Chapter 23

Convergence Analysis of Iterative Algorithms for Phase Retrieval

D. Russell Luke and Anna-Lena Martins

Abstract This chapter surveys the analysis of the phase retrieval problem as an inconsistent and nonconvex feasibility problem. We apply a convergence framework for iterative mappings developed by Luke, Tam and Thao in 2018 to the inconsistent and nonconvex phase retrieval problem and establish the convergence properties (with rates) of popular projection methods for this problem. Although our main purpose is to illustrate the convergence results and their underlying concepts, we demonstrate how our theoretical analysis aligns with practical numerical computation applied to laboratory data.

Mathematics Subject Classification: 65K10 · 49K40 · 49M05 · 65K05 · 90C26 · 49M20 · 49J53

23.1 Introduction

We highlight recent theoretical advances that have opened the door to a quantitative convergence analysis of well-known phase retrieval algorithms. As shown in Chap. 6, phase retrieval problems have a natural and easy characterization as feasibility problems, and issues like noise and model misspecification do not effect the abstract regularity of the problem formulation. This was also observed in studies by Bauschke et al. [1] and Marchesini [2] reviewing phase retrieval algorithms in the context of fixed point iterations, though in those works the theory only provided convex heuristics for understanding the most successful algorithms. A slow progression of the theory for nonconvex feasibility culminating in the work by Luke et al. in [3] now provides a firm theoretical basis for understanding most of the standard algorithms for phase retrieval.

D. R. Luke (✉) · A.-L. Martins
Institute for Numerical and Applied Mathematics, Universität Göttingen,
Lotzestr. 16-18, 37083 Göttingen, Germany
e-mail: r.luke@math.uni-goettingen.de

A.-L. Martins
e-mail: a.martins@math.uni-goettingen.de

The approach is fix-point theoretic and is based on a framework introduced by Luke et al. in [3]. Given some (set-valued) mapping $T : \mathcal{E} \rightrightarrows \mathcal{E}$, where $\mathcal{E}$ is a finite-dimensional Euclidean space, the algorithms are studied as mere generators of sequences $(x^k)_{k \in \mathbb{N}}$ through the fixed point iteration $x^{k+1} \in Tx^k$ ($\forall k \in \mathbb{N}$) with $x^k \rightarrow x^*$ where $x^* = Tx^*$. We demonstrate the convergence framework of [3] on a few of the more prevalent iterative phase retrieval algorithms introduced in Chap. 6.

The analysis is based on two main properties. The first of these is the regularity of the mapping defining the fixed point iteration; the second property concerns the stability of the fixed points of the mapping. The first property is covered by the notion of *pointwise almost averagedness*, a generalization of regularity concepts like (firm) nonexpansiveness. Already in the 1960s Opial [4] showed that an iterative sequence defined by an averaged self-mapping with nonempty fixed point set converges to a fixed point. It is no surprise, then, that generalizations of averagedness should play a central role in convergence for more general fixed point mappings. In the setting of feasibility problems, i.e. finding a point in the intersection of a collection of sets, pointwise almost averagedness of the fixed point mapping is inherited from the regularity of the sets.

The other concept that is central to the analysis concerns stability of the fixed points. This is characterized by the notion of *metric subregularity* as presented in Dontchev and Rockafellar [5], and Ioffe [6, 7]. Metric subregularity of the mapping at fixed points guarantees quantitative estimates for the rate of convergence of the iterates. This is closely related to the existence of error bounds, and weak-sharp minima, among other equivalent notions that provide a path to a quantitative convergence analysis.

In Sect. 23.2 we remind the reader of the phase retrieval problem. Section 23.3 and its subsections introduce basic notations and concepts. This is followed by a toolkit for convergence in Sect. 23.4 that describes the convergence framework we are working with. The use of this theoretical toolkit is demonstrated on two of the most prevalent algorithms for phase retrieval. We conclude this chapter with some numerical remarks in Sect. 23.8.

23.2 Phase Retrieval as a Feasibility Problem

The phase retrieval problem reviewed in Chaps. 2 and 6 involves reconstructing a complex valued field in a plane (the object plane) from measurements of its amplitude under a unitary mapping in a plane somewhere downstream from the object plane (the image plane). We use the notation for the phase retrieval problem already introduced in previous chapters. For a detailed description see Sect. 6.1.1. The measurements are represented by the sets

$$M_j := \left\{ u \in \mathbb{C}^n \mid \|(\mathcal{F}u)_i\| = \sqrt{I_{j,i}}, (i = 1, 2, \dots, n) \right\} \quad (j = 1, 2, \dots, m). \quad (23.1)$$

The problem of recovering the phase from just the modulus of unitary transformed measurements is impossible to solve uniquely. Usually nonuniqueness is associated with *ill-posedness*, but for feasibility problems it is rather *existence* that is the source of difficulty. In real-world problems measurement errors and model misspecification have profound implications for feasibility models, but not for the reasons that one might expect. The geometry of the individual measurement sets does not change in the presence of noise or model misspecification. The issue is that the measurements are not *consistent* with one another. In other words, there is no solution that satisfies the measurements and other model requirements (like nonnegativity, in the case of real objects). A solution from the provided information is then only an approximation to the actual signal. Mathematically these characteristics translate into an *inconsistent* feasibility problem. That is, the intersection of the sets in the feasibility model is empty. Inconsistency has been investigated in many works (see for instance [8–11]) but most of these studies consider convex sets. Unfortunately, the sets involved in the phase retrieval problem are mostly nonconvex and have empty intersection. In [3] the authors provided a scheme to handle even this case. The following sections are devoted to their work and present the most important concepts.

To avoid ambiguities recovering the phase, one often uses a priori information about the model. Common examples are the knowledge of a support of the signal, real-valuedness, non-negativity, sparsity or the information about an amplitude:

support constraint	$\mathfrak{S} := \{y \in \mathbb{C}^n \mid y_i = 0 \ \forall i \notin I\}$
real-valued support constraint	$\mathfrak{S}_r := \{y \in \mathbb{R}^n \mid y_i = 0 \ \forall i \notin I\}$
non-negative support constraint	$\mathfrak{S}_+ := \{y \in \mathbb{R}_+^n \mid y_i = 0 \ \forall i \notin I\}$
amplitude constraint	$A := \{y \in \mathbb{C}^n \mid y_k = a_k, \ 1 \leq k \leq n\}$
sparsity constraint	$\mathcal{A}_s := \{y \in \mathbb{R}^n \mid \ y\ _0 \leq s\}$

(23.2)

for a set of indices $I \subset \{1, 2, \dots, n\}$, $a \in \mathbb{R}_+^n$ and $s \in \{1, 2, \dots, n\}$, where $\mathbb{R}_+^n = \{x \in \mathbb{R}^n \mid x_i \geq 0, \ 1 \leq i \leq n\}$. In the following we focus on the (non-negative) support constraint.

23.3 Notation and Basic Concepts

Our setting throughout this chapter is a finite dimensional real Euclidean space $\mathcal{E}$ equipped with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. The open unit ball is denoted by $\mathbb{B}$, whereas $\mathbb{S}$ stands for the unit sphere in $\mathcal{E}$. The open ball with radius δ and center x is denoted by $\mathbb{B}_\delta(x)$. The iterative algorithms we analyze can be represented by mappings $T : \mathcal{E} \rightrightarrows \mathcal{E}$, where $\rightrightarrows$ indicates that T is a point-to-set mapping. $\mathbb{N}$ denotes the natural numbers. The *inverse mapping* T^{-1} at a point y in the range of T is defined as the set of all points x such that $y \in T(x)$.

23.3.1 Projectors

We follow in this section the definitions introduced in Chap. 6. As a reminder: the distance of a point x to a set $\Omega \subset \mathcal{E}$ is defined by $\text{dist}(x, \Omega) := \inf_{y \in \Omega} \{\|y - x\|\}$. The corresponding *projector* onto the set Ω is given by $\mathcal{P}_\Omega : \mathcal{E} \rightrightarrows \mathcal{E}$, $x \mapsto \{y \in \Omega \mid \text{dist}(x, \Omega) = \|y - x\|\}$. A single element of $\mathcal{P}_\Omega x$ is called a *projection*. Similarly to the projector, the *reflector* onto a set Ω is defined by $\mathcal{R}_\Omega : \mathcal{E} \rightrightarrows \mathcal{E}$, $x \mapsto 2\mathcal{P}_\Omega x - x$, which is again a set. A single element in $\mathcal{R}_\Omega$ is called a *reflection*.

The regularity of a set influences the properties of the corresponding projector onto the set. The best properties are generated by *convex* sets. A convex set Ω is defined as a set that contains the line segment between any two points $x, y \in \Omega$. The projector onto a convex set is not only single-valued, but can be characterized by a variational inequality (see for instance [12, Theorem 3.14]). As we see in Sect. 23.3.2 the algorithms considered here are all composed of projectors and reflectors. This leads to an analysis of the projectors onto the sets introduced in Sect. 23.2. The projector onto the measurement sets $\mathcal{M}_j$, defined in (23.1) was already discussed in Sect. 6.1.2. The projectors onto the support constraint sets are even simpler. The following statement is taken from [1, Example 3.14].

Lemma 23.1 (projectors onto support constraints) *Let $y \in \mathbb{C}^n$, $I \subset \{1, 2, \dots, n\}$. Then the following hold.*

$$\mathcal{P}_{\mathbb{S}} y = z \quad \text{where } z_j = \begin{cases} y_j & \text{if } j \in I \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \leq i \leq n, \quad (23.3)$$

$$\mathcal{P}_{\mathbb{S}_+} y = z \quad \text{where } z_j = \begin{cases} \max\{\text{Re}(y_j), 0\} & \text{if } j \in I \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \leq i \leq n. \quad (23.4)$$

The projectors onto other constraint sets can be found, for instance, in [13] or [14] for a sparsity constraint, or in [1, Example 3.14] for an amplitude constraint or real-valued sparsity constraint. Except for the amplitude and sparsity constraint, all other mentioned constraint sets are closed and convex. The type of regularity of the constraint sets is later discussed in Remark 23.5.1.

Another concept closely related to that of projectors are normal cones.

Definition 23.2 (*Normal cones*) Let $\Omega \subseteq \mathcal{E}$. Define the cone containing Ω by

$$\text{cone}(\Omega) := \mathbb{R}_+ \cdot \Omega := \{\kappa s \mid \kappa \in \mathbb{R}_+, s \in \Omega\}.$$

Let $\Omega \subseteq \mathcal{E}$ and let $x \in \Omega$.

(i) The *proximal normal cone* of Ω at $\bar{x}$ is defined by

$$N_\Omega^{\text{prox}}(x) = \text{cone}(\mathcal{P}_\Omega^{-1}x - x).$$

Equivalently, $x^* \in N_\Omega^{\text{prox}}(x)$ whenever there exists $\sigma \geq 0$ such that

$$\langle x^*, y - x \rangle \leq \sigma \|y - x\|^2 \quad (\forall y \in \Omega).$$

(ii) The *limiting (proximal) normal cone* of Ω at x is defined by

$$N_{\Omega}(x) = \limsup_{z \rightarrow x} N_{\Omega}^{\text{prox}}(z),$$

where the limit superior is taken in the sense of *Painlevé-Kuratowski outer limit* (for more details on the outer limit see for instance [15, Chap. 4]).

When $x \notin \Omega$ all normal cones at x are empty (by definition). If the set Ω is convex, the given definitions of the normal cones coincide (see for instance [16]).

23.3.2 Algorithms

In the context of feasibility problems, a prominent class of iterative algorithms are projection algorithms. Under these, the most prominent and probably one of the easiest to compute is the *method of cyclic projections* as introduced in Sect. 6.2.1. Given a finite number of closed sets $\Omega_1, \Omega_2, \dots, \Omega_m \subseteq \mathcal{E}$ and a point it generates the next iterate by consecutively projecting onto each of the individual sets. For only two sets the algorithm reduces to the method of alternating projections. In Sect. 6.2.3 the error reduction algorithm was identified with the method of alternating projections applied to a measurement and a support constraint. This connection was first made by Levi and Stark in [17]. Considering again only two sets, Sect. 6.1.2 introduced the well-known *Douglas-Rachford algorithm* as well as its relaxed version, the *relaxed averaged alternating reflection* algorithm introduced by Luke in [10]. For one magnitude constraint and a support constraint Douglas-Rachford yields *Fienup's hybrid input output method (HIO)* [18]. The connection of HIO and Douglas-Rachford was already observed by Bauschke et al. [1]. These three algorithms are the ones we want to focus on here. Nevertheless, we want to emphasize that the analysis shown below can be applied also to other projection methods.

Our survey is far from complete. Other approaches worthy of mention are several of the algorithms discussed in Chap. 5 and those in Chap. 6. Readers familiar with the physics literature will also miss the Hybrid Projection Reflection algorithm, [19], difference map, [20], solvent flipping algorithm, [21], and Fienup's Basic Input-Output algorithm (BIO). BIO is, in fact, nothing more than Dykstra's algorithm, see [1]. Like the BIO algorithm, most of the known approaches to phase retrieval fit into a concise scheme presented in [22].

23.3.3 Fixed Points and Regularities of Mappings

We refer to $\text{Fix } T$ as the set of fixed points of the mapping T , i.e. $x \in \text{Fix } T$ if and only if $x \in Tx$. The continuity of set-valued mappings is a well-developed concept and follows the familiar patterns of continuity for single-valued functions. One key

property is *nonexpansiveness*, which nothing more than being Lipschitz continuous with constant 1. That is, given two points, their images under the mapping T are no further away from each other than the initial points. A slightly stronger notion than nonexpansiveness is *averagedness*. For set-valued mappings, a finer distinction of the types of continuity, whether pointwise, or uniform, for example, is necessary. The following definition captures the crucial types of continuity and regularity of set-valued mappings that lie at the heart of numerical analysis of algorithms for phase retrieval.

Definition 23.3 (*almost nonexpansive/averaged mappings*) Let $D \subseteq \mathcal{E}$ and $T : D \rightrightarrows \mathcal{E}$.

- (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

$$\|x^+ - y^+\| \leq \sqrt{1 + \epsilon} \|x - y\| \quad (\forall y^+ \in Ty)(\forall x^+ \in Tx)(\forall x \in D). \quad (23.5)$$

If (23.5) 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* if there is an averaging constant $\alpha \in (0, 1)$ and a violation constant $\epsilon \in [0, 1)$ 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 .

Similarly, 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 $\alpha\epsilon$)*.

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

From the above definition it can easily be seen that if a set-valued mapping is nonexpansive at a point, then it is single-valued there. This is a crucial property for our analytical framework, but should not be confused with uniqueness of fixed points: a multi-valued operator can be single-valued at its fixed points without having unique fixed points.

Proposition 23.4 (single-valuedness, Proposition 2.2 of [3]) Let $T : \mathcal{E} \rightrightarrows \mathcal{E}$ be pointwise almost averaged on $D \subset \mathcal{E}$ at $\bar{x} \in D$ with violation $\epsilon \geq 0$. Then T is single-valued at $\bar{x}$. In particular, if $\bar{x} \in \text{Fix } T$, then $T\bar{x} = \{\bar{x}\}$.

Averaged mappings do not enjoy as nice a calculus as nonexpansive mappings, but the next proposition shows that averagedness of some sort is preserved under addition and composition.

Proposition 23.5 (compositions, Proposition 2.4 of [3]) *Let $T_j : \mathcal{E} \rightrightarrows \mathcal{E}$ for $j = 1, 2, \dots, m$ be pointwise almost averaged on U_j at all $y_j \in S_j \subset \mathcal{E}$ 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 $\epsilon = \sum_{j=1}^m w_j \epsilon_j$ with 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 averaged at all $y \in S_m$ on U_m with violation at most $\epsilon = \prod_{j=1}^m (1 + \epsilon_j) - 1$. and averaging constant at least $\alpha = m / \left(m - 1 + \frac{1}{\max_{j=1,2,\dots,m} \{\alpha_j\}} \right)$.*

23.4 A Toolkit for Convergence

With the characterization of algorithms as simply self mappings with certain regularity properties, we show in this section how those properties come together to guarantee convergence of the algorithm iterations to fixed points. The fixed points need not be solutions to the feasibility problem (indeed, this does not exist for phase retrieval) but will in general be a point that allows one to compute *another point* that does have some physical significance, such as a *local best approximation point*.

It turns out that convergence itself is provided by regularity properties introduced in Sect. 23.3.3. The basic convergence idea goes back to Opial [4]. It says that averagedness of a single-valued mapping T and nonemptiness of the fixed point set imply convergence of the iterative sequence $(T^k x^0)_{k \in \mathbb{N}}$ to a point in $\text{Fix } T$ for any $x^0 \in \mathcal{E}$. Henceforth, we will see that averagedness of T and a nonempty fixed point set is enough to get convergence. As one would expect, it can be difficult for a map to satisfy these properties globally. Nevertheless, this is often the case in nonconvex problem instances. Thus, we seek a statement that includes local properties. That is in our case pointwise almost averagedness as introduced in Definition 23.3.

But convergence alone for iterative procedures is not enough: eventually one has to stop the iteration and without knowing the *rate* of convergence it is impossible to estimate how far a given iterate must be to the solution. A quantitative convergence analysis is achieved with the second essential property: *metric (sub-)regularity*. This concept has been studied by many authors in the literature (see for instance [5–7, 15, 23, 24]). For the definition of metric regularity we need *gauge functions*. A function $\mu : [0, \infty) \rightarrow [0, \infty)$ is a gauge function if it is continuous and strictly increasing with $\mu(0) = 0$ and $\lim_{t \rightarrow \infty} \mu(t) = \infty$. The following definition is taken from [3, Definition 2.5].

Definition 23.6 (*metric regularity on a set*) Let $\Phi : \mathcal{E} \rightrightarrows \mathbb{Y}$, $U \subset \mathcal{E}$, $V \subset \mathbb{Y}$. The mapping Φ is called *metrically regular with gauge μ on $U \times V$ relative to $\Lambda \subset \mathcal{E}$* if

$$\text{dist}(x, \Phi^{-1}(y) \cap \Lambda) \leq \mu(\text{dist}(y, \Phi(x))) \quad (23.6)$$

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

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 $\Lambda = \mathcal{E}$, the quantifier “relative to” is dropped. When μ is linear, the smallest constant κ for which (23.6) holds is called *modulus* of metric regularity.

While this definition might seem abstract there are properties that directly imply metric regularity or reformulations that allow to prove metric regularity. One of these is *polyhedrality* (see [3, Proposition 2.6]). A mapping $T : \mathcal{E} \rightrightarrows \mathcal{E}$ is called polyhedral if its graph is the union of finitely many sets that can be expressed as the intersection of finitely many closed half-spaces and/or hyper-planes [5].

Collecting the concepts we have established so far, we present the following convergence result that goes back to Luke et al. in [3, Theorem 2.2] and was later refined in [25] by Luke et al. to convergence to a specific point.

Theorem 23.4.1 (basic convergence template with metric regularity) *Let $T : \Lambda \rightrightarrows \Lambda$ for $\Lambda \subset \mathcal{E}$, $\Phi := T - \text{Id}$ 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 scalar ϵ_i and a positive constant α bounded above by 1, such that,*

- (i) *T is pointwise almost averaged at all $y \in S$ with averaging constant α and violation ϵ on $S_{\gamma^i \bar{\delta}}$, and*
- (ii) *for $R_i := S_{\gamma^i \bar{\delta}} \setminus (\text{Fix } T \cap S + \gamma^{i+1} \bar{\delta} \mathbb{B})$,*

(i)

$$\text{dist}(x, S) \leq \text{dist}(x, \Phi^{-1}(\bar{y}) \cap \Lambda)$$

for all $x \in R_i$ and $\bar{y} \in \Phi(\mathcal{P}_S x) \setminus \Phi(x)$,

- (ii) *Φ is metrically regular with gauge μ_i relative to Λ on $R_i \times \Phi(\mathcal{P}_S(R_i))$, where μ_i satisfies*

$$\sup_{x \in R_i, \bar{y} \in \Phi(\mathcal{P}_S(R_i)), \bar{y} \notin \Phi(x)} \frac{\mu_i(\text{dist}(\bar{y}, \Phi(x)))}{\text{dist}(\bar{y}, \Phi(x))} \leq \kappa_i < \sqrt{\frac{1 - \alpha_i}{\epsilon_i \alpha_i}}. \quad (23.7)$$

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

$$\text{dist}(x^{j+1}, \text{Fix } T \cap S) \leq c_i \text{dist}(x^j, S) \quad (\forall x^j \in R_i), \quad (23.8)$$

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

In particular, if $\kappa_i \leq \bar{\kappa} < \sqrt{\frac{1-\alpha}{\alpha\epsilon}}$ for all i large enough, then convergence is eventually at least R -linear with rate at most $\bar{c} := \sqrt{1 + \bar{\epsilon} - \left(\frac{1-\alpha}{\bar{\kappa}^2\alpha}\right)}$ to some point in $\text{Fix } T \cap S$. If $S \cap \Lambda$ is a singleton, then (iii) is redundant and convergence is Q -linear.

In both Opial's original statement as well as Theorem 23.4.1 averagedness is the essential property for convergence of iterative algorithms. Whereas assumption (ii) of Theorem 23.4.1 serves to quantify the convergence.

23.5 Regularities of Sets and Their Collection

In this section we connect the regularities of sets to regularities of the projectors on these, which effect the regularity of the mapping T . When dealing with nonconvex sets there are numerous set-regularity definitions available. A recent survey by Kruger et al. [26], sorted the different classes of nonconvex sets to highlight their dependencies and differences. Uniting several concepts of regularity, we propose to use the notion of ϵ -set regularity as introduced in [26] and refined in [27].

Definition 23.7 (ϵ -set regularity) Let $\Omega \subset \mathcal{E}$ be nonempty and let $\bar{x} \in \Omega$. The set Ω is said to be ϵ -subregular relative to Λ at $\bar{x}$ for $(\bar{y}, \bar{v}) \in \text{gph}(N_\Omega)$ if it is locally closed at $\bar{x}$ and there exists an $\epsilon > 0$ together with a neighborhood U of $\bar{x}$ such that

$$\langle \bar{v} - (x - x^+), x^+ - \bar{y} \rangle \leq \epsilon \|\bar{v} - (x - x^+)\| \|x^+ - \bar{y}\|, \quad (\forall x \in \Lambda \cap U)(x^+ \in \mathcal{P}_\Omega x). \quad (23.9)$$

if for every $\epsilon > 0$ there is a neighborhood (depending on ϵ) such that (23.9) holds, then Ω is said to be subregular relative to Λ at $\bar{x}$ for $(\bar{y}, \bar{v}) \in \text{gph}(N_\Omega)$. If $\Lambda = \{\bar{x}\}$, then the qualifier “relative to” is dropped.

In the phase retrieval problem one type of nonconvexity, that is also covered by ϵ -subregularity, is prox-regularity.

Definition 23.8 (prox-regular sets) A closed set Ω is prox-regular at $\bar{x} \in \Omega$ if for $\bar{v} \in N_\Omega(\bar{x})$ there exist $\epsilon, \delta > 0$ such that

$$\frac{\epsilon}{2} \|x - c\|^2 \geq \langle v, x - c \rangle \quad (\forall x, c \in \Omega \cap \mathbb{B}_\delta(\bar{x}))(\forall v \in N_\Omega(c) \cap \mathbb{B}_\delta(\bar{v})).$$

This definition dates back to Federer [28] who called the property *sets with positive reach*. The definition presented here is taken from [29, Proposition 1.2]. The authors in [29] showed that their definition of prox-regularity at $\bar{x} \in C$ is equivalent to several statements. One of the most prominent might be local single-valuedness of the projector [29, Theorem 1.3] around $\bar{x}$. Kruger et al. showed that prox-regularity implies ϵ -subregularity in [26, Proposition 4(vi)]. As the next remark shows all constraint sets involved in the phase retrieval problem are, in fact, prox-regular.

Remark 23.5.1 (phase retrieval constraint sets are prox-regular) Of great importance for the convergence analysis of the introduced algorithms is the ϵ -subregularity of the measurement sets defined in (23.1). By [3, Example 3.1.b] circles are subregular at any of their points $\bar{x}$ for all $(\bar{x}, v)$ in the graph of the normal cone of the sets. As mentioned before ϵ -subregularity covers a diverse range of regularity notions for sets. The measurement sets investigated here are in fact shown to be semi-algebraic [30, Proposition 3.5] and prox-regular by [29, Theorem 1.3] and (6.11).

The other sets that are involved in the phase retrieval problem are the qualitative constraints introduced in (23.2) or mentioned before. Except for the amplitude constraint and the sparsity constraint all of these sets are convex and thus by [3, Proposition 3.1 (vii)] subregular. Fortunately, the amplitude constraint describes coordinatewise circles when the other coordinates are fixed, like the measurement constraint. Hence, the amplitude constraint is ϵ -subregular as well (and additionally semi-algebraic and prox-regular). The sparsity constraint $\mathcal{A}_s$ is prox-regular at all points $\bar{x}$ satisfying $\|\bar{x}\|_0 = s$ (similar to the proof in [14, Proposition 4.4]).

By [12, Proposition 4.8] the projector onto a closed convex set is averaged with constant $\alpha = 1/2$. Allowing sets to have a more general regularity, here prox-regularity, yield regularity of the projectors as well.

Proposition 23.9 (projectors and reflectors onto prox-regular sets) *Let $\Omega \subset \mathcal{E}$ be nonempty closed, and let U be a neighborhood of $\bar{x} \in C$. Let $\Lambda \subset \Omega \cap U$. If Ω is prox-regular at $\bar{x}$ with constant ϵ on the neighborhood U , then the following hold.*

- (i) *Let $\epsilon \in [0, 1)$. The projector $\mathcal{P}_\Omega$ is pointwise almost firmly nonexpansive at each $y \in \Lambda$ with violation $\epsilon_2 := 2\epsilon + 2\epsilon^2$ on U . That is, at each $y \in \Lambda$*

$$\|x - y\|^2 + \|x' - x\| \leq (1 + \epsilon_2) \|x' - y'\|^2 \quad (\forall x' \in U) \quad (\forall x \in \mathcal{P}_\Omega x').$$

- (ii) *The reflector $\mathcal{R}_\Omega$ is pointwise almost nonexpansive at each $y \in \Lambda$ with violation $\epsilon_3 := 4\epsilon + 4\epsilon^2$ on U ; that is, for all $y \in \Lambda$*

$$\|x - y\| \leq \sqrt{1 + \epsilon_3} \|x' - y\| \quad (\forall x' \in U) \quad (\forall x \in \mathcal{P}_\Omega x').$$

Proof By [26, Proposition 4(vi)] prox-regularity of Ω at $\bar{x}$ implies that the set Ω is ϵ -subregular at $\bar{x}$ for all $(c, v) \in \text{gph} N_\Omega$, where $c \in U$. The result follows then from [3, Theorem 3.1].

Note that Proposition 23.9 presents a special case of [3, Theorem 3.1], where the authors allowed their sets to be ϵ -subregular for certain normal vectors. By Proposition 23.5 compositions and convex combinations of averaged mappings are again averaged. Combining this with Proposition 23.9 implies that compositions of projectors are averaged. Thus, the algorithms presented in Sect. 23.3.2 are pointwise almost averaged as we see in Sect. 23.7.

Whereas the regularity of the individual sets imply almost averagedness of the mapping T , metric regularity relies on the regularity of the whole collection of sets

$\{\Omega_1, \Omega_2, \dots, \Omega_m\}$. The idea of regularities of collections of sets traces back to [26, Theorem 3] by Kruger, Luke and Thao, but the analysis there covers only consistent feasibility problems, i.e. the intersection of sets is nonempty. A generalized notion of *subtransversality* proposed in [3, Definition 3.2] includes inconsistent settings too.

Definition 23.10 (*subtransversal collection of sets*) Let $\{\Omega_1, \dots, \Omega_m\}$ be a collection of nonempty closed subsets of $\mathcal{E}$ and define $\Upsilon : \mathcal{E}^m \rightrightarrows \mathcal{E}^m$ by $\Upsilon(x) := \mathcal{P}_\Omega(\Pi x) - \Pi x$ where $\Omega := \Omega_1 \times \Omega_2 \times \dots \times \Omega_m$, the projection $\mathcal{P}_\Omega$ is with respect to the Euclidean norm on $\mathcal{E}^m$ and $\Pi : x = (x_1, x_2, \dots, x_m) \mapsto (x_2, x_3, \dots, x_m, x_1)$ is the permutation mapping on the product space $\mathcal{E}^m$ for $x_j \in \mathcal{E}$ ($j = 1, 2, \dots, m$). Let $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m) \in \mathcal{E}^m$ and $\bar{y} \in \Upsilon(\bar{x})$. The collection of sets is said to be *subtransversal with gauge μ relative to $\Lambda \subset \mathcal{E}^m$ at $\bar{x}$ for $\bar{y}$* if Υ 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 . As in Definition 23.6, when $\mu(t) = \kappa t$, $\forall t \in [0, \infty)$, one says “constant κ ” instead of “gauge $\mu(t) = \kappa t$ ”. When $\Lambda = \mathcal{E}$, the quantifier “relative to” is dropped.

In [3, Proposition 3.3] Luke et al. showed that for a *consistent* feasibility problem subtransversality of the collection of sets is equivalent to what is elsewhere recognized as *linear regularity* of the collection [31].

23.6 Analysis of Cyclic Projections

Having introduced the main tools for convergence, this section is devoted to an explicit demonstration of how this framework can be applied. In particular, we present the main steps of the convergence analysis of the cyclic projection mapping as done by Luke et al. in [3].

As introduced in Algorithm 6.2.1 the method of cyclic projections on a finite collection of closed subsets of $\mathcal{E}$, $\{\Omega_1, \Omega_2, \dots, \Omega_m\}$ ($m \geq 2$), is defined by the mapping

$$\mathcal{P}_0 := \mathcal{P}_{\Omega_1} \mathcal{P}_{\Omega_2} \cdots \mathcal{P}_{\Omega_m} \mathcal{P}_{\Omega_1}, \quad (23.10)$$

that we denote for notational simplicity by $\mathcal{P}_0$. For an initial point u^0 the algorithm generates a sequence $(u^k)_{k \in \mathbb{N}}$ by $u^{k+1} \in \mathcal{P}_0 u^k$. For the analysis of $\mathcal{P}_0$ it is convenient to introduce some auxiliary sets. We denote by Ω the product of the sets Ω_j on $\mathcal{E}^m$,

$$\Omega := \Omega_1 \times \Omega_2 \times \dots \times \Omega_m.$$

Let $\bar{u} \in \text{Fix } \mathcal{P}_0$ and fix $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ where

$$\mathcal{Z}(u) := \{\zeta := z - \Pi z \mid z \in W_0 \subset \mathcal{E}^m, z_1 = u\} \quad (23.11)$$

for

$$W_0 := \{x \in \mathcal{E}^m \mid x_m \in \mathcal{P}_{\Omega_m} x_1, x_j \in \mathcal{P}_{\Omega_j} x_{j+1}, j = 1, 2, \dots, m-1\}. \quad (23.12)$$

Note that $\sum_{j=1}^m \bar{\zeta}_j = 0$. The elements of W_0 are all cycles of the cyclic projection method, where each coordinate of x corresponds to an inner iterate of $\mathcal{P}_0$. The first coordinate x_1 of $x \in W_0$ is, thus, a fixed point of $\mathcal{P}_0$. The vectors $\zeta \in \mathcal{Z}(u)$ are called *difference vectors*. Their coordinate entries provide information about the gaps between the inner iterates of a cycle of the mapping $\mathcal{P}_0$.

To monitor the inner iterations, we consider the cyclic projection algorithm lifted to the product space $\mathcal{E}^m$. That is, generate the sequence $(x^k)_{k \in \mathbb{N}}$ by $x^{k+1} \in T_{\bar{\zeta}} x^k$ with

$$T_{\bar{\zeta}} : \mathcal{E}^m \rightrightarrows \mathcal{E}^m x \mapsto \left\{ \left(x_1^+, x_1^+ - \bar{\zeta}_1, \dots, x_1^+ - \sum_{j=1}^{m-1} \bar{\zeta}_j \right) \mid x_1^+ \in \mathcal{P}_0 x_1 \right\} \quad (23.13)$$

for $\bar{\zeta} \in \mathcal{Z}(\bar{u})$ where $\bar{u} \in \text{Fix } \mathcal{P}_0$. Thus, the first entry of $T_{\bar{\zeta}}$ belongs to the cyclic projection mapping $\mathcal{P}_0$. Whereas the other entries of $T_{\bar{\zeta}} x$ indicate how close or distant x_1^+ is from a certain cycle specified by $\bar{\zeta}$. In order to isolate cycles, we restrict our attention to relevant subsets of $\mathcal{E}^m$. These are

$$W(\bar{\zeta}) := \{x \in \mathcal{E}^m \mid x - \Pi x = \bar{\zeta}\}, \quad (23.14)$$

$$L := \text{an affine subspace with } T_{\bar{\zeta}} : L \rightrightarrows L, \quad (23.15)$$

$$\Lambda := L \cap W(\bar{\zeta}). \quad (23.16)$$

The set $W(\bar{\zeta})$ contains all points whose entries have a certain distance to each other, namely $\bar{\zeta}_i$. In particular, $W(\bar{\zeta})$ contains all fixed points of $T_{\bar{\zeta}}$. The affine subspace L is used to restrict the analysis to an affine subspace that contains the iterates x^k of $T_{\bar{\zeta}}$.

To apply the convergence framework, Theorem 23.4.1, there are two major steps we have to take. First, we have to show that the mapping is averaged. Since the cyclic projection mapping is, as its name suggests, a composition of projectors averagedness, this not hard to show by the concepts presented in Sect. 23.5. Second, metric subregularity needs to be proven. For this, we state an auxiliary result that relates metric subregularity to subtransversality of the collection of sets (see [3, Proposition 3.4]).

Proposition 23.11 (metric subregularity of cyclic projections) *Let $\bar{u} \in \text{Fix } \mathcal{P}_0$ 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 mappings for $\Phi_{\bar{\zeta}} := T_{\bar{\zeta}} - \text{Id}$ and $\Upsilon := (\mathcal{P}_{\Omega} - \text{Id}) \circ \Pi$. Suppose the following hold:*

- (i) *the collection of sets $\{\Omega_1, \Omega_2, \dots, \Omega_m\}$ is subtransversal at $\bar{x}$ for $\bar{\zeta}$ relative to $\Lambda := L \cap W(\bar{\zeta})$ with constant κ and neighborhood U of $\bar{x}$;*
- (ii) *there exists a positive constant σ such that*

$$\text{dist}(\bar{\zeta}, \mathcal{Y}(x)) \leq \sigma \text{dist}(0, \Phi_{\bar{\zeta}}(x)), \quad \forall x \in \Lambda \cap U \text{ with } x_1 \in \Omega_1.$$

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

Proposition 23.11 indicates that subtransversality plus the additional assumption (ii) are enough to deduce metric subregularity of $\Phi_{\bar{\zeta}} := T_{\bar{\zeta}} - \text{Id}$ as required in Theorem 23.4.1. Using this connection and the development in Sect. 23.5 about almost averagedness we can state the following convergence result which is an implication of Theorem 23.4.1.

Theorem 23.6.1 (convergence of cyclic projections) *Let $S_0 \subset \text{Fix } \mathcal{P}_0 \neq \emptyset$ and $Z := \cup_{u \in S_0} \mathcal{Z}(u)$. Define*

$$S_j := \cup_{\zeta \in Z} \left(S_0 - \sum_{i=1}^{j-1} \zeta_i \right) \quad j = 1, 2, \dots, m. \quad (23.17)$$

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

$$\mathcal{P}_{\Omega_j} \left(u - \sum_{i=1}^j \zeta_i \right) \subseteq S_0 - \sum_{i=1}^{j-1} \zeta_i \quad \forall u \in S_0, \quad \forall \zeta \in Z \text{ for each } j = 1, 2, \dots, m, \quad (23.18)$$

$$\mathcal{P}_{\Omega_j} U_{j+1} \subseteq U_j \text{ for each } j = 1, 2, \dots, m \quad (U_{m+1} := U_1). \quad (23.19)$$

For fixed $\bar{\zeta} \in Z$ and $\bar{x} \in S$ with $\bar{\zeta} = \bar{x} - \Pi \bar{x}$, 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 (23.13), seeded by a point $x^0 \in W(\bar{\zeta}) \cap U$ for $W(\bar{\zeta})$ defined by (23.14) with $x_1^0 \in \Omega_1 \cap U_1$.

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:

- (i) Ω_j is prox-regular at all $\hat{x}_j \in S_j$ with constant $\epsilon_j \in (0, 1)$ on the neighborhood U_j for $j = 1, 2, \dots, m$;
- (ii) for each $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m) \in S$, the collection of sets $\{\Omega_1, \Omega_2, \dots, \Omega_m\}$ is subtransversal at $\hat{x}$ for $\bar{\zeta} := \hat{x} - \Pi \hat{x}$ relative to Λ with constant κ on the neighborhood U ;
- (iii) for $\Upsilon_{\bar{\zeta}} := T_{\bar{\zeta}} - \text{Id}$ and $\Psi := (\mathcal{P}_{\Omega} - \text{Id}) \circ \Pi$ there exists a positive constant σ such that for all $\bar{\zeta} \in Z$

$$\text{dist}(\hat{\zeta}, \mathcal{Y}(x)) \leq \sigma \text{dist}(0, \Phi_{\bar{\zeta}}(x))$$

holds whenever $x \in \Lambda \cap U$ with $x_1 \in \Omega_1$;

- (iv) $\text{dist}(x, S) \leq \text{dist}\left(x, \Phi_{\bar{\zeta}}^{-1}(0) \cap \Lambda\right)$ for all $x \in U \cap \Lambda$, for all $\hat{\zeta} \in Z$.

Then the sequence $(x^k)_{k \in \mathbb{N}}$ initialized by a point $x^0 \in W(\bar{\zeta}) \cap U$ with $x_1^0 \in \Omega_1 \cap U_1$ 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 $c := \sqrt{1 + \bar{\epsilon} - \frac{1-\alpha}{\alpha \bar{\kappa}^2}}$ where $\bar{\epsilon} := \prod_{j=1}^m (1 + \tilde{\epsilon}_j) - 1$, $\tilde{\epsilon}_j := 4\epsilon_j \frac{1+\epsilon_j}{(1-\epsilon_j)^2}$, $\alpha := \frac{m}{m+1}$ and $\bar{\kappa} := \kappa\sigma$. If, in addition, $\bar{\kappa} < \sqrt{\frac{1-\alpha}{\epsilon\alpha}}$, then $\text{dist}(x^k, \text{Fix } T_{\bar{\zeta}} \cap S) \rightarrow 0$, and hence $\text{dist}(x_1^k, \text{Fix } \mathcal{P}_0 \cap S_1) \rightarrow 0$ at least linearly with rate $c < 1$.

Proof This is a special case of [3, Theorem 3.2] when the sets are prox-regular.

Remark 23.6.2 Theorem 23.6.1 is rather long and technical at first sight, though the pieces are easily parsed. Equations (23.17)–(23.19) force the iterations to stay in specific neighborhoods. This is needed to apply Proposition 23.9 with the help of (i) to deduce pointwise almost averagedness of $\mathcal{P}_0$ and likewise of $T_{\bar{\zeta}}$. Assumptions (ii) and (iii) then yield metric subregularity of $\Phi_{\bar{\zeta}} = T_{\bar{\zeta}} - \text{Id}$ by Proposition 23.11. This is where the construction in the product space comes into play. Working on $\mathcal{E}^m$, we were able to use subtransversality to show metric subregularity of $\Phi_{\bar{\zeta}}$. It is worth mentioning that, until now, we were not able to show metric subregularity for the mapping directly associated to $\mathcal{P}_0$. Adding assumption (iv) in Theorem 23.6.1 we can finally apply Theorem 23.4.1 and deduce convergence of $T_{\bar{\zeta}}$ with the given constants. At this point the definition of $T_{\bar{\zeta}}$ becomes crucial. Since the first iterate of the sequence x^k generated under the mapping $T_{\bar{\zeta}}$ is nothing more than applying the method of cyclic projections $\mathcal{P}_0$, convergence of x^k implies convergence of x_1^k , that is, the sequence generated by cyclic projections. In [25] Luke et al. discussed the necessity of subtransversality for alternating projections to converge R-linearly.

23.7 Application to Phase Retrieval Algorithms

In Sect. 23.6 we have seen how to apply Theorem 23.4.1 on the method of cyclic projections. This section is devoted to the analysis of other well known algorithms which we introduced in Sect. 23.3.2. The analysis in Sect. 23.6 focuses on showing how to satisfy the assumptions of Theorem 23.4.1 in the context of set-feasibility. This section aims to provide a broad intuition of the convergence of projection based algorithms used to solve the phase retrieval problem. This explains also why the statements given next are presented in a cartoon-like manner. The statements include only the most important parts that yield local convergence, but not how to construct it nor at which rate. Nevertheless, these are verifiable by following the approach in Sect. 23.6.

Corollary 23.12 (convergence of the error reduction algorithm) *Let $\text{Fix } \mathcal{P}_{\ominus} \mathcal{P}_{\mathcal{M}_1} \neq \emptyset$. The error reduction algorithm, that is alternating projections as discussed*

in Sect. 6.2.3 on the sets $\mathfrak{S}$ and $\mathcal{M}_1$, converges locally linearly to a point $\tilde{x} \in \text{Fix } \mathcal{P}_{\mathfrak{S}}\mathcal{P}_{\mathcal{M}_1}$ whenever the mapping $\Phi = \mathcal{P}_{\mathfrak{S}}\mathcal{P}_{\mathcal{M}_1} - \text{Id}$ is locally metrically subregular at its zeros.

Proof Following Luke et al. in [32, Sect. 3.2.2], we represent $\mathbb{C}$ as $\mathbb{R}^2$ and reformulate the phase retrieval problem as a feasibility problem with entrywise values in $\mathbb{R}^2$. Then this is an application of Theorem 23.4.1 using Remark 23.5.1.

Remark 23.7.1 In contrast to Theorem 23.6.1 metric subregularity is required directly in Theorem 23.12. Equivalently, we could demand subtransversality of the collection of sets $\{\mathfrak{S}, \mathcal{M}_1\}$ plus the additional assumption (iii) in Theorem 23.6.1. The problem here is, that, until now, it is not clear when and where these two assumptions are satisfied. Illustrative examples and numerical simulations indicate that they hold in many instances. Nevertheless, there are certain situations when at least one of the two assumptions is violated (see for instance [33]). Moreover, allowing metric subregularity under some gauge depicts the reality sometimes better than restricting the analysis to a linear setting. One example is the setting of alternating projections applied to the sphere $\mathbb{S}$ and a line tangent to $\mathbb{S}$ at $\bar{x} = (0, -1)$. In this instance the algorithm does not converge linearly to $\bar{x}$, although it converges depending on the initial point (see for instance [3]). This problem is not only interesting for the type of convergence, but also when it comes to the actual numerical implementation of algorithms. Although sets in real-life applications intersect tangentially on a set of measure zero, beyond a certain numerical accuracy the distinction between tangential intersection and linear convergence with a rate constant within 15 digits of 1 is rather academic. Having a relatively large gap between sets for inconsistent feasibility is in fact an advantage for the numerical performance of an algorithm.

Theorem 23.7.2 (convergence of Fienup's HIO method) *Let $\beta_n = 1$ for all n and $\text{Fix } \frac{1}{2}(\mathcal{R}_{\mathfrak{S}}\mathcal{R}_{\mathcal{M}_1} + \text{Id}) \neq \emptyset$. The HIO algorithm, defined in (6.9) that is Douglas-Rachford as defined in (6.15) on the sets $\mathfrak{S}$ and $\mathcal{M}_1$, converges locally linearly to a point $\tilde{x} \in \text{Fix } \frac{1}{2}(\mathcal{R}_{\mathfrak{S}}\mathcal{R}_{\mathcal{M}_1} + \text{Id})$ whenever the mapping $\Phi = \frac{1}{2}(\mathcal{R}_{\mathfrak{S}}\mathcal{R}_{\mathcal{M}_1} + \text{Id}) - \text{Id}$ is locally metrically subregular at its zeros.*

Proof Since Fienup's HIO for $\beta_n = 1$ for all n can be identified with the Douglas-Rachford method the result follows from [3, Theorem 3.4].

Even if one had an infinite detector, noisy measurements make the phase retrieval problem almost always inconsistent. It is easy to prove [8, Theorem 3.13] that, in this case, $\text{Fix } \frac{1}{2}(\mathcal{R}_{\mathfrak{S}}\mathcal{R}_{\mathcal{M}_1} + \text{Id}) = \emptyset$ and so Φ does not possess zeros. Consequently, Fienup's HIO algorithm *cannot* converge. To circumvent this problem, one can use a relaxed version of Douglas-Rachford, the relaxed averaged alternating reflections method (RAAR), that we introduced in Sect. 6.1.2 which is adapted to inconsistent feasibility.

Theorem 23.7.3 (convergence of RAAR) *relaxed averaged alternating reflections. Let $\bar{x} \in \text{Fix } T_{\text{RAAR}}$ for T_{RAAR} defined in (6.22). The relaxed averaged alternating reflections applied to a phase retrieval problem converges locally linearly to a point*

$\tilde{x} \in \text{Fix } T_{RAAR}$ whenever the mapping $\Phi = \frac{\lambda}{2} (\mathcal{R}_{\mathcal{S}} \mathcal{R}_{\mathcal{M}_1} + \text{Id}) + (1 - \lambda) \mathcal{P}_{\mathcal{M}_1} - \text{Id}$ is locally metrically subregular at its zeros.

A detailed proof of the convergence analysis for the relaxed averaged alternating reflection algorithm can be found in [33] by the authors of this chapter. There we use subtransversality of the collections of sets in general feasibility problems to make the connection to metric subregularity of the algorithm in question. The analysis does not use prox-regularity as the desired type of regularity for sets yielding the almost averaging property, but rather the property of being *super-regular at a distance*. This extends notions of regularity of sets to their effect on points that are not in the sets. Their definition is in line with ϵ -subregularity and is thus connected to the analysis of [3].

Remark 23.7.4 In [33] we not only provided a convergence statement for the relaxed averaged alternating reflections method, but also gave a description of the fixed point set of the underlying mapping. For super-regular sets at a distance, the fixed points, if they exist, are either points in the intersection of both sets or relate to the local gap between these, if the intersection of the sets is empty. This result is in line with [11] where Luke studied the case of one set being convex and the other prox-regular. In contrast to the original Douglas-Rachford algorithm, the main advantage of the relaxed version is that existence of fixed points does not depend on whether the feasibility problem is consistent. Connecting this observation to the convergence analysis presented here, in practice the Douglas-Rachford/HIO is much less stable than the relaxed version.

Following the ideas above, it is not hard to show that most projection methods are pointwise almost averaged mappings when applied to the phase retrieval problem. Nonetheless, the property of metric subregularity is still an open problem in some important cases. Thus, local convergence can be easily verified, but it is hard to quantify.

23.8 Final Remarks

When it comes to computing (see Remark 23.7.1), whether a method converges, let alone determining the rate depends on the numerical precision. But also inconsistency has an impact on the numerical performance. Closely related to this, we want to stress another feature of the analysis surveyed here. That is, sometimes less information can lead to better performance of an algorithm. For a demonstration we analyze a data set recorded by undergraduates at the X-Ray Physics Institute at the University of Göttingen. It is an optical diffraction image with model constraints $\sqrt{I_{j,i}}$, $j = 1, 2, \dots, m$, as in (23.1) with $m = 1$ and n the dimension of the image and additional support constraint. The full data set has dimension $n = 1392 \times 1040$, the cropped data set $n = 128^2$. The graphs shown in Figs. 23.1 and 23.2 are produced by applying the alternating projection algorithm, i.e. error reduction, on the data sets individually.

As it turns out alternating projections on the full data set (Fig. 23.2) shows a worse convergence behavior than the image with the limited data set (Fig. 23.1). Not only that the algorithm needs more iterations to reach a certain accuracy (9.8485×10^4 instead of 666), but also the rate of linear convergence when the iterates reach a suitable neighborhood is worse. Noteworthy is the observed gap in both problem instances. In the full data set version the gap is smaller than in the version with a limited data set. We conjecture that this behavior is closely related to the property of metric subregularity, or in the context of set feasibility, subtransversality. The more, and better, information one has, the closer the constraint sets come to intersect. But this can included cases in which the sets intersect transversally as well. In cases like these the method of alternating projections does not have to converge locally linearly but can show a sublinear convergence behavior (see for instance [3, Remark 3.2]). The take home message in this context is that more information does not have to yield a better image when applying numerical algorithms. This is good news and bad news for these algorithms. The good news is that one can profit from implicit regularization with smaller problem sizes. The bad news is that this indicates a type of *dimension dependence* of these methods: the higher the dimension, the worse the constants in the linear convergence rates. This is not surprising and points to the need for models that lead to algorithms whose performance (that is, regularity) is dimension independent. While our discussion here focuses on the theoretical analysis rather than the comparison of the presented algorithms we point the reader to a study by Luke et al. [22], where the authors present a thorough review of first-order proximal methods for phase retrieval algorithms.

Fig. 23.1 Alternating projections applied to the data set “tasse” for full data (blue) and cropped data (orange). **a** Change in iterates until iteration 900. **b** Gap distance until iteration 50

Fig. 23.2 Change in iterates for full data set “tasse”

References

1. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *J. Opt. Soc. Am. A* **19**(7), 1334–1345 (2002)
2. Marchesini, S.: A unified evaluation of iterative projection algorithms for phase retrieval. *Rev. Sci. Instrum.* **78**(1), 011301 (2007)
3. Luke, D.R., Thao, N.H., Tam, M.K.: Quantitative convergence analysis of iterated expansive, set-valued mappings. *Math. Oper. Res.* **43**(4), 1143–1176 (2018). <https://doi.org/10.1287/moor.2017.0898>
4. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.* **73**(4), 591–597 (1967)
5. Dontchev, A.L., Rockafellar, R.T.: *Implicit Functions and Solution Mappings*, 2nd edn. Springer, Dordrecht (2014)
6. Ioffe, A.D.: Regularity on a fixed set. *SIAM J. Optim.* **21**(4), 1345–1370 (2011)
7. Ioffe, A.D.: Nonlinear regularity models. *Math. Program.* **139**(1–2), 223–242 (2013)
8. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. *J. Approx. Theory* **127**(2), 178–192 (2004)
9. Borwein, J.M., Tam, M.K.: The cyclic Douglas-Rachford method for inconsistent feasibility problems. *J. Nonlinear Convex Anal.* **16**, 537–584 (2015)
10. Luke, D.R.: Relaxed averaged alternating reflections for diffraction imaging. *Inverse Probl.* **21**(1), 37 (2005)
11. Luke, D.R.: Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space. *SIAM J. Optim.* **19**(2), 714–739 (2008)
12. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York (2011)
13. Hesse, R., Luke, D.R., Neumann, P.: Alternating projections and Douglas-Rachford for sparse affine feasibility. *IEEE Trans. Signal Process.* **62**(18), 4868–4881 (2014)
14. Tam, M.K.: Regularity properties of non-negative sparsity sets. *J. Math. Anal. Appl.* **447**(2), 758–777 (2017)
15. Rockafellar, R., Wets, R.: *Variational Analysis*. Springer, Berlin (1998)
16. Mordukhovich, B.S.: *Variational Analysis and Applications*. Springer Monographs in Mathematics. Springer International Publishing (2018). <https://books.google.de/books?id=6DxnDwAAQBAJ>
17. Levi, A., Stark, H.: Image restoration by the method of generalized projections with application to restoration from magnitude. *JOSA A* **1**(9), 932–943 (1984)
18. Fienup, J.R.: Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**(15), 2758–2769 (1982)

19. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Hybrid projection-reflection method for phase retrieval. *J. Opt. Soc. Am. A* **20**(6), 1025–1034 (2003)
20. Elser, V.: Phase retrieval by iterated projections. *JOSA A* **20**(1), 40–55 (2003)
21. Abrahams, J.P., Leslie, A.G.W.: Methods used in the structure determination of bovine mitochondrial F1 atpase. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **52**(1), 30–42 (1996)
22. Luke, D.R., Sabach, S., Teboulle, M.: Optimization on spheres: models and Proximal algorithms with computational performance comparisons. *SIAM J. Math. Data Sci.* **1**(3), 408–445 (2019)
23. Aze, D.: A unified theory for metric regularity of multifunctions. *J. Convex Anal.* **13**(2), 225 (2006)
24. Penot, J.P.: Metric regularity, openness and lipschitzian behavior of multifunctions. *Nonlinear Anal.: Theory Methods Appl.* **13**(6), 629–643 (1989)
25. Luke, D.R., Teboulle, M., Thao, N.H.: Necessary conditions for linear convergence of iterated expansive, set-valued mappings. *Math. Program.* 180:1–31 (2020). <https://doi.org/10.1007/s10107-018-1343-8>
26. Kruger, A.Y., Luke, D.R., Thao, N.H.: Set regularities and feasibility problems. *Math. Program.* **168**(1–2), 279–311 (2018)
27. Daniilidis, A., Luke, D.R., Tam, M.K.: Characterizations of super-regularity and its variants. In: *Splitting Algorithms. Modern Operator Theory and Applications*. Springer (2019). <https://arxiv.org/abs/1808.04978>
28. Federer, H.: Curvature measures. *Trans. Am. Math. Soc.* **93**(3), 418–491 (1959)
29. Poliquin, R.A., Rockafellar, R., Thibault, L.: Local differentiability of distance functions. *Trans. Am. Math. Soc.* **352**(11), 5231–5249 (2000)
30. Hesse, R., Luke, D.R., Sabach, S., Tam, M.K.: Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging. *SIAM J. Imaging Sci.* **8**(1), 426–457 (2015)
31. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. *SIAM Rev.* **38**(3), 367–426 (1996)
32. Luke, D.R., Burke, J.V., Lyon, R.G.: Optical wavefront reconstruction: theory and numerical methods. *SIAM Rev.* **44**(2), 169 (2002)
33. Luke, D.R., Martins, A.L.: Convergence analysis of the relaxed Douglas-Rachford algorithm. *SIAM J. Optim.* (to appear). <https://arxiv.org/abs/1811.11590>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.