity constrained optimization. A is elementally regular at any $\bar{x} \neq (0, 0)$, say $\|\bar{x}\| > \delta > 0$, for all $(a, v) \in \text{gph } N_A$ where $a \in \mathbb{B}_\delta(\bar{x})$ with constant $\varepsilon = 0$ and neighborhood $\mathbb{B}_\delta(\bar{x})$. The set A is not elementally regular at the point $\bar{x} = (0, 0)$ for any $((0, 0), v) \in \text{gph } N_A$ since $N_A(0, 0) = A$. However, A is elementally subregular at the point $\bar{x} = (0, 0)$ for all $(a, v) \in \text{gph } N_A$ with constant $\varepsilon = 0$ and neighborhood \mathbb{E} since all vectors $a \in A$ are orthogonal to $N_A(a)$.

- (b) (circle) The circle is central to the phase retrieval problem,

$$A = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}.$$

The set A is uniformly elementally regular at any $\bar{x} \in A$ for all $(\bar{x}, v) \in \text{gph } N_A$. Indeed, note first that for any $\bar{x} \in A$, $N_A(\bar{x})$ consists of the line passing through the origin and \bar{x} . Now, for any $\varepsilon \in (0, 1)$, we choose $\delta = \varepsilon$. Then for any $x \in A \cap \mathbb{B}_\delta(\bar{x})$, it holds $\cos \angle(-\bar{x}, x - \bar{x}) \leq \delta = \varepsilon$. Hence, for all $x \in A \cap \mathbb{B}_\delta(\bar{x})$ and $v \in N_A(\bar{x})$,

$$\langle v, x - \bar{x} \rangle = \cos \angle(v, x - \bar{x}) \|v\| \|x - \bar{x}\| \leq \cos \angle(-\bar{x}, x - \bar{x}) \|v\| \|x - \bar{x}\| \leq \varepsilon \|v\| \|x - \bar{x}\|.$$

- (c) Let us consider

$$A = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, -1/2x_1 \leq x_2 \leq x_1, x_1 \geq 0\} \subset \mathbb{R}^2,$$

$$B = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, x_1 \leq |x_2|\} \subset \mathbb{R}^2,$$

$$\bar{x} = (0, 0).$$

The set B is elementally subregular relative to A at $\bar{x} = (0, 0)$ for all $(b, v) \in \text{gph}(N_B \cap A)$ with constant $\varepsilon = 0$ and neighborhood \mathbb{E} since for all $a \in A$, $a_B \in P_B(a)$ and $v \in N_B(b) \cap A$, there holds

$$\langle v - (a - a_B), a_B - b \rangle = \langle v, a_B - b \rangle - \langle a - a_B, a_B - b \rangle = 0.$$

The set B , however, is not elementally regular at $\bar{x} = (0, 0)$ for any $((0, 0), v) \in \text{gph} N_B$ because by choosing $x = tv \in B$ (where $(0, 0) \neq v \in B \cap N_B((0, 0))$, $t \downarrow 0$), we get

$$\langle v, x \rangle = \|v\| \|x\| > 0.$$

The following equivalences explain how the language of elemental regularity to some extent unifies the existing regularity notions of sets.

PROPOSITION 2.1.4. [84, Proposition 4] Let A , A' and B be closed nonempty subsets of \mathbb{E} .

(i) Let $A \cap B \neq \emptyset$ and suppose that there is a neighborhood W of $\bar{x} \in A \cap B$ and a constant $\varepsilon > 0$ such that for each

$$(a, v) \in V := \left\{ (b_A, u) \in \text{gph} N_A^{\text{prox}} \left| \begin{array}{l} u = b - b_A, \quad \text{for } b \in B \cap W \\ \text{and } b_A \in P_A(b) \cap W \end{array} \right. \right\}, \quad (2.2)$$

it holds that

$$\bar{x} \in \mathbb{B}_{(1+\varepsilon^2)\|v\|}(a+v). \quad (2.3)$$

Then, A is σ -Hölder regular relative to B at \bar{x} with constant $c = \varepsilon^2$ and neighborhood W of \bar{x} if and only if A is elementally subregular of order σ relative to $A \cap P_B^{-1}(a+v)$ at \bar{x} for each $(a, v) \in V$ with constant $\varepsilon = \sqrt{c}$ and the respective neighborhood $U(a, v)$.

- (ii) Let $B \subset A$. The set A is (ε, δ) -subregular relative to B at $\bar{x} \in A$ if and only if A is elementally subregular relative to B at \bar{x} for all $(a, v) \in \text{gph} N_A^{\text{prox}}$ where $a \in \mathbb{B}_\delta(\bar{x})$ with constant ε and neighborhood $\mathbb{B}_\delta(\bar{x})$. Consequently, (ε, δ) -subregularity implies 0-Hölder regularity.
- (iii) If the set A is $(\mathbb{E}, \varepsilon, \delta)$ -regular at \bar{x} , then A is elementally regular at \bar{x} for all (\bar{x}, v) with constant ε , where $0 \neq v \in N_A^{\text{prox}}(\bar{x})$. Consequently, $(\mathbb{E}, \varepsilon, \delta)$ -regularity implies (ε, δ) -subregularity.
- (iv) The set A is Clarke regular at $\bar{x} \in A$ if and only if A is uniformly elementally regular at \bar{x} for all (\bar{x}, v) with $v \in N_A(\bar{x})$. Consequently, Clarke regularity implies (ε, δ) -regularity.
- (v) The set A is super-regular at $\bar{x} \in A$ if and only if for any $\varepsilon > 0$, there is a $\delta > 0$ such that A is elementally regular at \bar{x} for all $(a, v) \in \text{gph} N_A$ where $a \in \mathbb{B}_\delta(\bar{x})$ with constant ε and neighborhood $\mathbb{B}_\delta(\bar{x})$. Consequently, super-regularity implies Clarke regularity.

- (vi) If A is prox-regular at \bar{x} , then there exist positive constants $\bar{\varepsilon}$ and $\bar{\delta}$ such that, for any $\varepsilon > 0$ and $\delta := \frac{\varepsilon \bar{\delta}}{\bar{\varepsilon}}$ defined correspondingly, A is elementally regular at \bar{x} for all $(a, v) \in \text{gph } N_A$ where $a \in \mathbb{B}_\delta(\bar{x})$ with constant ε and neighborhood $\mathbb{B}_\delta(\bar{x})$. Consequently, prox-regularity implies super-regularity.
- (vii) If A is convex then it is elementally regular at all $x \in A$ for all $(a, v) \in \text{gph } N_A$ with constant $\varepsilon = 0$ and the neighborhood \mathbb{E} for both x and v .

The following relations reveal the almost (firm)-nonexpansiveness of the projector onto elementally subregular sets.

PROPOSITION 2.1.5 (characterizations of elemental subregularity). [103, Proposition 3.2]

- (i) A nonempty set $A \subset \mathbb{E}$ is elementally subregular at \bar{x} relative to Λ for $(y, v) \in \text{gph } (N_A^{\text{prox}})$ where $y \in P_A(y+v)$ if and only if there is a neighborhood U of \bar{x} together with a constant $\varepsilon \geq 0$ such that

$$\|x - y\|^2 \leq \varepsilon \|(y' - y) - (x' - x)\| \|x - y\| + \langle x' - y', x - y \rangle$$

holds with $y' = y + v$ whenever $x' \in U \cap \Lambda$ and $x \in P_A x'$.

- (ii) Let the nonempty set $A \subset \mathbb{E}$ be elementally subregular at \bar{x} relative to Λ for $(y, v) \in \text{gph } (N_A^{\text{prox}})$ where $y \in P_A(y+v)$ with the constant $\varepsilon \geq 0$ for the neighborhood U of \bar{x} . Then

$$\|x - y\| \leq \varepsilon \|(y' - y) - (x' - x)\| + \|x' - y'\|$$

holds with $y' = y + v$ whenever $x' \in U \cap \Lambda$ and $x \in P_A x'$.

The next theorem establishes the connection between elemental subregularity of a set and almost nonexpansiveness/averaging of the projector onto that set. Since the cyclic projections algorithm applied to inconsistent feasibility problems involves the properties of the projectors at points that are *outside* the sets, we show how the properties depend on whether the reference points are inside or outside of the sets. The theorem uses the symbol Λ to indicate subsets of the sets and the symbol Λ' to indicate points on some neighborhood whose projection lies in Λ . Later, the sets Λ' will be specialized in the context of cyclic projections to sets of points S_j whose projections lie in A_j . One thing to note in the theorem below is that the almost nonexpansive/averaging property degrades rapidly as the reference points move away from the sets.

THEOREM 2.1.6 (projectors and reflectors onto elementally subregular sets). [103, Theorem 3.1] Let $A \subset \mathbb{E}$ be nonempty closed, and let U be a neighborhood of $\bar{x} \in A$. Let $\Lambda \subset A \cap U$ and $\Lambda' := P_A^{-1}(\Lambda) \cap U$. If A is elementally subregular at \bar{x} relative to Λ' for each

$$(x, v) \in V := \{(z, w) \in \text{gph } N_A^{\text{prox}} \mid z + w \in U \quad \text{and} \quad z \in P_A(z + w)\}$$

with constant ε on the neighborhood U , then the following hold.

- (i) The projector P_A is pointwise almost nonexpansive at each $y \in \Lambda$ on U with violation $\varepsilon' := 2\varepsilon + \varepsilon^2$. That is, at each $y \in \Lambda$

$$\|x - y\| \leq \sqrt{1 + \varepsilon'} \|x' - y\| \quad \forall x' \in U, x \in P_A x'.$$

- (ii) Let $\varepsilon \in [0, 1)$. The projector P_A is pointwise almost nonexpansive at each $y' \in \Lambda'$ with violation $\tilde{\varepsilon}$ on U for $\tilde{\varepsilon} := 4\varepsilon / (1 - \varepsilon)^2$. That is, at each $y' \in \Lambda'$

$$\|x - y\| \leq \frac{1 + \varepsilon}{1 - \varepsilon} \|x' - y'\| \quad \forall x' \in U, x \in P_A x', y \in P_A y'.$$

- (iii) The projector P_A is pointwise almost firmly nonexpansive at each $y \in \Lambda$ with violation $\varepsilon'_2 := 2\varepsilon + 2\varepsilon^2$ on U . That is, at each $y \in \Lambda$

$$\|x - y\|^2 + \|x' - x\|^2 \leq (1 + \varepsilon'_2) \|x' - y\|^2 \quad \forall x' \in U, x \in P_A x'.$$

- (iv) Let $\varepsilon \in [0, 1)$. The projector P_A is pointwise almost firmly nonexpansive at each $y' \in \Lambda'$ with violation $\tilde{\varepsilon}_2 := 4\varepsilon(1 + \varepsilon) / (1 - \varepsilon)^2$ on U . That is, at each $y' \in \Lambda'$

$$\|x - y\|^2 + \|(x' - x) - (y' - y)\|^2 \leq (1 + \tilde{\varepsilon}_2) \|x' - y'\|^2 \quad \forall x' \in U, x \in P_A x', y \in P_A y'.$$

- (v) The reflector R_A is pointwise almost nonexpansive at each $y \in \Lambda$ (respectively, $y' \in \Lambda'$) with violation $\varepsilon'_3 := 4\varepsilon + 4\varepsilon^2$ (respectively, $\tilde{\varepsilon}_3 := 8\varepsilon(1 + \varepsilon) / (1 - \varepsilon)^2$) on U ; that is, for all $y \in \Lambda$ (respectively, $y' \in \Lambda'$)

$$\|x - y\| \leq \sqrt{1 + \varepsilon'_3} \|x' - y\| \quad \forall x' \in U, x \in R_A x'$$

(respectively, $\|x - y\| \leq \sqrt{1 + \tilde{\varepsilon}_3} \|x' - y'\| \quad \forall x' \in U, x \in R_A x', y \in R_A y'$.)

2.2 Metric (sub)regularity of set-valued mappings

The underlying spaces in this section are infinite dimensional normed linear spaces if not otherwise specified. For clarity, we use notation \mathbb{E} whenever presenting results in finite dimensional Euclidean spaces.

2.2.1 Primal characterizations

Metric regularity of set-valued mappings is one of the corner stones of variational analysis. The property is regarded as a natural extension to set-valued mappings of the regularity estimates provided by the classical Banach-Schauder open mapping theorem (for linear operators) and the Lyusternik-Graves theorem (for nonlinear operators) [47, 48, 65, 109, 129].

The Robinson-Ursescu theorem gives an important example of this property, in particular, a closed convex set-valued mapping F is metrically regular at a point $\bar{x} \in \text{dom } F$ for $\bar{y} \in F(\bar{x})$ if and only if \bar{y} is an interior point of $\text{range } F$.

The following concept of *metric regularity with functional modulus on a set* characterizes the stability of mappings at points in their image and has played a central role, implicitly and explicitly, in our analysis of convergence of Picard iterations [4, 59, 103]. In particular, the key insight into condition (b) of Theorem 3.1.1 is the connection to *metric regularity* of set-valued mappings (cf., [50, 129]). This approach to the study of algorithms has been advanced by several authors [2, 3, 70, 74, 122]. We modify the concept of *metric regularity with functional modulus on a set* suggested in [66, Definition 2.1 (b)] and [67, Definition 1 (b)] so that the property is relativized to appropriate sets for iterative methods.

DEFINITION 2.2.1 (metric regularity on a set). [103, Definition 2.5] Let $F : X \rightrightarrows Y$, $U \subset X$, $V \subset Y$. The mapping F is called metrically regular with gauge μ on $U \times V$ relative to $\Lambda \subset X$ if

$$\text{dist}(x, F^{-1}(y) \cap \Lambda) \leq \mu(\text{dist}(y, F(x))) \quad (2.4)$$

holds for all $x \in U \cap \Lambda$ and $y \in V$ with $0 < \mu(\text{dist}(y, F(x)))$. When the set V consists of a single point, $V = \{\bar{y}\}$, then F is said to be metrically subregular for \bar{y} on U with gauge μ relative to $\Lambda \subset X$.

When μ is a linear function (that is, $\mu(t) = \kappa t$, $\forall t \in [0, \infty)$), one says “with constant κ ” instead of “with gauge $\mu(t) = \kappa t$ ”. When $\Lambda = X$, the quantifier “relative to” is dropped. When μ is linear, the infimum of κ for which (2.4) holds is called the modulus of metric regularity on $U \times V$.

The conventional concept of *metric regularity* [10, 50, 129] (and *metric regularity of order ω* , respectively [86]) at a point $\bar{x} \in X$ for $\bar{y} \in F(\bar{x})$ corresponds to the setting in Definition 2.2.1 where $\Lambda = X$, U and V are neighborhoods of \bar{x} and \bar{y} , respectively, and the gauge function $\mu(t) = \kappa t$ ($\mu(t) = \kappa t^\omega$ for metric regularity of order $\omega < 1$) for all $t \in [0, \infty)$, with $\kappa > 0$. The infimum of κ over all neighborhoods U and V such that (2.4) is satisfied is the regularity modulus of F at \bar{x} for \bar{y} and denoted by $\text{reg}(F; \bar{x}|\bar{y})$.

The flexibility of choosing the sets U and V in Definition 2.2.1 allows the same definition and terminology to cover well-known relaxations of metric regularity such as *metric subregularity* (U is a neighborhood of \bar{x} and $V = \{\bar{y}\}$ [50]. In this case, the infimum of κ over all neighborhoods U of \bar{x} such that (2.4) is satisfied is the modulus of metric subregularity of F at \bar{x} for \bar{y} and denoted by $\text{subreg}(F; \bar{x}|\bar{y})$.) and *metric hemi/semiregularity* ($U = \{\bar{x}\}$ and V is a neighborhood of \bar{y} [109, Definition 1.47]). For our purposes, we will use the flexibility of choosing U and V in Definition 2.2.1 to *exclude* the reference point \bar{x} and to *isolate* the image point \bar{y} . This is reminiscent of the Kurdyka-Łojasiewicz (KL) property [25] for functions which requires that the subdifferential possesses a sharpness property near (but not at) critical points of the function. However, since the restriction of V to a point features prominently in our development, we retain the terminology *metric subregularity* to

ease the technicality of the presentation. The reader is cautioned, however, that our usage of metric subregularity does not precisely correspond to the usual definition (see [50]) since we do not require the domain U to be a neighborhood.

The metric regularity of a set-valued mapping F can be used for measuring the “conditioning” of the generalized equation: for a given $y \in Y$,

$$\text{find } x \in X \text{ such that } y \in F(x). \quad (2.5)$$

Inequality (2.4) then provides an estimate of how far a point x can be from the solution set of (2.5) corresponding to the right-hand side y ; this distance is bounded from above by a multiple κ of the “residual” $\text{dist}(y, F(x))$. In other words, the presence of metric regularity of F at \bar{x} for $\bar{y} \in F(\bar{x})$ means that (2.5) is, from a certain perspective, well-posed around there. This conditioning is stable under small perturbations on F [48, 49], where quantitative estimates of how large a perturbation can be before metric regularity breaks down are also established.

Metric regularity admits several equivalent descriptions to (2.4). Recall that [65, p.510] F is called *metrically graph-regular* at \bar{x} for $\bar{y} \in F(\bar{x})$ if there exist positive numbers κ and δ such that

$$\text{dist}(x, F^{-1}(y)) \leq d_\kappa((x, y), \text{gph } F), \quad \forall x \in B_\delta(\bar{x}), y \in B_\delta(\bar{y}), \quad (2.6)$$

where

$$d_\kappa((x, y), \text{gph } F) := \inf_{(u, w) \in \text{gph } F} (\text{dist}(x, u) + \kappa \text{dist}(y, w)).$$

The two descriptions (2.4) and (2.6) are equivalent with the same κ (and possibly different δ), in particular, metric regularity of F at \bar{x} for \bar{y} is equivalent to metric graph-regularity of F at \bar{x} for \bar{y} [65, Proposition 4, p.510]. We also refer the reader to that paper for other equivalent descriptions of metric regularity. The main idea for these possibilities is that the definition of the conventional metric regularity would be qualitatively unchanged when reasonable restriction on x and y was added, for example, $(x, y) \notin \text{gph } F$.

Dmitruk et al. [46] and Ioffe [63] showed the equivalence between the metric regularity and the linear openness property of a set-valued mapping F , which are determined by the first-order behaviour of the mapping and invariant under sufficiently small first-order perturbations [46, 48, 65]. The two properties are also equivalent to the Aubin property of the inverse mapping F^{-1} thanks to Borwein and Zhuang [31] and Penot [123].

DEFINITION 2.2.2. (i) *A set-valued mapping F is linearly open at \bar{x} for $\bar{y} \in F(\bar{x})$ if there exist $\kappa \geq 0$ and $\delta > 0$ such that*

$$F(x + \kappa \rho \text{int } \mathbb{B}) \supset [F(x) + \rho \text{int } \mathbb{B}] \cap B_\delta(\bar{y}), \quad \forall x \in B_\delta(\bar{x}), \forall \rho > 0. \quad (2.7)$$

The infimum of κ over all δ such that (2.7) is satisfied is the modulus of linear openness of F at \bar{x} for \bar{y} and denoted by $\text{lop}(F; \bar{x}|\bar{y})$.

(ii) F has the Aubin property at \bar{x} for $\bar{y} \in F(\bar{x})$ if there exist $\kappa \geq 0$ and $\delta > 0$ such that

$$\text{excess}(F(x) \cap B_\delta(\bar{y}), F(x')) \leq \kappa \text{dist}(x, x'), \quad \forall x, x' \in B_\delta(\bar{x}). \quad (2.8)$$

The infimum of κ over all combinations of κ and δ such that (2.8) is satisfied is the Lipschitz modulus of F at \bar{x} for $\bar{y} \in F(\bar{x})$ and denoted by $\text{lip}(F; \bar{x}|\bar{y})$.

PROPOSITION 2.2.3. *Metric regularity and linear openness of a set-valued mapping F at \bar{x} for $\bar{y} \in F(\bar{x})$ are equivalent. They are also equivalent to the Aubin property of the mapping F^{-1} at \bar{y} for \bar{x} . Moreover, it holds*

$$\text{reg}(F; \bar{x}|\bar{y}) = \text{lop}(F; \bar{x}|\bar{y}) = \text{lip}(F^{-1}; \bar{y}|\bar{x}).$$

Metric subregularity also enjoys the relationships analogous to those stated in Proposition 2.2.3 with the sub-versions of linear openness and Aubin properties. The interested reader is referred to [1, 50, 58, 129].

Metric subregularity can also be characterized via the concept of local error bound of extended real-valued functions. A function $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ having a local error bound at a point \bar{x} with $f(\bar{x}) = 0$ simply coincides with the set-valued mapping $x \mapsto [f(x), +\infty)$ ($\forall x \in X$) being metrically regular at \bar{x} for 0 [80, Proposition 9(ii)]. This transition allows one to deduce criteria for local error bounds of l.s.c. extended real-valued functions from those for metric subregularity.

In the finite dimensional setting \mathbb{E} , the following proposition, taken from [50], characterizes metric subregularity in terms of the graphical derivative defined by (1.2).

PROPOSITION 2.2.4 (characterization of metric subregularity). *Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ have locally closed graph at $(\bar{x}, \bar{y}) \in \text{gph} T$, $F := T - \text{Id}$, and $\bar{z} := \bar{y} - \bar{x}$. Then F is metrically subregular at \bar{x} for \bar{z} with constant κ and some neighborhood U of \bar{x} satisfying $U \cap F^{-1}(\bar{z}) = \{\bar{x}\}$ if and only if the graphical derivative satisfies*

$$DF(\bar{x}|\bar{z})^{-1}(0) = \{0\}. \quad (2.9)$$

If, in addition, T is single-valued and continuously differentiable on U , then the two properties hold if and only if ∇F has rank n at \bar{x} with $\left\| \left[[\nabla F(x)]^T \right]^{-1} \right\| \leq \kappa$ for all x on U .

While the characterization (2.9) appears daunting, the property comes almost for free for polyhedral mappings.

PROPOSITION 2.2.5 (polyhedrality implies metric subregularity). [103, Proposition 2.6] *Let $\Lambda \subset \mathbb{E}$ be an affine subspace and $T : \Lambda \rightrightarrows \Lambda$. If T is polyhedral and $\text{Fix } T \cap \Lambda$ is an isolated point, $\{\bar{x}\}$, then $F := T - \text{Id}$ is metrically subregular at \bar{x} for 0 relative to Λ with some constant κ and some neighborhood U of \bar{x} satisfying $U \cap F^{-1}(0) = \{\bar{x}\}$.*

The property characterized in Proposition 2.2.4 is known as the *strong metric subregularity* [50, Section 3I] while Proposition 2.2.5 characterizes its relative version. For completeness, F is strongly metrically subregular at \bar{x} for $\bar{y} \in F(\bar{x})$ (relative to Λ , respectively) if it is metrically subregular at \bar{x} for \bar{y} and \bar{x} is an isolated point of $F^{-1}(\bar{y})$ (relative to Λ , respectively). For certain applications in stability and numerical analysis, the strong metric subregularity is needed instead of metric subregularity due to its persistence under small perturbations on F .

2.2.2 Dual characterizations

Metric regularities of set-valued mappings defined in Definition d:(str)metric (sub)reg are obviously properties in the primal space. They can also be characterized via objectives of dual spaces [69, 107, 136]. The following coderivative [106] and outer coderivative [69] of set-valued mappings are the central concepts in this subsection. Similar to speaking of subdifferentials of functions, the adjective “outer” means that the sequence (x_n) in Definition 2.2.6 (ii) is outside the solution set of the inverse problem of finding x such that $\bar{y} \in F(x)$. The latter problem is one of primal motivations for the development of the theory of metric subregularity. For the history of coderivative, we refer the reader to the monograph [109].

DEFINITION 2.2.6. *Let $F : X \rightrightarrows Y$ and $\bar{x} \in \text{dom } F$.*

- (i) *The (limiting or Mordukhovich) coderivative of F at \bar{x} for $\bar{y} \in F(\bar{x})$ is the set-valued mapping $D^*F(\bar{x}|\bar{y}) : Y \rightrightarrows X$ defined by*

$$x^* \in D^*F(\bar{x}|\bar{y})(y^*) \iff (x^*, -y^*) \in N_{\text{gph } F}(\bar{x}, \bar{y}).$$

- (ii) *The outer coderivative of F at \bar{x} for $\bar{y} \in F(\bar{x})$ is the set-valued mapping $D^*_>F(\bar{x}|\bar{y}) : Y \rightrightarrows X$ defined by $x^* \in D^*_>F(\bar{x}|\bar{y})(y^*)$ if there exists a sequence of quadruples (x_n, y_n, x_n^*, y_n^*) converging to $(\bar{x}, \bar{y}, x^*, y^*)$ such that, for $n = 1, 2, \dots$,*

$$\bar{y} \notin F(x_n), y_n \in P_{F(x_n)}(\bar{y}), y_n^* = \lambda_n(y_n - \bar{y}), \lambda_n > 0, x_n^* \in D^*F(x_n|y_n)(y_n^*).$$

The coderivative mapping $D^*F(\bar{x}|\bar{y})$ is positively homogeneous, i.e., its graph is a cone. Recall that [48] the outer norm of a positively homogeneous set-valued mapping S is defined by

$$\|S\|^+ := \sup_{\|x\| \leq 1} \sup_{y \in S(x)} \|y\|.$$

The following famous Mordukhovich criterion provides not only a handy test for metric regularity of F at \bar{x} for $\bar{y} \in F(\bar{x})$ (equivalently, the linear openness and Aubin properties) but also an estimate of the regularity modulus $\text{reg}(F; \bar{x}|\bar{y})$ via the knowledge of the coderivative mapping $D^*F(\bar{x}|\bar{y})$. This criterion also encompasses dual characterizations of transversality of collections of sets thanks to the relationships that we will discuss in Section 2.3.

PROPOSITION 2.2.7 (Mordukhovich criterion). [107] *Let $F : X \rightrightarrows Y$ be a set-valued mapping between Euclidean spaces. Suppose that $\text{gph } F$ is locally closed at a point $(\bar{x}, \bar{y}) \in \text{gph } F$. Then F is metrically regular at \bar{x} for \bar{y} if and only if*

$$D^*F(\bar{x}|\bar{y})^{-1}(0) = \{0\}. \quad (2.10)$$

In that case,

$$\text{reg}(F; \bar{x}|\bar{y}) = \|D^*F(\bar{x}|\bar{y})^{-1}\|^+.$$

It is worth mentioning that the Mordukhovich criterion is also true in the much more general setting of Asplund spaces provided that a more general (mixed) coderivative is used and F^{-1} satisfies a certain compactness assumption (the partial sequential normal compactness) [110]. Sufficient and/or necessary conditions for metric regularity in the infinite dimensional setting were also established, for example, in [77, 108].

Dual characterizations of metric subregularity can often be obtained in two directions. The first direction is based on the fact that (2.10) is straightforwardly a sufficient condition for metric subregularity. By reducing in an appropriate way the size of (the graph of) the mapping $D^*F(\bar{x}|\bar{y})$ in the Mordukhovich criterion (2.10), for example, to the outer coderivative - Proposition 2.2.8, one can naturally expect to come up with dual necessary and/or sufficient conditions [69, 136, 137]. The second direction is based on the equivalence between the metric subregularity of F at \bar{x} for $\bar{y} \in F(\bar{x})$ and the existence of a local error bound for the function $x \mapsto \text{dist}(\bar{y}, F(x))$ at \bar{x} . Whenever this function is l.s.c. around \bar{x} , for example, when F is outer semicontinuous), subdifferential criteria for local error bounds can automatically be interpreted as dual characterizations for metric subregularity [81]. Intimate relationships between subdifferentials of a function $x \mapsto f(x)$ and the corresponding coderivatives of the mapping $x \mapsto [f(x), +\infty)$ to some extent unify the two directions, see [81].

For closed convex set-valued mappings, the following criterion, which is analogous to (2.10) for metric regularity, for metric subregularity was proved in [136]. The statement also holds true when X is an Asplund space.

PROPOSITION 2.2.8. [136, Corollary 3.2] *Suppose that $F : X \rightrightarrows Y$ is convex and $\text{gph } F$ is locally closed at a point $(\bar{x}, \bar{y}) \in \text{gph } F$. Then F is metrically subregular at \bar{x} for \bar{y} if and only if*

$$D_{>}^*F(\bar{x}|\bar{y})^{-1}(0) = \{0\}.$$

PROPOSITION 2.2.9. [136, Theorem 3.6 (ii)] *Let $\text{gph } F$ be locally closed at a point $(\bar{x}, \bar{y}) \in \text{gph } F$. Suppose that there exist positive numbers γ, δ such that*

$$\text{dist}(0, D^*F(x|y)(y - \bar{y})) \geq \gamma \|y - \bar{y}\|,$$

for all $x \in B_\delta(\bar{x}) \setminus F^{-1}(\bar{y})$ and $y \in B_\delta(\bar{y}) \cap P_{F(x)}(\bar{y})$. Then F is metrically subregular at \bar{x} for \bar{y} with the modulus of metric subregularity not greater than $\frac{1}{\gamma}$.

Proposition 2.2.9 obviously yields an estimate of modulus of metric subregularity:

$$\text{subreg}(F; \bar{x}|\bar{y}) \leq \liminf_{\delta \downarrow 0} \left\{ \frac{\text{dist}(0, D^*F(x|y)(y - \bar{y}))}{\|y - \bar{y}\|} : x \in B_\delta(\bar{x}) \setminus F^{-1}(\bar{y}), y \in B_\delta(\bar{y}) \cap P_{F(x)}(\bar{y}) \right\}.$$

The following estimate of modulus of metric subregularity was also proved in [136, Theorem 3.1], where the authors made use of the Fréchet coderivative \widehat{D}^*F , whose definition is similar to Definition 2.2.6 (i) for the (limiting) coderivative D^*F with the only change that the Fréchet normal cone is used instead of the (limiting) one. Note that the inequality can be strict.

$$\text{subreg}(F; \bar{x}|\bar{y}) \leq \inf_{\delta > 0} \sup \left\{ \left\| \widehat{D}^*F(x|y)^{-1} \right\|^+ : x \in B_\delta(\bar{x}) \setminus F^{-1}(\bar{y}), y \in B_\delta(\bar{y}) \cap F(x) \right\}.$$

Further discussion regarding dual characterizations of metric subregularity of set-valued mappings in more general settings can be found in [81, 136, 137, 138].

We refer the readers to the monographs [50, 109] and surveys [1, 10, 65, 68] for a comprehensive exposition of the properties of set-valued mappings in variational analysis.

2.3 (Sub)transversality of collections of sets

The underlying spaces in this section are infinite dimensional normed linear spaces if not otherwise specified. For clarity, we use notation \mathbb{E} whenever presenting results in finite dimensional Euclidean spaces. The content of this section is taken from our joint papers with Prof. Alexander Y. Kruger [84, 83] except Definition 2.3.12 and Proposition 2.3.13 taken from our joint work with Dr. Matthew K. Tam [103].

2.3.1 Primal characterizations

In this section we discuss these two standard regularity properties of a pair of sets, namely *transversality* and *subtransversality* (also known under other names). Subtransversality of collections of sets has emerged as a key - by some estimates *the key* - notion in the analysis of convergence of simple iterative methods for solving feasibility problems. The origins of the concept can be traced back to that of transversality in differential geometry which deals of course with smooth manifolds (see, for instance, [57, 61]). The notion of transversality in differential geometry is motivated by the problem of determining when the intersection of two smooth manifolds is also a manifold near some point in the intersection. A sufficient condition for this to happen is when the collection $\{A, B\}$ of smooth manifolds is *transversal* at $\bar{x} \in A \cap B$, i.e., the sum of the tangent spaces to A and B at \bar{x} generates the ambient space. Under this assumption, $A \cap B$ is a smooth manifold around \bar{x} and the tangent space

to the intersection is equal to the intersection of the tangent spaces at \bar{x} and the normal spaces to the sets at \bar{x} have only the origin in common (cf. [57, 68, 91]). The tangent space intersection property is only a necessary condition and is in general weaker than the condition on the normal spaces.

DEFINITION 2.3.1 (transversality and subtransversality). [84, Definition 6]

(i) $\{A, B\}$ is subtransversal at \bar{x} if there exist numbers $\alpha > 0$ and $\delta > 0$ such that

$$(A + (\alpha\rho)\mathbb{B}) \cap (B + (\alpha\rho)\mathbb{B}) \cap \mathbb{B}_\delta(\bar{x}) \subseteq (A \cap B) + \rho\mathbb{B} \quad (2.11)$$

for all $\rho \in (0, \delta)$.

If, additionally, \bar{x} is an isolated point of $A \cap B$, then $\{A, B\}$ is called strongly subtransversal at \bar{x} . The (possibly infinite) supremum of all α above is denoted $\text{sr}[A, B](\bar{x})$ with the convention that the supremum of the empty set is zero.

(ii) $\{A, B\}$ is transversal at \bar{x} if there exist numbers $\alpha > 0$ and $\delta > 0$ such that

$$(A - a - x_1) \cap (B - b - x_2) \cap (\rho\mathbb{B}) \neq \emptyset \quad (2.12)$$

for all $\rho \in (0, \delta)$, $a \in A \cap \mathbb{B}_\delta(\bar{x})$, $b \in B \cap \mathbb{B}_\delta(\bar{x})$, and all $x_1, x_2 \in \mathbb{E}$ with

$$\max\{\|x_1\|, \|x_2\|\} < \alpha\rho.$$

The (possibly infinite) supremum of all α above is denoted by $\text{r}[A, B](\bar{x})$ with the convention that the supremum of the empty set is zero.

REMARK 2.3.2. [84, Remark 3] The maximum of the norms in Definition 2.3.1 – explicitly present in part ((ii)) and implicitly also in part ((i)) – corresponds to the maximum norm in \mathbb{R}^2 employed in these definitions and subsequent assertions. It can be replaced everywhere by the sum norm (pretty common in this type of definition in the literature) or any other equivalent norm. All the assertions that follow including the quantitative characterizations remain valid (as long as the same norm is used everywhere), although the exact values $\text{sr}[A, B](\bar{x})$ and $\text{r}[A, B](\bar{x})$ do depend on the chosen norm and some estimates can change. Note that the currently used maximum norm is not Euclidean. These details become important in the context of applications where one norm may be more appropriate than another.

Definition 2.3.1((i)) was introduced recently in [87] and can be viewed as a local analogue of the global *uniform normal property* introduced in the convex setting in [11, Definition 3.1(4)] as a generalization of the *property (N)* of convex cones by Jameson [71]. A particular case of the Jameson property (N) for convex cones A and B such that $B = -A$ and $A \cap (-A) = \{0\}$ was studied by M. Krein in the 1940s. Definition 2.3.1((ii)) first appeared in [78] (see also [79, 80]) in the normed linear space setting, where the property was referred to as simply *regularity* (and later as *strong regularity* and *uniform regularity*). In [90], the property is called *linearly regular intersection*.

EXAMPLE 2.3.3. [84, Example 3] If $\bar{x} \in \text{int}(A \cap B)$, then $\{A, B\}$ is trivially transversal (and consequently subtransversal) at \bar{x} with any $\alpha > 0$. Thus, $r[A, B](\bar{x}) = \text{sr}[A, B](\bar{x}) = \infty$.

EXAMPLE 2.3.4. [84, Example 4] If $A = B$ and $\bar{x} \in \text{bd}(A \cap B)$, then $A + (\alpha\rho)\mathbb{B} = B + (\alpha\rho)\mathbb{B}$ and $A \cap B + \rho\mathbb{B} = A + \rho\mathbb{B}$. Hence, condition (2.11) holds (with any $\delta > 0$) if and only if $\alpha \leq 1$. Thus, $\{A, B\}$ is subtransversal at \bar{x} and $\text{sr}[A, B](\bar{x}) = 1$.

Note that, under the conditions of Example 2.3.4, $\{A, B\}$ does not have to be transversal at \bar{x} .

EXAMPLE 2.3.5. [84, Example 5] Let $\mathbb{E} = \mathbb{R}^2$, $A = B = \mathbb{R} \times \{0\}$ and $\bar{x} = (0, 0)$. Thanks to Example 2.3.4, $\{A, B\}$ is subtransversal at \bar{x} and $\text{sr}[A, B](\bar{x}) = 1$. At the same time, $A - a = B - b = \mathbb{R} \times \{0\}$ for any $a \in A$ and $b \in B$. If $x_1 = (0, \varepsilon)$ and $x_2 = (0, 0)$, then condition (2.12) does not hold for any $\varepsilon > 0$ and $\rho > 0$. Thus, $\{A, B\}$ is not transversal at \bar{x} and $r[A, B](\bar{x}) = 0$.

The next two results are a catalog of the main primal characterizations of subtransversality and transversality, respectively.

THEOREM 2.3.6 (characterizations of subtransversality). [84, Theorem 1] The following statements are equivalent to $\{A, B\}$ being subtransversal at \bar{x} .

(i) There exist numbers $\delta > 0$ and $\alpha > 0$ such that

$$(A - x) \cap (B - x) \cap (\rho\mathbb{B}) \neq \emptyset$$

for all $x \in \mathbb{B}_\delta(\bar{x})$ such that $x = a + x_1 = b + x_2$ for some $a \in A$, $b \in B$ and $x_1, x_2 \in \mathbb{E}$ with $\max\{\|x_1\|, \|x_2\|\} < \alpha\rho$. Moreover, $\text{sr}[A, B](\bar{x})$ is the exact upper bound of all numbers α such that the condition above is satisfied.

(ii) There exist numbers $\delta > 0$ and $\alpha > 0$ such that

$$\alpha \text{dist}(x, A \cap B) \leq \max\{\text{dist}(x, A), \text{dist}(x, B)\} \quad \text{for all } x \in \mathbb{B}_\delta(\bar{x}). \quad (2.13)$$

Moreover, $\text{sr}[A, B](\bar{x})$ is the exact upper bound of all numbers α such that (2.13) is satisfied.

(iii) There exist numbers $\alpha \in (0, 1)$ and $\delta > 0$ such that

$$\alpha \text{dist}(x, A \cap B) \leq \text{dist}(x, B) \quad \text{for all } x \in A \cap \mathbb{B}_\delta(\bar{x}). \quad (2.14)$$

Moreover,

$$\frac{1}{2(\text{sr}'[A, B](\bar{x})^{-1} + 1)} \leq \text{sr}[A, B](\bar{x}) \leq \text{sr}'[A, B](\bar{x}),$$

where $\text{sr}'[A, B](\bar{x})$ is the exact upper bound of all numbers $\alpha \in (0, 1)$ such that condition (2.14) is satisfied, with the convention that the supremum of the empty subset of \mathbb{R}_+ equals 0.

REMARK 2.3.7 (Historical remarks and further relations). [84, Remark 4] Thanks to characterization (ii) of Theorem 2.3.6, subtransversality of a collection of sets can be recognized as a well known regularity property that has been around for more than 20 years under the names of (local) linear regularity, metric regularity, linear coherence, metric inequality, and subtransversality; cf. [11, 12, 13, 51, 59, 64, 65, 68, 75, 92, 117, 121, 124, 135, 139]. It has been used as the key assumption when establishing linear convergence of sequences generated by cyclic projections algorithms and a qualification condition for subdifferential and normal cone calculus formulae. This property is implied by the bounded linearly regularity [13]. If A and B are closed convex sets and the collection $\{A, B\}$ is subtransversal at any point in $A \cap B$, then it is boundedly linear regular; cf. [11, Remark 6.1(d)]. Characterization (iii) of Theorem 2.3.6 can be considered as a nonconvex extension of [113, Theorem 3.1].

One can also observe that condition (2.13) is equivalent to the function

$$x \mapsto \max\{\text{dist}(x, A), \text{dist}(x, B)\}$$

having a local error bound [9, 52, 81]/weak sharp minimum [35, 36, 37] at \bar{x} with constant α .

In the finite dimensional setting \mathbb{E} , the geometrical property (2.13) of a collection of sets $\{A, B\}$ can also be viewed as a certain property of a collection of distance functions $x \mapsto \text{dist}(x, A)$ and $x \mapsto \text{dist}(x, B)$. It is sufficient to notice that

$$A \cap B = \{x \in \mathbb{E} \mid \max\{\text{dist}(x, A), \text{dist}(x, B)\} \leq 0\}.$$

One can study regularity properties of collections of arbitrary (not necessarily distance) functions. Such an attempt has been made recently in the convex setting by Pang [121]. Given a collection of convex functions $\{f_1, f_2\}$, the following analogue of condition (2.13) is considered in [121]:

$$\alpha \text{dist}(x, C) \leq \max\{\text{dist}(x, H_1(x)), \text{dist}(x, H_2(x))\} \quad \text{for all } x \in \mathbb{E},$$

where $C := \{u \in \mathbb{E} \mid \max\{f_1(u), f_2(u)\} \leq 0\}$, $H_i(x) := \{u \in \mathbb{E} \mid f_i(x) + \langle v_i, u - x \rangle \leq 0\}$ for some chosen $v_i \in \partial f_i(x)$ if $f_i(x) > 0$ and $H_i(x) := \mathbb{E}$ otherwise, $i = 1, 2$. It is easy to check that, in the case of distance functions, this property reduces to (2.13).

THEOREM 2.3.8 (metric characterizations of transversality). [84, Theorem 2 (i)–(ii)] The following statements are equivalent to $\{A, B\}$ being transversal at \bar{x} .

(i) There exist numbers $\delta > 0$ and $\alpha > 0$ such that

$$\alpha \text{dist}(x, (A - x_1) \cap (B - x_2)) \leq \max\{\text{dist}(x, A - x_1), \text{dist}(x, B - x_2)\}, \quad (2.15)$$

for all $x \in \mathbb{B}_\delta(\bar{x})$ and $x_1, x_2 \in \delta\mathbb{B}$. Moreover, $r[A, B](\bar{x})$ is the exact upper bound of all numbers α such that (2.15) is satisfied.

(ii) There exist numbers $\delta > 0$ and $\alpha > 0$ such that

$$\alpha \operatorname{dist}(x, (A - x_1) \cap (B - x_2)) \leq \operatorname{dist}(x, B - x_2), \quad \forall x \in (A - x_1) \cap \mathbb{B}_\delta(\bar{x}), \quad x_1, x_2 \in \delta \mathbb{B}. \quad (2.16)$$

Moreover,

$$\frac{r'[A, B](\bar{x})}{r'[A, B](\bar{x}) + 2} \leq r[A, B](\bar{x}) \leq r'[A, B](\bar{x}),$$

where $r'[A, B](\bar{x})$ is the exact upper bound of all numbers α such that condition (2.16) is satisfied.

REMARK 2.3.9. [84, Remark 5] Characterization (i) of Theorem 2.3.8 reveals that the transversality of a collection of sets corresponds to subtransversality/linear regularity of all their small translations holding uniformly (cf. [51, page 1638]). Property (2.15) was referred to in [78, 79, 80] as strong metric inequality. If A and B are closed convex sets and $\operatorname{int} A \neq \emptyset$, then the transversality of the collection $\{A, B\}$ is equivalent to the conventional qualification condition: $\operatorname{int} A \cap B \neq \emptyset$; cf. [78, Proposition 14]. One can think of condition (2.15) as a kind of uniform local error bound/relaxed weak sharp minimum property; cf. [79].

The characterization of subtransversality given in Theorem 2.3.6(i) and the definition of transversality shows that transversality implies subtransversality (see Theorem 2.3.10 below). Alternatively, the implication is also immediate from Theorem 2.3.6(ii) and Theorem 2.3.8(i). There are a number of other useful sufficient conditions for subtransversality, detailed in the next theorem.

THEOREM 2.3.10 (primal sufficient conditions for subtransversality). [84, Theorem 4 (i) & (iii)–(iv)] If one of the following hold, then $\{A, B\}$ is subtransversal at \bar{x} .

- (i) The collection $\{A, B\}$ is transversal at \bar{x} . Moreover, $r[A, B](\bar{x}) \leq \operatorname{sr}[A, B](\bar{x})$.
- (ii) The sets A and B are intrinsically transversal at \bar{x} .
- (iii) The set B intersects A separably at \bar{x} and B is 0-Hölder regular relative to A at \bar{x} with an adequate compromise between the constants.

REMARK 2.3.11 (entanglement of elemental regularity and regularity of collections of sets). [84, Remark 12] Theorem 2.3.10(iii) demonstrates that regularity of individual sets has implications for the regularity of the collection of sets. The converse entanglement has also been observed in [118, Proposition 8]: if A and B are intrinsically transversal at \bar{x} with constant α , then A is σ -Hölder regular at \bar{x} relative to B for every $\sigma \in [0, 1)$ with any constant $c < \frac{\alpha^2}{1 - \alpha^2}$.

As a consequence of Proposition 2.1.4(i), if A and B are intrinsically transversal at \bar{x} with constant $\alpha \in (0, 1]$ and, in addition, there is a neighborhood W of \bar{x} and a positive

constant $\varepsilon < \frac{\alpha}{\sqrt{1-\alpha^2}}$ such that for each $(a, v) \in V$ defined in (2.2), condition (2.3) holds true, then A is elementally subregular of any order $\sigma \in [0, 1)$ relative to $A \cap P_B^{-1}(a + v)$ at \bar{x} for each $(a, v) \in V$ with constant ε and the respective neighborhood $U(a, v)$.

For analyzing convergence of algorithms dealing with inconsistent feasibility in Section 3.2, one needs to extend the (sub)transversality concepts to collections of sets that don't necessarily intersect. The idea behind the following definition stems from the equivalence between metric subregularity of an appropriate set-valued mapping on the product space and subtransversality of sets at common points [84, Theorem 3]. The trick to extending this to points that do not belong to all the sets is to define the correct set-valued mapping.

DEFINITION 2.3.12 (subtransversality of (inconsistent) collections of sets). [103, Definition 3.2] Let $\{A_1, A_2, \dots, A_m\}$ be a collection of nonempty closed subsets of \mathbb{E} and define $\Psi : \mathbb{E}^m \rightrightarrows \mathbb{E}^m$ by $\Psi(x) := P_A(\Pi x) - \Pi x$ where $A := A_1 \times A_2 \times \dots \times A_m$, the projection P_A is with respect to the Euclidean norm on \mathbb{E}^m and Π is the permutation mapping given by (1.12). Let $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m) \in \mathbb{E}^m$ and $\bar{y} \in \Psi(\bar{x})$.

- (i) The collection of sets is said to be subtransversal with gauge μ relative to $\Lambda \subset \mathbb{E}^m$ at \bar{x} for \bar{y} if Ψ is metrically subregular at \bar{x} for \bar{y} on some neighborhood U of \bar{x} (metrically regular on $U \times \{\bar{y}\}$) with gauge μ relative to Λ .
- (ii) The collection of sets is said to be transversal with gauge μ relative to $\Lambda \subset \mathbb{E}^m$ at \bar{x} for \bar{y} if Ψ is metrically regular with gauge μ relative to Λ on $U \times V$, for some neighborhoods U of \bar{x} and V of \bar{y} .

As in Definition 2.2.1, when $\mu(t) = \kappa t$, $\forall t \in [0, \infty)$, one says “constant κ ” instead of “gauge $\mu(t) = \kappa t$ ”. When $\Lambda = \mathbb{E}$, the quantifier “relative to” is dropped.

Consistent with the terminology of metric regularity and subregularity, the prefix “sub” is meant to indicate the pointwise version of the more classical, though restrictive, idea of transversality. When the point $\bar{x} = (\bar{u}, \dots, \bar{u})$ for $\bar{u} \in \bigcap_{j=1}^m A_j$ the following characterization of subtransversality holds.

Examples of subtransversality for inconsistent collections of sets are later given in Examples 3.2.10 and 3.2.11.

PROPOSITION 2.3.13 (subtransversality at common points). [103, Proposition 3.3] Let \mathbb{E}^m be endowed with 2-norm, that is, $\|(x_1, x_2, \dots, x_m)\|_2 = \left(\sum_{j=1}^m \|x_j\|_{\mathbb{E}}^2\right)^{1/2}$. A collection $\{A_1, A_2, \dots, A_m\}$ of nonempty closed subsets of \mathbb{E} is subtransversal relative to

$$\Lambda := \{x = (u, u, \dots, u) \in \mathbb{E}^m \mid u \in \mathbb{E}\}$$

at $\bar{x} = (\bar{u}, \dots, \bar{u})$ with $\bar{u} \in \bigcap_{j=1}^m A_j$ for $\bar{y} = 0$ with gauge μ if there exists a neighborhood U' of \bar{u} together with a gauge μ' satisfying $\sqrt{m}\mu' \leq \mu$ such that

$$\text{dist}(u, \bigcap_{j=1}^m A_j) \leq \mu' \left(\max_{j=1, \dots, m} \text{dist}(u, A_j) \right), \quad \forall u \in U'. \quad (2.17)$$

Conversely, if $\{A_1, A_2, \dots, A_m\}$ is subtransversal relative to Λ at \bar{x} for $\bar{y} = 0$ with gauge μ , then (2.17) is satisfied with any gauge μ' for which $\mu(\sqrt{mt}) \leq \sqrt{m}\mu'(t)$ for all $t \in [0, \infty)$.

Note that if one endows \mathbb{E}^m with the maximum norm,

$$\|(x_1, x_2, \dots, x_m)\|_{\mathbb{E}^m} := \max_{1 \leq j \leq m} \|x_j\|_{\mathbb{E}},$$

then it holds that

$$\begin{aligned} \text{dist}(x, (\bigcap_{j=1}^m A_j, \bigcap_{j=1}^m A_j, \dots, \bigcap_{j=1}^m A_j) \cap \Lambda) &= \text{dist}(u, \bigcap_{j=1}^m A_j); \\ \text{dist}(x, A) &= \max_{j=1, \dots, m} \text{dist}(u, A_j) \quad \text{for all } u \text{ and } x \text{ as above.} \end{aligned}$$

Then the two properties in Proposition 2.3.13 are equivalent for the same gauge $\mu' = \mu$.

2.3.2 (Sub)transversality versus metric (sub)regularity

The regularity/transversality properties of the collection $\{A, B\}$ in X can be understood in terms of the corresponding properties of the set-valued mapping $F : X \rightrightarrows X^2$ defined by (cf. [51, 65, 68, 90])

$$F(x) := (A - x) \times (B - x). \quad (2.18)$$

For $x \in X$ and $u = (u_1, u_2) \in X^2$, we have

$$x \in A \cap B \iff (0, 0) \in F(x), \quad F^{-1}(u) = (A - u_1) \cap (B - u_2), \quad \text{and} \quad F^{-1}(0) = A \cap B.$$

The mapping (2.18) is not the only possibility. Another useful construction is given by the set-valued mapping $G : X^2 \rightrightarrows X$ (cf. [91, page 226], [51, page 1638], [68, Corollary 7.13]) defined as follows:

$$G(x_1, x_2) := \begin{cases} \{x_1 - x_2\} & \text{if } x_1 \in A \text{ and } x_2 \in B, \\ \emptyset & \text{otherwise.} \end{cases} \quad (2.19)$$

Obviously,

$$0 \in G(x_1, x_2) \iff x_1 = x_2 \in A \cap B.$$

Conversely, the regularity of certain set-valued mappings can be understood in terms of the corresponding properties of collections of sets. Indeed, given a set-valued mapping $F : X \rightrightarrows Y$, its regularity properties at a point (\bar{x}, \bar{y}) in its graph $\text{gph } F := \{(x, y) \in X \times Y \mid y \in F(x)\}$ are connected to those of the collection of sets (cf. [78, Corollary 2.1])

$$A := \text{gph } F \quad \text{and} \quad B := X \times \{\bar{y}\} \quad (2.20)$$

in $X \times Y$. One can check that $(\bar{x}, \bar{y}) \in A \cap B = F^{-1}(\bar{y}) \times \{\bar{y}\}$.

THEOREM 2.3.14 (characterizations of (sub)transversality via set-valued mappings). [84, Theorem 3] Subtransversality of the collection $\{A, B\}$ at a point $\bar{x} \in A \cap B$ has the following equivalent characterizations.

- (i) The set-valued mapping $F : X \rightrightarrows X^2$ given by (2.18) with the max norm on X^2 is metrically subregular at \bar{x} for $(0, 0)$. Moreover,

$$\text{sr}[A, B](\bar{x}) = \text{sr}[F](\bar{x}, (0, 0)).$$

The mapping F is strongly metrically subregular at \bar{x} for $(0, 0)$ if and only if the collection $\{A, B\}$ is strongly subtransversal there.

- (ii) The set-valued mapping $G : X^2 \rightrightarrows X$ given by (2.19) with the l_2 -norm on X^2 (i.e., $\|(x, x')\| = \sqrt{\|x\|^2 + \|x'\|^2}$ for all $x, x' \in X$) is metrically subregular at (\bar{x}, \bar{x}) for 0. Moreover,

$$\sqrt{\frac{2}{1 + (\text{sr}[A, B](\bar{x}))^{-2}}} \leq \text{sr}[G](\bar{x}, \bar{x}, 0) \leq \frac{2}{[(\text{sr}[A, B](\bar{x}))^{-1} - 1]_+}. \quad (2.21)$$

Transversality of the collection $\{A, B\}$ at a point $\bar{x} \in A \cap B$ has the following equivalent characterizations.

- (i') The set-valued mapping $F : X \rightrightarrows X^2$ given by (2.18) with the max norm on X^2 is metrically regular at \bar{x} for $(0, 0)$. Moreover,

$$\text{r}[A, B](\bar{x}) = \text{r}[F](\bar{x}, (0, 0)).$$

- (ii') The set-valued mapping $G : X^2 \rightrightarrows X$ given by (2.19) with the l_2 -norm on X^2 is metrically regular at (\bar{x}, \bar{x}) for 0. Moreover,

$$\sqrt{\frac{2}{1 + (\text{r}[A, B](\bar{x}))^{-2}}} \leq \text{r}[G](\bar{x}, \bar{x}, 0) \leq \frac{2}{[(\text{r}[A, B](\bar{x}))^{-1} - 1]_+}. \quad (2.22)$$

THEOREM 2.3.15. [86, Theorem 5.1]

- (i) F is metrically subregular at (\bar{x}, \bar{y}) if and only if the collection of sets $\{A, B\}$ defined at (2.20) is subregular at (\bar{x}, \bar{y}) . Moreover,

$$\frac{\text{sr}[F](\bar{x}, \bar{y})}{\text{sr}[F](\bar{x}, \bar{y}) + 2} \leq \text{sr}[A, B](\bar{x}) \leq \min\{\text{sr}[F](\bar{x}, \bar{y})/2, 1\}.$$

- (ii) F is metrically regular at (\bar{x}, \bar{y}) if and only if the collection of sets $\{A, B\}$ defined at (2.20) is transversal at (\bar{x}, \bar{y}) . Moreover,

$$\frac{\text{r}[F](\bar{x}, \bar{y})}{\text{r}[F](\bar{x}, \bar{y}) + 2} \leq \text{r}[A, B](\bar{x}) \leq \min\{\text{r}[F](\bar{x}, \bar{y})/2, 1\}.$$

REMARK 2.3.16 (collections of sets and set-valued mappings). [84, Remark 10] *The characterizations in Theorem 2.3.14 provide a one-to-one correspondence between regularity properties of collections of sets and the corresponding ones of set-valued mappings. They remain true for arbitrarily finite collections of sets.*

The ‘positive part’ sign in the right-hand sides of the conditions (2.21) and (2.22) is used to accommodate for the case when $\bar{x} \in \text{int } A \cap \text{int } B$ and, hence, $r[A, B](\bar{x}) = +\infty$. In this case, conditions (2.21) and (2.22) impose no upper bound on the values $\text{sr}[G](\bar{x}, \bar{x}, 0)$ and $r[G](\bar{x}, \bar{x}, 0)$.

In view of characterization (ii’), the property of regular intersection of sets considered in [91, Section 5] is equivalent to their collection being transversal. This fact also follows from [91, Theorem 5.1].

Thanks to characterizations of Theorem 2.3.14, when investigating regularity properties of collections of sets one can employ the well developed regularity theory of set-valued mappings, particularly, the celebrated coderivative criterion for metric regularity [50, 109, 129] (see also [77]) as well as criteria of metric subregularity based on outer coderivatives (see [69, 81, 136, 137, 138]). On the other hand, related studies in [80, Theorem 7], [87, Theorem 5.1(ii)] have shown that regularity criteria developed for collections of sets can be used when studying the corresponding properties of set-valued mappings. The coderivatives (Fréchet, limiting or other) of the mappings (2.18) and (2.19) employed in Theorem 2.3.14 admit simple representations in terms of the corresponding normal cones to the sets involved in their definitions; see [80, the proof of Theorem 8], [91, Lemma 5.1], [90, page 491], and [68, Theorem 7.12 and Corollary 7.13]. As a consequence, the coderivative criteria of regularity of set-valued mappings easily translate into the dual characterizations of the corresponding regularity properties of collections of sets. Not surprisingly, this way one rediscovers (some of) the known characterizations collected in Theorem 2.3.14; see [80, Theorem 8], [91, Theorem 5.1], and [68, Theorems 7.12 and 7.15].

2.3.3 Dual characterizations

The dual criterion for the transversality in Asplund spaces is well known, see [78, 79, 80, 85, 87].

THEOREM 2.3.17 (dual characterizations of transversality). [84, Theorem 2 (iii)–(vii)] *The following statements are equivalent to $\{A, B\}$ being transversal at \bar{x} .*

- (i) *There exist numbers $\alpha > 0$ and $\delta > 0$ such that $\|v_1 + v_2\| > \alpha$ for all $a \in A \cap \mathbb{B}_\delta(\bar{x})$, $b \in B \cap \mathbb{B}_\delta(\bar{x})$, $v_1 \in \widehat{N}_A(a)$ and $v_2 \in \widehat{N}_B(b)$ with $\|v_1\| + \|v_2\| = 1$. Moreover, $r[A, B](\bar{x})$ is the exact upper bound of all numbers α above.*
- (ii) *There exists a number $\alpha > 0$ such that $\|v_1 + v_2\| > \alpha$ for all $v_1 \in N_A(a)$ and $v_2 \in N_B(b)$ with $\|v_1\| + \|v_2\| = 1$. Moreover, $r[A, B](\bar{x})$ is the exact upper bound of all such numbers α .*

(iii) $N_A(\bar{x}) \cap (-N_B(\bar{x})) = \{0\}$.

(iv) There is a number $\alpha > 0$ such that $d^2(v, N_A(\bar{x})) + d^2(v, -N_B(\bar{x})) > \alpha^2$ for all $v \in \mathbb{S}$. Moreover, the exact upper bound of all such numbers α , denoted $r_v[A, B](\bar{x})$, satisfies $r_v[A, B](\bar{x}) = \sqrt{2}r[A, B](\bar{x})$.

(v) There is a number $\alpha < 1$ such that $-\langle v_1, v_2 \rangle < \alpha$ for all $v_1 \in N_A(\bar{x})$ and $v_2 \in N_B(\bar{x})$ with $\|v_1\| = \|v_2\| = 1$. Moreover, the exact lower bound of all such numbers α , denoted $r_a[A, B](\bar{x})$, satisfies $r_a[A, B](\bar{x}) + 2(r[A, B](\bar{x}))^2 = 1$.

REMARK 2.3.18. [83, page 705] Characterization (iii) is a well known qualification condition/nonseparability property that has been around for about 30 years under various names, e.g., transversality [40], normal qualification condition [109, 124], linearly regular intersection [90], alliedness property [124], and transversal intersection [51, 68].

REMARK 2.3.19 (characterization (i) and Jameson's property). [84, Remark 6] Characterization (i) in Theorem 2.3.17 can be formulated equivalently as follows:

there exist numbers $\alpha > 0$ and $\delta > 0$ such that $\|v_1 + v_2\| \geq \alpha(\|v_1\| + \|v_2\|)$ for all $a \in A \cap \mathbb{B}_\delta(\bar{x})$, $b \in B \cap \mathbb{B}_\delta(\bar{x})$, $v_1 \in \widehat{N}_A(a)$ and $v_2 \in \widehat{N}_B(b)$.

This characterization can be interpreted as a strengthened version of the *Jameson's property (G)* [71] (cf. [11, 15, 113]). As with all dual characterizations, it basically requires that among all admissible pairs of nonzero normals to the sets there is no pair of normals which are oppositely directed.

REMARK 2.3.20 (characterization (iii) and related notions). [84, Remark 7] Note that, unlike the other characterizations, (iii) provides only a qualitative criterion of transversality. It has the interpretation that the cones $N_A(\bar{x})$ and $N_B(\bar{x})$ are strongly additively regular [36], and has been described as a “concise, fundamental, and widely studied geometric property” [51] extensively used in nonconvex optimization and calculus.

An immediate consequence of characterization (iii) is the following crucial inclusion expressed in terms of the limiting normal cones (cf. [40, page 99], [129, Theorem 6.42], [109, page 142]):

$$N_{A \cap B}(\bar{x}) \subseteq N_A(\bar{x}) + N_B(\bar{x}), \quad (2.23)$$

which can be considered as an extension of the strong conical hull intersection property (strong CHIP) [45] (cf. [11, Definition 5.1(2)]) to nonconvex sets. Indeed, since the opposite inclusion in terms of Fréchet normal cones holds true trivially:

$$\widehat{N}_{A \cap B}(\bar{x}) \supset \widehat{N}_A(\bar{x}) + \widehat{N}_B(\bar{x}), \quad (2.24)$$

and both cones reduce in the convex case to the normal cone (1.4), inclusion (2.23) is equivalent in the convex setting to the strong CHIP:

$$N_{A \cap B}(\bar{x}) = N_A(\bar{x}) + N_B(\bar{x}). \quad (2.25)$$

The last equality has proved to be a fundamental regularity property in several areas of convex optimization; see the discussion of the role of this property (and many other regularity properties of collections of convex sets) in [11, 15]. Inclusion (2.23) plays a similar role in nonconvex optimization and calculus. Thus, thanks to Theorem 2.3.17(iii), transversality implies the extended strong CHIP (2.23).

In fact, it is now well recognized that inclusion (2.23) is ensured by the weaker subtransversality property. The next proposition is a consequence of [69, Proposition 3.2] (or [124, Theorem 6.41]) and the characterization of subtransversality in Theorem 2.3.6(ii).

PROPOSITION 2.3.21. [84, Proposition 5] If $\{A, B\}$ is subtransversal at \bar{x} , then inclusion (2.23) holds true.

In the convex case, a nonlocal version of Proposition 2.3.21 together with certain quantitative estimates can be found in [11, 15].

If a stronger than (2.23) condition (2.25) is satisfied in the nonconvex case (with Fréchet subdifferentials), then this property is referred to in [114] as the strong Fréchet-CHIP. Since inclusion (2.24) always holds, this is equivalent to inclusion (2.23) with Fréchet subdifferentials in place of the limiting ones. A quantitative (by a positive number α) version of the strong Fréchet-CHIP property was studied in the convex and nonconvex settings in [114, 135]:

$$\widehat{N}_{A \cap B}(\bar{x}) \cap \mathbb{B} \subseteq \alpha \left(\left(\widehat{N}_A(\bar{x}) \cap \mathbb{B} \right) + \left(\widehat{N}_B(\bar{x}) \cap \mathbb{B} \right) \right).$$

A number of important links with other regularity properties were established there, and variants of the above property involving Clarke normal cones were also considered.

REMARK 2.3.22 (characterizations in the Euclidean space setting). The following equivalent characterizations of transversality have been established in [85, Theorem 2]).

- (i) There exists a number $\alpha > 0$ such that $\|v_1 + v_2\| > 2\alpha$ for all $v_1 \in N_A(\bar{x})$ and $v_2 \in N_B(\bar{x})$ with $\|v_1\| = \|v_2\| = 1$. Moreover, the exact upper bound of all such numbers α equals $r[A, B](\bar{x})$.
- (ii) There exists a number $\alpha < 1$ such that $\|v_1 - v_2\| < 2\alpha$ for all $v_1 \in N_A(\bar{x})$ and $v_2 \in N_B(\bar{x})$ with $\|v_1\| = \|v_2\| = 1$. Moreover, the exact lower bound of all such numbers α , denoted $r_d[A, B](\bar{x})$, satisfies $(r[A, B](\bar{x}))^2 + (r_d[A, B](\bar{x}))^2 = 1$.

For brevity, the characterizations above are in terms of limiting normals only. The corresponding (approximate) statements in terms of Fréchet and proximal normals can be formulated in a similar way. These characterizations as well as that of Theorem 2.3.17(iv) for the proximal normal cone only hold in Euclidean spaces.

REMARK 2.3.23. [84, Remark 9] Theorem 2.3.17(v) also has analogues in terms of Fréchet and proximal normals. The expression $-\langle v_1, v_2 \rangle$ can be interpreted as the cosine of the angle between the vectors v_1 and $-v_2$. Note that, unlike $r[A, B](\bar{x})$, $r_d[A, B](\bar{x})$, and $r_v[A, B](\bar{x})$, constant $r_a[A, B](\bar{x})$ can be negative. Constant $r_a[A, B](\bar{x})$ is a modification of another one:

$$\bar{c} := \max \{ -\langle v_1, v_2 \rangle \mid v_1 \in N_A(\bar{x}) \cap \mathbb{B}, v_2 \in N_B(\bar{x}) \cap \mathbb{B} \},$$

used in [90] for characterizing transversality. It is easy to check that $\bar{c} = (r_a[A, B](\bar{x}))_+$, and $\bar{c} < 1$ if and only if $r_a[A, B](\bar{x}) < 1$.

The next characterization of transversality following from Theorem 2.3.17(ii) and the relationships among normal cones is needed for the analysis in Section 3.5.

PROPOSITION 2.3.24. [90, Theorem 5.16] Let A and B be two closed sets in \mathbb{E} . Suppose that $\{A, B\}$ is transversal at $\bar{x} \in A \cap B$, or equivalently,

$$\bar{\theta} := \sup \{ \langle u, v \rangle \mid u \in N_A(\bar{x}), v \in N_B(\bar{x}), \|u\| = \|v\| = 1 \} < 1. \quad (2.26)$$

Then for any $\theta \in (\bar{\theta}, 1)$, there exists a $\delta > 0$ such that

$$\left. \begin{array}{l} a \in A \cap B_\delta(\bar{x}), b \in B \cap B_\delta(\bar{x}), \\ u \in N_A^{\text{prox}}(a), v \in N_B^{\text{prox}}(b) \end{array} \right\} \implies \langle u, v \rangle \geq -\theta \|u\| \cdot \|v\|.$$

The next theorem deals with the subtransversality property. It provides a dual sufficient condition for this property in an Asplund space.

THEOREM 2.3.25. [83, Theorem 2] Suppose X is Asplund, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then $\{A, B\}$ is subtransversal at \bar{x} if there exist numbers $\alpha \in (0, 1)$ and $\delta > 0$ such that, for all $a \in (A \setminus B) \cap \mathbb{B}_\delta(\bar{x})$, $b \in (B \setminus A) \cap \mathbb{B}_\delta(\bar{x})$ and $x \in \mathbb{B}_\delta(\bar{x})$ with $\|x - a\| = \|x - b\|$, there exists an $\varepsilon > 0$ such that $\|x_1^* + x_2^*\| > \alpha$ for all $a' \in A \cap \mathbb{B}_\varepsilon(a)$, $b' \in B \cap \mathbb{B}_\varepsilon(b)$, $x'_1 \in \mathbb{B}_\varepsilon(a)$, $x'_2 \in \mathbb{B}_\varepsilon(b)$, $x' \in \mathbb{B}_\varepsilon(x)$, and $x_1^*, x_2^* \in X^*$ satisfying

$$\|x' - x'_1\| = \|x' - x'_2\|, \quad (2.27)$$

$$\|x_1^*\| + \|x_2^*\| = 1, \quad \langle x_1^*, x' - x'_1 \rangle = \|x_1^*\| \|x' - x'_1\|, \quad \langle x_2^*, x' - x'_2 \rangle = \|x_2^*\| \|x' - x'_2\|, \quad (2.28)$$

$$\text{dist}(x_1^*, N_A(a')) < \delta, \quad \text{dist}(x_2^*, N_B(b')) < \delta. \quad (2.29)$$

Moreover, $\text{sr}[A, B](\bar{x}) \geq \alpha$.

In the convex case, one can formulate a necessary and sufficient dual criterion of subtransversality in general Banach spaces which takes a simpler form.

THEOREM 2.3.26. [83, Theorem 3] Suppose X is a Banach space, $A, B \subset X$ are closed and convex, and $\bar{x} \in A \cap B$. Then $\{A, B\}$ is subtransversal at \bar{x} if and only if there exist numbers

$\alpha \in (0, 1)$ and $\delta > 0$ such that $\|x_1^* + x_2^*\| > \alpha$ for all $a \in (A \setminus B) \cap \mathbb{B}_\delta(\bar{x})$, $b \in (B \setminus A) \cap \mathbb{B}_\delta(\bar{x})$, $x \in \mathbb{B}_\delta(\bar{x})$ with $\|x - a\| = \|x - b\|$, and $x_1^*, x_2^* \in X^*$ satisfying

$$\begin{aligned} \|x_1^*\| + \|x_2^*\| &= 1, \quad \langle x_1^*, x - a \rangle = \|x_1^*\| \|x - a\|, \quad \langle x_2^*, x - b \rangle = \|x_2^*\| \|x - b\|, \\ \text{dist}(x_1^*, N_A(a)) &< \delta, \quad \text{dist}(x_2^*, N_B(b)) < \delta. \end{aligned}$$

Moreover, the exact upper bound of all such α equals $\text{sr}[A, B](\bar{x})$.

REMARK 2.3.27. [83, Remark 3]

- (i) It is sufficient to check the conditions of Theorems 2.3.25 and 2.3.26 only for $x_1^* \neq 0$ and $x_2^* \neq 0$. Indeed, if one of the vectors x_1^* and x_2^* equals 0, then by the normalization condition $\|x_1^*\| + \|x_2^*\| = 1$, the norm of the other one equals 1, and consequently $\|x_1^* + x_2^*\| = 1$, i.e., such pairs x_1^*, x_2^* do not impose any restrictions on α .
- (ii) Similarly to the classical characterization (iii) in Theorem 2.3.17 of transversality, the subtransversality characterizations in Theorems 2.3.25 and 2.3.26 require that among all admissible (i.e., satisfying all the conditions of the theorems) pairs of nonzero elements x_1^* and x_2^* there is no one with x_1^* and x_2^* oppositely directed.
- (iii) The sum $\|x_1^*\| + \|x_2^*\|$ in Theorems 2.3.25 and 2.3.26 corresponds to the sum norm on \mathbb{R}^2 , which is dual to the maximum norm on \mathbb{R}^2 used in the definitions of subtransversality. It can be replaced by $\max\{\|x_1^*\|, \|x_2^*\|\}$ (cf. [124, (6.11)]) or any other norm on \mathbb{R}^2 .

The proof of Theorem 2.3.25 follows the sequence proposed in [81] when deducing metric subregularity characterizations for set-valued mappings and consists of a series of propositions providing lower primal and dual estimates for the constant $\text{sr}[A, B](\bar{x})$ and, thus, sufficient conditions for the subtransversality of the pair $\{A, B\}$ at \bar{x} which can be of independent interest.

First observe that constant $\text{sr}[A, B](\bar{x})$ characterizing subtransversality and introduced in Definition 2.3.1 can be written explicitly as

$$\text{sr}[A, B](\bar{x}) = \liminf_{\substack{a \rightarrow \bar{x}, b \rightarrow \bar{x}, x \rightarrow \bar{x} \\ a \in A, b \in B, x \notin A \cap B}} \frac{f(a, b, x)}{\text{dist}(x, A \cap B)} = \liminf_{\substack{a \rightarrow \bar{x}, b \rightarrow \bar{x}, x \rightarrow \bar{x} \\ x \notin A \cap B}} \frac{\hat{f}(a, b, x)}{\text{dist}(x, A \cap B)} \quad (2.30)$$

with the convention that the infimum over the empty set equals 1, and the functions $f : X^3 \rightarrow \mathbb{R}$ and $\hat{f} : X^3 \rightarrow \mathbb{R}_\infty$ defined, respectively, by

$$f(x_1, x_2, x) := \max\{\|x_1 - x\|, \|x_2 - x\|\}, \quad x_1, x_2, x \in X, \quad (2.31)$$

$$\hat{f}(x_1, x_2, x) := f(x_1, x_2, x) + i_{A \times B}(x_1, x_2), \quad x_1, x_2, x \in X, \quad (2.32)$$

where $i_{A \times B}$ is the indicator function of $A \times B$: $i_{A \times B}(x_1, x_2) = 0$ if $x_1 \in A$, $x_2 \in B$ and $i_{A \times B}(x_1, x_2) = +\infty$ otherwise.

Below, we are going to use two different norms on X^3 : a norm depending on a parameter $\rho > 0$ and defined as follows:

$$\|(x_1, x_2, x)\|_\rho := \max\{\|x\|, \rho\|x_1\|, \rho\|x_2\|\}, \quad x_1, x_2, x \in X, \quad (2.33)$$

and the conventional maximum norm $\|(\cdot, \cdot, \cdot)\|$ corresponding to $\rho = 1$ in the above definition; we drop the subscript ρ in this case. It is easy to check that the dual norm corresponding to (2.33) has the following form:

$$\|(x_1^*, x_2^*, x^*)\|_\rho = \|x^*\| + \rho^{-1}(\|x_1^*\| + \|x_2^*\|), \quad x_1^*, x_2^*, x^* \in X^*.$$

The next proposition provides an equivalent primal space representation of the subtransversality constant (2.30).

PROPOSITION 2.3.28. [83, Proposition 7] *Suppose X is a Banach space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then the following representation of the subtransversality constant (2.30) is true:*

$$\text{sr}[A, B](\bar{x}) = \lim_{\rho \downarrow 0} \inf_{\substack{a \in A \cap \mathbb{B}_\rho(\bar{x}), b \in B \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x}), \max\{\|x-a\|, \|x-b\|\} > 0}} \sup_{\substack{a' \in A, b' \in B, u \in X \\ (a', b', u) \neq (a, b, x)}} \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho}, \quad (2.34)$$

with the convention that the infimum over the empty set equals 1.

REMARK 2.3.29. [83, Remark 4]

- (i) *The right-hand side of (2.34) is the uniform strict outer slope [81] of the function (2.32) (considered as a function of two variables x and (x_1, x_2)) at $(\bar{x}, (\bar{x}, \bar{x}))$.*
- (ii) *The inequality ' \leq ' in (2.34) is valid in arbitrary (not necessarily complete) normed linear spaces. The completeness of the space X is only needed for the inequality ' \geq ', the proof of which is based on the application of the Ekeland variational principle.*

The next proposition provides another two primal space representations of the subtransversality constant (2.30) which impose additional restrictions on the choice of a, b and x under the inf in (2.34).

PROPOSITION 2.3.30. [83, Proposition 8] *Suppose X is a Banach space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then the following representations of the subtransversality constant (2.30) are true:*

$$\begin{aligned} \text{sr}[A, B](\bar{x}) &= \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x})}} \sup_{\substack{a' \in A, b' \in B, u \in X \\ (a', b', u) \neq (a, b, x)}} \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho} \\ &= \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x}), \|x-a\| = \|x-b\|}} \sup_{\substack{a' \in A, b' \in B, u \in X \\ (a', b', u) \neq (a, b, x)}} \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho}, \end{aligned} \quad (2.35)$$

with the convention that the infimum over the empty set equals 1.

REMARK 2.3.31. [83, Remark 5] The expression after sup in the right-hand sides of (2.34) and (2.35) can be greater than 1. Nevertheless, $\text{sr}[A, B](\bar{x})$ computed in accordance with (2.34) or (2.35) (under the conventions employed in Propositions 2.3.28 and 2.3.30) is always less than or equal to 1.

Now we define a ‘localized’ subtransversality constant:

$$\text{str}_1[A, B](\bar{x}) := \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), \\ x \in \mathbb{B}_\rho(\bar{x})}} \inf_{\substack{b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ a' \rightarrow a, b' \rightarrow b, u \rightarrow x \\ a' \in A, b' \in B \\ (a', b', u) \neq (a, b, x)}} \limsup \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho}, \quad (2.36)$$

with the convention that the infimum over the empty set equals 1. It corresponds to the first expression in (2.35) with sup replaced by lim sup. Observe that

$$\limsup_{\substack{u \rightarrow x, a' \rightarrow a, b' \rightarrow b \\ a' \in A, b' \in B \\ (a', b', u) \neq (a, b, x)}} \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho}.$$

in the above definition is the ρ -slope [81] (i.e., the slope [9, 44, 52, 65] with respect to the distance in X^3 corresponding to the norm defined by (2.33)) at $(x, (a, b))$ of the function $(u, (a', b')) \mapsto f(a', b', u)$.

PROPOSITION 2.3.32. [83, Proposition 9] Suppose X is a normed linear space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then the following representation of the subtransversality constant (2.36) is true:

$$\text{str}_1[A, B](\bar{x}) = \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), \\ x \in \mathbb{B}_\rho(\bar{x}), \|x - a\| = \|x - b\|}} \inf_{\substack{b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ a' \rightarrow a, b' \rightarrow b, u \rightarrow x \\ a' \in A, b' \in B \\ (a', b', u) \neq (a, b, x)}} \limsup \frac{(f(a, b, x) - f(a', b', u))_+}{\|(a', b', u) - (a, b, x)\|_\rho},$$

with the convention that the infimum over the empty set equals 1.

REMARK 2.3.33. [83, Remark 6] One can define an analogue of $\text{str}_1[A, B](\bar{x})$ using the limiting procedure in the representation of $\text{sr}[A, B](\bar{x})$ in (2.34). Unlike the ‘nonlocal’ case in Propositions 2.3.28 and 2.3.30, such an analogue does not coincide in general with $\text{str}_1[A, B](\bar{x})$ defined by (2.36), although it can still be used for formulating sufficient conditions of subtransversality.

The next proposition clarifies the relationship between $\text{str}_1[A, B](\bar{x})$ and $\text{sr}[A, B](\bar{x})$.

PROPOSITION 2.3.34. [83, Proposition 10] Suppose X is a Banach space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then

$$(i) \text{str}_1[A, B](\bar{x}) \leq \text{sr}[A, B](\bar{x});$$

(ii) if A and B are convex, then (i) holds as equality.

REMARK 2.3.35. [83, Remark 7] Proposition 2.3.21 is valid in arbitrary (not necessarily complete) normed linear spaces if $\text{sr}[A, B](\bar{x})$ is defined by one of the expressions in (2.35) (see Remark 2.3.29(ii)).

To proceed to dual characterizations of subtransversality, we need a representation of the subdifferential of the convex function f given by (2.31).

LEMMA 2.3.36. [83, Lemma 3] Let X be a normed space and f be given by (2.31). Then

$$\partial f(x_1, x_2, x) = \{(x_1^*, x_2^*, -x_1^* - x_2^*) \in (X^*)^3 \mid (x_1^*, x_2^*) \in \partial g(x_1 - x, x_2 - x)\}$$

for all $x_1, x_2, x \in X$, where g is the maximum norm on X^2 :

$$g(x_1, x_2) := \max\{\|x_1\|, \|x_2\|\}, \quad x_1, x_2 \in X.$$

If $x_1 \neq x$ or $x_2 \neq x$, then $(x_1^*, x_2^*, x^*) \in \partial f(x_1, x_2, x)$ if and only if the following conditions are satisfied:

$$\begin{aligned} x_1^* + x_2^* + x^* &= 0, & \|x_1^*\| + \|x_2^*\| &= 1, \\ \langle x_1^*, x_1 - x \rangle &= \|x_1^*\| \|x_1 - x\|, & \langle x_2^*, x_2 - x \rangle &= \|x_2^*\| \|x_2 - x\|, \\ \text{if } \|x_1 - x\| < \|x_2 - x\|, & \text{ then } x_1^* &= 0, \\ \text{if } \|x_2 - x\| < \|x_1 - x\|, & \text{ then } x_2^* &= 0. \end{aligned}$$

The subtransversality constant (2.36) admits dual estimates which are crucial for the conclusions of Theorems 2.3.25 and 2.3.26. In what follows we will use notations $\text{itr}_w[A, B](\bar{x})$ and $\text{itr}_c[A, B](\bar{x})$ for the supremum of all α in Theorems 2.3.25 and 2.3.26, respectively, with the convention that the supremum over the empty set equals 0. It is easy to check the following explicit representations of the two constants:

$$\begin{aligned} \text{itr}_w[A, B](\bar{x}) &:= \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x}), \|x - a\| = \|x - b\|}} \liminf_{\substack{x' \rightarrow x, x'_1 \rightarrow a, x'_2 \rightarrow b, a' \rightarrow a, b' \rightarrow b \\ a' \in A, b' \in B, \|x' - x'_1\| = \|x' - x'_2\| \\ \text{dist}(x_1^*, N_A(a')) < \rho, \text{dist}(x_2^*, N_B(b')) < \rho, \|x_1^*\| + \|x_2^*\| = 1 \\ \langle x_1^*, x' - x'_1 \rangle = \|x_1^*\| \|x' - x'_1\|, \langle x_2^*, x' - x'_2 \rangle = \|x_2^*\| \|x' - x'_2\|}} \|x_1^* + x_2^*\|, \end{aligned} \quad (2.37)$$

$$\begin{aligned} \text{itr}_c[A, B](\bar{x}) &:= \liminf_{\substack{x \rightarrow \bar{x}, a \rightarrow \bar{x}, b \rightarrow \bar{x} \\ a \in A \setminus B, b \in B \setminus A, \|x - a\| = \|x - b\| \\ \text{dist}(x_1^*, N_A(a)) \rightarrow 0, \text{dist}(x_2^*, N_B(b)) \rightarrow 0, \|x_1^*\| + \|x_2^*\| = 1 \\ \langle x_1^*, x - a \rangle = \|x_1^*\| \|x - a\|, \langle x_2^*, x - b \rangle = \|x_2^*\| \|x - b\|}} \|x_1^* + x_2^*\| \end{aligned} \quad (2.38)$$

with the convention that the infimum over the empty set equals 1.

PROPOSITION 2.3.37. [83, Proposition 11] Suppose X is a Banach space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$.

(i) If either X is Asplund or A and B are convex, then the following dual representations of the subtransversality constant (2.36) are true:

$$\begin{aligned} \text{str}_1[A, B](\bar{x}) &= \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x}) \\ (x_1^*, x_2^*, x^*) \in \partial \hat{f}(a, b, x), \|x_1^*\| + \|x_2^*\| < \rho}} \|x^*\| \\ &= \lim_{\rho \downarrow 0} \inf_{\substack{a \in (A \setminus B) \cap \mathbb{B}_\rho(\bar{x}), b \in (B \setminus A) \cap \mathbb{B}_\rho(\bar{x}) \\ x \in \mathbb{B}_\rho(\bar{x}), \|x-a\| = \|x-b\| \\ (x_1^*, x_2^*, x^*) \in \partial \hat{f}(a, b, x), \|x_1^*\| + \|x_2^*\| < \rho}} \|x^*\|, \end{aligned} \quad (2.39)$$

where the function $\hat{f} : X^3 \rightarrow \mathbb{R}_\infty$ is defined by (2.32) and the convention that the infimum over the empty set equals 1 is in force. Moreover,

(ii) if X is Asplund, then $\text{str}_1[A, B](\bar{x}) \geq \text{itr}_w[A, B](\bar{x})$;

(iii) if A and B are convex, then $\text{str}_1[A, B](\bar{x}) = \text{itr}_c[A, B](\bar{x})$.

REMARK 2.3.38. [83, Remark 8] The inequality ' \leq ' in both representations in (2.39) as well as the opposite inequalities in the convex case are valid in arbitrary (not necessarily complete) normed linear spaces.

PROPOSITION 2.3.39. [83, Proposition 12] Suppose X is a Banach space, $A, B \subset X$ are closed and convex, and $\bar{x} \in A \cap B$. Then $\text{sr}[A, B](\bar{x}) = \text{str}_1[A, B](\bar{x}) = \text{itr}_c[A, B](\bar{x})$.

REMARK 2.3.40. [83, Remark 9] Using the representations in Propositions 2.3.28, 2.3.30, 2.3.32 and 2.3.37, one can formulate several intermediate sufficient (and in some cases also necessary) conditions of subtransversality.

The two-limit definition (2.37) as well as the corresponding dual space characterization of subtransversality in Theorem 2.3.25 look complicated and difficult to verify. The following one-limit modification of (2.37) in terms of Fréchet normals can be useful:

$$\begin{aligned} \text{itr}[A, B](\bar{x}) := & \liminf_{\substack{a \rightarrow \bar{x}, b \rightarrow \bar{x}, x \rightarrow \bar{x} \\ a \in A \setminus B, b \in B \setminus A, x \neq a, x \neq b \\ x_1^* \in N_A(a) \setminus \{0\}, x_2^* \in N_B(b) \setminus \{0\}, \|x_1^*\| + \|x_2^*\| = 1 \\ \frac{\|x-a\|}{\|x-b\|} \rightarrow 1, \frac{\langle x_1^*, x-a \rangle}{\|x_1^*\| \|x-a\|} \rightarrow 1, \frac{\langle x_2^*, x-b \rangle}{\|x_2^*\| \|x-b\|} \rightarrow 1}} \|x_1^* + x_2^*\| \end{aligned} \quad (2.40)$$

with the convention that the infimum over the empty set equals 1. The relationship between the constants (2.37), (2.38) and (2.40) is given by the next proposition.

PROPOSITION 2.3.41. [83, Proposition 13] Suppose X is a Banach space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$.

$$(i) \quad 0 \leq \text{itr}[A, B](\bar{x}) \leq \text{itr}_w[A, B](\bar{x}) \leq \text{itr}_c[A, B](\bar{x}) \leq 1;$$

(ii) if $\dim X < \infty$, then

$$\text{itr}_w[A, B](\bar{x}) = \liminf_{\substack{a \rightarrow \bar{x}, b \rightarrow \bar{x}, x \rightarrow \bar{x} \\ a \in A \setminus B, b \in B \setminus A, \|x-a\| = \|x-b\| \\ \text{dist}(x_1^*, \bar{N}_A(a)) \rightarrow 0, \text{dist}(x_2^*, \bar{N}_B(b)) \rightarrow 0, \|x_1^*\| + \|x_2^*\| = 1 \\ \langle x_1^*, x-a \rangle = \|x_1^*\| \|x-a\|, \langle x_2^*, x-b \rangle = \|x_2^*\| \|x-b\|}} \|x_1^* + x_2^*\|$$

with the convention that the infimum over the empty set equals 1;

(iii) if $\dim X < \infty$, and A and B are convex, then

$$\text{itr}_w[A, B](\bar{x}) = \text{itr}_c[A, B](\bar{x}) = \text{sr}[A, B](\bar{x}).$$

The property introduced in Theorem 2.3.25 as a sufficient dual space characterization of subtransversality and corresponding to the condition $\text{itr}_w[A, B](\bar{x}) > 0$ as well as the stronger property corresponding to the condition $\text{itr}[A, B](\bar{x}) > 0$ are themselves important transversality properties of the pair $\{A, B\}$ at \bar{x} . Borrowing partially the terminology from [51], we are going to call these properties *weak intrinsic transversality* and *intrinsic transversality*, respectively.

DEFINITION 2.3.42. [83, Definition 4] Suppose X is a normed linear space, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. The pair $\{A, B\}$ is

(i) *weakly intrinsically transversal at \bar{x}* if $\text{itr}_w[A, B](\bar{x}) > 0$, i.e., there exist numbers $\alpha \in (0, 1)$ and $\delta > 0$ such that, for all $a \in (A \setminus B) \cap \mathbb{B}_\delta(\bar{x})$, $b \in (B \setminus A) \cap \mathbb{B}_\delta(\bar{x})$ and $x \in \mathbb{B}_\delta(\bar{x})$ with $\|x - a\| = \|x - b\|$, one has $\|x_1^* + x_2^*\| > \alpha$ for some $\varepsilon > 0$ and all $a' \in A \cap \mathbb{B}_\varepsilon(a)$, $b' \in B \cap \mathbb{B}_\varepsilon(b)$, $x_1' \in \mathbb{B}_\varepsilon(a)$, $x_2' \in \mathbb{B}_\varepsilon(b)$, $x' \in \mathbb{B}_\varepsilon(x)$, and $x_1^*, x_2^* \in X^*$ satisfying conditions (2.27), (2.28) and (2.29);

(ii) *intrinsically transversal at \bar{x}* if $\text{itr}[A, B](\bar{x}) > 0$, i.e., there exist numbers $\alpha \in (0, 1)$ and $\delta > 0$ such that $\|x_1^* + x_2^*\| > \alpha$ for all $a \in (A \setminus B) \cap \mathbb{B}_\delta(\bar{x})$, $b \in (B \setminus A) \cap \mathbb{B}_\delta(\bar{x})$, $x \in \mathbb{B}_\delta(\bar{x})$, $x_1^* \in N_A(a) \setminus \{0\}$ and $x_2^* \in N_B(b) \setminus \{0\}$ satisfying

$$x \neq a, \quad x \neq b, \quad 1 - \delta < \frac{\|x - a\|}{\|x - b\|} < 1 + \delta, \quad (2.41)$$

$$\|x_1^*\| + \|x_2^*\| = 1, \quad \frac{\langle x_1^*, x - a \rangle}{\|x_1^*\| \|x - a\|} > 1 - \delta, \quad \frac{\langle x_2^*, x - b \rangle}{\|x_2^*\| \|x - b\|} > 1 - \delta. \quad (2.42)$$

REMARK 2.3.43. The properties introduced in Definition 2.3.42 are less restrictive than the dual criterion of transversality in Theorem 2.3.17.

In view of Definition 2.3.42, Theorem 2.3.25 says that in Asplund spaces weak intrinsic transversality (and consequently intrinsic transversality) implies subtransversality. Thanks to Proposition 2.3.41(i) and Remark 2.3.43, we have the following chain of implications in Asplund spaces:

$$\begin{aligned} \text{transversality} &\implies \text{intrinsic transversality} \\ &\implies \text{weak intrinsic transversality} \implies \text{subtransversality.} \end{aligned}$$

By Proposition 2.3.41(iii), when the space is finite dimensional and the sets are convex, the last two properties are equivalent.

For a further discussion on the intrinsic transversality, we would like to refer to the paper [51, 82].

As a consequence of Proposition 2.3.41(i), we obtain the following dual sufficient condition of subtransversality of a pair of closed sets in an Asplund space.

COROLLARY 2.3.44. [83, Corollary 2] *Suppose X is Asplund, $A, B \subset X$ are closed, and $\bar{x} \in A \cap B$. Then $\{A, B\}$ is subtransversal at \bar{x} if there exist numbers $\alpha \in (0, 1)$ and $\delta > 0$ such that $\|x_1^* + x_2^*\| > \alpha$ for all $a \in (A \setminus B) \cap \mathbb{B}_\delta(\bar{x})$, $b \in (B \setminus A) \cap \mathbb{B}_\delta(\bar{x})$, $x \in \mathbb{B}_\delta(\bar{x})$, $x_1^* \in N_A(a) \setminus \{0\}$ and $x_2^* \in N_B(b) \setminus \{0\}$ satisfying (2.41) and (2.42).*

2.3.4 Special cases: convex sets, cones and manifolds

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} .

A number of simplifications are possible in the convex setting, for cones and for manifolds.

The next representations follow from the simplified representations for $r[A, B](\bar{x})$ that are possible for convex sets or cones (cf. [78, Propositions 13 and 15]).

PROPOSITION 2.3.45 (collections of convex sets). [84, Proposition 6] *Suppose A and B are convex. The collection $\{A, B\}$ is transversal at \bar{x} if and only if one of the next two equivalent conditions holds true:*

(i) *there exists a number $\alpha > 0$ such that*

$$(A - x_1) \cap (B - x_2) \cap \mathbb{B}_\rho(\bar{x}) \neq \emptyset \tag{2.43}$$

for all $\rho > 0$ and all $x_1, x_2 \in \mathbb{E}$ with $\max\{\|x_1\|, \|x_2\|\} < \alpha\rho$;

(ii) *there exists a number $\alpha > 0$ such that condition (2.43) is satisfied for some $\rho > 0$ and all $x_1, x_2 \in \mathbb{E}$ with $\max\{\|x_1\|, \|x_2\|\} < \alpha\rho$.*

Moreover, the exact upper bound of all numbers α in any of the above conditions equals $r[A, B](\bar{x})$.

PROPOSITION 2.3.46 (cones). [84, Proposition 7] Suppose A and B are cones. The collection $\{A, B\}$ is transversal at 0 if and only if there exists a number $\alpha > 0$ such that

$$(A - a - x_1) \cap (B - b - x_2) \cap \mathbb{B} \neq \emptyset$$

for all $a \in A$, $b \in B$ and all $x_1, x_2 \in \mathbb{E}$ with $\max\{\|x_1\|, \|x_2\|\} < \alpha$. Moreover, the exact upper bound of all numbers α in any of the above conditions equals $r[A, B](0)$.

In the case when A and B are smooth manifolds, one can deduce the Friedrichs angle characterization of transversality established in [91, Theorem 5.2].

PROPOSITION 2.3.47 (manifolds). [84, Proposition 8] Let A and B be smooth manifolds around a point $\bar{x} \in A \cap B$. Then

$$r_a[A, B](\bar{x}) = c(A, B, \bar{x}),$$

$c(A, B, \bar{x})$ is the Friedrichs angle between the two normal spaces $N_A(\bar{x})$ $N_B(\bar{x})$.

REMARK 2.3.48. [84, Remark 14] Some sufficient and also necessary characterizations of the subtransversality property in terms of the Fréchet subdifferentials of the function $x \mapsto \text{dist}(x, A) + \text{dist}(x, B)$ were formulated [117, Theorem 3.1].

The next example illustrates the computation of the constants characterizing regularity.

EXAMPLE 2.3.49. [84, Example 6] Let $\mathbb{E} = \mathbb{R}^2$, $A = \mathbb{R} \times \{0\}$, $B = \{(t, t) \mid t \in \mathbb{R}\}$, $\bar{x} = (0, 0)$. A and B are linear subspaces. We have $A \cap B = \{(0, 0)\}$, $T_A(\bar{x}) = A$, $T_B(\bar{x}) = B$, $T_{A \cap B}(\bar{x}) = \{(0, 0)\}$, $N_A(\bar{x}) = A^\perp = \{0\} \times \mathbb{R}$, $N_B(\bar{x}) = B^\perp = \{(t, -t) \mid t \in \mathbb{R}\}$. The collection $\{A, B\}$ is transversal at \bar{x} in the classical sense and, thanks to Proposition 2.3.47, also transversal at \bar{x} in the sense of Definition 2.3.1(ii). By the representations in Theorem 2.3.8(i)-(v), after performing some simple computations, we obtain:

$$\begin{aligned} r[A, B](\bar{x}) &= \frac{1}{2} \left\| \left(\frac{1}{\sqrt{2}} - 1, \frac{1}{\sqrt{2}} \right) \right\| = t_2, \\ r_d[A, B](\bar{x}) &= \frac{1}{2} \left\| \left(\frac{1}{\sqrt{2}} + 1, \frac{1}{\sqrt{2}} \right) \right\| = t_1, \\ r_v[A, B](\bar{x}) &= \sqrt{d^2((t_1, t_2), A) + d^2((t_1, t_2), B)} \\ &= \sqrt{\|(t_1, t_2) - (t_1, 0)\|^2 + \left\| (t_1, t_2) - \left(\frac{t_1 + t_2}{2}, \frac{t_1 + t_2}{2} \right) \right\|^2} = t_2 \sqrt{2}, \\ r_a[A, B](\bar{x}) &= \left\langle \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), (1, 0) \right\rangle = \frac{1}{\sqrt{2}}, \end{aligned}$$

where $t_1 := \frac{\sqrt{2+\sqrt{2}}}{2}$ and $t_2 := \frac{\sqrt{2-\sqrt{2}}}{2}$. It is easy to check that all the relations in Theorem 2.3.8(i)-(v) are satisfied.

Chapter 3

Convergence analysis

In recent years there has been a tremendous interest in *first-order methods* for solving variational problems. As the name suggests, these methods only use information that, in some way, encodes the gradient of a function to be minimized. Often one has in mind the following universal optimization problem for such methods

$$\underset{x \in \mathbb{E}}{\text{minimize}} \sum_{j=1}^m f_j(x) \tag{3.1}$$

where f_j are scalar extended-valued functions, not necessarily smooth or convex, on a Hilbert space. This specializes to constrained optimization in the case that one or more of the functions f_j is an indicator function for a set.

Based on the knowledge of regularity notions discussed in Chapter 2, several abstract programs of analysis are studied in this chapter. As consequences, a number of convergence results are derived for a variety of projection algorithms for solving the *feasibility* problem

$$\text{find } \bar{x} \in \bigcap_{j=1}^m A_j,$$

which is the specialization of (3.1) to the case

$$f_j(x) = \iota_{A_j}(x) := \begin{cases} 0 & \text{if } x \in A_j \\ +\infty & \text{else} \end{cases} \quad (j = 1, 2, \dots, m).$$

3.1 Abstract convergence of Picard iterations

Regarding the underlying space in this section, \mathbb{E} stands for a Euclidean space while \mathcal{H} stands for an infinite dimensional space. The content of this section is taken from our joint papers with Dr. Matthew K. Tam [103, 102].

The next theorem serves as the basic template for the quantitative convergence analysis of fixed point iterations and generalizes [59, Lemma 3.1]. By the notation $T : \Lambda \rightrightarrows \Lambda$ where Λ is a subset or an affine subspace of \mathbb{E} , we mean that $T : \mathbb{E} \rightrightarrows \mathbb{E}$ and $T(x) \subset \Lambda$ for all $x \in \Lambda$. This simplification of notation should not lead to any confusion if one keeps in mind that there may exist fixed points of T that are not in Λ . For the importance of the use of Λ in isolating the desirable fixed point, we refer the reader to [4, Example 1.8].

THEOREM 3.1.1. [103, Theorem 2.1] *Let $T : \Lambda \rightrightarrows \Lambda$ for $\Lambda \subset \mathbb{E}$ and let $S \subset \text{ri } \Lambda$ be closed and nonempty with $Ty \subset \text{Fix } T \cap S$ for all $y \in S$. Let \mathcal{O} be a neighborhood of S such that $\mathcal{O} \cap \Lambda \subset \text{ri } \Lambda$. Suppose*

- (a) *T is pointwise almost averaged at all points $y \in S$ with violation ε and averaging constant $\alpha \in (0, 1)$ on $\mathcal{O} \cap \Lambda$, and*
- (b) *there exists a neighborhood \mathcal{V} of $\text{Fix } T \cap S$ and a $\kappa > 0$, such that for all $y^+ \in Ty$, $y \in S$, and all $x^+ \in Tx$ the estimate*

$$\text{dist}(x, S) \leq \kappa \| (x - x^+) - (y - y^+) \| \quad (3.2)$$

holds true whenever $x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$.

Then for all $x^+ \in Tx$

$$\text{dist}(x^+, \text{Fix } T \cap S) \leq \sqrt{1 + \varepsilon - \frac{1 - \alpha}{\kappa^2 \alpha}} \text{dist}(x, S) \quad (3.3)$$

whenever $x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$.

In particular, if $\kappa < \sqrt{\frac{1 - \alpha}{\varepsilon \alpha}}$, then for all $x^0 \in \mathcal{O} \cap \Lambda$ the iteration $x^{j+1} \in Tx^j$ satisfies

$$\text{dist}(x^{j+1}, \text{Fix } T \cap S) \leq c^j \text{dist}(x^0, S)$$

with $c := (1 + \varepsilon - \frac{1 - \alpha}{\alpha \kappa^2})^{1/2} < 1$ for all j such that $x^i \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$ for $i = 1, 2, \dots, j$.

Some remarks will help clarify the technicalities. The role of assumption (a) is clear in the two-property scheme we have set up. The second assumption (b) is a characterization of the required stability of the fixed points and their preimages. It is helpful to consider a specialization of this assumption which simplifies things considerably. First, by Proposition 1.3.6, since T is almost averaged at all points in S , then it is single-valued there and one can simply write Ty for all $y \in S$ instead of $y^+ \in Ty$. The real simplification comes when one considers the case $S = \text{Fix } T$. In this case $Ty = y$ for all $y \in S$ and condition (3.2) simplifies to

$$\text{dist}(x, \text{Fix } T) \leq \kappa \text{dist}(0, x - Tx) \quad \iff \quad \text{dist}(x, F^{-1}(0)) \leq \kappa \text{dist}(0, F(x))$$

for all $x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$ where $F := T - \text{Id}$. The statement on annular regions $(\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$ can be viewed as an assumption about the existence of an *error bound* on that region. For earlier manifestations of this and connections to previous work on error bounds see [104] and [116]. In the present context, this condition has been identified in Section 2.2 with *metric subregularity* of F .

The assumptions lead to the conclusion that the iterates approach the set of fixed points at some rate that can be bounded below by a linear characterization on the region $(\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$. This will lead to *convergence* in Corollary 3.1.2 where on all such annular regions there is some lower linear convergence bound.

The possibility to have $S \subset \text{Fix } T$ and not $S = \text{Fix } T$ allows one to sidestep complications arising from the not-so-exotic occurrence of fixed point mappings that are almost nonexpansive at some points in $\text{Fix } T$ and not at others (see Example 1.3.5(ii)). It would be too restrictive in the statement of the theorem, however, to have $S \subseteq \text{Fix } T$, since this does not allow one to tackle inconsistent feasibility, studied in depth in Section 3.2. In particular, we have in mind the situation where sets A and B do not intersect, but still the alternating projections mapping $T_{AP} := P_A P_B$ has nice properties at points in B that, while not fixed points, at least locally are nearest to A . The full richness of the structure is used in Theorem 3.2.7 where we establish, for the first time, sufficient conditions for local linear convergence of the method of cyclic projections for nonconvex inconsistent feasibility.

The inequality (3.3) by itself says nothing about convergence of the iteration $x^{j+1} = Tx^j$, but it does clearly indicate what needs to hold in order for the iterates to move closer to a fixed point of T . This is stated explicitly in the next corollary.

COROLLARY 3.1.2 (convergence). [103, Corollary 2.2] *Let $T : \Lambda \rightrightarrows \Lambda$ for $\Lambda \subset \mathbb{E}$ and let $S \subset \text{ri } \Lambda$ be closed and nonempty with $T\bar{x} \subset \text{Fix } T \cap S$ for all $\bar{x} \in S$. Define $\mathcal{O}_\delta := S + \delta\mathbb{B}$ and $\mathcal{V}_\delta := \text{Fix } T \cap S + \delta\mathbb{B}$. Suppose that for $\gamma \in (0, 1)$ fixed and for all $\bar{\delta} > 0$ small enough, there is a triplet $(\varepsilon, \delta, \alpha) \in \mathbb{R}_+ \times (0, \gamma\bar{\delta}) \times (0, 1)$ such that*

(a) *T is pointwise almost averaged at all $y \in S$ with violation ε and averaging constant α on $\mathcal{O}_{\bar{\delta}} \cap \Lambda$, and*

(b) *at each $y^+ \in Ty$ for all $y \in S$ there exists a $\kappa \in \left[0, \sqrt{\frac{1-\alpha}{\varepsilon\alpha}}\right)$ such that*

$$\text{dist}(x, S) \leq \kappa \| (x - x^+) - (y - y^+) \|$$

at each $x^+ \in Tx$ for all $x \in (\mathcal{O}_{\bar{\delta}} \cap \Lambda) \setminus (\mathcal{V}_{\bar{\delta}} \cap \Lambda)$.

Then for any x^0 close enough to S the iterates $x^{i+1} \in Tx^i$ satisfy $\text{dist}(x^i, \text{Fix } T \cap S) \rightarrow 0$ as $i \rightarrow \infty$.

An interesting avenue of investigation would be to see to what extent the proof mining techniques of [76] could be applied to quantify convergence in the present setting.

THEOREM 3.1.3 ((sub)linear convergence with metric regularity). [103, Theorem 2.2] Let $T : \Lambda \rightrightarrows \Lambda$ for $\Lambda \subset \mathbb{E}$, $F := T - \text{Id}$ and let $S \subset \text{ri} \Lambda$ be closed and nonempty with $TS \subset \text{Fix } T \cap S$. Denote $(S + \delta\mathbb{B}) \cap \Lambda$ by S_δ for a nonnegative real δ . Suppose that, for all $\bar{\delta} > 0$ small enough, there are $\gamma \in (0, 1)$, a nonnegative sequence of scalars $(\varepsilon_i)_{i \in \mathbb{N}}$ and a sequence of positive constants α_i bounded above by $\bar{\alpha} < 1$, such that, for each $i \in \mathbb{N}$,

(a) T is pointwise almost averaged at all $y \in S$ with averaging constant α_i and violation ε_i on $S_{\gamma^i \bar{\delta}}$, and

(b) for

$$R_i := S_{\gamma^i \bar{\delta}} \setminus (\text{Fix } T \cap S + \gamma^{i+1} \bar{\delta} \mathbb{B}),$$

(i) $\text{dist}(x, S) \leq \text{dist}(x, F^{-1}(\bar{y}) \cap \Lambda)$ for all $x \in R_i$ and $\bar{y} \in F(P_S(x)) \setminus F(x)$,

(ii) F is metrically regular with gauge μ_i relative to Λ on $R_i \times F(P_S(R_i))$, where μ_i satisfies

$$\sup_{x \in R_i, \bar{y} \in F(P_S(R_i)), \bar{y} \notin F(x)} \frac{\mu_i(\text{dist}(\bar{y}, F(x)))}{\text{dist}(\bar{y}, F(x))} \leq \kappa_i < \sqrt{\frac{1 - \alpha_i}{\varepsilon_i \alpha_i}}. \quad (3.4)$$

Then, for any $x^0 \in \Lambda$ close enough to S , the iterates $x^{j+1} \in Tx^j$ satisfy $\text{dist}(x^j, \text{Fix } T \cap S) \rightarrow 0$ and

$$\text{dist}(x^{j+1}, \text{Fix } T \cap S) \leq c_i \text{dist}(x^j, S) \quad \forall x^j \in R_i,$$

where $c_i := \sqrt{1 + \varepsilon_i - \left(\frac{1 - \alpha_i}{\kappa_i^2 \alpha_i}\right)} < 1$.

In particular, if ε_i is bounded above by $\bar{\varepsilon}$ and $\kappa_i \leq \bar{\kappa} < \sqrt{\frac{1 - \bar{\alpha}}{\bar{\alpha} \bar{\varepsilon}}}$ for all i large enough, then convergence is eventually at least linear with rate at most $\bar{c} := \sqrt{1 + \bar{\varepsilon} - \left(\frac{1 - \bar{\alpha}}{\bar{\kappa}^2 \bar{\alpha}}\right)} < 1$.

The first inequality in (3.4) is a condition on the gauge function μ_i and would not be needed if the statement were limited to linearly metrically regular mappings. Essentially, it says that the gauge function characterizing metric regularity of F can be bounded above by a linear function. The second inequality states that the constant of metric regularity κ_i is small enough relative to the violation of the averaging property ε_i to guarantee a linear progression of the iterates through the region R_i .

When $S = \text{Fix } T \cap \Lambda$ in Theorem 3.1.3, the condition (b)(i) can be dropped from the assumptions, as the next corollary shows.

COROLLARY 3.1.4. [103, Corollary 2.3] Let $T : \Lambda \rightrightarrows \Lambda$ for $\Lambda \subset \mathbb{E}$ with $\text{Fix } T$ nonempty and closed, $F := T - \text{Id}$. Denote $(\text{Fix } T + \delta\mathbb{B}) \cap \Lambda$ by S_δ for a nonnegative real δ . Suppose that, for all $\bar{\delta} > 0$ small enough, there are $\gamma \in (0, 1)$, a nonnegative sequence of scalars $(\varepsilon_i)_{i \in \mathbb{N}}$ and a sequence of positive constants α_i bounded above by $\bar{\alpha} < 1$, such that, for each $i \in \mathbb{N}$,

(a) T is pointwise almost averaged at all $y \in \text{Fix } T \cap \Lambda$ with averaging constant α_i and violation ε_i on $S_{\gamma^i \bar{\delta}}$, and

(b) for

$$R_i := S_{\gamma^i \bar{\delta}} \setminus (\text{Fix } T + \gamma^{i+1} \bar{\delta} \mathbb{B}),$$

F is metrically subregular for 0 on R_i (metrically regular on $R_i \times \{0\}$) with gauge μ_i relative to Λ , where μ_i satisfies

$$\sup_{x \in R_i} \frac{\mu_i(\text{dist}(0, F(x)))}{\text{dist}(0, F(x))} \leq \kappa_i < \sqrt{\frac{1 - \alpha_i}{\varepsilon_i \alpha_i}}.$$

Then, for any $x^0 \in \Lambda$ close enough to $\text{Fix } T \cap \Lambda$, the iterates $x^{j+1} \in Tx^j$ satisfy

$$\text{dist}(x^j, \text{Fix } T \cap \Lambda) \rightarrow 0$$

and

$$\text{dist}(x^{j+1}, \text{Fix } T \cap \Lambda) \leq c_i \text{dist}(x^j, \text{Fix } T \cap \Lambda) \quad \forall x^j \in R_i,$$

where $c_i := \sqrt{1 + \varepsilon_i - \left(\frac{1 - \alpha_i}{\kappa_i^2 \alpha_i}\right)} < 1$.

In particular, if ε_i is bounded above by $\bar{\varepsilon}$ and $\kappa_i \leq \bar{\kappa} < \sqrt{\frac{1 - \bar{\alpha}}{\bar{\alpha} \bar{\varepsilon}}}$ for all i large enough, then convergence is eventually at least linear with rate at most $\bar{c} := \sqrt{1 + \bar{\varepsilon} - \left(\frac{1 - \bar{\alpha}}{\bar{\kappa}^2 \bar{\alpha}}\right)} < 1$.

The following example explains why gauge metric regularity on a set (Definition 2.2.1) fits well in the framework of Theorem 3.1.3, whereas the conventional metric (sub)regularity does not.

EXAMPLE 3.1.5 (a line tangent to a circle). [103, Example 2.4] In \mathbb{R}^2 , consider the two sets

$$\begin{aligned} A &:= \{(u, -1) \in \mathbb{R}^2 \mid u \in \mathbb{R}\}, \\ B &:= \{(u, v) \in \mathbb{R}^2 \mid u^2 + v^2 = 1\}, \end{aligned}$$

and the point $\bar{x} = (0, -1)$. It is well known that the alternating projections algorithm $T := P_A P_B$ does not converge linearly to \bar{x} unless with the starting points on $\{(0, v) \in \mathbb{R}^2 : v \in \mathbb{R}\}$ (in this special case, the method reaches \bar{x} in one step). Note that T behaves the same if B is replaced by the closed unit ball (the case of two closed convex sets). In particular, T is averaged with constant $\alpha = 2/3$ by Proposition 1.3.10(iii). Hence, the absence of linear convergence of T here can be explained as the lack of regularity of the fixed point set $A \cap B = \{\bar{x}\}$. In fact, the mapping $F := T - \text{Id}$ is not (linearly) metrically subregular at \bar{x} for 0 on any set $\mathbb{B}_\delta(\bar{x})$, for any $\delta > 0$. However, T does converge sublinearly to \bar{x} . This can be characterized in two different ways.

- Using Corollary 3.1.4, we characterize sublinear convergence in this example as linear convergence on annular sets. To proceed, we set

$$R_i := \mathbb{B}_{2^{-i}}(\bar{x}) \setminus \mathbb{B}_{2^{-(i+1)}}(\bar{x}), \quad (i = 0, 1, \dots).$$

This corresponds to setting $\bar{\delta} = 1$ and $\gamma = 1/2$ in Corollary 3.1.4. The task that remains is to estimate the constant of metric subregularity, κ_i , of F on each R_i . Indeed, we have

$$\inf_{x \in R_i \cap A} \frac{\|x - Tx\|}{\|x - \bar{x}\|} = \frac{\|x^* - Tx^*\|}{\|x^* - \bar{x}\|} = 1 - \frac{1}{\sqrt{2^{-2(i+1)} + 1}} := \kappa_i > 0, \quad (i = 0, 1, \dots),$$

where $x^* = (2^{-(i+1)}, -1)$.

Hence, on each ring R_i , T converges linearly to a point in $\mathbb{B}_{2^{-(i+1)}}(\bar{x})$ with rate c_i not worse than $\sqrt{1 - 1/(2\kappa_i^2)} < 1$ by Corollary 3.1.4.

- The discussion above uses the linear gauge functions $\mu_i(t) := \frac{t}{\kappa_i}$ on annular regions, and hence a piecewise linear gauge function for the characterization of metric subregularity. Alternatively, we can construct a smooth gauge function μ that works on neighborhoods of the fixed point. For analyzing convergence of P_AP_B , we must have F metrically subregular at 0 with gauge μ on \mathbb{R}^2 relative to A . But we have

$$\text{dist}(0, F(x)) = \|x - x^+\| = f(\|x - \bar{x}\|) = f(\text{dist}(x, F^{-1}(0))), \quad \forall x \in A, \quad (3.5)$$

where $f : [0, \infty) \rightarrow [0, \infty)$ is given by $f(t) := t \left(1 - 1/\sqrt{t^2 + 1}\right)$. The function f is continuous strictly increasing and satisfies $f(0) = 0$ and $\lim_{t \rightarrow \infty} f(t) = \infty$. Hence, f is a gauge function.

We can now characterize sublinear convergence of P_AP_B explicitly without resorting to annular sets. Note first that since $f(t) < t$ for all $t \in (0, \infty)$ the function $g : [0, \infty) \rightarrow [0, \infty)$ given by

$$g(t) := \sqrt{t^2 - \frac{1}{2}(f(t))^2}$$

is a gauge function and satisfies $g(t) < t$ for all $t \in (0, \infty)$. Note next that $T := P_AP_B$ is (for all points in A) averaged with constant $2/3$ together with (3.5), we get for any $x \in A$

$$\begin{aligned} \|x^+ - \bar{x}\|^2 &\leq \|x - \bar{x}\|^2 - (1/2) \|x - x^+\|^2 \\ &= \|x - \bar{x}\|^2 - (1/2) (f(\|x - \bar{x}\|))^2. \end{aligned}$$

This implies

$$\begin{aligned} \text{dist}(x^+, S) = \|x^+ - \bar{x}\| &\leq \sqrt{\|x - \bar{x}\|^2 - (1/2) (f(\|x - \bar{x}\|))^2} \\ &= g(\|x - \bar{x}\|) = g(\text{dist}(x, S)), \quad \forall x \in A. \end{aligned}$$

△

REMARK 3.1.6 (global (sub)linear convergence of pointwise averaged mappings). [103, Remark 2.2] As Example 3.1.5 illustrates, Theorem 3.1.3 is not an asymptotic result and does not gainsay the possibility that the required properties hold with neighborhood $U = \mathbb{E}$, which would then lead to a global quantification of convergence. First order methods for convex problems lead generically to globally averaged fixed point mappings T . Convergence for convex problems can be determined from the averaging property of T and existence of fixed points. Hence in order to quantify convergence the only thing to be determined is the gauge of metric regularity at the fixed points of T . In this context, see [28]. Example 3.1.5 illustrates how this can be done. This instance will be revisited in Example 3.2.11.

PROPOSITION 3.1.7 (local linear convergence: polyhedral fixed point iterations). [103, Proposition 2.7] Let $\Lambda \subset \mathbb{E}$ be an affine subspace and $T : \Lambda \rightrightarrows \Lambda$ be pointwise almost averaged at $\{\bar{x}\} = \text{Fix } T \cap \Lambda$ on Λ with violation constant ε and averaging constant α . If T is polyhedral, then there is a neighborhood U of \bar{x} such that

$$\|x^+ - \bar{x}\| \leq c \|x - \bar{x}\| \quad \forall x \in U \cap \Lambda, \quad x^+ \in Tx,$$

where $c = \sqrt{1 + \varepsilon - \frac{1-\alpha}{\kappa^2\alpha}}$ and κ is the modulus of metric subregularity of $F := T - \text{Id}$ for 0 on U relative to Λ . If, in addition $\kappa < \sqrt{(1-\alpha)/(\alpha\varepsilon)}$, then the fixed point iteration $x^{j+1} \in Tx^j$ converges linearly to \bar{x} with rate $c < 1$ for all $x^0 \in U \cap \Lambda$.

In what follows, the n -fold composition of a function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is denoted

$$\varphi^n := \underbrace{\varphi \circ \dots \circ \varphi}_{n \text{ times}}.$$

THEOREM 3.1.8 (error bound estimate for convergence rate). [102, Theorem 2] Let D be a nonempty closed convex subset of \mathcal{H} and let $T : D \rightarrow D$ be averaged with $\text{Fix } T \neq \emptyset$. Suppose that, on each bounded subset U of D , there exists a gauge function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that condition

$$\text{dist}(x, \text{Fix } T) \leq \kappa(\|x - Tx\|) \quad \forall x \in U.$$

is satisfied and

$$\lim_{n \rightarrow \infty} \varphi^n(t) = 0 \quad \forall t \geq 0 \quad \text{where} \quad \varphi(t) := \sqrt{t^2 - \gamma \kappa^{-1}(t)^2}.$$

For any $x_0 \in D$, define $x_{n+1} := Tx_n$ for all $n \in \mathbb{N}$. Then $x_n \rightarrow x^* \in \text{Fix } T$ and

$$\|x_n - x^*\| \leq 2\varphi^n(\text{dist}(x_0, \text{Fix } T)) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In other words, (x_n) converges strongly to x^* with rate no worse than the rate at which $\varphi^n(\text{dist}(x_0, \text{Fix } T)) \searrow 0$.

REMARK 3.1.9. [102, Remark 1] We discuss some important special cases of Theorem 3.1.8.

- (i) (linear regularity). The setting in which κ is linear (i.e., $\kappa(t) = Kt$ for some $K > 0$) corresponds to bounded linear regularity of T as discussed in [23, 89]. In this case, $\kappa^{-1}(t) = t/K$ and so

$$\varphi(t) = \sqrt{t^2 - \gamma \frac{t^2}{K^2}} = t \sqrt{1 - \frac{\gamma}{K^2}} \implies \varphi^n(t) = t \left(\sqrt{1 - \frac{\gamma}{K^2}} \right)^n.$$

Theorem 3.1.8 implies R -linear convergence with rate no worse than $\sqrt{1 - \frac{\gamma}{K^2}}$ which recovers the single operator specialization of [23].

- (ii) (Hölder regularity). The case in which κ is a ‘‘Hölder-type function’’ (i.e., $\kappa(t) = Kt^\tau$ for constants $K > 0$ and $\tau \in (0, 1]$) corresponds to bounded Hölder regularity of T as was discussed in [28]. In this case, $\kappa^{-1}(t) = \sqrt[\tau]{t/K}$ and so

$$\varphi(t) = \sqrt{t^2 - \frac{\gamma}{K^{\frac{2}{\tau}}} t^{\frac{2}{\tau}}} = t \sqrt{1 - \frac{\gamma}{K^{\frac{2}{\tau}}} t^{\alpha}},$$

where $\alpha := 2/\tau - 2 = 2(1 - \tau)/\tau > 0$. By [29, §4] this yields

$$\varphi^n(t) \leq \left(t^{-\alpha} + \alpha n \frac{\gamma}{K^{2/\tau}} \right)^{-\alpha} = O(n^{-1/\alpha}) = O\left(n^{-\frac{\tau}{2(1-\tau)}}\right).$$

Theorem 3.1.8 then implies convergence with order $O\left(n^{-\frac{\tau}{2(1-\tau)}}\right)$ which recovers [28, Proposition 3.1].

As the following example shows, at least in principle, Theorem 3.1.8 opens the possibility of characterizing different convergence rates by choosing U appropriately.

EXAMPLE 3.1.10 (convergence rate by regions of a fixed point). [102, Example 1] Consider the alternating projections operator $T := P_A P_B$ for the two convex subsets A and B of \mathbb{R}^2 given by

$$A := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}, \quad B := \text{epi}(f) \text{ where } f(t) = \begin{cases} t & \text{if } t \geq 0, \\ t^2 & \text{if } t < 0. \end{cases}$$

In this setting, we have $\text{Fix } T = A \cap B = \{0\}$. The alternating projections sequence given by $x_{n+1} := T x_n$ always converges to 0. However, the rate which it does so depends on the starting point $x_0 \in \mathbb{R}^2$. We consider two cases:

- (i) Let $U_1 := \mathbb{R}_+ \times \mathbb{R}$. Then the linear error bound condition is satisfied on U_1 and (x_n) converges linearly.
- (ii) Let $U_2 := \mathbb{R}_- \times \mathbb{R}$. Then there is a Hölder-type gauge function κ such that the error bound condition with gauge κ is satisfied on U_2 and (x_n) converges sublinearly.

3.2 Cyclic projections

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} . The content of this section is taken from our joint work with Dr. Matthew K. Tam [103] except Theorem 3.2.13.

Having established the basic geometric language of set feasibility and its connection to the averaging and stability properties of fixed point mappings, we can now present convergence results for cyclic projections between sets with possibly empty intersection, Theorem 3.2.7 and Corollary 3.2.8. The majority of the work, and the source of technical complications, lies in constructing an appropriate fixed point mapping in the right space in order to be able to apply Theorem 3.1.3. As we have already said, establishing the extent of almost averaging is a straight-forward application of Theorem 2.1.6. Thanks to Proposition 1.3.10 this can be stated in terms of the more primitive property of elemental set regularity. The challenging part is to show that subtransversality as introduced above leads to metric subregularity of an appropriate fixed point surrogate for cyclic projections, Proposition 3.2.4. In the process we show in Proposition 3.2.6 that elemental regularity and subtransversality become entangled and it is not clear whether they can be completely separated when it comes to necessary conditions for convergence of cyclic projections.

Given a collection of closed subsets of \mathbb{E} , $\{A_1, A_2, \dots, A_m\}$ ($m \geq 2$), and an initial point u^0 , the cyclic projections algorithm generates the sequence $(u^k)_{k \in \mathbb{N}}$ by

$$u^{k+1} \in T_{CP}u^k, \quad T_{CP} := P_{A_1}P_{A_2} \cdots P_{A_m}.$$

We will assume throughout this section that $\text{Fix } T_{CP} \neq \emptyset$.

Our analysis proceeds on an appropriate product space designed for the cycles associated with a given fixed point of T_{CP} . As above we will use A to denote the product set on \mathbb{E}^m : $A := A_1 \times A_2 \times \cdots \times A_m$. Let $\bar{u} \in \text{Fix } T_{CP}$ and let $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ where

$$\mathcal{Z}(u) := \{\zeta := z - \Pi z \mid z \in W_0 \subset \mathbb{E}^m, z_1 = u\}$$

for the permutation mapping Π given by (1.12) and

$$W_0 := \{x \in \mathbb{E}^m \mid x_m \in P_{A_m}x_1, x_j \in P_{A_j}x_{j+1}, j = 1, 2, \dots, m-1\}.$$

Note that $\sum_{j=1}^m \bar{\zeta}_j = 0$. The vector $\bar{\zeta}$ is a *difference vector* which gives information regarding the intra-steps of the cyclic projections operator T_{CP} at the fixed point \bar{u} . In the case of only two sets, a difference vector is frequently called a *gap vector* [12, 17, 22, 96]. This is unique in the convex case, but need not be in the nonconvex case (see Lemma 3.2.3 below). In the more general setting we have here, this corresponds to nonuniqueness of cycles for cyclic projections. This greatly complicates matters since the fixed points associated with T_{CP} will not, in general, be associated with cycles that are the same length and orientation. Consequently, the usual trick of looking at the zeros of $T_{CP} - \text{Id}$ is rather uninformative,

and another mapping needs to be constructed which distinguishes fixed points associated with different cycles. The following development establishes some of the key properties of difference vectors and cycles which then motivates the mapping that we construct for this purpose.

To analyze the cyclic projections algorithm we consider the sequence on the product space on \mathbb{E}^m , $(x^k)_{k \in \mathbb{N}}$ generated by $x^{k+1} \in T_{\bar{\zeta}} x^k$ with

$$T_{\bar{\zeta}} : \mathbb{E}^m \rightrightarrows \mathbb{E}^m : x \mapsto \left\{ \left(x_1^+, x_1^+ - \bar{\zeta}_1, \dots, x_1^+ - \sum_{j=1}^{m-1} \bar{\zeta}_j \right) \mid x_1^+ \in T_{CP} x_1 \right\} \quad (3.6)$$

for $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ where $\bar{u} \in \text{Fix } T_{CP}$. In order to isolate cycles we restrict our attention to relevant subsets of \mathbb{E}^m . These are

$$\begin{aligned} W(\bar{\zeta}) &:= \{x \in \mathbb{E}^m \mid x - \Pi x = \bar{\zeta}\}, \\ L &:= \text{an affine subspace with } T_{\bar{\zeta}} : L \rightrightarrows L, \\ \Lambda &:= L \cap W(\bar{\zeta}). \end{aligned} \quad (3.7)$$

The set $W(\bar{\zeta})$ is an affine transformation of the diagonal of the product space and thus an affine subspace: for $x, y \in W(\bar{\zeta})$, $z = \lambda x + (1 - \lambda)y$ satisfies $z - \Pi z = \bar{\zeta}$ for all $\lambda \in \mathbb{R}$. This affine subspace is used to characterize the local geometry of the sets in relation to each other at fixed points of the cyclic projections operator.

Points in $\text{Fix } T_{CP}$ can correspond to cycles of different lengths, hence an element $x \in \text{Fix } T_{\bar{\zeta}}$ need not be in W_0 and vice versa, as the next example demonstrates.

EXAMPLE 3.2.1 ($\text{Fix } T_{\bar{\zeta}}$ and W_0). [103, Example 3.2] Consider the sets $A_1 = \{0, 1\}$ and $A_2 = \{0, 3/4\}$. The cyclic projections operator T_{CP} has fixed points $\{0, 1\}$ and two corresponding cycles, $\mathcal{Z}(0) = \{(0, 0)\}$ and $\mathcal{Z}(1) = \{(1/4, -1/4)\}$. Let $\bar{\zeta} = (1/4, -1/4)$. Then $(0, -1/4) \in \text{Fix } T_{\bar{\zeta}}$ but $(0, -1/4) \notin W_0$. Conversely, the vector $(0, 0) \in W_0$, but $(0, 0) \notin \text{Fix } T_{\bar{\zeta}}$. The point $(1, 3/4)$, however, belongs to both W_0 and $\text{Fix } T_{\bar{\zeta}}$.

The example above shows that what distinguishes elements in $\text{Fix } T_{\bar{\zeta}}$ from each other is whether or not they also belong to W_0 . The next lemma establishes that, on appropriate subsets, a fixed point of $T_{\bar{\zeta}}$ can be identified meaningfully with a vector in the image of the mapping Ψ in Definition 2.3.12 which is used to characterize the alignment of the sets A_j to each other at points of interest (in particular, fixed points of the cyclic projections operator).

LEMMA 3.2.2. [103, Lemma 3.1] Let $\bar{u} \in \text{Fix } T_{CP}$ and let $\bar{\zeta} \in \mathcal{Z}(\bar{u})$. Define $\Psi := (P_A - \text{Id}) \circ \Pi$ and $F_{\bar{\zeta}} := T_{\bar{\zeta}} - \text{Id}$.

(i) $T_{\bar{\zeta}}$ maps $W(\bar{\zeta})$ to itself. Moreover $x \in \text{Fix } T_{\bar{\zeta}}$ if and only if $x \in W(\bar{\zeta})$ with $x_1 \in \text{Fix } T_{CP}$. Indeed,

$$\text{Fix } T_{\bar{\zeta}} = \left\{ x = (x_1, x_2, \dots, x_m) \in \mathbb{E}^m \mid x_1 \in \text{Fix } T_{CP}, x_j = x_1 - \sum_{i=1}^{j-1} \bar{\zeta}_i, j = 2, 3, \dots, m \right\}.$$

(ii) A point $\bar{z} \in \text{Fix } T_{\bar{\zeta}} \cap W_0$ if and only if $\bar{\zeta} \in \Psi(\bar{z})$ if and only if $\bar{\zeta} \in (F_{\bar{\zeta}} \circ \Pi)(\bar{z})$.

(iii) $\Psi^{-1}(\bar{\zeta}) \cap W(\bar{\zeta}) \subseteq F_{\bar{\zeta}}^{-1}(0) \cap W(\bar{\zeta})$.

(iv) If the distance is with respect to the Euclidean norm then

$$\text{dist}(0, F_{\bar{\zeta}}(x)) = \sqrt{m} \text{dist}(x_1, T_{CP}x_1).$$

LEMMA 3.2.3 (difference vectors: cyclic projections). [103, Lemma 3.2] Let $A_j \subseteq \mathbb{E}$ be nonempty and closed ($j = 1, 2, \dots, m$). Let $S_0 \subset \text{Fix } T_{CP}$, let U_0 be a neighborhood of S_0 and define $U := \{z = (z_1, z_2, \dots, z_m) \in W_0 \mid z_1 \in U_0\}$. Fix $\bar{u} \in S_0$ and the difference vector $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ with $\bar{\zeta} = \bar{z} - \Pi\bar{z}$ for the point $\bar{z} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_m) \in W_0$ having $\bar{z}_1 = \bar{u}$. If A_j is elementally subregular at \bar{z}_j for $(\bar{z}_j, 0) \in \text{gph } N_{A_j}^{\text{prox}}$ with constant $\bar{\varepsilon}_j$ and neighborhood $U_j := p_j(U)$ of \bar{z}_j (where p_j is the j th coordinate projection operator), then

$$\|\bar{\zeta} - \zeta\|^2 \leq \sum_{j=1}^m \varepsilon_j \|\bar{z}_j - z_j\|^2 \quad (\varepsilon_j := 2\bar{\varepsilon}_j + 2\bar{\varepsilon}_j^2)$$

for the difference vector $\zeta \in \mathcal{Z}(u)$ with $u \in S_0$ and $\zeta = z - \Pi z$ where $z = (z_1, z_2, \dots, z_m) \in W_0$ with $z_1 = u$. If the sets A_j ($j = 1, 2, \dots, m$) are in fact convex, then the difference vector is unique and independent of the initial point \bar{u} , that is, $\mathcal{Z}(u) = \{\bar{\zeta}\}$ for all $u \in S_0$.

PROPOSITION 3.2.4 (metric subregularity of cyclic projections). [103, Proposition 3.4] Let $\bar{u} \in \text{Fix } T_{CP}$ and $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ and let $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m) \in W_0$ satisfy $\bar{\zeta} = \bar{x} - \Pi\bar{x}$ with $\bar{x}_1 = \bar{u}$. For L an affine subspace containing \bar{x} , let $T_{\bar{\zeta}} : L \rightrightarrows L$ and define the mapping $F_{\bar{\zeta}} := T_{\bar{\zeta}} - \text{Id}$. Suppose the following hold:

- (a) the collection of sets $\{A_1, A_2, \dots, A_m\}$ is subtransversal at \bar{x} for $\bar{\zeta}$ relative to $\Lambda := L \cap W(\bar{\zeta})$ with constant κ and neighborhood U of \bar{x} ;
- (b) there exists a positive constant σ such that

$$\text{dist}(\bar{\zeta}, \Psi(x)) \leq \sigma \text{dist}(0, F_{\bar{\zeta}}(x)), \quad \forall x \in \Lambda \cap U \text{ with } x_1 \in A_1. \quad (3.8)$$

Then F is metrically subregular for 0 on U (metrically regular on $U \times \{0\}$) relative to Λ with constant $\bar{\kappa} = \kappa\sigma$.

EXAMPLE 3.2.5 (two intersecting sets). [103, Example 3.3] To provide some insight into condition (b) of Proposition 3.2.4 it is instructive to examine the case of two sets with nonempty intersection. Let $\bar{x} = (\bar{u}, \bar{u})$ with $\bar{u} \in A_1 \cap A_2$ and the difference vector $\bar{\zeta} = 0 \in \mathcal{Z}(\bar{x})$. To simplify the presentation, let us consider $L = \mathbb{E}^2$ and $U = U' \times U'$, where U' is a neighborhood of \bar{u} . Then, one has $\Lambda = W(0) = \{(u, u) : u \in \mathbb{E}\}$ and, hence, $x \in \Lambda \cap U$ with $x_1 \in A_1$ is equivalent to $x = (u, u) \in U$ with $u \in A_1 \cap U'$. For such a point $x = (u, u)$, one has

$$\begin{aligned} \text{dist}(0, \Psi(x)) &= \text{dist}(u, A_2), \\ \text{dist}(0, F_0(x)) &= \sqrt{2} \text{dist}(u, P_{A_1} P_{A_2}(u)), \end{aligned}$$

where the last equality follows from the representation

$$F_0(x) = \{(z - u, z - u) \in \mathbb{E}^2 : z \in P_{A_1} P_{A_2}(u)\}.$$

Condition (b) of Proposition 3.2.4 becomes

$$\text{dist}(u, A_2) \leq \gamma \text{dist}(u, P_{A_1} P_{A_2}(u)), \quad \forall u \in A_1 \cap U'. \quad (3.9)$$

where $\gamma := \sqrt{2}\sigma > 0$. In [84, Remark 12] the phenomenon of entanglement of elemental subregularity and regularity of collections of sets is briefly discussed in the context of other notions of regularity in the literature. Inequality (3.9) serves as a type of conduit for this entanglement of regularities as Proposition 3.2.6 demonstrates.

PROPOSITION 3.2.6 (elemental subregularity and (3.9) imply subtransversality). [103, Proposition 3.5] Let $\bar{u} \in A_1 \cap A_2$ and U' be the neighborhood of \bar{u} as in Example 3.2.5. Suppose that condition (3.9) holds and that the set A_1 is elementally subregular relative to A_2 at \bar{u} for all $(\bar{y}, 0)$ with $\bar{y} \in A_1 \cap U'$ with constant $\varepsilon < 1/(1 + \gamma^2)$ and the neighborhood U' . Then $\{A_1, A_2\}$ is subtransversal at \bar{u} .

The main result of this section can now be presented. This statement uses the full technology of regularities relativized to certain sets of points S_j introduced in Definitions 2.1.2 and 1.3.3 and used in Proposition 1.3.10, as well as the expanded notion of subtransversality of inconsistent collections of sets introduced in Definition 2.3.12 and applied in Proposition 3.2.4.

THEOREM 3.2.7 (convergence of cyclic projections). [103, Theorem 3.2] Let $S_0 \subset \text{Fix } T_{CP} \neq \emptyset$ and $Z := \cup_{u \in S_0} \mathcal{Z}(u)$. Define

$$S_j := \bigcup_{\zeta \in Z} \left(S_0 - \sum_{i=1}^{j-1} \zeta_i \right) \quad (j = 1, 2, \dots, m). \quad (3.10)$$

Let $U := U_1 \times U_2 \times \cdots \times U_m$ be a neighborhood of $S := S_1 \times S_2 \times \cdots \times S_m$ and suppose that

$$P_{A_j} \left(u - \sum_{i=1}^j \zeta_i \right) \subseteq S_0 - \sum_{i=1}^{j-1} \zeta_i \quad \forall u \in S_0, \forall \zeta \in Z \quad \text{for each } j = 1, 2, \dots, m, \quad (3.11a)$$

$$P_{A_j} U_{j+1} \subseteq U_j \quad \text{for each } j = 1, 2, \dots, m \quad (U_{m+1} := U_1). \quad (3.11b)$$

For $\bar{\zeta} \in Z$ fixed and $\bar{x} \in S$ with $\bar{\zeta} = \Pi \bar{x} - \bar{x}$, generate the sequence $(x^k)_{k \in \mathbb{N}}$ by $x^{k+1} \in T_{\bar{\zeta}} x^k$ for $T_{\bar{\zeta}}$ defined by (3.6), seeded by a point $x^0 \in W(\bar{\zeta}) \cap U$ for $W(\bar{\zeta})$ defined by (3.7) with $x_1^0 \in A_1 \cap U_1$.

Suppose that, for $\Lambda := L \cap \text{aff}(\cup_{\zeta \in Z} W(\zeta)) \supset S$ such that $T_{\zeta} : \Lambda \rightrightarrows \Lambda$ for all $\zeta \in Z$ and an affine subspace $L \supset \text{aff}(x^k)_{k \in \mathbb{N}}$, the following hold:

(a) the set A_j is elementally subregular at all $\hat{x}_j \in S_j$ relative to S_j for each

$$(x_j, v_j) \in V_j := \left\{ (z, w) \in \text{gph } N_{A_j}^{\text{prox}} \mid z + w \in U_j \quad \text{and} \quad z \in P_{A_j}(z + w) \right\}$$

with constant $\varepsilon_j \in (0, 1)$ on the neighborhood U_j for $j = 1, 2, \dots, m$;

(b) for each $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m) \in S$, the collection of sets $\{A_1, A_2, \dots, A_m\}$ is subtransversal at \hat{x} for $\hat{\zeta} := \hat{x} - \Pi \hat{x}$ relative to Λ with constant κ on the neighborhood U ;

(c) for $F_{\hat{\zeta}} := T_{\hat{\zeta}} - \text{Id}$ and $\Psi := (P_A - \text{Id}) \circ \Pi$ there exists a positive constant σ such that for all $\hat{\zeta} \in Z$

$$\text{dist}(\hat{\zeta}, \Psi(x)) \leq \sigma \text{dist}(0, F_{\hat{\zeta}}(x))$$

holds whenever $x \in \Lambda \cap U$ with $x_1 \in A_1$;

(d) $\text{dist}(x, S) \leq \text{dist}(x, F_{\hat{\zeta}}^{-1}(0) \cap \Lambda)$ for all $x \in U \cap \Lambda$, for all $\hat{\zeta} \in Z$.

Then the sequence $(x^k)_{k \in \mathbb{N}}$ seeded by a point $x^0 \in W(\bar{\zeta}) \cap U$ with $x_1^0 \in A_1 \cap U_1$ satisfies

$$\text{dist}(x^{k+1}, \text{Fix } T_{\bar{\zeta}} \cap S) \leq c \text{dist}(x^k, S)$$

whenever $x^k \in U$ with

$$c := \sqrt{1 + \bar{\varepsilon} - \frac{1 - \alpha}{\alpha \bar{\kappa}^2}}$$

for

$$\bar{\varepsilon} := \prod_{j=1}^m (1 + \tilde{\varepsilon}_j) - 1, \quad \tilde{\varepsilon}_j := 4\varepsilon_j \frac{1 + \varepsilon_j}{(1 - \varepsilon_j)^2}, \quad \alpha := \frac{m}{m + 1} \quad (3.12)$$

and $\bar{\kappa} = \kappa\sigma$. If, in addition,

$$\bar{\kappa} < \sqrt{\frac{1-\alpha}{\bar{\varepsilon}\alpha}},$$

then $\text{dist}(x^k, \text{Fix } T_{\bar{\zeta}} \cap S) \rightarrow 0$, and hence $\text{dist}(x_1^k, \text{Fix } T_{CP} \cap S_1) \rightarrow 0$, at least linearly with rate $c < 1$.

COROLLARY 3.2.8 (global R-linear convergence of convex cyclic projections). *[103, Corollary 3.1] Let the sets A_j ($j = 1, 2, \dots, m$) be nonempty, closed and convex, let $S_0 = \text{Fix } T_{CP} \neq \emptyset$ and let $S = S_1 \times S_2 \times \dots \times S_m$ for S_j defined by (3.10). Let $\Lambda := W(\bar{\zeta})$ for $\bar{\zeta} \in \mathcal{Z}(u)$ and any $u \in S_0$. Suppose, in addition, that*

(b') *for each $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m) \in S$, the collection of sets $\{A_1, A_2, \dots, A_m\}$ is subtransversal at \hat{x} for $\bar{\zeta} = \hat{x} - \Pi\hat{x}$ relative to Λ with neighborhood $U \supset S$;*

(c') *there exists a positive constant σ such that*

$$\text{dist}(\bar{\zeta}, \Psi(x)) \leq \sigma \text{dist}(0, F_{\bar{\zeta}}(x))$$

holds whenever $x \in \Lambda \cap U$ with $x_1 \in A_1$.

Then the sequence $(x^k)_{k \in \mathbb{N}}$ generated by $x^{k+1} \in T_{\bar{\zeta}}x^k$ seeded by any point $x^0 \in W(\bar{\zeta})$ with $x_1^0 \in A_1$ satisfies

$$\text{dist}(x^{k+1}, \text{Fix } T_{\bar{\zeta}} \cap S) \leq c \text{dist}(x^k, S)$$

for all k large enough where

$$c := \sqrt{1 - \frac{1-\alpha}{\alpha\bar{\kappa}^2}} < 1$$

with $\bar{\kappa} = \kappa\sigma$ for κ a constant of metric subregularity of Ψ for $\bar{\zeta}$ on U relative to Λ and α given by (3.12). In other words, $\text{dist}(x^k, \text{Fix } T_{\bar{\zeta}} \cap S) \rightarrow 0$, and hence $\text{dist}(x_1^k, \text{Fix } T_{CP} \cap S_0) \rightarrow 0$, at least R-linearly with rate $c < 1$.

When the sets A_j are affine, then it is easy to see that the sets are subtransversal to each other at collections of nearest points corresponding to the gap between the sets. If the cyclic projection algorithm does not converge in one step (which it will in the case of either parallel or orthogonally arranged sets) the above corollary shows that cyclic projections converge linearly with rate $\sqrt{1-\bar{\kappa}}$ where $\bar{\kappa}$ is the constant of metric subregularity, reflecting the angle between the affine subspaces. This much for the affine case has already been shown in [14, Theorem 5.7.8].

REMARK 3.2.9 (global convergence for nonconvex alternating projections). *[103, Remark 3.1] Convexity is not necessary for global linear convergence of alternating projections. This has been demonstrated using earlier versions of the theory presented here for sparse affine*

feasibility in [60, Corollary III.13 and Theorem III.15]. A sufficient property for global results in sparse affine feasibility is a common restricted isometry property [60, Eq. (32)] familiar to experts in signal processing with sparsity constraints. The restricted isometry property was shown in [60, Proposition III.14] to imply transversality of the affine subspace with all subspaces of a certain dimension.

EXAMPLE 3.2.10 (an equilateral triangle – three affine subspaces with a hole). [103, Example 3.4] Consider the problem specified by the following three sets in \mathbb{R}^2

$$\begin{aligned} A_1 &= \mathbb{R}(1, 0) &&= \{x \in \mathbb{R}^2 \mid \langle (0, 1), x \rangle = 0\}, \\ A_2 &= (0, -1) + \mathbb{R}(-\sqrt{3}, 1) &&= \{x \in \mathbb{R}^2 \mid \langle (-\sqrt{3}, 1), x \rangle = \sqrt{3}\}, \\ A_3 &= (0, 1) + \mathbb{R}(\sqrt{3}, 1) &&= \{x \in \mathbb{R}^2 \mid \langle (\sqrt{3}, 1), x \rangle = 1\}. \end{aligned}$$

The following statements regarding the assumptions of Corollary 3.2.8 are easily verified.

- (i) The set $S_0 = \text{Fix } T_{CP} = \{(-1/3, 0)\}$.
- (ii) There is a unique fixed point $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3) = ((-1/3, 0), (-1/3, 2/\sqrt{3}), (2/3, 1/\sqrt{3}))$.
- (iii) The set of difference vectors is a singleton:

$$Z = \{\bar{\zeta}\} = \{(\bar{\zeta}_1, \bar{\zeta}_2, \bar{\zeta}_3)\} = \left\{ \left((0, -2/\sqrt{3}), (-1, 1/\sqrt{3}), (1, 1/\sqrt{3}) \right) \right\}.$$

- (iv) The sets S_1, S_2 and S_3 are given by

$$\begin{aligned} S_1 &= S_0 - \bar{\zeta}_1 &&= \left\{ (-1/3, 2/\sqrt{3}) \right\} \\ S_2 &= S_0 - \bar{\zeta}_1 - \bar{\zeta}_2 &&= \left\{ (2/3, 1/\sqrt{3}) \right\} \\ S_3 &= S_0 &&= \{(-1/3, 0)\}. \end{aligned}$$

- (v) Condition (3.11a) is satisfied and condition (3.11b) is satisfied with $U_j = \mathbb{R}^2$ ($j = 1, 2, 3$).
- (vi) For $j \in \{1, 2, 3\}$, A_j is convex and hence elementally regular at \bar{x}_j with constant $\varepsilon_j = 0$ [84, Proposition 4].
- (vii) The mapping Ψ is metrically subregular for $\bar{\zeta}$ on $(\mathbb{R}^2)^3$ with constant $\kappa = \sqrt{2}$ relative to $W(\bar{\zeta})$:

$$\text{dist}(x, \Psi^{-1}(\bar{\zeta}) \cap W(\bar{\zeta})) \leq \sqrt{2} \text{dist}(\bar{\zeta}, \Psi(x)) \quad \forall x \in (\mathbb{R}^2)^3.$$

(viii) For all $x \in W(\bar{\zeta})$, the inequality $\text{dist}(\bar{\zeta}, \Psi(x)) \leq \sigma \text{dist}(0, F_{\bar{\zeta}}(x))$ holds with $\sigma = 4\sqrt{2}/9$.

The assumptions of Corollary 3.2.8 are satisfied. Furthermore, Proposition 3.2.4 shows that the mapping $F_{\bar{\zeta}}$ is metrically subregular for 0 on $(\mathbb{R}^2)^3$ relative to $W(\bar{\zeta})$ with constant $\bar{\kappa} = \kappa\sigma = \sqrt{2} \times 4\sqrt{2}/9 = 8/9$. Altogether, Corollary 3.2.8 yields that, from any starting point, the cyclic projections method converges linearly to \bar{u} with rate at most $c = \sqrt{37}/8$.

The next example is new and rather unexpected.

EXAMPLE 3.2.11 (two non-intersecting circles). [103, Example 3.5] Fix $r > 0$ and consider the problem specified by the following two sets in \mathbb{R}^2

$$\begin{aligned} A_1 &= \{x \in \mathbb{R}^2 \mid \|x\| = 1\}, \\ A_2 &= \{x \in \mathbb{R}^2 \mid \|x + (0, 1/2 + r)\| = 2 + r\}. \end{aligned}$$

In this example we focus on (local) behavior around the point $\bar{u} = (0, 1)$. For U_1 , a sufficiently small neighborhood of \bar{u} , the following statements regarding the assumptions of Theorem 3.2.7 can be verified.

- (i) $S_0 = \text{Fix } T_{CP} \cap U_1 = \{\bar{u}\} = \{(0, 1)\}$;
- (ii) $\bar{x} = (\bar{x}_1, \bar{x}_2) = (\bar{u}, (0, 3/2)) = ((0, 1), (0, 3/2))$;
- (iii) $\mathcal{Z} = \{\bar{\zeta}\} = \{(\bar{\zeta}_1, \bar{\zeta}_2)\} = \{((0, -1/2), (0, 1/2))\}$;
- (iv) the sets S_1 and S_2 are given by

$$\begin{aligned} S_1 &= S_0 - \bar{\zeta}_1 = \{(0, 1/2)\} \\ S_2 &= S_0 - \bar{\zeta}_1 - \bar{\zeta}_2 = \{(0, 1)\}; \end{aligned}$$

- (v) (3.11a) is satisfied, and (3.11b) holds with U_1 already given and U_2 equal to a scaled-translate of U_1 —more precisely, U_1 and U_2 are related by

$$U_2 = \frac{2+r}{\text{dist}(\bar{u}, (0, -\frac{1}{2}-r))} U_1 + (0, 1/2);$$

- (vi) $L = \mathbb{R}^2 \times \mathbb{R}^2$;
- (vii) for $j \in \{1, 2\}$, A_j is uniformly elementally regular at \bar{x}_j for any $\varepsilon_j \in (0, 1)$ [84, Example 2(b)];

In order to verify the remaining conditions of Theorem 3.2.7, we use the following parametrization: any $x = (x_1, x_2) \in W(\bar{\zeta})$ with $x_1 \in A_1$ may be expressed in the form $x_1 = (b, \sqrt{1-b^2}) \in A_1$ where $b \in \mathbb{R}$ is a parameter.

(viii) $\{A_1, A_2\}$ is subtransversal at \bar{x} relative to $W(\bar{\zeta})$, i.e., Ψ is metrically subregular at \bar{x} for $\bar{\zeta}$ on U (metrically regular at $(\bar{x}, \bar{\zeta})$ on $U \times \{\bar{\zeta}\}$) relative to $W(\bar{\zeta})$ with constant

$$\kappa \lim_{b \rightarrow 0} \frac{\text{dist}(x, \Psi^{-1}(\bar{\zeta}) \cap W(\bar{\zeta}))}{\text{dist}(\bar{\zeta}, \Psi(x))} = \frac{3(2r+3)}{\sqrt{2r^2+6r+9}}.$$

(ix) For any $\rho > 0$ such that

$$\rho > \lim_{b \rightarrow 0} \frac{\text{dist}(\bar{\zeta}, \Psi(x))}{\text{dist}(0, F_{\bar{\zeta}}(x))} = \frac{\sqrt{2}\sqrt{2r^2+6r+9}(2r+3)}{2\sqrt{4r^2+12r+13}(r+2)},$$

the following inequality holds

$$\text{dist}(\bar{\zeta}, \Psi(x)) \leq \rho \text{dist}(0, F_{\bar{\zeta}}(x))$$

for all $x \in W(\bar{\zeta})$ sufficiently close to \bar{x} .

The assumptions of Theorem 3.2.7 are satisfied. Furthermore, Proposition 3.2.4 shows that the mapping $F_{\bar{\zeta}}$ is metrically subregular at \bar{x} for 0 relative to $W(\bar{\zeta})$ on U with the constant $\bar{\kappa}$ equal to the product of constant of subtransversality κ in (viii) and ρ . That is,

$$\bar{\kappa} = \frac{3\sqrt{2}(2r+3)^2}{2\sqrt{4r^2+12r+13}(r+2)}.$$

Altogether, Theorem 3.2.7 yields that, for any c with

$$1 > c > \sqrt{1 - \frac{(4r^2+12r+13)(r+2)^2}{9(2r+3)^4}},$$

there exists a neighborhood of \bar{u} such that the cyclic projections method converges linearly to \bar{u} with rate c .

REMARK 3.2.12 (non-intersecting circle and line). [103, Remark 3.2] A analysis similar to Example 3.2.11 can be performed for the case in which the second circle A_2 is replaced with the line $(0, 3/2) + \mathbb{R}(1, 0)$. Formally, this corresponds to setting the parameter $r = +\infty$ in Example 3.2.11. Although there are some technicalities involved in order to make such an argument fully rigorous, a separate computation has verified the constants obtained in this way agree with those obtained from a direct computation. When the circle and line are tangent, then Example 3.1.5 shows how sublinear convergence of alternating projections can be quantified.

We conclude this section with a more intuitive result which will be applied directly to the source location problem in Section 5.1.

THEOREM 3.2.13 (linear convergence from strong subtransversality and prox-regularity). *Consider a collection of prox-regular sets $\{A_1, A_2, \dots, A_m\}$ and suppose that it is strongly subtransversal at $\bar{x} \in \cap_{i=1}^m A_i$. Then every sequence generated by T_{CP} converges linearly to \bar{x} provided that the initial point is sufficiently close to \bar{x} .*

Proof. By the strong subtransversality assumption, there exist $\kappa > 0$ and $\Delta > 0$ such that $(\cap_{i=1}^m A_i) \cap \mathbb{B}_{2\Delta}(\bar{x}) = \{\bar{x}\}$ and

$$\|x - \bar{x}\| = \text{dist}(x, \cap_{i=1}^m A_i) \leq \kappa \max_{1 \leq i \leq m} \text{dist}(x, A_i) \quad \forall x \in \mathbb{B}_{\Delta}(\bar{x}). \quad (3.13)$$

By the prox-subregularity assumption, for any given $\varepsilon \in (0, 1)$, there exists $\delta_\varepsilon > 0$ such that

$$\langle x - P_{A_i}x, \bar{x} - P_{A_i}x \rangle \leq \varepsilon \|x - P_{A_i}x\| \|\bar{x} - P_{A_i}x\| \quad \forall x \in \mathbb{B}_{\delta_\varepsilon}(\bar{x}), \forall i = \overline{1, m}. \quad (3.14)$$

Fix a number

$$0 < \varepsilon < \frac{1}{2m(\kappa + 1)} \quad (3.15)$$

and a corresponding $\delta_\varepsilon > 0$ satisfying condition (3.14). Let us define $\delta = \min\{\delta_\varepsilon, \Delta\} > 0$ and show that every sequence generated by T_{CP} starting in $\mathbb{B}_\delta(\bar{x})$ converges linearly to \bar{x} .

Indeed, let any $x \in A_m \cap \mathbb{B}_\delta(\bar{x})$ and $x^+ \in T_{CP}x$. Let us denote $x_i \in P_{A_i}x_{i-1}$ ($i = 1, 2, \dots, m$), where $x_0 = x$ and $x_m = x^+$. By (3.13) and the choice of δ , we have that

$$\|x - \bar{x}\| \leq \kappa \max_{1 \leq i \leq m} \text{dist}(x, A_i). \quad (3.16)$$

Since $x_i \in A_i$ ($1 \leq i \leq m$),

$$\text{dist}(x, A_i) = \text{dist}(x_0, A_i) \leq \sum_{j=0}^{i-1} \|x_j - x_{j+1}\| \leq \sum_{j=0}^{m-1} \|x_j - x_{j+1}\|. \quad (3.17)$$

Plugging (3.17) into (3.16) yields

$$\|x - \bar{x}\| \leq \kappa \sum_{j=0}^{m-1} \|x_j - x_{j+1}\|. \quad (3.18)$$

Note also that

$$\|x - x_{i+1}\| = \|x_0 - x_{i+1}\| \leq \sum_{j=0}^i \|x_j - x_{j+1}\| \leq \sum_{j=0}^{m-1} \|x_j - x_{j+1}\|. \quad (3.19)$$

Using (3.14), triangle inequality, (3.18) and (3.19) successively yields that for each $i = 0, 1, \dots, m-1$,

$$\begin{aligned}
\langle x_i - x_{i+1}, \bar{x} - x_{i+1} \rangle &\leq \varepsilon \|x_i - x_{i+1}\| \|\bar{x} - x_{i+1}\| \\
&\leq \varepsilon \|x_i - x_{i+1}\| (\|\bar{x} - x\| + \|x - x_{i+1}\|) \\
&\leq \varepsilon \|x_i - x_{i+1}\| \left(\kappa \sum_{j=0}^{m-1} \|x_j - x_{j+1}\| + \sum_{j=0}^{m-1} \|x_j - x_{j+1}\| \right) \\
&= \varepsilon(\kappa + 1) \|x_i - x_{i+1}\| \sum_{j=0}^{m-1} \|x_j - x_{j+1}\|.
\end{aligned}$$

Then we get

$$\begin{aligned}
\|x_i - \bar{x}\|^2 &= \|x_{i+1} - \bar{x}\|^2 + \|x_i - x_{i+1}\|^2 - 2 \langle x_i - x_{i+1}, \bar{x} - x_{i+1} \rangle \\
&\geq \|x_{i+1} - \bar{x}\|^2 + \|x_i - x_{i+1}\|^2 - 2\varepsilon(\kappa + 1) \|x_i - x_{i+1}\| \sum_{j=0}^{m-1} \|x_j - x_{j+1}\|.
\end{aligned}$$

Adding the above inequalities over $i = 0, 1, \dots, m-1$, we obtain

$$\|x - \bar{x}\|^2 \geq \|x^+ - \bar{x}\|^2 + \sum_{i=0}^{m-1} \|x_i - x_{i+1}\|^2 - 2\varepsilon(\kappa + 1) \left(\sum_{i=0}^{m-1} \|x_i - x_{i+1}\| \right)^2.$$

Thanks to the Cauchy-Schwarz inequality and (3.18), the last estimate implies

$$\begin{aligned}
\|x - \bar{x}\|^2 &\geq \|x^+ - \bar{x}\|^2 + \frac{1}{m} \left(\sum_{i=0}^{m-1} \|x_i - x_{i+1}\| \right)^2 - 2\varepsilon(\kappa + 1) \left(\sum_{i=0}^{m-1} \|x_i - x_{i+1}\| \right)^2 \\
&= \|x^+ - \bar{x}\|^2 + \left(\frac{1}{m} - 2\varepsilon(\kappa + 1) \right) \left(\sum_{i=0}^{m-1} \|x_i - x_{i+1}\| \right)^2 \\
&\geq \|x^+ - \bar{x}\|^2 + \left(\frac{1}{m} - 2\varepsilon(\kappa + 1) \right) \frac{1}{\kappa^2} \|x - \bar{x}\|^2.
\end{aligned}$$

Hence

$$\|x^+ - \bar{x}\|^2 \leq \left(1 + \frac{2\varepsilon(\kappa + 1)}{\kappa^2} - \frac{1}{m\kappa^2} \right) \|x - \bar{x}\|^2.$$

Due to (3.15) we have $c := \left(1 + \frac{2\varepsilon(\kappa + 1)}{\kappa^2} - \frac{1}{m\kappa^2} \right) < 1$ and the proof is complete. \square

REMARK 3.2.14. *Since the parameter $\varepsilon \downarrow 0$ as $x^k \rightarrow \bar{x}$, the rate c estimated above tends to $1 - \frac{1}{m\kappa^2}$ which is governed by the modulus of the strong subtransversality property.*

REMARK 3.2.15. *It is not difficult to see that Theorem 3.2.13 is encompassed in the framework of Theorem 3.2.7. The proof given above can be viewed as a shortcut for verifying the assumptions of that theorem.*

3.3 Alternating projections

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} .

Given two sets A and B , the feasibility problem consists in finding a point in their intersection $A \cap B$. If these are closed sets in finite dimensions, alternating projections are determined by a sequence (x_k) starting with some point x_0 and such that

$$x_{k+1} \in P_A P_B(x_k) \quad (k = 0, 1, \dots).$$

In analyzing convergence of the alternating projections (x_k) , it is usually helpful to look at the sequence of intermediate points (b_k) with $b_k \in P_B(x_k)$ and $x_{k+1} \in P_A(b_k)$ ($k = 0, 1, \dots$). We denote the joining sequence by (z_k) , that is

$$z_{2n} = x_n \text{ and } z_{2n+1} = b_n, \quad (n = 0, 1, \dots). \quad (3.20)$$

For simplicity of presentation let us assume throughout the discussion, without loss of generality, that $x_0 \in A$.

Bregman [32] and Gubin et al [56] showed that, if $A \cap B \neq \emptyset$ and the sets are closed and convex, the sequence converges to a point in $A \cap B$. In the case of two subspaces, this fact was established by von Neumann in the mid-1930s; that is why the method of alternating projections is sometimes referred to as *von Neumann's method*. It was noted in [118] that alternating projections can be traced back to the 1869 work by Schwarz. It was shown in [56] that, if $\text{ri } A \cap \text{ri } B \neq \emptyset$, the convergence is R-linear. A systematic analysis of the convergence of alternating projections in the convex setting was done by Bauschke and Borwein [12, 13], who demonstrated that it is the subtransversality property that is needed to ensure R-linear convergence.

In this section, let us consider the consistent feasibility problem of finding $x \in A \cap B$. Let $\bar{x} \in A \cap B$.

PROPOSITION 3.3.1. *Let $A, B \subset \mathbb{E}$ be closed and convex, and $\bar{x} \in A \cap B$. If $\{A, B\}$ is subtransversal at \bar{x} , then alternating projections converge locally linearly with rate at most $1 - \text{sr}[A, B](\bar{x})^2$.*

In fact, Section 4.3 will show that subtransversality in the convex setting is not only sufficient but also necessary for linear convergence of alternating projections. The picture becomes much more complicated if the convexity assumption is dropped. We next recall the two approaches for proving linear convergence of nonconvex alternating projections.

It was established in [59] that subtransversality of the collection of sets (with a reasonably good quantitative constant as always for convergence analysis of nonconvex alternating projections) is sufficient for linear monotonicity of the method for (ε, δ) -subregular sets. This result is updated here in light of more recent terminology.

PROPOSITION 3.3.2 (convergence of alternating projections with nonempty intersection). [101, Proposition 4.9] Let $S \subset A \cap B \neq \emptyset$. Let U be a neighborhood of S and suppose that

$$P_A U \subseteq U \quad \text{and} \quad P_B U \subseteq U.$$

Let Λ be an affine subspace of \mathbb{E} with $\Lambda \supset S$ such that $T_{AP} := P_A P_B : \Lambda \rightrightarrows \Lambda$. Define $F := T_{AP} - \text{Id}$. Let the sets A and B be elementally subregular at all $\bar{x} \in S$ relative to S respectively for each

$$\begin{aligned} (x, v_A) \in V_A &:= \{(z, w) \in \text{gph } N_A^{\text{prox}} \mid z + w \in U \quad \text{and} \quad z \in P_A(z + w)\} \\ (x, v_B) \in V_B &:= \{(z, w) \in \text{gph } N_B^{\text{prox}} \mid z + w \in U \quad \text{and} \quad z \in P_B(z + w)\} \end{aligned}$$

with respective constants $\varepsilon_A, \varepsilon_B \in [0, 1)$ on the neighborhood U . Suppose that the following hold:

- (a) for each $\bar{x} \in S$, the collection $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant κ on the neighborhood U ;
- (b) there exists a positive constant σ such that condition (3.8) holds true;
- (c) $\text{dist}(x, S) \leq \text{dist}(x, A \cap B \cap \Lambda)$ for all $x \in U \cap \Lambda$.

Then every sequence $(x_k)_{k \in \mathbb{N}}$ generated by $x_{k+1} \in T_{AP} x_k$ seeded by any point $x_0 \in A \cap U \cap \Lambda$ is linearly monotone with respect to S with constant

$$c := \sqrt{1 + \tilde{\varepsilon}_A + \tilde{\varepsilon}_B + \tilde{\varepsilon}_A \tilde{\varepsilon}_B - \frac{1}{(\kappa \sigma)^2}}$$

with

$$\tilde{\varepsilon}_{A/B} := 4\varepsilon_{A/B} \frac{1 + \varepsilon_{A/B}}{(1 - \varepsilon_{A/B})^2}.$$

Consequently, if

$$\tilde{\varepsilon}_A + \tilde{\varepsilon}_B + \tilde{\varepsilon}_A \tilde{\varepsilon}_B < \frac{1}{(\kappa \sigma)^2}.$$

then $\text{dist}(x_k, S) \rightarrow 0$ at least linearly with rate $c < 1$.

If $S = A \cap B \cap \Lambda$ in the above theorem, then assumption (d) in Proposition 3.2.7 can obviously be removed.

The next proposition catalogs the mentioned results which complements Proposition 3.2.7.

PROPOSITION 3.3.3 (R-linear convergence of nonconvex alternating projections). *For A and B closed with nonempty intersection and $\bar{x} \in A \cap B$, the alternating projections algorithm locally converges R -linearly if one of the following collection of conditions holds.*

- (i) [91, Theorem 4.3] A and B are smooth manifolds around \bar{x} and $\{A, B\}$ is transversal at \bar{x} .
- (ii) [51, Theorem 6.1] $\{A, B\}$ is intrinsically transversal at \bar{x} .
- (iii) [90, Theorem 5.16] A is super-regular at \bar{x} and $\{A, B\}$ is transversal at \bar{x} .
- (iv) [20, Theorem 3.17] A is (B, ε, δ) -regular at \bar{x} and the (A, B) -qualification condition holds at \bar{x} .
- (v) [118, Theorem 2] A is 0-Hölder regular relative to B at \bar{x} and $\{A, B\}$ intersects separably at \bar{x} .

REMARK 3.3.4. *Amongst criteria listed in Proposition 3.3.3, item (v) has been known to use weakest regularity assumptions [118].*

An attempt on comparing the two approaches of Propositions 3.3.2 and 3.3.3 leads to the following unified criterion for linear convergence of alternating projections for consistent feasibility. The following theorem indeed establishes the weakest known sufficient condition for linear convergence of alternating projections.

THEOREM 3.3.5 (unified criterion for linear convergence of T_{AP}). *Suppose that*

- (i) A is 0-Hölder at \bar{x} relative to B with constant $\gamma \in [0, 1)$ on a neighborhood U of \bar{x} ;
- (ii) there is a positive constant $\kappa < \sqrt{\frac{1-\gamma}{\gamma}}$ such that

$$\text{dist}(x, A \cap B) \leq \kappa \|x - x^+\| \quad \forall x \in U \text{ and } x^+ \in T_{AP}x. \quad (3.21)$$

Then every sequence $(x_k)_{k \in \mathbb{N}}$ generated by $x_{k+1} \in T_{AP}x_k$ seeded by any point sufficiently close to \bar{x} is linearly convergent to a point in $A \cap B$.

Proof. Let us denote $c := \max \left\{ \sqrt{\frac{1}{1-\gamma} - \frac{1}{\kappa^2}}, \frac{1}{1+\gamma^2} \right\}$ and note that $c < 1$ as $\kappa < \sqrt{\frac{1-\gamma}{\gamma}}$ as assumed. Without loss of generality, it can be assumed that $c \in (0, 1)$. Let $\delta > 0$ be so small that $\mathbb{B}_{\frac{2\delta}{1-c}}(\bar{x}) \subset U$.

Take any $x \in A \cap \mathbb{B}_\delta(\bar{x})$ and $x^+ \in T_{AP}x$ and let $b \in P_B(x)$ such that $x^+ \in P_A(b)$. We consider the two cases of b relative to x and x^+ as follows.

Case 1: $\|b - x\| \geq (1 + \gamma^2)\|b - x^+\|$. Then

$$\|b - x^+\| \leq \frac{1}{1 + \gamma^2} \|b - x\|. \quad (3.22)$$

Case 2: $\|b - x\| < (1 + \gamma^2)\|b - x^+\|$. Note that both x and b are in U . By definition of 0-Hölder regularity, we have that

$$\langle b - x^+, x - x^+ \rangle \leq \gamma \|b - x^+\| \|x - x^+\|.$$

Then

$$\begin{aligned} \|x - b\|^2 &= \|x - x^+\|^2 + \|b - x^+\|^2 - 2 \langle b - x^+, x - x^+ \rangle \\ &\geq \|x - x^+\|^2 + \|b - x^+\|^2 - 2\gamma \|b - x^+\| \|x - x^+\| \\ &= (1 - \gamma) (\|x - x^+\|^2 + \|b - x^+\|^2) + \gamma (\|x - x^+\| - \|b - x^+\|)^2. \\ &\geq (1 - \gamma) (\|x - x^+\|^2 + \|b - x^+\|^2). \end{aligned}$$

This together with inequality (3.21) implies that

$$\begin{aligned} \frac{1}{1 - \gamma} \|x - b\|^2 &\geq \|x - x^+\|^2 + \|b - x^+\|^2 \\ &\geq \frac{1}{\kappa^2} \text{dist}(x, A \cap B) + \|b - x^+\|^2 \\ &\geq \frac{1}{\kappa^2} \text{dist}(x, B) + \|b - x^+\|^2 \\ &= \frac{1}{\kappa^2} \|x - b\|^2 + \|b - x^+\|^2. \end{aligned}$$

Hence

$$\|b - x^+\| \leq \sqrt{\frac{1}{1 - \gamma} - \frac{1}{\kappa^2}} \|b - x\|. \quad (3.23)$$

A combination of (3.22) and (3.23) yields that

$$\|b - x^+\| \leq c \|b - x\| \quad \forall x \in A \cap \mathbb{B}_\delta(\bar{x}), b \in P_B(x), x^+ \in P_A(b). \quad (3.24)$$

Using (3.24) and noting the choice of δ , one can infer from the induction procedure in [90, Theorem 2] that verify that every sequence (x_k) generated by $x_{k+1} \in T_{AP}x_k$ seeded by any point $A \cap \mathbb{B}_\delta(\bar{x})$ is linearly extendible with frequency 2 and rate c . Proposition 2.6 of [101] then implies that (x_k) converges linearly with rate \sqrt{c} to a point \tilde{x} , which belongs to $A \cap B$ due to the closeness of the sets and the nature of alternating projections. The proof is complete. \square

Condition (3.21) is also known as the *coercivity* property of T_{AP} .

We next clarify the relationships amongst regularity notions imposed in Propositions 3.3.2 and 3.3.3 and Theorem 3.3.5 in order to demonstrate the unification of the latter result. The elemental regularity condition imposed in Proposition 3.3.2 implies the 0-Hölder regularity imposed in the other two results [84, Proposition 4]. The relationships amongst the subtransversality, separate intersection and coercivity properties are more fundamental for explaining the relationships amongst the above results.

LEMMA 3.3.6. *The two regularity conditions imposed in Proposition 3.3.2 imply the coercivity property of T_{AP} . As a consequence, Theorem 3.3.5 theoretically encompasses Proposition 3.3.2.*

Proof. Let $x \in A$ be sufficiently close to \bar{x} , $b \in P_B(x)$ and $x^+ \in P_A(b)$. The assumptions of Proposition 3.3.2 imply the linear monotonicity of T_{AP} with respect to $A \cap B$ with some constant $c \in [0, 1)$ (see the proof of [59, Corollary 3.13]), in particular,

$$\text{dist}(x^+, A \cap B) \leq c \text{dist}(x, A \cap B).$$

This property obviously implies the coercivity of T_{AP} since

$$\|x - x^+\| \geq \text{dist}(x, A \cap B) - \text{dist}(x^+, A \cap B) \geq (1 - c) \text{dist}(x, A \cap B).$$

The proof is complete. □

LEMMA 3.3.7. *The two regularity conditions imposed in Proposition 3.3.3 (v) imply the coercivity property of T_{AP} . As a consequence, Theorem 3.3.5 theoretically encompasses Proposition 3.3.3 in view of Remark 3.3.4.*

Proof. Let $x \in A$ be sufficiently close to \bar{x} , $b \in P_B(x)$ and $x^+ \in P_A(b)$. The assumptions of Proposition Proposition 3.3.3 (v) imply the linear extendibility of T_{AP} with frequency 2 and some constant $c \in [0, 1)$ (see the proof of [118, Theorem 2]), in particular,

$$\|b - x^+\| \leq c \|b - x\|.$$

Thanks to [101, Theorem 4.16], linear extendibility of T_{AP} with frequency 2 and some constant c also implies the subtransversality of $\{A, B\}$ at \bar{x} , in particular,

$$\text{dist}(x, A \cap B) \leq \frac{2}{1 - c} \text{dist}(x, B).$$

A combination of the two inequalities yields the coercivity property as claimed:

$$\|x - x^+\| \geq \|b - x\| - \|b - x^+\| \geq (1 - c) \|b - x\| = (1 - c) \text{dist}(x, B) \geq \frac{(1 - c)^2}{2} \text{dist}(x, A \cap B).$$

The proof is complete. □

LEMMA 3.3.8. *When the sets A and B are convex, the coercivity property of T_{AP} is equivalent to the subtransversality of $\{A, B\}$ at \bar{x} .*

Proof. The implication that subtransversality implies the coercivity property is covered in Lemma 3.3.6 since convexity implies elemental regularity, see also the original work in the convex setting [13]. Thanks to Lemma 3.1 of [59], the coercivity property implies linear monotonicity of all iterations generated by T_{AP} seeded sufficiently close to \bar{x} . The latter property in turn implies subtransversality of $\{A, B\}$ at \bar{x} as claimed thanks to Theorem 4.12 of [101]. \square

In view of Lemmas 3.3.6, 3.3.7 and 3.3.8, Theorem 3.3.5 theoretically unifies all existing criteria for linear convergence of alternating projections for consistent feasibility in both convex and nonconvex settings. We note that each of the above convergence criteria requires its own technical constraint on the quantitative constants of the relevant regularity notions, however, it seems challenging to make a rigorous comparison amongst such technical constraints.

We conclude this section with a specific result about alternating projections on the product space which will be applied directly to the source location problem in Section 5.1. Given a collection of sets $\{A_1, A_2, \dots, A_m\}$, we define the two sets in the cartesian product space \mathbb{E}^m as follows:

$$\mathcal{A} := \prod_{i=1}^m A_i, \quad \mathcal{D} := \{(x, x, \dots, x) \in \mathbb{E}^m \mid x \in \mathbb{E}\}.$$

It is well known that alternating projections for the two sets \mathcal{A} and \mathcal{D} corresponds exactly to the averaged projections for the m sets $\{A_1, A_2, \dots, A_m\}$ [126]:

$$P_{\mathcal{D}}P_{\mathcal{A}}([x]_m) = \left[\frac{1}{m} \sum_{i=1}^m P_{A_i} x \right]_m \quad \forall x \in \mathbb{E}.$$

THEOREM 3.3.9 (linear convergence from strong subtransversality and prox-regularity). *Consider a collection of prox-regular sets $\{A_1, A_2, \dots, A_m\}$ and suppose that it is strongly subtransversal at $\bar{x} \in \cap_{i=1}^m A_i$. Then every sequence generated by $P_{\mathcal{D}}P_{\mathcal{A}}$ converges linearly to $[\bar{x}]_m$ provided that the initial point is sufficiently close to $[\bar{x}]_m$.*

Proof. By the strong subtransversality assumption, there exist $\kappa > 0$ and $\Delta > 0$ such that $(\cap_{i=1}^m A_i) \cap \mathbb{B}_{2\Delta}(\bar{x}) = \{\bar{x}\}$ and

$$\|x - \bar{x}\| = \text{dist}(x, \cap_{i=1}^m A_i) \leq \kappa \max_{1 \leq i \leq m} \text{dist}(x, A_i) \quad \forall x \in \mathbb{B}_{\Delta}(\bar{x}). \quad (3.25)$$

By the prox-subregularity assumption, for any given $\varepsilon \in (0, 1)$, there exists $\delta_{\varepsilon} > 0$ such that

$$\langle x - P_{A_i} x, \bar{x} - P_{A_i} x \rangle \leq \varepsilon \|x - P_{A_i} x\| \|\bar{x} - P_{A_i} x\| \quad \forall x \in \mathbb{B}_{\delta_{\varepsilon}}(\bar{x}), \forall i = \overline{1, m}.$$

This in particular implies that for any given $\varepsilon \in (0, 1)$, there exists $\delta_\varepsilon > 0$ such that

$$\langle u - P_{\mathcal{A}}u, \bar{u} - P_{\mathcal{A}}u \rangle \leq \varepsilon \|u - P_{\mathcal{A}}u\| \|\bar{u} - P_{\mathcal{A}}u\| \quad \forall x \in \mathbb{B}_{\delta_\varepsilon}(\bar{x}), u = [x]_m, \bar{u} = [\bar{x}]_m. \quad (3.26)$$

Fix a number $\varepsilon > 0$ satisfying

$$\varepsilon + \varepsilon^2 + \frac{\varepsilon}{2(1-\varepsilon)} < \frac{1}{m\kappa^2} \quad (3.27)$$

and a corresponding $\delta_\varepsilon > 0$ satisfying condition (3.26). Let us define $\delta = \min\{\delta_\varepsilon, \Delta\} > 0$ and show that every sequence generated by T_{AP} starting at $[x]_m$ with $x \in \mathbb{B}_\delta(\bar{x})$ converges linearly to $[\bar{x}]_m$.

Indeed, consider any $x \in \mathbb{B}_\delta(\bar{x})$, $u = [x]_m$, $\bar{u} = [\bar{x}]_m$ and $u^+ \in P_{\mathcal{D}}P_{\mathcal{A}}u$. Due to the prox-regularity assumption, we can assume the singleton of the projections involved in this proof. By (3.25) and the choice of δ , we have that

$$\|x - \bar{x}\| \leq \kappa \max_{1 \leq i \leq m} \text{dist}(x, A_i). \quad (3.28)$$

Since $\max_{1 \leq i \leq m} \text{dist}(x, A_i) \leq \text{dist}(u, \mathcal{A})$, the inequality (3.28) implies that

$$\|u - \bar{u}\| = \sqrt{m} \|x - \bar{x}\| \leq \sqrt{m\kappa} \max_{1 \leq i \leq m} \text{dist}(x, A_i) \leq \sqrt{m\kappa} \text{dist}(u, \mathcal{A}) = \sqrt{m\kappa} \|u - P_{\mathcal{A}}u\|. \quad (3.29)$$

Using the Cauchy-Schwarz inequality and (3.26), we get

$$\begin{aligned} \|u - \bar{u}\|^2 &= \|u - P_{\mathcal{A}}u\|^2 + \|P_{\mathcal{A}}u - \bar{u}\|^2 - 2 \langle u - P_{\mathcal{A}}u, \bar{u} - P_{\mathcal{A}}u \rangle \\ &\geq \|u - P_{\mathcal{A}}u\|^2 + \|P_{\mathcal{A}}u - \bar{u}\|^2 - 2\varepsilon \|u - P_{\mathcal{A}}u\| \|\bar{u} - P_{\mathcal{A}}u\| \\ &\geq (1 - \varepsilon) \left(\|u - P_{\mathcal{A}}u\|^2 + \|P_{\mathcal{A}}u - \bar{u}\|^2 \right). \end{aligned} \quad (3.30)$$

Plugging (3.29) into (3.30) we get

$$\begin{aligned} \|P_{\mathcal{A}}u - \bar{u}\|^2 &\leq \frac{1}{1-\varepsilon} \|u - \bar{u}\|^2 - \|u - P_{\mathcal{A}}u\|^2 \\ &\leq \frac{1}{1-\varepsilon} \|u - \bar{u}\|^2 - \frac{1}{m\kappa^2} \|u - \bar{u}\|^2 \\ &= \left(\frac{1}{1-\varepsilon} - \frac{1}{m\kappa^2} \right) \|u - \bar{u}\|^2. \end{aligned} \quad (3.31)$$

Since $u^+ = P_{\mathcal{D}}P_{\mathcal{A}}u$ and \mathcal{D} is a subspace containing u and \bar{u} , we have

$$\begin{aligned} \|u - u^+\|^2 &= \|u - P_{\mathcal{A}}u\|^2 - \|u^+ - P_{\mathcal{A}}u\|^2 \\ &= \|u - P_{\mathcal{A}}u\|^2 - \|P_{\mathcal{A}}u - \bar{u}\|^2 + \|u^+ - \bar{u}\|^2. \end{aligned} \quad (3.32)$$

Plugging (3.29) and (3.31) into (3.32) we get

$$\begin{aligned} \|u - u^+\|^2 &\geq \frac{1}{m\kappa^2} \|u - \bar{u}\|^2 - \left(\frac{1}{1-\varepsilon} - \frac{1}{m\kappa^2} \right) \|u - \bar{u}\|^2 + \|u^+ - \bar{u}\|^2 \\ &= \left(\frac{2}{m\kappa^2} - \frac{1}{1-\varepsilon} \right) \|u - \bar{u}\|^2 + \|u^+ - \bar{u}\|^2. \end{aligned} \quad (3.33)$$

Thanks to Theorem 2.1.6(iii) the projectors P_{A_i} ($i = 1, 2, \dots, m$) are almost firmly nonexpansive at \bar{x} with violation at most $2\varepsilon + 2\varepsilon^2$ on $\mathbb{B}_\delta(\bar{x})$. Proposition 1.3.10(i) then ensures that

$$\|u^+ - \bar{u}\|^2 + \|u - u^+\|^2 \leq (1 + 2\varepsilon + 2\varepsilon^2) \|u - \bar{u}\|^2. \quad (3.34)$$

Plugging (3.33) into (3.34), we obtain

$$\|u^+ - \bar{u}\|^2 + \left(\frac{2}{m\kappa^2} - \frac{1}{1-\varepsilon} \right) \|u - \bar{u}\|^2 + \|u^+ - \bar{u}\|^2 \leq (1 + 2\varepsilon + 2\varepsilon^2) \|u - \bar{u}\|^2.$$

Equivalently,

$$\|u^+ - \bar{u}\|^2 \leq \left(1 + \varepsilon + \varepsilon^2 + \frac{\varepsilon}{2(1-\varepsilon)} - \frac{1}{m\kappa^2} \right) \|u - \bar{u}\|^2.$$

Due to (3.27) we have $c := \left(1 + \varepsilon + \varepsilon^2 + \frac{\varepsilon}{2(1-\varepsilon)} - \frac{1}{m\kappa^2} \right) < 1$ and the proof is complete. \square

REMARK 3.3.10. *Since the parameter $\varepsilon \downarrow 0$ as $u^k \rightarrow \bar{u}$, the rate c estimated above tends to $1 - \frac{1}{m\kappa^2}$ which is governed by the modulus of the strong subtransversality property. Compared to Remark 3.2.14, we see that the estimated rates for the cyclic projections and the averaged projections are very much the same.*

REMARK 3.3.11. *Again, Theorem 3.3.9 is encompassed in the framework of Theorem 3.2.7 and the proof given above can be viewed as a shortcut for verifying the assumptions of that theorem.*

3.4 Forward–backward algorithms

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} . The content of this section is taken from our joint work with Dr. Matthew K. Tam [103].

We consider the structured optimization problem

$$\underset{x \in \mathbb{E}}{\text{minimize}} \quad f(x) + g(x) \quad (\mathcal{P})$$

under different assumptions on the functions f and g . At the very least, we will assume that these functions are proper, lower semicontinuous functions.

We consider the ubiquitous *forward-backward* algorithm: given $x^0 \in \mathbb{E}$, generate the sequence $(x^k)_{k \in \mathbb{N}}$ via

$$x^{k+1} \in T_{\text{FB}}(x^k) := \text{prox}_{1,g}(x^k - t\nabla f(x^k)). \quad (3.35)$$

We keep the step-length fixed for simplicity. This is a reasonable strategy, obviously, when f is continuously differentiable with Lipschitz continuous gradient and when g is convex (not necessarily smooth), which we will assume throughout this subsection. For the case that g is the indicator function of a set C , that is $g = \iota_C$, then (3.35) is just the projected gradient algorithm for constrained optimization with a smooth objective. For simplicity, we will take the proximal parameter $\lambda = 1$ and use the notation prox_g instead of $\text{prox}_{1,g}$. The following discussion uses the property of hypomonotonicity (Definition 1.3.9(b)).

PROPOSITION 3.4.1 (almost averaged: steepest descent). *[103, Proposition 3.6] Let U be a nonempty open subset of \mathbb{E} . Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a continuously differentiable function with calm gradient at \bar{x} and calmness modulus L on the neighborhood U of \bar{x} . In addition, let ∇f be pointwise hypomonotone at \bar{x} with violation constant τ on U . Choose $\beta > 0$ and let $t \in (0, \beta)$. Then the mapping $T_{t,f} := \text{Id} - t\nabla f$ is pointwise almost averaged at \bar{x} with averaging constant $\alpha = t/\beta \in (0, 1)$ and violation constant $\varepsilon = \alpha(2\beta\tau + \beta^2 L^2)$ on U . If ∇f is pointwise strongly monotone at \bar{x} with modulus $|\tau| > 0$ (that is, pointwise hypomonotone with constant $\tau < 0$) and calm with modulus L on U and $t < 2|\tau|/L^2$, then $T_{t,f}$ is pointwise averaged at \bar{x} with averaging constant $\alpha = tL^2/(2|\tau|) \in (0, 1)$ on U .*

Note the trade-off between the step-length and the averaging property: the smaller the step, the smaller the averaging constant. In the case that ∇f is not monotone, the violation constant of nonexpansivity can also be chosen to be arbitrarily small by choosing β arbitrarily small, regardless of the size of the hypomonotonicity constant τ or the Lipschitz constant L . This will be exploited in Theorem 3.4.4 below. If ∇f is strongly monotone, the theorem establishes an upper limit on the stepsize for which nonexpansivity holds, but this does not rule out the possibility that, even for nonexpansive mappings, it might be more efficient to take a larger step that technically renders the mapping only *almost nonexpansive*. As we have seen in Theorem 3.1.3, if the fixed point set is *attractive* enough, then linear convergence of the iteration can still be guaranteed, even with this larger stepsize. This yields a local justification of *extrapolation*, or excessively large stepsizes.

PROPOSITION 3.4.2 (almost averaged: nonconvex forward-backward). *[103, Proposition 3.7] Let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper and l.s.c. with nonempty, pointwise Type-I nonmonotone subdifferential at all points on $S'_g \subset U'_g$ with violation τ_g on U'_g , that is, at each $w \in \partial g(v)$ and $v \in S'_g$ the inequality*

$$-\tau_g \|(u+z) - (v+w)\|^2 \leq \langle z-w, u-v \rangle$$

holds whenever $z \in \partial g(u)$ for $u \in U'_g$. Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a continuously differentiable function with calm gradient (modulus L) which is also pointwise hypomonotone at all $\bar{x} \in S_f \subset U_f$ with violation constant τ_f on U_f . For $T_{t,f} := \text{Id} - t\nabla f$, suppose that $T_{t,f}U_f \subset U_g$ where $U_g := \{u + z \mid u \in U'_g, z \in \partial g(u)\}$ and that $T_{t,f}S_f \subset S_g$ where $S_g := \{v + w \mid v \in S'_g, w \in \partial g(v)\}$. Choose $\beta > 0$ and $t \in (0, \beta)$. Then the forward–backward mapping $T_{\text{FB}} := \text{prox}_g(\text{Id} - t\nabla f)$ is pointwise almost averaged at all $\bar{x} \in S_f$ with violation constant $\varepsilon = (1 + 2\tau_g)(1 + t(2\tau_f + \beta L^2)) - 1$ and averaging constant α on U_f where

$$\alpha = \begin{cases} \frac{2}{3}, & \text{for all } \alpha_0 \leq \frac{1}{2}, \\ \frac{2\alpha_0}{\alpha_0 + 1}, & \text{for all } \alpha_0 > \frac{1}{2}, \end{cases} \quad \text{and} \quad \alpha_0 = \frac{t}{\beta}. \quad (3.36)$$

COROLLARY 3.4.3 (almost averaged: semi-convex forward–backward). [*103, Corollary 3.2*] Let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper, l.s.c. and convex. Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a continuously differentiable function with calm gradient (calmness modulus L) which is also pointwise hypomonotone at all $\bar{x} \in S_f \subset U_f$ with violation constant τ_f on U_f . Choose $\beta > 0$ and $t \in (0, \beta)$. Then the forward–backward mapping $T_{\text{FB}} := \text{prox}_g(\text{Id} - t\nabla f)$ is pointwise almost averaged at all $\bar{x} \in S_f$ with violation constant $\varepsilon = t(2\tau_f + \beta L^2)$ and averaging constant α given by (3.36) on U_f .

As the above proposition shows, the almost averaging property comes relatively naturally. A little more challenging is to show that Assumption (b) of Theorem 3.1.3 holds for a given application. The next theorem is formulated in terms of metric subregularity, but for the forward–backward iteration, the graphical derivative characterization given in Proposition 2.2.4 can allow for a direct verification of the regularity assumptions.

THEOREM 3.4.4 (local linear convergence: forward–backward). [*103, Theorem 3.3*] Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a continuously differentiable function with calm gradient (modulus L) which is also pointwise hypomonotone at all $\bar{x} \in \text{Fix } T_{\text{FB}} \subset U_f$ with violation constant τ_f on U_f . Let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper and l.s.c. with nonempty, pointwise Type-I nonmonotone subdifferential at all $v \in S'_g \subset U'_g$, with violation τ_g on U'_g whenever $z \in \partial g(u)$ for $u \in U'_g$. For $T_{t,f} := \text{Id} - t\nabla f$ let $T_{t,f}U_f \subset U_g$ where $U_g := \{u + z \mid u \in U'_g, z \in \partial g(u)\}$ and let $T_{t,f}\text{Fix } T_{\text{FB}} \subset S_g$ where $S_g := \{v + w \mid v \in S'_g, w \in \partial g(v)\}$. If, for all $t \geq 0$ small enough, $F_{\text{FB}} := T_{\text{FB}} - \text{Id}$ is metrically subregular for 0 on U_f with modulus $\kappa \leq \bar{\kappa} < 1/(2\sqrt{\tau_g})$, then for all t small enough, the forward–backward iteration $x^{k+1} \in T_{\text{FB}}x^k$ satisfies $\text{dist}(x^k, \text{Fix } T_{\text{FB}}) \rightarrow 0$ at least linearly for all x^0 close enough to $\text{Fix } T_{\text{FB}}$. In particular, if g is convex, and $\bar{\kappa}$ is finite, then the distance of the iterates to $\text{Fix } T_{\text{FB}}$ converges linearly to zero from any initial point x^0 close enough provided that the stepsize t is sufficiently small.

COROLLARY 3.4.5 (global linear convergence: convex forward–backward). [*103, Corollary 3.3*] Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a continuously differentiable function with calm gradient (modulus L)

which is also pointwise strongly monotone at all $\bar{x} \in \text{Fix } T_{\text{FB}}$ on \mathbb{R}^n . Let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper, convex and l.s.c. Let $T_{t,f} \text{Fix } T_{\text{FB}} \subset S_g$ where

$$S_g := \{v + w \mid v \in S'_g, w \in \partial g(v)\}.$$

If, for all $t \geq 0$ small enough, $F_{\text{FB}} := T_{\text{FB}} - \text{Id}$ is metrically subregular for 0 on \mathbb{R}^n with modulus $\kappa \leq \bar{\kappa} < +\infty$, then for all fixed step-length t small enough, the forward-backward iteration $x^{k+1} = T_{\text{FB}}x^k$ satisfies $\text{dist}(x^k, \text{Fix } T_{\text{FB}}) \rightarrow 0$ at least linearly for all $x^0 \in \mathbb{R}^n$.

REMARK 3.4.6 (extrapolation). In Corollary 3.4.5 it is not necessary to choose the stepsize small enough that T_{FB} is pointwise averaged. It suffices to choose the stepsize t small enough that $c := \sqrt{1 + \varepsilon - \frac{1}{2\bar{\kappa}^2}} < 1$ where $\varepsilon = \beta/2(2\tau_f + \beta L^2)$. In this case, T_{FB} is only almost pointwise averaged with violation ε on \mathbb{R}^n .

REMARK 3.4.7. Optimization problems involving the sum of a smooth function and a nonsmooth function are commonly found in applications and accelerations to forward-backward algorithms have been a subject of intense study [6, 24, 38, 112]. To this point the theory on quantitative convergence of the iterates is limited to the convex setting under the additional assumption of strong convexity/strong monotonicity. Theorem 3.4.4 shows that locally, convexity of the smooth function plays no role in the convergence of the iterates or the order of convergence, and strong convexity, much less convexity, of the function g is also not crucial - it is primarily the regularity of the fixed points that matters locally. This agrees nicely with recent global linear convergence results of a primal-dual method for saddle point problems that uses pointwise quadratic supportability in place of the much stronger strong convexity assumption [100]. Moreover, local linear convergence is guaranteed by metric subregularity on an appropriate set without any fine-tuning of the only algorithm parameter t , other than assuring that this parameter is small enough. When the nonsmooth term is the indicator function of some constraint set, then the regularity assumption can be replaced by the characterization in terms of the graphical derivative (2.9) to yield a familiar constraint qualification at fixed points.

If the functions in (\mathcal{P}) are piecewise linear-quadratic, then the forward-backward mapping has polyhedral structure (Proposition 3.4.9), which, following Proposition 3.1.7, allows for easy verification of the conditions for linear convergence (Proposition 3.4.10).

DEFINITION 3.4.8 (piecewise linear-quadratic functions). A function $f : \mathbb{R}^n \rightarrow [-\infty, +\infty]$ is called piecewise linear-quadratic if $\text{dom } f$ can be represented as the union of finitely many polyhedral sets, relative to each of which $f(x)$ is given by an expression of the form $\frac{1}{2}\langle x, Ax \rangle + \langle a, x \rangle + \alpha$ for some scalar $\alpha \in \mathbb{R}$ vector $a \in \mathbb{R}^n$, and symmetric matrix $A \in \mathbb{R}^{n \times n}$. If f can be represented by a single linear-quadratic equation on \mathbb{R}^n , then f is said to be linear-quadratic.

For instance, if f is piecewise linear-quadratic, then the subdifferential of f and its proximal mapping prox_f are polyhedral [129, Proposition 12.30].

PROPOSITION 3.4.9 (polyhedral forward–backward). [103, Proposition 3.8] *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be quadratic and let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper, l.s.c. and piecewise linear-quadratic convex. The mapping T_{FB} defined by (3.35) is single-valued and polyhedral.*

PROPOSITION 3.4.10 (linear convergence of polyhedral forward–backward). [103, Proposition 3.9] *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be quadratic and let $g : \mathbb{E} \rightarrow (-\infty, +\infty]$ be proper, l.s.c. and piecewise linear-quadratic convex. Suppose $\text{Fix } T_{\text{FB}}$ is an isolated point $\{\bar{x}\}$, where $T_{\text{FB}} := \text{prox}_g(\text{Id} - t\nabla f)$. Suppose also that the modulus of metric subregularity κ of $F := T_{\text{FB}} - \text{Id}$ at \bar{x} for 0 is bounded above by some constant $\bar{\kappa}$ for all $t > 0$ small enough. Then, for all t small enough, the forward–backward iteration $x^{k+1} = T_{\text{FB}}(x^k)$ converges at least linearly to \bar{x} whenever x^0 is close enough to \bar{x} .*

EXAMPLE 3.4.11 (iterative soft-thresholding). [103, Example 3.7] *Let $f(x) = x^T A x + x^T b$ and $g(x) = \alpha \|Bx\|_1$ for $A \in \mathbb{R}^{n \times n}$ symmetric and $B \in \mathbb{R}^{m \times n}$ full rank. The forward–backward algorithm applied to the problem minimize $f(x) + g(x)$ is the iterative soft-thresholding algorithm [43] with fixed step-length t in the forward step $x - t\nabla f(x) = x - t(2Ax + b)$. The function g is piecewise linear, so prox_g is polyhedral hence the forward–backward fixed point mapping T_{FB} is single-valued and polyhedral. As long as $\text{Fix } T_{\text{FB}}$ is an isolated point relative to the affine hull of the iterates $x^{k+1} = T_{\text{FB}}x^k$, and the modulus of metric subregularity is independent of the stepsize t for all t small enough, then, by Proposition 3.4.10 for small enough stepsize t the iterates x^k converge linearly to $\text{Fix } T_{\text{FB}}$ for all starting points close enough to $\text{Fix } T_{\text{FB}}$. If A is positive definite (i.e., f is convex) then the set of fixed points is a singleton and convergence is linear from any starting point x^0 .*

As a special case, the forward–backward algorithm with parameter $\lambda \in (0, 2]$ for feasibility of two sets takes the form

$$x_{k+1} \in T_{\text{FB}}(x_k) := P_A((1 - \lambda)x_k + \lambda P_B(x_k)), \quad (k = 0, 1, \dots). \quad (3.37)$$

Following the analysis of [90, Theorem 5.2], one can obtain the following convergence result.

THEOREM 3.4.12 (linear convergence: forward–backward for feasibility). *Suppose that $\{A, B\}$ is transversal at \bar{x} and A is super-regular at \bar{x} . Then the forward–backward algorithm (3.37) converges locally linearly around \bar{x} .*

3.5 Douglas–Rachford algorithm and its relaxations

The underlying space in this section is a finite dimensional Euclidean space \mathbb{E} . The first half of this section is taken from our joint work with Dr. Matthew K. Tam [103] while the rest has been published recently in [133].

The Douglas–Rachford algorithm is commonly encountered in one form or another for solving both feasibility problems and structured optimization. In the context of problem

(\mathcal{P}) the iteration takes the form

$$x^{k+1} \in T_{\text{DR}}(x^k) := \frac{1}{2} (R_f R_g + \text{Id})(x^k). \quad (3.38)$$

where $R_f := 2 \text{prox}_f - \text{Id}$ (i.e., the proximal reflector) and R_g is similarly given.

Revisiting the setting of [96], we use the tools developed in the present paper to show when one can expect local linear convergence of the Douglas–Rachford iteration. For simplicity, as in [96], we will assume that f is convex in order to arrive at a clean final statement, though convexity is not needed for local linear convergence.

PROPOSITION 3.5.1. [103, Proposition 3.10] *Let $g = \iota_C$ for $C \subset \mathbb{E}$ a manifold, and let $f : \mathbb{E} \rightarrow \mathbb{R}$ be convex and linear-quadratic. Fix $\bar{x} \in \text{Fix } T_{\text{DR}}$. Then for any $\varepsilon > 0$ small enough, there exists $\delta > 0$ such that T_{DR} is single-valued and almost firmly nonexpansive with violation $\varepsilon_g = 4\varepsilon + 4\varepsilon^2$ on $\mathbb{B}_\delta(\bar{x})$.*

THEOREM 3.5.2. [103, Theorem 3.4] *Let $g = \iota_C$ for $C \subset \mathbb{E}$ a manifold and let $f : \mathbb{E} \rightarrow \mathbb{R}$ be linear-quadratic convex. Let $(x^k)_{k \in \mathbb{N}}$ be iterates of the Douglas–Rachford (3.38) algorithm and let $\Lambda = \text{aff}(x^k)$. If $T_{\text{DR}} - \text{Id}$ is metrically subregular at all points $\bar{x} \in \text{Fix } T_{\text{DR}} \cap \Lambda \neq \emptyset$ relative to Λ then for all x^0 close enough to $\text{Fix } T_{\text{DR}} \cap \Lambda$, the sequence x^k converges linearly to a point in $\text{Fix } T \cap \Lambda$ with constant at most $c = \sqrt{1 + \varepsilon} - 1/\kappa^2 < 1$ where κ is the constant of metric subregularity for $F := T_{\text{DR}} - \text{Id}$ on some neighborhood U containing the sequence and ε is the violation of almost firm nonexpansiveness on the neighborhood U .*

REMARK 3.5.3. [103, Remark 3.5] *Assuming that the fixed points, restricted to the affine hull of the iterates, are isolated points, polyhedrality was used in [4] to verify that the Douglas–Rachford mapping is indeed metrically subregular at the fixed points. While in principle the graphical derivative formulas (see Proposition 2.2.4) could be used for more general situations, it is not easy to compute the graphical derivative of the Douglas–Rachford operator, even in the simple setting above. This is a theoretical bottleneck for the practical applicability of metric subregularity for more general algorithms.*

Applied to feasibility problems, the Douglas–Rachford algorithm is also described as averaged alternating reflections [17]. Here, both $f = \iota_A$ and $g = \iota_B$ are the indicator functions of individual constraint sets. When the sets A and B are sufficiently regular, as they certainly are in the phase retrieval problem, and intersect transversally, local linear convergence of the Douglas–Rachford algorithm in this instance was established in [125]. As discussed in Example 3.1.5, however, for any phase retrieval problem arising from a physical noncrystallographic diffraction experiment, the constraint sets cannot intersect when finite support is required of the reconstructed object. This fact, seldom acknowledged in the phase retrieval literature, is borne out in the observed instability of the Douglas–Rachford algorithm applied to phase retrieval [95]: it cannot converge when the constraint sets do not intersect [17, Theorem 3.13].

To address this issue, a relaxation for nonconvex feasibility was studied in [95, 96] that amounts to (3.38) where f is the Moreau envelope of a nonsmooth function and g is the

indicator function of a sufficiently regular set. Optimization problems with this structure are guaranteed to have solutions. In particular, when f is the Moreau envelope to ι_A with parameter λ , the corresponding iteration given by (3.38) can be expressed as a convex combination of the underlying basic Douglas–Rachford operator and the projector of the constraint set encoded by g [96, Proposition 2.5]:

$$x^{k+1} \in T_{RAAR}x^k := \frac{1}{2(\lambda+1)}(R_AR_B + \text{Id})(x^k) + \frac{\lambda}{\lambda+1}P_Bx^k$$

where $R_A = 2P_A - \text{Id}$, $R_B = 2P_B - \text{Id}$ and $\beta = \frac{1}{\lambda+1}$. In [95] and the physics literature this is known as *relaxed alternating averaged reflections*.

We consider RAAR for the special structure of feasibility problem of

$$\text{finding } \bar{y} \in A\chi \cap \mathcal{Y}, \quad (3.39)$$

where $\chi = \mathbb{C}^n$ and $A : \mathbb{C}^n \rightarrow \mathbb{C}^N$ ($N \geq 2n$) is linear isometric and $\mathcal{Y} \subset \mathbb{C}^N$ given by

$$\mathcal{Y} := \{y \in \mathbb{C}^N \mid |y| = b \text{ pointwise}\}, \quad \text{for a given } b \in \mathbb{R}_+^N.$$

This model is for the phase retrieval problem in Fourier domain and the RAAR algorithm for solving (3.39) takes the explicit form

$$T_{RAAR}(y) = \beta \left(y + AA^* \left(2b \odot \frac{y}{|y|} - y \right) - \frac{2\beta - 1}{\beta} b \odot \frac{y}{|y|} \right). \quad (3.40)$$

The next linear convergence result is formulated in [93] which extends the analogous result for the Douglas–Rachford algorithm [39, Theorem 5.1].

THEOREM 3.5.4. [93, Theorem 4] *Let $\bar{x} \in \mathbb{C}^n$ and $A \in \mathbb{C}^{N \times n}$ isometric with $N \geq 2n$. Let $\bar{y} = A\bar{x}$, $b = |\bar{y}|$ and suppose*

$$b_{\min} = \min_{1 \leq j \leq N} b(j) > 0. \quad (3.41)$$

Let

$$B := \text{diag} \left(\frac{\bar{y}^*}{|\bar{y}|} \right) A \quad (3.42)$$

and suppose that

$$\lambda_2 := \max \{ \|\text{Im}(Bu)\| : u \in \mathbb{C}^n, \|u\| = 1, u \perp i\bar{x} \} < 1. \quad (3.43)$$

Let y^k be an RAAR iteration sequence with $y^1 = Ax^1$ and $x^k = A^*y^k$. If x^1 is sufficiently close to \bar{x} , then for some constant $\eta < 1$,

$$\|x^k - \bar{x}\|_{\text{Opt}} \leq \eta^{k-1} \|x^1 - \bar{x}\|_{\text{Opt}},$$

where

$$\|x^k - \bar{x}\|_{\text{Opt}} := \min_{c \in \mathbb{C}, |c|=1} \|cx^k - \bar{x}\|.$$

We consider another relaxed version of the Douglas–Rachford method [133]:

$$T_{DRAP}x = P_B(P_Ax + \lambda(P_Ax - x)) - \lambda(P_Ax - x),$$

where $\lambda \in (0, 1)$ is a parameter.

REMARK 3.5.5. T_{DRAP} with $\lambda = 0$ is the alternating projections method and T_{DRAP} with $\lambda = 1$ is the Douglas–Rachford method.

Following the lines of [125, Theorem 4.3] and [59, Theorem 3.18], as preliminary results we obtain local linear convergence of T_{DRAP} under the assumption of transversality.

Let us consider the consistent feasibility for two sets $\{A, B\}$ and let $\bar{x} \in A \cap B$.

The next lemma is an extension of [125, Lemma 4.2].

LEMMA 3.5.6. [133, Lemma 3] Suppose that $\{A, B\}$ is transversal at \bar{x} , i.e., $\bar{\theta} < 1$ where $\bar{\theta}$ is defined by (2.26). Then for any $\theta \in (\bar{\theta}, 1)$, there exists a number $\delta > 0$ such that for all $x \in \mathbb{B}_\delta(\bar{x})$ and $x^+ \in T_{DRAP}x$,

$$\kappa \operatorname{dist}(x, A \cap B) \leq \|x - x^+\|,$$

where κ is defined by

$$\kappa := \frac{(1 - \theta)\sqrt{1 + \bar{\theta}}}{\sqrt{2} \max\{1, \lambda + \sqrt{1 - \theta^2}\}} > 0. \quad (3.44)$$

We are now ready to prove local linear convergence for algorithm T_{DRAP} which generalizes the corresponding results established in [59, 125] for the DR method.

THEOREM 3.5.7 (linear convergence of algorithm T_{DRAP}). [133, Theorem 4] Suppose that $\{A, B\}$ is transversal at \bar{x} . Let $\theta \in (\bar{\theta}, 1)$ where $\bar{\theta}$ is defined by (2.26). Suppose that A and B are (ε, δ) -regular at \bar{x} with $\tilde{\varepsilon} < \frac{(1 + \lambda)\kappa^2}{2}$, where κ and $\tilde{\varepsilon}$ are respectively given by (3.44) and

$$\tilde{\varepsilon} := 2(2\varepsilon + 2\varepsilon^2) + (1 + \lambda)(2\varepsilon + 2\varepsilon^2)^2.$$

Then every iteration $x_{k+1} \in T_{DRAP}x_k$ starting sufficiently close to \bar{x} converges R -linearly to a point in $A \cap B$.

REMARK 3.5.8. Theorem 3.5.7 remains valid if the transversality assumption is weakened to the transversality relative to the affine hull of $A \cup B$.

3.6 ADMM algorithms

The underlying space in this section is a finite dimensional Euclidean space.

The following minimization problem which covers both source location and phase retrieval problems discussed in Section 5.

$$\underset{\mathbb{E} \times \mathbb{E}}{\text{minimize}} F(x, u) := \frac{1}{2} \|x - a\|^2 - \langle u, x - a \rangle + \iota_{\mathcal{A}}(x) + \iota_{\mathcal{B}}(u), \quad (3.45)$$

where \mathcal{A} and \mathcal{B} are convex sets in \mathbb{E} .

The augmented Lagrangian for the problem (3.45) is

$$L_{\rho}(x, u, v, w) = \frac{1}{2} \|x - a\|^2 - \langle v, x - a \rangle + \langle w, u - v \rangle + \frac{\rho}{2} \|u - v\|^2.$$

Here $\rho > 0$ is the penalty parameter and $w \in \mathbb{E}$ is the multiplier corresponding to the constraint $u - v = 0$. We will always assume that $\rho > 2$.

The basic ADMM algorithm for solving (3.45) can be rewritten as a projection algorithm.

Algorithm 3.6.1. [99] *For any starting point $(x^0, u^0, v^0, w^0) \in \mathbb{E}^4$, one generates an iteration $\{\mathbf{y}^k := (x^k, u^k, v^k, w^k)\}_{k \in \mathbb{N}}$ as follows:*

$$\begin{aligned} x^{k+1} &\in P_{\mathcal{A}}(a + v^k), \\ u^{k+1} &\in P_{\mathcal{B}}(v^k - \rho^{-1}w^k), \\ v^{k+1} &= \frac{1}{\rho} \left(\rho u^{k+1} + x^{k+1} - a + w^k \right), \\ w^{k+1} &= a - x^{k+1}. \end{aligned}$$

Algorithm 3.6.1 determines a set-valued operator $T : \mathbb{E}^4 \rightrightarrows \mathbb{E}^4$ which assigns each input (x, u, v, w) with $T(x, u, v, w)$ consisting of all points (x^+, u^+, v^+, w^+) generated by the main loop of Algorithm 3.6.1. It was shown in [99] that fixed points of T are critical points of F .

Denote $\mathbf{y}^k = (x^k, u^k, v^k, w^k)$, $k \in \mathbb{N}$.

The following global convergence of Algorithm 3.6.1 was established in [99].

THEOREM 3.6.2 (Global convergence of Algorithm 3.6.1). *Let $\{\mathbf{y}^k\}_{k \in \mathbb{N}}$ be a sequence generated by Algorithm 3.6.1. Then the sequence $\{\mathbf{y}^k\}_{k \in \mathbb{N}}$ converges globally to some point $\mathbf{y}^* = (\bar{x}, \bar{u}, \bar{v}, \bar{w})$ with (\bar{x}, \bar{u}) being a critical point of F .*

Chapter 4

Necessary conditions for convergence

In recent years there has been a lot of progress in determining ever weaker conditions to guarantee local linear convergence of elementary fixed point algorithms, with particular attention given to the method of alternating projections and the Douglas–Rachford iteration [20, 51, 59, 90, 91, 118, 125]. These works beg the question: what are necessary conditions for linear convergence? We shed some light on this question for expansive fixed point iterations and show how our theory specializes for the alternating projections iteration in nonconvex and convex settings. The content of this chapter is taken from our joint papers with Prof. Marc Teboulle and Dr. Matthew K. Tam [101, 102].

4.1 Existence of implicit error bounds

The underlying space in this section is an infinite dimensional Hilbert space if not otherwise specified.

We first present necessary conditions for the existence of a gauge-type subregularity property – what we refer to as an *implicit error bound*.

The next lemma will be referred to frequently in the subsequent development.

LEMMA 4.1.1. [102, Lemma 1] *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ satisfy $\text{Fix } T \neq \emptyset$. Let $U \subset \mathcal{H}$ with $U \cap \text{Fix } T \neq \emptyset$. Define the set-valued map $S : \mathbb{R}_+ \rightrightarrows \mathcal{H}$ by*

$$S(t) := \{y \in \mathcal{H} : \text{dist}(y, Ty) \leq t\}$$

and define the function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ by

$$\kappa(t) := \sup_{y \in S(t) \cap U} \{\text{dist}(y, \text{Fix } T)\}. \tag{4.1}$$

The following assertions hold.

(i) The set $S(t)$ is a nonempty subset of $\text{dom } T$ for all $t \geq 0$ and satisfies

$$\emptyset \neq \text{Fix } T = S(0) \subset S(s) \subset S(t) \quad \forall t \geq s \geq 0.$$

(ii) The function κ is nonnegative, nondecreasing, $\kappa(0) = 0$ and satisfies

$$\text{dist}(x, \text{Fix } T) \leq \kappa(\|x - Tx\|) \quad \forall x \in U. \quad (4.2)$$

If any of the following hold, then κ is bounded:

- (a) there is a bounded set V with $S(t) \cap U \subset V$ for all t ;
- (b) the function $\text{dist}(\cdot, \text{Fix } T)$ is bounded on U .

The next results show that nonexpansiveness alone is enough to guarantee the existence of an error bound. This is remarkable since, without asymptotic regularity, the fixed point iteration need not even converge.

THEOREM 4.1.2 (error bounds of nonexpansive operators: finite dimensional version). [102, Theorem 3] *Let \mathcal{H} be a finite dimensional Hilbert space. Suppose that $T : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive with $\text{Fix } T \neq \emptyset$. Then, for each bounded set U containing a fixed point of T , the nondecreasing function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defined by (4.1) is bounded, right-continuous at $t = 0$ with $\kappa(0) = 0$ and satisfies (4.2).*

Note that the proof Theorem 4.1.2 is not valid in infinite dimensions, since in this case the bounded sequence (y_n) need only contain a weakly convergent subsequence and the $\text{dist}(\cdot, \text{Fix } T)$ need not be weakly (sequentially) continuous.

REMARK 4.1.3 (Infinite dimensional counterexamples). [102, Remark 2] *In general, the assumption of finite dimensionality of \mathcal{H} in Theorem 4.1.2 cannot be dropped. Indeed, if \mathcal{H} is infinite dimensional, then a concrete counterexample is provided by any averaged operator with a fixed point, T , for which there is a starting point, $x_0 \in \mathcal{H}$, such that the sequence $(T^n x_0)_{n=0}^\infty$ converges weakly but not strongly. The explicit constructions of such an examples can be found, for instance, in [55] and in [62].*

We make the following observation.

LEMMA 4.1.4. [102, Lemma 2] *Let \mathcal{H} be a Hilbert space, and let $T : \mathcal{H} \rightarrow \mathcal{H}$ be averaged with $\text{Fix } T \neq \emptyset$. For each Picard iteration (x_n) generated by T from a starting point $x_0 \in \mathcal{H}$, let us define $d_0 := \text{dist}(x_0, \text{Fix } T)$ and $\underline{d} := \lim_{n \rightarrow \infty} \text{dist}(x_n, \text{Fix } T)$. Then there exists a continuous and nondecreasing function $\mu : [\underline{d}, d_0] \rightarrow [\underline{d}, d_0]$ satisfying $\mu(t) < t$ for all $t \in (\underline{d}, d_0]$ such that*

$$\text{dist}(x_{n+1}, \text{Fix } T) = \mu(\text{dist}(x_n, \text{Fix } T)) \quad \forall n \in \mathbb{N}. \quad (4.3)$$

Proof. Since the proof is constructive and needed in the subsequent analysis, it is presented here for completeness. Let us denote $d_n := \text{dist}(x_n, \text{Fix } T)$ for all $n \in \mathbb{N}$. We first claim that there exists a sequence $(c_n) \subset [0, 1)$, dependent on x_0 , such that

$$d_{n+1} = c_n d_n \quad \forall n \in \mathbb{N}. \quad (4.4)$$

For any $N \in \mathbb{N}$, if $x_{N+1} \in \text{Fix } T$, then one can take $c_n = 0$ for all $n > N$. Suppose, then, that $x_{n+1} \notin \text{Fix } T$, hence $x_n \notin \text{Fix } T$ and $x_n \neq x_{n+1}$. In particular, $\|x_n - x_{n+1}\| > 0$. Since T is averaged, there is a constant $\gamma > 0$ such that

$$d_{n+1}^2 \leq d_n^2 - \gamma \|x_n - x_{n+1}\|^2.$$

Consequently, we have $0 < d_{n+1} < d_n$ and it follows that

$$c_n := \frac{d_{n+1}}{d_n} \in (0, 1)$$

is well-defined and satisfies (4.4).

We next define the piecewise linear function, μ , on $[\underline{d}, d_0]$ such that

$$\mu(\underline{d}) := \underline{d}, \quad \mu(d_n) := c_n d_n \quad \forall n \in \mathbb{N}. \quad (4.5)$$

and, on each interval of the form $[d_{n+1}, d_n]$, the value of μ is given by a linear interpolation of its values defined by (4.5).

To complete the proof, we check that μ is nondecreasing on $[\underline{d}, d_0]$. By the construction of μ , the sequence $(\mu(d_n))$ is nonincreasing as $n \rightarrow \infty$. It suffices to check that μ is nondecreasing on each (nontrivial) interval $[d_{n+1}, d_n]$. Indeed, let $d_{n+1} \leq t_1 < t_2 \leq d_n$, then

$$\begin{aligned} \mu(t_1) &= \mu(d_{n+1}) + \frac{t_1 - d_{n+1}}{d_n - d_{n+1}} (\mu(d_n) - \mu(d_{n+1})) \\ &\leq \mu(d_{n+1}) + \frac{t_2 - d_{n+1}}{d_n - d_{n+1}} (\mu(d_n) - \mu(d_{n+1})) = \mu(t_2). \end{aligned}$$

□

PROPOSITION 4.1.5. [102, Proposition 1] *Let \mathcal{H} be a Hilbert space and consider an operator $T : \mathcal{H} \rightarrow \mathcal{H}$ with $\text{Fix } T \neq \emptyset$. Let $(x_n)_{n \in \mathbb{N}}$ be a Picard sequence such that $\text{dist}(x_n, \text{Fix } T) \rightarrow 0$. Then the function κ defined by (4.1) with $U := (x_n)_{n \in \mathbb{N}}$ is nonnegative, nondecreasing, bounded, $\kappa(0) = 0$ and satisfies*

$$\text{dist}(x_n, \text{Fix } T) \leq \kappa(\|x_n - Tx_n\|) \quad \forall n \in \mathbb{N}.$$

In addition, if T is averaged, then the sequence $(x_n)_{n \in \mathbb{N}}$ converges strongly to some point x in $\text{Fix } T$ and the function κ is right continuous at 0.

It is clear from the above observation that, in order to obtain a meaningful error bound, a suitable function κ needs to be found for *all* possible starting points on a bounded set containing fixed points of T . Nevertheless, the sequence (c_n) given by Lemma 4.1.4 does characterize strong convergence of the corresponding iteration (x_n) . More specifically, we have the following.

PROPOSITION 4.1.6 (equivalences). [102, Proposition 2] *Let \mathcal{H} be a Hilbert space, let $T : \mathcal{H} \rightarrow \mathcal{H}$ be averaged with $\text{Fix } T \neq \emptyset$ and let (x_n) be a Picard iteration generated by T with initial point $x_0 \in \mathcal{H}$. The following statements are equivalent.*

- (i) (x_n) converges strongly to a point x in \mathcal{H} .
- (ii) (x_n) converges strongly to a point x in $\text{Fix } T$.
- (iii) $\text{dist}(x_n, \text{Fix } T)$ converges to zero.
- (iv) There exists a nondecreasing function $\mu : [0, d_0] \rightarrow [0, d_0]$ satisfying $\mu(t) < t$ for all $t \in [0, d_0]$ such that (4.3) holds and $\mu^n(\text{dist}(x_0, \text{Fix } T)) \rightarrow 0$ as $n \rightarrow \infty$.

REMARK 4.1.7. [102, Remark 3] *The function μ in Proposition 4.1.6(iv) characterizes the convergence rate of (x_n) .*

- (i) *When μ is majorized by a linear function with slope $c \in [0, 1)$ on some interval $[0, \tau)$ where $\tau > 0$, that is,*

$$\mu(\text{dist}(x_n, \text{Fix } T)) \leq c \text{dist}(x_n, \text{Fix } T) \quad \forall n \text{ sufficiently large}$$

– *equivalently, the sequence (c_n) defined in (4.4) satisfies $c := \sup_{n \in \mathbb{N}} c_n < 1$ – then we have a linearly monotone sequence as defined in [101] and R-linear convergence as detailed in [16, Theorem 5.12].*

- (ii) *When $\mu^n(\text{dist}(x_0, \text{Fix } T))$ tends to zero slower or faster than a linear rate, the sequence (x_n) is said to converge sublinearly or superlinearly, respectively. An example of sublinear convergence corresponding to $\mu(t) = \frac{t}{\sqrt{t^2+1}}$ for all $t \in [0, \text{dist}(x_0, \text{Fix } T)]$ is detailed in Example 4.1.10 below.*

In order to deduce a uniform version of the previous results, a property which holds uniformly on U is needed.

THEOREM 4.1.8 (sufficient condition for an error bound). [102, Theorem 4] *Let \mathcal{H} be a Hilbert space, let $T : \mathcal{H} \rightarrow \mathcal{H}$ with $\text{Fix } T \neq \emptyset$, let U be a bounded subset of \mathcal{H} containing a fixed point of T . Suppose that there exists a function $c : [0, \infty) \rightarrow [0, 1]$ which is upper semi-continuous on $(0, \text{exc}(U, \text{Fix } T))$ and satisfies $c(t) < 1$ for all t in this interval such that*

$$\text{dist}(Tx, \text{Fix } T) \leq c(\text{dist}(x, \text{Fix } T)) \text{dist}(x, \text{Fix } T) \quad \forall x \in U. \quad (4.6)$$

Then the nonnegative, nondecreasing function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defined by (4.1) is bounded, right-continuous at $t = 0$ and satisfies (4.2).

EXAMPLE 4.1.9 (arbitrarily slow convergence). [102, Example 2] There are two things to point out about the theorem above, both hinging on the choice of the subset U . The first point is that it is possible to choose U such that no c satisfying the requirements of the theorem exists. We demonstrate this when U is simply a ball. Such a phenomenon shows that uniform linear error bounds are not always possible. The second point, however, is that when an iteration converges it is always possible to choose a set U such that a function c exists satisfying the requirements of Theorem 4.1.8, but the resulting error bound may not always be informative. We also show an example of this below.

To put the above results in context, consider the method of alternating projections for finding the intersection of two closed subspaces of a Hilbert space, call them A and B . The alternating projections fixed point mapping is $T := P_A P_B$ with $\text{Fix } T_{AB} = A \cap B$. Von Neumann showed that the iterates of the method of alternating projections converges strongly to the projection of the starting point onto the intersection [134]. In the mid 1950's a rate was established in terms of what is known as the Friedrich's angle [54] between the sets defined as the number in $[0, \frac{\pi}{2}]$ whose cosine is given by

$$c(A, B) := \sup \left\{ |\langle a, b \rangle| \mid \begin{array}{l} a \in A \cap (A \cap B)^\perp, \quad \|a\| \leq 1, \\ b \in B \cap (A \cap B)^\perp, \quad \|b\| \leq 1. \end{array} \right\}$$

It is straightforward to see that $c(A, B) \leq 1$. Moreover, $c(A, B) < 1$ if and only if $A + B$ is closed [12, Lemma 4.10]. In this case, a bound on the rate of convergence in terms of the Friedrichs angle follows from the fact that [72]

$$\|T^n - P_{A \cap B}\| = c(A, B)^{2n-1} \quad \forall n \in \mathbb{N}. \quad (4.7)$$

If $A + B$ is not closed, then it was shown in [18] (i.e., $c(A, B) = 1$) that convergence can be arbitrarily slow in the sense that for any nonincreasing sequence $\lambda_n \rightarrow 0$ with $\lambda_0 < 1$, there is a starting point x_λ such that

$$\|T^n x_\lambda - P_{A \cap B} x_\lambda\| \geq \lambda_n \quad \forall n \in \mathbb{N}$$

In the context of Theorem 4.1.8, if $A + B$ is closed, then the function $c : [0, \infty) \rightarrow [0, 1]$ can be simply chosen to be the cosine of the Friedrichs angle [19, Theorem 3.16].

On the other hand, if $A + B$ is not closed, then no such function exists as soon as the bounded set U contains dilate of the sphere $S := \{x \in \mathcal{H} : \|x\| = 1\}$. To see this, suppose on the contrary, that there exists a function c satisfying Theorem 4.1.8. In particular, we have that $c(t) < 1$ ($t > 0$). Then for any $x \in S \subseteq U$ we have

$$\begin{aligned} \|Tx - P_{A \cap B} x\| &= \text{dist}(Tx, \text{Fix } T) \leq c(\text{dist}(x, \text{Fix } T)) \text{dist}(x, \text{Fix } T) \\ &= c(\text{dist}(x, \text{Fix } T)) \|x - P_{A \cap B} x\| \\ &\leq c(\text{dist}(x, \text{Fix } T)) \|x\|. \end{aligned}$$

Dividing both sides of the inequality by $\|x\|$, taking the supremum over S , and substituting (4.7) gives

$$1 \leq \sup_{x \in S} c(\text{dist}(x, \text{Fix } T)),$$

which contradicts the assumption that $c(t) < 1$ (as c satisfies Theorem 4.1.8). The choice of U to be a scaled ball is the natural choice when one is interested in uniform error bounds. This example shows that even for the simple alternating projections algorithm, such bounds are not always possible.

To the second point, if for the above example, instead of choosing U to be a ball, we restrict U to be the iterates x_n of the alternating projections sequence together with their limit x_∞ for a fixed x_0 , then we can construct a function c satisfying the assumptions of Theorem 4.1.8. Indeed, choose $c(t)$ to be a linear interpolation of the points

$$c(t_n) := \frac{\|Tx_n - x_\infty\|}{\|x_n - x_\infty\|} \quad \text{for } t_n = \|x_n - x_\infty\| \text{ whenever } \|x_n - x_\infty\| > 0.$$

Such a function satisfies the requirements of Theorem 4.1.8 and hence guarantees the existence of an error bound. But this is not informative, because the error bound depends on the iteration itself, and hence the initial guess x_0 . Returning to the fact that if $A + B$ is not closed the alternating projections algorithm exhibits arbitrarily slow convergence, then even though we have an error bound for a particular instance we cannot say anything about uniform rates of convergence.

The following example illustrates the role of the function c satisfying condition (4.6) as in Theorem 4.1.8.

EXAMPLE 4.1.10. [102, Example 3] Consider the alternating projections operator $T := P_A P_B$ for the two convex subsets A and B of \mathbb{R}^2 given by

$$A := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}, \quad B := \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + (x_2 - 1)^2 \leq 1\}.$$

Then we have $\text{Fix } T = A \cap B = \{0\}$ and the only set U of interest is $U = A$. For each $x \in U$, say $x = (t, 0)$, it holds $Tx = \left(\frac{t}{\sqrt{t^2+1}}, 0\right)$ and consequently

$$\text{dist}(x, \text{Fix } T) = |t|, \quad \text{dist}(Tx, \text{Fix } T) = \frac{|t|}{\sqrt{t^2+1}}, \quad \|x - Tx\| = |t| \left(1 - \frac{1}{\sqrt{t^2+1}}\right).$$

In this setting, we now can directly check the following statements.

(i) The function c defined by

$$c(t) := \frac{1}{\sqrt{t^2+1}}, \quad \forall t \in \mathbb{R}_+$$

satisfies all the assumptions of Theorem 4.1.8. It is worth emphasizing that for each $\alpha > 0$,

$$c_\alpha := \sup\{c(t) : t \geq \alpha\} = \frac{1}{\sqrt{\alpha^2 + 1}} < 1 \text{ while } \sup\{c(t) : t \geq 0\} = 1.$$

(ii) The function $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defined by

$$\varphi(t) := t \left(1 - \frac{1}{\sqrt{t^2 + 1}} \right), \quad \forall t \in \mathbb{R}_+,$$

is a gauge function and the desired function, κ , defined by (4.1) is the inverse function φ^{-1} which is also a gauge function.

(iii) This development is an extension of μ -monotonicity introduced in [101]. A sequence (x_k) on \mathcal{H} , is said to be μ -monotone with respect to Ω ($\emptyset \neq \Omega \subset \mathcal{H}$) if there exists a nonnegative function $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying $\mu(0) = 0$ and

$$\bar{\mu}(t_1, k_1) < \bar{\mu}(t_2, k_2) \text{ when } (t_1 < t_2 \text{ and } k_1 = k_2) \text{ or } (t_1 = t_2 \neq 0 \text{ and } k_1 > k_2)$$

with

$$(\forall k \in \mathbb{N}) \quad \text{dist}(x_{k+1}, \Omega) \leq \mu(\text{dist}(x_k, \Omega)).$$

In the present example, the sequence (x_n) generated by T is μ -monotone with respect to $\text{Fix } T$, where $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is given by

$$\mu(t) := \frac{t}{\sqrt{t^2 + 1}}, \quad \forall t \in \mathbb{R}_+.$$

REMARK 4.1.11. [102, Remark 4] Condition (4.6) can be viewed as the functional extension of the linear result in [101, Theorem 3.12] where linear monotonicity (part (ii) of Example 4.1.10) was shown to be sufficient for the existence of linear error bounds. Indeed, (4.6) is a realization of the notion of μ -monotonicity introduced in [101] in which the function μ has the form $\mu(t) := c(t) \cdot t$ for all $t \geq 0$. In particular, if $c(t) := c_0$ for some constant $c_0 < 1$, Theorem 4.1.8 recovers [101, Theorem 3.12].

Note that in Theorem 4.1.8, condition (4.6) is the only assumption required to obtain the error bound. An implicit consequence of the condition is that the distance of Picard iterates to $\text{Fix } T$ converges to zero as soon as T has a fixed point and that the initial point of the iteration is in a set U which satisfies $T(U) \subset U$.

PROPOSITION 4.1.12 (convergence to zero of the distance to fixed points). [102, Proposition 3] Let \mathcal{H} be a Hilbert space, let $T : \mathcal{H} \rightarrow \mathcal{H}$ with $\text{Fix } T \neq \emptyset$, and let U be a bounded subset containing a fixed point of T and $T(U) \subset U$. Suppose that there exists a function $c : [0, \infty) \rightarrow [0, 1]$ being upper semi-continuous on $(0, \text{exc}(U, \text{Fix } T)]$ and satisfying $c(t) < 1$ for all t in this interval such that condition (4.6) is satisfied. Then every Picard iteration (x_n) with $x_0 \in U$ generated by T satisfies $\text{dist}(x_n, \text{Fix } T) \rightarrow 0$ as $n \rightarrow \infty$.

In light of Proposition 4.1.12, Theorem 4.1.8 can be viewed as a uniform version of Proposition 4.1.5.

We discuss some insights of condition (4.6) in the averaged operator setting.

REMARK 4.1.13. [102, Remark 5] Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be averaged with $\text{Fix } T \neq \emptyset$.

(i) Lemma 4.1.4 implies that, for each $x \in \mathcal{H}$, there exists a number $c_x < 1$ such that

$$\text{dist}(Tx, \text{Fix } T) \leq c_x \text{dist}(x, \text{Fix } T).$$

Note that, the existence of a function c satisfying condition (4.6) would require that the supremum of all such numbers c_x taken over each level set $\mathcal{L}_t := \{x : \text{dist}(x, \text{Fix } T) = t\}$ exists and is less than 1. In this case, c can be any function which is upper semi-continuous on $(0, \text{exc}(U, \text{Fix } T)]$ and satisfies

$$\sup\{c_x : x \in \mathcal{L}_t\} \leq c(t) < 1, \quad \forall t > 0.$$

Note that the function $f : \mathcal{H} \rightarrow \mathbb{R}_+$ given by

$$f(x) := \begin{cases} \frac{\text{dist}(Tx, \text{Fix } T)}{\text{dist}(x, \text{Fix } T)} & \text{if } x \notin \text{Fix } T, \\ 0 & \text{if otherwise} \end{cases}$$

is continuous at all points $x \notin \text{Fix } T$ as a quotient of two continuous functions $\text{dist}(\cdot, \text{Fix } T)$ and $\text{dist}(T\cdot, \text{Fix } T)$ (because T is averaged). Thus, in particular, if \mathcal{H} is finite dimensional and $\text{Fix } T$ is bounded, then \mathcal{L}_t is compact and hence, for all $t > 0$, $\sup\{c_x : x \in \mathcal{L}_t\}$ is trivially less than one. In other words, for an averaged operator in a finite dimensional space, condition (4.6) in Theorem 4.1.8 is superfluous and only upper semi-continuity of c need be assumed.

- (ii) Condition (4.6) quantifies the rate of decrease of $\text{dist}(\cdot, \text{Fix } T)$ on each level set \mathcal{L}_t . More precisely, if $x_n \in \mathcal{L}_t$, then the distance to $\text{Fix } T$ will decrease by a factor of at least $c(t)$ in the next iterate x_{n+1} . Furthermore, a closer look at the proof of Proposition 4.1.12 shows that condition (4.6) can actually provide an estimate of the rate at which $\text{dist}(T^n x, \text{Fix } T) \rightarrow 0$ even in infinite dimensional setting.
- (iii) On one hand, Theorem 4.1.8 can be viewed as an attempt to extend Theorem 4.1.2 to infinite dimensional settings. On the other hand, it shows that an error bound in the form of (4.2) is a necessary condition for a certain type of μ -monotonicity (see Example 4.1.10 and Remark 4.1.11). More precisely, μ -monotonicity with μ of the form $\mu(t) = c(t)t$ for all $t \geq 0$ where c denotes the function in (4.6).

We next discuss the linear metric subregularity/error bounds as necessary conditions for linear convergence of fixed point iterations.

The following result shows that metric subregularity is *necessary* for linearly monotone sequences, without any assumptions about the averaging properties of T , almost or otherwise.

THEOREM 4.1.14 (necessity of metric subregularity). [101, Theorem 3.12] Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$, fix $\Omega \subset \Lambda \subset \mathbb{E}$ where $\text{Fix } T \cap \Lambda$ is closed and nonempty. If for each $x_0 \in \Omega$, every sequence $(x_k)_{k \in \mathbb{N}}$ generated by $x_{k+1} \in Tx_k \subset \Lambda$ is linearly monotone with respect to $\text{Fix } T \cap \Lambda$ with constant $c \in [0, 1)$, then the mapping $\Phi := T - \text{Id}$ is metrically subregular on Ω for 0 relative to Λ with constant $\kappa \leq \frac{1}{1-c}$.

COROLLARY 4.1.15 (necessary conditions for linear convergence). [101, Corollary 3.13] For a fixed number $\delta \in (0, \infty]$ let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ be almost averaged with violation ε and averaging constant α on $(\text{Fix } T + \mathbb{B}_\delta) \cap \Lambda$ where $\text{Fix } T$ is assumed closed and nonempty. If, for each $x_0 \in ((\text{Fix } T + \mathbb{B}_\delta) \cap \Lambda) \setminus \text{Fix } T$, every sequence $(x_k)_{k \in \mathbb{N}}$ generated by $x_{k+1} \in Tx_k \subset \Lambda$ is linearly monotone with respect to $\text{Fix } T \cap \Lambda$ with constant $c \in [0, 1)$, then all such sequences converge R -linearly with rate c to some point in $\text{Fix } T \cap \Lambda$ and $\Phi := T - \text{Id}$ is metrically subregular on $(\text{Fix } T + \mathbb{B}_\delta) \setminus \text{Fix } T$ for 0 relative to Λ with constant $\kappa \leq \frac{1}{1-c}$.

4.2 Necessary conditions for linear convergence of alternating projections

The underlying space in this section is a finite dimensional Hilbert space.

The next theorem shows that the converse to Proposition 3.3.2 holds more generally without any assumption on the elemental regularity of the individual sets. Its proof uses the idea in the proof of [51, Theorem 6.2].

THEOREM 4.2.1 (necessary condition for linear monotonicity). [101, Theorem 4.12 with $n = 1$] Let A and B be closed sets with $\bar{x} \in S \subset A \cap B$. Let Λ be an affine subspace containing S and $c \in [0, 1)$. Suppose that every sequence of alternating projections starting in Λ and sufficiently close to \bar{x} is contained in Λ and is linearly monotone with respect to S with constant c . Then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant $\text{sr}'[A, B](\bar{x}) \geq \frac{1-c}{2}$.

The next statement is an immediate consequence of Proposition 3.3.2 and Theorem 4.2.1.

COROLLARY 4.2.2 (subtransversality is necessary and sufficient for linear monotonicity). [101, Corollary 4.13] Let $\Lambda \subset \mathbb{E}$ be an affine subspace and let A and B be closed subsets of \mathbb{E} that are elementally subregular relative to $S \subset A \cap B \cap \Lambda$ at $\bar{x} \in S$ with constant ε and neighborhood $\mathbb{B}_\delta(\bar{x}) \cap \Lambda$ for all $(a, v) \in \text{gph } N_A^{\text{prox}}$ with $a \in \mathbb{B}_\delta(\bar{x}) \cap \Lambda$.

Suppose that every sequence of alternating projections with the starting point sufficiently close to \bar{x} is contained in Λ . All such sequences of alternating projections are linearly monotone with respect to S with constant $c \in [0, 1)$ if and only if the collection of sets is subtransversal at \bar{x} relative to Λ (with an adequate balance of quantitative constants).

The next technical lemma allows us to formally avoid the restriction “monotone” in Theorem 4.2.1.

LEMMA 4.2.3. [101, Lemma 4.14] *Let $(x_k)_{k \in \mathbb{N}}$ be a sequence generated by T_{AP} that converges R -linearly to $\bar{x} \in A \cap B$ with rate $c \in [0, 1)$. Then there exists a subsequence $(x_{k_n})_{n \in \mathbb{N}}$ that is linearly monotone with respect to any set $S \subset A \cap B$ with $\bar{x} \in S$.*

Proof. We present the proof for subsequent discussion. By definition of R -linear convergence, there is $\gamma < +\infty$ such that $\|x_k - \bar{x}\| \leq \gamma c^k$ for all $k \in \mathbb{N}$. Let S be any set such that $\bar{x} \in S \subset A \cap B$. If $x_{k_0} := x_0 \notin S$, i.e., $\text{dist}(x_{k_0}, S) > 0$, then there exists an iterate of $(x_k)_{k \in \mathbb{N}}$ (we choose the first one) relabeled x_{k_1} such that

$$\text{dist}(x_{k_1}, S) \leq \|x_{k_1} - \bar{x}\| \leq \gamma c^{k_1} \leq c \text{dist}(x_{k_0}, S). \quad (4.8)$$

Repeating this argument for x_{k_1} in place of x_{k_0} and so on, we extract a subsequence $(x_{k_n})_{n \in \mathbb{N}}$ satisfying

$$\text{dist}(x_{k_{n+1}}, S) \leq c \text{dist}(x_{k_n}, S) \quad \forall n \in \mathbb{N}.$$

The proof is complete. \square

The above observation allows us to obtain the statement about necessary conditions for linear convergence of the alternating projections algorithm which extends Theorem 4.2.1. Here, the index number k_1 depending on the sequence $(x_k)_{k \in \mathbb{N}}$ will come into play in determining the constant of linear regularity.

THEOREM 4.2.4 (subtransversality is necessary for linear convergence). [101, Theorem 4.15] *Let $m \in \mathbb{N}$ be fixed and $c \in [0, 1)$. Let Λ , A and B be closed subsets of \mathbb{E} and let $\bar{x} \in S \subset A \cap B \cap \Lambda$. Suppose that any alternating projections sequence $(x_k)_{k \in \mathbb{N}}$ starting in $A \cap \Lambda$ and sufficiently close to \bar{x} is contained in Λ , converges R -linearly to a point in S with rate c , and satisfies $k_1 \leq m$ where k_1 is determined as in (4.8). Then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant $\text{sr}'[A, B](\bar{x}) \geq \frac{1-c}{2m}$.*

THEOREM 4.2.5 (necessary condition for linear extendability). [101, Theorem 4.16 with $n = 1$] *Let Λ be an affine subspace, and let A and B be closed sets, $\bar{x} \in A \cap B \cap \Lambda$ and $c \in [0, 1)$. Suppose that for any alternating projections sequence $(x_k)_{k \in \mathbb{N}}$ starting in Λ and sufficiently close to \bar{x} , the joining sequence $(z_k)_{k \in \mathbb{N}}$ given by (3.20) is a linear extension of $(x_k)_{k \in \mathbb{N}}$ on Λ with frequency 2 and rate c . Then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant $\text{sr}'[A, B](\bar{x}) \geq \frac{1-c}{2}$.*

The joining alternating projections sequence $(z_k)_{k \in \mathbb{N}}$ given by (3.20) often plays a role as an intermediate step in the analysis of alternating projections. As we shall see, property of linear extendability itself can also be of interest when dealing with the alternating projections algorithm, especially for nonconvex setting. This observation can be seen for example in [20, 51, 90, 91, 118].

Theorems 4.2.1 and 4.2.5 remain valid if instead of the whole alternating projections sequence $(x_k)_{k \in \mathbb{N}}$, one supposes there exists a subsequence of form $(x_{j+nk})_{k \in \mathbb{N}}$ for some $j \in \{0, 1, \dots, n-1\}$ that fulfills the required property.

THEOREM 4.2.6 (subtransversality is necessary for linear monotonicity of subsequences). *[101, Theorem 4.12] Let Λ , A , and B be closed subsets of \mathbb{E} , let $\bar{x} \in S \subset A \cap B \cap \Lambda$, and let $1 \leq n \in \mathbb{N}$ and $c \in [0, 1)$ be fixed. Suppose that for any sequence of alternating projections $(x_k)_{k \in \mathbb{N}}$ starting in Λ and sufficiently close to \bar{x} , there exists a subsequence of the form $(x_{j+nk})_{k \in \mathbb{N}}$ for some $j \in \{0, 1, \dots, n-1\}$ that remains in Λ and is linearly monotone with respect to S with constant c . Then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant $\text{sr}'[A, B](\bar{x}) \geq \frac{1-c}{2(2n^2-1-c(n-1))}$.*

THEOREM 4.2.7 (subtransversality is necessary for linear extendability of subsequences). *[101, Theorem 4.16] Let Λ , A , and B be closed subsets of \mathbb{E} , let $\bar{x} \in A \cap B \cap \Lambda$, and let $1 \leq n \in \mathbb{N}$ and $c \in [0, 1)$ be fixed. Suppose that every alternating projections sequence $(x_k)_{k \in \mathbb{N}}$ starting in $A \cap \Lambda$ and sufficiently close to \bar{x} has a subsequence of the form $(x_{j+nk})_{k \in \mathbb{N}}$ for some $j \in \{0, 1, \dots, n-1\}$ such that the joining sequence $(z_k)_{k \in \mathbb{N}}$ given by (3.20) is a linear extension of (x_{j+nk}) on Λ with frequency $2n$ and rate c . Then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} relative to Λ with constant $\text{sr}'[A, B](\bar{x}) \geq \frac{1-c}{2(2n-1-c(n-1))}$.*

Note that Theorems 4.2.1 and 4.2.5 turn out to be special cases of Theorems 4.2.6 and 4.2.7, respectively with $n = 1$, i.e., the desired subsequence is actually the whole alternating projections sequence.

In general, subtransversality is not a sufficient condition for an alternating projections sequence to converge to a point in the intersection of the sets. For example, let us define the function $f : [0, 1] \rightarrow \mathbb{R}$ by $f(0) = 0$ and on each interval of form $(1/2^{n+1}, 1/2^n]$,

$$f(t) = \begin{cases} -t + 1/2^{n+1}, & \text{if } t \in (1/2^{n+1}, 3/2^{n+2}], \\ t - 1/2^n, & \text{if } t \in (3/2^{n+2}, 1/2^n], \end{cases} \quad (\forall n \in \mathbb{N})$$

and consider the sets: $A = \text{gph } f$ and $B = \{(t, t/3) \mid t \in [0, 1]\}$ and the point $\mathbf{x} = (0, 0) \in A \cap B$ in \mathbb{R}^2 . Then it can be verified that the collection of sets $\{A, B\}$ is subtransversal at \mathbf{x} while the alternating projections method gets stuck at points $(1/2^n, 0) \notin A \cap B$.

In the remainder of this section, we show that the property of subtransversality of the collection of sets has been imposed either explicitly or implicitly in all existing linear convergence criteria for the alternating projections method that we are aware of.

It can be recognized without much effort that under any item of Proposition 3.3.3, the sequences generated by alternating projections starting sufficiently close to \bar{x} are actually linearly extendible.

PROPOSITION 4.2.8 (ubiquity of subtransversality in linear convergence criteria). *[101, Proposition 4.18] Suppose that one of the conditions (i)–(v) of Proposition 3.3.3 is satisfied.*

Then for any alternating projections sequence $(x_k)_{k \in \mathbb{N}}$ starting sufficiently close to \bar{x} , the corresponding joining sequence $(z_k)_{k \in \mathbb{N}}$ given by (3.20) is a linear extension of $(x_k)_{k \in \mathbb{N}}$ with frequency 2 and rate $c \in [0, 1)$.

Taking Theorem 4.2.5 into account we conclude that subtransversality of the collection of sets $\{A, B\}$ at \bar{x} is a consequence of each item listed in Proposition 3.3.3. This observation gives some insights about relationships between various regularity notions of collections of sets and has been formulated partly in [51, Theorem 6.2] and [84, Theorem 4]. Hence, the subtransversality property lies at the foundation of all linear convergence criteria for the method of alternating projections for both convex and nonconvex sets appearing in the literature to this point.

Based on the results obtained in this section we conjecture that, for alternating projections applied to consistent feasibility, subtransversality is necessary for R-linear convergence of the iterates to fixed points, but not sufficient unless the sets are convex. On the other hand, transversality is sufficient, but is far from being necessary even in the convex case. For example, transversality always fails when the affine span of the union of the sets is not equal to the whole space, while alternating projections can still converge linearly as in the case when the sets are convex with nonempty intersection of their relative interiors. A quest has started for the weakest regularity property lying between transversality and subtransversality and being sufficient for the local linear convergence of alternating projections. We mention here the articles by Bauschke et al. [20, 21] utilizing *restricted normal cones*, Drusvyatskiy et al. [51] introducing and successfully employing *intrinsic transversality*, Noll and Rondepierre [118] introducing a concept of *separable intersection*, with 0-separability being a weaker property than intrinsic transversality and still implying the local linear convergence of alternating projections under the additional assumption that one of the sets is 0-Hölder regular at the reference point with respect to the other.

4.3 Further discussion on convex alternating projections

The underlying space in this section is a finite dimensional Hilbert space.

In the convex setting, statements with sharper convergence rate estimates are possible. This is the main goal of the present section. Note that a convex set is elementally regular at all points in the set for all normal vectors with constant $\varepsilon = 0$ and neighborhood \mathbb{E} [83, Proposition 4(vii)]. We can thus, without loss of generality, remove the restriction to the subset Λ that is omnipresent in the nonconvex setting. We also write $P_A x$ and $P_B x$ for the projections since the projectors are single-valued.

The next technical lemma is fundamental for the subsequent analysis.

LEMMA 4.3.1 (nondecrease of rate). [101, Lemma 5.1] *Let A and B be two closed convex sets in \mathbb{E} . We have*

$$\|P_B P_A P_B x - P_A P_B x\| \cdot \|P_B x - x\| \geq \|P_A P_B x - P_B x\|^2 \quad \forall x \in A.$$

Lemma 4.3.1 implies that for any sequence $(x_k)_{k \in \mathbb{N}}$ of alternating projections for convex sets, the rate $\frac{\|x_{k+1} - x_k\|}{\|x_k - x_{k-1}\|}$ is nondecreasing when k increases. This allows us to deduce the following fact about the algorithm.

THEOREM 4.3.2 (lower bound of complexity). *[101, Theorem 5.2] Consider the alternating projections algorithm for two closed convex sets A and B with a nonempty intersection. Then one of the following statements holds true.*

- (i) *The alternating projections method finds a solution after one iterate.*
- (ii) *Alternating projections will not reach a solution after any finite number of iterates.*

REMARK 4.3.3. *[101, Remark 5.3] In contrast to Theorem 4.3.2 for convex sets, there are simple examples of nonconvex sets such that for any given number $n \in \mathbb{N}$, the alternating projections method will find a solution after exactly n iterates. For instance, let us consider a geometric sequence $z_k = \left(\frac{1}{3}\right)^k z_0$ where $0 \neq z_0 \in \mathbb{E}$. For any number $n \in \mathbb{N}$, one can construct the two finite sets by $A := \{z_{2k} \mid k = 0, 1, \dots, n\}$ and $B := \{z_{2n}\} \cup \{z_{2k+1} \mid k = 0, 1, \dots, n-1\}$. Then the alternating projections method starting at z_0 will find the unique solution z_{2n} after exactly n iterates.*

THEOREM 4.3.4 (necessary and sufficient condition: local version). *[101, Theorem 5.4] Let A and B be closed convex sets and $\bar{x} \in A \cap B$. If the collection of sets $\{A, B\}$ is subtransversal at \bar{x} with constant $\text{sr}'[A, B](\bar{x}) \in (0, 1)$, then for any number $c \in (1 - \text{sr}'[A, B](\bar{x})^2, 1)$, all alternating projections sequences starting sufficiently close to \bar{x} are linearly monotone with respect to $A \cap B$ with rate not greater than c .*

Conversely, if there exists a number $c \in [0, 1)$ such that every alternating projections iteration starting sufficiently close to \bar{x} converges R -linearly to some point in $A \cap B$ with rate not greater than c , then the collection of sets $\{A, B\}$ is subtransversal at \bar{x} with constant $\text{sr}'[A, B](\bar{x}) \geq 1 - c$.

The next theorem is a global version of Theorem 4.3.4.

THEOREM 4.3.5 (necessary and sufficient condition: global version). *[101, Theorem 5.5] Let A and B be closed convex sets with nonempty intersection. If the collection of sets $\{A, B\}$ is subtransversal at every point of (the boundary of) $A \cap B$ with constants bounded from below by $\kappa \in (0, 1)$, then for any number $c \in (1 - \kappa^2, 1)$, every alternating projections iteration converges R -linearly to a point in $A \cap B$ with rate not greater than c .*

Conversely, if there exists a number $c \in [0, 1)$ such that every alternating projections sequence eventually converges R -linearly to a point in $A \cap B$ with rate not greater than c , then the collection of sets $\{A, B\}$ is globally subtransversal with constant $\kappa \geq 1 - c$, that is,

$$(1 - c) \text{dist}(x, A \cap B) \leq \text{dist}(x, B) \quad \forall x \in A.$$

It is clear that Theorem 4.3.4 does not cover Theorem 4.3.5. The following example also rules out the inverse inclusion.

EXAMPLE 4.3.6 (Theorem 4.3.5 does not cover Theorem 4.3.4). [101, Example 5.6] Consider the convex function $f : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$f(t) = \begin{cases} t^2, & \text{if } t \in [0, \infty), \\ 0, & \text{if } t \in [-1, 0), \\ -t - 1, & \text{if } t \in (-\infty, -1). \end{cases}$$

In \mathbb{R}^2 , we define two closed convex sets $A := \text{epi } f$ and $B := \mathbb{R} \times \mathbb{R}_-$ and a point $\bar{x} = (-1, 0) \in A \cap B$. Then the two equivalent properties (transversality of $\{A, B\}$ at \bar{x} and local linear convergence of T_{AP} around \bar{x}) involved in Theorem 4.3.4 hold true while the two global ones involved in Theorem 4.3.5 do not.

To establish global convergence of a fixed point iteration, one normally needs some kind of global regularity behavior of the fixed point set. In Theorem 4.3.5, we formally impose only subtransversality in order to deduce global R-linear convergence and vice versa. Beside the global behavior of convexity, the hidden reason behind this seemingly contradicting phenomenon is a well known fact about subtransversality of collections of convex sets. We next deduce this result from the convergence analysis above.

COROLLARY 4.3.7. [94, Theorem 8] Let A and B be closed and convex subsets of \mathbb{E} with nonempty intersection. The collection of sets $\{A, B\}$ is globally subtransversal, that is, there is a constant $\kappa > 0$ such that

$$\kappa \text{dist}(x, A \cap B) \leq \text{dist}(x, B) \quad \forall x \in A,$$

if and only if $\{A, B\}$ is subtransversal at every point in $\text{bd}(A \cap B)$ with constants bounded from below by some $\underline{\kappa} > 0$.

The convergence counterpart of Corollary 4.3.7 can also be of interest.

COROLLARY 4.3.8. [101, Corollary 5.8] Let $(x_k)_{k \in \mathbb{N}}$ be an alternating projections sequence for two closed convex subsets of \mathbb{E} with nonempty intersection and $c \in [0, 1)$. If there exists a natural number $p \in \mathbb{N}$ such that $\|x_k - \tilde{x}\| \leq \gamma c^k$ for all $k \geq p$, then $\|x_k - \tilde{x}\| \leq \gamma c^k$ for all $k \in \mathbb{N}$.

We emphasize that the two convergence properties appearing in Corollary 4.3.8 are always equivalent (by the argument for the second part of Theorem 4.3.5) if the constant γ is not required to be the same. However, this requirement becomes important when one wants to estimate global rate of convergence via the local rate of convergence. The next statement can easily be observed as a by-product via the proof of Theorem 4.3.4.

PROPOSITION 4.3.9 (equivalence of linear monotonicity and R-linear convergence). [101, Proposition 5.9] For sequences of alternating projections between convex sets, R-linear convergence and linear monotonicity of the sequence of iterates are equivalent.

The next statement can serve as a motivation for Definition 1.2.5.

PROPOSITION 4.3.10 (Q-linear convergence implies linear extendability). [*101, Proposition 5.10*] *Let $(x_k)_{k \in \mathbb{N}}$ be a sequence of alternating projections for two closed convex sets $A, B \subset \mathbb{E}$ with nonempty intersection. If $(x_k)_{k \in \mathbb{N}}$ converges Q-linearly to a point $\tilde{x} \in A \cap B$ with rate $c \in [0, 1)$, then $(x_k)_{k \in \mathbb{N}}$ is linearly extendible with frequency 2 and rate c , and the corresponding joining sequence $(z_k)_{k \in \mathbb{N}}$ is such a linear extension sequence.*

Chapter 5

Applications

The algorithms discussed in Chapter 3 are simulated for the source location and phase retrieval problems. Regularity properties from the problem data are discussed in accordance with the convergence theory of each method.

5.1 Source location problem

The ideal mathematical model for source location problem is geometrically very simple, find the unique common point of a collection of spheres

$$\text{find } \bar{x} \in \bigcap_{j=1}^m \mathcal{S}_j, \quad (5.1)$$

where \mathcal{S}_j ($j = 1, 2, \dots, m$) is the sphere in \mathbb{R}^n centered at a_j and with radius $r_j > 0$.

The simplicity of (5.1) provides a useful intuition of rather technical regularity notions involved in the convergence theory in Chapter 3.

Let us consider (5.1) in \mathbb{R}^3 and make the following natural assumption on the sensors. The treatment for the problem in n -dimensional case is analogous.

Assumption 5.1.1. *There are always three sensors $\{a_{j_1}, a_{j_2}, a_{j_3}\}$ that together with the true source \bar{x} are affinely independent.*

The following facts follow from the prox-regularity of the spheres and Assumption 5.1.1.

Fact 5.1.2 (prox-regularity of spheres). *Each \mathcal{S}_j ($j = 1, 2, \dots, m$) is prox-regular at \bar{x} , i.e., for any given $\varepsilon \in (0, 1)$ it holds*

$$\langle x - P_{\mathcal{S}_j}x, \bar{x} - P_{\mathcal{S}_j}x \rangle \leq \varepsilon \|x - P_{\mathcal{S}_j}x\| \|\bar{x} - P_{\mathcal{S}_j}x\| \quad \forall x \in \mathbb{B}_\delta(\bar{x}), \quad (5.2)$$

where $\delta := 2r_j\varepsilon\sqrt{1 - \varepsilon^2} > 0$.

For all $\delta > 0$ sufficiently small, the constant ε in (5.2) can be represented as a functional of δ ,

$$\varepsilon = f(\delta) := \frac{1}{\sqrt{2}} \left(1 - \sqrt{1 - \frac{\delta^2}{r_j^2}} \right)^{1/2} \in (0, 1/\sqrt{2}). \quad (5.3)$$

This function will be needed for estimating radius of linear convergence of algorithms. It is important to note that $f(\delta) \downarrow 0$ as $\delta \downarrow 0$.

Fact 5.1.3 (strong subtransversality). *Assumption 5.1.1 implies that $\{\mathcal{S}_{j_1}, \mathcal{S}_{j_2}, \mathcal{S}_{j_3}\}$ is strongly subtransversal at \bar{x} , that is, there exist $\kappa, \Delta > 0$ such that $(\cap_{i=1}^3 \mathcal{S}_{j_i}) \cap \mathbb{B}_{2\Delta}(\bar{x}) = \{\bar{x}\}$ and*

$$\|x - \bar{x}\| = \text{dist}(x, \cap_{i=1}^3 \mathcal{S}_{j_i}) \leq \kappa \max_{i=1,2,3} \text{dist}(x, \mathcal{S}_{j_i}) \quad \forall x \in \mathbb{B}_{\Delta}(\bar{x}).$$

Let us denote

$$r := \min\{r_j > 0 : 1 \leq j \leq m\} > 0.$$

5.1.1 Cyclic and averaged projections

The following theorem guarantees local linear convergence of T_{CP} for solving (5.1) under Assumption 5.1.1.

THEOREM 5.1.4 (linear convergence for T_{CP}). *Let $\delta \in (0, \min\{r, \Delta\})$ satisfy*

$$f(\delta) < \frac{1}{2m(\kappa + 1)},$$

where $f(\delta)$ is given by (5.3). Then for any starting point in $\mathbb{B}_{\delta}(\bar{x})$, the method T_{CP} for solving (5.1) converges linearly to \bar{x} with rate at most

$$c = \left(1 + \frac{2f(\delta)(\kappa + 1)}{\kappa^2} - \frac{1}{m\kappa^2} \right)^{1/2} \in (0, 1).$$

Proof. Assumption 5.1.1 implies the strong subtransversality of $\{\mathcal{S}_{j_1}, \mathcal{S}_{j_2}, \mathcal{S}_{j_3}\}$ at \bar{x} by Fact 5.1.3. The latter in turn implies the strong subtransversality of the collection $\{\mathcal{S}_i\}$ at \bar{x} . The statement now follows from Theorem 3.2.13 in view of Fact 5.1.2. \square

The following theorem guarantees local linear convergence of T_{AP} for solving (5.1) under Assumption 5.1.1.

THEOREM 5.1.5 (linear convergence for T_{AP}). *Let $\delta \in (0, \min\{r, \Delta\})$ satisfy*

$$f(\delta) + f(\delta)^2 + \frac{f(\delta)}{2(1 - f(\delta))} \leq \frac{1}{m\kappa^2},$$

where $f(\delta)$ is given by (5.3). Then for any starting point in $\mathbb{B}_\delta(\bar{x})$, the method T_{AP} for solving (5.1) converges linearly to \bar{x} with rate at most

$$c = \left(1 + f(\delta) + f(\delta)^2 + \frac{f(\delta)}{2(1-f(\delta))} - \frac{1}{m\kappa^2} \right)^{1/2}.$$

Proof. Assumption 5.1.1 implies the strong subtransversality of $\{\mathcal{S}_{j_1}, \mathcal{S}_{j_2}, \mathcal{S}_{j_3}\}$ at \bar{x} by Fact 5.1.3. The latter in turn implies the strong subtransversality of the collection $\{\mathcal{S}_i\}$ at \bar{x} . The statement now follows from Theorem 3.3.9 in view of Fact 5.1.2. \square

REMARK 5.1.6. Since $f(\delta) \downarrow 0$ as $\delta \downarrow 0$, the rate c estimated in Theorems 5.1.4 and 5.1.5 must be strictly smaller than 1 when the starting point is sufficiently close to \bar{x} . This convergence rate is improving (smaller) when the iterates get closer to \bar{x} (i.e., δ is decreasing).

5.1.2 Forward–backward algorithm and variants of the DR method

We discuss the source location problem with three sensors in the product space

$$\text{find } \bar{u} \in \Lambda \cap \mathcal{S},$$

where Λ is the diagonal and $\mathcal{S} := \prod_{j=1}^3 \mathcal{S}_j$ in $\mathbb{R}^{3 \times 3}$.

The following lemma ensures that $\{\Lambda, \mathcal{S}\}$ is transversal at the solution under Assumption 5.1.1.

LEMMA 5.1.7. *If the three sensors satisfy Assumption 5.1.1, then $\{\Lambda, \mathcal{S}\}$ is transversal at the solution.*

Proof. The statement follows from the linear independence of three nonzero normal vectors of the spheres at the solution. \square

In turn, the transversality of $\{\Lambda, \mathcal{S}\}$ implies the metric subregularity condition imposed in Corollary 3.1.4 for these algorithms, and as a consequence these algorithms are locally linearly convergent.

5.1.3 ADMM algorithm

Let us consider the source location problem with noise, that is to find an appropriate approximation of the true source \bar{x} that is described by the following system of equations

$$r_j = \|\bar{x} - a_j\| + \varepsilon_j \quad (j = 1, 2, \dots, m), \quad (5.4)$$

where ε_j is the j -th unknown noise. The parameters $a = (a_1, a_2, \dots, a_m) \in \mathbb{R}^{nm}$ and $r = (r_1, r_2, \dots, r_m) \in \mathbb{R}_+^m$ are the receiver locations and distances of the receivers to the unknown sender \bar{x} , respectively.

One strategy proposed in [99] to address (5.4) is to find a solution to the following minimization problem

$$\min_{(x,u) \in \mathbb{R}^n \times \mathbb{R}^{nm}} f(x,u) := \sum_{j=1}^m \left(\frac{1}{2} \|x - a_j\|^2 - r_j \langle u_j, x - a_j \rangle + \iota_{\mathbb{B}}(u_j) \right). \quad (5.5)$$

Let us denote

$$\begin{aligned} \mathcal{A} &:= \{x \in \mathbb{R}^{nm} \mid \exists z \in \mathbb{R}^n \text{ such that } x + a = (z, z, \dots, z)\}, \\ \mathcal{B} &:= (r_1 \mathbb{B}) \times \dots \times (r_m \mathbb{B}), \\ \mathbb{E} &:= (\mathbb{R}^n)^m \text{ the product space endowed with the 2-norm.} \end{aligned}$$

The problem (5.5) then takes the form of (3.45), and hence Algorithm 3.6.1 applied to this problem converges globally thanks to Theorem 3.6.2.

5.1.4 Numerical simulation

We do simulation for the source location problem in \mathbb{R}^3 with $m = 20$ sensors.

- Generate randomly the sensor locations a_j , $j = 1, 2, \dots, m$, and the true source location \bar{x} , from a uniform distribution over the box $[-10000, 10000]^3$.
- Compute the ranges r_j , $j = 1, 2, \dots, m$, using the relation

$$r_j = \|\bar{x} - a_j\| + \varepsilon_j,$$

where ε_j are noise.

- Generate random starting point again from uniform distribution over the box for all methods.

The stopping criterion $\|x - x^+\| < 10^{-10}$ is used. We run all the above algorithms in Matlab and observe their convergence behaviors which appear to be consistent with the convergence theory discussed above. The parameter is chosen with seemingly best performance for each method: $\lambda = .15$ for FB, $\beta = .8$ for RAAR, $\lambda = .15$ for DRAP, and $\rho = 1.15$ for ADMM.

The *change* of the distance between two consecutive iterates is of interest. When linear convergence appears to be the case, this observable information may provide the rate of convergence. Under the assumption that the iteration will remain in the convergent area, one can obtain practically useful error bound for the distance from the iterate to a solution.

We also pay attention to the iterate *gap* that in a sense measures the infeasibility at the iterates. If we think feasibility as the problem of minimizing the function that is the sum of (the squares of) the distance functions to the sets, then iterate gaps are simply the values of

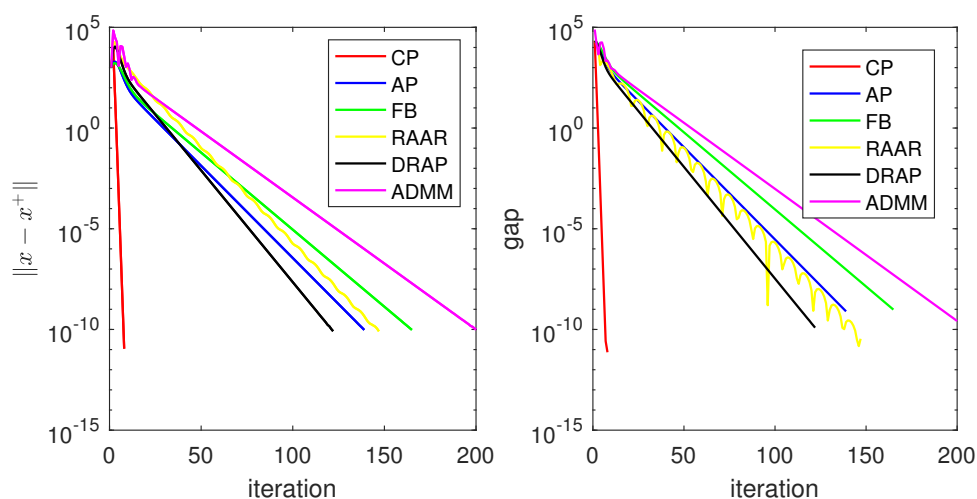


Figure 5.1: Source location problem without noise: the change in iterates (left) and the gap in iterates (right).

that function evaluated at the iterates. For the RAAR and DRAP algorithms, the iterates are themselves not informative but their shadows, by which we mean the projections of the iterates on one of the sets. Hence, the iterate gap corresponding to these methods is calculated for the shadow iterates instead of the iterates themselves.

Figures 5.1 and 5.2 present the changes and the gaps of the algorithms for solving the source location problem both without noise and with noise, respectively. In the simulation with noise, the noise ε_j ($j = 1, 2, \dots, m$) are generated from a normal distribution with zero mean and standard deviation 20.

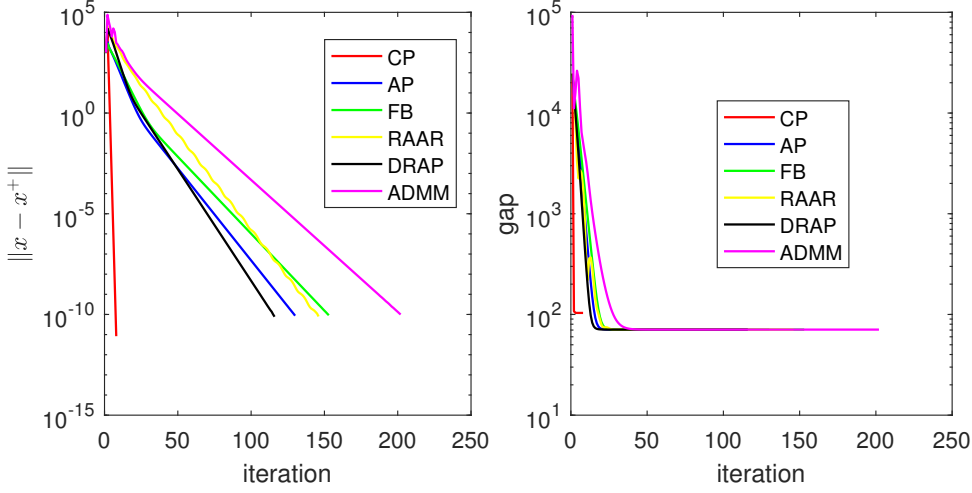


Figure 5.2: Source location problem with noise: the change in iterates (left) and the gap in iterates (right).

5.2 Phase retrieval problem

The phase retrieval problem of recovering a complex signal from several measured intensity patterns has great interests in optics and imaging science [53, 95, 98, 131].

For an unknown object $\bar{x} \in \mathbb{C}^n$, let $A \in \mathbb{C}^{N \times n}$ be the propagation matrix (A can be normalized to be isometric) and $b \in \mathbb{R}^N$ be the data vector of $|A\bar{x}|^2$ (elementwise squared amplitude). We consider the phase retrieval problem of finding a solution to the equation

$$|Ax|^2 = b. \quad (5.6)$$

We first consider

$$A = \frac{1}{\sqrt{m}}(F_1^T \ F_2^T \ \cdots \ F_m^T)^T,$$

where $F_j \in \mathbb{C}^{n \times n}$ are unitary matrices and $b = (b_1^T \ b_2^T \ \cdots \ b_m^T)^T$. For each $j = 1, 2, \dots, m$, let us define

$$\mathcal{S}_j := \{x \in \mathbb{C}^n : |F_j x|^2 = b_j\}. \quad (5.7)$$

Then the phase retrieval problem (5.6) can be reformulated as a nonconvex feasibility:

$$\text{find } \bar{x} \in \bigcap_{j=0}^m \mathcal{S}_j, \quad (5.8)$$

where \mathcal{S}_0 describes a priori information.

5.2.1 Cyclic and averaged projections

Geometrically, each of the sets \mathcal{S}_j defined by (5.7) is the cartesian product of a number of circles in the complex plane and hence it is prox-regular and semi-algebraic. It is worth mentioning that the analysis regarding prox-regularity in the context of the phase retrieval problem was first given in [96].

The sets \mathcal{S}_j on the one hand can be identified as prox-regular and semi-algebraic sets in \mathbb{R}^{2n} . On the other hand, for any $m \geq 2$, it is almost surely that $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m\}$ is subtransversal at points of intersection. In fact, the subtransversality property holds pairwise almost surely, that is for all subcollections of two sets. Hence, these algorithms converge locally linearly thanks to Chapter 3. In view of Example 3.2.11, the subtransversality in the extended sense (Definition 2.3.12 (i)) holds true almost surely, and as a result, these convergence results are also valid for the inconsistent case of (5.8) which corresponds to the presence of noise or model mismatch of phase retrieval.

5.2.2 Forward–backward algorithm and variants of the DR method

Reformulate problem (5.8) in the product space

$$\text{find } \bar{u} \in \Lambda \cap \mathcal{S},$$

where Λ is the diagonal and $\mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_m$ in $(\mathbb{C}^n)^m$.

Again the sets \mathcal{S} and Λ can be identified as prox-regular and semi-algebraic sets in \mathbb{R}^{2mn} . Under the assumption of metric subregularity of $T - \text{Id}$ as in Corollary 3.1.4, these algorithms are known to be locally linearly convergent by Chapter 3. In the special case of $m = 2$, the transversality of $\{\Lambda, \mathcal{S}\}$ at the solution implies the metric subregularity of $T - \text{Id}$, and as a result this assumption is satisfied almost surely. However, for $m > 2$, the transversality property of $\{\Lambda, \mathcal{S}\}$ becomes infeasible. This phenomenon will be investigated in a future research.

In (5.6), let us consider $A \in \mathbb{C}^{N \times n}$ ($N \geq 2n$) an isometric propagation matrix. Then we can consider the phase retrieval problem in the Fourier domain as follows:

$$\text{find } \bar{y} \in \mathcal{A} \cap \mathcal{B}, \tag{5.9}$$

where $\mathcal{A} := A(\mathbb{C}^n)$ and \mathcal{B} is the set of points satisfying the Fourier domain constraint, i.e.,

$$\mathcal{B} = \{y \in \mathbb{C}^N \mid |y|^2 = b\}.$$

For (5.9), let us assume that the mask functions, which together with the Fourier/Fresnel transform compose the propagation matrix A , are continuous random variables. Then $\bar{y} = A\bar{x}$ almost surely vanishes nowhere, i.e. condition (3.41) in Theorem 3.5.4 is satisfied. The number λ_2 defined by (3.43) in that theorem is indeed the second largest singular value of the matrix $(\text{Re}(B) - \text{Im}(B)) \in \mathbb{R}^{N \times 2n}$ (the largest one is 1), where B is defined

by (3.42). Condition (3.43) essentially requires the spectral gap of this real matrix. In the current setting, condition (3.43) is satisfied by, for example, [39, Proposition 6.1]. As a result, Theorem 3.5.4 yields local linear convergence of the *RAAR* algorithm (3.40).

5.2.3 ADMM algorithm

We consider the minimization problem for solving (5.6) as follows:

$$\min_{x \in \mathbb{C}^n} f(x) := \sum_{j=1}^m \| |F_j x| - b_j \|^2. \quad (5.10)$$

Decompose the variable $x = \operatorname{Re}(x) + i\operatorname{Im}(x)$, where $\operatorname{Re}(x), \operatorname{Im}(x) \in \mathbb{R}^n$. Let $z = p(x)$, where $p : \mathbb{C}^n \rightarrow \mathbb{R}^{2n}$ is defined by

$$\operatorname{Re}(x) + i\operatorname{Im}(x) = x \mapsto p(x) := (\operatorname{Re}(x)^T, \operatorname{Im}(x)^T)^T.$$

Decompose uniquely also $F_j = \operatorname{Re}(F_j) + i\operatorname{Im}(F_j)$, ($j = 1, \dots, m$), where $\operatorname{Re}(F_j)$ and $\operatorname{Im}(F_j)$ are real matrices in $\mathbb{R}^{n \times n}$. Define the linear operators $L_j : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ by

$$L_j = \begin{pmatrix} \operatorname{Re}(F_j) & -\operatorname{Im}(F_j) \\ \operatorname{Im}(F_j) & \operatorname{Re}(F_j) \end{pmatrix} = p \circ F_j \circ p^{-1}.$$

Note that L_j ($j = 1, \dots, m$) are isomorphic linear mappings since both F_j and p are so.

Denote also $L_j^i : \mathbb{R}^{2n} \rightarrow \mathbb{R}^2$ the linear mappings consisting of the i and $i+n$ rows of L_j , ($i = 1, \dots, n, j = 1, \dots, m$). Then

$$|(F_j x)(i)| = \|L_j^i z\|.$$

Hence, the problem (5.10) is equivalent to

$$\min_{z \in \mathbb{R}^{2n}} \sum_{j=1}^m \sum_{i=1}^n \left(\frac{1}{2} \|L_j^i z\|^2 - b_j^i \|L_j^i z\| \right). \quad (5.11)$$

Note that

$$\|L_j^i z\| = \max_{u_j^i \in \mathbb{B}} \langle u_j^i, L_j^i z \rangle \quad (i = 1, \dots, n, j = 1, \dots, m).$$

The problem (5.11) is equivalent to

$$\min_{z \in \mathbb{R}^{2n}} \left\{ \sum_{j=1}^m \sum_{i=1}^n \left(\frac{1}{2} \|L_j^i z\|^2 - \langle u_j^i, L_j^i z \rangle \right), u_j^i \in b_j^i \mathbb{B} \right\}. \quad (5.12)$$

Define the two closed and convex subsets in $\mathbb{E} = \mathbb{R}^{2mn}$ as follows:

$$\begin{aligned}\mathcal{A} &= \{(z_j^i) \in (\mathbb{R}^2)^{mn} \mid \exists z \in \mathbb{R}^{2n} : z_j^i = L_j^i z\}, \\ \mathcal{B} &= \prod_{1 \leq j \leq m, 1 \leq i \leq n} b_j^i \mathbb{B}.\end{aligned}$$

Then (5.12) is equivalently rewritten as

$$\min_{x, u \in \mathbb{E}} \left\{ \frac{1}{2} \|x\|^2 - \langle u, x \rangle : x \in \mathcal{A}, u \in \mathcal{B} \right\}$$

which turns out to be problem (3.45) with $a = 0$.

Hence, Algorithm 3.6.1 is known to converge globally thanks to Theorem 3.6.2.

5.2.4 Numerical simulation

Consider a complex object $\bar{x} \in \mathbb{C}^{128 \times 128}$ with the support constraint χ . We can scale and normalize the data such that

$$\bar{x} = \iota_\chi \odot \exp(2\pi\theta) / \|\iota_\chi \odot \exp(2\pi\theta)\|.$$

Let us consider the phase retrieval problem with four images generated via the corresponding unitary transforms

$$F_1 = F \circ \exp(-2\pi\theta), \quad F_2 = F, \quad F_3 = F \circ \exp(2\pi\Theta), \quad F_4 = F \circ \exp(-2\pi\Theta),$$

where F is the Fourier transform (normalized to be unitary) and $\Theta \in (-1, 1]^{128 \times 128}$ is a given defocus.

The stopping criteria $\|x - x^+\| < 10^{-15}$ is used. The parameter is chosen with seemingly best performance for each method: $\lambda = .45$ for FB, $\beta = .8$ for RAAR, $\lambda = .35$ for DRAP, and $\rho = 1.25$ for ADMM.

Due to the ambiguity up to a total piston term of phase retrieval, the iterative gap is measured up the optimal total phase shift:

$$\|x - \bar{x}\|_{Opt} = \|(x^* \bar{x})x / |x^* \bar{x}| - \bar{x}\|.$$

In the experiment with noise a white Gaussian noise at $30dB$ was added to the intensity measurement and the negative entries of the obtained images were then reset to zeros. Figures 5.3 and 5.4 present the iterative change and gap of the algorithms for solving this phase retrieval problem without noise and with noise, respectively. The reconstruction phase up to an optimal total phase shift for the problem with noise is presented in Figure 5.5.

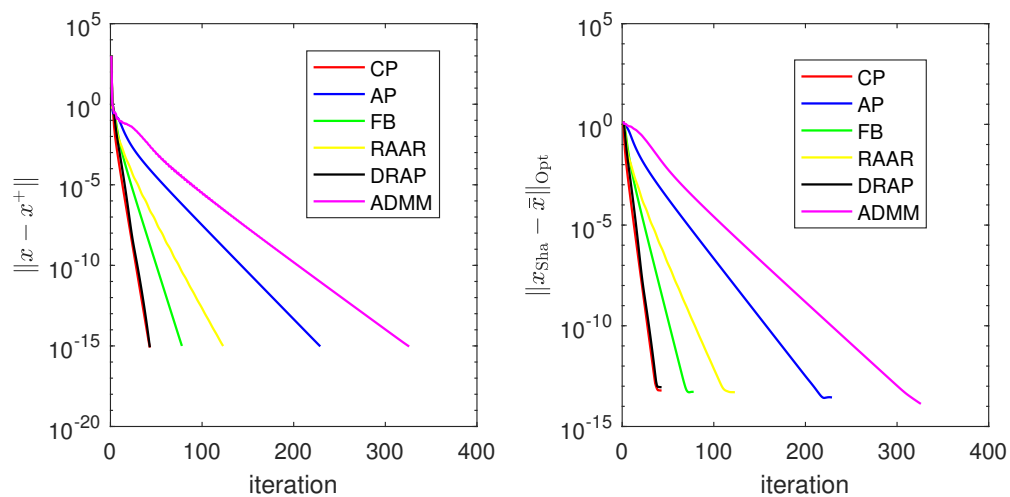


Figure 5.3: Phase retrieval JWST experiment without noise: the change in iterates (left) and the gap in iterates (right).

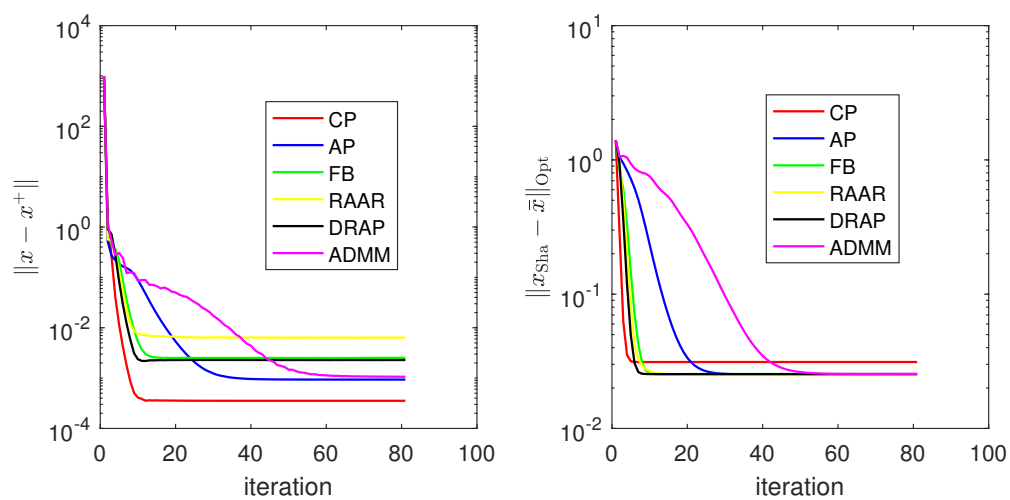


Figure 5.4: Phase retrieval JWST experiment with noise: the change in iterates (left) and the gap in iterates (right).

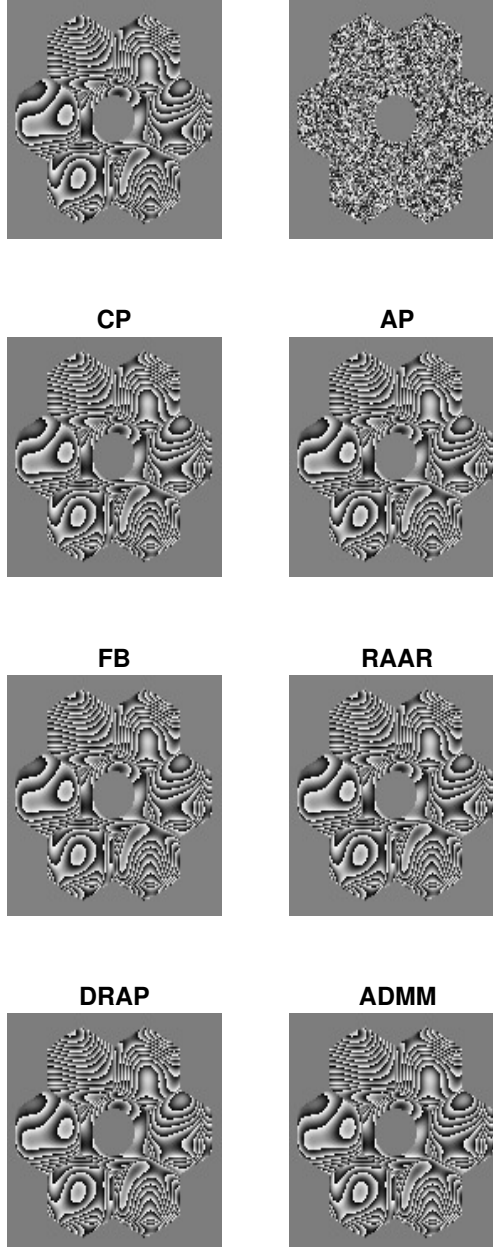


Figure 5.5: Phase retrieval JWST experiment with noise: reconstruction up to a total piston term.

Chapter 6

Conclusion

A case study on algorithms for structured nonconvex optimization has been conducted in the thesis. Its contribution to the field of convergence analysis is twofold: 1) regularity theory essential for convergence analysis, and 2) convergence criteria of numerical methods with application.

We synthesize and unify notions of regularity, especially those of individual sets and of collections of sets, as they appear in the convergence theory of projection methods for feasibility problems. Several new characterizations of regularity notions are presented. A number of new relationships amongst regularity properties are established. Based on the knowledge of regularity notions, we develop a framework for quantitative convergence analysis of fixed point iterations with a number of subsequent results showing convergence of fundamental optimization algorithms. Several new convergence criteria for projection methods are presented. New understanding on regularity theory also paves the way to a development on necessary conditions for local linear convergence of fundamental algorithms. Metric subregularity is shown to be necessary for linear monotonicity of Picard iterations. An intensive discussion on subtransversality as necessary conditions for linear convergence of alternating projections is presented. In particular, subtransversality is shown to be not only sufficient but also necessary for linear convergence of convex consistent alternating projections. We apply and illustrate the theory to the source location and phase retrieval problems.

In summary, the thesis contributes new insight into the bilateral research topics that, on the one hand, understanding of regularity properties of the problem data allows one to establish convergence criteria for optimization algorithms, and on the other hand, analyzing convergence of numerical methods often leads to a search for more subtle characterizations of the input data, and hence provides a fruitful platform for investigating regularity notions.

Bibliography

- [1] M. Apetrii, M. Durea, and R. Strugariu. On subregularity properties of set-valued mappings. *Set-Valued Var. Anal.*, 21(1):93–126, 2013.
- [2] F. J. Aragón Artacho, A. L. Dontchev, and M. H. Geoffroy. Convergence of the proximal point method for metrically regular mappings. *ESAIM: Proc.*, 17:1–8, 2007.
- [3] F. J. Aragón Artacho and M. H. Geoffroy. Uniformity and inexact version of a proximal method for metrically regular mappings. *J. Math. Anal. Appl.*, 335(1):168–183, 2007.
- [4] T. Aspelmeier, C. Charitha, and D. R. Luke. Local linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging. *SIAM J. Imaging Sci.*, 9(2):842–868, 2016.
- [5] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Lojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.
- [6] H. Attouch and J. Peypouquet. The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.*, 26(3):1824–1834, 2016.
- [7] J.-P. Aubin. *Contingent Derivatives of Set-valued Maps and Existence of Solutions to Nonlinear Inclusions and Differential Inclusions*. Cahiers du CEREMADE. Mathematics Research Center, University of Wisconsin, 1980.
- [8] J.-P. Aubin and H. Frankowska. *Set-valued analysis*. Birkhäuser, Boston, 1990.
- [9] D. Azé. A survey on error bounds for lower semicontinuous functions. In *Proceedings of 2003 MODE-SMAI Conference*, volume 13 of *ESAIM Proc.*, pages 1–17. EDP Sci., Les Ulis, 2003.
- [10] D. Azé. A unified theory for metric regularity of multifunctions. *J. Convex Anal.*, 13:225–252, 2006.

- [11] A. Bakan, F. Deutsch, and W. Li. Strong CHIP, normality, and linear regularity of convex sets. *Trans. Amer. Math. Soc.*, 357(10):3831–3863, 2005.
- [12] H. H. Bauschke and J. M. Borwein. On the convergence of von Neumann’s alternating projection algorithm for two sets. *Set-Valued Anal.*, 1(2):185–212, 1993.
- [13] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.
- [14] H. H. Bauschke, J. M. Borwein, and A. S. Lewis. The method of cyclic projections for closed convex sets in Hilbert space. In *Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995)*, pages 1–38. Amer. Math. Soc., Providence, RI, 1997.
- [15] H. H. Bauschke, J. M. Borwein, and W. Li. Strong conical hull intersection property, bounded linear regularity, Jameson’s property (G), and error bounds in convex optimization. *Math. Program., Ser. A*, 86(1):135–160, 1999.
- [16] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books Math./Ouvrages Math. SMC. Springer, New York, 2011.
- [17] 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.
- [18] H. H. Bauschke, F. Deutsch, and H. Hundal. Characterizing arbitrarily slow convergence in the method of alternating projections. *Int. Trans. Oper. Res.*, 16(4):413–425, 2009.
- [19] H. H. Bauschke, F. Deutsch, H. Hundal, and S.-H. Park. Accelerating the convergence of the method of alternating projections. *Trans. Amer. Math. Soc.*, 355(9):3433–3461, 2003.
- [20] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections: Applications. *Set-Valued Var. Anal.*, 21:475–501, 2013.
- [21] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections: Theory. *Set-Valued Var. Anal.*, 21:431–473, 2013.
- [22] H. H. Bauschke and W. M. Moursi. The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces. *SIAM J. Optim.*, 26(2):968–985, 2016.

- [23] H. H. Bauschke, D. Noll, and H. M. Phan. Linear and strong convergence of algorithms involving averaged nonexpansive operators. *J. Math. Anal. Appl.*, (421):1–20, 2015.
- [24] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [25] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Trans. Amer. Math. Soc.*, 362(6):3319–3363, 2010.
- [26] J. Bolte, Nguyen T. Phong, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program., Ser. A*, 165(2):471–507, 2017.
- [27] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program., Ser. A*, 146(1-2):459–494, 2014.
- [28] J. M. Borwein, G. Li, and M. K. Tam. Convergence rate analysis for averaged fixed point iterations in common fixed point problems. *SIAM J. Optim.*, 27(1):1–33, 2017.
- [29] J. M. Borwein, G. Li, and L. Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM J. Optim.*, 24(1):498–527, 2014.
- [30] J. M. Borwein and Q. J. Zhu. *Techniques of Variational Analysis*. Springer, New York, 2005.
- [31] J. M. Borwein and D. M. Zhuang. Verifiable necessary and sufficient conditions for openness and regularity of set-valued and single-valued maps. *J. Math. Anal. Appl.*, 134(2):441–459, 1988.
- [32] L. M. Bregman. The method of successive projection for finding a common point of convex sets. *Soviet Mathematics - Doklady*, 6:688–692, 1965.
- [33] R. E. Bruck and S. Reich. Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston J. Math.*, 3(4):459–470, 1977.
- [34] L. N. H. Bunt. *Bitdrage tot de theorie der konvekse puntverzamelingen*. PhD thesis, Univ. of Groningen, Amsterdam, 1934.
- [35] J. V. Burke and S. Deng. Weak sharp minima revisited. I. Basic theory. *Control Cybernet.*, 31(3):439–469, 2002.
- [36] J. V. Burke and S. Deng. Weak sharp minima revisited. II. Application to linear regularity and error bounds. *Math. Program., Ser. B*, 104(2-3):235–261, 2005.

- [37] J. V. Burke and M. C. Ferris. Weak sharp minima in mathematical programming. *SIAM J. Control Optim.*, 31:1340–1359, 1993.
- [38] A. Chambolle and C. Dossal. On the convergence of the iterates of the “fast iterative shrinkage/thresholding algorithm”. *J. Optim. Theory Appl.*, 166(3):968–982, 2015.
- [39] P. Chen and A. Fannjiang. Fourier phase retrieval with a single mask by Douglas–Rachford algorithms. *Appl. Comput. Harmon. Anal.*, 44(3):665–699, 2018.
- [40] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*, volume 178 of *Graduate Texts in Mathematics*. Springer, New York, 1998.
- [41] P. L. Combettes and T. Pennanen. Proximal methods for cohypomonotone operators. *SIAM J. Control Optim.*, 43(2):731–742, 2004.
- [42] A. Daniilidis and P. Georgiev. Approximate convexity and submonotonicity. *J. Math. Anal. Appl.*, 291(1):292–301, 2004.
- [43] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57:1413–1457, 2004.
- [44] E. De Giorgi, A. Marino, and M. Tosques. Evolution problems in in metric spaces and steepest descent curves. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.
- [45] F. Deutsch, W. Li, and J. D. Ward. A dual approach to constrained interpolation from a convex subset of Hilbert space. *J. Approx. Theory*, 90(3):385–414, 1997.
- [46] A. V. Dmitruk, A. A. Milyutin, and N. P. Osmolovsky. Lyusternik’s theorem and the theory of extrema. *Russian Math. Surveys*, 35:11–51, 1980.
- [47] A. L. Dontchev. The Graves theorem revisited. *J. Convex Anal.*, 3(1):45–53, 1996.
- [48] A. L. Dontchev, A. S. Lewis, and R. T. Rockafellar. The radius of metric regularity. *Trans. Amer. Math. Soc.*, 355:493–517, 2003.
- [49] A. L. Dontchev and R. T. Rockafellar. Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Anal.*, 12(1-2):79–109, 2004.
- [50] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*. Srpinge-Verlag, New York, second edition, 2014.
- [51] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis. Transversality and alternating projections for nonconvex sets. *Found. Comput. Math.*, 15(6):1637–1651, 2015.

- [52] M. J. Fabian, R. Henrion, A. Y. Kruger, and J. V. Outrata. Error bounds: necessary and sufficient conditions. *Set-Valued Var. Anal.*, 18(2):121–149, 2010.
- [53] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21:2758–2769, 1982.
- [54] K. Friedrichs. On certain inequalities and characteristic value problems for analytic functions and for functions of two variables. *Trans. Amer. Math. Soc.*, 41:321–364, 1937.
- [55] A. Genel and J. Lindenstrauss. An example concerning fixed points. *Israel Journal of Mathematics*, 22:81–86, 1975.
- [56] L. Gubin, B. Polyak, and E. Raik. The method of projections for finding the common point of convex sets. *USSR Comput. Math and Math Phys.*, 7(6):1–24, 1967.
- [57] V. Guillemin and A. Pollack. *Differential Topology*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974.
- [58] R. Henrion and J. V. Outrata. Calmness of constraint systems with applications. *Math. Program., Ser. B*, 104(2-3):437–464, 2005.
- [59] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM J. Optim.*, 23(4):2397–2419, 2013.
- [60] R. Hesse, D. R. Luke, and P. Neumann. Alternating projections and Douglas–Rachford for sparse affine feasibility. *IEEE Trans. Signal Process.*, 62(18):4868–4881, 2014.
- [61] M. Hirsch. *Differential Topology*. Springer Verlag, New York, 1976.
- [62] H. Hundal. An alternating projection that does not converge in norm. *Nonlinear Anal.*, 57:35–61, 2004.
- [63] A. D. Ioffe. Nonsmooth analysis: differential calculus of nondifferentiable mappings. *Trans. Amer. Math. Soc.*, 266:1–56, 1981.
- [64] A. D. Ioffe. Approximate subdifferentials and applications: III. *Mathematika*, 36(71):1–38, 1989.
- [65] A. D. Ioffe. Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):103–162, 2000.
- [66] A. D. Ioffe. Regularity on a fixed set. *SIAM J. Optim.*, 21(4):1345–1370, 2011.
- [67] A. D. Ioffe. Nonlinear regularity models. *Math. Program., Ser. B*, 139(1-2):223–242, 2013.

- [68] A. D. Ioffe. Metric regularity – a survey. Part I. Theory. *J. Aust. Math. Soc.*, 101(2):188–243, 2016.
- [69] A. D. Ioffe and J. V. Outrata. On metric and calmness qualification conditions in subdifferential calculus. *Set-Valued Anal.*, 16(2-3):199–227, 2008.
- [70] 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.
- [71] G. J. O. Jameson. The duality of pairs of wedges. *Proc. London Math. Soc.*, 24:531–547, 1972.
- [72] S. Kayalar and H. Weinert. Error bounds for the method of alternating projections. *Math. Control Signals Syst.*, 1:43–59, 1988.
- [73] Phan Q. Khanh, A. Y. Kruger, and Nguyen H. Thao. An induction theorem and nonlinear regularity models. *SIAM J. Optim.*, 25(4):2561–2588, 2015.
- [74] D. Klatte and B. Kummer. Optimization methods and stability of inclusions in Banach spaces. *Math. Program., Ser. B*, 117(1-2):305–330, 2009.
- [75] D. Klatte and W. Li. Asymptotic constraint qualifications and global error bounds for convex inequalities. *Math. Program., Ser. A*, 84(1):137–160, 1999.
- [76] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Quantitative asymptotic regularity results for the composition of two mappings. *Optimization*, 66(8):1291–1299, 2017.
- [77] A. Y. Kruger. A covering theorem for set-valued mappings. *Optimization*, 19(6):763–780, 1988.
- [78] A. Y. Kruger. Stationarity and regularity of set systems. *Pac. J. Optim.*, 1(1):101–126, 2005.
- [79] A. Y. Kruger. About regularity of collections of sets. *Set-Valued Anal.*, 14:187–206, 2006.
- [80] A. Y. Kruger. About stationarity and regularity in variational analysis. *Taiwanese J. Math.*, 13(6A):1737–1785, 2009.
- [81] A. Y. Kruger. Error bounds and metric subregularity. *Optimization*, 64(1):49–79, 2015.
- [82] A. Y. Kruger. About intrinsic transversality of pairs of sets. *Set-Valued Var. Anal.*, 26(1):111–142, 2018.
- [83] A. Y. Kruger, D. R. Luke, and Nguyen H. Thao. *Set-Valued Var. Anal.*, 25(4):701–729, 2017.

- [84] A. Y. Kruger, D. R. Luke, and Nguyen H. Thao. Set regularities and feasibility problems. *Math. Program., Ser. B*, 168(1):1–33, 2018.
- [85] A. Y. Kruger and Nguyen H. Thao. About uniform regularity of collections of sets. *Serdica Math. J.*, 39:287–312, 2013.
- [86] A. Y. Kruger and Nguyen H. Thao. About $[q]$ -regularity properties of collections of sets. *J. Math. Anal. Appl.*, 416(2):471–496, 2014.
- [87] A. Y. Kruger and Nguyen H. Thao. Quantitative characterizations of regularity properties of collections of sets. *J. Optim. Theory and Appl.*, 164:41–67, 2015.
- [88] A. Y. Kruger and Nguyen H. Thao. Regularity of collections of sets and convergence of inexact alternating projections. *J. Convex Anal.*, 23(3):823–847, 2016.
- [89] D. Leventhal. Metric subregularity and the proximal point method. *J. Math. Anal. Appl.*, 360(2):681–688, 2009.
- [90] 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.
- [91] A. S. Lewis and J. Malick. Alternating projections on manifolds. *Math. Oper. Res.*, 33:216–234, 2008.
- [92] C. Li, K. F. Ng, and T. K. Pong. The SECQ, linear regularity, and the strong CHIP for an infinite system of closed convex sets in normed linear spaces. *SIAM J. Optim.*, 18(2):643–665, 2007.
- [93] J. Li and T. Zhou. On relaxed averaged alternating reflections (RAAR) algorithm for phase retrieval with structured illumination. *Inverse Problems*, 33(2):025012 (20pp), 2017.
- [94] W. Li. Abadie’s constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM J. Optim.*, 7(4):966–978, 1997.
- [95] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21:37–50, 2005.
- [96] D. R. Luke. Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space. *SIAM J. Optim.*, 19(2):714–739, 2008.
- [97] D. R. Luke. Lecture Notes in Numerical Variational Analysis. *Institute for Numerical and Applied Mathematics, Univ. of Gottingen*, 2017.
- [98] D. R. Luke, J. V. Burke, and R. G. Lyon. Optical wavefront reconstruction: theory and numerical methods. *SIAM Rev.*, 44(2):169–224, 2002.

- [99] D. R. Luke, S. Sabach, M. Teboulle, and K. Zatlaway. A simple globally convergent algorithm for the nonsmooth nonconvex single source localization problem. *J. Global Optim.*, 69(4):889–909, 2017.
- [100] D. R. Luke and R. Shefi. A globally linearly convergent method for pointwise quadratically supportable convex-concave saddle point problems. *J. Math. Anal. Appl.*, 457(2):1568–1590, 2018.
- [101] D. R. Luke, M. Teboulle, and Nguyen H. Thao. Necessary conditions for linear convergence of iterated expansive, set-valued mappings with application to alternating projections. Submitted January 2017.
- [102] D. R. Luke, Nguyen H. Thao, and M. K. Tam. Implicit error bounds for Picard iterations on Hilbert spaces. *Vietnam J. Math.*, 46(2):243–258, 2018.
- [103] D. R. Luke, Nguyen H. Thao, and M. K. Tam. Quantitative convergence analysis of iterated expansive, set-valued mappings. *Math. Oper. Res.*, to appear.
- [104] Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Ann. Oper. Res.*, 46/47(1-4):157–178, 1993.
- [105] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29(3):341–346, 1962.
- [106] B. S. Mordukhovich. *Approximation Methods in Problems of Optimization and Control*. Nauka, Moscow, 1988.
- [107] B. S. Mordukhovich. Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions. *Trans. Amer. Math. Soc.*, 340(1):1–35, 1993.
- [108] B. S. Mordukhovich. Coderivatives of set-valued mappings: calculus and applications. In *Proceedings of the Second World Congress of Nonlinear Analysts, Part 5 (Athens, 1996)*, volume 30, pages 3059–3070, 1997.
- [109] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation. I: Basic Theory*, volume 330 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Berlin, 2006.
- [110] B. S. Mordukhovich and Y. Shao. Extremal characterizations of Asplund spaces. *Proc. Amer. Math. Soc.*, 124(1):197–205, 1996.
- [111] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *Comptes Rendus de l'Académie des Sciences de Paris*, 255:2897–2899, 1962.

- [112] Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, CORE Discussion Papers, 2007.
- [113] K. F. Ng and W. H. Yang. Regularities and their relations to error bounds. *Math. Program., Ser. A*, 99(3):521–538, 2004.
- [114] K. F. Ng and R. Zang. Linear regularity and ϕ -regularity of nonconvex sets. *J. Math. Anal. Appl.*, 328(1):257–280, 2007.
- [115] Huynh V. Ngai, Dinh T. Luc, and M. Théra. Approximate convex functions. *J. Nonlinear Convex Anal.*, 1:155–176, 2000.
- [116] Huynh V. Ngai, Dinh T. Luc, and M. Théra. Extensions of Fréchet ϵ -subdifferential calculus and applications. *J. Math. Anal. Appl.*, 268(1):266–290, 2002.
- [117] Huynh V. Ngai and M. Théra. Metric inequality, subdifferential calculus and applications. *Set-Valued Analysis*, 9:187–216, 2001.
- [118] D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. *Found. Comput. Math.*, 16(2):425–455, 2016.
- [119] R. D. Nussbaum. Degree theory for local condensing maps. *J. Math. Anal. and Appl.*, 37:741–766, 1972.
- [120] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
- [121] C. H. Jeffrey Pang. First order constrained optimization algorithms with feasibility updates. *arXiv:1506.08247v1*, 2015.
- [122] T. Pennanen. Local convergence of the proximal point algorithm and multiplier methods without monotonicity. *Math. Oper. Res.*, 27:170–191, 2002.
- [123] J.-P. Penot. Metric regularity, openness and Lipschitzian behavior of multifunctions. *Nonlinear Anal.*, 13(6):629–643, 1989.
- [124] J.-P. Penot. *Calculus Without Derivatives*. Springer, New York, 2013.
- [125] H. M. Phan. Linear convergence of the Douglas–Rachford method for two closed sets. *Optimization*, 65:369–385, 2016.
- [126] G. Pierra. Decomposition through formalization in a product space. *Math. Programming*, 28(1):96–115, 1984.
- [127] R. A. Poliquin, R. T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

- [128] S. Reich. Fixed points of condensing functions. *J. Math. Anal. Appl.*, 41:460–467, 1973.
- [129] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Grundlehren Math. Wiss. Springer-Verlag, Berlin, 1998.
- [130] B. D. Rouhani. Asymptotic behaviour of almost nonexpansive sequences in a Hilbert space. *J. Math. Anal. Appl.*, 151(1):226–235, 1990.
- [131] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.
- [132] J. E. Spingarn. Submonotone subdifferentials of Lipschitz functions. *Trans. Amer. Math. Soc.*, 264:77–89, 1981.
- [133] Nguyen H. Thao. A convergent relaxation of the Douglas–Rachford algorithm. *Comput. Optim. Appl.*, 70(3):841–863, 2018.
- [134] J. von Neumann. *Functional Operators, Vol II. The geometry of orthogonal spaces*, volume 22 of *Ann. Math Stud.* Princeton University Press, 1950.
- [135] X. Y. Zheng and K. F. Ng. Linear regularity for a collection of subsmooth sets in Banach spaces. *SIAM J. Optim.*, 19(1):62–76, 2008.
- [136] X. Y. Zheng and K. F. Ng. Metric subregularity and calmness for nonconvex generalized equations in Banach spaces. *SIAM J. Optim.*, 20(5):2119–2136, 2010.
- [137] X. Y. Zheng and K. F. Ng. Metric subregularity for nonclosed convex multifunctions in normed spaces. *ESAIM Control Optim. Calc. Var.*, 16(3):601–617, 2010.
- [138] X. Y. Zheng and K. F. Ng. Metric subregularity for proximal generalized equations in Hilbert spaces. *Nonlinear Anal.*, 75(3):1686–1699, 2012.
- [139] X. Y. Zheng, Z. Wei, and J.-C. Yao. Uniform subsmoothness and linear regularity for a collection of infinitely many closed sets. *Nonlinear Anal.*, 73(2):413–430, 2010.