Fixed Point Algorithms for Nonconvex Feasibility with Applications

Dissertation zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
Doctor rerum naturalium
im Promotionsprogramm *PhD School of Mathematical Sciences*
der Georg-August University School of Science (GAUSS)

vorgelegt von Robert Hesse aus Northeim

Göttingen 2014

Betreuungsausschuss

- Prof. Dr. Russell Luke, Institut für numerische und angewandte Mathematik, Georg-August-Universität Göttingen
- Prof. Dr. Thorsten Hohage, Institut für numerische und angewandte Mathematik, Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission

- Referent:
  Prof. Dr. Russell Luke, Institut für numerische und angewandte Mathematik, Georg-August-Universität Göttingen
- Korreferent:
  Prof. Dr. Gerlind Plonka-Hoch, Institut für numerische und angewandte Mathematik, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission

- Prof. Dr. Thomas Schick, Mathematisches Institut, Georg-August-Universität Göttingen
- Prof. Dr. Valentin Blomer, Mathematisches Institut, Georg-August-Universität Göttingen
- Jun.-Prof. Dr. Andrea Krajina, Institut für Mathematische Stochastik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 14.07.2014

"Philosophy is written in that great book which ever lies before our eyes –I mean the universe – but we cannot understand it if we do not first learn the language and grasp the symbols, in which it is written. This book is written in the mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth." - Galileo Galilei

"Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius –and a lot of courage – to move in the opposite direction." - Albert Einstein

Contents

Acknowledgements

First of all, I would like to thank my advisor and mentor Prof. Dr. Russell Luke, for introducing me to the fascinating fields of nonconvex feasibility and optical wavefront reconstruction and for the many hours he has spent making me a better scientist and mathematician. I am very grateful for his encouragement, motivation and enthusiasm.

I want to express my gratitude to Prof. Dr. Gerlind Plonka-Hoch for her effort as a referee for this thesis.

My thanks go to the group of Prof. Dr. Tim Salditt at the IRP Göttingen for providing valuable support and resources. I am particularly thankful to Robin Wilke for interesting discussions about his experiences, intuitive explanations about experimental set-ups, and providing me with “real data”.

I thank Prof. Dr. Heinz Bauschke for his hospitality at UBC Kelowna. I benefited a lot from his resourceful work and input. He and his group gave me continuous encouragement and inspired me to reach for even stronger results.

Matthew K. Tam and Pär Persson Mattsson deserve great credit for optimizing and maintaining code in our workgroups “ProxToolbox”.

Many other people at the Institute for Numerical and Applied Mathematics earn my thanks for many discussions and making the institute an enjoyable place to work. In particular I always had nice discussion with Carolin Homann, Marius Wischerhoff and Patrick Neumann and they always lend me an ear. The latter and Stefan Loock earn credit for proofreading parts of my manuscript.

Ein besonderer Danke gilt meinen Eltern, meiner Familie und meinen Freunden, für ihre Unterstützung während meines gesamten Studiums und ihr Verständnis in letzter Zeit.

Special thanks go to my wife Amélie Carolina for her support, comfort, tolerance and love.

This work has been supported financially by the DFG Collaborative Research Center SFB 755 “Nanoscale Photonic Imaging” and the Graduiertenkolleg 1023 “Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden”.

Introduction

This work lies at the intersection of Optimization, Variational Analysis and applications in a wide spectrum of many disparate disciplines such as imaging science, signal processing and inverse scattering, to mention just a few.

Mainly, I am interested in studying and developing qualitative and quantitative characterizations of convergence behavior of numerical methods and using theoretical insight to implement efficient algorithms for solving real-world problems and to legitimate and speed up existing algorithmic schemes. I am inspired by algorithms that have been applied to solve practical problems without theoretical justification or explanation and eager to identify structures that lead to the success of these methods.

This thesis covers both the theoretical (Part I) and practical (Part II) aspects of these fascinating areas.

Fixed Point Theory - Feasibility Problems

Projection algorithms for solving (nonconvex) feasibility problems provide powerful and computationally efficient schemes for a wide variety of applications. Algorithms as *Alternating Projections* (**AP**) and *Douglas–Rachford* (**DR**) are two of the more prominent projection algorithms in imaging sciences and signal processing (Part II). These methods also have been applied successfully to sparse image reconstruction (Bauschke et al., 2013b; Hesse et al., 2014a; Demanet and Zhang, 2013) and combinatorial optimization problems (Artacho et al., 2013; Elser et al., 2006). An introduction to feasibility problems, the fundamental algorithmic schemes **AP** and **DR** will be given in Chapter 1.

In Part I of this thesis a nonconvex framework is introduced that enables a general and new approach to characterizing the convergence behavior of general fixed point operators. In classical fixed point theory, firm nonexpansiveness of mappings is a property that is often used to show convergence of a broad class of algorithms. An overview over the classical concepts is given in Chapter 2.

Firm nonexpansiveness of projectors onto convex sets is closely related to the *best approximation property* for convex sets. As our main interest is dealing with nonconvex feasibility, the described methodsno longer match the notion of firm nonexpansiveness. In Chapter 3 several nonconvex notions of set regularity are introduced and discussed, in order to provide reasonable assumptions and an essential fundament for the following analysis.

The framework, theorems and concepts developed in Chapter 4 then generalize the tools from convex analysis for the analysis of fixed-point iterations of operators that violate the classical property of firm nonexpansiveness in some quantifiable fashion.

Chapter 5 provides quantitative characterizations of regularity of collections of sets and regularity of fixed point sets. This theory is essential in characterizing the convergence behavior of algorithms and in achieving (optimal) rates of convergence. In some of the applications the current convergence rates are still not optimal due to the lack of appropriate quantitative characterizations in the literature. However, Chapter 5 provides some new results, relating and unifying different notions of regularity, that are *primal notions* like *uniform* and *linear regularity*, *metric (sub-)regularity* and more *dual notions* like *normal cone conditions*.

Using these techniques, Chapter 6 then carries out the convergence analysis on **AP** and **DR**. A preliminary version of this result was published in (Hesse and Luke, 2013). In the nonconvex setting, to the best of our knowledge, these results were the most general at that time, and they are by now complemented by several authors (Bauschke et al., 2013a; Phan, 2014; Bauschke and Noll, 2014). Incorporating the results from Chapter 5 the main (nonconvex) convergence results of Chapter 6 are simplified in comparison to other results in the current literature.

Compressed Sensing - new results on an alternative approach

The problem of finding a vector with the fewest nonzero elements that satisfies an under-determined system of linear equations is an NP-complete problem that is typically solved numerically via convex heuristics or nicely-behaved nonconvex relaxations. The nonconvex notions of regularity described Chapter 3 fit naturally in the framework of sparse image reconstruction. In Chapter 8 elementary methods based on projections for solving the *sparse feasibility* problem are considered. In contrast to methods based on convex heuristics, these results provide an interesting and more direct approach than usual convex relaxations.

Phase Retrieval and Ptychographic Imaging

The *Phase Retrieval Problem* is an ill-posed inverse problem, where one seeks to determine the shape – or more precise the amplitude and complex phase – of an unknown object from its intensity measurement in the measurement plane (detector). The reconstruction of the object from one intensity measurement is not possible, so usually we need to incorporate additional, *a priori information*, about the object, *i.e.,* given support or amplitude or *sparsity* in some basis.

The theory developed in Part I provides insight into the behavior of classical algorithms such as the *Gerchberg-Saxton-*, *Error Reduction-* and *Hybrid-Input-Output-Algorithm* as well as to more advanced schemes as the *Difference Map Algorithm* or the *Relaxed Averaged Alternating Reflection Algorithm* (RAAR) (*regularized Douglas–Rachford*). All of these methods are basically variants of **AP** and **DR** and some of them are still state-of-the-art methods in this field, which will be explored in Section 9.4. *Ptychographic Imaging* for simultaneous probe and object reconstruction in complex wave fronts in X-ray microscopy are also modeled within our above mentioned theoretical framework (Section 9.6).

The characterization of the convergence behavior developed in the fixed point theory allows us to speed up existing algorithmic schemes. The flexibility of the *feasibility problem framework* allows us to easily incorporate different new physical constraints as additional a priori information into the existing algorithms. The commonly used more heuristic schemes lack this adaptivity and theoretical foundation.

Part I.

Projection Methods - Local Geometry and Convergence

1. Feasibility Problems – Projection Algorithms

1.1. Notation – Foundations

Most of the notation in this work is standard and should be clear from the context. Throughout this work $\mathcal{H}$ is a Hilbert space equipped with a real inner product $\langle\cdot,\cdot\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{R}$, while $\mathbb{E}$ is an Euclidean space, *i.e.,* a finite dimensional space equipped with a real inner product. If not explicitly stated otherwise, norms and distance function are implicitly referred to Euclidean norm and distance function $\|x\| := \sqrt{\langle x,x\rangle}$ and $d\,(x,y) := \|x-y\|$ (Exceptions are in some of the Remark and at the beginning of the Part, where they are explained explicitly to circumvent any confusions). $\mathbb{B}_\delta(\bar{x}) := \{x\in\mathcal{H} \mid \ d\,(x,\bar{x})\leq\delta\}$ is the closed ball with radius δ centered at $\bar{x}$. We will use the notation $A:\mathcal{H}\rightrightarrows\mathbb{Y}$ to indicate a *set-valued* operator A that maps $\mathcal{H}$ to subsets of a Hilbert space $\mathbb{Y}$. For an operator $A:\mathcal{H}\rightrightarrows\mathbb{Y}$ its *graph* is given by

$$\mathrm{gph}A := \{(x,y)\in\mathcal{H}\times\mathbb{Y} \mid y\in Ax\}\,. \tag{1.1}$$

$\mathbb{R}_+ := \{\lambda\in\mathbb{R}\mid\lambda\geq 0\}$ is the nonnegative real line, while

$$\mathbb{R}_{++} := \mathbb{R}_+\backslash\{0\} = \{\lambda\in\mathbb{R}\mid\lambda>0\}\,.$$

For a subset $\Omega\subset\mathcal{H}$ we define

$$\mathrm{cone}\Omega := R_{++}\Omega = \{\lambda x\mid x\in\Omega,\ \lambda\in\mathbb{R}_{++}\}$$

and imply that K is a cone if and only if $K=\mathrm{cone}K$. According to (Rockafellar and Wets, 1998, Equation 4(2)) for any sequence of sets $\{\Omega_n\}_{n\in\mathbb{N}}$ the *outer limit* and *inner limit* are given by

$$\limsup_{n\to\infty}\Omega_n := \left\{x \;\middle|\; \begin{array}{c}\forall\varepsilon>0, \exists\,(\Omega_{n_k})_{k\in\mathbb{N}} \text{ subsequence of } (\Omega_n)_{n\in\mathbb{N}}:\\ d\,(x,\Omega_{n_k})\leq\varepsilon,\ \forall k\in\mathbb{N}\end{array}\right\}, \tag{1.2}$$

$$\liminf_{n\to\infty}\Omega_n := \{x\mid\forall\varepsilon>0,\ \exists N\in\mathbb{N}:\ \ d\,(x,\Omega_n)\leq\varepsilon,\ \forall n\geq N\}\,, \tag{1.3}$$

where $d\,(x,\Omega)=\inf_{y\in\Omega}\|x-y\|$ is the *distance* of x to Ω.

1.2. The Two Set Feasibility Problem

Given two (possibly nonconvex) nonempty subsets $\mathbf{A}, \mathbf{B}$ of a Hilbert space $\mathcal{H}$ one wants to find a point $\bar{x}$ that lies in the intersection $\mathbf{A} \cap \mathbf{B}$, i.e. the two set feasibility problem is

$$\text{Find } \bar{x} \in \mathbf{A} \cap \mathbf{B}. \tag{1.4}$$

1.2.1. More than two sets - Pierra's trick

If one wants to find a point in the intersection of more than two, say m, sets Ω_i one faces the feasibility problem

$$\bar{x} \in \cap_{i=1}^{m} \Omega_i. \tag{1.5}$$

Using *Pierra's product space trick* (Pierra, 1984) this can be reformulated as a two set feasibility problem (1.4).

Note that $\bar{x}$ solves the m-set feasibility problem (1.5) if and only if $\bar{x} \in \Omega_i$ for all $i = 1, \ldots, m$ which is equivalent to

$$\underbrace{(\bar{x}, \ldots, \bar{x})}_{m-\text{times}} \in \Omega_1 \times \cdots \times \Omega_m.$$

By defining the product set $\mathbf{A}$ and the diagonal space $\mathbf{B}$ by

$$\mathbf{A} := \Omega_1 \times \cdots \times \Omega_m, \tag{1.6}$$

$$\mathbf{B} := \left\{ (x_1, \ldots, x_m) \mid x_i \in \mathcal{H} \text{ and } x_1 = \cdots = x_m \right\}. \tag{1.7}$$

The corresponding projectors onto $\mathbf{A}$ and $\mathbf{B}$ are then given by

$$P_{\mathbf{A}}\left(x_1, \ldots, x_m\right) = \left(P_{\Omega_1} x_1, \ldots, P_{\Omega_m} x_m\right), \tag{1.8}$$

$$P_{\mathbf{B}}\left(x_1, \ldots, x_m\right) = \sum_{k=1}^{m} \frac{1}{m} x_k \tag{1.9}$$

Using this product space formulation one can reduce any m-set feasibility problem to a two set feasibility problem in order to apply two set algorithmic schemes as von Neumanns Alternating Projections sequence or the Douglas–Rachford algorithm.

1.3. Distance Function, Proximity Operators

Definition 1.1 (distance function)**:**
Let $\Omega \subset \mathcal{H}$ be nonempty, $x \in \mathcal{H}$. The *distance* of x to Ω is defined by

$$d\left(x, \Omega\right) := \inf_{y \in \mathcal{H}} \left\| x - y \right\|. \tag{1.10}$$

Definition 1.2 (Best approximation points and projector)**:**
Let $\Omega \subset \mathcal{H}$ be nonempty and $x \in \mathcal{H}$. An element $\bar{x} \in \Omega$ is a *best approximation* to x in Ω if

$$\bar{x} \in \arg\min_{y \in \Omega} \|x - y\| . \tag{1.11}$$

The (possibly empty) set of all best approximation points from x to Ω denoted $P_\Omega(x)$, is given by

$$P_\Omega x := \{y \in \Omega \mid \|x - y\| = d(x, \Omega)\} . \tag{1.12}$$

The mapping $P_\Omega : \mathcal{H} \rightrightarrows \Omega$ ist called *projector* onto Ω.

Remark 1.3*: The projector is also called* metric projection*,* proximity map *or* projection operator*. The term* metric projection *was first used in (Aronszajn and Smith, 1954). A best approximation point,* i.e., *a point in the projector is also called* nearest point to *or* projection of x*. In the following, if the projector $P_\Omega x$ of a point x onto a set Ω defines a singleton,* i.e., $P_\Omega x = \{\bar{x}\}$*, the notation $P_\Omega x := \bar{x}$ will be used which is [if any] a slight abuse of notation.* $\Diamond$

Definition 1.4 (Proximinal (Phelps, 1957, p. 790) and Chebyshev (Efimov and Steckkin, 1958) sets)**:**
A set $\Omega \subset \mathcal{H}$ is called *proximinal* if

$$P_\Omega(x) \neq \emptyset \quad \text{for all } x \in \mathcal{H}. \tag{1.13}$$

If in addition the projection is *single-valued* the set is called *Chebyshev*.

Theorem 1.5 ((Deutsch, 2001, Theorem 3.1))**:**
A proximinal set $\Omega \subset \mathcal{H}$ (and thereby a Chebyshev set) is closed.

Proof. Assume Ω is not closed. Then there is a sequence $(x_n)_{n\in\mathbb{N}} \subset \Omega$ such that $x_n \to x$ as $n \to \infty$ but $x \notin \Omega$. By the definition of x one has $x \in \operatorname{cl}(\Omega)$ and therefore $P_\Omega(x) = \emptyset$ which contradicts the set Ω being proximinal. □

Proposition 1.6 ((Deutsch, 2001, Theorem 3.8))**:**
On a finite dimensional Hilbert space $\mathcal{H}$ a nonempty set $\Omega \subset \mathcal{H}$ is proximinal if and only if it is closed.

1.4. Von Neumann's Alternating Projection Algorithm

Definition 1.7 (Alternating Projections)**:**
For two nonempty, closed sets $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ the mapping

$$\mathrm{T}_{\mathrm{AP}}\, x := P_{\mathbf{A}} P_{\mathbf{B}} x \tag{1.14}$$

is called the *alternating projections operator.*

For given initial $x_0 \in \mathcal{H}$ any sequence generated by

$$x_{n+1} \in \mathrm{T}_{\mathrm{AP}}\, x_n \tag{1.15}$$

is called *von Neumann's alternating projection algorithm* or simply *von Neumann sequence* or *Alternating Projections.*

1.4.1. Convergence of Alternating Projections: History and known Results

Remark 1.8 (Convergence results for convex sets)*: We summarize some of the results on von Neumanns Alternating Projections:*

- *(von Neumann, 1933): Let* $\mathbf{A}, \mathbf{B}$ *be closed subspaces of* $\mathcal{H}$*. For any* $x \in \mathcal{H}$

$$\lim_{n\to\infty} \left(P_{\mathbf{A}} P_{\mathbf{B}}\right)^n x = P_{\mathbf{A}\cap\mathbf{B}} x. \tag{1.16}$$

- *(Aronszajn, 1950): Let* $\mathbf{A}, \mathbf{B}$ *be closed subspaces of* $\mathcal{H}$*. Then*

$$\sup_{\|x\|=1} \left\| \left(\left(P_{\mathbf{A}} P_{\mathbf{B}}\right)^n - P_{\mathbf{A}\cap\mathbf{B}} \right) x \right\| \leq c_F \left(\mathbf{A}, \mathbf{B}\right)^{2n-1}, \tag{1.17}$$

  where $c_F(\mathbf{A}, \mathbf{B})$ *is the* Friedrichs angle *between* $\mathbf{A}$ *and* $\mathbf{B}$*, compare Definition 5.1 equation* (5.1)*.*

- *(Cheney and Goldstein, 1959, Theorem 4) Let* $\mathbf{A}, \mathbf{B}$ *be closed and convex and let either* $\mathbf{A}$ *or* $\mathbf{B}$ *be compact* or *finite dimensional with* $d(\mathbf{A}, \mathbf{B}) = \|a - b\|$ *for some* $a \in \mathbf{A}$, $b \in \mathbf{B}$ *then*

$$\lim_{n\to\infty} \left(P_{\mathbf{A}} P_{\mathbf{B}}\right)^n x = \bar{x} \in \mathrm{Fix}\left(P_{\mathbf{A}} P_{\mathbf{B}}\right). \tag{1.18}$$

  Note that the result does not necessarily require $\mathbf{A} \cap \mathbf{B} \neq \emptyset$*.*

- *(Gubin et al., 1967): Convergence for* m *closed convex sets* $\Omega_1, \ldots, \Omega_m$ *of the* cyclic projections algorithm*,* i.e., $P_{\Omega_1} \cdots P_{\Omega_m}$ *with a linear rate to a point in* $\bigcap_{i=1}^m \Omega_i$*.*

- *(Bauschke and Borwein, 1993): Introduction of linear and bounded linear regularity for convex sets. See Chapter 5 and Remark 5.8. The authors prove linear convergence with rates for general closed, convex sets. Compare Corollary 6.4* $\mathbf{A}, \mathbf{B}$*.*

- *(Deutsch, 2001; Deutsch, 1991): Detailed survey on the method of alternating projections.*

- *(Deutsch and Hundal, 2006a; Deutsch and Hundal, 2006b; Deutsch and Hundal, 2008): Study of regularity of set intersection and characterization of convergence of cyclic projection sequences.*

◊

Remark 1.9 (Convergence results for nonconvex sets)*: We summarize some of the nonconvex convergence results for von Neumanns Alternating Projections:*

- *(Lewis and Malick, 2008; Lewis et al., 2009): First* nonconvex *convergence results for the alternating projection algorithm and introduction of new nonconvex notions of sets. That is,* super-regularity *[Definition 3.14 (a)], and a transversality conditions for the collection* $\{\mathbf{A}, \mathbf{B}\}$ *[Remark 5.22 equation* (5.31)*].*
- *(Bauschke et al., 2013d; Bauschke et al., 2013c; Bauschke et al., 2013b):* Quantification *of (Lewis et al., 2009),* i.e., *introduction of* (ε, δ)*-regularity [Definition 3.14 (b)], CQ-number* Θ_δ *[Definition 5.40].*
- *(Hesse and Luke, 2013): Introduction of* (ε, δ)*-subregularity [Definition 3.14 (c)] and introduction of a* nonconvex *approach different from (Lewis et al., 2009) more related to the approach in (Bauschke and Borwein, 1993). Compare Corollary 6.4.*

◊

1.5. Douglas–Rachford Algorithm

Definition 1.10:
Let $\Omega \subset \mathcal{H}$ be nonempty and closed. The reflector $R_\Omega : \mathcal{H} \rightrightarrows \mathcal{H}$ to the set Ω ist defined by

$$R_\Omega x := 2P_\Omega x - x, \tag{1.19}$$

for all $x \in \mathcal{H}$.

Definition 1.11 (Douglas–Rachford Algorithm/Averaged Alternating Reflections)**:**
For two nonempty, closed sets $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ the mapping

$$\mathrm{T}_{\mathrm{DR}}\, x := \frac{1}{2}\left(R_{\mathbf{A}} R_{\mathbf{B}} x + x\right) \tag{1.20}$$

is called *Douglas–Rachford Operator.*

For given initial $x_0 \in \mathcal{H}$ any sequence generated by

$$x_{n+1} \in \mathrm{T}_{\mathrm{DR}}\, x_n \tag{1.21}$$

is called *Douglas–Rachford algorithm* or *Averaged Alternating Reflections* .

Remark 1.12*: What we are calling Douglas–Rachford algorithm was first introduced in (Douglas and Rachford, 1956) as an operator splitting technique for partial differential equations. In fact the original definition is more related to the equivalent formula* (1.22) *that we will see in Lemma 1.13. The definition of the Douglas–Rachford operator given in equation* (1.20) *is motivated by the rather geometric interpretation in the case of set feasibility, that is,* Averaged Alternating Reflections*.*

For a detailed study on operator splitting and Douglas–Rachford aside set feasibility see (Lions and Mercier, 1979; Eckstein and Bertsekas, 1992) and the more recent comprehensive works (Eckstein and Svaiter, 2008) and (Eckstein and Svaiter, 2009). The Douglas–Rachford algorithm owes its prominence in large part to its relation via duality to the alternating directions method of multipliers (ADMM) for solving constrained optimization problems, see (Gabay, 1983).

Due to its success in solving nonconvex feasibility problems (see for an interesting survey for instance (Artacho et al., 2013) and for concrete examples (Demanet and Zhang, 2013; Hesse et al., 2014a)) the Douglas–Rachford algorithm has evolved into a topic of intense research during the last years (Borwein and Sims, 2011; Bauschke et al., 2013a; Bauschke and Noll, 2014; Phan, 2014). ◊

Lemma 1.13 (Equivalent definition of the Douglas–Rachford Operator)**:**
For all $x \in \mathbb{E}$

$$\mathrm{T}_{\mathrm{DR}}\, x = \left\{ P_{\mathbf{A}}(2z - x) - z + x \mid z \in P_{\mathbf{B}} x \right\}. \tag{1.22}$$

Proof.

$$\begin{aligned}
\mathrm{T}_{\mathrm{DR}}\, x &= \left\{ \frac{1}{2}\left(R_A v + x\right) \,\middle|\, v \in R_B x \right\} \\
&= \left\{ \frac{1}{2}\left(R_A(2z - x) + x\right) \,\middle|\, z \in P_B x \right\}, \\
&= \left\{ \frac{1}{2}\left(2P_A(2z - x) - (2z - x) + x\right) \,\middle|\, z \in P_B x \right\} \\
&= \left\{ P_A(2z - x) - z + x \mid z \in P_B x \right\}.
\end{aligned}$$

□

1.5.1. Convergence of the Douglas–Rachford Algorithm: History and known Results

Remark 1.14 (Some convergence results on the Douglas–Rachford Algorithm)*: Due to its success in solving non-convex feasibility problems there has been an increased interest in Douglas–Rachford type methods in the last decade. Some of the most interesting results in the literature are:*

- *(Douglas and Rachford, 1956): Introduction of the original operator splitting scheme for partial differential equations.*
- *(Lions and Mercier, 1979, Consequence of Corollary 1): Let* $\mathbf{A}, \mathbf{B}$ *be closed and convex subsets of* $\mathcal{H}$ *and let* $\mathbf{A} \cap \mathbf{B} \neq \emptyset$*. Then for* $x_0 \in \mathcal{H}$ *the sequence* $x_{n+1} = \mathrm{T}_{DR}\, x_n$ *converges weakly to* $\bar{x} \in \mathrm{Fix}\,(\mathrm{T}_{DR})$.
- *(Bauschke et al., 2004): Characterization of fixed point set of* T_{DR} *and weak convergence of the* shadow sequence *(compare* (2.20) *) for convex subsets of* $\mathcal{H}$*. The weak convergence result covers the case* $\mathbf{A} \cap \mathbf{B} = \emptyset$.
- *(Eckstein and Svaiter, 2009): Generalization to a splitting scheme for* m *operators similar to* Pierra's formulation (1.6)*. Weak convergence of the iterates generated by* (1.21)*, provided* $\mathbf{A} \cap \mathbf{B} \neq \emptyset$.
- *(Borwein and Sims, 2011): Discussion of a two dimensional example, that is, the intersection of a sphere and a line as a model case.*
- *(Hesse and Luke, 2013): Local linear convergence on Euclidean spaces for a* super-regular *set* $\mathbf{A}$ *and a subspace* $\mathbf{B}$ *based on* uniform regularity.
- *(Phan, 2014): Local convergence on Euclidean spaces or two* super-regular *sets* $\mathbf{A}, \mathbf{B}$ *based on a variant of* uniform regularity.

$\diamondsuit$

1.5.2. Feasibility as a special case of Operator Splitting

To illustrate the connection between feasibility problems and operator splitting or more general the theory of monotone operators, we give a short introduction, focusing on the main connections between the fields. For the sake of simplicity in this introduction, we will focus on the special setting of *prox-regular* sets, which implies that locally the Projectors onto the sets are single valued and hence the different notions of the normal cones coincide and can be described by $N_\Omega\,(\bar{x}) := \mathrm{cone}(P_\Omega^{-1}\bar{x} - \bar{x})$ (compare Definition 3.1).

Good sources on a general and detailed theory are -among others- (Bauschke and Combettes, 2011) (Luke, 2008) and the references therein.

The idea of operator splitting is the following: Given two operators $T_1 : \mathcal{H} \rightrightarrows \mathcal{H}$ and $T_2 : \mathcal{H} \rightrightarrows \mathcal{H}$ one asks for a point $\bar{x}$ such that

$$0 \in T_1(\bar{x}) + T_2(\bar{x}). \tag{1.23}$$

For $x_0 \in \mathcal{H}$ the Douglas–Rachford algorithm is given by

$$x_{n+1} = J^{\lambda}_{T_1}(2J^{\lambda}_{T_2} - \mathrm{Id})x_n + (\mathrm{Id} - J^{\lambda}_{T_2})x_n, \tag{1.24}$$

where for an operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ and $\lambda > 0$

$$J^{\lambda}_{T} := (\mathrm{Id} + \lambda T)^{-1}. \tag{1.25}$$

is the *resolvent* of T. Note that the *feasibility problem* (1.4) can be equivalently restated as

$$\min_{x \in \mathcal{H}} \iota_A(x) + \iota_B(x), \tag{1.26}$$

where $\iota_\Omega : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is the indicator function of the set Ω, *i.e.*,

$$\iota_\Omega(x) := \begin{cases} 0 & \text{if} \quad x \in \Omega \\ \infty & \text{if} \quad x \notin \Omega \end{cases}. \tag{1.27}$$

A necessary condition for $\bar{x}$ to solve equation (1.26) –and hence equation (1.4)– is

$$0 \in \partial \iota_A(\bar{x}) + \partial \iota_B(\bar{x}), \tag{1.28}$$

where $\partial \iota_\Omega$ is the subdifferential of the indicator function. Note that for any $\lambda > 0$ the resolvent of the normal cone is exactly the projector onto the set Ω

$$J^{\lambda}_{\partial \iota_\Omega} = J^{\lambda}_{N_\Omega(\cdot)} = P_\Omega. \tag{1.29}$$

Remark 1.15*: DR is actually* engineered *to find a point that solves equation* (1.28)*,* i.e., *find $\bar{x}$ such that*

$$0 \in N_A(\bar{x}) + N_B(\bar{x}). \tag{1.30}$$

and not (1.4)*.*

We will later state conditions that characterize whether or not the solution sets of (1.4) *and* (1.30) *coincide.* ◊

1.6. Examples

Example 1.16*: The following easy examples will appear throughout this work and serve to illustrate the regularity concepts we introduce and the convergence behavior of the algorithms under consideration.*

(a) *Two lines in* $\mathbb{R}^2$*:*

$$\begin{aligned}\mathbf{A} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\right\} \subset \mathbb{R}^2 \\ \mathbf{B} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2\right\} \subset \mathbb{R}^2.\end{aligned}$$

We will see that for any $x_0 \in \mathbb{R}^2$ *Alternating Projections and Douglas–Rachford converge with a linear rate to the intersection.*

(b) *Two lines in* $\mathbb{R}^3$*:*

$$\begin{aligned}\mathbf{A} &= \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 = 0, x_3 = 0\right\} \subset \mathbb{R}^3 \\ \mathbf{B} &= \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 = x_2, x_3 = 0\right\} \subset \mathbb{R}^3.\end{aligned}$$

After the first iteration step Alternating Projections shows exactly the same convergence behavior as in the first example. Douglas–Rachford does not converge to $\{0\} = A \cap B$*. All iterates from starting points on the line* $\{t(0,0,1) \mid t \in \mathbb{R}\}$ *are fixed points of the Douglas Rachford operator. On the other hand, iterates from starting points in* $A + B$ *stay in* $A + B$*, and the case then reduces to example ((a)).*

(c) *A line and a ball intersecting in one point:*

$$\begin{aligned}\mathbf{A} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\right\} \subset \mathbb{R}^2 \\ \mathbf{B} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + (x_2 - 1)^2 \leq 1\right\}.\end{aligned}$$

Alternating Projection converges to the intersection, but not with a linear rate. Douglas–Rachford has fixed points that lie outside the intersection, namely

$$\operatorname{Fix}(\mathrm{T}_{DR}) = P_{\mathbf{B}}^{-1}(0) = \{0\} \times \mathbb{R}_+$$

(cf. Proposition 2.18).

(d) *A* cross *and a subspace in* $\mathbb{R}^2$*:*

$$\begin{aligned}\mathbf{A} &= \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R} \\ \mathbf{B} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2\right\}.\end{aligned}$$

This example relates to the problem of sparse-signal recovery*. Both AP and Douglas–Rachford converge globally to the intersection* $\{0\} = A \cap B$*, though* A *is nonconvex. The convergence of both methods is covered by the theory built up in this work (cf. Chapter 8).*

(e) *A circle and a line:*

$$\begin{aligned}\mathbf{A} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = \sqrt{2}/2\right\} \subset \mathbb{R}^2 \\ \mathbf{B} &= \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\right\}.\end{aligned}$$

This example is of our particular interest, since it is a simple model case of the phase retrieval problem. *Until the publication of (Hesse and Luke, 2013) the only* direct *nonconvex convergence results for Douglas–Rachford were related to this model case, see (Artacho and Borwein, 2013; Borwein and Sims, 2011). Local convergence of Alternating Projections is covered by (Lewis et al., 2009; Bauschke et al., 2013c) as well as by the results in this work.*

△

2. Classical Convex Results

2.1. Best Approximation

Definition 2.1 (convex sets)**:**
A subset $C \subset \mathcal{H}$ is called *convex* if

$$\lambda x + (1-\lambda)y \in C, \quad \text{for all } x, y \in C,\ \lambda \in [0,1]. \tag{2.1}$$

Theorem 2.2 (uniqueness of best approximations, (Deutsch, 2001, Theorem 2.4))**:**
Let $C \subset \mathcal{H}$ be convex. Each $x \in \mathcal{H}$ has at most one best approximation in C. In particular, every proximinal convex set is Chebyshev.

Proof. For any $x \in \mathcal{H}$ suppose $y_1, y_2 \in P_C(x)$. By convexity $\frac{1}{2}y_1 + \frac{1}{2}y_2 \in C$ and

$$d(x, C) \le \left\| x - \frac{1}{2}(y_1 + y_1) \right\| \le \frac{1}{2}\|x - y_1\| + \frac{1}{2}\|x - y_2\| = d(x, C)$$

which implies $y_1 = y_2$. □

Theorem 2.3 ((Aronszajn, 1950), (Cheney and Goldstein, 1959))**:**
Let $C \subset \mathcal{H}$ be nonempty and convex, $x \in \mathcal{H}$ and $\bar{x} \in C$. Then $\bar{x}$ is the best approximation point $\bar{x} = P_C(x)$ if and only if

$$\langle x - \bar{x}, y - \bar{x} \rangle \le 0 \quad \text{for all } y \in C. \tag{2.2}$$

Proof. Let $\bar{x} = P_C(x)$ and assume $\langle x - \bar{x}, y - \bar{x} \rangle > 0$ for some $y \in C$. For $\lambda \in (0,1)$ define $y_\lambda := \lambda y + (1-\lambda)\bar{x}$ and note that by convexity of C $y_\lambda \in C$. For sufficiently small λ one achieves

$$\begin{aligned} \|x - y_\lambda\|^2 &= \|x - \lambda y - (1-\lambda)\bar{x}\|^2 \\ &= \|x - \bar{x}\|^2 - \lambda \Bigg(\underbrace{2\langle x - \bar{x}, y - \bar{x} \rangle - \lambda \|y - \bar{x}\|^2}_{>0} \Bigg). \end{aligned}$$

This implies $\|x - y_\lambda\| < \|x - \bar{x}\|$ and therefore contradicts the assumption that $\bar{x} = P_C(x)$.

Now let $\langle x-\bar{x}, y-\bar{x}\rangle \leq 0$ for all $y \in C$. Using this and the *Cauchy-Schwarz* inequality one achieves

$$\begin{aligned}\|x-\bar{x}\|^2 &= \langle x-\bar{x}, x-\bar{x}\rangle \\ &= \langle x-\bar{x}, x-y\rangle + \langle x-\bar{x}, y-\bar{x}\rangle \\ &\leq \langle x-\bar{x}, x-y\rangle \\ &\leq \|x-\bar{x}\|\,\|x-y\|\,.\end{aligned}$$

Therefore one can conclude $\|x-\bar{x}\| \leq \|x-y\|$ for all $y \in C$, which yields $\bar{x} = P_C(x)$. ☐

Theorem 2.4 ((Bauschke and Combettes, 2011, Proposition 6.27 and Theorem 6.29), (Moreau, 1962))**:**
Let $\mathbf{K}$ be a closed convex cone. For all $x \in \mathcal{H}$ the following properties hold

$$\langle x - P_{\mathbf{K}}x, P_{\mathbf{K}}x\rangle = 0, \tag{2.3}$$

$$\langle x - P_{\mathbf{K}}x, y\rangle \leq 0 \quad \text{for all } y \in \mathbf{K}, \tag{2.4}$$

$$\|x-P_{\mathbf{K}}x\|^2 + \|P_{\mathbf{K}}x\|^2 = \|x\|^2\,. \tag{2.5}$$

Proof. The first equation follows by the best approximation property (2.2) by choosing $y = 2P_{\mathbf{K}}x$ and $y = 0$ and combining the resulting inequalities. The inequality (2.4) then is a direct consequence of the first equation. Equation (2.5) follows by expanding

$$\|x\|^2 = \|x-P_{\mathbf{K}}x\|^2 + \|P_{\mathbf{K}}x\|^2 + 2\langle x-P_{\mathbf{K}}x, P_{\mathbf{K}}x\rangle \tag{2.6}$$

and the use of (2.3). ☐

Corollary 2.5 ((Bauschke and Combettes, 2011, Corollary 3.20))**:**
Let $\mathbf{L} \subset \mathcal{H}$ be an affine subspace, $x \in \mathcal{H}$ and $\bar{x} \in \mathbf{L}$. The following hold:

(a) $\bar{x}$ is the best approximation point $\bar{x} = P_{\mathbf{L}}(x)$ if and only if

$$\langle x-\bar{x}, y-z\rangle = 0 \quad \text{for all } y, z \in \mathbf{L}. \tag{2.7}$$

(b) For all $x, y \in \mathcal{H}$, $\lambda \in \mathbb{R}$:

$$P_{\mathbf{L}}\left(\lambda x + (1-\lambda)y\right) = \lambda P_{\mathbf{L}}x + (1-\lambda)P_{\mathbf{L}}y. \tag{2.8}$$

Proof. (a) By the best approximation property (2.2) one has

$$\langle x-\bar{x}, y-\bar{x}\rangle \leq 0 \quad \text{for all } y \in \mathbf{L}.$$

Since $\mathbf{L}$ is an affine subspace for any $y \in \mathbf{L}$ one has $\tilde{y} := 2\bar{x} - y \in \mathbf{L}$. Inserting this in the last inequality achieves $\langle x-\bar{x}, y-\bar{x}\rangle = 0$. Likewise one has for any $z \in \mathbf{L}$ $\langle x-\bar{x}, z-\bar{x}\rangle = 0$ and hence

$$\langle x-\bar{x}, y-z\rangle = 0, \quad \text{for all } y, z \in \mathbf{L}.$$

(b) Since $\mathbf{L}$ is an affine subspace for any $x_1, x_2 \in \mathcal{H}$, $\lambda \in \mathbb{R}$ the point $\bar{x}_\lambda := \lambda P_{\mathbf{L}} x_1 + (1-\lambda) P_{\mathbf{L}} x_2$ is an element of $\mathbf{L}$. Now by (2.7) for any $y, z \in \mathbf{L}$

$$\begin{aligned}\langle \lambda x_1 + (1-\lambda)x_2 - \bar{x}_\lambda, y - z\rangle =& \langle \lambda\left(x_1 - P_{\mathbf{L}}x_1\right) + (1-\lambda)\left(x_2 - P_{\mathbf{L}}x_2\right), y - z\rangle \\ =& \lambda \underbrace{\langle x_1 - P_{\mathbf{L}}x_1, y - z\rangle}_{=0} + (1-\lambda)\underbrace{\langle x_2 - P_{\mathbf{L}}x_2, y - z\rangle}_{=0}.\end{aligned}$$

Hence by (2.7) $\bar{x}_\lambda$ is best approximation to $\lambda x_1 + (1-\lambda)x_2$. □

Proposition 2.6:
Let $\Omega \subset \mathcal{H}$ be closed and nonempty. Let $\mathbf{L}$ be an affine subspace such that $\Omega \subseteq \mathbf{L}$. Then

$$P_\Omega P_{\mathbf{L}} = P_\Omega = P_{\mathbf{L}} P_\Omega \tag{2.9}$$

$$R_\Omega P_{\mathbf{L}} = P_{\mathbf{L}} R_\Omega \tag{2.10}$$

Proof. (2.9) follows by (Bauschke et al., 2013d, Lemma 3.3).

To show (2.10) note that then

$$P_{\mathbf{L}} R_\Omega = P_{\mathbf{L}}\left(2P_\Omega - \mathrm{Id}\right) \overset{(2.8)}{=} 2P_{\mathbf{L}}P_\Omega - P_{\mathbf{L}} \overset{(2.9)}{=} 2P_\Omega P_{\mathbf{L}} - P_{\mathbf{L}} = R_\Omega P_{\mathbf{L}}.$$

□

Remark 2.7: *Equation* (2.9) *appeared in (Bauschke et al., 2013d, Lemma 3.3). Equation* (2.10) *is discussed for two linear subspaces in (Hesse et al., 2014a, Lemma 4.4 and Proposition 4.5) and in a general version in(Phan, 2014, Lemma 2.5).* ◊

2.2. Nonexpansiveness and Firm Nonexpansiveness of Operators

Definition 2.8:
Let $\mathbf{D} \subset \mathcal{H}$ be nonempty.

$T : \mathbf{D} \to \mathcal{H}$ is called *nonexpansive* if

$$\|Tx - Ty\| \leq \|x - y\| \tag{2.11}$$

holds for all $x, y \in \mathbf{D}$.

$T : \mathbf{D} \to \mathcal{H}$ is called *firmly nonexpansive* if

$$\|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leq \|x - y\|^2 \tag{2.12}$$

holds for all $x, y \in \mathbf{D}$.

Lemma 2.9 ((Bauschke and Combettes, 2011, Proposition 4.2))**:**
Let $\mathbf{D} \subset \mathcal{H}$ be nonempty and let $T : \mathbf{D} \to \mathcal{H}$. The following are equivalent

(i) T is firmly nonexpansive on $\mathbf{D}$.

(ii) T is 1/2-averaged, *i.e.,* $T = \frac{1}{2}\left(\operatorname{Id} + \hat{T}\right)$ and the mapping $\hat{T} : \mathbf{D} \to \mathcal{H}$, $x \mapsto (2T - \operatorname{Id})x$ is nonexpansive on $\mathbf{D}$.

(iii) $\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle$ for all $x, y \in \mathbf{D}$.

Proof. To show that (ii) is equivalent to (iii) one observes

$$\begin{aligned} & \|(2T - \operatorname{Id})\, x - (2T - \operatorname{Id})\, y\|^2 \\ = \; & 4\,\|Tx - Ty\|^2 - 4\langle Tx - Ty, x - y\rangle + \|x - y\|^2 . \end{aligned}$$

The definition of nonexpansiveness

$$\|(2T - \operatorname{Id})\, x - (2T - \operatorname{Id})\, y\|^2 \leq \|x - y\|^2 ,$$

holds if and only if

$$\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle.$$

To see the equivalence of (i) and (iii) write

$$\|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2 = \|Tx - Ty\| - 2\langle Tx - Ty, x - y\rangle + \|x - y\|^2$$

and insert this in (i) (see equation (2.12)) to get

$$2\,\|Tx - Ty\|^2 - 2\langle Tx - Ty, x - y\rangle + \|x - y\|^2 \leq \|x - y\|^2 .$$

This then holds if and only if (iii) holds. □

Remark 2.10*:* Firm nonexpansiveness *of mappings is a property closely related to the* best approximation *property* (2.2)*, as for instance Theorem 2.11 will indicate. In the literature firm nonexpansiveness is often defined by one of the characterization in Lemma 2.9. The term* pseudocontractive *also appears in the literature – compare (Eckstein, 1989, p. 43) or (Reinermann and Schöneberg, 1967) – to describe equation* (2.12)*. For a detailed study of firmly nonexpansive mappings see (Zarantonello, 1971, Section 1), (Goebel and Reich, 1984; Goebel and Kirk, 1990). A detailed modern treatment of firmly nonexpansive mappings can be found (Bauschke and Combettes, 2011, Chapter 4).* ◊

2.3. Basic Properties of Projectors and Reflectors onto Convex Sets

Theorem 2.11 ((Nashed, 1968, Equation (10)))**:**
Let C be a closed, nonempty and convex set. The projector $P_C : \mathcal{H} \to C$ is firmly nonexpansive.

Proof. We use the best approximation property (2.2) to achieve

$$\begin{aligned} & \langle P_C x - P_C y, x - y\rangle \\ = \; & \underbrace{\langle P_C x - P_C y, x - P_C x\rangle}_{\geq 0} + \underbrace{\langle P_C x - P_C y, P_C y - y\rangle}_{\geq 0} + \langle P_C x - P_C y, P_C x - P_C y\rangle \\ \geq \; & \|P_C x - P_C y\|^2 . \end{aligned}$$

□

Lemma 2.12:
Let C be nonempty, closed and convex. The reflector $R_C : \mathcal{H} \to \mathcal{H}$ is nonexpansive.

Proof. By firm nonexpansiveness of the projector P_C, or more precise Lemma 2.9 (iii), one gets $\|P_C x - P_C y\|^2 \leq \langle P_C x - P_C y, x - y\rangle$, which then yields

$$\begin{aligned} \|R_C x - R_C y\|^2 &= \|2P_C x - 2P_C y - (x - y)\|^2 \\ &= 4\|P_C x - P_C y\|^2 - 4\langle P_C x - P_C y, x - y\rangle + \|x - y\|^2 \\ &\leq \|x - y\|^2 . \end{aligned}$$

□

Corollary 2.13 (Projectors and reflectors on subspaces)**:**
Let $\mathbf{L}$ be an affine subspace. The following conditions hold

(a) $P_{\mathbf{L}}$ is firmly nonexpansive with equality, *i.e.,*

$$\|P_{\mathbf{L}} x - P_{\mathbf{L}} y\|^2 + \|(\operatorname{Id} - P_{\mathbf{L}})x - (\operatorname{Id} - P_{\mathbf{L}})y\|^2 = \|x - y\|^2 \tag{2.13}$$

for all $x \in \mathcal{H}$.

(b) For all $x \in \mathcal{H}$, $y \in \mathbf{L}$ the following identity holds

$$\|R_{\mathbf{L}} x - y\| = \|x - y\| . \tag{2.14}$$

Remark 2.14*: Corollary 2.13 (a) is actually a restatement of the* Moreau decomposition (2.5) *on subspaces.* ◊

Proof. (a) is achieved by replacing (2.2) by the stronger equation (2.7) analog to Theorem 2.11.

(b) follows by the use of equation (2.7):

$$\begin{aligned}
\|R_{\mathbf{L}}x - y\|^2 &= \|2P_{\mathbf{L}}x - 2x + x - y\|^2 \\
&= \|x - y\|^2 + 4\|P_{\mathbf{L}}x - x\|^2 + 4\langle P_{\mathbf{L}}x - x, x - y\rangle \\
&= \|x - y\|^2 + 4\|P_{\mathbf{L}}x - x\|^2 \\
&\quad +4\Big(\underbrace{\langle P_{\mathbf{L}}x - x, x - P_{\mathbf{L}}x\rangle}_{=-\|P_{\mathbf{L}}x - x\|^2} + \underbrace{\langle P_{\mathbf{L}}x - x, P_{\mathbf{L}}x - y\rangle}_{=0}\Big) \\
&= \|x - y\|^2 .
\end{aligned}$$

□

2.4. Basic Properties of the Douglas–Rachford Operator

Theorem 2.15 ((Lions and Mercier, 1979, Proposition 2))**:**
Let $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ be closed, convex and nonempty. The Douglas–Rachford operator T_{DR} (1.20) is firmly nonexpansive.

Proof. By Lemma 2.12 the reflectors $R_{\mathbf{A}}$ and $R_{\mathbf{B}}$ are nonexpansive. For $x \in \mathcal{H}$ define $\hat{T}x := R_{\mathbf{A}}R_{\mathbf{B}}x$ and note that $\hat{T}$ as a composition of nonexpansive mappings is nonexpansive. Firm nonexpansiveness of T_{DR} is then a consequence of Theorem 2.9 (ii). □

Definition 2.16 (The gap vector, (Bauschke and Borwein, 1993))**:**
Let $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ be nonempty, closed and convex. Let

$$v := P_{\mathrm{cl}(\mathbf{B}-\mathbf{A})}(0). \tag{2.15}$$

We call v the *displacement vector* or the *gap* vector between the sets $\mathbf{A}$ and $\mathbf{B}$.

Proposition 2.17 (Fixed points of Alternating Projections, (Bauschke and Borwein, 1993, Section 5))**:**
Assume that $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ are closed and nonempty. Then

$$\mathrm{Fix}\,(\mathrm{T}_{\mathrm{AP}}) = \mathbf{A} \cap (\mathbf{B} - v), \tag{2.16}$$

where v is the displacement vector given by (2.15). Furthermore assume that there is $\hat{x} \in \mathbf{A} \cap \mathbf{B}$. Then $\mathrm{Fix}\,(\mathrm{T}_{\mathrm{AP}}) = \mathbf{A} \cap \mathbf{B}$.

Proposition 2.18 (Fixed points of Douglas–Rachford, (Bauschke et al., 2004))**:**
Assume that $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ are nonempty, closed and convex. Let v be the displacement vector given by (2.15) and for any closed, convex set Ω let $N_\Omega(\bar{x}) := \text{cone}(P_\Omega^{-1}\bar{x} - \bar{x})$ be the normal cone to Ω at $\bar{x}$ (compare Definition 3.1). The following properties hold:

(a) Let $x_0 \in \mathcal{H}$ and x_n be the sequence generated by (1.21), *i.e.,* $x_{n+1} = \mathrm{T}_{\mathrm{DR}}\, x_n$. Then

$$\mathrm{T}_{\mathrm{DR}}\, x_n - x_n \to -v, \quad n \to \infty. \tag{2.17}$$

(b)

$$\begin{aligned} (\mathbf{A} + v) \cap \mathbf{B} + N_{\mathrm{cl}(\mathbf{B}-\mathbf{A})}(v) \quad &\subset \quad \mathrm{Fix}\,(\mathrm{T}_{\mathrm{DR}} + v) \\ \mathrm{Fix}\,(\mathrm{T}_{\mathrm{DR}} + v) \quad &\subset \quad (\mathbf{A} + v) \cap \mathbf{B} + N_{\mathrm{cl}(\mathbf{B}-\mathbf{A})}(v) + v. \end{aligned} \tag{2.18}$$

(c) If $\mathbf{A} \cap \mathbf{B} \neq \emptyset$, then

$$\mathrm{Fix}\,(\mathrm{T}_{\mathrm{DR}}) = \mathbf{A} \cap \mathbf{B} + N_{\mathrm{cl}(\mathbf{B}-\mathbf{A})}(0)\,. \tag{2.19}$$

Proof. For (a) see (Bauschke et al., 2004, Fact 3.2, Theorem 3.4). (b) is (Bauschke et al., 2004, Theorem 3.5), whilst (a) is (Bauschke et al., 2004, Corollary 3.5). □

Remark 2.19*: A nonconvex analog to Proposition 2.18 can be found in (Luke, 2008, Lemma 3.8).*

Proposition 2.18 indicates why the Douglas–Rachford algorithm is notoriously difficult to analyze. If $\mathbf{A} \cap \mathbf{B} = \emptyset$ *the algorithm does not converge at all, and even if there are points* $\hat{x} \in \mathbf{A} \cap \mathbf{B}$ *the set* $\mathrm{Fix}\,(\mathrm{T}_{DR})$ *does not necessarily coincide with the intersection. This was already pointed out in remark 1.15. We will characterize conditions that guarantee* $\mathrm{Fix}\,(\mathrm{T}_{DR}) = \mathbf{A} \cap \mathbf{B}$ *in Chapter 5.*

Proposition 2.18 also suggests that it may be reasonable for x_n *generated by the Douglas–Rachford algorithm* (1.21) *to monitor the* shadow sequence $P_B x_n$ *rather than the sequence* x_n*. See for instance (Bauschke et al., 2004, Remark 3.10).* ◊

Definition 2.20 (shadow sequence)**:**
For $x_0 \in \mathcal{H}$ $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ closed let x_n be a sequence generated by the Douglas–Rachford operator, *i.e.,* a sequence according to (1.21). The *shadow sequence* of x_n is defined by

$$P_{\mathbf{B}}(x_n) = P_{\mathbf{B}}\left((\mathrm{T}_{\mathrm{DR}}\, x_0)^n\right), \quad \text{for } n \in \mathbb{N}. \tag{2.20}$$

2.5. On the Douglas–Rachford Operator on Parallel Subspaces

Theorem 2.21:
Let $\Omega, \mathbf{A}, \mathbf{B}$ be closed, nonempty subsets of $\mathcal{H}$ and let T_{DR} be the Douglas–Rachford

operator defined by (1.20) and let $\mathbf{L}$ be an affine subspace such that $\mathbf{A}\cap\mathbf{B} \subseteq \operatorname{aff}(\mathbf{A}\cup\mathbf{B}) \subseteq \mathbf{L}$. Then

$$P_{\mathbf{L}}\,\mathrm{T}_{\mathrm{DR}} = \mathrm{T}_{\mathrm{DR}}\,P_{\mathbf{L}}. \tag{2.21}$$

Furthermore for any $\tilde{x} \in \mathbf{L}$ one has $\mathrm{T}_{\mathrm{DR}}\,\tilde{x} \subset \mathbf{L}$.

Proof. Since $\mathbf{A}, \mathbf{B} \subset \mathbf{L}$ (2.21) follows by applying (2.10)

$$\begin{aligned} P_{\mathbf{L}}\,\mathrm{T}_{\mathrm{DR}} &= P_{\mathbf{L}}\frac{1}{2}\left(\mathrm{Id} + R_{\mathbf{A}}R_{\mathbf{B}}\right) \\ &= \frac{1}{2}\left(P_{\mathbf{L}} + P_{\mathbf{L}}R_{\mathbf{A}}R_{\mathbf{B}}\right) \\ &\overset{(2.10)}{=} \frac{1}{2}\left(P_{\mathbf{L}} + R_{\mathbf{A}}R_{\mathbf{B}}P_{\mathbf{L}}\right) \\ &= \mathrm{T}_{\mathrm{DR}}\,P_{\mathbf{L}}. \end{aligned}$$

It is then a direct consequence of (2.21) that if $x \in \mathbf{L}$ then $P_{\mathbf{L}}\,\mathrm{T}_{\mathrm{DR}}\,x = \mathrm{T}_{\mathrm{DR}}\,x$ and hence $\mathrm{T}_{\mathrm{DR}}\,x \subset \mathbf{L}$. □

Remark 2.22*: A similar result to equation* (2.21) *is discussed in (Hesse et al., 2014a, Lemma 4.4 and Proposition 4.5) and (Phan, 2014, Theorem 3.14), where the latter provides a more general discussion for the Douglas–Rachford operator on parallel subspaces and leads to the following interesting result.* ◊

Proposition 2.23 ((Phan, 2014, Theorem 3.16))**:**
Let $\mathbf{A}$ and $\mathbf{B}$ be closed and nonempty and let $\hat{x} \in \mathbf{A}\cap\mathbf{B} \neq \emptyset$, $\mathbf{L} := \operatorname{aff}(\mathbf{A}\cup\mathbf{B})$. For $x_0 \in \mathcal{H}$ let x_n be a Douglas–Rachford sequence generated by (1.21), *i.e.,*

$$x_{n+1} \in \mathrm{T}_{\mathrm{DR}}\,x_n, \quad n \in \mathbb{N}$$

Define $\tilde{x}_n := P_{\mathbf{L}}x_n$, for $n \in \mathbb{N}$. Then

(a) For all $n \in \mathbb{N}$
$$\tilde{x}_n \in \mathrm{T}_{\mathrm{DR}}\,\tilde{x}_{n-1}. \tag{2.22}$$

(b) For all $n \in \mathbb{N}$
$$\tilde{x}_n - x_n = \tilde{x}_0 - x_0. \tag{2.23}$$

(c) If $y_n \to \bar{y} \in \mathbf{A}\cap\mathbf{B}$ for $n \to \infty$ then $x_n \to \bar{x} \in \operatorname{Fix}(\mathrm{T}_{\mathrm{DR}})$.

3. Set Regularity

3.1. Foundations – Normal and Tangent Cones

From now on, if not stated otherwise, $\mathbb{E}$ is a Euclidean space. Ω_1, Ω_2 are closed and nonempty subsets of $\mathbb{E}$.

Most of the following definitions can be found in (Rockafellar and Wets, 1998) in more detail.

Definition 3.1 (normal cones, (Rockafellar and Wets, 1998, Definition 6.3 and Example 6.16))**:**
The *proximal normal cone* $N^{\mathrm{P}}_{\Omega}(\bar{x})$, the *Fréchet normal cone* $\widehat{N}_{\Omega}(\bar{x})$ and the *limiting normal cone* $N_{\Omega}(\bar{x})$ to a set $\Omega \subset \mathbb{E}$ at a point $\bar{x} \in \Omega$ are defined by

$$
\begin{aligned}
N^{\mathrm{P}}_{\Omega}(\bar{x}) &:= \operatorname{cone}(P_{\Omega}^{-1}(\bar{x}) - \bar{x}), && (3.1)\\
\widehat{N}_{\Omega}(\bar{x}) &:= \left\{ v \in \mathbb{E} \;\middle|\; \limsup_{x \xrightarrow{\Omega} \bar{x}} \frac{\langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|} \leq 0 \right\}, && (3.2)\\
N_{\Omega}(\bar{x}) &:= \limsup_{x \xrightarrow{\Omega} \bar{x}} \widehat{N}_{\Omega}(x). && (3.3)
\end{aligned}
$$

Remark 3.2*: The construction of the* limiting normal cone *goes back to Mordukhovich (see (Rockafellar and Wets, 1998, Chap. 6 Commentary)). Definition* (3.3) *is the most conventional definition of the limiting normal cone. However, for our purposes here, the following equivalent definition is more appropriate.* ◊

Proposition 3.3 (Mordukhovich normal cone (Mordukhovich, 2006, Theorem 1.6))**:**
The *limiting normal cone* or *Mordukhovich normal cone* is the smallest cone satisfying the two properties

(a) $P_{\Omega}^{-1}(\bar{x}) \subseteq (\mathrm{Id} + N_{\Omega}(\cdot))(\bar{x})$ and in particular $\bar{x} \in P_{\Omega}(x) \Rightarrow x - \bar{x} \in N_{\Omega}(\bar{x})$,

(b) for any sequence $x_i \to \bar{x}$ in Ω any limit of a sequence of normals $\nu_i \in N_{\Omega}(x_i)$ must lie in $N_{\Omega}(\bar{x})$.

In fact the *limiting normal cone* $N_{\Omega}(\bar{x})$ to a set $\Omega \subset \mathbb{E}$ at a point $\bar{x} \in \Omega$ can equivalently be defined as any vector that can be written as the limit of proximal normals; that is,

$\bar{\nu} \in N_\Omega(\bar{x})$ if and only if there exist sequences $(x_k)_{k\in\mathbb{N}}$ in Ω and $(\nu_k)_{k\in\mathbb{N}}$ in $N^{\mathrm{P}}_\Omega(x_k)$ such that $x_k \to \bar{x}$ and $\nu_k \to \bar{\nu}$.

Remark 3.4*: As a consequence of the previous Proposition*

$$N_\Omega(\bar{x}) = \limsup_{x \overset{\Omega}{\to} \bar{x}} N^{\mathrm{P}}_\Omega(x).$$

This only holds on finite dimensional spaces $\mathbb{E}$ *and is not true on general Hilbert spaces* $\mathcal{H}$ *and in fact is one of the main reasons, most of the nonconvex convergence theory for projection algorithms is formulated on Euclidean spaces* $\mathbb{E}$. ♢

Proposition 3.5 ((Rockafellar and Wets, 1998, Proposition 6.5))**:**
At any point $\bar{x} \in \Omega$ the sets $N^{\mathrm{P}}_\Omega(\bar{x})$, $\widehat{N}_\Omega(\bar{x})$ and $N_\Omega(\bar{x})$ are closed cones. In addition the Fréchet normal cone $\widehat{N}_\Omega(\bar{x})$ is always convex, and

$$\widehat{N}_\Omega(\bar{x}) \subset N_\Omega(\bar{x}) \quad = \quad \limsup_{x \overset{\Omega}{\to} \bar{x}} \widehat{N}_\Omega(x).$$

Definition 3.6 (tangent cones, (Rockafellar and Wets, 1998, Definition 6.1 and 6.25))**:**
A vector $w \in \mathbb{E}$ is *tangent* to a set $\Omega \subset \mathbb{E}$ at a point $\bar{x}$, if there are sequences $(\bar{x}_n)_{n\in\mathbb{N}} \subset \Omega$, $\bar{x}_n \overset{n\to\infty}{\to} \bar{x}$ and $\lambda_n > 0$, $\lambda_n \overset{n\to\infty}{\to} 0$, such that

$$\frac{\bar{x}_n - \bar{x}}{\lambda_n} \overset{n\to\infty}{\to} w. \tag{3.4}$$

The set of all tangent vectors of Ω at $\bar{x}$ is called *tangent cone* $T_\Omega(\bar{x})$.

A vector $\hat{w} \in \mathbb{E}$ is a *regular tangent* vector to $\Omega \subset \mathbb{E}$ at $\bar{x}$, if for any sequence $(\bar{x}_n)_{n\in\mathbb{N}} \subset \Omega$, $\bar{x}_n \overset{n\to\infty}{\to} \bar{x}$ and $\lambda_n > 0$, $\lambda_n \overset{n\to\infty}{\to} 0$, there exists a sequence $(x_n)_{n\in\mathbb{N}} \subset \Omega$, $x_n \overset{n\to\infty}{\to} \bar{x}$, such that

$$\frac{x_n - \bar{x}}{\lambda_n} \overset{n\to\infty}{\to} \hat{w}. \tag{3.5}$$

The set of all regular tangent vectors of Ω at $\bar{x}$ is called *regular tangent cone* $\widehat{T}_\Omega(\bar{x})$.

Proposition 3.7 ((Rockafellar and Wets, 1998, Theorem 6.26))**:**
Let $\Omega \subset \mathbb{E}$ and $\bar{x} \in \Omega$. Both $T_\Omega(\bar{x})$ and $\widehat{T}_\Omega(\bar{x})$ are closed cones. In addition $\widehat{T}_\Omega(\bar{x}) \subset T_\Omega(\bar{x})$ and $\widehat{T}_\Omega(\bar{x})$ is always convex. If Ω is closed the following relation holds

$$\widehat{T}_\Omega(\bar{x}) = \liminf_{x \overset{\Omega}{\to} \bar{x}} T_\Omega(x).$$

Definition 3.8 ((Rockafellar and Wets, 1998, Corollary 6.21))**:**
For any cone $K \subset \mathbb{E}$ the *polar* of K is defined to be the cone

$$K^\ominus := \{y \in \mathbb{E} \mid \langle x, y\rangle \le 0 \text{ for all } x \in K\}. \tag{3.6}$$

$K^\ominus$ is closed and convex. The *bipolar* is the cone $K^{\ominus\ominus} := (K^\ominus)^\ominus$. Note that $K^{\ominus\ominus} = \mathrm{cl}\,(\mathrm{conv})\,K$ and that for two cones $K_1, K_2 \subset \mathbb{E}$, $K_1 \subset K_2$ implies $K_1^\ominus \supset K_2^\ominus$.

Theorem 3.9 (tangent-normal polarity, (Rockafellar and Wets, 1998, Theorem 6.28))**:**
Let $\Omega \subset \mathbb{E}$ be closed and $\bar{x} \in \Omega$. The following properties hold

$$\begin{aligned} \widehat{N}_\Omega(\bar{x}) &= T_\Omega(\bar{x})^\ominus, \\ \widehat{T}_\Omega(\bar{x}) &= N_\Omega(\bar{x})^\ominus. \end{aligned}$$

3.2. Nonconvex Notions of Regularity

Definition 3.10 (Clarke-regularity)**:**
A nonempty set $\Omega \subset \mathbb{E}$ is *Clarke-regular* at a point $\bar{x} \in \Omega$ if

$$N_\Omega(\bar{x}) = \widehat{N}_\Omega(\bar{x}). \tag{3.7}$$

Definition 3.11 (Prox-regularity, (Poliquin et al., 2000, Definition 1.1))**:**
A nonempty, closed set $\Omega \subset \mathbb{E}$ is *prox-regular* at a point $\bar{x} \in \Omega$ for $\bar{\nu} \in N_\Omega(\bar{x})$, if there is $\delta > 0$ and $\rho > 0$, such that for all $x \in \Omega \cap \mathbb{B}_\delta(\bar{x})$, $\nu \in N_\Omega(x) \cap \mathbb{B}_\delta(\bar{\nu})$

$$x = P_{\Omega \cap \mathbb{B}_\delta(\bar{x})}\left(x + \rho^{-1}\nu\right). \tag{3.8}$$

The set Ω is simply *prox-regular* at $\bar{x}$ if (3.8) holds for all $\bar{\nu} \in N_\Omega(\bar{x})$.

Remark 3.12*: Early results relating to the notion of* prox-regularity *were already introduced in (Federer, 1959). We stick to the notation of (Poliquin et al., 2000) as the following result is one of the basic properties in our intended applications.* $\diamondsuit$

Theorem 3.13 (Prox-regularity, (Poliquin et al., 2000, Theorem 1.3))**:**
A nonempty, closed set $\Omega \subset \mathbb{E}$ is *prox-regular* at a point $\bar{x} \in \Omega$ if and only if the projector P_Ω is single-valued around $\bar{x}$.

Definition 3.14 (Super-regularity and its variants)**:**
A nonempty set $\Omega \subset \mathbb{E}$ is

(a) *super-regular* at a point $\bar{x} \in \Omega$ if, for all $\varepsilon > 0$, there exists a δ such that the inequality

$$\langle \nu_x, y - x \rangle \leq \varepsilon \|\nu_x\| \|y - x\| \tag{3.9}$$

holds for all points $y, x \in \mathbb{B}_\delta(\bar{x}) \cap \Omega$ and all vectors $\nu_x \in N^{\mathrm{P}}_\Omega(x)$.

(b) *(ε, δ)-regular* at $\bar{x}$, if there exists $\varepsilon > 0, \delta > 0$ such that (3.9) holds for all $y, x \in \mathbb{B}_\delta(\bar{x}) \cap \Omega$, $\nu_x \in N^{\mathrm{P}}_\Omega(x)$.

(c) *(ε, δ)-subregular* at $\bar{x}$ with respect to a set S, if there exists $\varepsilon > 0, \delta > 0$ such that (3.9) holds for all $x \in \mathbb{B}_\delta(\bar{x}) \cap \Omega$, $y \in S \cap \mathbb{B}_\delta(\bar{x})$, $\nu_x \in N^{\mathrm{P}}_\Omega(x)$. The set Ω is simply said to be (ε, δ)-subregular at $\bar{x}$ if $S = \{\bar{x}\}$.

If a set Ω is (ε, δ)-(sub)-regular at $\bar{x}$ for any $\delta > 0$, we call it (ε, ∞)-(sub)-regular at $\bar{x}$.

Remark 3.15*: Super-regularity was introduced in (Lewis et al., 2009, Definition 4.3). It was introduced as one of the fundamental tools to achieve local linear convergence of the alternating projections algorithm. (ε, δ)-regularity was introduced as a generalization of super-regularity in (Bauschke et al., 2013d, Definition 8.1). The notion of (ε, δ)-subregularity first appeared in (Hesse and Luke, 2013, Definition 2.9) and is weaker still than (ε, δ)-regularity.*

A similar condition to subregularity appears in the context of regularized inverse problems (Jin and Lorenz, 2010, Corollary 3.6). $\Diamond$

Theorem 3.16 (Prox-regularity implies super-regularity, (Lewis et al., 2009, Proposition 4.9))**:**
Let $\Omega \subset \mathbb{E}$ be closed and nonempty. If Ω is prox-regular at a $\bar{x} \in \Omega$, then Ω is super-regular at $\bar{x}$.

Theorem 3.17 (Super-regularity implies Clarke-regularity, (Lewis et al., 2009, Corollary 4.5))**:**
Let $\Omega \subset \mathbb{E}$ be closed and nonempty. If Ω is super-regular at a $\bar{x} \in \Omega$, then Ω is Clarke-regular at $\bar{x}$.

Proof. (Lewis et al., 2009, Corollary 4.5) . $\square$

Remark 3.18*: As mentioned before, we are interested in providing a reasonable qualitative and quantitative framework to prove convergence of projection algorithms.* Being able *to efficiently calculate* projectors *or more precisely* best approximation *points is usually closely related to prox-regularity or as we will see later at least $(\varepsilon, \delta)-$subregularity.* $\Diamond$

Super-regularity is something between Clarke regularity and amenability or prox-regularity. (ε, δ)-regularity is still weaker than Clarke regularity (and hence super-regularity) as the next example shows.

Remark 3.19*:* [Example 1.16 (d) revisited] The set

$$A := \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R} \tag{3.10}$$

is a particularly easy pathological set that illustrates the distinction between our new notion of subregularity and previous notions found in the literature. Note that for $x_1 \in \mathbb{R} \times \{0\}$, $N_A(x_1) = N_A^P(x_1) = \{0\} \times \mathbb{R}$ and for $x_2 \in \{0\} \times \mathbb{R}$, $N_A(x_2) = N_A^P(x_2) = \mathbb{R} \times \{0\}$ and that $N_A(0) = A$ and $N_A^P(0) = 0$ which implies that at the origin A is not Clarke regular and therefore neither super-regular nor prox-regular there. In fact, it is not even (ε, δ)-regular at the origin for any $\varepsilon < 1$ and any $\delta > 0$. The set A is, however, $(0, \infty)$-*sub*regular at $\{0\}$. Indeed, for any $x_1 \in \mathbb{R} \times \{0\}$ one has $\nu_1 \in N_A(x_1) = \{0\} \times \mathbb{R}$ and therefore $\langle \nu_1, x_1 - 0 \rangle = 0$. Analogously for $x_2 \in \{0\} \times \mathbb{R}$, $\nu_2 \in N_A(x_2) = \mathbb{R} \times \{0\}$ and it follows that $\langle \nu_2, x_2 - 0 \rangle = 0$, which shows that A is $(0, \infty)$-subregular at 0. $\Diamond$

Remark 3.20 ((ε, δ)-subregularity as a quantitative measure)*: Remark 3.19 illustrates that* (ε, δ)*-subregularity with respect to a point* $\bar{x}$ *in some sense forms a* quantitative measure *on the degree of* violation of convexity*. In terms of the point* 0 *of* A *as in the last example every convex combination of points* $x \in A$ *and* 0 *are again points of* A*. Indeed, in applications* (ε, δ)*-subregularity* with respect to the set of feasible points *leads to improved convergence results in many application. See for instance the application to* sparse affine feasibility *in Chapter 8.* ♢

Lemma 3.21 (Projectors under unitary transformations)**:**
Let $N \subset \mathbb{E}$ be closed and let $U : \mathbb{E} \to \mathbb{E}$ be an unitary linear operator. Then the projector P_M onto set M defined by

$$M := \{x \in \mathbb{E} \mid Ux \in N\} \tag{3.11}$$

is given by

$$P_M = U^* P_N U. \tag{3.12}$$

As a consequence the proximal normal cone fulfills

$$UN^{\mathrm{P}}_M (x) = N^{\mathrm{P}}_N (Ux). \tag{3.13}$$

for any $x \in M$.

Proof. First note that $N = U(M)$ and $M = U^*(N)$ and hence, as N is closed so is M. By Proposition 1.6 then $P_N y \neq \emptyset$ and $P_M x \neq \emptyset$ for any $x, y \in \mathbb{E}$.

Choose any $x \in \mathbb{E}$. Then

$$\bar{x} \in P_M x \quad \Leftrightarrow \quad \|x - \bar{x}\| = d(x, M).$$

On the other hand, since U is an isometry

$$d(x, M) = d(Ux, U(M)) = d(Ux, N),$$

and hence

$$\|Ux - U\bar{x}\| = d(Ux, N) \quad \Leftrightarrow \quad U\bar{x} \in P_N Ux$$

which is equivalent to $\bar{x} \in U^* P_N Ux$.

Equation (3.13) is then a consequence of the definition of the proximal normal cone (3.1). □

Theorem 3.22 (subregularity under unitary transformations)**:**
Let $U : \mathbb{E} \to \mathbb{E}$ be an unitary linear operator. Let $N \subset \mathbb{E}$ be (ε, δ)-subregular at $\bar{y}$ with respect to $S \subset N$. Then the set M defined by

$$M := \{x \in \mathbb{E} \mid Ux \in N\}. \tag{3.14}$$

is (ε, δ)-subregular at $\bar{x} := U^*\bar{y}$ with respect to $U^*S \subset M$.

Proof. Note since N is (ε, δ)-subregular at $\bar{y}$

$$\langle \nu_u, v - u \rangle \leq \varepsilon \left\| \nu_u \right\| \left\| v - u \right\| \tag{3.15}$$

holds for all points $u \in \mathbb{B}_\delta(\bar{y}) \cap N$, $v \in S$ and all vectors $\nu_u \in N_N^{\mathrm{P}}(u)$.

Now for $\bar{x} = U^*\bar{y}$ choose any $x \in M \cap \mathbb{B}_\delta(\bar{x})$, $y \in U^*S$ and $\nu_x \in N_M^{\mathrm{P}}(x)$. By definition of M then $Ux, Uy \in N$ and by Lemma 3.21 $UN_M^{\mathrm{P}}(x) = N_N^{\mathrm{P}}(Ux)$. (ε, δ)-subregularity of N then implies

$$\begin{aligned}\langle \nu_x, y - x \rangle =& \langle \underbrace{U\nu_x}_{\in N_N^{\mathrm{P}}(Ux)}, \underbrace{Uy}_{\in S} - \underbrace{Ux}_{\in N} \rangle \\ \overset{(3.15)}{\leq}& \varepsilon \left\| U\nu_x \right\| \left\| Uy - Ux \right\| \\ =& \varepsilon \left\| \nu_x \right\| \left\| y - x \right\|,\end{aligned}$$

which completes the proof of (ε, δ)-subregularity of M at $\bar{x}$. □

Corollary 3.23 (super-regularity under unitary transformations)**:**
Let $U : \mathbb{E} \to \mathbb{E}$ be an unitary linear operator. Let N be super-regular at $\bar{y}$. Then the set M defined by

$$M := \{x \in \mathbb{E} \mid Ux \in N\}. \tag{3.16}$$

is super-regular at $\bar{x} := U^*\bar{y}$.

4. $(\mathrm{S}, \varepsilon)$-Firm Nonexpansiveness

Up to this point, the results have concerned mostly convex sets, and hence the projector and related fixed point algorithms have all been single-valued. In what follows, we generalize to nonconvex sets and therefore allow multi-valuedness of the projectors.

4.1. Definition and Basic Properties

We define next an analog to firm nonexpansiveness in the nonconvex case with respect to a set $\mathbf{S}$.

Definition 4.1 (($\mathbf{S}, \varepsilon$)-(firmly-)nonexpansive mappings, (Hesse and Luke, 2013, Definition 2.3))**:**
Let $\mathbf{D}$ and $\mathbf{S}$ be nonempty subsets of $\mathbb{E}$.

$T : \mathbf{D} \rightrightarrows \mathbb{E}$ is called $(\mathbf{S}, \varepsilon)$-*nonexpansive on* $\mathbf{D}$ if

$$\begin{gathered} \forall x \in \mathbf{D},\ \forall \bar{x} \in \mathbf{S},\ \forall x_+ \in Tx,\ \forall \bar{x}_+ \in T\bar{x} : \\ \|x_+ - \bar{x}_+\| \leq \sqrt{1+\varepsilon}\, \|x - \bar{x}\| . \end{gathered} \tag{4.1}$$

If (4.1) holds with $\varepsilon = 0$ then we say that T is $\mathbf{S}$-*nonexpansive* on $\mathbf{D}$.

$T : \mathbf{D} \rightrightarrows \mathbb{E}$ is called $(\mathbf{S}, \varepsilon)$-*firmly nonexpansive* on $\mathbf{D}$ if

$$\begin{gathered} \forall x \in \mathbf{D},\ \forall \bar{x} \in \mathbf{S},\ \forall x_+ \in Tx,\ \forall \bar{x}_+ \in T\bar{x} : \\ \|x_+ - \bar{x}_+\|^2 + \|(x - x_+) - (\bar{x} - \bar{x}_+)\|^2 \leq (1+\varepsilon)\, \|x - \bar{x}\|^2 . \end{gathered} \tag{4.2}$$

If (4.2) holds with $\varepsilon = 0$ then we say that T is $\mathbf{S}$-*firmly nonexpansive* on $\mathbf{D}$.

Note that, as with (firmly) nonexpansive mappings, the mapping T need not be a self-mapping from $\mathbf{D}$ to itself. The classical (firmly) nonexpansive operator on $\mathbf{D}$ is $(\mathbf{D}, 0)$-(firmly) nonexpansive on $\mathbf{D}$.

Remark 4.2*: In the special case where* $\mathbf{S} = \operatorname{Fix}(T)$*, mappings satisfying* (4.1) *are also called* quasi-(firmly-)nonexpansive *(Bauschke and Combettes, 2011). Quasi-nonexpansiveness is a restriction of another well-known concept, Fejér monotonicity, to*

$\mathbf{S} = \mathrm{Fix}\,(T)$. *Equation* (4.2) *is a relaxed version of firm nonexpansiveness* (2.12). $(\mathbf{S},\varepsilon)$-*(firmly-)nonexpansive mappings were introduced (Hesse and Luke, 2013, Definition 2.3). A similar concept* γ-quasi firm nonexpansiveness*, which is motivated by* $(\mathbf{S},\varepsilon)$*-(firmly-)nonexpansiveness, was introduced (Phan, 2014, Definition 4).* ◊

Analogous to the relation between firmly nonexpansive mappings and 1/2-averaged mappings in Theorem 2.9 we have the following relationship between $(\mathbf{S},\varepsilon)$-firmly nonexpansive mappings (4.2) and their 1/2-averaged companion mapping. For a detailed discussion on general *averaged mappings* see (Bauschke and Combettes, 2011, Chapter 4) and references therein.

Lemma 4.3 (1/2−averaged mappings, (Hesse and Luke, 2013, Lemma 2.4))**:**
Let $\mathbf{D},\mathbf{S} \subset \mathbb{E}$ be nonempty and $T : \mathbf{D} \rightrightarrows \mathbb{E}$. The following are equivalent

(a) $T : \mathbf{D} \rightrightarrows \mathbb{E}$ is $(\mathbf{S},\varepsilon)$-firmly nonexpansive on $\mathbf{D}$.

(b) The mapping $\widetilde{T} : \mathbf{D} \rightrightarrows \mathbb{E}$ given by

$$\widetilde{T}x := (2Tx - x) \quad \forall x \in \mathbf{D}$$

is $(\mathbf{S},2\varepsilon)$-nonexpansive on $\mathbf{D}$, i.e. T can be written as

$$Tx = \frac{1}{2}\left(x + \widetilde{T}x\right) \quad \forall x \in \mathbf{D}. \tag{4.3}$$

Proof. For $x \in \mathbf{D}$ choose $x_+ \in Tx$. Observe that, by the definition of $\widetilde{T}$, there is a corresponding $\tilde{x} \in \widetilde{T}x$ such that $x_+ = \frac{1}{2}(x + \tilde{x})$, which is just formula (4.3). Let y be any point in $\mathbf{S}$ and select any $y_+ \in Ty$ (respectively choose $y_+, \tilde{y}$). Then

$$\begin{aligned}
&\|x_+ - y_+\|^2 + \|x - x_+ - (y - y_+)\|^2 \\
&\quad = \left\|\tfrac{1}{2}(x+\tilde{x}) - \tfrac{1}{2}(y+\tilde{y})\right\|^2 + \left\|\tfrac{1}{2}(x-\tilde{x}) - \tfrac{1}{2}(y-\tilde{y})\right\|^2 \\
&\quad = \tfrac{1}{4}\left[\|x-y\|^2 + 2\langle x-y, \tilde{x}-\tilde{y}\rangle + \|\tilde{x}-\tilde{y}\|^2\right] \\
&\qquad + \tfrac{1}{4}\left[\|x-y\|^2 - 2\langle x-y, \tilde{x}-\tilde{y}\rangle + \|\tilde{x}-\tilde{y}\|^2\right] \\
&\quad = \tfrac{1}{2}\|x-y\|^2 + \tfrac{1}{2}\|\tilde{x}-\tilde{y}\|^2 \\
&\quad \overset{!}{\leq} \tfrac{1}{2}\|x-y\|^2 + \tfrac{1}{2}(1+2\varepsilon)\|x-y\|^2 \\
&\quad = (1+\varepsilon)\|x-y\|^2,
\end{aligned}$$

where the inequality holds if and only if $\widetilde{T}$ is $(\mathbf{S},2\varepsilon)$−nonexpansive. By definition, it then holds that T is $(\mathbf{S},\varepsilon)$-firmly nonexpansive if and only if $\widetilde{T}$ is $(\mathbf{S},2\varepsilon)$−nonexpansive, as claimed. □

The $(\mathbf{S},\varepsilon)$-firm nonexpansiveness is preserved under convex combination of operators.

Theorem 4.4 ((Hesse and Luke, 2013, Theorem 2.5))**:**
Let T_1 be $(\mathbf{S},\varepsilon_1)$-firmly nonexpansive and T_2 be $(\mathbf{S},\varepsilon_2)$-firmly nonexpansive on $\mathbf{D}$. The convex combination $\lambda T_1 + (1-\lambda)T_2$ is $(\mathbf{S},\varepsilon)$-firmly nonexpansive on $\mathbf{D}$ where $\varepsilon = \max\{\varepsilon_1,\varepsilon_2\}$.

Proof. Let $x, y \in \mathbf{D}$. Let

$$\begin{array}{rrcll} & x_+ & \in & \lambda T_1 x + (1-\lambda)\, T_2 x, & \text{and} \\ & y_+ & \in & \lambda T_1 y + (1-\lambda)\, T_2 y, & \\ \Rightarrow & x_+ & = & \lambda x_+^{(1)} + (1-\lambda) x_+^{(2)}, & \text{where } x_+^{(1)} \in T_1 x,\ x_+^{(2)} \in T_2 x \\ & y_+ & = & \lambda y_+^{(1)} + (1-\lambda) y_+^{(2)}, & \text{where } y_+^{(1)} \in T_1 y,\ y_+^{(2)} \in T_2 y. \end{array}$$

By Lemma 4.3 (b) one has nonexpansiveness of the mappings $\widehat{T}_1, \widehat{T}_2$ given by $\widehat{T}_1 x = 2T_1 x - x$ and $\widehat{T}_2 x = 2T_2 x - x$, $x \in \mathbf{D}$ that is

$$\begin{aligned} \left\| \left[2x_+^{(1)} - x\right] - \left[2y_+^{(1)} - y\right] \right\| &\le \sqrt{1+2\varepsilon_1}\, \|x - y\|, \\ \left\| \left[2x_+^{(2)} - x\right] - \left[2y_+^{(2)} - y\right] \right\| &\le \sqrt{1+2\varepsilon_2}\, \|x - y\|. \end{aligned}$$

This implies

$$\begin{aligned} &\|(2x_+ - x) - (2y_+ - y)\| \\ &\quad = \left\| \left(2\left[\lambda x_+^{(1)} + (1-\lambda)x_+^{(2)}\right] - x\right) - \left(2\left[\lambda y_+^{(1)} + (1-\lambda) y_+^{(2)}\right] - y\right) \right\| \\ &\quad = \left\| \lambda\left(\left[2x_+^{(1)} - x\right] - \left[2y_+^{(1)} - y\right]\right) - (1-\lambda)\left(\left[2x_+^{(2)} - x\right] - \left[2y_+^{(2)} - y\right]\right) \right\| \\ &\quad \le \lambda \left\| \left[2x_+^{(1)} - x\right] - \left[2y_+^{(1)} - y\right] \right\| + (1-\lambda) \left\| \left[2x_+^{(2)} - x\right] - \left[2y_+^{(2)} - y\right] \right\| \\ &\quad \le \sqrt{1+2\varepsilon}\, \|x-y\|. \end{aligned}$$

The use of Lemma 4.3 (b) then completes the proof. □

4.2. Projectors and Reflectors

We show in this section how $(\mathbf{S},\varepsilon)$(firm)-nonexpansiveness of projectors and reflectors is a consequence of (sub-)regularity of the underlying sets.

Theorem 4.5 (projectors and reflectors onto (ε,δ)-subregular sets, (Hesse and Luke, 2013, Theorem 2.14))**:**
Let $\Omega \subset \mathbb{E}$ be nonempty, closed and (ε,δ)-subregular at $\hat{x}$ with respect to $\mathbf{S} \subseteq \Omega \cap \mathbb{B}_\delta(\hat{x})$ and define

$$U := \{x \in \mathbb{E} \mid P_\Omega x \subset \mathbb{B}_\delta(\hat{x})\}. \tag{4.4}$$

The following properties hold.

(a) The projector P_Ω is $(\mathbf{S},\tilde{\varepsilon}_1)$-nonexpansive on U, that is,

$$\begin{array}{c}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \\ \|x_+ - \bar{x}\| \leq \sqrt{1+\varepsilon}\,\|x-\bar{x}\|.\end{array} \tag{4.5}$$

where $\tilde{\varepsilon}_1 := 2\varepsilon + \varepsilon^2$.

(b) The projector P_Ω is $(\mathbf{S},\tilde{\varepsilon}_2)$-firmly nonexpansive on U, that is,

$$\begin{array}{c}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \\ \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \leq (1+\tilde{\varepsilon}_2)\,\|x-\bar{x}\|^2,\end{array} \tag{4.6}$$

where $\tilde{\varepsilon}_2 := 2\varepsilon + 2\varepsilon^2$.

(c) The reflector R_Ω is $(\mathbf{S},\tilde{\varepsilon}_3)$-nonexpansive on U, that is,

$$\forall x \in U,\ \forall x_+ \in R_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \|x_+ - \bar{x}\| \leq \sqrt{1+\tilde{\varepsilon}_3}\,\|x-\bar{x}\|, \tag{4.7}$$

where $\tilde{\varepsilon}_3 := 4\varepsilon + 4\varepsilon^2$.

Proof. (a) The projector is nonempty since Ω is closed. Then by the Cauchy-Schwarz inequality

$$\begin{array}{rcl}\|x_+ - \bar{x}\|^2 & = & \langle x-\bar{x}, x_+-\bar{x}\rangle + \langle x_+ - x, x_+ - \bar{x}\rangle \\ & \leq & \|x-\bar{x}\|\,\|x_+-\bar{x}\| + \langle x_+ - x, x_+ - \bar{x}\rangle.\end{array} \tag{4.8}$$

Now for $x \in U$ we have also that $x_+ \in \mathbb{B}_\delta(\hat{x})$ and thus, by the definition of (ε,δ)-subregularity with respect to $\mathbf{S}$,

$$\begin{array}{c}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \\ \begin{array}{rcl}\langle x_+ - x, x_+ - \bar{x}\rangle & \leq & \varepsilon\,\|x-x_+\|\,\|x_+-\bar{x}\| \\ & \leq & \varepsilon\,\|x-\bar{x}\|\,\|x_+-\bar{x}\|.\end{array}\end{array}$$

Combining this with (4.8) yields

$$\begin{array}{c}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \\ \begin{array}{rcl}\|x_+ - \bar{x}\| & \leq & (1+\varepsilon)\,\|x-\bar{x}\| \\ & = & \sqrt{1+(2\varepsilon+\varepsilon^2)}\,\|x-\bar{x}\|\end{array}\end{array} \tag{4.9}$$

as claimed.

(b) Expanding and rearranging the norm yields

$$\begin{array}{l}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S}: \\ \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \\ \quad = \|x_+ - \bar{x}\|^2 + \|x - \bar{x} + \bar{x} - x_+\|^2 \\ \quad = \|x_+ - \bar{x}\|^2 + \|x - \bar{x}\|^2 + 2\langle x-\bar{x}, \bar{x}-x_+\rangle + \|x_+ - \bar{x}\|^2 \\ \quad = 2\,\|x_+ - \bar{x}\|^2 + \|x - \bar{x}\|^2 + 2\underbrace{\langle x_+-\bar{x}, \bar{x}-x_+\rangle}_{=-\|x_+-\bar{x}\|^2} + 2\underbrace{\langle x-x_+, \bar{x}-x_+\rangle}_{\leq\varepsilon\|x-x_+\|\|x_+-\bar{x}\|} \\ \quad \leq \|x-\bar{x}\|^2 + 2\varepsilon\,\|x_+-\bar{x}\|\,\|x-x_+\|\end{array}$$

where the last inequality follows from the definition of (ε, δ)-subregularity with respect to $\mathbf{S}$. By definition, $\|x - x_+\| = d(x, \Omega) \le \|x - \bar{x}\|$. Combining the last inequality and inequality (4.5) yields

$$\begin{gathered}\forall x \in U,\ \forall x_+ \in P_\Omega x,\ \forall \bar{x} \in \mathbf{S} :\\ \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \le (1 + 2\varepsilon(1+\varepsilon)) \|x - \bar{x}\|^2 .\end{gathered}$$

(c) By (b) the projector is $(\mathbf{S}, 2\varepsilon + 2\varepsilon^2)$-firmly nonexpansive on U, and so by Lemma 4.3 (b) $R_\Omega = 2P_\Omega - \mathrm{Id}$ is $(\mathbf{S}, 4\varepsilon + 4\varepsilon^2)$-nonexpansive on U. □

Note that $\tilde{\varepsilon}_1 < \tilde{\varepsilon}_2$ $(\varepsilon > 0)$ in the above theorem, in other words, the *degree* to which classical firm nonexpansiveness is violated is greater than the degree to which classical nonexpansiveness is violated. This is as one would expect since firm nonexpansiveness is a stronger property than nonexpansiveness.

We can now characterize the degree to which the Douglas–Rachford operator violates firm-nonexpansiveness on neighborhoods of (ε, δ)-subregular sets.

Theorem 4.6 ($(\mathbf{S}, \tilde{\varepsilon})$-firm nonexpansiveness of T_{DR}, (Hesse and Luke, 2013, Theorem 2.15))**:**
Let $\mathbf{A}, \mathbf{B} \subset \mathbb{E}$ be closed and nonempty. Let $\mathbf{A}$ and $\mathbf{B}$ be $(\varepsilon_{\mathbf{A}}, \delta)$- and $(\varepsilon_{\mathbf{B}}, \delta)$-subregular respectively at $\hat{x}$ with respect to $\mathbf{S} \subset \mathbb{B}_\delta(\hat{x}) \cap (\mathbf{A} \cap \mathbf{B})$. Let $\mathrm{T}_{\mathrm{DR}} : \mathbb{E} \rightrightarrows \mathbb{E}$ be the Douglas–Rachford operator defined by (1.20) and define

$$U := \{z \in \mathbb{E} \mid P_{\mathbf{B}} z \subset \mathbb{B}_\delta(\hat{x}) \text{ and } P_{\mathbf{A}} R_{\mathbf{B}} z \subset \mathbb{B}_\delta(\hat{x})\}. \tag{4.10}$$

Then T_{DR} is $(\mathbf{S}, \tilde{\varepsilon})$-firmly nonexpansive on U, *i.e.,*

$$\begin{gathered}\forall x \in U,\ \forall x_+ \in \mathrm{T}_{\mathrm{DR}}\, x,\ \forall \bar{x} \in \mathbf{S} :\\ \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \le (1 + \tilde{\varepsilon}) \|x - \bar{x}\|^2 ,\end{gathered} \tag{4.11}$$

where

$$\tilde{\varepsilon} = 2\varepsilon_{\mathbf{A}}(1 + \varepsilon_{\mathbf{A}}) + 2\varepsilon_{\mathbf{B}}(1 + \varepsilon_{\mathbf{B}}) + 8\varepsilon_{\mathbf{A}}(1 + \varepsilon_{\mathbf{A}})\varepsilon_{\mathbf{B}}(1 + \varepsilon_{\mathbf{B}}). \tag{4.12}$$

Proof. Define $U_{\mathbf{A}} := \{z \mid P_{\mathbf{A}} z \subset \mathbb{B}_\delta(\hat{x})\}$. By Theorem 4.5 (c)

$$\begin{gathered}\forall y \in U_A,\ \forall \tilde{x} \in R_{\mathbf{A}} y,\ \forall \bar{x} \in \mathbf{S} :\\ \|\tilde{x} - \bar{x}\| \le \sqrt{1 + 4\varepsilon_A(1 + \varepsilon_A)}\, \|y - \bar{x}\| .\end{gathered} \tag{4.13}$$

Similarly, define $U_{\mathbf{B}} := \{z \mid P_{\mathbf{B}} z \subset \mathbb{B}_\delta(\hat{x})\}$ and again apply Theorem 4.5 (c) to get

$$\begin{gathered}\forall x \in U_{\mathbf{B}},\ \forall y \in R_{\mathbf{B}} x,\ \forall \bar{x} \in \mathbf{S} :\\ \|y - \bar{x}\| \le \sqrt{1 + 4\varepsilon_{\mathbf{B}}(1 + \varepsilon_{\mathbf{B}})}\, \|x - \bar{x}\| .\end{gathered} \tag{4.14}$$

Now, we choose any $x \in U_{\mathbf{B}}$ such that $R_{\mathbf{B}}x \in U_{\mathbf{A}}$, that is $x \in U$, so that we can combine (4.13)-(4.14) to get

$$\begin{array}{rcl} & & \forall x \in U,\ \forall \tilde{x} \in R_{\mathbf{A}}R_{\mathbf{B}}x,\ \forall \bar{x} \in \mathbf{S} : \\ \|\tilde{x} - \bar{x}\| & \leq & \sqrt{1+4\varepsilon_{\mathbf{A}}(1+\varepsilon_{\mathbf{A}})}\sqrt{1+4\varepsilon_{\mathbf{B}}(1+\varepsilon_{\mathbf{B}})}\,\|x-\bar{x}\| \\ & = & \sqrt{1+2\tilde{\varepsilon}}\,\|x-\bar{x}\|\,. \end{array} \tag{4.15}$$

Note that $R_{\mathbf{A}}R_{\mathbf{B}}\bar{x} = R_{\mathbf{B}}\bar{x} = \bar{x}$ since $\bar{x} \in \mathbf{A} \cap \mathbf{B}$, so (4.15) says that the operator $\widetilde{T} := R_{\mathbf{A}}R_{\mathbf{B}}$ is $(\mathbf{S},\tilde{\varepsilon})$-nonexpansive on U. Hence by Lemma 4.3 $\mathrm{T}_{\mathrm{DR}} = \frac{1}{2}\left(\widetilde{T}+I\right)$ is $(\mathbf{S},2\tilde{\varepsilon})$-firmly nonexpansive on U, as claimed. □

If one of the sets above is convex, say $\mathbf{B}$ for instance, the constant $\tilde{\varepsilon}$ simplifies to $\tilde{\varepsilon} = 2\varepsilon_{\mathbf{A}}(1+\varepsilon_{\mathbf{A}})$ since $\mathbf{B}$ is $(0,\infty)$-subregular at $\bar{x}$.

5. Regularity of Collections of Sets

5.1. Principal angles

The first idea on *principal angles* was introduced in (Jordan, 1875) and different notions have been (re-)discovered at several times in the literature. At this point, we will only introduce the following definition, as its connections to the convergence of Alternating Projections in well understood (Deutsch, 2001) and serves it serves as a reference for optimal convergence rates.

Definition 5.1 (Friedrichs angle, Dixmier angle on subspaces, (Friedrichs, 1937))**:**
The angle between two closed subspaces M and N is the angle in the interval $[0, \pi/2]$ whose cosine is given by

$$c_F(M, N) := \sup \left\{ \langle x, y \rangle \middle| \begin{matrix} x \in M \cap (M \cap N)^\perp \cap \mathbb{B} \\ y \in N \cap (M \cap N)^\perp \cap \mathbb{B} \end{matrix} \right\}. \tag{5.1}$$

The minimal or Dixmier angle between two subspaces is the number in $[0, \pi/2]$ whose cosine is given by

$$c_F^{(0)}(M, N) := \sup \left\{ \langle x, y \rangle \middle| x \in M \cap \mathbb{B}, y \in N \cap \mathbb{B} \right\}. \tag{5.2}$$

Theorem 5.2 ((Deutsch, 1995, Theorem 2.16))**:**
Let M and N be two closed subspaces. Then

$$c_F(N, M) = c_F\left(N^\perp, M^\perp\right). \tag{5.3}$$

5.2. Uniform and Linear Regularity

Definition 5.3 (uniform regularity, (Kruger, 2004, Definition 2 and Proposition 4))**:**
A collection of m closed, nonempty sets $\{\Omega_1, \Omega_2, \dots, \Omega_m\}$ is *uniformly regular at* $\hat{x}$ if there exists an $\alpha > 0$ and a $\delta > 0$ such that for all $\rho \in (0, \delta]$, $\omega_i \in \Omega_i \cap \mathbb{B}_\delta(\hat{x})$, $a_i \in \mathbb{B}_{\alpha\rho}(0)$, $i = 1, 2, \dots, m$:

$$\left(\bigcap_{i=1}^{m} (\Omega_i - \omega_i - a_i) \right) \cap \mathbb{B}_\rho \neq \emptyset. \tag{5.4}$$

Theorem 5.4 ((Kruger, 2006, Theorem 1))**:**
A collection of closed, nonempty sets $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is *uniformly regular* at $\hat{x}$ if and only if there exists a $\kappa > 0$ and a $\delta > 0$ such that for all $x \in \mathbb{B}_\delta(\hat{x})$, $x_i \in \mathbb{B}_\delta(0)$, $i = 1, \ldots, m$

$$d\left(x, \bigcap_{i=1}^{m}(\Omega_i - x_i)\right) \leq \kappa \max_{i=1,\ldots,m} d\left(x + x_i, \Omega_i\right). \tag{5.5}$$

Theorem 5.5 ((Kruger, 2006, Corollary 2))**:**
A collection of closed sets $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is *uniformly regular* (5.4) at a point $\hat{x} \in \bigcap_{i=1}^m \Omega_i$ if and only if the only solution to the system

$$\sum_{i=1}^{m} \nu_i = 0, \qquad \text{with } \nu_i \in N_{\Omega_i}(\hat{x}) \quad \text{for } i = 1, 2, \ldots, m \tag{5.6}$$

is $\nu_i = 0$ for $i = 1, 2, \ldots, m$.

Definition 5.6 ((Local) linear regularity, (Bauschke and Borwein, 1993, Definition 3.11 and 3.13))**:**
A collection of closed, nonempty sets $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is *locally linearly regular* at $\hat{x} \in \bigcap_{i=1}^m \Omega_i$ if there exists $\kappa > 0$ and $\delta > 0$ such that for all $x \in \mathbb{B}_\delta(\hat{x})$:

$$d\left(x, \bigcap_{i=1}^{m} \Omega_i\right) \leq \kappa \max_{i=1,\ldots,m} d\left(x, \Omega_i\right). \tag{5.7}$$

If for a given $\delta > 0$ there is a κ such that (5.7) holds, we say that the collection is locally linearly regular *with respect to* δ. If (5.7) holds for any $\delta > 0$ the collection of sets is called *linearly regular.*

Example 5.7*:* [example 1.16 revisited.] The collection of sets in example 1.16 (a) is uniformly regular at 0 and linearly regular. The same collection of sets embedded in a higher-dimensional space is still linearly regular, but looses its uniform regularity. This can be seen by shifting one of the sets in example 1.16 (b) in x_3-direction, as this renders the intersection empty. This shows that *linear regularity does not imply uniform regularity.* The collection of sets in example (c) is neither uniformly regular nor linearly regular. The collection of sets in example 1.16 (d) is uniformly regular at the intersection. One has $N_B(0) = \{(\lambda, -\lambda) |\ \lambda \in \mathbb{R}\}$ and by Remark 3.19 $N_A(0) = A$ and this directly shows $N_A(0) \cap -N_B(0) = \{0\}$ and hence implies condition (5.6). In example 1.16 (e) one of the sets is nonconvex, but the collection of sets is still well-behaved in the sense that it is both uniformly and linearly regular. It is worth emphasizing, however, that the set A in example 1.16 (d) is not Clarke regular at the origin (Remark 3.19). This illustrates the fact that collections of classically "irregular" sets can still be quite regular at points of intersection. △

Remark 5.8*: The notions of* strong, linear *and* uniform *are overused in literature which results in inconsistent nomenclature. Since equation* (5.5) *directly implies equation* (5.7)*, uniform regularity is indeed a more restrictive notion than linear regularity.*

Uniform regularity appears in (Lewis et al., 2009) as local linear regularity, *while it is called* strong regularity *in (Kruger, 2006, Propostion 2). Also compare this to the* basic constraint qualification for sets *in (Mordukhovich, 2006, Definition 3.2). The notion appeared first in (Kruger, 2004, Proposition 4) and was introduced as a characterization of the* absence of weak stationary.

Based on uniform regularity (or more specific characterizations (5.6)*) Lewis Luke and Malick proved local linear convergence of* $\boldsymbol{AP}$ *in the nonconvex setting, where both sets* $\mathbf{A}, \mathbf{B}$ *are closed and one of the sets is* superregular. *See (Lewis et al., 2009). This was, at the time, the most general* nonconvex *convergence result providing a linear convergence rate, compare Remark 1.8. However, as we will see at several points in this work, the notion of uniform regularity is in fact too restrictive and does not cover some easy convex examples. Compare for instance 5.7. The ideas of (Lewis et al., 2009) were refined in (Bauschke et al., 2013d; Bauschke et al., 2013c), compare Definition 5.40. The approach in (Lewis et al., 2009; Bauschke et al., 2013d; Bauschke et al., 2013c) differs from the approach in this work. Whenever possible connections between the different techniques and frameworks are given.*

The initial definition of (bounded) linear regularity *goes back to (Bauschke and Borwein, 1993, Definition 3.11 and 3.13). Compare this to (Bauschke and Borwein, 1996, Definition 5.6). What we are calling (local) linear regularity has appeared in various forms elsewhere. See for instance (Ioffe, 2000, Proposition 4), (Ngai and Théra., 2001, Section 3), and (Kruger, 2006, Equation (15)). Based on (bounded) linear regularity (Bauschke and Borwein, 1993) showed linear convergence of alternating projections in the convex setting for a detailed survey see also (Bauschke and Borwein, 1996).*◊

5.3. Metric Regularity

Definition 5.9 ((strong) metric (sub)-regularity)**:**

(a) The mapping $\Phi : \mathbb{E} \rightrightarrows \mathbb{Y}$ is called *metrically regular* at $\hat{x}$ for $\hat{y}$ if there is a finite scalar $\kappa > 0$ together with neighborhood $U_{\hat{x}}$ of $\hat{x}$ and $V_{\hat{y}}$ of $\hat{y}$ such that

$$d\left(x, \Phi^{-1}(y)\right) \leq \kappa\, d\left(y, \Phi(x)\right) \quad \text{for all } (x, y) \in U_{\hat{x}} \times V_{\hat{y}}. \tag{5.8}$$

The *regularity modulus* $\operatorname{reg}\Phi(\hat{x}|\hat{y})$ is the infimum of $\kappa > 0$ over all such combinations $\kappa, U_{\hat{x}}, V_{\hat{y}}$ that (5.8) holds.

(b) The mapping $\Phi : \mathbf{X} \rightrightarrows \mathbb{Y}$ is called *metrically subregular* at $\hat{x}$ for $\hat{y}$ if $(\hat{x}, \hat{y}) \in \operatorname{gph}\Phi$ and there is a finite scalar $\kappa > 0$ and neighborhoods $U_{\hat{x}}$ of $\hat{x}$ and $V_{\hat{y}}$ of $\hat{y}$ such that

$$d\left(x, \Phi^{-1}(\hat{y})\right) \leq \kappa\, d\left(\hat{y}, \Phi(x) \cap V_{\hat{y}}\right) \quad \text{for all } x \in U_{\hat{x}}. \tag{5.9}$$

The *subregularity modulus* $\operatorname{subreg}\Phi(\hat{x}|\hat{y})$ is the infimum of $\kappa > 0$ over all such combinations $\kappa, U_{\hat{x}}, V_{\hat{y}}$ that (5.9) holds.

(c) The mapping $\Phi : \mathbf{X} \rightrightarrows \mathbb{Y}$ is *strongly metrically subregular* at $\hat{x}$ for $\hat{y}$ if $(\hat{x}, \hat{y}) \in \operatorname{gph}\Phi$ and there is a finite scalar $\kappa > 0$ along with neighborhoods $U_{\hat{x}}$ of $\hat{x}$ and $V_{\hat{y}}$ of $\hat{y}$ such that

$$\|x - \hat{x}\| \leq \kappa\, d\left(\hat{y}, \Phi(x) \cap V\right) \quad \text{for all } x \in U_{\hat{x}}. \tag{5.10}$$

Remark 5.10*: The nomenclature* metric regularity *goes back to (Borwein, 1986), where the concept itself goes back far earlier. See for instance the independent works (Ursecu, 1975) and (Robinson, 1976). In this work we follow the terminology of (Rockafellar and Wets, 1998; Dontchev and Rockafellar, 2008). For detailed historical remarks on* metric regularity *see (Rockafellar and Wets, 1998, Commentary to Chapter 9) and on* metric regularity *and* (strong) metric subregularity *see (Dontchev and Rockafellar, 2008, Commentary to Chapter 5) and the references therein.* $\Diamond$

Corollary 5.11:
Strong metric subregularity of Φ at a point $\hat{x}$ for $\hat{y}$ is equivalent to metric subregularity of Φ at $\hat{x}$ for $\hat{y}$ and $\hat{x}$ being an *isolated point* of Φ^{-1} at $\hat{y}$.

Proof. This is an equivalent definition and can be seen by (5.9). For a detailed study see (Dontchev and Rockafellar, 2008, pp. 186/187). $\square$

Using the above definitions and theorems from metric regularity, we can now establish the connection to the definitions of *uniform* and *local linear* regularity.

Proposition 5.12:
For $\Omega_1, ..., \Omega_m$ closed and nonempty subsets of $\mathbb{E}$ consider the set-valued mapping $\Phi : \mathbb{E} \rightrightarrows \mathbb{E}^m$.

$$\Phi(x) := (\Omega_1 - x) \times \cdots \times (\Omega_m - x). \tag{5.11}$$

Then

$$\Phi^{-1}((y_1, \ldots, y_m)) = \bigcap_{i=1}^{m} \{x \mid y_i \in \Omega_i - x\} = \bigcap_{i=1}^{m} \Omega_i - y_i, \tag{5.12}$$

and

$$\Phi^{-1}((0, \ldots, 0)) = \bigcap_{i=1}^{m} \Omega_i.$$

Hence finding $(\hat{x}, 0) \in \mathrm{gph}\Phi$ is equivalent to finding $\hat{x} \in \bigcap_{i=1}^{m} \Omega_i$.

Theorem 5.13:
Let $\Omega_1, ..., \Omega_m$ be closed and nonempty subsets of $\mathbb{E}$ and let $\Phi : \mathbb{E} \rightrightarrows \mathbb{E}^m$ be set-valued mapping given by equation (5.11), *i.e.,*

$$\Phi(x) = (\Omega_1 - x) \times \cdots \times (\Omega_m - x).$$

We have the following characterizations:

(a) Φ is metrically regular at $\hat{x}$ for 0 if and only if $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is uniformly regular at $\hat{x}$.

(b) Φ is metrically subregular at $\hat{x}$ for 0 if and only if $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is locally linearly regular.

(c) Φ is strongly metrically subregular at $\hat{x}$ for 0 if and only if $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ is locally linearly regular and $\{\hat{x}\} = \bigcap_{i=1}^{m} \Omega_i$.

Proof. (a) Let Φ be metrically regular at $\hat{x}$ for 0. There are $U_{\hat{x}} \subset \mathbb{E}$ and $V_0 \subset \mathbb{E}^m$ such that

$$d\left(x, \Phi^{-1}(y)\right) \leq \kappa\, d\left(y, \Phi(x)\right) \quad \text{for all } (x, y) \in U_{\hat{x}} \times V_0.$$

Now choose $\delta > 0$, such that $\mathbb{B}_\delta(\hat{x}) \subset U_{\hat{x}}$ and $\mathbb{B}_\delta(0) \times \cdots \times \mathbb{B}_\delta(0) \subset V_0$ and substitute $(x_1, x_2, \ldots, x_m) := y$ to achieve

$$d\left(x, \bigcap_{i=1}^{m}(\Omega_i - x_i)\right) \leq \kappa \sqrt{\sum_{i=1}^{m} d^2\left(x + x_i, \Omega_i\right)}, \quad \forall x \in \mathbb{B}_\delta(\hat{x}). \tag{5.13}$$

By norm equivalence on $\mathbb{E}^m$ this implies existence of $\tilde{\kappa}$ such that (5.5) holds. On the other hand if the collection $\{\Omega_1, \ldots, \Omega_m\}$ is uniformly regular then, also by norm equivalence the strong metric inequality (5.8) holds for $U_{\hat{x}} := \mathbb{B}_\delta(\hat{x}) \subset \mathbb{E}$, $V_0 := \mathbb{B}_\delta(0) \times \cdots \times \mathbb{B}_\delta(0) \subset \mathbb{E}^m$.

(b) For $y = 0$ (resp. $x_i = 0$ for $i = 1, \ldots, m$) the equivalence between metric subregularity of Φ at $\hat{x}$ for 0 and local linear regularity of $\{\Omega_1, \ldots, \Omega_m\}$ and therefore (b) follows analogous to the first part.

(c) This is a consequence of Proposition 5.11. □

Remark 5.14*: Theorem 5.13 (a) was first mentioned in (Kruger, 2006, Proposition 9) and applied to the context of* nonconvex feasibility *in (Lewis et al., 2009, Section 3). The properties (b) and (c) were proven in the prepint (Hesse and Luke, 2012). However this proposition is not included in the final version of the article (Hesse and Luke, 2013).* ◊

The following statement is an immediate consequence of Theorem 5.13.

Corollary 5.15:
Let $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \bigcap_{i=1}^m \Omega_i$. The collection is

(a) *uniformly regular* at $\hat{x}$ if and only if there exists a $\kappa > 0$ and a $\delta > 0$ such that, for all $x \in \mathbb{B}_\delta(\hat{x})$, $x_i \in \mathbb{B}_\delta(0)$, $i = 1, \ldots, m$,

$$d\left(x, \bigcap_{i=1}^{m} (\Omega_i - x_i)\right) \leq \kappa \sqrt{\sum_{i=1}^{m} d^2\left(x + x_i, \Omega_i\right)}. \tag{5.14}$$

(b) *locally linearly regular* at $\hat{x}$ if and only if there exists a $\kappa > 0$ and a $\delta > 0$ such that, for all $x \in \mathbb{B}_\delta(\hat{x})$

$$d\left(x, \bigcap_{i=1}^{m} \Omega_i\right) \leq \kappa \sqrt{\sum_{i=1}^{m} d^2\left(x, \Omega_i\right)}. \tag{5.15}$$

In order to calculate the moduli of regularity $\mathrm{reg}\Phi(\hat{x}, 0)$ and $\mathrm{subreg}\Phi(\hat{x}, 0)$ of the function Φ given by (5.11) we need some additional tools from variational analysis. We will also provide some interesting properties of the function Φ, which will help us to develop a framework that allows us to *quantify* the definitions of *uniform* and *linear regularity* in a sense that gives insight in the local geometry of the intersections of the sets.

Definition 5.16 (Graphical derivative and coderivative, (Rockafellar and Wets, 1998, Definition 8.33))**:**
Let $\Phi : \mathbb{E} \rightrightarrows \mathbb{Y}$ be a multi valued function. The *graphical derivative* $\mathrm{D}\Phi$ and the *graphical coderivative* $\mathrm{D}^*\Phi$ are defined by

$$z \in \mathrm{D}\Phi(\hat{x} \mid \hat{y})(w) \quad :\Leftrightarrow \quad (w,\ z) \in T_{\mathrm{gph}\Phi}(\hat{x}|\hat{y}), \tag{5.16}$$

$$v \in \mathrm{D}^*\Phi(\hat{x} \mid \hat{y})(y) \quad :\Leftrightarrow \quad (v, -y) \in N_{\mathrm{gph}\Phi}\left(\hat{x}|\hat{y}\right). \tag{5.17}$$

Proposition 5.17 ((Rockafellar and Wets, 1998, Theorem 9.43) and (Dontchev and Rockafellar, 2008, Theorem 4C.1))**:**
Let $\Phi : \mathbb{E} \rightrightarrows \mathbb{Y}$ be a multi valued function and let $(\hat{x}, \hat{y}) \in \mathrm{gph}\Phi$ and $\mathrm{gph}\Phi$ be closed around $(\hat{x}, \hat{y})$. The following statements hold:

(a) If Φ is metrically regular then

$$\mathrm{reg}\Phi(\hat{x}, \hat{y}) = 1/\min\left\{ d\left(0, \mathrm{D}^*\Phi(\hat{x}|\hat{y})(y)\right) \mid \|y\| = 1\right\}. \tag{5.18}$$

(b) Φ is strongly metrically subregular at $\hat{x}$ for $\hat{y}$ if and only if

$$\mathrm{D}\Phi(\hat{x}|\hat{y})^{-1}(0) = \{0\}, \tag{5.19}$$

which is equivalent to $|\mathrm{D}\Phi(\hat{x}|\hat{y})^{-1}|^+ < \infty$, and in that case

$$\mathrm{subreg}\Phi(\hat{x}, \hat{y}) = \left|\mathrm{D}\Phi(\hat{x}|\hat{y})^{-1}\right|^+, \tag{5.20}$$

where for any sublinear mapping $F : \mathbb{E} \to \mathbb{Y}$

$$|F|^+ := \sup_{x \in \mathbb{B}} \sup_{y \in F(x)} \|y\| \tag{5.21}$$

is the *outer norm.*

Proposition 5.18 ((Rockafellar and Wets, 1998), Exercise 10.43)**:**
Let $\Psi : \mathbb{E} \rightrightarrows \mathbb{Y}$ and let $f : \mathbb{E} \to \mathbb{Y}$ be continuously differentiable. For $\hat{x} \in \mathbb{E}$, $\hat{u} \in \Phi(\hat{x})$ define $\hat{u}_0 := \hat{u} - \Phi(\hat{x})$. For $\Phi := \Psi + f$ following identities hold:

$$\mathrm{D}\Phi(\hat{x}|\hat{u})(w) = \mathrm{D}\Psi(\hat{x}|\hat{u}_0)(w) + \nabla_{\hat{x}} f \cdot w \quad \text{for all } w \in \mathbb{E}, \tag{5.22}$$

$$\mathrm{D}^*\Phi(\hat{x}|\hat{u})(y) = \mathrm{D}^*\Psi(\hat{x}|\hat{u}_0)(y) + [\nabla_{\hat{x}} f]^* \cdot y \quad \text{for all } y \in \mathbb{Y}. \tag{5.23}$$

Theorem 5.19 (Properties of Φ)**:**
Let $\Omega_1, \dots, \Omega_m$, be closed and nonempty subsets of $\mathbb{E}$. Define function $\Phi : \mathbb{E} \rightrightarrows \mathbb{E}^m$, by (5.11), *i.e.,*

$$\Phi(x) = (\Omega_1 - x) \times \cdots \times (\Omega_m - x).$$

The following statements hold:

(a) The graphical derivative and coderivative of Φ are given by

$$\mathrm{D}\Phi(\hat{x} \mid 0)(w) = (T_{\Omega_1}(\hat{x}) - w) \times \cdots \times (T_{\Omega_i}(\hat{x}) - w), \tag{5.24}$$

$$\mathrm{D}^*\Phi(\hat{x} \mid 0)(y) = \begin{cases} -\sum_{i=1}^m y_i & \text{if } y_i \in -N_{\Omega_i}(\hat{x}), \ \forall i = 1, \dots, m \\ \emptyset & \text{else} \end{cases}. \tag{5.25}$$

(b)

$$\frac{1}{\mathrm{reg}\Phi(\hat{x} \mid 0)} = \min\left\{ \left\| \sum_{i=1}^m \nu_i \right\| \;\middle|\; \sum_{i=1}^m \|\nu_i\|^2 = 1, \ \nu_i \in N_{\Omega_i}(\hat{x}) \right\} \tag{5.26}$$

(c) The collection $\{\Omega_1, \dots, \Omega_m\}$ is locally linearly regular and $\bigcap_{i=1}^m \Omega_i = \{\hat{x}\}$ if and only if

$$\bigcap_{i=1}^{m} T_{\Omega_i}(\hat{x}) = \{0\}. \tag{5.27}$$

This then implies

$$\begin{array}{rl} & \mathrm{subreg}\Phi(\hat{x} \mid 0) \\ = & \max\left\{\|w\| \mid (T_{\Omega_1}(\hat{x}) - w) \times \cdots \times (T_{\Omega_m}(\hat{x}) - w) \cap \mathbb{B} \neq \emptyset\right\}. \end{array}$$

Proof. (a) To compute the graphical derivative $\mathrm{D}\Phi(\hat{x} \mid 0)$, we decompose the mapping Φ as $\Psi - A$, where, for points $x \in \mathbb{E}$,

$$\begin{array}{rcl} \Psi(x) & = & \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m \\ Ax & = & \underbrace{(x, x, \dots, x)}_{m-\text{times}}. \end{array}$$

One has

$$\mathrm{gph}\Psi = \mathbb{E} \times \Omega_1 \times \cdots \times \Omega_m,$$

and therefore

$$\begin{array}{rcl} T_{\mathrm{gph}\Psi}(\hat{x} \mid A\hat{x}) & = & \mathbb{E} \times T_{\Omega_1}(\hat{x}) \times \cdots \times T_{\Omega_m}(\hat{x}), \\ N_{\mathrm{gph}\Psi}(\hat{x} \mid A\hat{x}) & = & \{0\} \times N_{\Omega_1}(\hat{x}) \times \cdots \times N_{\Omega_m}(\hat{x}). \end{array}$$

By Definition (5.16) and (5.17) this yields

$$\begin{array}{rcl} \mathrm{D}\Psi(\hat{x} \mid 0)(w) & = & T_{\Omega_1}(\hat{x}) \times \cdots \times T_{\Omega_m}(\hat{x}), \\ \mathrm{D}^*\Psi(\hat{x} \mid 0)(y) & = & \left\{ \begin{array}{ll} \{0\} & \text{if } y_i \in -N_{\Omega_i}(\hat{x}),\ \forall i = 1, \dots, m \\ \emptyset & \text{else} \end{array} \right. . \end{array}$$

An application of (5.22) and (5.23) then yields (5.24) and (5.25). Application of (5.18) then shows (5.26) and therefore (b) is complete.

(c) Note that

$$\mathrm{D}\Phi(\hat{x} \mid 0)^{-1}(y_1, \dots, y_m) = \bigcap_i (T_{\Omega_i}(\hat{x}) - y_i).$$

By Proposition 5.17 equation (5.19) strong metric subregularity of Φ at $\hat{x}$ for 0 is equivalent to

$$\mathrm{D}\Phi(\hat{x} \mid 0)^{-1}(0) = \bigcap_i T_{\Omega_i}(\hat{x}) = \{0\}.$$

By Theorem 5.13 (c) this is equivalent to local linear regularity and $\bigcap_{i=1}^{m} \Omega_i = \{\hat{x}\}$. Also by Proposition 5.17 equation (5.20) one achieves

$$\begin{aligned}
\operatorname{subreg}(\Phi; \hat{x} \mid 0) &= |\mathrm{D}\Phi(\hat{x} \mid 0)^{-1}|^{+} \\
&= \sup_{\|y\|\leq 1} \sup_{v \in \mathrm{D}\Phi(\hat{x}|0)^{-1}(y)} \|v\| \\
&= \sup\left\{\|w\| \;\middle|\; w \in \mathrm{D}\Phi(\hat{x} \mid 0)^{-1}(\mathbb{B}) \neq \emptyset\right\} \\
&= \sup\{\|w\| \mid \mathrm{D}\Phi(\hat{x} \mid 0)(w) \cap \mathbb{B} \neq \emptyset\} \\
&= \max\{\|w\| \mid \mathrm{D}\Phi(\hat{x} \mid 0)(w) \cap \mathbb{B} \neq \emptyset\}.
\end{aligned}$$

where we used that $w \in \mathrm{D}\Phi(\hat{x} \mid 0)^{-1}(\mathbb{B})$ if and only if $\mathrm{D}\Phi(\hat{x} \mid 0)(w) \cap \mathbb{B} \neq \emptyset$. The maximum is then attained by compactness of the unit ball $\mathbb{B}$. □

Remark 5.20*: Theorem 5.19 complements (Lewis et al., 2009, Section 3) where both* (5.25) *and* (5.26) *were achieved.* ♢

Theorem 5.21:
Consider two closed, nonempty sets Ω_1, Ω_2. The following statements are equivalent:

i) The family $\{\Omega_1, \Omega_2\}$ is uniformly regular (5.4) at $\hat{x}$.

ii) The constant

$$c_D := \max\left\{-\langle \nu_1, \nu_2\rangle \;\middle|\; \begin{array}{ll} \nu_1 \in & N_{\Omega_1}(\hat{x}) \cap \mathbb{B} \\ \nu_2 \in & N_{\Omega_2}(\hat{x}) \cap \mathbb{B} \end{array}\right\} \tag{5.28}$$

is stricly less than 1.

iii) The constant

$$c_D = \max\left\{2\frac{-\langle \nu_1, \nu_2\rangle}{\|\nu_1\|^2 + \|\nu_2\|^2} \;\middle|\; \begin{array}{ll} \nu_1 \in & N_{\Omega_1}(\hat{x}) \\ \nu_2 \in & N_{\Omega_2}(\hat{x}) \end{array}\right\} \tag{5.29}$$

is stricly less than 1.

iv) The regularity modulus $\operatorname{reg}\Phi(\hat{x}|\hat{y})$ is finite and

$$\operatorname{reg}\Phi(\hat{x}|\hat{y}) = \frac{1}{\sqrt{1 - c_D}}. \tag{5.30}$$

The constants (5.28) and (5.29) coincide.

Proof. [i) ⇔ ii)] Note that by equation (5.6) for $m = 2$ the collection $\{\Omega_1, \Omega_2\}$ is uniformly regular at $\hat{x}$ if and only if $N_{\Omega_1}(\hat{x}) \cap N_{\Omega_2}(\hat{x}) = \{0\}$. By compactness of the unit ball the maximum in (5.28) is attained for some $\tilde{\nu}_1, \tilde{\nu}_2$ and by the *Cauchy-Schwarz* inequality $c_D = -\langle \tilde{\nu}_1, \tilde{\nu}_2\rangle < 1$ holds if and only if $\tilde{\nu}_1 \neq -\tilde{\nu}_2$, *i.e.*, the collection is uniformly regular.

[ii) ⇔ iii)] To see that (5.28) and (5.29) coincide note that

$$\begin{aligned}\frac{2}{\|\nu_1\|^2+\|\nu_2\|^2}|\langle\nu_1,\nu_2\rangle| &\leq \frac{1}{\|\nu_1\|\,\|\nu_2\|}|\langle\nu_1,\nu_2\rangle| \\ &\leq c_D\end{aligned}$$

for all $\nu_i \in N_{\Omega_i}(\hat{x})$, $i=1,2$. By compactness of $\mathbb{B}$ the maximum in (5.28) is attained and there are $\bar{\nu}_1 \in N_{\Omega_1}(\hat{x})$, $\|\bar{\nu}_1\|=1$ and $\bar{\nu}_2 \in N_{\Omega_2}(\hat{x})$, $\|\bar{\nu}_2\|=1$ such that $c_D=-\langle\bar{\nu}_1,\bar{\nu}_2\rangle$ and one achieves

$$\frac{2}{\|\bar{\nu}_1\|^2+\|\bar{\nu}_2\|^2}\langle\bar{\nu}_1,\bar{\nu}_2\rangle = \langle\bar{\nu}_1,\bar{\nu}_2\rangle = c_D,$$

which completes (5.29).

[iii) ⇔ iv)] By Theorem 5.13 and the definition of metric regularity $\operatorname{reg}\Phi(\hat{x}|0)<\infty$ is equivalent to uniform regularity of $\{\Omega_1,\Omega_2\}$. To show (5.30) we use (5.25) and apply (5.18) to achieve

$$\begin{aligned}\frac{1}{\operatorname{reg}\Phi(\hat{x}\mid 0)} &= \min\left\{\|\nu_1+\nu_2\| \;\middle|\; \|\nu_1\|^2+\|\nu_2\|^2=1,\ \nu_i\in N_{\Omega_i}(\hat{x}),\ i=1,2\right\} \\ &= \min\left\{\sqrt{1+2\langle\nu_1,\nu_2\rangle} \;\middle|\; \|\nu_1\|^2+\|\nu_2\|^2=1,\ \nu_i\in N_{\Omega_i}(\hat{x}),\ i=1,2\right\} \\ &= \max\left\{\sqrt{1-2\langle\nu_1,\nu_2\rangle} \;\middle|\; \|\nu_1\|^2+\|\nu_2\|^2=1,\ \begin{array}{rr}\nu_1\in & N_{\Omega_1}(\hat{x}) \\ \nu_2\in & -N_{\Omega_2}(\hat{x})\end{array}\right\} \\ &\overset{(5.29)}{=} \sqrt{1-c_D}.\end{aligned}$$

□

Remark 5.22: *Theorem 5.21 is a generalization of (Lewis et al., 2009, Section 5). Lewis, Luke and Malick used the idea that for two sets $\Omega_1,\Omega_2\subset\mathbb{E}$ uniform regularity, or more precise the characterization in equation* (5.6), *can be restated as*

$$N_{\Omega_1}(\hat{x})\cap -N_{\Omega_2}(\hat{x})=\{0\}. \tag{5.31}$$

They introduced the dual *definition of an angle between two closed sets* (5.28) *and showed that c_D is less then* 1 *if and only if* (5.31) *holds. (Lewis et al., 2009) proved local linear convergence of AP in the nonconvex setting, where both sets* **A**, **B** *are closed and one of the sets is* superregular *and the rate of linear convergence depends on c_D. They furthermore were the first to observe the relation* (5.30), i.e., *the relation between the* modulus of uniform regularity (5.14) *and the dual definition of the angle c_D* (5.28). *We will complement their ideas throughout section 5.4.* ♢

Theorem 5.23:
Consider two closed, nonempty sets Ω_1, Ω_2. The following statements are equivalent:

i) The collection $\{\Omega_1, \Omega_2\}$ is locally linearly regular (5.7) at $\hat{x}$ and $\{\hat{x}\} = \Omega_1 \cap \Omega_2$.

ii) The constant

$$c_P := \max\left\{ \langle u, v\rangle \,\middle|\, \begin{array}{lcl} u & \in & T_{\Omega_1}(\hat{x}) \cap \mathbb{B} \\ v & \in & T_{\Omega_2}(\hat{x}) \cap \mathbb{B} \end{array} \right\} \tag{5.32}$$

is strictly less than 1.

iii) The mapping Φ given by (5.11) is strongly metrically subregular.

i) ⇔ ii). By (5.27) strong metric subregularity of Φ at $\hat{x}$ for 0 is equivalent to

$$\bigcap_{i=1}^{2} T_{\Omega_i}(\hat{x}) = \{0\}. \tag{5.33}$$

By the fact that the intersections $T_{\Omega_1}(\hat{x}) \cap \mathbb{B}$, $T_{\Omega_2}(\hat{x}) \cap \mathbb{B}$ are compact in $\mathbb{E}$ and the fact that the Cauchy-Schwarz inequality holds with equality if and only if u and v are linearly dependent $c_P < 1$ holds if and only if (5.33) holds, which completes the proof.

[i) ⇔ iii)] This is a consequence of Theorem 5.13 and the definition of strong metric subregularity.

□

Remark 5.24*: Theorem 5.23 ii) seems to be new. The definition* (5.32) *relates to the* Dixmier angle*, which initially was defined on subspaces* (5.2)*. See Definition 5.1* (5.2)*.* ◊

5.4. Quantitative Notions of Regularity

We have already seen in Theorem 5.13 that for any finite collection $\{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ of m sets, one can replace the maximum norm $\max_{i=1}^{m} d(x, \Omega_i)$ on the product space $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_m \subset \mathbb{E}^m$ (see (5.5) and (5.7)) by the euclidean norm $\sqrt{\sum_{i=1}^{m} d^2(x, \Omega_i)}$ without any change to the qualitative characteristics given by the previous definitions and theorems.

The tools from *metric regularity* indicate some nice properties between the regularity moduli and the idea of an angle (compare (5.28), (5.29)). However, these characteristics, or more precisely the regularity moduli and the constant c_D are qualitative notions in the sense, that they are only true asymptotically, *i.e.,* they are defined as infima (resp. suprema) over all neighborhoods or equivalently the limits over neighborhoods $\mathbb{B}_\delta$ as $\delta \to 0$. In the following section we will build up an analysis and tools that locally *quantify* the results from metric regularity in a sense more appropriate for local convergence analysis of our intended applications.

Definition 5.25 (local regularity moduli)**:**
Let $\{\Omega_1, \Omega_2\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \Omega_1 \cap \Omega_2$.

(a) For $\delta > 0$ the *local modulus of uniform regularity* $\kappa_u^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}$ is defined as the smallest constant $\kappa > 0$ such that

$$\begin{gathered} x_i \in \mathbb{B}_\delta(\hat{x}),\ \nu_i \in N^{\mathrm{P}}_{\Omega_i}(x_i),\ \text{ for } i = 1,2: \\ \sqrt{\|\nu_1\|^2 + \|\nu_2\|^2} \le \kappa \|\nu_1 + \nu_2\|. \end{gathered} \tag{5.34}$$

(b) For $\delta > 0$ the *local modulus of linear regularity* $\kappa_l^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}$ is defined as the smallest constant κ such that

$$d\left(x, \bigcap_{i=2}^{m} \Omega_i\right) \le \kappa \sqrt{d^2(x, \Omega_1) + d^2(x, \Omega_2)} \tag{5.35}$$

holds for all $x \in \mathbb{B}_\delta(\hat{x})$.

By definition of linear regularity it is an immediate consequence that existence of a δ such that $\kappa_l^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}$ is finite is equivalent to local linear regularity [compare characterization (5.15)]. The local modulus of uniform regularity is a *quantification* of the constraint qualification (5.6).

Definition 5.26:
Let $\{\Omega_1, \Omega_2\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \Omega_1 \cap \Omega_2$. Define

$$c_D^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\} := \sup\left\{ -2\frac{\langle \nu_2, \nu_2 \rangle}{\|\nu_1\|^2 + \|\nu_2\|^2} \;\middle|\; \begin{array}{c} \nu_i \in N^{\mathrm{P}}_{\Omega_i}(x_i) \\ x_i \in \mathbb{B}_\delta(\hat{x}) \end{array},\ i = 1,2 \right\}. \tag{5.36}$$

Theorem 5.27:
Let $\{\Omega_1, \Omega_2\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \Omega_1 \cap \Omega_2$. The collection is uniformly regular at $\hat{x}$ if and only if there exists $\bar{\delta} > 0$ such that for all $\delta \in (0, \bar{\delta}]$ the *local modulus of uniform regularity* $\kappa_u^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}$ given by (5.34) is finite. Furthermore for all $\delta \in (0, \bar{\delta}]$ the following relations hold

$$\frac{1}{\kappa_u^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}} = \inf\left\{ \|\nu_1 + \nu_2\| \;\middle|\; \begin{array}{cc} \|\nu_1\|^2 + \|\nu_2\|^2 = 1, & x_i \in \mathbb{B}_\delta(\hat{x}), \\ \nu_i \in N^{\mathrm{P}}_{\Omega_i}(x_i), & i = 1,2 \end{array} \right\} \tag{5.37}$$

$$\kappa_u^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\} = \frac{1}{\sqrt{1 - c_D^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}}}, \tag{5.38}$$

where $c_D^{(\delta)}(\hat{x})\{\Omega_1, \Omega_2\}$ is defined by (5.36).

Proof. Assume that $\operatorname{reg}\Phi(\hat{x} \mid 0) < \infty$ [The case $\operatorname{reg}\Phi(\hat{x} \mid 0) = \infty$, *i.e.*, the case where the collection is not uniformly regular, is trivial by (5.6)]. Remember that by (5.26)

$$\frac{1}{\mathrm{reg}\Phi(\hat{x}\mid 0)} = \min\left\{\|\nu_1+\nu_2\| \;\middle|\; \|\nu_1\|^2+\|\nu_2\|^2=1,\ \nu_i\in N_{\Omega_i}(\hat{x}),\ i=1,2\right\}.$$

By compactness of the unit sphere and closedness of $N_{\Omega_i}(\bar{x})$ there are $\bar{\nu}_i \in N_{\Omega_i}(\bar{x})$, $\|\bar{\nu}_1\|^2+\|\bar{\nu}_2\|^2=1$, $(i=1,2)$ such that the minimum is attained, *i.e.*,

$$1/\mathrm{reg}\Phi(\hat{x}\mid 0) = \|\bar{\nu}_1+\bar{\nu}_2\|.$$

Now by Proposition 3.3 there are $x_i^{(k)}\in\Omega_i$ and $\nu_i^{(k)}\in N_\Omega^{\mathrm{P}}\left(x_i^{(k)}\right)$, $\left\|\nu_1^{(k)}\right\|^2+\left\|\nu_2^{(k)}\right\|^2=1$ such that $x_i^{(k)}\to\bar{x}$ and $\nu_i^{(k)}\to\bar{\nu}_i$, as $k\to\infty$, for $i=1,2$. As a consequence then

$$\frac{1}{\mathrm{reg}\Phi(\hat{x}\mid 0)} = \lim_{k\to\infty}\left\|\nu_1^{(k)}+\nu_2^{(k)}\right\|.$$

On the other hand for any $\delta>0$

$$\frac{1}{\mathrm{reg}\Phi(\hat{x}\mid 0)} \geq \inf\left\{\|\nu_1+\nu_2\| \;\middle|\; \begin{array}{cc}\|\nu_1\|^2+\|\nu_2\|^2=1, & x_i\in\mathbb{B}_\delta(\hat{x}),\\ \nu_i\in N_{\Omega_i}^{\mathrm{P}}(x_i), & i=1,2\end{array}\right\}.$$

Combining these last two statements one achieves

$$\frac{1}{\mathrm{reg}\Phi(\hat{x}\mid 0)} = \liminf_{\delta\searrow 0}\left\{\|\nu_1+\nu_2\| \;\middle|\; \begin{array}{cc}\|\nu_1\|^2+\|\nu_2\|^2=1, & x_i\in\mathbb{B}_\delta(\hat{x}),\\ \nu_i\in N_{\Omega_i}^{\mathrm{P}}(x_i), & i=1,2\end{array}\right\}.$$

Hence for any $\kappa>\mathrm{reg}\Phi(\hat{x}\mid 0)$ [*i.e.*, $1/\kappa<1/\mathrm{reg}\Phi(\hat{x}\mid 0)$] there is $\bar{\delta}>0$ such that for any $\delta\in(0,\bar{\delta}]$

$$\frac{1}{\mathrm{reg}\Phi(\hat{x}\mid 0)} \geq \inf\left\{\|\nu_1+\nu_2\| \;\middle|\; \begin{array}{cc}\|\nu_1\|^2+\|\nu_2\|^2=1, & x_i\in\mathbb{B}_\delta(\hat{x}),\\ \nu_i\in N_{\Omega_i}^{\mathrm{P}}(x_i), & i=1,2\end{array}\right\} \geq \frac{1}{\kappa} \tag{5.39}$$

which shows (5.34). Now, for $\delta\in(0,\bar{\delta}]$, let $\kappa_u^{(\delta)}(\hat{x})\{\Omega_1,\Omega_2\}$ be the smallest κ such that (5.34) holds, which is equivalent to equality in the second inequality of (5.39). A short calculation then shows

$$\begin{aligned}\frac{1}{\kappa_u^{(\delta)}(\hat{x})\{\Omega_1,\Omega_2\}} &= \inf\left\{\|\nu_1+\nu_2\| \;\middle|\; \begin{array}{cc}\|\nu_1\|^2+\|\nu_2\|^2=1, & x_i\in\mathbb{B}_\delta(\hat{x}),\\ \nu_i\in N_{\Omega_i}^{\mathrm{P}}(x_i), & i=1,2\end{array}\right\}\\ &= \inf\left\{\sqrt{1+2\langle\nu_1,\nu_2\rangle} \;\middle|\; \begin{array}{cc}\|\nu_1\|^2+\|\nu_2\|^2=1, & x_i\in\mathbb{B}_\delta(\hat{x}),\\ \nu_i\in N_{\Omega_i}^{\mathrm{P}}(x_i), & i=1,2\end{array}\right\}\\ &\overset{(5.36)}{=} \sqrt{1-c_D^{(\delta)}(\hat{x})\{\Omega_1,\Omega_2\}}.\end{aligned}$$

□

Corollary 5.28:
Let $\{\Omega_1, \Omega_2\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \Omega_1 \cap \Omega_2$. The collection is uniformly regular at $\hat{x}$ if and only if there exists $\delta > 0$ such that $c_D^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\}$ given by (5.34) is strictly less then 1.

Remark 5.29*: Corollary 5.28 is an immediate consequence of Theorem 5.27. The Corollary also is a generalization of (Lewis et al., 2009, Theorem 5.16), where the authors showed that uniform regularity implies existence of a* $\delta > 0$ *such that* $c_D^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\} < 1$.
$\diamondsuit$

5.5. Norm Equivalence

The following statement is a quantified analogue to Corollary 5.15.

Corollary 5.30:
Let $\{\Omega_1, \Omega_2\}$ be a collection of closed, nonempty sets and let $\hat{x} \in \Omega_1 \cap \Omega_2$.

(a) Let the collection be *uniformly regular* at $\hat{x}$ for $\delta > 0$ and $\kappa_u^{(\delta)}(\hat{x})\,\{\Omega_2, \Omega_2\}$ the local regularity modulus (5.34). Then for all $x_i \in \mathbb{B}_\delta(\hat{x})$, $\nu_i \in N_{\Omega_i}(x_i)$ for $i = 1, 2$:

$$\max\{\|\nu_1\|, \|\nu_2\|\} \leq \kappa_u^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\}\,\|\nu_1 + \nu_2\|, \tag{5.40}$$
$$\|\nu_1\| + \|\nu_2\| \leq \sqrt{2}\kappa_u^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\}\,\|\nu_1 + \nu_2\|. \tag{5.41}$$

(b) Let the collection be *locally linearly regular* at $\hat{x}$ for $\delta > 0$ and $\kappa_l^{(\delta)}(\hat{x})\,\{\Omega_2, \Omega_2\}$ the local regularity modulus Then for all $x \in \mathbb{B}_\delta(\hat{x})$

$$d(x, \Omega_1 \cap \Omega_2) \leq \sqrt{2}\kappa_l^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\}\ \max\{d(x, \Omega_1), d(x, \Omega_2)\}, \tag{5.42}$$
$$d(x, \Omega_1 \cap \Omega_2) \leq \kappa_l^{(\delta)}(\hat{x})\,\{\Omega_1, \Omega_2\}\ (d(x, \Omega_1) + d(x, \Omega_2)). \tag{5.43}$$

Remark 5.31*: An observation that relates to equation* (5.41) *can be found in (Kruger and Thao, 2014, Remark 1).* $\diamondsuit$

5.6. Linear Regularity under Unions of Sets

Lemma 5.32 (linear regularity under unions, (Hesse et al., 2014a, Lemma 3.5))**:**
Let $\{\Omega_1, \Omega_2, \ldots, \Omega_m, \Omega_{m+1})$ be a collection of nonempty subsets of $\mathbb{E}$ with nonempty intersection.

Let $\hat{x} \in \left(\cap_{j=1}^{m+1} \Omega_j\right)$. Suppose that for any $j = 1, 2, \ldots, m$ that the collection $\{\Omega_j, \Omega_{m+1}\}$ is locally linearly regular with $\kappa_l^{(\delta)}(\hat{x})\,\{\Omega_j, \Omega_{m+1}\}$ local regularity modulus (5.35). Then the collection $\left\{\bigcup_{j=1}^m \Omega_j, \Omega_{m+1}\right\}$ is locally linearly regular at $\hat{x}$ with local modulus

$$\kappa_l^{(\delta)}(\hat{x}) \left\{ \bigcup_{j=1}^{m} \Omega_j, \Omega_{m+1} \right\} \leq \max_{j=1,\ldots,m} \{\kappa_l^{(\delta)}(\hat{x})\,\{\Omega_j, \Omega_{m+1}\}\}.$$

Proof. Denote $\tilde{\Omega} := \bigcup_{j=1}^m \Omega_j$. Let $\overline{\kappa} \geq \max_{j=1,\ldots,m}\{\kappa_l^{(\delta)}(\hat{x})\,\{\Omega_j, \Omega_{m+1}\}\}$ and note that for all $x \in \mathbb{B}_\delta(\hat{x})$ we have

$$\begin{aligned} d\left(x, \tilde{\Omega} \cap \Omega_{m+1}\right) &= \min_{j=1,\ldots,m} \{\, d\left(x, \Omega_j \cap \Omega_{m+1}\right)\} \\ &\overset{(5.35)}{\leq} \min_{j=1,\ldots,m} \left\{ \kappa_l^{(\delta)}(\hat{x})\,\{\Omega_j, \Omega_{m+1}\} \sqrt{d^2\left(x, \Omega_j\right) + d^2\left(x, \Omega_{m+1}\right)} \right\} \\ &\leq \overline{\kappa} \sqrt{\min_{j=1,\ldots,m} d^2\left(x, \Omega_j\right) + d^2\left(x, \Omega_{m+1}\right)} \\ &\leq \overline{\kappa} \sqrt{d^2\left(x, \tilde{\Omega}\right) + d^2\left(x, \Omega_{m+1}\right)} \end{aligned}$$

This completes the proof. □

5.7. Restricted Regularity

Definition 5.33 (restricted uniform regularity)**:**
Let $\hat{x} \in \Omega_1 \cap \Omega_2$. Let $\mathbf{L_0}$ be a subspace of $\mathbb{E}$ and let $\mathbf{L} := \mathbf{L_0} + \hat{x}$, *i.e.,* an affine subspace parallel to $\mathbf{L_0}$. Define $\tilde{\Omega}_1 := (\Omega_1 - \hat{x})|_{\mathbf{L_0}}$, $\tilde{\Omega}_2 := (\Omega_2 - \hat{x})|_{\mathbf{L_0}}$.

Let $\tilde{\kappa}_u^{(\delta)}(\hat{x}) \left\{\tilde{\Omega}_1, \tilde{\Omega}_2\right\} := \kappa_u^{(\delta)}(\hat{x}) \left\{\tilde{\Omega}_1, \tilde{\Omega}_2\right\}\Big|_{\mathbf{L_0}}$, *i.e.,* the local modulus of uniform regularity with respect to the Euclidean space $\mathbf{L_0}$.

On $\mathbb{E}$ we define the *restricted modulus of uniform regularity* with respect to the affine subspace $\mathbf{L}$

$$\kappa_u^{(\delta,\mathbf{L})}(\hat{x})\,\{\Omega_1, \Omega_2\} \;:=\; \tilde{\kappa}_u^{(\delta)}(\hat{x}) \left\{\tilde{\Omega}_1, \tilde{\Omega}_2\right\}. \tag{5.44}$$

Furthermore let

$$c_D^{(\delta,\mathbf{L})}(\hat{x})\,\{\Omega_1, \Omega_2\} := \sup \left\{ -2 \frac{\langle \nu_2, \nu_2 \rangle}{\|\nu_1\|^2 + \|\nu_2\|^2} \;\middle|\; \begin{array}{c} \nu_i \in N^{\mathrm{P}}_{\Omega_i}(x_i) \cap \mathbf{L_0} \\ x_i \in \mathbb{B}_\delta(\hat{x}) \end{array} ,\ i = 1, 2 \right\}. \tag{5.45}$$

Note by definition there is a $\delta > 0$ such that $\kappa_u^{(\delta,\mathbf{L})}(\hat{x})\,\{\Omega_1, \Omega_2\}$ is finite if and only if the collection $\left\{\tilde{\Omega}_1, \tilde{\Omega}_2\right\}$ is uniformly regular with respect to the Euclidean subspace $\mathbf{L_0} \subset \mathbb{E}$.

Corollary 5.34 (restricted transversality condition)**:**
The collection $\{(\Omega_1 - \hat{x}) \cap \mathbf{L_0}, (\Omega_2 - \hat{x}) \cap \mathbf{L_0}\}$ is uniformly regular at $\hat{x}$ with respect to the Euclidean subspace $\mathbf{L_0} \subset \mathbb{E}$ if and only if

$$N_{\Omega_1}(\hat{x}) \cap N_{\Omega_2}(\hat{x}) \cap \mathbf{L_0} = \{0\}. \tag{5.46}$$

Under that assumption

$$\frac{1}{\kappa_u^{(\delta,\mathbf{L})}(\hat{x})\{\Omega_1, \Omega_2\}} = \inf\left\{ \|\nu_1 + \nu_2\| \;\middle|\; \begin{array}{ll} \|\nu_1\|^2 + \|\nu_2\|^2 = 1, & x_i \in \mathbb{B}_\delta(\hat{x}), \\ \nu_i \in N^{\mathrm{P}}_{\Omega_i}(x_i) \cap \mathbf{L_0}, & i = 1,2 \end{array} \right\} \tag{5.47}$$

$$\kappa_u^{(\delta,\mathbf{L})}(\hat{x})\{\Omega_1, \Omega_2\} = \frac{1}{\sqrt{1 - c_D^{(\delta,\mathbf{L})}(\hat{x})\{\Omega_1, \Omega_2\}}}, \tag{5.48}$$

where $c_D^{(\delta,\mathbf{L})}(\hat{x})\{\Omega_1, \Omega_2\}$ is given by (5.45)

Proof. By shifting to the origin and restricting to the euclidean subspace $\mathbf{L_0}$ the first part follows by (5.6). Analogously the second part of the Corollary follows by Theorem 5.27. □

Corollary 5.35 (restricted uniform regularity and Friedrichs angle)**:**
Let N, M be affine subspaces of $\mathbb{E}$ such that $N \cap M \neq \emptyset$ and let $\hat{x} \in N \cap M$. Let $\mathbf{A}$ be the linear subspaces parallel to N and $\mathbf{B}$ be the linear subspace parallel to M and define the linear subspace $\mathbf{L_0}$ by $\mathbf{L_0} := N - M = \mathbf{A} + \mathbf{B}$ and the affine subspace $\mathbf{L} := \hat{x} + \mathbf{L_0} = \operatorname{aff}(N \cup M)$. Then for all $\delta > 0$

$$c_D^{(\delta,\mathbf{L})}(\hat{x})\{N, M\} = c_F(\mathbf{A}, \mathbf{B}). \tag{5.49}$$

As a consequence any two linear subspaces $\mathbf{A}, \mathbf{B}$ of an euclidean space are uniformly regular with respect to the subspace $\mathbf{A} + \mathbf{B}$.

Proof. Note first that for any $x \in \mathbf{A}$, $y \in \mathbf{B}$ one has $N_{\mathbf{A}}(x) = \mathbf{A}^\perp$ and $N_{\mathbf{B}}(y) = \mathbf{B}^\perp$. Furthermore observe that $\left(\mathbf{A}^\perp \cap \mathbf{B}^\perp\right)^\perp = \mathbf{A} + \mathbf{B} = \mathbf{L_0}$. Then by Theorem 5.2 equation (5.3)

$$\begin{aligned}
c_F(\mathbf{A}, \mathbf{B}) &= c_F\left(\mathbf{A}^\perp, \mathbf{B}^\perp\right) \\
&= c_F\left(\mathbf{A}^\perp, -\mathbf{B}^\perp\right) \\
&= \max\left\{ \langle \nu_{\mathbf{A}}, -\nu_{\mathbf{B}} \rangle \;\middle|\; \begin{array}{lll} \nu_{\mathbf{A}} & \in & \mathbf{A}^\perp \cap \left(\mathbf{A}^\perp \cap \mathbf{B}^\perp\right)^\perp \cap \mathbb{B} \\ \nu_{\mathbf{B}} & \in & \mathbf{B}^\perp \cap \left(\mathbf{A}^\perp \cap \mathbf{B}^\perp\right)^\perp \cap \mathbb{B} \end{array} \right\} \\
&= \max\left\{ -\langle \nu_{\mathbf{A}}, \nu_{\mathbf{B}} \rangle \;\middle|\; \begin{array}{lll} \nu_{\mathbf{A}} & \in & \mathbf{A}^\perp \cap \mathbf{L_0} \cap \mathbb{B} \\ \nu_{\mathbf{B}} & \in & \mathbf{B}^\perp \cap \mathbf{L_0} \cap \mathbb{B} \end{array} \right\} \\
&= \max\left\{ -\frac{2\langle \nu_{\mathbf{A}}, \nu_{\mathbf{B}} \rangle}{\|\nu_{\mathbf{A}}\|^2 + \|\nu_B\|^2} \;\middle|\; \begin{array}{lll} \nu_{\mathbf{A}} & \in & \mathbf{A}^\perp \cap \mathbf{L_0} \\ \nu_{\mathbf{B}} & \in & \mathbf{B}^\perp \cap \mathbf{L_0} \end{array} \right\} \\
&= c_D^{(\delta,\mathbf{L})}(\hat{x})\{N, M\}.
\end{aligned}$$

□

Corollary 5.36 (modulus of linear regularity)**:**
Let N, M be affine subspaces of $\mathbb{E}$ such that $N \cap M \neq \emptyset$ and let $\hat{x} \in N \cap M$. Let $\mathbf{A}$ be the linear subspaces parallel to N and $\mathbf{B}$ be the linear subspace parallel to M and define the linear subspace $\mathbf{L_0}$ by $\mathbf{L_0} := N - M = \mathbf{A} + \mathbf{B}$ and the affine subspace $\mathbf{L} := \hat{x} + \mathbf{L_0} = \operatorname{aff}(N \cup M)$. Then for all $\delta > 0$

$$\kappa_l^{(\delta)}(\hat{x})\{N, M\} \leq \frac{1}{\sqrt{1 - c_F(\mathbf{A}, \mathbf{B})}}. \tag{5.50}$$

Proof. Define $\tilde{\Phi} : \mathbf{L_0} \rightrightarrows \mathbf{L_0} \times \mathbf{L_0}$, $\tilde{\Phi}(\tilde{x}) := (\mathbf{A} - \tilde{x}) \times (\mathbf{B} - \tilde{x})$. By Corollary 5.35 $\{\mathbf{A}, \mathbf{B}\}$ is always uniformly regular with respect to $\mathbf{L_0}$ and therefor the modulus of metric regularity $\operatorname{reg}\tilde{\Phi}(0|0)$ is finite and does not depend on specific neighborhoods. Furthermore by definition of metric regularity and metric subregularity one has for any $\delta > 0$

$$\operatorname{subreg}\tilde{\Phi}(0|0) \leq \operatorname{reg}\tilde{\Phi}(0|0) = \kappa_u^{(\delta,\mathbf{L})}(\hat{x})\{N, M\}.$$

Hence, by (5.30) and the fact that $c_D|_{\mathbf{L_0}} = c_F(\mathbf{A}, \mathbf{B})$ one achieves

$$\operatorname{subreg}\tilde{\Phi}(0|0) \leq \frac{1}{\sqrt{1 - c_F(\mathbf{A}, \mathbf{B})}}.$$

This then shows that for all $x \in \mathbf{L}$

$$d^2(x, \mathbf{A} \cap \mathbf{B}) \leq \frac{1}{1 - c_F(\mathbf{A}, \mathbf{B})}\Big(d^2(x, \mathbf{A}) + d^2(x, \mathbf{B})\Big). \tag{5.51}$$

If $\mathbf{L_0} = \mathbb{E}$ the proof is complete. Assume $\mathbf{L_0} \subsetneq \mathbb{E}$ and let $x \in \mathbb{E}$, then

$$\begin{aligned}
& d^2(x, \mathbf{A} \cap \mathbf{B}) \\
\overset{(2.7)}{=} \; & \|x - P_{\mathbf{L}}x\|^2 + \|P_{\mathbf{L}}x - P_{\mathbf{A}\cap\mathbf{B}}x\|^2 \\
\overset{(5.51)}{\leq} \; & \|x - P_{\mathbf{L}}x\|^2 + \frac{1}{1 - c_F(\mathbf{A}, \mathbf{B})}\Big(d^2(P_{\mathbf{L}}x, \mathbf{A}) + d^2(P_{\mathbf{L}}x, \mathbf{B})\Big) \\
\leq \; & \frac{1}{1 - c_F(\mathbf{A}, \mathbf{B})}\Big(\|x - P_{\mathbf{L}}x\|^2 + d^2(P_{\mathbf{L}}x, \mathbf{A}) + \|x - P_{\mathbf{L}}x\|^2 + d^2(P_{\mathbf{L}}x, \mathbf{B})\Big) \\
= \; & \frac{1}{1 - c_F(\mathbf{A}, \mathbf{B})}\Big(d^2(x, \mathbf{A}) + d^2(x, \mathbf{B})\Big).
\end{aligned}$$

This completes the proof. □

Remark 5.37*: The last Corollary is also a restatement of the fact, that on Euclidean spaces, every two closed subspaces are linearly regular.* ♢

5.8. Other Existing Notions of Regularity

Remark 5.38*: Another concept that relates to the last corollary can be found in (Bauschke et al., 2013d; Bauschke et al., 2013c) and will be discussed in the following* ◊

Definition 5.39 ((Bauschke et al., 2013d; Bauschke et al., 2013c))**:**
The *S-restricted proximal normal cone* to a set $\Omega \subset \mathbb{E}$ at a point $\bar{x} \in \Omega$ is given by

$$N_{\Omega}^{S}(\bar{x}) \quad := \quad \operatorname{cone}\left((S \cap P_{\Omega}^{-1}\bar{x}) - \bar{x}\right) = \operatorname{cone}\left((S - \bar{x}) \cap (P_{\Omega}^{-1}\bar{x} - \bar{x})\right). \tag{5.52}$$

for $S = \mathbb{E}$ this reduces to the proximal normal cone $N_{\Omega}^{\mathrm{P}}(\bar{x})$ [Compare (3.1)].

Definition 5.40 (CQ-number, (Bauschke et al., 2013d; Bauschke et al., 2013c))**:**
Let $\mathbf{A}, \widetilde{\mathbf{A}}, \mathbf{B}, \widetilde{\mathbf{B}}$ be nonempty subsets of $\mathbb{E}$ and let $\hat{x} \in \mathbb{E}$ and $\delta \geq 0$. The *CQ-number* at $\hat{x}$ associated with $(\mathbf{A}, \widetilde{\mathbf{A}}, \mathbf{B}, \widetilde{\mathbf{B}})$ and δ is

$$\Theta_{\delta}(\mathbf{A}, \widetilde{\mathbf{A}}, \mathbf{B}, \widetilde{\mathbf{B}}) := \sup\left\{ \langle u, v\rangle \,\middle|\, \begin{array}{lll} u \in & N_{\mathbf{A}}^{\widetilde{\mathbf{B}}}(a) \cap \mathbb{B}, & a \in \mathbb{B}_{\delta}(\hat{x}) \\ v \in & -\; N_{\mathbf{B}}^{\widetilde{\mathbf{A}}}(b) \cap \mathbb{B}, & b \in \mathbb{B}_{\delta}(\hat{x}) \end{array} \right\} \tag{5.53}$$

where $N_{\mathbf{A}}^{\mathbf{B}}(a)$ is the *B-restricted proximal normal cone* of $\mathbf{A}$ at a (5.52).

Remark 5.41*: Note that*

$$\Theta_{\delta}(A, B, \mathbb{E}, \mathbb{E}) = c_D^{(\delta)}(\hat{x})\,\{A, B\} \tag{5.54}$$

and

$$\Theta_{\delta}(A, B, \mathbf{L}, \mathbf{L}) = c_D^{(\delta,\mathbf{L})}(\hat{x})\,\{A, B\} \tag{5.55}$$

but

$$\Theta_{\delta}(A, B, \operatorname{aff}(A \cup B), \operatorname{aff}(A \cup B)) \neq \Theta_{\delta}(A, B, A, B). \tag{5.56}$$

◊

Example 5.42*: Let* $\mathbb{E} = \mathbb{R}^2$*,*

$$\begin{aligned} \mathbf{A} \quad &:= \quad \left\{(x_1, x_2) \subset \mathbb{R}^2 \,\middle|\, x_1 \leq |x_2|\right\} && (5.57) \\ \mathbf{B} \quad &:= \quad \left\{(x_1, x_2) \subset \mathbb{R}^2 \,\middle|\, x_1 \leq 0\right\} && (5.58) \end{aligned}$$

Then for any $\delta > 0$

$$\begin{aligned} c_D^{(\delta)}(0)\,\{\mathbf{A}, \mathbf{B}\} \quad &= \quad \Theta_{\delta}(\mathbf{A}, \mathbf{B}, \operatorname{aff}(\mathbf{A} \cup \mathbf{B}), \operatorname{aff}(\mathbf{A} \cup \mathbf{B})) = 1, && (5.59) \\ \Theta_{\delta}(\mathbf{A}, \mathbf{B}, \mathbf{A}, \mathbf{B}) \quad &= \quad \sqrt{2}/2. && (5.60) \end{aligned}$$

△

6. Linear Convergence of Iterated $(\mathbf{S}, \varepsilon)$-Firmly Nonexpansive Operators

Our main goal in this section is to establish the weakest conditions we can (at the moment) under which the AP and Douglas–Rachford algorithms converge locally linearly.

In addition to regularity of the operators –Introduced in Chapter 4–, we need regularity of the fixed point sets of the operators. This is developed next.

The general framework is not restricted to projection-type algorithms. The main results apply for a general class of fixed point operators, yielding *local attractivity* of mappings based on *fixed point properties.*

Some of the results in this chapter were published in (Hesse and Luke, 2013), as well as the basic ideas this chapter is based on.

Despite its simplicity, the following Lemma is one of our fundamental tools.

Lemma 6.1 ((Hesse and Luke, 2013, Lemma 3.1))**:**
Let $\mathbf{D} \subset \mathbb{E}$, let $\mathbf{S} \subset \operatorname{Fix}(T)$ be closed, $T : \mathbf{D} \rightrightarrows \mathbb{E}$ and $U \subset \mathbf{D}$. If

(a) T is $(\mathbf{S}, \varepsilon)$-firmly nonexpansive on U and

(b) for some $\lambda > 0$, T satisfies the *coercivity condition*

$$\begin{aligned} &\forall x \in U,\ \forall\ x_+ \in Tx : \\ &\|x - x_+\| \geq \lambda\ d(x, \mathbf{S}). \end{aligned} \tag{6.1}$$

Then

$$\begin{aligned} &\forall x \in U,\ \forall\ x_+ \in Tx : \\ &d(x_+, \mathbf{S}) \leq \sqrt{(1 + \varepsilon - \lambda^2)}\ d(x, \mathbf{S}). \end{aligned} \tag{6.2}$$

Proof. For $x \in U$ choose any $x_+ \in Tx$, and any $\bar{x} \in P_{\mathbf{S}}x$. Combining equations (6.1) and (4.2) yields

$$\begin{aligned} \|x_+ - \bar{x}\|^2 + (\lambda \|x - \bar{x}\|)^2 &\overset{(6.1)}{\leq} \|x_+ - \bar{x}\|^2 + \|x - x_+\|^2 \\ &\overset{(4.2)}{\leq} (1+\varepsilon)\|x - \bar{x}\|^2, \end{aligned}$$

which immediately yields

$$\|x_+ - \bar{x}\|^2 \leq (1+\varepsilon-\lambda^2)\|x - \bar{x}\|^2. \tag{6.3}$$

Since $\bar{x} \in \mathbf{S}$ by definition one has $d(x_+, \mathbf{S}) \leq \|x_+ - \bar{x}\|$. Inserting this in (6.3) and using the fact that $\|x - \bar{x}\| = d(x, \mathbf{S})$ then proves (6.2). □

6.1. Linear Convergence of Alternating Projections

In the case of the alternating projections operator, the connection between local linear regularity of the collection of sets and the coercivity of the operator with respect to the intersection is natural, as the next result shows.

Proposition 6.2 (coercivity of the projector)**:**
Let $\mathbf{A}, \mathbf{B}$ be nonempty and closed subsets of $\mathbb{E}$, $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A} \cap \mathbf{B}$ and let the collection $\mathbf{A}, \mathbf{B}$ be *locally linearly regular* at $\hat{x}$ for $\delta > 0$. One has

$$\begin{gathered} \forall x_+ \in P_{\mathbf{B}}x,\ \forall x \in \mathbf{A} \cap \mathbb{B}_\delta(\hat{x}) : \\ \|x - x_+\| \geq \frac{1}{\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\, d\left(x, \hat{\mathbf{S}}\right), \end{gathered}$$

where $\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}$ is the local regularity modulus.

Proof. By the definition of the distance and the projector one has, for $x \in \mathbf{A}$ and any $x_+ \in P_{\mathbf{B}}x$,

$$\begin{aligned} \|x - x_+\| &= d(x, \mathbf{B}) \\ &= \left(d^2(x, \mathbf{B}) + \underbrace{d^2(x, \mathbf{A})}_{=0} \right)^{\frac{1}{2}} \\ &\overset{(5.15)}{\geq} \frac{1}{\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\, d\left(x, \hat{\mathbf{S}}\right). \end{aligned}$$

□

Theorem 6.3 (projections onto a (ε, δ)-subregular set)**:**
Let $\mathbf{A}, \mathbf{B}$ be nonempty and closed subsets of $\mathbb{E}$ and let $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A} \cap \mathbf{B}$. If

(a) $\mathbf{B}$ is (ε, δ)-subregular at $\hat{x}$ with respect to $\hat{\mathbf{S}}$ and

(b) the collection $\{\mathbf{A}, \mathbf{B}\}$ is locally linearly regular at $\hat{x}$ on $\mathbb{B}_\delta(\hat{x})$

then

$$d\left(x_+, \hat{\mathbf{S}}\right) \leq \sqrt{1 + \tilde{\varepsilon} - \gamma^2}\; d\left(x, \hat{\mathbf{S}}\right), \quad \forall x_+ \in P_{\mathbf{B}}x,\ \forall x \in U \tag{6.4}$$

where $\gamma = 1/\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A}, \mathbf{B}\}$ with $\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A}, \mathbf{B}\}$ the regularity modulus on $\mathbb{B}_\delta(\hat{x})$, $\tilde{\varepsilon} = 2\varepsilon + 2\varepsilon^2$ and

$$U \subset \{x \in \mathbf{A} \cap \mathbb{B}_\delta(\hat{x}) \mid P_{\mathbf{B}}x \subset \mathbb{B}_\delta(\hat{x})\}. \tag{6.5}$$

Proof. Since $\mathbf{B}$ is (ε, δ)-subregular at $\hat{x}$ with respect to $\hat{\mathbf{S}}$ one can apply Theorem 4.5 to show that the projector $P_{\mathbf{B}}$ is $(\hat{\mathbf{S}}, 2\varepsilon + 2\varepsilon^2)$-firmly nonexpansive on U. Moreover, condition (b) and Proposition 6.2 yield

$$\|x_+ - x\| \geq \gamma\, d\left(x, \hat{\mathbf{S}}\right) \quad \forall x_+ \in P_{\mathbf{B}}x,\ \forall x \in U.$$

Combining (a) and (b) and applying Lemma 6.1 then gives

$$d\left(x_+, \hat{\mathbf{S}}\right) \leq \sqrt{1 + 2\tilde{\varepsilon} - \gamma^2}\, d\left(x, \hat{\mathbf{S}}\right), \quad \forall x_+ \in P_{\mathbf{B}}x,\ \forall x \in U.$$

□

Corollary 6.4 (projections onto a convex set, (Bauschke and Borwein, 1993, Corollary 3.14))**:**
Let $\mathbf{A}$ and $\mathbf{B}$ be nonempty, closed subsets of $\mathbb{E}$. If

(a) the collection $\{\mathbf{A}, \mathbf{B}\}$ is locally linearly regular at $\hat{x} \in \mathbf{A} \cap \mathbf{B}$ on $\mathbb{B}_\delta(\hat{x})$ with local regularity modulus $\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A}, \mathbf{B}\} > 0$ and

(b) $\mathbf{B}$ is convex

then

$$d\left(x_+, \hat{\mathbf{S}}\right) \leq \sqrt{1 - \gamma^2}\; d\left(x, \hat{\mathbf{S}}\right), \quad \forall x_+ \in P_{\mathbf{B}}x,\ \forall x \in \mathbf{A} \cap \mathbb{B}_\delta(\hat{x}) \tag{6.6}$$

where $\gamma = 1/\kappa$.

Proof. By convexity of $\mathbf{B}$ the projector $P_{\mathbf{B}}$ is nonexpansive and it follows that $P_{\mathbf{B}}x \in \mathbb{B}_\delta(\hat{x})$ for all $x \in \mathbb{B}_\delta(\hat{x})$. Convexity of $\mathbf{B}$ is equivalent to $\mathbf{B}$ beeing $(0, +\infty)$-regular and hence $\tilde{\varepsilon} = 0$ in Theorem 6.3. □

Theorem 6.5 (linear convergence of von Neumann sequences)**:**
Let $\mathbf{A}, \mathbf{B}$ be closed nonempty subsets of $\mathbb{E}$ and let the collection $\{\mathbf{A}, \mathbf{B}\}$ be *locally linearly regular* at $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A} \cap \mathbf{B}$ on $\mathbb{B}_\delta(\hat{x})$ with regularity modulus $\kappa > 0$. Define $\gamma := 1/\kappa_l^{(\delta)}(\hat{x})\,\{\mathbf{A}, \mathbf{B}\}$ and let $x_0 \in \mathbf{A}$. Generate the sequence $\{x_n\}_{n\in\mathbb{N}}$ by

$$x_{2n+1} \in P_{\mathbf{B}} x_{2n} \quad \text{and} \quad x_{2n+2} \in P_{\mathbf{A}} x_{2n+1} \quad \forall n = 0, 1, 2, \ldots. \tag{6.7}$$

(a) If $\mathbf{A}$ and $\mathbf{B}$ are $(\varepsilon, \delta)-$subregular at $\hat{x}$ with respect to $\hat{\mathbf{S}}$ and $\tilde{\varepsilon} := 2\varepsilon + 2\varepsilon^2 \le \gamma^2$, then

$$d\left(x_{2n+2}, \hat{\mathbf{S}}\right) \le (1 - \gamma^2 + \tilde{\varepsilon})\, d\left(x_{2n}, \hat{\mathbf{S}}\right) \quad \forall n = 0, 1, 2, \ldots$$

for all $x_0 \in \mathbb{B}_{\delta/2}(\hat{x}) \cap \mathbf{A}$.

(b) If $\mathbf{A}$ is $(\varepsilon, \delta)-$subregular with respect to $\hat{\mathbf{S}}$, $\mathbf{B}$ is convex and $\tilde{\varepsilon} := 2\varepsilon + 2\varepsilon^2 \le (2\gamma - \gamma^2)/(1 - \gamma^2)$, then

$$d\left(x_{2n+2}, \hat{\mathbf{S}}\right) \le \sqrt{1 - \gamma^2 + \tilde{\varepsilon}}\sqrt{1 - \gamma^2}\, d\left(x_{2n}, \hat{\mathbf{S}}\right) \quad \forall n = 0, 1, 2, \ldots,$$

for all $x_0 \in \mathbb{B}_{\delta/2}(\hat{x}) \cap \mathbf{A}$.

(c) If $\mathbf{A}$ and $\mathbf{B}$ are convex, then

$$d\left(x_{2n+2}, \hat{\mathbf{S}}\right) \le (1 - \gamma^2)\, d\left(x_{2n}, \hat{\mathbf{S}}\right) \quad \forall n = 0, 1, 2, \ldots$$

for all $x_0 \in \mathbb{B}_\delta(\hat{x}) \cap \mathbf{A}$.

Proof. (a) First one has to show that all iterates remain close to $\hat{x}$ for x_0 close to $\hat{x}$, that is, we have to show that all iterates remain in the set U defined by (6.5). Note that for any $x_0 \in \mathbb{B}_{\delta/2}(\hat{x})$, and $x_1 \in P_{\mathbf{B}} x_0$ one has

$$\|x_0 - x_1\| = d(x_0, \mathbf{B}) \le \|x_0 - \hat{x}\|.$$

since $\hat{x} \in \mathbf{B}$. Thus

$$\|x_1 - \hat{x}\| \le \|x_0 - x_1\| + \|x_0 - \hat{x}\| \le \|x_0 - \hat{x}\| + \|x_0 - \hat{x}\| \le \delta,$$

which shows that $P_{\mathbf{B}} x_0 \subset \mathbb{B}_\delta(\hat{x})$, $\forall x_0 \in \mathbb{B}_{\delta/2}(\hat{x})$. One can now apply Theorem 6.3 to conclude that

$$d\left(x_1, \hat{\mathbf{S}}\right) \le \sqrt{1 - \gamma^2 + \tilde{\varepsilon}}\; d\left(x_0, \hat{\mathbf{S}}\right).$$

The last equation then implies that $x_1 \in \mathbb{B}_{\delta/2}(\hat{x})$ as long as $\gamma^2 \ge \tilde{\varepsilon}$ and therefore the same argument can be applied to x_1 to conclude that

$$d\left(x_2, \hat{\mathbf{S}}\right) \;\le\; \sqrt{1 - \gamma^2 + \tilde{\varepsilon}}\; d\left(x_1, \hat{\mathbf{S}}\right). \tag{6.8}$$

Combining the last two equations, (a) then follow by induction.

(b) Applying Corollary 6.4 yields

$$d\left(x_1,\hat{\mathbf{S}}\right) \;\leq\; \sqrt{1-\gamma^2}\; d\left(x_0,\hat{\mathbf{S}}\right)$$

and analogous to (a) note that (6.8) is still valid for $\tilde{\varepsilon} \leq (2\gamma-\gamma^2)/(1-\gamma^2)$. With $\tilde{\varepsilon} \leq (2\gamma-\gamma^2)/(1-\gamma^2)$ it follows that

$$\begin{aligned}\sqrt{1-\gamma^2+\tilde{\varepsilon}}\sqrt{1-\gamma^2} &\leq \sqrt{1-\gamma^2+(2\gamma-\gamma^2)/(1-\gamma^2)}\sqrt{1-\gamma^2}\\ &\leq \sqrt{1-2\gamma+\gamma^2+(2\gamma-\gamma^2)}\\ &\leq 1\end{aligned}$$

and therefore (b) follows by induction.

(c) is an immediate consequence of Corollary 6.4. □

Corollary 6.6 (von Neumann sequences on subspaces)**:**
Let N, M be two affine subspaces with $N \cap M \neq \emptyset$. Let $\mathbf{A}$ be the subspace parallel to N and $\mathbf{B}$ be the subspace parallel to M and $c_F(\mathbf{A},\mathbf{B})$ the corresponding Friedrich angle (compare (5.1)). Then for any $x_0 \in \mathbb{E}$

$$d\left(x_{2n+2}, N\cap M\right) \leq \left(c_F(\mathbf{A},\mathbf{B})\right) d\left(x_{2n}, N\cap M\right) \quad \forall n = 0,1,2,\ldots.$$

Proof. By Corollary 5.36

$$\gamma = \frac{1}{\kappa_l^{(\delta)}(\hat{x})\left\{N,M\right\}} = \sqrt{1-c_F(\mathbf{A},\mathbf{B})}$$

and hence by Theorem 6.5 (c) for any $x_0 \in \mathbb{E}$.

$$d\left(x_{2n+2}, N\cap M\right) \leq \left(c_F(\mathbf{A},\mathbf{B})\right) d\left(x_{2n}, N\cap M\right) \quad \forall n = 0,1,2,\ldots.$$

□

Remark 6.7*: According to classical results (Deutsch and Hundal, 2006a; Deutsch and Hundal, 2006b; Deutsch and Hundal, 2008) the convergence rate of alternating projections on subspaces becomes* $(c_F(\mathbf{A},\mathbf{B}))^2$*. This indicates that the rate achieved in Corollary 6.6 is not optimal in the case of linear subspaces.* ◊

Corollary 6.8:
Let $\mathbf{A}$, $\mathbf{B}$ be closed, nonempty and super-regular. Let $\{\mathbf{A},\mathbf{B}\}$ be locally linearly regular at $\hat{x} \in \mathbf{A}\cap\mathbf{B}$. Then there is a $\delta > 0$ such that for all $x_0 \in \mathbb{B}_\delta(\hat{x}) \cap \mathbf{A}$ any *von Neumann sequence* generated by (1.15) converges with linear rate to $\mathbf{A}\cap\mathbf{B}$.

Proof. By local linear regularity, there is $\delta_l > 0$ such that $\kappa_l^{(\delta_l)}(\hat{x})\{\mathbf{A},\mathbf{B}\} < \infty$. Now, by super-regularity at $\hat{x}$, for any ε there exists $\delta_{\mathbf{A}}$, such that $\mathbf{A}$ is $(\varepsilon,\delta_{\mathbf{A}})$-subregular at $\hat{x}$. Respectively for any ε, $\mathbf{B}$ is $(\varepsilon,\delta_{\mathbf{B}})$-subregular at $\hat{x}$ for some $\delta_{\mathbf{B}}$. In other words, for $\kappa_l^{(\delta_l)}(\hat{x})\{\mathbf{A},\mathbf{B}\}$ determined by the regularity of the collection $\{\mathbf{A},\mathbf{B}\}$ at $\hat{x}$, we can always choose ε, such that the requirement of Theorem 6.5 (a), *i.e.,* $\tilde{\varepsilon} := 2\varepsilon + 2\varepsilon^2 < (1/\kappa_l^{(\delta_l)}(\hat{x})\{\mathbf{A},\mathbf{B}\})^2$, is satisfied on $\mathbb{B}_{(2\min\{\delta_{\mathbf{A}},\delta_{\mathbf{B}}\})}(\hat{x})$. This completes the proof of linear convergence on $\mathbb{B}_\delta(\hat{x})$. □

6.2. Linear Convergence of Douglas–Rachford

We now turn to the Douglas–Rachford algorithm. This algorithm is notoriously difficult to analyze and our results reflect this in considerably more circumscribed conditions than are required for the AP algorithm. There are also recent papers (Bauschke et al., 2013a), (Phan, 2014), (Bauschke and Noll, 2014) that complement some of the results of (Hesse and Luke, 2013), which results will be complemented at this point.

Nevertheless, to the best of our knowledge some of the following convergence results are the most general to date.

The first result gives sufficient conditions for the coercivity condition (6.1) to hold.

Lemma 6.9:
Let the collection of closed subsets $\mathbf{A},\mathbf{B}$ of $\mathbb{E}$ be uniformly regular at $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A}\cap\mathbf{B}$ on $\mathbb{B}_\delta(\hat{x})$ with constant, $\delta > 0$ and $\kappa_u^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\} > 0$ according to (5.34). Let

$$U \subset \left\{x \in \mathbb{B}_\delta(\hat{x}) \mid P_{\mathbf{B}}x \subset \mathbb{B}_\delta(\hat{x}),\ P_{\mathbf{A}}R_{\mathbf{B}}x \subset \mathbb{B}_\delta(\hat{x})\right\}. \tag{6.9}$$

Then $\mathrm{T_{DR}}$ satisfies

$$\begin{gathered}\forall\, x \in U,\ \forall x_+ \in \mathrm{T_{DR}}\,x: \\ \|x - x_+\| \geq \left(\frac{1}{\kappa_u^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\right)\max\left\{d\left(R_{\mathbf{B}}x,\hat{\mathbf{S}}\right), \tfrac{1}{\sqrt{10}}\,d\left(x,\hat{\mathbf{S}}\right)\right\}.\end{gathered} \tag{6.10}$$

Proof. Let $x \in U$ and choose any $x_+ \in \mathrm{T_{DR}}\,x$. For some $z \in P_{\mathbf{A}}(R_{\mathbf{B}}x)$ by the definition of the reflector (1.19) there exists $y \in P_{\mathbf{B}}x$ such that $z \in P_{\mathbf{A}}(2y - x)$ and we can write

$$x_+ = x + z - y.$$

By construction of U: $2(y - x) \in R_{\mathbf{B}}x \subset \mathbb{B}_\delta(\hat{x})$ and $z \in P_{\mathbf{A}}R_{\mathbf{B}}x \subset \mathbb{B}_\delta(\hat{x})$, and hence by uniform regularity, or more precisely, characterization (5.34), one achieves

$$\begin{aligned}\|x_+ - x\|^2 &= \|z-y\|^2 \\ &= \|\underbrace{z-(2y-x)}_{\in -N_{\mathbf{A}}(2y-x)} + \underbrace{y-x}_{\in -N_{\mathbf{B}}(y)}\|^2 \\ &\geq \left(\frac{1}{\kappa_u^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\right)^2 \left(\|z-(2y-x)\|^2 + \|y-x\|^2\right).\end{aligned} \tag{6.11}$$

Uniform regularity for given $\delta > 0$ implies local linear regularity and therefore

$$\begin{aligned}\|z-(2y-x)\|^2 + \|y-x\|^2 &= \|z-(2y-x)\|^2 + \|(2y-x)-y\|^2 \\ &\geq d^2(2y-x,\mathbf{A}) + d^2(2y-x,\mathbf{B}) \\ &\geq \left(\frac{1}{\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\right)^2 d^2\left(2y-x,\hat{\mathbf{S}}\right)\end{aligned}$$

Combining the last inequality and (6.11), we achieve the first part, that is

$$\|x_+ - x\|^2 \geq \left(\frac{1}{\kappa_u^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\,\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}}\right)^2 d^2\left(2y-x,\hat{\mathbf{S}}\right).$$

The triangle inequality shows

$$\begin{aligned}d(x,\mathbf{A}) &\leq \|x-(2y-x)\| + d(2y-x,\mathbf{A}) \\ &= 2\|y-x\| + \|(2y-x)-z\|.\end{aligned} \tag{6.12}$$

And by linear regularity (5.43) and the *Cauchy-Schwartz* inequality this then yields

$$\begin{aligned}d\left(x,\hat{\mathbf{S}}\right) &\overset{(5.43)}{\leq} \kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\left(d(x,\mathbf{B}) + d(x,\mathbf{A})\right) \\ &\overset{(6.12)}{\leq} \kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\left(3\|y-x\| + \|(2y-x)-z\|\right) \\ &\leq \kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\sqrt{\left(3\|y-x\| + \|(2y-x)-z\|\right)^2} \\ &\overset{\text{C.S.}}{\leq} \kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\sqrt{(3^2+1^2)\left(\|y-x\|^2 + \|(2y-x)-z\|^2\right)}.\end{aligned}$$

Combining the last inequality and inequality (6.11) we achieve

$$\|x_+ - x\|^2 \geq \frac{1}{10\left(\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}\right)^2\left(\kappa_u^{(\delta)}(\mathbf{A})\{\mathbf{B},\Omega_2\}\right)^2} d^2\left(x,\hat{\mathbf{S}}\right)$$

□

Remark 6.10*: Some of the ideas on some of the estimates of Lemma 6.9 origin in (Phan, 2014, Theorem 6.1) which itself was motivated by (Hesse and Luke, 2013, Lemma 3.14). The techniques are quite similar, however the main difference is the use of different* product space estimates *in the definition of linear regularity, which results in the relation $\mu\sqrt{2} = \kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}$ between the constant μ (Phan, 2014, Definition 2.11) and $\kappa_l^{(\delta)}(\hat{x})\{\mathbf{A},\mathbf{B}\}$. Compare for instance Corollary 5.30.* ◊

Lemma 6.9 with the additional assumption of (ε,δ)-subregularity of the nonconvex sets $\mathbf{A},\mathbf{B}$ yields local linear convergence of the Douglas–Rachford algorithm.

Theorem 6.11:
Let the collection $\{\mathbf{A},\mathbf{B}\}$ of closed subsets $\mathbf{A},\mathbf{B}$ of $\mathbb{E}$ be uniformly regular at $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A}\cap\mathbf{B}$ for $\delta > 0$ with local regularity modulus $\kappa_u^{(\delta)}(\hat{x})\,\{\mathbf{A},\mathbf{B}\}$. Suppose that $\mathbf{A}$, $\mathbf{B}$ are (ε,δ)-subregular at $\hat{x}$ with respect to $\hat{\mathbf{S}}$. Let $\tilde{\delta} := \delta/[2(1+2\varepsilon)]$, $\tilde{\varepsilon}$ be given by (4.12) and $\eta := 1/\left[10\left(\kappa_u^{(\delta)}(\hat{x})\,\{\mathbf{A},\mathbf{B}\}\,\kappa_l^{(\delta)}(\hat{x})\,\{\mathbf{A},\mathbf{B}\}\right)^2\right]$. Then $\mathrm{T_{DR}}$ satisfies

$$\begin{aligned}&\forall x \in \mathbb{B}_{\tilde{\delta}}(\hat{x}),\ \forall x_+ \in \mathrm{T_{DR}}\,x :\\ &d\left(x_+,\hat{\mathbf{S}}\right) \le \sqrt{1+\tilde{\varepsilon}-\eta}\ \ d\left(x,\hat{\mathbf{S}}\right).\end{aligned} \tag{6.13}$$

Proof. First one has to show requirement (6.9). Choose any $x_+ \in \mathrm{T_{DR}}\,x$. For some $z \in P_{\mathbf{A}}(R_{\mathbf{B}}x)$ by the definition of the reflector (1.19) there exists $y \in P_{\mathbf{B}}x$ such that $z \in P_{\mathbf{A}}(2y-x)$ and we can write

$$x_+ = x + z - y.$$

Note that by triangle inequality, the fact

$$\|z-(2y-x)\| \le\ d\,(2y-x,\mathbf{A}) \le \|(2y-x)-\hat{x}\|$$

and by (4.7) we achieve

$$\begin{aligned}\|z-\hat{x}\| &\le \|z-(2y-x)\| + \|(2y-x)-\hat{x}\|\\ &\le 2\,\|(2y-x)-\hat{x}\|\\ &\le 2\sqrt{1+4\varepsilon+4\varepsilon^2}\,\|(2y-x)-\hat{x}\|\\ &= 2(1+2\varepsilon))\,\|(2y-x)-\hat{x}\|\end{aligned}$$

and therefore $z \in \mathbb{B}_{\delta}(\hat{x})$, *i.e.*, $\mathbb{B}_{\tilde{\delta}}(\hat{x}) \subset U$ as required by (6.9). So by Lemma 6.9 the coercivity condition (6.1)

$$\begin{aligned}&\forall x \in \mathbb{B}_{\tilde{\delta}}(\hat{x}),\ \forall x_+ \in \mathrm{T_{DR}}\,x :\\ &\|x-x_+\| \ge \sqrt{\eta}\,d\left(x,\hat{\mathbf{S}}\right)\end{aligned}$$

is satisfied on $\mathbb{B}_{\tilde{\delta}}(\hat{x})$ for $\sqrt{\eta} = 1/\left[\sqrt{10}\kappa_u^{(\delta)}(\hat{x})\,\{\mathbf{A},\mathbf{B}\}\,\kappa_l^{(\delta)}(\hat{x})\,\{\mathbf{A},\mathbf{B}\}\right]$. Moreover, since $\mathbf{A}$ and $\mathbf{B}$ are (ε,δ)-subregular by Theorem 4.6 $\mathrm{T_{DR}}$ is $(\hat{\mathbf{S}},\tilde{\varepsilon})$-firmly nonexpansive with $\tilde{\varepsilon}$ given by (4.12), that is

$$\begin{aligned}&\forall x \in \mathbb{B}_{\tilde{\delta}}(\hat{x}),\ \forall x_+ \in \mathrm{T_{DR}}\,x,\ \forall \bar{x} \in \hat{\mathbf{S}} :\\ &\|x_+-\bar{x}\|^2 + \|x-x_+\|^2 \le (1+\tilde{\varepsilon})\,\|x-\bar{x}\|^2,\end{aligned}$$

Lemma 6.1 then applies to yield (6.13). □

We summarize the discussion on Douglas–Rachford with the following convergence result.

Theorem 6.12:
Assume $\mathbf{A}, \mathbf{B} \subset \mathbb{E}$ are closed and super-regular at $\hat{x} \in \hat{\mathbf{S}} := \mathbf{A} \cap \mathbf{B}$ and that the collection $\{\mathbf{A}, \mathbf{B}\}$ is uniformly regular at $\hat{x}$. Then there is a $\delta > 0$ such that for all $x_0 \in \mathbb{B}_\delta(\hat{x})$ the Douglas–Rachford algorithm converges to $\hat{\mathbf{S}}$ with a linear rate.

More precisely, there is a δ such that $\mathbf{A}$ is $(\varepsilon_{\mathbf{A}}^{(\delta)}, \delta)$-subregular at $\hat{x}$, $\mathbf{B}$ is $(\varepsilon_{\mathbf{B}}^{(\delta)}, \delta)$-subregular at $\hat{x}$,

$$\alpha_\delta := \sqrt{\frac{1 + (1 + 2\varepsilon_{\mathbf{A}})^2 (1 + 2\varepsilon_{\mathbf{B}})^2}{2} - \frac{1}{10 \left(\kappa_u^{(\delta)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\} \, \kappa_l^{(\delta)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\} \right)^2}} < 1 \qquad (6.14)$$

and

$$d\left(\mathrm{T}_{\mathrm{DR}}\, x_n, \hat{\mathbf{S}}\right) \leq \alpha_\delta \, d\left(x_n, \hat{\mathbf{S}}\right), \quad n \in \mathbb{N}. \qquad (6.15)$$

Proof. By uniform regularity at $\hat{x}$, and hence local linear regularity, there is $\delta_u > 0$ such that $\kappa_l^{(\delta_u)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\}, \kappa_u^{(\delta_u)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\} < \infty$. Now, by super-regularity at $\hat{x}$, for any $\varepsilon_{\mathbf{A}}$ there exists $\delta_{\mathbf{A}}$, such that $\mathbf{A}$ is $(\varepsilon_{\mathbf{A}}, \delta_{\mathbf{A}})$-subregular at $\hat{x}$. Respectively for any $\varepsilon_{\mathbf{B}}$, $\mathbf{B}$ is $(\varepsilon_{\mathbf{B}}, \delta_{\mathbf{B}})$-subregular at $\hat{x}$ for some $\delta_{\mathbf{B}}$. In other words, for $\kappa_u^{(\delta_u)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\}$ determined by the regularity of the collection $\{\mathbf{A}, \mathbf{B}\}$ at $\hat{x}$, we can always choose $\varepsilon_{\mathbf{A}}$, $\varepsilon_{\mathbf{B}}$ (generating corresponding $\delta_{\mathbf{A}}$, $\delta_{\mathbf{B}}$ radius) such that

$$\begin{aligned} \tilde{\varepsilon} &= 2\varepsilon_{\mathbf{A}}(1 + \varepsilon_{\mathbf{A}}) + 2\varepsilon_{\mathbf{B}}(1 + \varepsilon_{\mathbf{B}}) + 8\varepsilon_{\mathbf{A}}(1 + \varepsilon_{\mathbf{A}})\varepsilon_{\mathbf{B}}(1 + \varepsilon_{\mathbf{B}}) \\ &< \eta = 1/\left[10 \left(\kappa_u^{(\delta)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\} \, \kappa_l^{(\delta)}(\hat{x}) \{\mathbf{A}, \mathbf{B}\} \right)^2\right]. \end{aligned}$$

is satisfied on $\mathbb{B}_{\min\{\delta_{\mathbf{A}}, \delta_{\mathbf{B}}\}}(\hat{x})$. This is equivalent to

$$1 + \tilde{\varepsilon} \;=\; \frac{1 + (1 + 2\varepsilon_{\mathbf{A}})^2 (1 + 2\varepsilon_{\mathbf{B}})^2}{2} < 1 + \eta,$$

and hence $\alpha_{\min\{\delta_{\mathbf{A}}, \delta_{\mathbf{B}}\}} := \sqrt{1 + \tilde{\varepsilon} - \eta} < 1$. Then for $\delta := \min\{\delta_u, \delta_{\mathbf{A}}, \delta_{\mathbf{B}}\} / [2(1 + \varepsilon)]$, the requirements of Theorem 6.11 are satisfied on $\mathbb{B}_{2(1+\varepsilon)\delta}(\hat{x})$, which completes the proof of linear convergence on $\mathbb{B}_\delta(\hat{x})$. □

6.2.1. Douglas–Rachford on Subspaces

We finish this section with the fact that strong regularity of the intersection is *necessary*, not just sufficient for convergence of the iterates of the Douglas–Rachford algorithm to the intersection in the affine case.

Corollary 6.13:
Let N, M be two affine subspaces with $N \cap M \neq \emptyset$. Let $\mathbf{A}$ be the subspace parallel to N and $\mathbf{B}$ be the subspace parallel to M and $c_F(\mathbf{A}, \mathbf{B})$ the corresponding Friedrich angle (compare (5.1)).

(a) For any starting point $x_0 \in \mathbb{E}$ the Douglas–Rachford Algorithm converges with linear rate $\sqrt{c_F(\mathbf{A}, \mathbf{B})(2 - c_F(\mathbf{A}, \mathbf{B}))}$ to $\operatorname{Fix}(\mathrm{T}_{\mathrm{DR}}) = N \cap M + (N - M)^{\perp}$.

(b) If $A^{\perp} \cap B^{\perp} = \{0\}$ then the Douglas–Rachford Algorithm converges $\mathbf{A} \cap \mathbf{B}$ to for any starting point $x_0 \in \mathbb{E}$ $N \cap M$ with linear rate $\sqrt{c_F(\mathbf{A}, \mathbf{B})(2 - c_F(\mathbf{A}, \mathbf{B}))}$.

Proof. For $x_0 \in \mathbb{E}$ let x_n be a Douglas–Rachford sequence generated by (1.21), *i.e.,*

$$x_{n+1} \in \mathrm{T}_{\mathrm{DR}}\, x_n.$$

Define $\tilde{x}_n := P_{\mathbf{L}} x_n - \hat{x}$, where $\mathbf{L} = \operatorname{aff}(N \cup M)$. Note by (2.23) and (2.22) that $\tilde{x}_n$ is a Douglas–Rachford sequence on the euclidean space $\mathbf{L_0} := \mathbf{L} - \hat{x} = \mathbf{A} + \mathbf{B}$. Once we show convergence of $\tilde{x}_n$ to $\mathbf{A} \cap \mathbf{B}$ on $\mathbf{L_0}$, (a) follows by Proposition 2.23 (c) and (b) becomes the special case $\mathbf{L_0} = \mathbb{E}$.

To show that $\tilde{x}_n$ converges to $\mathbf{A} \cap \mathbf{B}$ note that by Corollary 5.35, $c_F(\mathbf{A}, \mathbf{B}) < 1$ and by Theorem 5.27 (5.37)

$$\begin{aligned}
\kappa_u^{(\delta, \mathbf{L})}(\hat{x})\{N, M\} &= \kappa_u^{(\delta)}(\hat{x})\{\mathbf{A}, \mathbf{B}\}\Big|_{\mathbf{L_0}} \\
&= \frac{1}{\sqrt{1 - c_D^{(\delta)}(\hat{x})\{\mathbf{A}, \mathbf{B}\}\Big|_{\mathbf{L_0}}}} \\
&= \frac{1}{\sqrt{1 - c_D^{(\delta, \mathbf{L})}(\hat{x})\{N, M\}}} \\
&= \frac{1}{\sqrt{1 - c_F(\mathbf{A}, \mathbf{B})}}.
\end{aligned}$$

Furthermore by Corollary 5.36

$$\kappa_l^{(\delta)}(\hat{x})\{N, M\} = \frac{1}{\sqrt{1 - c_F(\mathbf{A}, \mathbf{B})}}$$

and hence the coercivity condition (6.10) becomes

$$\begin{aligned}
\|\tilde{x}_n - \mathrm{T}_{\mathrm{DR}}\, \tilde{x}_n\|^2 &\geq (1 - c_F(\mathbf{A}, \mathbf{B}))^2\, d^2(R_{\mathbf{B}} \tilde{x}_n, \mathbf{A} \cap \mathbf{B}) \\
&\overset{(2.8)}{=} (1 - c_F(\mathbf{A}, \mathbf{B}))^2\, d^2(\tilde{x}_n, \mathbf{A} \cap \mathbf{B}). \\
&\overset{(2.8)}{=} \left(1 - 2c_F(\mathbf{A}, \mathbf{B}) + [c_F(\mathbf{A}, \mathbf{B})]^2\right) d^2(\tilde{x}_n, \mathbf{A} \cap \mathbf{B}).
\end{aligned}$$

The last equation combined with the firm nonexpansiveness of T_{DR} under the application of Theorem 6.1 then yields

$$d\left(\mathrm{T}_{\mathrm{DR}}\,\tilde{x}_n, \mathbf{A} \cap \mathbf{B}\right) \leq \sqrt{c_F\left(\mathbf{A}, \mathbf{B}\right)\left(2 - c_F\left(\mathbf{A}, \mathbf{B}\right)\right)}\, d\left(\tilde{x}_n, \mathbf{A} \cap \mathbf{B}\right),$$

and hence the proof is complete. □

Remark 6.14*: In (Bauschke et al., 2013a) the authors proof linear convergence of the Douglas–Rachford algorithm to* $\operatorname{Fix}\left(\mathrm{T}_{DR}\right)$ *in the case of linear subspaces on general Hilbert spaces. They achieve an optimal rate of convergence, which is given by the Friedrichs angle* $c_F\left(\mathbf{A}, \mathbf{B}\right)$*. This indicates, that the rate achieved in Corollary 6.13 are not optimal in the case of linear subspaces.* ♢

6.3. Conclusion on the Theory

Remark 6.15 (Severeness of Douglas–Rachford)*: Remark 1.15, Theorem 2.18 along with Remark 2.19 and Theorem 6.12 indicate that the Douglas–Rachford operator is in some sense* severe *to regularity of the collection* $\{\mathbf{A}, \mathbf{B}\}$ *and hence not likely to converge to* $\mathbf{A} \cap \mathbf{B}$*. However, there are several reasons that indicate, that this* severeness *in fact is not a disadvantage, but a benefit of this algorithmic scheme. We point the reader to chapter 7, that is the existence of several different techniques, either* (re-)establishing *well behavedness of the collections of sets or* regularization *the algorithmic schemes to overcome this difficulty. In fact we will see throughout the applications part at several points, that many algorithmic schemes can be reformulated as (regularized) Douglas–Rachford type algorithms, and that typically this schemes emerge in applications, where other algorithmic schemes get stuck in local minima, whilst Douglas–Rachford type algorithms* escape *(cf. Remark 2.18 (a)) local minima.* ♢

Part II.

Applications - Sparse Affine Feasibility and X-Ray Imaging

7. Regularization

In application one often deals with small changes in the model that can render the intersection empty, *i.e.,*

$$\mathbf{A} \cap \mathbf{B} = \emptyset. \tag{7.1}$$

Small changes are manifested by various factors during the collection and recording of measurement data as well as approximation errors caused by numerical approximations.

Especially the Douglas–Rachford method is known to be sensitive to small changes in the intersection of the constraint sets. There are several different strategies in applications and in the literature that deal with this kind of *ill-posedness*, that will be discussed within this section.

7.1. Projections onto Regularized Sets

Probably the most intuitive scheme to overcome noisy measurements is to incorporate a small perturbation into the feasibility problem formulation.

Definition 7.1 (Bregman-regularized sets)**:**
Let $\psi : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be strictly convex and differentiable on the interior of its domain. The *Bregman distance* is defined by

$$d_\psi\left(x, y\right) := \psi(x) - \psi(y) - \operatorname{grad}\psi(y)\left(x - y\right). \tag{7.2}$$

For $\gamma \geq 0$ the $\gamma-$*regularized set* is defined by

$$\Omega^{(\gamma)} := \left\{x \mid\ d_\psi\left(x, \Omega\right) \leq \gamma\right\}. \tag{7.3}$$

The γ-projector onto Ω is defined by

$$P_\Omega^{(\gamma)} x := \begin{cases} (1-\lambda)P_\Omega x + \lambda x & \text{if } d_\psi\left(x, \Omega\right) \geq \gamma \\ x & \text{if } d_\psi\left(x, \Omega\right) < \gamma \end{cases}, \tag{7.4}$$

where $\lambda := \frac{\gamma}{d_\psi(x,\Omega)}$.

Example 7.2: For $\psi = \frac{1}{2}\|\cdot\|^2$ *the Bregman distance corresponds to the squared Euclidean distance, which is an appropriate error measure, when dealing with* Gaussian noise.

For $\mathbb{E} = \mathbb{R}^n$ *and*

$$\psi(x) = \sum_{i=1}^{n} h(x_i) \quad \text{for } h(\lambda) := \begin{cases} \lambda \log \lambda - \lambda & \text{if } \lambda > 0 \\ 0 & \text{if } \lambda = 0 \\ \infty & \text{if } \lambda < 0 \end{cases} \tag{7.5}$$

The Bregman distance becomes

$$d_\psi(x, y) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} + y_i - x_i \tag{7.6}$$

which corresponds to the Kullback-Leibler divergence. *The Kullback-Leibler divergence is an appropriate error measure, when dealing with* Poisson noise. △

Corollary 7.3:
Let $d_\psi(x, y) := d(x, y) = \|x - y\|$ be the Euclidean distance function. For any nonempty and closed set Ω the $\gamma-$projector coincides with the projector onto the $\gamma-$regularized set Ω^γ, *i.e.,*

$$P_\Omega^{(\gamma)} x = P_{\Omega^{(\gamma)}} x, \quad \text{for all } x \in \mathbb{E}. \tag{7.7}$$

Proof. The first part follows by easy calculus.

To show the second part assume $d(x, \Omega) \geq \gamma$ (The case $d(x, \Omega) < \gamma$ is trivial). By definition of the projector

$$\bar{x}_\gamma \in P_{\Omega^\gamma} \quad \Leftrightarrow \quad \|x - \bar{x}_\gamma\| = d\left(x, \Omega^{(\gamma)}\right).$$

Furthermore definition of $\Omega^{(\gamma)}$ it holds that

$$d\left(x, \Omega^{(\gamma)}\right) = d(x, \Omega) - \gamma$$

and hence for any $\bar{x} \in P_\Omega x$

$$\|x - \bar{x}\| - \gamma = d\left(x, \Omega^{(\gamma)}\right).$$

Reformulating

$$\|x - \bar{x}\| - \gamma = (1 - \lambda)\|x - \bar{x}\| = \|x - \underbrace{((1 - \lambda)\bar{x} - \lambda x)}_{\in P_\Omega^{(\gamma)}}\|.$$

then yields the statement that $P_{\Omega^{(\gamma)}}$ and $P_\Omega^{(\gamma)}$ coincide. □

Definition 7.4 (relaxed feasibility problem)**:**
Let **A**, **B** nonempty and closed. The relaxed feasibility problem is given by

$$\text{Find } \bar{x} \in \mathbf{A} \cap \mathbf{B}^{(\gamma)}. \tag{7.8}$$

Remark 7.5*: An analysis of the convergence behavior of* Approximate *Alternating Projections onto (Bregman) regularized sets can be found in (Luke, 2012a). To be more accurate the setting in (Luke, 2012a) incorporates regularizing a set M, which is the preimage of a set N under a uniform linear transform U,* i.e., *$M = U^* N$. In that setting the set N is disturbed by Poisson noise. In terms of the notation used in this work the regularized set there becomes $U^* N^\gamma U$, which can be handled by application of Lemma 3.21.*

In case of the Douglas–Rachford Operator alternative strategies will be discussed in the following. ♢

7.2. Regularized Douglas–Rachford

Another approach to deal with the shortcoming of small perturbations in the model, as suggested in (Bauschke et al., 2004) (Luke, 2008), is to regularize the setting by instead applying a relaxation of the Douglas–Rachford algorithm known as "*RAAR*" in the literature. Again, this can be viewed as a kind of regularization and will be introduced in the following.

Definition 7.6 (Regularized Douglas–Rachford Algorithm/Relaxed Averaged Alternating Reflections)**:**
For two nonempty, closed sets $\mathbf{A}, \mathbf{B} \subset \mathcal{H}$ the mapping

$$\mathrm{T}_{\mathrm{DR}}^{(\lambda)} x \quad = \quad \lambda \, \mathrm{T}_{\mathrm{DR}} \, x + (1 - \lambda) \, P_{\mathbf{B}} x \tag{7.9}$$

is called *regularized Douglas-Rachford Operator* or *Relaxed Averaged Alternating Reflections Operator.*

For given initial $x_0 \in \mathcal{H}$ the sequence generated by

$$x_{n+1} = \mathrm{T}_{\mathrm{DR}}^{(\lambda)} x_n \tag{7.10}$$

is called *regularized Douglas–Rachford algorithm* or *Relaxed Averaged Alternating Reflections.*

Remark 7.7*: In the literature* (7.10) *is usually referred to as RAAR. RAAR was introduced in (Luke, 2005) and has proven to be an efficient regularization strategy to overcome difficulties if the set* **B** *is affected by small perturbations caused by noisy measurement data. A detailed analysis on RAAR can be found in (Luke, 2008).*

Theorem 4.4 indicates that the same convergence analysis of the operator T_{DR} *can also be extended to* $\mathrm{T}_{DR}^{(\lambda)}$. $\Diamond$

8. Sparse Affine Feasibility

This chapter summaries some of the results of (Hesse et al., 2014a), that are based on the convergence results of Chapter 6 and (Hesse and Luke, 2013).

We explicitly focus on the results that reflect the geometric interpretation of the local properties defined in Chapter 3. Basically, that is, the observation that the notion of (ε, δ)-subregularity covers the model of sparse affine feasibility naturally and hence one can apply the results of Chapter 6.

For a detailed survey, a more general discussion and motivation see (Hesse et al., 2014a), (Luke, 2013), (Beck and Teboulle, 2011) and the references therein.

8.1. A Short Introduction to Sparse Affine Feasibility

We consider the problem of sparsity optimization with affine constraints:

$$\text{minimize } \|x\|_0 \text{ subject to } Mx = p \tag{8.1}$$

where $m, n \in \mathbb{N}$, $m < n$, $M \in \mathbb{R}^{m \times n}$ is a real m−by−n matrix and $\|x\|_0 := \sum_{j=1}^{n} |\operatorname{sgn} x_j|$ is the number of nonzero entries of a real vector $x \in \mathbb{R}^n$ of dimension n. Given an a priori bound $s \in \mathbb{N}$ on the desired sparsity of the solution one can relax problem (8.1) to the feasibility problem

$$\text{find } \bar{x} \in \mathbf{A_s} \cap \mathbf{L}, \tag{8.2}$$

where $\mathbf{A_s} := \{x \in \mathbb{R}^n |\ \|x\|_0 \leq s\}$, and $\mathbf{L} := \{x \in \mathbb{R}^n |\ Mx = p\}$. The set $\mathbf{L}$ is an affine subspace, whilst $\mathbf{A_s}$ is a non-convex set. However the set $\mathbf{A_s}$ locally has a nice structure in the sense that one can explicitly calculate the projector onto the set.

We are now ready to apply the above general results to affine sparse feasibility. We begin with characterization of the regularity of the sets involved.

8.2. The Sparse Set

Definition 8.1:
For a fixed $s \leq n$ the set of s-sparse vectors , *i.e.,* the set of vectors with at most s

nonzero elements is given by

$$\mathbf{A_s} := \{x \in \mathbb{E}|\ \|x\|_0 \le s\}\,. \tag{8.3}$$

Theorem 8.2 ((Bauschke et al., 2013b, Equation (27d)))**:**
Define

$$\mathcal{J} := 2^{\{1,2,\ldots,n\}} \text{ and } \ \mathcal{J}_s := \{J \in \mathcal{J}|\ \#J = s\}\,. \tag{8.4}$$

The set $\mathbf{A_s}$ can be written as the union of all subspaces indexed by $J \in \mathcal{J}_s$

$$\mathbf{A_s} = \bigcup_{J\in\mathcal{J}_s} A_J, \tag{8.5}$$

where $A_J := \operatorname{span}\{e_i|\ i \in J\}$ and e_i is the $i-$th standard unit vector in $\mathbb{R}^n$.

Proposition 8.3 ((Bauschke et al., 2013b, Proposition 3.6))**:**
The projector onto $\mathbf{A_s}$ is given by

$$P_{\mathbf{A_s}}(x) = \bigcup_{J\in C_s(x)} P_{A_J}x, \tag{8.6}$$

where for $x \in \mathbb{R}^n$

$$C_s(x) := \left\{J \in \mathcal{J}_s\middle|\ \min_{i\in J}|x_i| \ge \max_{i\notin J}|x_i|\right\} \tag{8.7}$$

is the set of s largest coordinates in absolute value and

$$(P_{A_J}x)_i = \begin{cases} x_i, & i \in J, \\ 0, & i \notin J \end{cases}. \tag{8.8}$$

Theorem 8.4 (regularity of $\mathbf{A_s}$, (Hesse et al., 2014a, Theorem 3.4)):

Let

$$\begin{aligned} I : \mathbb{R}^n &\to \{1,\ldots,n\}, \\ x &\mapsto \{i \in \{1,\ldots,n\}|\ x_i \neq 0\}\,. \end{aligned} \tag{8.9}$$

At any point $\bar{x} \in \mathbf{A_s}\backslash\{0\}$ the set $\mathbf{A_s}$ is $(0,\delta)$-subregular at $\bar{x}$ for

$$\delta \in (0, \min\{|\bar{x}_j|\ |\ j \in I(\bar{x})\}).$$

On the other hand, the set $\mathbf{A_s}$ is not $(0,\delta)$-subregular at $\bar{x} \in \mathbf{A_s}\backslash\{0\}$ for any

$$\delta \ge \min\{|\bar{x}_j|\ |\ j \in I(\bar{x})\}).$$

In contrast, at 0 the set $\mathbf{A_s}$ is $(0,\infty)$-subregular.

Theorem 8.5 (regularity of $(\mathbf{A_s}, \mathbf{L})$):
Let $\mathbf{A_s}$ be defined by (8.3), let $\mathbf{L}$ be an affine subspace such that $\mathbf{A_s} \cap \mathbf{L} \neq \emptyset$. At any $\bar{x} \in \mathbf{A_s} \cap \mathbf{L}$ and for any $\delta \in (0, \min\{|\bar{x}_j|\ |\ j \in I(\bar{x})\})$ the collection $\{\mathbf{A_s}, \mathbf{L}\}$ is locally linearly regular on $\mathbb{B}_{\delta/2}(\bar{x})$ with local modulus of regularity $\overline{\kappa} = \max_{J\in\mathcal{J}_s, I(\bar{x})\subseteq J}\{\kappa_l^{(\delta)}(\bar{x})\,\{A_J, \mathbf{L}\}\}$ where $\kappa_l^{(\delta)}(\bar{x})\,\{A_J, \mathbf{L}\}$ is the local modulus of linear regularity (5.35) of the collection $\{A_J, \mathbf{L}\}$.

Proof. For any $\bar{x} \in \mathbf{A_s} \cap \mathbf{L}$ we have $\bar{x} \in A_J \cap \mathbf{L}$ for all $J \in \mathcal{J}_s$ with $I(\bar{x}) \subseteq J$ and thus $(A_J, \mathbf{L})$ is linearly regular (Bauschke and Borwein, 1996, Proposition 5.9 and Remark 5.10). Now let $\kappa_l^{(\delta)}(\bar{x})\{\mathbf{A_s}, \mathbf{L}\}$ be the local modulus of linear regularity (5.35) and define

$$\overline{\mathbf{A_s}} := \bigcup_{J \in \mathcal{J}_s,\ I(\bar{x}) \subseteq J} A_J.$$

Then by Lemma 5.32 the collection $\left(\overline{\mathbf{A_s}}, B\right)$ is linearly regular at $\bar{x}$ with modulus of regularity $\overline{\kappa} := \max_{J \in \mathcal{J}_s,\ I(\bar{x}) \subseteq J} \{\kappa_l^{(\delta)}(\bar{x})\{A_J, \mathbf{L}\}\}$.

□

Remark 8.6*: Theorem 8.5 is actually a slightly improved version of (Hesse et al., 2014a, Theorem 3.6). The main difference lies in the application of a equivalent definition of local linear regularity. Compare Definition 5.6, Definition 5.25 and Corollary 5.30* ♢

8.3. Local Linear Convergence of Alternating Projections

Theorem 8.7:
Let $\mathbf{A_s}$ be defined by (8.3), let $\mathbf{L}$ be an affine subspace and let $\bar{x} \in \mathbf{A_s} \cap \mathbf{L} \neq \emptyset$. Choose $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. For $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$ the alternating projection algorithm, *i.e.,* any sequence generated by

$$x_{n+1} \in \mathrm{T_{AP}}\, x_n = P_{\mathbf{A_s}} P_{\mathbf{L}} x_n,$$

converges locally linearly to the intersection $\mathbf{A_s} \cap \mathbf{L}$ with linear rate

$$\bar{c} = \max_{J \in \mathcal{J}_s, I(\bar{x}) \subseteq J} c_F\left(A_J, \mathbf{L}\right)$$

where $c_F\left(A_J, \mathbf{L}\right)$ is the cosine of the Friedrichs angle between A_J and $\mathbf{L}$, see (5.1).

Proof. We will check the requirements of Theorem 6.5 (b). First note that $\mathbf{L}$ is convex and $\mathbf{A_s}$ by Theorem 8.4 $(0, \delta)$-subregular at $\bar{x}$ for any $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$, hence $\tilde{\varepsilon} = 0$. By Theorem 8.5, *i.e.,* linear regularity of $\{\mathbf{A_s}, \mathbf{L}\}$ we achieve $1 - 1/\overline{\kappa}^2 < 1$ and hence linear convergence of any alternating projection sequence on $\mathbb{B}_{\delta/2}(\bar{x})$.

To get a explicit estimate of the linear rate note that by Corollary 5.36

$$\kappa_l^{(\delta)}(\hat{x})\{A_J, \mathbf{L}\} \leq \frac{1}{\sqrt{1 - c_F\left(A_J, \mathbf{L}\right)}}$$

for any $J \in \mathcal{J}_s$. Hence $(1 - 1/\left(\kappa_l^{(\delta)}(\hat{x})\{A_J, \mathbf{L}\}\right)^2 \leq c_F(A_J, \mathbf{L})$ for any $J \in \mathcal{J}_s$ which implies $1 - 1/\overline{\kappa}^2 \leq \overline{c} < 1$ and yields the requested estimate on the linear rate. □

Remark 8.8*: Theorem 8.7 was also shown in (Bauschke et al., 2013b, Theorem 3.19) using very different techniques. The approach taken in (Hesse et al., 2014a) is based on the local modulus of regularity* $\kappa_u^{(\delta)}(\bar{x})\{\mathbf{A_s}, \mathbf{L}\}$*, whilst the approach in (Bauschke et al., 2013b) is based on the Friedrichs angle* (5.1)*. The relation between this different techniques is not fully understood. Theorem 8.7 is an interesting result that establishes am interesting connection between the two different approaches. However convergence rate in (Bauschke et al., 2013b) is* $\overline{c}^2$ *rather than* $\overline{c}$*, which indicates that in some applications the framework used in this work is not optimal, which is the price of its more general nature. Other results relating the different concepts of regularity of collections and intersections of sets are given in Chapter 5. Of special interest in this context are for instance Corollary 5.35 and Corollary 5.36.* ♢

8.4. Local Linear Convergence of Douglas–Rachford

The convergence analysis on the Douglas–Rachford operator requires a broader discussion on the geometric properties of the set $\mathbf{A_s}$. At this point we just state one of the most recent and interesting results and refer the interested reader to (Hesse et al., 2014a, Section 4) for a broad discussion.

Theorem 8.9 ((Hesse et al., 2014a, Theorem 4.7))**:**
Let $\mathbf{A_s}$ be defined by (8.3), let $\mathbf{L}$ be an affine subspace and let $\bar{x} \in \mathbf{A_s} \cap \mathbf{L} \neq \emptyset$ with $\|\bar{x}\|_0 = s$. Choose $0 < \delta < \min\{|\bar{x}_j| \mid j \in I(\bar{x})\}$. For $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$ the corresponding Douglas–Rachford algorithm, *i.e.,* any sequence generated by

$$x_{n+1} \in \mathrm{T_{DR}}\, x_n = \frac{1}{2}\left(R_{\mathbf{A_s}} R_{\mathbf{L}} x_n + x_n\right),$$

converge with linear rate to $\mathrm{Fix}\,(\mathrm{T_{DR}})$. Moreover, for any $\hat{x} \in \mathrm{Fix}\,(\mathrm{T_{DR}}) \cap \mathbb{B}_{\delta/2}(\bar{x})$, we have $P_{\mathbf{L}}\hat{x} \in \mathbf{A_s} \cap \mathbf{L}$.

9. Diffraction Imaging

In this section we provide a short summary on scattering theory and *Rayleigh–Sommerfeld diffraction theory*. For a nice survey see (Luke et al., 2002). A detailed study on scattering theory can be found in (Kress and Colton, 1998). Detailed information on Coherent X-Ray Diffraction can be found in (Paganin, 2006), (Giewekemeyer, 2011).

Let $u : \mathbb{R}^3 \to \mathbb{C}$ be a complex valued function. For a partially differentiable function $u : \mathbb{R}^3 \to \mathbb{C}$ let $\operatorname{grad} u := (\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_1})$ be the *gradient* of u and for a partially differentiable vector field $v : \mathbb{C}^3 \to \mathbb{C}^3$ let $\operatorname{div} v := \sum_{i=1}^{3} \frac{\partial v_i}{\partial x_i}$ be the divergence. The Laplace Operator is defined by $\Delta u := \operatorname{div}(\operatorname{grad} u) = \sum_{i=1}^{3} \frac{\partial^2 u}{\partial x_i^2}$ for any two times partially differentiable function $u : \mathbb{R}^3 \to \mathbb{C}$.

9.1. Helmholtz Equation

Helmholtz equation

$$\Delta u + k^2 u = 0 \tag{9.1}$$

with positive constant wave number k

The *Sommerfeld Radiation condition*

$$\frac{\partial u}{\partial r} - \mathrm{i}ku = \mathcal{O}\left(\frac{1}{r}\right), \quad r = \|x\| \to \infty, \tag{9.2}$$

characterizes the behavior of outgoing solutions to the Helmholtz equation in great distance to the origin.

Definition 9.1:
Solutions to the Helmholtz equation that are defined all over $\mathbb{R}^3$ are called *entire* solutions. A Solution that fulfills the Sommerfeld Radiation condition (9.2) is called *radiating*.

Proposition 9.2:
The fundamental solution to the Helmholtz equation (9.1)

$$\Phi(x, y) = \frac{1}{4\pi} \frac{\mathrm{e}^{\mathrm{i}k\|x-y\|}}{\|x-y\|}. \tag{9.3}$$

The fundamental solution is a solution to the Helmholtz equation for all $x \in \mathbb{R}^3 \backslash \{y\}$ and it is radiating.

9.2. Greens Representation

9.2.1. Greens Theorem

$D \subset \mathbb{R}^3$ bounded volume with orientable boundary ∂D and *inward* normal ν. $u, v \in C^2(\overline{D})$.

$$\int_D (u\Delta v + \operatorname{grad} u \cdot \operatorname{grad} v)\, \mathrm{d}x = -\int_{\partial D} u \frac{\partial v}{\partial \nu}\, \mathrm{d}s. \tag{9.4}$$

Green's first theorem

$$\int_D (u\Delta v - v\Delta u)\, \mathrm{d}x = \int_{\partial D} -u \frac{\partial v}{\partial \nu} + v \frac{\partial u}{\partial \nu}\, \mathrm{d}s. \tag{9.5}$$

Green's second theorem

Proof. We define $F = u \operatorname{grad} v$

$$\begin{aligned}
\operatorname{div} F &= \operatorname{div}\, (u \operatorname{grad} v) \\
&= \operatorname{grad} u \cdot \operatorname{grad} v + u \operatorname{div} \operatorname{grad} v \\
&= \operatorname{grad} u \cdot \operatorname{grad} v + u\, \Delta v
\end{aligned}$$

Using *Gauß' Integral theorem*

$$\int_D \operatorname{div} F \,\mathrm{d}x = \int_{\partial D} F \cdot \nu \,\mathrm{d}s$$

we get (9.4). Interchanging of u and v und subtraction shows (9.5). □

9.2.2. Green's Formula

Theorem 9.3:
Let D be a bounded domain. For $u \in C^2(\overline{D})$ one has

$$\begin{aligned}
u(x) \quad = \quad & \int_{\partial D} \left(-\frac{\partial u}{\partial \nu}(y) \Phi(x, y) + u(y) \frac{\partial \Phi(x, y)}{\partial \nu(y)} \right) \mathrm{d}s(y) \\
& - \int_D \{\Delta u(y) + k^2 u(y)\} \Phi(x, y) \,\mathrm{d}y, \quad x \in D.
\end{aligned}$$

In particular, if u is a a solution to the Helmholtz equation in D we have

$$u(x) \quad = \int_{\partial D} \left(-\frac{\partial u}{\partial \nu}(y)\Phi(x,y) + u(y)\frac{\partial \Phi(x,y)}{\partial \nu(y)}\right) \mathrm{d}s(y), \quad x \in D. \tag{9.6}$$

Proof. We can apply *Green's second theorem* (9.5) to u and $\Phi(x,\cdot\,)$ in the area $\{y \in D : |x-y| > r\}$ and therefore get

$$\begin{aligned}&\int_D \{\Delta u(y) + k^2 u(y)\}\Phi(x,y)\,\mathrm{d}y \\ = &\int_{\partial D \cup \partial B(x;r)} \left(-u(y)\,\frac{\partial \Phi(x,y)}{\partial \nu(y)} + \frac{\partial u}{\partial \nu}(y)\,\Phi(x,y)\right) \mathrm{d}s(y),\end{aligned}$$

where ν is directed into this area. For $r \to 0$ using the mean value theorem we get

$$\lim_{r\to 0}\int_{|x-y|=r} \left\{-u(y)\frac{\partial \Phi(x,y)}{\partial \nu(y)} + \frac{\partial u}{\partial \nu}(y)\Phi(x,y)\right\} \mathrm{d}s(y) = u(x),$$

with the aid of

$$\Phi(x,y) = \mathcal{O}\left(\frac{1}{r}\right) \quad \text{and} \quad \operatorname{grad}_y\Phi(x,y) = \frac{1}{4\pi r^3}(x-y) + \mathcal{O}\left(\frac{1}{r}\right)$$

□

9.2.3. Green's Formula in a Half-Space

Let $\mathbb{T}$ be the x_1x_2-plane. To satisfy the Dirichlet boundary conditions

$$\begin{aligned} G = 0 \qquad & \text{on} \quad \mathbb{T} \\ \|x-y\|\left(\frac{\partial G}{\partial \nu} - \mathrm{i}kG\right) \to 0 \qquad & \text{as} \quad \|x-y\| \to \infty \end{aligned}$$

we define the field

$$G(x,y,x') := \Phi(x,y) - \Phi(x',y) \tag{9.7}$$

where the mirror point source x' is defined by the condition $\|x'-y\| = \|x-y\|$ for all $y \in \mathbb{T}$.

If $x' \notin D$ one can use Green's second theorem (9.5) and Green's formula (9.6) to establish

$$u(x) = \int_{\partial D} G(x,y,x')\frac{\partial u}{\partial \nu}(y) - u(y)\frac{\partial G(x,y,x')}{\partial \nu(y)}\,\mathrm{d}s(y) \quad x \in D. \tag{9.8}$$

9.3. Rayleigh-Sommerfeld Diffraction

Rayleigh-Sommerfeld diffraction theory is derived by considering the specific domain Let

$$D_r := \left\{ x \in \mathbb{R}^3 : \|x\| \leq r, x_3 \geq 0 \right\}.$$

Although the domain is depending on r we will just write $D = D_r$ The boundary ∂D consist of the hemisphere $\mathbb{E} = \{x : \|x\| = r, x_3 > 0\}$ and the disk $\mathbb{D} = \{x \in \mathbb{T} : \|x\| < r\}$.

Since we postulate that the unknown field u is only nonzero on a compact support $\mathbb{A} \subset \mathbb{D}$ centered at the origin and that it satisfies the Sommerfeld radiation condition (9.2) by passing the limit $r \to \infty$ Green's formula (9.8) reduces to

$$u(x) = \int_{\mathbb{A}} -u(y) \frac{\partial G(x, x', y)}{\partial \nu(y)} \, \mathrm{d}s(y) \quad x \in D \tag{9.9}$$

Let $\mathbb{I} \subset D$ be a screen parallel to $\mathbb{T}$. For $x \in \mathbb{I}$, $y \in \mathbb{A}$ we assume $\|x - y\| \gg \lambda$ and approximate

$$\begin{aligned} \frac{\partial G(x, x', y)}{\partial \nu(y)} &= 2 \frac{\exp(\,\mathrm{i}k \|x - y\|)}{4\pi \|x - y\|} \left(\mathrm{i}k - \frac{1}{\|x - y\|} \right) \alpha(\nu, y - x) \\ &\approx \underbrace{\frac{\exp(\,\mathrm{i}k \|x - y\|)}{-\,\mathrm{i}\lambda \|x - y\|} \alpha(\nu, y - x)}_{=:-h(x,y)} \end{aligned}$$

where $\alpha(x, y)$ is the cosine of the angle between two points x, y

$$\alpha(x, y) := \frac{x \cdot y}{\|x\| \, \|y\|}$$

and $\lambda = 2\pi/k$ is the wavelength.

Substituting this in Equation (9.9) yields the following mathematical formulation of *Huygen's principle*

$$u(x) \approx \int_{\mathbb{A}} u(y) h(x, y) \, \mathrm{d}s(y) \quad x \in \mathbb{I}. \tag{9.10}$$

9.3.1. Fresnel Approximation

For any point $x \in \mathbb{I}$ x_3 is the distance between $\mathbb{A}$ and $\mathbb{I}$. We assume that two points $x \in \mathbb{I}$ and $y \in \mathbb{A}$ satisfy the condition

$$\|(x_1 - y_1, x_2 - y_2, 0)\| \ll x_3. \tag{9.11}$$

Therefore we have $|x_1 - y_1| \ll x_3$ and $|x_2 - y_2| \ll x_3$ and can use in the binominal expansion to obtain

$$\|x - y\| \approx x_3 \left(1 + \frac{1}{2x_3^2}(x_1 - y_1)^2 + \frac{1}{2x_3^2}(x_2 - y_2)^2\right) \tag{9.12}$$

Using $\alpha(\nu, x - y) \approx 1$ and (9.12) and by neglecting the quadratics in the denominater the kernel h reduces to the *Fresnel* kernel

$$h_{Fre} := \frac{\exp(\,\mathrm{i}kx_3)}{\mathrm{i}\lambda x_3} \exp\left(\frac{\mathrm{i}k}{2x_3}\left((x_1 - y_1)^2 + (x_2 - y_2)^2\right)\right). \tag{9.13}$$

This kernel satisfies what is known as the *parabolic* wave equation

$$\left(\frac{\partial}{\partial x_3} - \frac{\mathrm{i}}{2k}\Delta_{\mathbb{T}} - \mathrm{i}k\right) h_{Fre} \tag{9.14}$$

where $\Delta_{\mathbb{T}}$ is the Laplacian in the $\mathbb{T}$ plane, i.e $\Delta_{\mathbb{T}} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$. Using the Fresnel kernel in (9.10) we obtain the Fresnel diffraction field

$$u_{Fre}(x) = \int_{\mathbb{A}} u(y) h_{Fre}(x, y)\, \mathrm{d}y_1\, \mathrm{d}y_2. \tag{9.15}$$

Interchanging the Integral- and the Differential-Operator show that this field also satisfies (9.14)

With the notation $h_{Fre}(x - y) := h_{Fre}(x - y, 0) = h_{Fre}(x, y)$ one can rewrite this as

$$u_{Fre}(x) = u(x) *_{\perp} h(x) \tag{9.16}$$

9.3.2. Fraunhofer Approximation

If $\mathbb{A}$ is small compared to $\mathbb{I}$ we can approximate

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 \approx x_1^2 + x_2^2 - 2(x_1 y_2 + x_2 y_2)$$

and the Fresnel kernel (9.13) reduces to the *Fraunhofer* approximation

$$h_{Fra}(x, y) = \frac{\exp(\,\mathrm{i}kx_3)}{\mathrm{i}\lambda x_3} \exp\left(\frac{\mathrm{i}k}{2x_3}(x_1^2 + x_2^2)\right) \exp\left(\frac{\mathrm{i}k}{x_3}(x_1 y_1 + x_2 y_2)\right). \tag{9.17}$$

The *Fraunhofer transform* of a field u across an aperture $\mathbb{A}$ is therefore given by

$$u_{Fra}(x) = \int_{\mathbb{A}} u(y) h_{Fra}(x, y)\, \mathrm{d}y_1\, \mathrm{d}y_2. \tag{9.18}$$

We rewrite

$$\begin{aligned} u_{Fra}(x) &= c(x)\int_{\mathbb{A}} u(y)\exp\left(\frac{\mathrm{i}k}{x_3}(x_1y_1+x_2y_2)\right)\,\mathrm{d}y_1\,\mathrm{d}y_2 \\ &= c(x)\int_{\mathbb{R}^2} u(y)\chi_{\mathbb{A}}(y)\exp\left(-2\pi\,\mathrm{i}(\xi_1y_1+\xi_2y_2)\right)\,\mathrm{d}y_1\,\mathrm{d}y_2. \end{aligned}$$

with $\xi_i = \frac{1}{\lambda x_3}x_i$ for $i = 1, 2$.

$$c(x) := \frac{\exp(\,\mathrm{i}kx_3)}{\mathrm{i}\lambda x_3}\exp\left(\frac{\mathrm{i}k}{2x_3}(x_1^2+x_2^2)\right) \tag{9.19}$$

9.3.3. Coherent Diffraction Measurements

Due to wavelengths in the nanometer scale and partial incoherences in the field u measurement devices are not able to measure the complex valued wave field $F_\perp u$ in the detector plane. A broad discussion on the sampling of highly coherent fields and statistical properties of waves is beyond the scope of this work. The interested readers are referred to (Paganin, 2006).

From our mathematical point of modeling the problem, we focus on the fact that diffraction patterns collect by measurement devices are given in terms of absolute counts of electrons or photons. That is, the measured intensity I of the wave field in the x_3-plane, *i.e.,* its *Fourier* (9.18) *or Fresnel* (9.15) *transform* $F_\perp$, is actually given by its squared amplitude.

$$|F_\perp(u)|^2 = I(\cdot). \tag{9.20}$$

This results in an ill-posed inverse problem as any phase could be assigned to the amplitudes prior to an inverse Fourier or Fresnel transform to real space. Hence additional *a priori* information on u are needed in order to perform a reasonable reconstruction on u. We will see in the next chapter, that classical algorithmic schemes can be interpreted in terms of the theory introduced in Part I

9.4. The Phase Retrieval Problem

A detailed survey on Phase Retrieval and numerical methods can be found in (Luke et al., 2002).

As the theoretical foundations of the first part of this work provide tools for a detailed analysis on Euclidean spaces we will carry out our analysis on the discretized model spaces. The motivation on this is also that at is we are able to provide a good understanding of the algorithms behavior within the modeled problem, rather than the accuracy of the modeling process itself. The operator $\mathbb{F} : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2}$ becomes the discretized version of the Fourier transform (9.18) or discretized Fresnel transform (9.15). Important properties of $\mathbb{F}$ are that it is an unitary, linear operator.

9.4.1. Preliminaries

$\mathbb{C}^{n_1 \times n_2}$ equipped with the real inner product

$$\langle x, y \rangle := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \operatorname{Re}\left((x_{ij})^* \; y_{ij}\right) \tag{9.21}$$

is an Euclidean space.

We define the *pointwise multiplication* $\odot : \mathbb{C}^{n_1 \times n_2} \times \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2}$, $(x, y) \mapsto x \odot y$ by

$$[x \odot y]_{ij} := x_{ij} \, y_{ij}, \quad i = 1, \ldots n_1, \; j = 1, \ldots, n_2. \tag{9.22}$$

Throughout the following sections $\mathbb{I} \subset \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ is a index set and we will use the notation

$$[1_{\mathbb{I}}]_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathbb{I} \\ 0 & \text{if } (i,j) \notin \mathbb{I} \end{cases}. \tag{9.23}$$

For any $x \in \mathbb{C}^{n_1 \times n_2}$ we use the notation $|x|$ as the pointwise absolute value, *i.e.*,

$$|x| := \begin{pmatrix} |x_{11}| & \ldots & |x_{n_1 1}| \\ \vdots & \ddots & \vdots \\ |x_{1 n_2}| & \ldots & |x_{n_1 n_2}| \end{pmatrix} \tag{9.24}$$

9.5. Phase Retrieval as a Feasibility Problem

To model the phase retrieval problem as a feasibility problem, we define the *magnitude constraint set* **M** by

$$\mathbf{M} := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; |\mathbb{F}\phi| = \mathbf{b} \right\}. \tag{9.25}$$

As mentioned at the end of section 9.3.3 the object ϕ cannot be recovered from a single magnitude measurement $\mathbf{b}$.

In order to achieve a reasonable reconstruction on ϕ one has to incorporate additional *a priori* information about ϕ into the model. In classical applications this additional information is for instance that the illuminated object has compact support, real-valuedness of ϕ (non-scattering, absorbing), known amplitude (non-absorbing) or combinations of these.

The support constraint, the combined support and real-valuedness and the combined support and nonnegativity constraint can be formulated as follows

$$\mathbf{S} := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; \phi_{ij} = 0,\ (i,j) \notin \mathbb{I} \right\}, \tag{9.26}$$

$$\mathbf{S}_* := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; \begin{array}{ll} \phi_{ij} \in \mathbb{R}, & (i,j) \in \mathbb{I} \\ \phi_{ij} = 0, & (i,j) \notin \mathbb{I} \end{array} \right\}, \tag{9.27}$$

$$\mathbf{S}_+ := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; \begin{array}{ll} \phi_{ij} \in [0,\infty), & (i,j) \in \mathbb{I} \\ \phi_{ij} = 0, & (i,j) \notin \mathbb{I} \end{array} \right\} \tag{9.28}$$

where $\mathbb{I} \subset \{1,\dots,n_1\} \times \{1,\dots,n_2\}$ is an arbitrary index set.

Corollary 9.4:
The sets $\mathbf{S}$, $\mathbf{S}_*$ and $\mathbf{S}_+$ are convex The corresponding projectors onto these sets are pointwise given by

$$[P_{\mathbf{S}}\phi]_{ij} = \begin{cases} \phi_{ij} & \text{if } (i,j) \in \mathbb{I} \\ 0 & \text{if } (i,j) \notin \mathbb{I} \end{cases}, \tag{9.29}$$

$$[P_{\mathbf{S}}\phi]_{ij} = \begin{cases} \operatorname{Re}\phi_{ij} & \text{if } (i,j) \in \mathbb{I} \\ 0 & \text{if } (i,j) \notin \mathbb{I} \end{cases}, \tag{9.30}$$

$$[P_{\mathbf{S}}\phi]_{ij} = \begin{cases} (\operatorname{Re}\phi_{ij})_+ & \text{if } (i,j) \in \mathbb{I} \\ 0 & \text{if } (i,j) \notin \mathbb{I} \end{cases}. \tag{9.31}$$

Note that using notation (9.22) and (9.23) the Projector $P_{\mathbf{S}}$ onto the support set $\mathbf{S}$ for given $\phi \in \mathbb{C}^{n_1 \times n_2}$ can be written as $P_{\mathbf{S}}\phi = 1_{\mathbb{I}} \odot \phi$.

In contrast to the above mentioned convex constraint sets the amplitude constraint

$$\mathbf{N} := \{\phi \mid |\phi| = \mathbf{a}\} \tag{9.32}$$

is nonconvex.

Corollary 9.5:
Let $\mathbf{N}$ be the amplitude constraint set given by (9.32). The projector $P_{\mathbf{N}}$ can be component wise defined by

$$[P_{\mathbf{N}}\phi]_{ij} := \begin{cases} \mathbf{a}_{ij}\phi_{ij}/|\phi_{ij}| & \text{if } \phi_{ij} \neq 0 \\ \mathbf{a}_{ij}\exp(\,\mathrm{i}\theta), \quad \theta \in [0, 2\pi) & \text{if } \phi_{ij} = 0 \end{cases} \quad \text{for } i = 1, \ldots, n_1,\ j = 1, \ldots, n_2. \tag{9.33}$$

Theorem 9.6:
The magnitude set $\mathbf{M}$ given by (9.25), *i.e.,*

$$\mathbf{M} := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; |\mathbb{F}\phi| = \mathbf{b} \right\}$$

is prox-regular. The corresponding projector is given by

$$P_{\mathbf{M}}\phi = \mathbb{F}^{-1}\hat{\phi} \tag{9.34}$$

where for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$

$$\hat{\phi}_{ij} = \begin{cases} \mathbf{b}_{ij}[\mathbb{F}\phi]_{ij}/|[\mathbb{F}\phi]_{ij}| & \text{if } [\mathbb{F}\phi]_{ij} \neq 0 \\ \mathbf{b}_{ij}\exp(\,\mathrm{i}\theta), \quad \theta \in [0, 2\pi) & \text{if } [\mathbb{F}\phi]_{ij} = 0 \end{cases}. \tag{9.35}$$

Proof. That equation (9.34) indeed forms a projector follows by Lemma 3.21 and Corollary 9.5, as $\mathbb{F}$ is an unitary operator. By this given closed form of the projector $P_{\mathbf{M}}$ it then follows that the projection is unique for points close enough to $\mathbf{M}$ and hence prox-regular by Theorem 3.13. □

9.5.1. Physical Methods vs. Mathematical Description

In (Gerchberg and Saxton, 1972), independently of previous mathematical results for projections onto convex sets, the authors proposed a simple algorithm for solving phase retrieval problems in two dimensions. This algorithm was recognized in (Levi and Stark, 1984) a projection algorithm.

Algorithm 9.7 (Gerchberg-Saxton Algorithm, (Gerchberg and Saxton, 1972))**:**
For given $\phi_0 \in \mathbb{C}^{n_1 \times n_2}$ the *Gerchberg-Saxton* algorithm is defined by

$$\phi_{n+1} \in P_{\mathbf{N}}P_{\mathbf{M}}\phi_n, \tag{9.36}$$

where $\mathbf{N}$ and $\mathbf{M}$ are given by (9.32) and (9.25).

Clearly the algorithm is a special instance of *von Neumanns* method of alternating projections (1.15)

Algorithm 9.8 (Error Reduction Algorithm, (Fienup, 1982))**:**
For given $\phi_0 \in \mathbb{C}^{n_1 \times n_2}$ the *Error Reduction* algorithm with pure support constraint is defined by

$$[\phi_{n+1}]_{ij} := \begin{cases} [P_{\mathbf{M}}\phi_n]_{ij} & \text{if } (i,j) \in \mathbb{I} \\ 0 & \text{if } (i,j) \notin \mathbb{I} \end{cases} \tag{9.37}$$

for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$.

The Error Reduction algorithm can be rewritten as a *von Neumann sequence*

$$\phi_{n+1} = (P_{\mathbf{S}} P_{\mathbf{M}}) \phi_n. \tag{9.38}$$

9.5.2. Hybrid Input-Output Algorithm

Algorithm 9.9 (Hybrid Input-Output Algorithm, (Fienup, 1982))**:**
Choose $x_0 \in \mathbb{C}^{n_1 \times n_2}$. For a given $\beta \in (0, 1]$ the Hybrid Input-Output algorithm (HIO) with pure support constraint is defined by

$$[\phi_{n+1}]_{ij} = \begin{cases} [P_{\mathbf{M}}\phi_n]_{ij} & \text{if } (i,j) \in \mathbb{I} \\ [\phi_n]_{ij} - \beta\, [P_{\mathbf{M}}\phi_n]_{ij} & \text{if } (i,j) \notin \mathbb{I} \end{cases}, \tag{9.39}$$

for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$.

Remark 9.10*: Error Reduction and Hybrid Input-Output were introduced and discussed in (Fienup, 1982) in a setting which incorporates slightly more general type of constraints.* ◊

Corollary 9.11 ((Bauschke et al., 2002, Observation 5.10))**:**
For $\beta = 1$ HIO with a pure support constraint is equivalent to the Douglas–Rachford Projection algorithm (1.21).

Proof. Note that for given $\phi \in \mathbb{C}^{n_1 \times n_2}$ update rule (9.39) can be rewritten as

$$\begin{aligned} & 1_{\mathbb{I}} \odot P_{\mathbf{M}}\phi + (1_{\mathbb{C}^{n_1 \times n_2}} - 1_{\mathbb{I}}) \odot (\phi - P_{\mathbf{M}}\phi) \\ =& 1_{\mathbb{I}} \odot (2P_{\mathbf{M}}\phi - \phi) + \phi - P_{\mathbf{M}}\phi \\ =& 1_{\mathbb{I}} \odot (2P_{\mathbf{M}}\phi - \phi) + \phi - P_{\mathbf{M}}\phi \\ =& P_{\mathbf{S}} (2P_{\mathbf{M}}\phi - \phi) + \phi - P_{\mathbf{M}}\phi \end{aligned}$$

and by Lemma 1.13 this coincides with the Douglas–Rachford Operator (1.20). □

Remark 9.12*: The equivalence in Corollary 9.11 relies on the property that* $\mathbf{S}$ *is a subspace and hence* $P_{\mathbf{S}}$ *is a linear mapping (cf. Corollary 2.5 (b)). For other sets as* $\mathbf{S}_*$ *or* $\mathbf{S}_+$ *this observation is not true (cf. (Bauschke et al., 2003)).* ◊

9.5.3. The Bigger Picture

Remark 9.13 (Physical Methods vs mathematical prescriptions of algorithms)*: Within the last sections we have summarized and derived mathematical prescriptions for some standard physical phase retrieval schemes. It is worth mentioning that some of the methods, mentioning especially HIO type algorithm, are still* state-of-the-art *techniques in phase retrieval. The benefit to develop a detailed mathematical framework and convergence analysis is manifold. On the one hand the projection methods framework covers some state-of-the-art methods* out of the box*, whilst it is flexible in the sense, that most of the reasonable* physical constraints *or rather* a priori information *can be formulated as constraint sets and can easily be incorporated in the algorithms. On the other hand the framework enables a qualitative and even quantitative description of the convergence behavior of the applied methods.*

A nice series of papers that describes phase retrieval from the viewpoint of convex analysis is (Bauschke et al., 2002; Bauschke et al., 2003; Bauschke et al., 2004). ◊

Remark 9.14 (Alternating Projection schemes vs. Douglas–Rachford type methods)*: The problem with* Alternating Projection schemes *as* Gerchberg-Saxton *or* Error Reduction *is that they* tend to get stuck in local minima*. Proposition 2.17 and Proposition 2.18 (c) emphasize that the set of fixed points of AP type methods is bigger than the set of fixed points of Douglas–Rachford type methods. Proposition 2.18 (a) also illustrates, that in fact, even if the problem locally reduces to a convex problem, the iterates generated by the Douglas–Rachford Operator are* repelled *by local minima.*

A detailed survey on a numerical analysis of iterative projection algorithms for phase retrieval can be found in (Marchesini, 2007). ◊

9.5.4. Incorporating Sparsity

The discussion of chapter 8 also allows to incorporate *sparsity type constraints* as a priori information. A discussion on reconstruction schemes using a sparse shearlet constraint for Fresnel measurements can be found in (Loock and Plonka, 2014).

Proposition 9.15:
Let $\mathbf{A_s}$ be the set of s-sparse vectors given by (8.3), *i.e.,*

$$\mathbf{A_s} := \{x \in \mathbb{E} |\ \|x\|_0 \leq s\}$$

and let U be an unitary linear operator. The set of s-sparse vectors under the transform U is given by

$$\mathbf{A_s^U} := U(\mathbf{A_s}) = \{x \in \mathbb{E} \mid \|Ux\|_0 \leq s\} . \tag{9.40}$$

The Projector onto $\mathbf{A_s^U}$ is given by

$$P_{\mathbf{A_s^U}} = U^* P_{\mathbf{A_s}} U \tag{9.41}$$

where $P_{\mathbf{A_s}}$ is defined by Proposition 8.3 equation (8.6). Furthermore at any point $\bar{x} \in \mathbf{A_s^U} \backslash \{0\}$ the set $\mathbf{A_s^U}$ is $(0, \delta)$-subregular at $\bar{x}$ for $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(U\bar{x})\})$.

Proof. The identity (9.41) follows by Lemma 3.21.

Subregularity follows by Theorem 8.4 (subregularity of $\mathbf{A_s}$) and Theorem 3.22 (subregularity under unitary transforms). □

Remark 9.16*: Equation* (9.41) *is* hard-thresholding *in the* basis U*.* ◊

9.5.5. Local Linear Convergence of Projection Algorithms in Phase Retrieval

Theorem 9.17 (linear convergence of classical phase retrieval algorithms)**:**
Let $\mathbf{B} = \mathbf{M}$ be the magnitude constraint set (9.25) and let $\mathbf{A} \in \{\mathbf{S}, \mathbf{S}_*, \mathbf{S}_+, \mathbf{N}\}$ be a constraint set. Assume the collection $\{\mathbf{A}, \mathbf{B}\}$ is locally linearly regular then alternating projections (1.15) converges locally with a linear rate.

Proof. By Theorem 9.6 the set $\mathbf{M}$ is prox-regular, hence by Theorem 3.16 super-regular. The set $\mathbf{A}$ is super-regular (for any possible choice). Using local linear regularity of $\{\mathbf{A}, \mathbf{B}\}$ local convergence now follows by Theorem 6.8. □

Theorem 9.18 (linear convergence using a sparsity constraint)**:**
Let $\mathbf{B} = \mathbf{M}$ be the magnitude constraint set (9.25) and let $\mathbf{A} = \mathbf{A_s^U}$ be a sparsity constraint set (9.40). Assume the collection $\{\mathbf{M}, \mathbf{A_s^U}\}$ is locally linearly regular then alternating projections (1.15) converges locally with a linear rate.

Proof. By Theorem 9.6 the set $\mathbf{M}$ is prox-regular, hence by Theorem 3.16 super-regular, *i.e.,* for any $\varepsilon > 0$ there exists $\delta >$ such that $\mathbf{M}$ is (ε, δ)-subregular on $\mathbb{B}_\delta(\hat{x})$ for $\hat{x} \in \mathbf{M} \cap \mathbf{A_s^U}$.

The set $\mathbf{A_s^U}$ is $(0, \delta)$-regular according to Proposition 9.15. Assuming local linear regularity of $\{\mathbf{A}, \mathbf{B}\}$ local convergence now follows by Theorem 6.5 and appropriate choice of $\varepsilon > 0$. □

Theorem 9.19 (linear convergence of **DR** for phase retrieval)**:**
Let $\mathbf{B} = \mathbf{M}$ be the magnitude constraint set (9.25) and let $\mathbf{A} \in \{\mathbf{S}, \mathbf{S}_*, \mathbf{S}_+, \mathbf{N}\}$ be a constraint set. Assume the collection $\{\mathbf{A}, \mathbf{B}\}$ is uniformly regular then Douglas–Rachford (1.21) converges locally with a linear rate.

Proof. By Theorem 9.6 the set $\mathbf{M}$ is prox-regular, hence by Theorem 3.16 super-regular. The set $\mathbf{A}$ is super-regular (any possible set). Assuming uniform regularity of $\{\mathbf{A}, \mathbf{B}\}$ local convergence now follows by Theorem 6.12. □

Theorem 9.20 (linear convergence of **DR** using a sparsity constraint)**:**
Let $\mathbf{B} = \mathbf{M}$ be the magnitude constraint set (9.25) and let $\mathbf{A} = \mathbf{A}_{\mathbf{s}}^{\mathbf{U}}$ be a sparsity constraint set (9.40). Assume the collection $\{\mathbf{M}, \mathbf{A}_{\mathbf{s}}^{\mathbf{U}}\}$ is uniformly regular then Douglas–Rachford (1.21) converges locally with a linear rate.

Proof. By Theorem 9.6 the set $\mathbf{M}$ is prox-regular, hence by Theorem 3.16 super-regular, *i.e.,* for any $\varepsilon > 0$ there exists $\delta >$ such that $\mathbf{M}$ is (ε, δ)-subregular on $\mathbb{B}_\delta(\hat{x})$ for $\hat{x} \in \mathbf{M} \cap \mathbf{A}_{\mathbf{s}}^{\mathbf{U}}$.

The set $\mathbf{A}_{\mathbf{s}}^{\mathbf{U}}$ is $(0, \delta)$-regular according to Proposition 9.15. Assuming uniform regularity of $\{\mathbf{A}, \mathbf{B}\}$ local convergence now follows by Theorem 6.11 and appropriate choice of $\varepsilon > 0$. □

Remark 9.21*: It is likely that the problems in practice do not meet the required properties. Especially uniform regularity is a restrictive property, which is likely to to be not fulfilled in practice (cf. Example 5.7). Due to errors in the modeling process (idealized model) or in the measurements it may also be the case that the resulting* feasibility problem *does not have feasible points at all. We refer to Chapter 7, where the standard techniques to handle this* inconsistency *are summarized.* ◊

9.6. Ptychographic Imaging

The pure phase retrieval problem is in fact a special case of more general diffraction imaging. In applications it is usually related to the setting where the object ϕ is illuminated by an incoming plane wave.

A more general problem is the problem of simultaneously object and probe retrieval. In this section we will discuss simultaneous object and probe discussion by Ptychographic Imaging. In contrast to the Phase retrieval problem several magnitude measurements are given in this case.

Given are m different measurements of the form (9.25), *i.e.,*

$$\mathbf{M}^{(k)} := \left\{ \phi \in \mathbb{C}^{n_1 \times n_2} \;\middle|\; |\mathbb{F}\phi^{(k)}| = \mathbf{b}^{(k)} \right\}, \quad k = 1, \ldots, m. \tag{9.42}$$

For any $k = 1, \ldots, m$ the complex wave field $\phi_k \in \mathbb{C}^{n_1 \times n_2}$ can be described as a pointwise product of $\mathbf{x} \in \mathbf{X} \subset \mathbb{C}^{N_1 \times N_2}$ and $\mathbf{y} \in \mathbf{Y} \subset \mathbb{C}^{n_1 \times n_2}$

$$\phi^{(k)} := S_k(\mathbf{x}) \odot \mathbf{y}, \tag{9.43}$$

where for $k = 1, \ldots, m$ $S_k : \mathbb{C}^{N_1 \times N_2} \to \mathbb{C}^{n_1 \times n_2}$ is a known operator with given adjoint $S_k^* : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{N_1 \times N_2}$ and $\mathbf{X} \subset \mathbb{C}^{N_1 \times N_2}$ and $\mathbf{Y} \subset \mathbb{C}^{n_1 \times n_2}$ are closed constrained sets.

Remark 9.22*: In this chapter, we focus on the setting of blind tomography, that is, an unknown object is illuminated m-times by a compactly supported wave that is shifted after each measurement along the $x-$ and $y-$axis of the object. As only a small piece of the object is illuminated in every measurement this is typically modeled in the setting where $N_1 > n_1$, $N_2 > n_2$, $\mathbf{x} \in \mathbb{C}^{N_1 \times N_2}$ describes the unknown transmission function of the object, $\mathbf{y} \in \mathbb{C}^{n_1 \times n_2}$ is the unknown illuminating wave and $S_k : \mathbb{C}^{N_1 \times N_2} \to \mathbb{C}^{n_1 \times n_2}$ an indexing operator that shapes the shift of the object to the probe,* i.e., $S_k(\mathbf{x}) = \mathbf{x}|_{\mathbb{I}_\mathbf{k}}$ *for a given index $\mathbb{I}_\mathbf{k} \subset \mathbb{N}^{n_1 \times n_2}$. The adjoint mapping $S_k^* : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{N_1 \times N_2}$ that embeds the probe function $\mathbf{y}$ into the higher dimensional space $\mathbb{C}^{N_1 \times N_2}$ pointwise by*

$$[S_k^*(\mathbf{y})]_{ij} = \begin{cases} \mathbf{y}_{\eta_k(i,j)} & \text{if } (i,j) \in \mathbb{I}_\mathbf{k} \\ 0 & \text{if } (i,j) \notin \mathbb{I}_\mathbf{k} \end{cases},$$

where $\mathbb{I}_\mathbf{k} \subset \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$ is an index set and $\eta_k : \mathbb{I}_\mathbf{k} \to \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ is the associated bijective mapping, that matches $\mathbb{C}^{N_1 \times N_2}|_{\mathbb{I}_\mathbf{k}}$ to $\mathbb{C}^{n_1 \times n_2}$.

However, the theory and suggested schemes in the following also cover other interesting ptychographic settings. One recent approach for instance is longitudinal ptychography *which is described in (Robisch and Salditt, 2013). That is, the object is moved along the z axis and different measurements with the same illuminating wave are taken. In this formulation then $\mathbf{y} \in \mathbb{C}^{n_1 \times n_2}$ is the unknown object, whilst $\mathbf{x} \in \mathbb{C}^{n_1 \times n_2}$ is the unknown probe and $S_k : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2}$ becomes a propagation operator that models the propagation of $\mathbf{x}$ along the $z-$axis. The corresponding adjoint mapping $S_k^* : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2}$ is then given by the corresponding back-propagation.* ♢

9.6.1. Thibaults Approach - Difference Map

In (Thibault et al., 2009) the authors formulate the Ptychography Problem in a way, that can be interpreted as a two set feasibility problem.

The Ptychography Problem is modeled on the product space

$$\mathbb{E} := \left(\mathbb{C}^{n_1 \times n_2}\right)^m = \underbrace{\mathbb{C}^{n_1 \times n_2} \times \cdots \times \mathbb{C}^{n_1 \times n_2}}_{m\text{-times}}$$

and one seeks for $(\phi_1, \ldots, \phi_m) \in \mathbb{E}$ that fulfills (9.42) and (9.43) using a specific version of the *Difference Map* Algorithm, that will be sketched in the following.

Within the framework of this work, their idea can be formulated in the following way. Condition (9.42) is embedded into $(\mathbb{C}^{n_1 \times n_2})^m$ analogous to the product space formulation of *Pierra* (1.6) by defining

$$\mathbf{A} := \mathbf{M}^{(1)} \times \cdots \times \mathbf{M}^{(m)}. \tag{9.44}$$

To formulate condition (9.43) we embed $\mathbf{X} \times \mathbf{Y} \subset \mathbb{C}^{N_1 \times N_2} \times \mathbb{C}^{n_1 \times n_2}$ into $(\mathbb{C}^{n_1 \times n_2})^m$ by setting

$$\mathbf{B} := \{(\phi_1, \ldots, \phi_m) \mid \phi_k = S_k(\mathbf{x}) \odot \mathbf{y},\ \mathbf{x} \in \mathbf{X},\ \mathbf{y} \in \mathbf{Y},\ \text{for all } k = 1, \ldots, m\}. \tag{9.45}$$

For $\Phi_0 := \left(\phi_1^{(0)}, \ldots, \phi_m^{(0)}\right) \in (\mathbb{C}^{n_1 \times n_2})^m$ the Difference Map Algorithm (9) in (Thibault et al., 2009) then is given by

$$\Phi_{n+1} = \Phi_n + P_{\mathbf{A}}[2P_{\mathbf{B}}\Phi_n - \Phi_n] - P_{\mathbf{B}}\Phi_n, \tag{9.46}$$

which is by Lemma 1.13 exactly the Douglas–Rachford algorithm.

Remark 9.23*: In (Thibault et al., 2009) condition* (9.43) *is actually modeled without any additional constraints on* $\mathbf{x}$ *and* $\mathbf{y}$*. As in practice usually additional a priori information onto* $\mathbf{x}$ *and* $\mathbf{y}$ *are provided we incorporate this in form of closed constraint sets* $\mathbf{X}$ *and* $\mathbf{Y}$ *within our analysis.* $\Diamond$

9.6.2. Projectors onto Thibaults Constraint Sets

The Projector onto the magnitude constraint set $\mathbf{A}$ given by (9.42) is according to (1.8) component-wise given by

$$P_{\mathbf{A}}(x_1, \ldots, x_m) = (P_{\mathbf{M}_1}x_1, \ldots, P_{\mathbf{M}_m}x_m), \tag{9.47}$$

(9.48)

where for $i = 1, \dots, m$ $P_{\mathbf{M}_i}$ is the Projector onto the measurement set $\mathbf{M}_i$ according to (9.34).

While it is not clear how to compute an exact Projector onto the set $\mathbf{A}$ given by (9.45), the following algorithmic subroutine is suggested by (Thibault et al., 2009) and can be interpreted as a heuristic for its approximation.

Subroutine 9.24 (Thibault [approximate] Projection $P_{\mathbf{B}}$):

Input. Current Iterate Φ_k and current approximation to probe and object $\mathbf{y}_k, \mathbf{x}_k$
Initialization. Define $\hat{\mathbf{x}}^{(0)} := \mathbf{x}_k$, $\hat{\mathbf{y}}^{(0)} := \mathbf{y}_k$.
For $l = 0, 1,$ **steps**

1. **For** $i = 1, \dots, m$**:** Define

$$\alpha_i^{(l)} := 2 \left[\sum_{j=1}^{m} S_j^* \left(\left(\hat{\mathbf{y}}^{(l)} \right)^* \odot \hat{\mathbf{y}}^{(l)} \right) \right]_i \tag{9.49}$$

   and update $\hat{\mathbf{x}}_i^{(k+1)}$ by

$$\hat{\mathbf{x}}_i^{(l+1)} := \frac{\left[\sum_{j=1}^{m} S_j^* \left(\left(\hat{\mathbf{y}}^{(l)} \right)^* \odot \phi_j^{(k)} \right) \right]_i}{\alpha_i^{(l)}} \tag{9.50}$$

2. **For** $i = 1, \dots, m$**:** Define

$$\beta_i^{(l)} := 2 \left[\sum_{j=1}^{m} S_j \left(\left(\hat{\mathbf{x}}^{(l)} \right)^* \odot \hat{\mathbf{x}}^{(l)} \right) \right]_i \tag{9.51}$$

   and update $\hat{\mathbf{y}}_i^{(k+1)}$ by

$$\hat{\mathbf{y}}_i^{l+1} := \frac{\left[\sum_{j=1}^{m} S_j \left(\left(\hat{\mathbf{x}}^{(l)} \right)^* \right) \odot \phi_j^{(k)} \right]_i}{\beta_i^l} \tag{9.52}$$

Final Step Define $\mathbf{x}_{k+1} := \hat{\mathbf{x}}^{(\text{steps}+1)}$, $y_{k+1} := \hat{\mathbf{y}}^{(\text{steps}+1)}$ and set

$$P_{\mathbf{B}} \phi^k := \left(S_1 \left(\mathbf{x}_{k+1} \right) \odot \mathbf{y}_{k+1}, \quad \cdots \quad, S_m \left(\mathbf{x}_{k+1} \right) \odot \mathbf{y}_{k+1} \right). \tag{9.53}$$

Remark 9.25*: The choice of parameters in subroutine 9.24 is indeed reasonable and in some sense optimal. In fact α_i (cf.* (9.49)*) is the Lipschitz modulus of the partial derivative mapping*

$$\frac{\partial}{\partial \mathbf{x}_i} \sum_{k=1}^{m} \left\| S_k \left(\mathbf{x} \right) \odot \mathbf{y} - \phi^{(k)} \right\|^2 .$$

A detailed analysis on this can be found in (Hesse et al., 2014b). ◊

Remark 9.26*: The authors of (Thibault et al., 2009) actually monitor* $\mathbf{x}_k$ *and* $\mathbf{y}_k$*,* i.e., *the object and illumination function, rather than the iterate* Φ^k *itself. This is similar to monitoring the* shadow sequence (2.20)*, as* x^k *and* y^k *are computed during the computation of* $P_{\mathbf{A}}\phi_k$*, which is a kind of regularization.* ◊

9.6.3. A method by Maiden and Rodenburg

In comparison to the method of Thibault, the distinctive feature of the method of Maiden and Rodenburg (Maiden and Rodenburg, 2009) is that only a single magnitude measurement in used in each step. Their method can be described as follows.

Algorithm 9.27 (Maiden & Rodenburg):
Input. $\lambda_1, \lambda_2 \in (0, 1]$.
Initialization. $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{C}^{N_1 \times N_2} \times \mathbb{C}^{n_1 \times n_2}$.
General Step ($k = 0, 1, \ldots$)

1. Set $I = \{1, \ldots, m\}$, $\hat{\mathbf{x}}_1 := \mathbf{x}_1$, $\hat{\mathbf{y}}_1 := \mathbf{y}_k$
   Inner Step ($l = 1, 2, \ldots, m$)

   1.1. Choose $j \in I$ and set $I \equiv I \backslash \{j\}$

   $$\phi_l \in P_{\mathbf{M}_j}\left(S_j\left(\hat{\mathbf{x}}_l\right) \odot \hat{\mathbf{y}}_l\right)$$

   1.2. Update $\mathbf{x}_{l+1}$, $\mathbf{y}_{l+1}$ by

   $$\hat{\mathbf{x}}_{l+1} = \hat{\mathbf{x}}_l + \frac{\lambda_1}{\max_{ij} |\hat{\mathbf{y}}_l|^2_{ij}} \left(S_j^*\left(\left(\hat{\mathbf{y}}_l\right)^* \odot \phi_l\right) - -S_j^*\left(\left(\hat{\mathbf{y}}_l\right)^* \odot \hat{\mathbf{y}}_l\right) \odot \hat{\mathbf{x}}_l\right), \quad (9.54)$$

   $$\hat{\mathbf{y}}_{l+1} = \hat{\mathbf{y}}_l + \frac{\lambda_2}{\max_{ij} |\hat{x}^l|^2_{ij}} \left(S_j\left(\left(\hat{\mathbf{x}}_l\right)^*\right) \odot \phi_l - S_j\left(\left(\hat{\mathbf{x}}_l\right)^* \odot \hat{\mathbf{x}}_l\right) \odot \hat{\mathbf{y}}_l\right). \quad (9.55)$$

2. Set $x^{k+1} \equiv \hat{x}^m$, $y^{k+1} \equiv \hat{y}^m$

Remark 9.28*: The choice of parameters in subroutine 9.24 is indeed reasonable and in some sense optimal. In fact for given* l $\max_{ij} |\hat{\mathbf{y}}_l|^2_{ij}$ *(cf.* (9.54)*) is the Lipschitz modulus of the gradient mapping*

$$\nabla_{\mathbf{x}} \left\| S_l\left(\mathbf{x}\right) \odot \mathbf{y} - \phi^{(l)} \right\|^2.$$

A detailed analysis on this can be found in (Hesse et al., 2014b). ◊

9.6.4. Residual error

For given reconstructions of probe and object function $\mathbf{x}, \mathbf{y}$ the *residual error* is given by

$$\sum_{k=1}^{m} \frac{\left\| |\mathbb{F}\left(S_k\left(\mathbf{x}\right) \odot \mathbf{y}\right)| - \mathbf{b}^{(k)} \right\|}{\sum_{j=1}^{m} \|\mathbf{b}^{(j)}\|}. \tag{9.56}$$

9.6.5. Reconstruction from Real Data

In this section we discuss the reconstruction from a dataset of 676 diffraction patterns provide by the IRP (*Institut für Röntgenphysik*) Göttingen. A detailed discussion on the experimental set-up, collection and examination of this dataset can be found in (Wilke et al., 2013).

The algorithmic schemes (Thibault et al., 2009) introduced in Section 9.6.1 and (Maiden and Rodenburg, 2009) introduced in Section 9.6.3 are leading strategies in the simultaneous recovery of probe and object function in Ptychographic Imaging. Motivated and circumstantiated by the theory developed in the first part of this work, we are able to analyze the convergence behavior and suggest improvements in particular for Difference map algorithm.

Given are $m = 26^2 = 676$ diffraction patterns (cf. (9.42)), each are taken on a screen of 192^2 points.

Our comparison include five different algorithms: The two popular in the literature discussed in Sections 9.6.1 & 9.6.3, and several variants based on the method of (Thibault et al., 2009). More specially:

1. **Rodenburg:** (Maiden and Rodenburg, 2009) as described in Section 9.6.3 Algorithm 9.27.
2. **DM/DR:** The original Difference Map scheme of (Thibault et al., 2009) described in Section 9.6.1 equation (9.46), which is equivalent to Douglas–Rachford (1.21).
3. **DR-λ, $\lambda = 0.7$:** Relaxed Averaged Alternating Reflection (7.9) with $\lambda = 0.7$, as a regularized version of Douglas–Rachford.
4. **DR-λ auto:** Relaxed Averaged Alternating Reflection (7.9). Following (Luke, 2005), sequence $(\lambda_n)_{n=0}^{\infty}$ is chosen to be

   $$\lambda_k = 0.9 \exp((-k)^3) + 0.5(1 - \exp((-k)^3)).$$

   In particular, $\lambda_0 = 0.9$ and $\lambda_k \to .5$ as $k \to \infty$.

5. **AP:** The Projection schemes of (Thibault et al., 2009) using von Neumanns method of alternating projections (AP) (1.15) in place of the difference map.

6. **Reference:** Reference sceme as computed in (Wilke et al., 2013) (Average over 100 iterations obtained by a *Rodenburg sceme*).

In all of the schemes (except the reference scheme by Maiden and Rodenburg) $P_{\mathbf{B}}$ is given by (9.47), while an approximation to $P_{\mathbf{A}}$ is evaluated using Subroutine 9.24 and a number of steps = 5 inner iteration steps.

The initial guess on the object $\mathbf{x}$ is constant amplitude with zero phase. The initial guess on the probe $\mathbf{y}$ is the *Fresnel propagation* of a disc of constant amplitude according to the aperture-size in the physical set up (Wilke et al., 2013), as in typical Ptychography experiments *a priori*, though its fine structure, due to instrumentation aberrations and the like, the general structure on the probe is unknown. The object, on the other hand, is assumed to be completely unknown except for certain qualitative properties – for example, that it is barely absorbing.

We run two different reconstructions imposing two different object constraints, that are

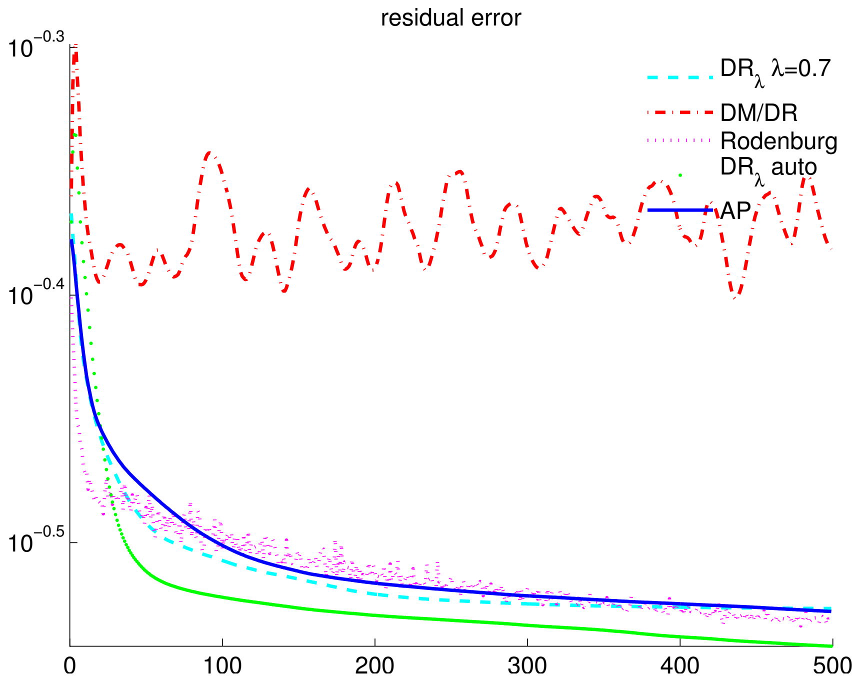
$$\mathbf{X}^{\text{run1}} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0,1]^{N_1 \times N_2} \right\}, \tag{9.57}$$

$$\mathbf{X}^{\text{run2}} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0.8,1]^{N_1 \times N_2} \right\}. \tag{9.58}$$

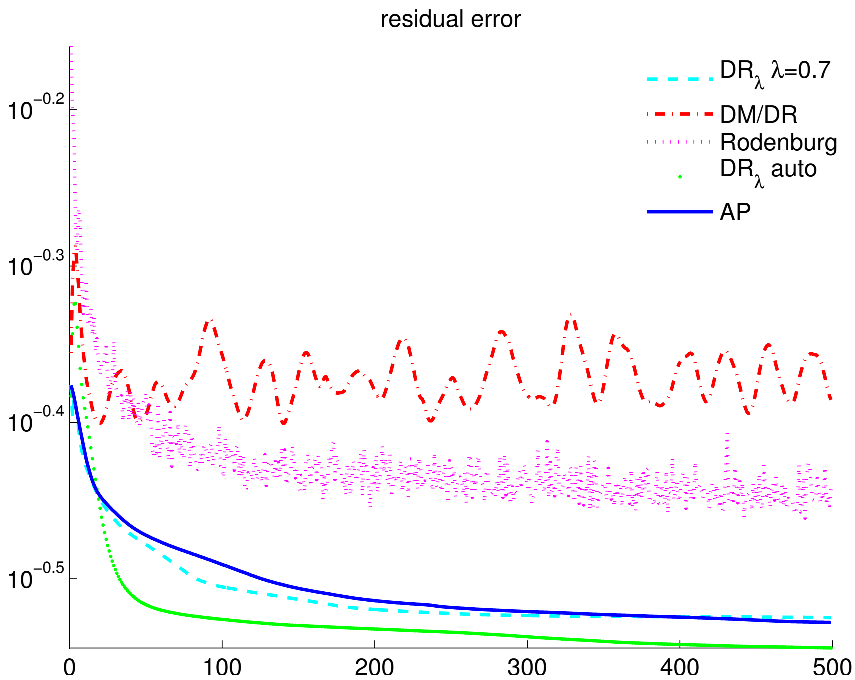
Constraint $\mathbf{X}^{\text{run1}}$ applies in general, as it models the absorbing nature in the experiment, while $\mathbf{X}^{\text{run2}}$ is a reasonable (additional) assumption on the object $\mathbf{x}$ in this specific setup.

In both runs a support constraint according to (9.26) is applied to the probe, which imposes that the probe is compactly supported within a disc in order to incorporate *a priori* information about the physical set-up.

Remark 9.29 (Remark on the plots)*: In the setting of blind ptychography the model does not circumvent non-uniqueness up to global shifts in* $\mathbf{x}$ *and* $\mathbf{y}$ *(cf. equation* (9.43)*). For a better comparison object and probe are translated by a global shift, computed using code written by Mauel Guizar (Guizar-Sicairos et al., 2008).* ◊

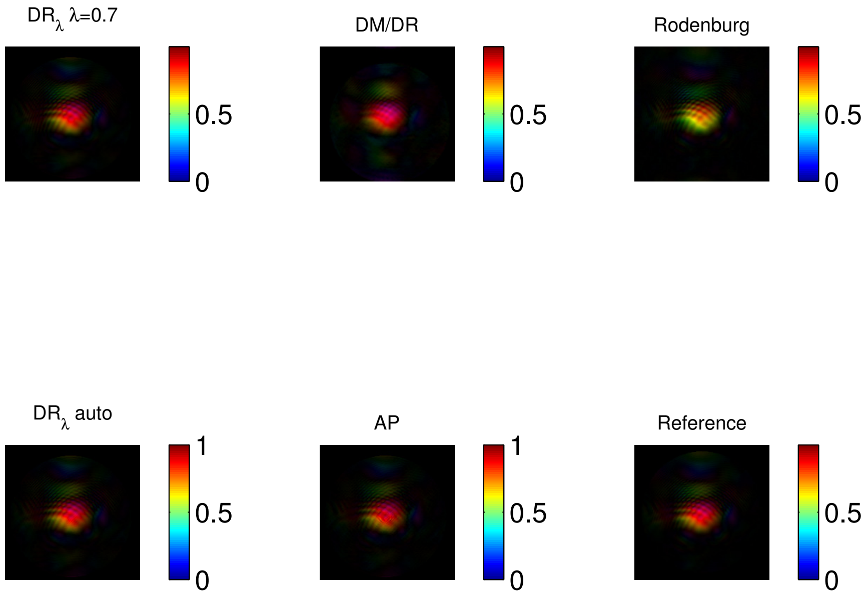


(a) $\mathbf{X} = \left\{\mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0,1]^{N_1 \times N_2}\right\}$

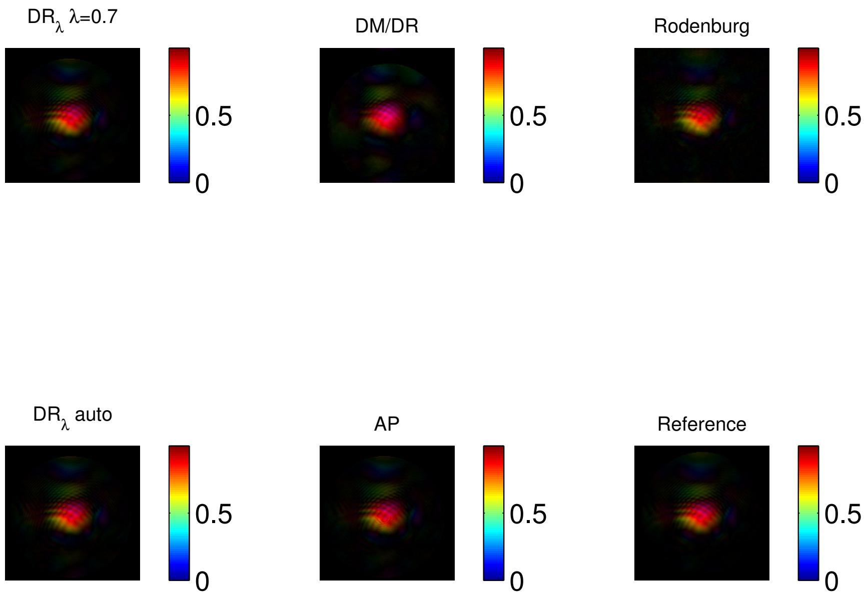


(b) $\mathbf{X} = \left\{\mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0.8,1]^{N_1 \times N_2}\right\}$

Figure 9.1.: Residual error (9.56) for two different object constraints

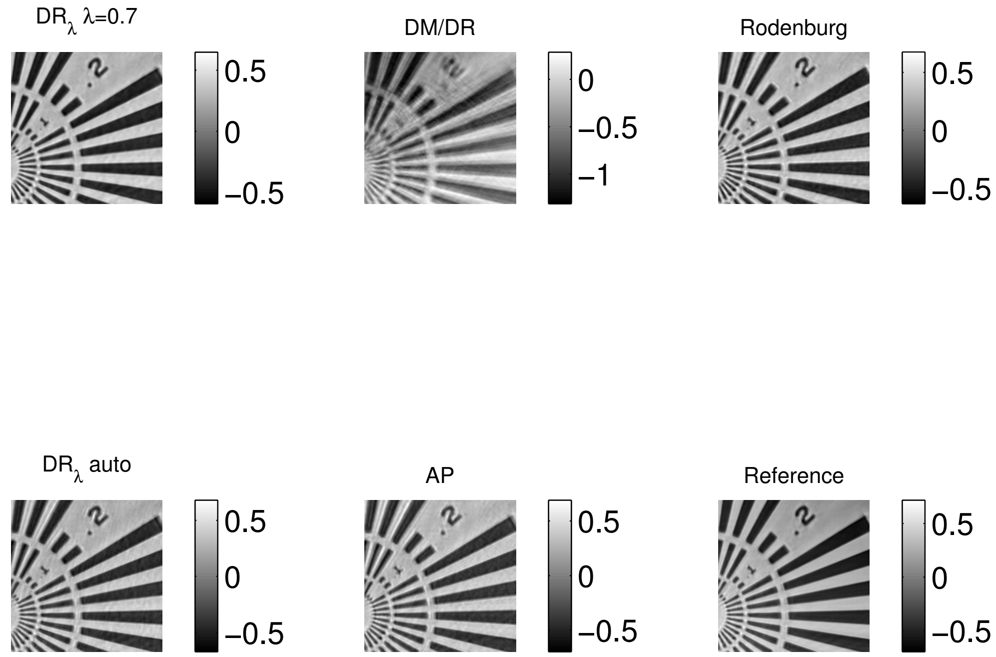


(a) Reconstructed illuminating wave $\mathbf{y}$ (amplitude), $\mathbf{X} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0,1]^{N_1 \times N_2} \right\}$.

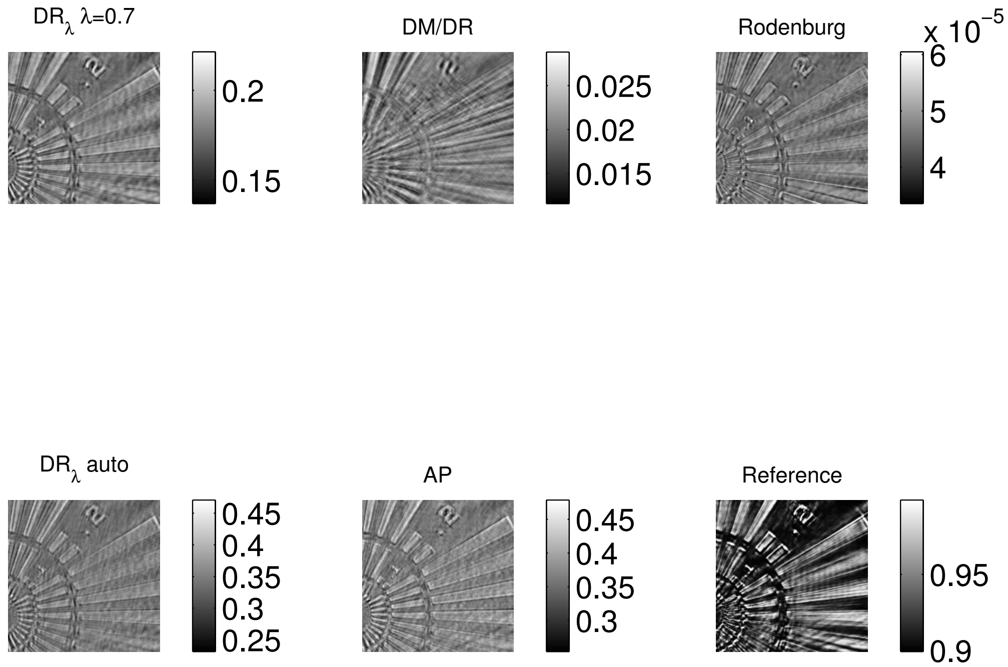


(b) Reconstructed illuminating wave $\mathbf{y}$ (amplitude), $\mathbf{X} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0.8,1]^{N_1 \times N_2} \right\}$.

Figure 9.2.: Reconstructed illuminating wave function $\mathbf{y}$ for different constraint sets (9.57) and (9.58)

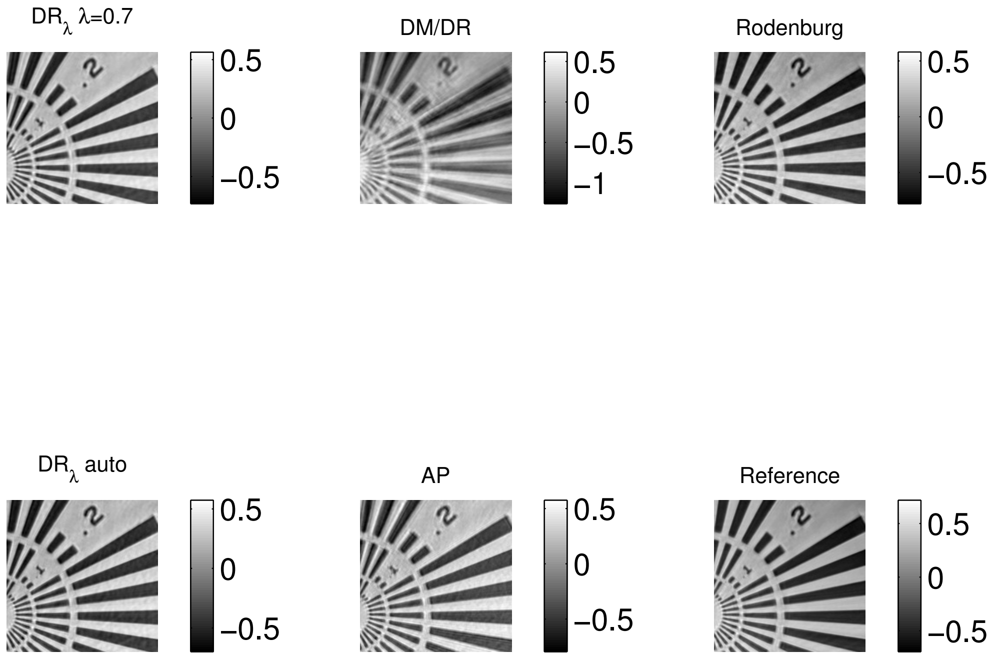


(a) Phase of reconstructed Object $\mathbf{x}$ (restricted to illuminated aperture)

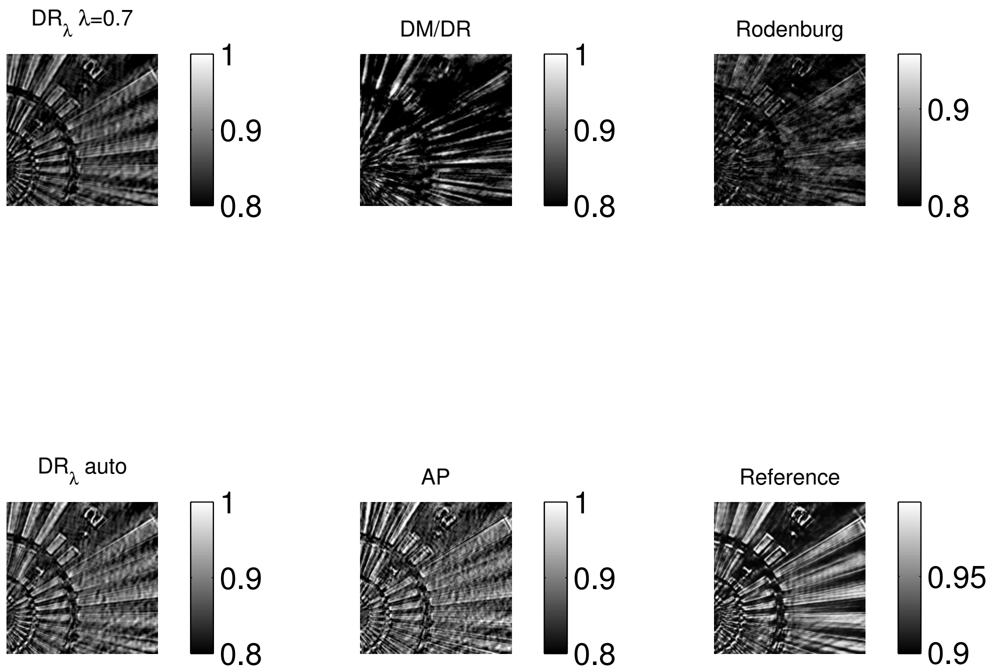


(b) Amplitude of reconstructed Object $\mathbf{x}$ (restricted to illuminated aperture)

Figure 9.3.: Reconstructed Object $\mathbf{x}$, $\mathbf{X} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0,1]^{N_1 \times N_2} \right\}$.



(a) Phase of reconstructed Object $\mathbf{x}$ (restricted to illuminated aperture)



(b) Amplitude of reconstructed Object $\mathbf{x}$ (restricted to illuminated aperture)

Figure 9.4.: Reconstructed Object $\mathbf{x}$, $\mathbf{X} = \left\{ \mathbf{x} \in \mathbb{C}^{N_1 \times N_2} \;\middle|\; |\mathbf{x}| \in [0.8, 1]^{N_1 \times N_2} \right\}$.

9.6.6. Comments on the Reconstructions

One of the main conclusions of the first part of this work is that **DM/DR** is *sensitive* to the intersection and does not lack the tendency to get stuck in local minima (cf. Remark 2.19 and Remark 6.15). Figure 9.1(a)/9.1(b) indicates that the residual of the iterates reconstructed by **DM/DR** oscillate (and hence in fact do not get stuck into global/local minima). Figure 9.1(a)/9.1(b) also shows that regularization of the operator (**DR-**λ) can restore the stability of the reconstruction and yield good results in terms of small residuals and *qualitative* good reconstructions. Especially when incorporating the stronger amplitude information (9.58) the reconstructed amplitudes of the **DR-**λ methods show more features than the other algorithm (Figure 9.4(b)).

In the **AP** scheme on the other hand the error in the residual is decreasing (Figure 9.1(a)/9.1(b)). Note that **AP** *by definition* always generates a decreasing sequence in $\|\Phi_n - \Phi_{n-1}\|$. The reconstructions are good, even though the residual stagnates at a higher level than the of **DR-**λ methods. This could indicate that the algorithm is stuck in a local minimum.

The *cyclic scheme* **Rodenburg** in both reconstruction does not form a non-decreasing sequence. Another interesting observation for this scheme is, that rather than benefiting from the additional information (constraint (9.58)) reconstruction actually slows down (Figure 9.1(b)).

Remark 9.30*: A comparison between the method of Thibault and Rodenburg for simulated data can be found in (Mattsson, 2013). The code was implemented using the* ProxToolbox*, which is available in (Luke, 2012b). Parts of the code were implemented in (Mattsson, 2013) and (Hesse et al., 2014b).* $\Diamond$

10. Concluding Remarks and Open Problems

We considered projection algorithms for solving (nonconvex) feasibility problems in Euclidean spaces. Of special interest were the Method of Alternating Projections (**AP**) and the Douglas–Rachford algorithm (**DR**). A notion of local sub-firm nonexpansiveness with respect to the intersection was introduced for consistent feasibility problems. This, together with a coercivity condition that relates to the regularity of the collection of sets at points in the intersection, yields local linear convergence of **AP** a for a wide class of nonconvex problems, and even local linear convergence of nonconvex instances of the Douglas–Rachford algorithm. We emphasize the following ideas of what could be subject to further research.

The moduli of linear and uniform regularity ((5.35) and (5.34)) do not recover optimal convergence results in some easy examples (see Remark 6.6 and Remark 6.13). We suspect this is an artifact of our choice of the *product-space-norm* in the definition of metric (sub)-regularity. We believe that there is a quantitative primal definition of a angle between two sets that recovers optimal results for AP on subspaces. Corollary 5.30 indicates that the choice of different metrics result in different regularity moduli. This could also be useful to achieve optimal linear convergence results for Douglas–Rachford.

Due to the *qualitative nature* of some of the definitions, which are used in parts of this work, the local geometry is introduced on local neighborhoods (δ-Balls). This analysis can also be carried out on more general structures (half-spaces, cones and the-like), which are more appropriate to identify the *proper regions of convergence* in nonconvex applications.

Another direction is to extend this analysis more generally to fixed point mappings built upon functions and more general set-valued mappings, but also in particular *proximal* operators and reflectors. The generality of our approach makes such extensions quite natural. Indeed, local linear as well as uniform regularity of collections of sets was shown in Chapter 5 to be related to *metric (sub-)regularity* of set-valued mappings which then guarantees that the condition (6.1) of Lemma 6.1 is satisfied. Of course, the difficulty remains to show that the mappings are indeed metrically subregular.

List of Figures

Bibliography

Aronszajn, N. (1950). *Introduction to the Theory of Hilbert Spaces*. Introduction to the Theory of Hilbert Spaces. Research [sic] Foundation.

Aronszajn, N. and Smith, K. T. (1954). Invariant Subspaces of Completely Continuous Operators. *Annals of Mathematics*, 60(2):pp. 345–350.

Artacho, F. J. A. and Borwein, J. M. (2013). Global convergence of a non-convex Douglas-Rachford iteration. *Journal of Global Optimization*, check:753–769.

Artacho, F. J. A., Borwein, J. M., and Tam, M. K. (2013). Recent Results on Douglas–Rachford Methods for Combinatorial Optimization Problems. *Journal of Optimization Theory and Applications*, accepted in November 2013:30.

Bauschke, H. H. and Borwein, J. M. (1993). On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1(2):185–212.

Bauschke, H. H. and Borwein, J. M. (1996). On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2002). Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *J. Opt. Soc. Am. A,*, 19(7):1334–1345.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2003). A Hybrid Projection Reflection Method for Phase Retrieval. *Journal of the Optical Society of America*, 20(6):1025–1034.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2004). Finding Best Approximation Pairs Relative to Two Closed Convex Sets in Hilbert Spaces. *Journal of Approximation Theory*, 127:178–314.

Bauschke, H. H., Cruz, J. Y. B., Nghia, T. T. A., Phan, H. M., and Wang, X. (2013a). The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle. arXiv:1309.4709v2.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2013b). Restricted normal cones and sparsity optimization with affine constraints. *Foundations of Computational Mathematics*, DOI: 10.1007/s1020801391610.:check.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2013c). Restricted Normal Cones and the Method of Alternating Projections: Applications. *Set-Valued and Variational Analysis*, 21:475–501.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2013d). Restricted Normal Cones and the Method of Alternating Projections: Theory. *Set-Valued and Variational Analysis*, 21:431–473.

Bauschke, H. H. and Noll, D. (2014). On the local convergence of the Douglas-Rachford algorithm. arXiv:1401.6188v1.

Beck, A. and Teboulle, M. (2011). A Linearly Convergent Algorithm for Solving a Class of Nonconvex/Affine Feasibility Problems. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 33–48. Springer New York.

Borwein, J. M. (1986). Stability and regular points of inequality systems. *Journal of Optimization Theory and Applications*, 48(1):9–52.

Borwein, J. M. and Sims, B. (2011). The Douglas-Rachford algorithm in the absence of convexity. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and its Applications*, 49:93–109.

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

Demanet, L. and Zhang, X. (2013). Eventual linear convergence of the Douglas Rachford iteration for basis pursuit. arXiv:1301.0542[math.NA].

Deutsch, F. (1991). The method of alternating orthogonal projections. *Approximation theory, spline functions and applications*, Maratea:105–121.

Deutsch, F. (1995). The Angle Between Subspaces of a Hilbert Space. In Singh, S., editor, *Approximation Theory, Wavelets and Applications*, volume 454 of *NATO Science Series*, pages 107–130. Springer Netherlands.

Deutsch, F. (2001). *Best Approximation in Inner Product Spaces*. Springer-Verlag, New York.

Deutsch, F. and Hundal, H. (2006a). The rate of convergence for the cyclic projections algorithm I: Angles between convex sets. *Journal of Approximation Theory*, 142(1):36 – 55.

Deutsch, F. and Hundal, H. (2006b). The rate of convergence for the cyclic projections algorithm II: Norms of nonlinear operators . *Journal of Approximation Theory*, 142(1):56 – 82.

Deutsch, F. and Hundal, H. (2008). The rate of convergence for the cyclic projections algorithm III: Regularity of convex sets. *Journal of Approximation Theory*, 155(1):155 – 184.

Dontchev, A. L. and Rockafellar, R. T. (2008). *Implicit Functions and Solution Mappings*. Springer.

Douglas, J. and Rachford, H. H. (1956). On the Numerical Solution of the Heat Conduction Problems in Two and Three Space Variables. *Trans. Amer. Math. Soc.*, 82:421–439.

Eckstein, J. (1989). Splitting Methods for Monotone Operators with Applications to Parallel Optimization. Master's thesis, Massachusetts Institue for Technology, Laboratory for Information and Decision Systems.

Eckstein, J. and Bertsekas, D. P. (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318.

Eckstein, J. and Svaiter, B. F. (2009). General projective splitting methods for sums of maximal monotone operators. *SIAM Journal on Control and Optimization*, 48(2):787–811.

Eckstein, J. and Svaiter, B. F. (January 2008). A family of projective splitting methods for the sum of two maximal monotone operators. *Mathematical Programming*, 111(1-2):173–199.

Efimov, N. V. and Steckkin, S. V. (1958). Some Properties of Chebyshev Sets. *Dokl. Akad. Nauk SSSR*, 118:17–19.

Elser, V., Rankenburg, I., and Thibault, P. (2006). Searching with iterated maps. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2):418.

Federer, H. (1959). Curvature measures. *Trans American Society*, 93:418–491.

Fienup, J. R. (1982). Phase retrieval algorithms: A comparison. *Applied Optices*, 21:2758–2769.

Friedrichs, K. (1937). On Certain Inequalities and Characteristic Value Problems for Analytic Functions and For Functions of Two Variables. *Transactions of the American Mathematical Society*, 41(3):pp. 321–364.

Gabay, D. (1983). Applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Solution of Boundary- Value Problems*, pages 299–340. North-Holland.

Gerchberg, R. W. and Saxton, W. O. (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246.

Giewekemeyer, K. (2011). *A study on new approaches in coherent x-ray microscopy of biological specimens*, volume 5. Universitätsverlag Göttingen.

Goebel, K. and Kirk, W. A. (1990). *Topics in Metric Fixed Point Theory*. Cambridge University Press, Cambridge.

Goebel, K. and Reich, S. (1984). *Uniform convexity, hyperbolic geometry, and nonexpansive mappings.*, volume 83. Pure and Applied Mathematics.

Gubin, L. G., Polyak, B. T., and Raik, E. V. (1967). The method of projections for finding the common point of convex sets. *Computational Mathematics and Mathematical Physics*, 7(6):1–24.

Guizar-Sicairos, M., Thurman, S., and Fienup, J. (2008). Efficient subpixel image registration algorithms. *Optics letters*, 33(2):156–158.

Hesse, R. and Luke, D. R. (2012). Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. arXiv:1212.3349v1.

Hesse, R. and Luke, D. R. (2013). Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization*, 23(4):2397–2419.

Hesse, R., Luke, D. R., and Neumann, P. (2014a). Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility. *IEEE Transactions on Signal Processing*, 99:1.

Hesse, R., Luke, D. R., Sabach, S., and Tam, M. K. (2014b). Proximal Heterogeneous Block Input-Output Method and application to Blind Ptychographic Diffraction Imaging. preprint.

Ioffe, A. D. (2000). Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501.

Jin, B. and Lorenz, D. A. (2010). Heuristic Parameter-Choice Rules for Convex Variational Regularization Based on Error Estimates. *SIAM Jounal on Numerical Analysis*, 48:1208–1229.

Jordan, C. (1875). Essai sur la géométrie á n dimensions. *Buletin de la Société Mathématique de France*, 3:103–174.

Kress, R. and Colton, D. (1998). *Inverse Acoustic and Electromagnetic Scattering Theory*, volume 93. Applied Mathematical Sciences, Springer, Heidelberg.

Kruger, A. Y. (2004). Weak Stationarity: Eliminating the gap between necessary and sufficient conditions. *Optimization*, 53:147–164.

Kruger, A. Y. (2006). About Regularity of Collections of Sets. *Set-Valued Analysis*, 14:187–206.

Kruger, A. Y. and Thao, N. H. (2014). About uniform regularity of collections of sets. arXiv:1305.0079v2.

Levi, A. and Stark, H. (1984). Image restoration by the method of generalized projections with application to restoration from magnitude. *Journal of the Optical Society of America A*, 1:932–943.

Lewis, A. S., Luke, D. R., and Malick, J. (2009). Local Linear Convergence for Alternating and Averaged Nonconvex Projections. *Foundations of Computational Mathematics*, 9(4):485–513.

Lewis, A. S. and Malick, J. (2008). Alternating Projections on manifolds. *Math. Oper. Res.*, 33:216–234.

Lions, P. L. and Mercier, B. (1979). Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979.

Loock, S. and Plonka, G. (2014). Phase retrieval for Fresnel measurements using a shearlet sparsity constraint. *Inverse Problems*, 30(5):055005.

Luke, D. R. (2005). Relaxed Averaged Alternating Reflections for Diffraction Imaging. *Inverser Problems*, 21:37–50.

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

Luke, D. R. (2012a). Local Linear Convergence of Approximate Projections onto Regularized Sets. *Nonlinear Analysis*, 75:1531–1546.

Luke, D. R. (2012b). ProxToolbox. A toolbox of algorithms and projection operators for implementing fixed point iterations based on the Prox operator, available at `http://num.math.uni-goettingen.de/~r.luke/publications/ProxToolbox.html`.

Luke, D. R. (2013). Prox-Regularity of Rank Constraint Sets and Implications for Algorithms. *Journal of Mathematical Imaging and Vision*, 47(3):231–238.

Luke, D. R., Burke, J. V., and Lyon, R. G. (2002). Optical Wavefront Reconstruction: Theory and Numerical Methods. *SIAM REVIEW*, 44(2):169–224.

Maiden, A. M. and Rodenburg, J. M. (2009). An improved ptychographical phase retrieval algorithm for diffractive imaging . *Ultramicroscopy*, 109(10):1256 – 1262.

Marchesini, S. (2007). Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78(1):011301.

Mattsson, P. P. (2013). *Projection Methods in Ptychographic Imaging.* Masterthesis, Institut für Numerische und Angewandte Mathematik Göttingen.

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

Moreau, J. J. (1962). Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *C. R. Acad. Sci. Paris*, 255:238–240.

Nashed, M. Z. (1968). A decomposition relative to convex sets. *Proceedings of the American Mathematical Society*, 19(4):pp. 782–786.

Ngai, H. V. and Théra., M. (2001). Metric Inequality, Subdifferential Calculus and Applications. *Set-Valued Analysis*, 9:187–216. 10.1023/A:1011291608129.

Paganin, D. M. (2006). *Coherent X-ray optics*, volume 1. Oxford University Press Oxford.

Phan, H. M. (2014). Linear Convergence of the Douglas-Rachford Method for Two Closed Sets. arXiv:1401.6509v1.

Phelps, R. R. (1957). Convex sets and nearest points. *Proceedings of the American Mathematical Society*, 8(4):790–797.

Pierra, G. (1984). Decomposition through formalization in a product space. *Math. Program.*, 28:96–115.

Poliquin, R. A., Rockafellar, R. T., and Thibault, L. (2000). Local differentiability of distance functions. *Transactions of the American Mathematical Society*, 352(11):5231–5249.

Reinermann, J. and Schöneberg, R. (1967). Some Results in Fixed Point Theory for Nonexpansive and Pseudocontractive Maps in Hilbert Space. *Fixed Point Theory and its Applications*, Academic Press:175–186.

Robinson, S. M. (1976). Stability theory for systems of inequalities. *SIAM Journal on Numerical Analysis*, 13:497–513.

Robisch, A.-L. and Salditt, T. (2013). Phase retrieval for object and probe using a series of defocus near-field images. *Optics Express*, 21(20):23345–23357.

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

Thibault, P., Dierolf, M., Bunk, O., Menzel, A., and Pfeiffer, F. (2009). Probe retrieval in ptychographic coherent diffractive imaging . *Ultramicroscopy*, 109(4):338 – 343.

Ursecu, C. (1975). Multifunction with closed convex graph. *Czechoslovak Mathematical Journal*, 25:438–441.

von Neumann, J. (1933). Functional Operators.

Wilke, R. N., Vassholz, M., and Salditt, T. (2013). Semi-transparent central stop in high-resolution X-ray ptychography using Kirkpatrick–Baez focusing. *Acta Crystallographica*, A69:490–497.

Zarantonello, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. In Zarantonello, E. H., editor, *Contributions to Nonlinear Functional Analysis*, pages 237–424. Academic Press, New York.

Curriculum Vitae

Robert Hesse, M.Sc.

Adress:	Institute for Numerical and Applied Mathematics University of Göttingen Lotzestraße 16-18 37083 Göttingen / Germany
Office:	+49/5561/394516
Email:	hesse@math.uni-goettingen.de
	`http://num.math.uni-goettingen.de/~hesse`

Personal Details

Date of birth	07/14/1985
Place of birth	Northeim
Nationality	German

Academic Education

05/2011-07/2014	PhD student of mathematics, University of Göttingen, Germany –advisor Professor Dr. D. Russell Luke
10/2006-04/2011	Student of mathematics, University of Göttingen, Germany
04/2011	**Master of Science** – advisor Professor Dr. Rainer Kress
09/2009	**Bachelor of Science** – advisor Professor Dr. Rainer Kress
08/2008-01/2009	Exchange semester, Lund University, Sweden

Research Experience

08/2011-07/2014	Collaborative Research Centre 755 *Nanoscale Photonic Imaging* German Research Foundation (DFG)
05/2011-07/2011	Institute for Numerical and Applied Mathematics University of Göttingen

Military Service

10/2005-06/2006	6./PzPiBtl. 1 - Holzminden

Teaching

10/2013-03/2014	*Convex Analysis and Applications,* leading of exercise session
04/2013-09/2013	*Introduction to Optimization,* leading of exercise session
10/2012-03/2013	*Seminar in Industrial Mathematics*
04/2012-09/2012	*Projection Methods and Applications to Image Processing,* assistant lecturer
05/2011-09/2011	*Numerical Mathematics* 2, assistant lecturer
04/2009-09/2010	*Numerical Mathematics* 1 & 2, student assistant/leading of exercise session

Conferences and Workshops

05/2014	*SIAM Conference on Optimization 2014*, San Diego, USA
07/2013	*International Conference on Continuous Optimization (ICCOPT 2013)*, Lisbon, Portugal
07/2013	*Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013)*, Lausanne, Switzerland
09/2012	Workshop *Advances in Image Processing*, Akademie Waldschlösschen (near Göttingen)
04/2012	*Spring School on Analysis: Variational Analysis and its Applications*, Paseky nad Jizerou, Czech Republic
10/2011	*Metric Regularity Days*, Paris, France

Talks

05/2014	The Douglas-Rachford Algorithm for Nonconvex Feasibility Problems, *SIAM Conference on Optimization 2014*, San Diego, USA
07/2013	Nonconvex feasibility problems, *International Conference on Continuous Optimization (ICCOPT 2013)*, Lisbon, Portugal
07/2013	Projection Methods for Sparse Affine Feasibility (poster), *Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013)*, Lausanne, Switzerland
10/2012	Projection Methods for nonconvex feasibility, *Centre for Optimization, Convex Analysis and Nonsmooth Analysis - Seminar*, University of British Columbia, Kelowna, Canada

Publications

R. Hesse and D. R. Luke, Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems, SIAM Journal on Optimization 23(4), 2397 – 2419 (2013).

R. Hesse, D. R. Luke and P. Neumann, Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility, IEEE Transactions on Signal Processing 99, DOI: 10.1109/TSP.2014.2339801 (2014).