Projection Methods in Sparse and Low Rank Feasibility

Dissertation zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

Doctor rerum naturalium

der Georg-August-Universität Göttingen

im Promotionsstudiengang *Mathematical Sciences*
der Georg-August University School of Science (GAUSS)

vorgelegt von Patrick Neumann aus Neuwied

Göttingen, 2015

Betreuungsausschuss

- Prof. Dr. Russell Luke,
  Institut für Numerische und Angewandte Mathematik,
  Georg-August-Universität Göttingen

- Prof. Dr. Max Wardetzky,
  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. Max Wardetzky,
  Institut für Numerische und Angewandte Mathematik,
  Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission

- Prof. Dr. Stephan Huckemann,
  Institut für Mathematische Stochastik,
  Georg-August-Universität Göttingen

- Prof. Dr. Victor Pidstrygach,
  Mathematisches Institut,
  Georg-August-Universität Göttingen

- Prof. Dr. Gerlind Plonka-Hoch,
  Institut für Numerische und Angewandte Mathematik,
  Georg-August-Universität Göttingen

- Prof. Dr. Anita Schöbel,
  Institut für Numerische und Angewandte Mathematik,
  Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 23.06.2015

„Jung, gieh raus un guck Dir de Welt an." - Franz Neumann (1925 − 2014)

Danksagung

In der Eröffnung meiner Dissertation möchte ich die Gelegenheit nutzen, Worte des Dankes zu formulieren. An erster Stelle gilt der Dank meinem Doktorvater Prof. Dr. D. Russell Luke, der sich im Frühling des Jahres 2011 dazu entschied, mich, als damaligen Masterstudenten im Bereich der geometrischen Gruppentheorie, auf dem Weg zur Promotion zu betreuen. Ich danke ihm für die hervorragende Betreuung, für die Geduld und für die vielen fruchtbaren und ermutigenden Diskussionen, die mich immer motivierten, das vorliegende Werk fertig zu stellen.

Ich danke ebenfalls Prof. Dr. Max Wardetzky, dass er sich bereit erklärte, zum einen im Werdegang meines Doktorandenstudiums als Zweitbetreuer zu fungieren, zum anderen für die Funktion als Zweitgutachter dieser Arbeit.

Diese Arbeit entstand unter der Finanzierung des DFG Graduiertenkollegs 1023 „Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden" und des DFG Sonderforschungsbereichs 755 „Nanoscale Photonic Imaging". Die Mitarbeit in den beiden Projekten ermöglichte mir die Einordnung der Arbeit in einen breiteren Kontext, insbesondere die Verknüpfung mit physikalischen Anwendungen. Ohne diese Förderung wäre es auch nicht möglich gewesen, meine Ergebnisse auf internationalen Konferenzen zu präsentieren.

Ich danke den Kollegen und Mitarbeitern, einschließlich den ehemaligen, am Institut für Numerische und Angewandte Mathematik für die angenehme Arbeitsatmosphäre und für stets offene Ohren.

Besonderer Dank gilt meinem guten Freund und ehemaligen Bürokollegen Dr. Robert Hesse, mit dem 12 m^2 niemals zu wenig wurden und mit dem mir die gemeinsame Arbeit und Zeit große Freude bereitete. Selbiger Dank gebührt meinem Freund und Kollegen Dr. Marius Wischerhoff, der immer zum gegenseitigen Austausch bereit ist, der gemeinsame Aktivitäten und Hilfsbereitschaft auch gern über eigene Interessen hebt und ohne dessen Korrekturarbeit dieses Werk nicht in dieser Form vorliegen würde.

I thank Hieu Thao Nguyen for suggestions and proofreading this manuscript. Weiterhin danke ich Corinna Krüger und Rebecca Nahme für das Gegenlesen dieser Arbeit und nicht zuletzt für angenehme und gemütliche Gespräche.

Für die Möglichkeit, meinen eigenen Weg gehen zu können, für immerwährende Unterstützung und Motivation danke ich meinen Eltern Marie-Theres und Werner Neumann sowie meiner Schwester Anke Neumann.

Zuletzt danke ich Meike für ihre bedingungslose Unterstützung, für ihre Liebe und dafür, mich immer, wenn ich den Weg nicht mehr sah, aufzuheben und mich auf diesen zurück zu führen.

Contents

Contents

List of Figures

1 Introduction

If we jump back and forth between two sets by using the shortest possible jumps, how does the outcome of this procedure depend on the sets? When does this iteration stagnate? These are the fundamental questions motivating this thesis.

More general, in this thesis, we give an analysis of fixed point algorithms involving projections onto closed, not necessarily convex, subsets of finite dimensional vector spaces. These methods are used in applications such as imaging science, signal processing, and inverse problems. The tools used in the analysis place this work at the intersection of optimization and variational analysis.

For example, in physical applications such as the phase retrieval problem, it is possible to model the problem as a problem of finding intersections of two or more closed subsets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$. The reason for this approach is that, in these models, we may be able to find closest points in each of the sets at a low cost, but finding the intersection immediately may be costly or impossible. A mathematical problem of the kind

$$\text{find } \bar{x} \in \Omega_1 \cap \Omega_2$$

is called a *feasibility problem*. We study theoretical properties of fixed point algorithms applied to nonconvex feasibility problems. Our study focusses on two prominent representatives, namely the *method of alternating projections* and the *Douglas-Rachford algorithm*. The alternating projections algorithm reads as follows. Denote for a point $x \in \mathbb{R}^n$ by $P_{\Omega_1} x$ the closest point in Ω_1 relative to x. Given an initial point $x^0 \in \mathbb{R}^n$, we generate the sequence $\{x^k\}_{k \in \mathbb{N}}$ via

$$x^{k+1} := P_{\Omega_1} P_{\Omega_2} x^k.$$

If we write $R_{\Omega_1} = 2P_{\Omega_1} - Id$, then for a given $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}_{k \in \mathbb{N}}$ of iterates generated by the Douglas-Rachford algorithm is given by

$$x^{k+1} := \frac{1}{2} \left(R_{\Omega_1} R_{\Omega_2} + Id \right) x^k.$$

We give the precise definitions in Chapter 4. The method of alternating projections goes back at least to von Neumann in (von Neumann, 1951), and since then it has been an object of broader research. The classical literature is restricted to the case where both sets Ω_1 and Ω_2 are convex subsets (Cheney and Goldstein, 1959), (Gubin et al., 1967), (Bauschke and Borwein, 1993), (Bauschke and Borwein, 1996), (Bauschke et al., 1997), (Bauschke et al., 2004), (Deutsch and Hundal, 2006a), (Deutsch and Hundal, 2006b), (Deutsch and Hundal, 2008).

Recent studies have identified the interplay between regularity of the sets and regularity of their intersection as the key to a general analysis. For instance, the aforementioned convexity is an assumption on the regularity of the sets. Additional assumptions, like a nonempty interior or a local angle between the sets are a key element to obtain rates of convergence.

Among the first studies of these methods in more general settings is the paper by Combettes and Trussel (Combettes and Trussell, 1990), but it was not until recently that a *quantitative* analysis was achieved (Hesse and Luke, 2013). This and additional tools from variational analysis gave rise to a deeper analysis of the nonconvex setting (Bauschke et al., 2013a), (Bauschke et al., 2013b), (Bauschke et al., 2014b), (Hesse et al., 2014).

For the Douglas-Rachford algorithm, which is known since work of Douglas and Rachford (Douglas and Rachford, 1956), and the work of Lions and Mercier (Lions and Mercier, 1979), the development has been slower. The work of Hesse and Luke (Hesse and Luke, 2013) is among the first results of local convergence with rates under absence of convexity.

We apply both methods to solve an optimization problem arising from the field of sparsity optimization. That is, we seek the vector with least possible nonzero entries satisfying an underdetermined system of linear equations. This problem, known as the *compressed sensing problem*, has gained a large popularity since the work of Candès and Tao (Candès and Tao, 2005), and it is especially not convex. There, the authors approach the problem by seeking the vector x satisfying $Mx = p$ with least possible ℓ_1-norm. Under suitable assumptions, this solution of ℓ_1-minimization coincides with the sparsest possible vector x satisfying $Mx = p$. In (Blumensath and Davies, 2009), (Blumensath and Davies, 2010), and (Beck and Teboulle, 2011), the ansatz of making an a priori assumption to the sparsity of the solution was suggested. We follow that ansatz in this work.

The second application of alternating projections comes from the physical problem of phase retrieval. Using the idea in (Candès et al., 2011), we obtain a formulation similar to the search for a vector with least possible nonzero entries satisfying a linear system. Instead of minimizing the nuclear norm, as done in (Candès et al., 2011), we propose again an a priori assumption to the solution and apply the method of alternating projections.

We give an analysis of projection methods in a nonconvex setting. In sparsity optimization, it is important to point out that up to now the sufficient conditions for convergence of projection methods to the correct solution are not competitive to those of ℓ_1-minimization. By weakening these sufficient conditions for convergence, projection methods may at least be on the same level of performance as other known methods. Also in phase retrieval, the analyzed setting in rank minimization suffers from the curse of dimensionality and yields high runtimes. This work is meant as a foundation for the development of new ways to solve existing problems.

However, the results shown in this thesis give new insights in the behavior of the method of alternating projections and of Douglas-Rachford, and these insights are the contribution of this work.

This thesis can be devided into two main parts, based on the underlying optimization problems. The first one is the compressed sensing problem, where we seek a solution to an underdetermined linear system with least possible nonzero entries. Because the problem is NP-hard, we relax it to a feasibility problem with two sets, namely, the set A_s of vectors with at most s nonzero entries and, for a linear mapping $M : \mathbb{R}^n \to \mathbb{R}^m$, the affine subspace B of vectors x satisfying $Mx = p$ for $p \in \mathbb{R}^m$ given. This problem will be referred to as the *sparse affine feasibility problem*. First, we name several geometric properties of the nonconvex set A_s, including (ε, δ)-subregularity, recently developed in (Hesse and Luke, 2013). Moreover, we show the explicit shape of the second-order subdifferential, defined in (Mordukhovich and Rockafellar, 2012), of the function that counts the number of nonzero entries in a vector.

For the Douglas-Rachford algorithm, we give the proof of linear convergence to a fixed point in the case of a feasibility problem of two affine subspaces. We show that the projection of this fixed point onto one of the affine subspaces is a solution to the feasibility problem. This result first appeared in (Hesse et al., 2014), and it is the first of several proofs of linear convergence of Douglas-Rachford that came up shortly after (Bauschke et al., 2014a), (Demanet and Zhang, 2013). It allows us to conclude a result of local linear convergence of the Douglas-Rachford algorithm in the sparse affine feasibility problem. Proceeding, we name sufficient conditions for the alternating projections algorithm to converge to the intersection of an affine subspace with lower level sets of point symmetric (i.e., $f(x) = f(-x)$ for all x), lower semicontinuous, subadditive functions. The theorem and its proof are inspired by (Beck and Teboulle, 2011) but shows convergence of alternating projections instead of iterative hard thresholding. Since the function that counts the number of nonzero entries of a vector satisfies all these properties, this implies convergence of alternating projections to a solution of the sparse affine feasibility problem. Together with a result of local linear convergence of the alternating projections algorithm in (Hesse et al., 2014), this allows us to deduce linear convergence after finitely many steps for *any* initial point of a sequence of points generated by the alternating projections algorithm. The conditions guaranteeing this convergence behavior are very strong, and we show some limitations of these conditions.

In contrast to the results on global convergence to the true solution of alternating projections in sparse affine feasibility, we generalize the setting to the search for the intersection of a polyhedral set and the set A_s of sparse vectors. Again, we show convergence to a fixed point of the alternating projections algorithm for any inital point. This results in a formulation of necessary conditions for global convergence of the method in the sparse affine feasibility problem.

The second part of this dissertation deals with the minimization of the rank of matrices satisfying a set of linear equations. As in the case of sparse affine feasibility, we relax the problem of minimizing the rank of a function to a feasibility problem between the set of matrices of fixed rank and the affine subspace given by the linear equations. This problem will be called *rank constrained affine feasibility problem*. The motivation for the analysis of the rank minimization problem comes from the physical application of phase retrieval and a reformulation of the same as a rank minimization problem (Candès et al., 2011). We show that, locally, the method of alternating projections must

converge at linear rate to a solution of the rank constrained affine feasibility problem. The final result is on sufficient conditions for global convergence of the same method, which are related to the analogous result in the sparse affine feasibility problem.

This work is organized as follows. In Chapter 2, we introduce several definitions including those of different notions of regularity. The notation used there is based on the book by Rockafellar and Wets, (Rockafellar and Wets, 1998). The presentation of different regularities is in the spirit of (Hesse and Luke, 2013). We give an introduction to sparsity optimization together with one of the main problems in Chapter 3. We check how the different regularity tools defined in Chapter 2 can be applied to the set of sparse vectors in the Euclidean space $\mathbb{R}^n$. To obtain convergence results for projection methods applied to sparsity optimization, we name several properties of projection operators in Chapter 4. We review classical and recent convergence results on alternating projections and a recent result on linear convergence of Douglas-Rachford in the case of affine subspaces. The latter first appeared in (Hesse et al., 2014). In Chapter 5, we present the first of three main results of this thesis. Namely, under strong assumptions, we show that the method of alternating projections converges to lower level sets of subadditive, lower semicontinuous, point symmetric functions for all initial points at a linear rate. We show the consequences of this result for sparsity optimization afterwards, as well as the behavior of Douglas-Rachford in the case of sparsity optimization.

A generalization of the case of alternating projections in sparsity optimization is presented in Chapter 6. We show that the method, if applied to a polyhedral set and the set of sparse vectors, generates a sequence of iterates which *always* converges to a finite set of cluster points. This is our second main result. Chapter 7 builds a link between sparsity optimization and rank minimization. We show there, via an embedding of the set of sparse vectors into the set of matrices of low rank, how regularity properties of sets of matrices translate to sets of vectors linked to these sets of matrices. A prominent representative of the latter, namely, the set of matrices of fixed rank, will be analyzed in more detail in Chapter 8. This analysis includes geometric properties and the formulation of the projector onto this set. In Chapter 9, we present the physical problem of phase retrieval, which is motivating the theoretical analysis in Chapter 10. In the latter, we study properties of the alternating projections operator, applied to a translation of the phase retrieval problem to a rank minimization problem. We show local linear convergence of alternating projections, which is our third main result. We also present a specialization of the first main result, presented in Chapter 5, to the problem of minimizing the rank of matrices with respect to affine constraints. Afterwards, in Chapter 11, we present numerical demonstrations of the theory developed in the former chapters.

2 Preliminaries

We start by introducing several notations and theoretical foundations for the following chapters.

2.1 Functions

The first definition is of set-valued mappings. It extends the notion of a classical function. Detailed studies on set-valued mappings are given in (Rockafellar and Wets, 1998, Chapter 5) and in (Dontchev and Rockafellar, 2014) with more examples and historical notes.

Definition 2.1.1 (set-valued mappings, (Rockafellar and Wets, 1998, p. 148)). *Let U, V be arbitrary sets. A mapping $F : U \rightrightarrows V, u \mapsto F(u) \subset V$, giving for every $u \in U$ a subset $F(u) \subset V$, is called a set-valued mapping. For a set-valued mapping F, it is always possible to give the inverse mapping $F^{-1} : V \rightrightarrows U$ assigning to every point $y \in V$ the set $\{x \in U \mid y \in F(x)\}$. The inverse map then is a set-valued mapping itself. A set-valued function $F : U \rightrightarrows V$ is single-valued at $x \in U$ if either $F(x) = \varnothing$ or there exists $y \in V$ such that $F(x) = \{y\}$. By abuse of notation and if it is clear from the context, write $F(x) = y$ if $F(x) = \{y\}$.*

With this notation, a classical function $f : U \to V$, where for each $u \in U$ there exists at most one $v \in V$ such that $f(u) = v$, becomes a single-valued function. Further, for any such f it is possible to give an (in general set-valued) inverse map $f^{-1} : V \rightrightarrows U$. In the following definitions we have to distinguish between set-valued functions and classical functions.

Definition 2.1.2 (domain, range, and graph of a set-valued mapping (Rockafellar and Wets, 1998, pp. 148–149)). *Define for a set-valued mapping $F : V \rightrightarrows W$ the sets*

$$\begin{aligned} \operatorname{dom}(F) &\coloneqq \{v \mid F(v) \neq \varnothing\}, \\ \operatorname{range}(F) &\coloneqq \{w \mid \exists v : \ w \in F(v)\}, \\ \operatorname{gph}(F) &\coloneqq \{(x, u) \mid u \in F(x)\}. \end{aligned} \tag{2.1}$$

Definition 2.1.3 (domain, range, and graph of a function (Rockafellar and Wets, 1998, p. 5)). *Define for a function $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ the sets*

$$\begin{aligned} \operatorname{dom}(f) &\coloneqq \{v \mid f(v) < \infty\}, \\ \operatorname{range}(f) &\coloneqq \{w \mid \exists v : \ w = f(v)\}, \\ \operatorname{gph}(f) &\coloneqq \{(x, u) \mid u = f(x)\}. \end{aligned} \tag{2.2}$$

Further, a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called proper *if $f(x) < \infty$ for at least one $x \in \mathbb{R}^n$ and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$.*

Definition 2.1.4 (fixed points). *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping. We define the set of* fixed points *of F by*

$$\text{Fix}(F) := \{x \in \mathbb{R}^n \mid x \in F(x)\}. \tag{2.3}$$

Further, we define the set of stable fixed points *of F via*

$$\text{StFix}(F) := \{x \in \text{Fix}(F) \mid F(y) = F(x) \quad \text{for all } y \in F(x)\}. \tag{2.4}$$

We refer to the example in Figure 6.1 to show the need for a definition of stable fixed points. Note that we always have $\text{StFix}(F) \subseteq \text{Fix}(F)$.

In the following, we give the definition of lower level sets and the lower semicontinuity of a function. According to the commentary in (Rockafellar and Wets, 1998, Chapter 1), the usage of these two constructions dates at least back to lectures of Fenchel (see (Fenchel, 1951)). Due to its general formulation, the following definition can be found in (Bauschke and Combettes, 2011, Definition 1.4).

Definition 2.1.5 (lower level sets). *Let $\mathbb{X}$ be a vector space. Define for a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ the* lower level set *of f at height $t \in \mathbb{R}$ by*

$$\text{lev}_{\leq t} f := \{x \in \mathbb{X} \mid f(x) \leq t\}. \tag{2.5}$$

Instead of giving the usual definition of lower semicontinuity in the literature, we cite the equivalence given for instance in (Bauschke and Combettes, 2011, Theorem 1.24). The reason for this is purely esthetic. It is also nicer to define continuity of functions via the property that the preimages of open sets are open sets instead of giving an ε, δ-criterion.

Definition 2.1.6 (subadditive function). *A function $f : \mathbb{X} \to \overline{\mathbb{R}}$ will be called* subadditive *if $f(x + y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{X}$.*

Definition 2.1.7 (lower semicontinuity). *A function $f : \mathbb{X} \to \overline{\mathbb{R}}$ will be called* lower semi-continuous *at every point in $\mathbb{X}$ if the lower level sets $\text{lev}_{\leq t} f$ are closed in $\mathbb{X}$ for all t.*

Definition 2.1.8 (epigraph). *For a function $f : \mathbb{X} \to \mathbb{R}$ define its* epigraph *by*

$$\text{epi} f := \{(x, \alpha) \in \mathbb{X} \times \mathbb{R} \mid \alpha \geq f(x)\}. \tag{2.6}$$

We introduce the notion of subdifferentials. Subdifferentials are a generalization of differentials for the cases when functions are not everywhere differentiable. For example, the indicator function ι_C of a closed set $C \subset \mathbb{R}^n$ is not differentiable at the boundary of C. Yet, the subdifferential of ι_C at boundary points exists (see (2.8) for the Definition of ι_C).

Definition 2.1.9 (subdifferential of a function (Rockafellar and Wets, 1998, Definition 8.3)). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a lower semicontinuous function with $\bar{x} \in \text{dom}(f)$.*

1. *The vector v is called a* regular subgradient *of f at $\bar{x}$, denoted by $v \in \hat{\partial} f(\bar{x})$, if*

$$\liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0. \tag{2.7}$$

2. *The vector v is a* limiting subgradient *of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if there are sequences $x^k \to \bar{x}$ with $f(x^k) \to f(\bar{x})$, and $v^k \in \hat{\partial} f(x^k)$ with $v^k \to v$.*

The set of regular subgradients and limiting subgradients at a point $\bar{x}$ will be called regular subdifferential *and* limiting subdifferential, *respectively.*

Remark 2.1.10. *The regular subddifferential in Definition 2.1.9 is also called Fréchet subdifferential. The limiting subdifferential is due to Mordukhovich. It can also be defined in terms of normal cones (Definition 2.3.3), see (Mordukhovich, 2006, Theorem 1.89).*

Definition 2.1.11 (subdifferentially regular function). *A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is subdifferentially regular if the subdifferentials $\hat{\partial} f$ and ∂f defined in Definition 2.1.9 coincide.*

2.2 Optimization

The definition of subdifferentials in Definition 2.1.9 leads to optimization problems since subdifferentials can be used to formulate optimality conditions. First, for a subset $\Omega \subset \mathbb{R}^n$, we define the *indicator function* of Ω via

$$\iota_\Omega : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad x \mapsto \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases} \tag{2.8}$$

In Chapter 3, we introduce the basic minimization problem of sparsity optimization. As a foundation, we define a more general family of optimization problems.

Definition 2.2.1. *Let $M : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an arbitrary function with $\text{dom}(f) \neq \emptyset$, and let $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be another function with $\text{dom}(g) \neq \emptyset$. Define now a fundamental optimization problem of finding a minimizer of the composition $f + g \circ M$, i.e.,*

$$\text{find} \operatorname*{argmin}_{x \in \mathbb{R}^n} \ \{ f(x) + g(Mx) \}. \tag{2.9}$$

Theorem 2.2.2 (Fermat's rule (Rockafellar and Wets, 1998, Theorem 10.1)). *If a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ has a local minimum at $\bar{x}$, then*

$$0 \in \hat{\partial} f(\bar{x}), \quad 0 \in \partial f(\bar{x}). \tag{2.10}$$

If f is convex (Definition 2.3.1), then (2.10) *is not only necessary for a local minimum but sufficient for a global minimum.*

A subclass of (2.9) are the so-called *feasibility problems*.

Definition 2.2.3 (feasibility problem). *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be nonempty. A mathematical problem of the form*

$$\text{find } x \in \Omega_1 \cap \Omega_2 \tag{2.11}$$

is called a feasibility problem.

There is a connection between feasibility problems and optimization problems of the type (2.9): a point x is a solution to (2.11) if and only if x is a zero of $f + g \circ M$ in (2.9), where $f = \iota_{\Omega_1}$, $g = \iota_{\Omega_2}$, and $M = \text{Id}$.

As we will see in the following chapters, solving feasibility problems highly depends on the geometries of the sets involved.

2.3 Geometry

"Eine Punktmenge heißt *konvex*, wenn sie mit zwei Punkten stets deren Verbindungs-strecke enthält"[1].

Definition 2.3.1 (convex set & convex function). *Let $\mathbb{X}$ be a vector space.*

- *A subset $\Omega \subset \mathbb{X}$ is* convex *if for any two points $x, y \in \Omega$ the relation*

$$tx + (1 - t)y \in \Omega \text{ for all } t \in [0, 1] \tag{2.12}$$

  holds.

- *For an arbitrary subset $D \subset \mathbb{R}^n$, define the* convex hull *of D (Rockafellar and Wets, 1998, Chapter 2.E.) as*

$$\text{con } D := \bigcap_{\substack{C \supseteq D \\ C \subset \mathbb{R}^n \text{ convex}}} C. \tag{2.13}$$

- *A function $f : \mathbb{X} \to \mathbb{R}$ is called* convex *if its epigraph is a convex set. For a function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, define its* convex hull *as, see (Rockafellar and Wets, 1998, Proposition 2.31),*

$$(\text{con } g)(x) := \inf \left\{ \sum_{j=0}^{n} \lambda_j g(x_j) \, \middle| \, \sum_{j=0}^{n} \lambda_j x_j = x, \lambda_j \geq 0, \sum_{j=0}^{n} \lambda_j = 1 \right\}, \tag{2.14}$$

  or, equivalently, con g *is the greatest convex function majorized by g.*

Convexity of sets and functions and its implications for optimization give rise to a broad field of mathematics. We refer to (Rockafellar, 1997) as a standard work on this

---

[1]See (Bonnesen and Fenchel, 1934, p. 3). This is of course not the first usage of the terminology "convex". But, the book of Bonnesen and Fenchel is one of the first works entirely devoted to the topic of convex geometry.

topic. At this point we introduce the *Minkowski sum* of sets. If A and B are subsets of a vector space $\mathbb{R}^n$, then we define the set

$$A + B := \{a + b \mid a \in A, b \in B\}. \tag{2.15}$$

Before proceeding with projections, we give the notion of the *affine hull* of a subset $\Omega \subset \mathbb{R}^n$

$$\text{aff } \Omega := \{\lambda x + (1 - \lambda)y \mid x, y \in \Omega, \lambda \in \mathbb{R}\}. \tag{2.16}$$

The most important tool analyzed in this work will be the projection onto a set, clarified in the following definition.

Definition 2.3.2 (projector and projection). *For $\Omega \subset \mathbb{R}^n$ closed and nonempty, the mapping $P_\Omega : \mathbb{R}^n \rightrightarrows \Omega$ defined by*

$$P_\Omega x := \underset{y \in \Omega}{\text{argmin}} \|x - y\| \tag{2.17}$$

is called the projector *onto Ω. A point $\bar{x} \in P_\Omega x$ is called a* projection *with respect to the norm $\|\cdot\|$. The projector is in close relation to the distance of a point $x \in \mathbb{R}^n$ to a set $\Omega \subset \mathbb{R}^n$, given by*

$$d_\Omega(x) := \min_{y \in \Omega} \|x - y\|. \tag{2.18}$$

By the mapping $R_\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, denote the reflector *on Ω defined as*

$$R_\Omega x := 2P_\Omega x - x. \tag{2.19}$$

As an initial intuitive example for the projector and the reflector, consider Figure 2.1.
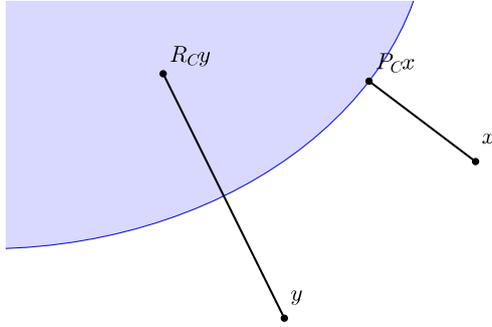
Figure 2.1: Points projected and reflected on a set

We restrict ourselves to the finite dimensional setting. Further, if not stated otherwise, the norm in (2.17) will always be the norm on $\mathbb{R}^n$ induced by the inner product. The expression in (2.17) is valid for any kind of set Ω as a set-valued mapping. Apparently, if Ω is an open subset of $\mathbb{R}^n$, then the minimizing argument does not exist. In that case, the projector would be an empty-valued mapping. In a finite dimensional Hilbert space, closedness of Ω is both necessary and sufficient for $P_\Omega x$ to be nonempty (see

(Bauschke and Combettes, 2011, Corollary 3.13)).

   Additional theory has been developed if the geometries of the sets at hand we would like to project onto are more restrictive. This leads to the question of regularity conditions of the sets. Some properties are already determined by a global structure, for example if the set Ω is a manifold or a submanifold (Lewis and Malick, 2008).

Definition 2.3.3 (normal cones (Bauschke et al., 2013b, Definition 2.1), based on (Mordukhovich, 2006, Definition 1.1 and Theorem 1.6)). *The* proximal normal cone $N_\Omega^P(\bar{x})$ *to a closed nonemtpy set $\Omega \subset \mathbb{R}^n$ at a point $\bar{x} \in \Omega$ is defined by*

$$N_\Omega^P(\bar{x}) := \begin{cases} \text{cone}(P_\Omega^{-1}(\bar{x}) - \bar{x}) & \text{if } \bar{x} \in \Omega, \\ \varnothing & \text{if } \bar{x} \notin \Omega. \end{cases} \tag{2.20}$$

The limiting normal cone $N_\Omega(\bar{x})$, *or simply the* normal cone, *is defined as the set of all vectors that can be written as the limit of proximal normals; that is, $\bar{v} \in N_\Omega(\bar{x})$ if and only if there exist sequences $(x^k)_{k \in \mathbf{N}}$ in Ω and $(v^k)_{k \in \mathbf{N}}$ in $N_\Omega^P(x^k)$ such that $x^k \to \bar{x}$ and $v^k \to \bar{v}$, as $k \to \infty$.*

   We refer to Figure 2.2 as an illustration of the different types of normal cones.

Definition 2.3.4 (tangent cones (Rockafellar and Wets, 1998, Definition 6.1)). *A vector $w \in \mathbb{R}^n$ is* tangent *to a set $\Omega \subset \mathbb{R}^n$ at a point $\bar{x} \in \Omega$, written $w \in T_\Omega(\bar{x})$, if there exists a sequence $(\tau^k)_{k \in \mathbb{N}} \subset \mathbb{R}$ with $\tau^k > 0, \tau^{k+1} \le \tau^k$ for all k and $\lim_{k \to \infty} \tau^k = 0$ as well as a sequence $(x^k)_{k \in \mathbb{N}} \subset \mathbb{R}^n$ such that $x^k \in \Omega$ for all k and $\lim_{k \to \infty} x^k = \bar{x}$ satisfying*

$$\frac{x^k - \bar{x}}{\tau^k} \to w, \quad \text{as } k \to \infty. \tag{2.21}$$

   We note that if Ω is a submanifold of $\mathbb{R}^n$, then the tangent cone in Definition 2.3.4 and the normal cone in Definition 2.3.3 as well as the tangent space and the normal space known from differential geometry coincide (see (Rockafellar and Wets, 1998, Example 6.8)).
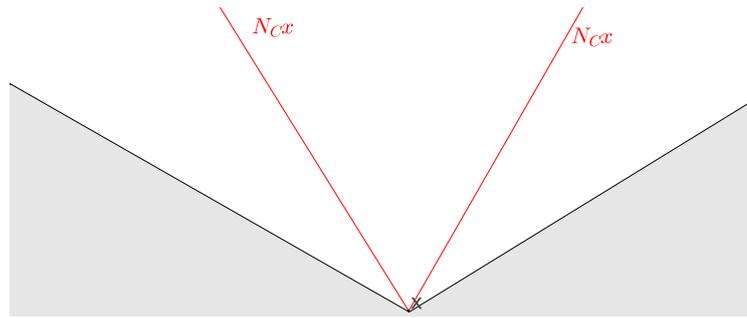
Figure 2.2: The red lines represent the limiting normal cone at the point x. Note that the proximal normal cone at x is 0.

The definitions of normal cones and tangent cones are formulated without any restrictions on the set Ω. If, for example, Ω is an open subset of $\mathbb{R}^n$, then we have $N_\Omega(x) = \{0\}$ and $T_\Omega(x) = \mathbb{R}^n$ at every point $x \in \Omega$.

Definition 2.3.5 (minimal distance of sets). *For two nonempty and closed subsets Ω_1, Ω_2 of $\mathbb{R}^n$ define the* minimal distance *between these two sets by*

$$\mathrm{d}(\Omega_1, \Omega_2) := \inf_{(x,y) \in \Omega_1 \times \Omega_2} \|x - y\|. \tag{2.22}$$

Remark 2.3.6. *If Ω_1, Ω_2 are closed subsets of $\mathbb{R}^n$, then this is not sufficient for $\mathrm{d}(\Omega_1, \Omega_2)$ to be a minimum instead of an infimum. A counterexample with two closed sets would be*

$$\Omega_1 := \left\{ x \in \mathbb{R}^2 \mid x_2 \geq \mathrm{e}^{x_1} \right\}, \ \Omega_2 := \left\{ x \in \mathbb{R}^2 \mid x_2 = 0 \right\}.$$

Both sets are closed, and for every $\varepsilon > 0$ there exists $x_1 \in \mathbb{R}$ such that $\mathrm{e}^{x_2} - x_1 \leq \varepsilon$. Thus, the infimum in (2.22) is zero.

We proceed with the definition of best approximation pairs between two closed sets. In (Bauschke et al., 2004), a formulation for closed convex sets C_1, C_2 as the pair of points

$$(v_1, v_2) \in C_1 \times C_2 \text{ such that } \|v_1 - v_2\| = \inf \|C_1 - C_2\| \tag{2.23}$$

has been given. Definition 2.3.5 is inspired by this formulation. The expression (2.23) is designed for convex sets since for convex sets C_1, C_2 the functions

$$f_1 : C_1 \to \mathbb{R}, v \mapsto d_{C_2}(v) \text{ and } f_2 : C_2 \to \mathbb{R}, w \mapsto d_{C_1}(w)$$

are convex functions. Hence, the function

$$f : C_1 \times C_2 \to \mathbb{R}, \quad (v_1, v_2) \mapsto \frac{1}{2} f_1(v_1) + \frac{1}{2} f_2(v_2) \tag{2.24}$$

is a convex function. A pair of points satisfying (2.23) is then a local minimum of (2.24). Because local minima of convex functions are global minima, it is equivalent to seek points as in (2.23) or to find local minima of (2.24). If the closed sets at hand are not convex anymore, the formulation in (2.23) will not cover local minima of (2.24). Because the fixed points of the algorithms we analyze in this thesis are related to all local minima of (2.24), it is necessary to have a local version of (2.23). See Figure 2.3 for an example of two nonconvex sets with local best approximation pairs.

Definition 2.3.7 (local best approximation pairs). *Given two closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, a pair of points $(v_1, v_2) \in \Omega_1 \times \Omega_2$ is a* local best approximation pair *if $v_2 \in P_{\Omega_2} v_1$ and $v_1 \in P_{\Omega_1} v_2$, and if there exist $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ such that*

$$d_{\Omega_2}(w_1) \geq d_{\Omega_2}(v_1) \text{ and } d_{\Omega_1}(w_2) \geq d_{\Omega_1}(v_2) \tag{2.25}$$

for all $w_1 \in \Omega_1 \cap \mathbb{B}_{\varepsilon_1}(v_1)$ and for all $w_2 \in \Omega_2 \cap \mathbb{B}_{\varepsilon_2}(v_2)$, cf. (2.18).

Remark 2.3.8. *The above defition on best approximation pairs does not exclude points in a possibly nonempty intersection of Ω_1 and Ω_2. On the other hand, it is not necessary for best approximation pairs to exist at all. See Remark 2.3.6 for an example. We compare this definition of local best approximation pairs with the definition of local best approximation points in (Luke, 2008, Definition 3.3). We note that the convexity of Ω_1 in (Luke, 2008, Definition 3.3) can be omitted (Luke, 2015). In contrast to Definition 2.3.7, the Definition in (Luke, 2008) is just one-sided. This means that, for a local best approximation point $x \in \Omega_2$, the point $P_{\Omega_1}x$ is not necessarily a local best approximation point to Ω_2. For our purposes, we define a two-sided version here.*

Definition 2.3.9 (local best approximation point (Luke, 2008, Definition 3.3)). *For a convex set Ω_1 and for a nonconvex set Ω_2, a point $x \in \Omega_2$ is a local best approximation point if there exists a neighborhood $\mathbb{B}_\varepsilon(x)$ such that $d_{\Omega_1}(x) \leq d_{\Omega_1}(y)$ for all $y \in \mathbb{B}_\varepsilon(x) \cap \Omega_2$.*

In a close relation to best approximation pairs stands the gap vector. When dealing with nonconvexity, we have to define a local version.

Definition 2.3.10 ((local) gap vector (Luke, 2008, Equation 1.7)). *Let Ω_1, Ω_2 be closed and convex subsets of $\mathbb{R}^n$. For the set $G_{12} := \overline{\Omega_1 - \Omega_2}$, define the* gap vector

$$g := P_{G_{12}}0. \tag{2.26}$$

For two closed subsets $\Omega_3, \Omega_4 \subset \mathbb{R}^n$, define

$$B(\Omega_3, \Omega_4) := \{(a,b) \in \Omega_3 \times \Omega_4 \mid (a,b) \text{ is a local best approximation pair}\} \tag{2.27}$$

The set of local gap vectors G *via*

$$G := \{a - b \mid (a,b) \text{ is a local best approximation pair}\}. \tag{2.28}$$

Lemma 2.3.11. *Let $U, V \subset \mathbb{R}^n$ be affine subspaces. Then there exist $u \in U$ and $v \in V$ such that (u,v) is a best approximation pair between U and V.*

Proof. First, we note that the Minkowski sum $U + V$ of the affine subspaces is again an affine subspace. Hence, the difference $U - V$ is an affine subspace and especially a closed, convex set. That means that there exists a point $g \in U - V$ such that $P_{U-V}0 = g$. Hence, there exist $u \in U$ and $v \in V$ such that $u - v = g$. Because g is the gap vector between U and V, we have a pair $(u,v) \in U \times V$ such that (u,v) is a best approximation pair between U and V. $\qquad\square$

2.4 Regularity

As already mentioned, the convergence behavior of the algorithms used in this thesis heavily relies on properties of the sets and their intersections involved in the feasibility problems on hand. These properties are called *regularity conditions*. We distinguish
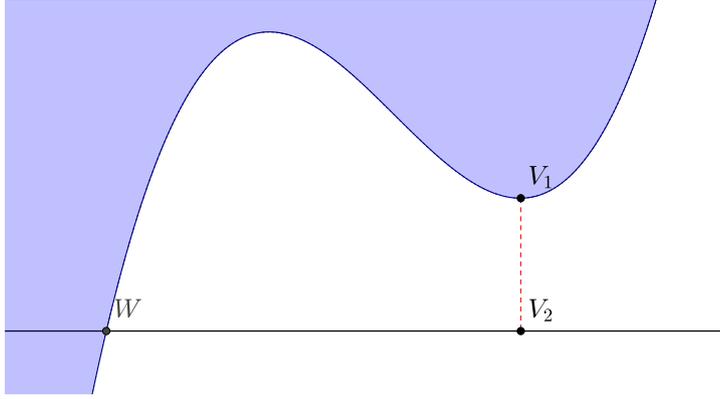
Figure 2.3: Two different kinds of best approximation pairs: the point W is in the intersection of the horizontal line and the blue set, the pair (V_1, V_2) attains a local minimum of the function in (2.24).

between two kinds of regularity conditions. On the one hand, we define regularities of sets in Section 2.4.1. These describe properties of a single set Ω at hand. On the other hand, we name regularities of collections of sets in Section 2.4.2. There, the interplay between different sets is studied.

2.4.1 Regularity of Sets

Definition 2.4.1. *Let $\Omega \subset \mathbb{R}^n$ be nonempty.*

1. *The set $\Omega \subset \mathbb{R}^n$ is (ε, δ)-subregular at $\bar{x}$ with respect to $U \subset \mathbb{R}^n$ if there exist $\varepsilon \geq 0$ and $\delta > 0$ such that*

$$\langle v, z - y \rangle \leq \varepsilon \|v\| \|z - y\| \tag{2.29}$$

   holds for all $y \in \Omega \cap \mathbb{B}_\delta(\bar{x})$, $z \in U \cap \mathbb{B}_\delta(\bar{x})$, $v \in N_\Omega^P(y)$. For simplicity, Ω is (ε, δ)-subregular at $\bar{x}$ if $U = \{\bar{x}\}$.

2. *If $U = \Omega$ in (1), then Ω is (ε, δ)-regular at $\bar{x}$.*

3. *If Ω is a closed set, then Ω is Clarke regular at $\bar{x} \in \Omega$ if, for all $\varepsilon > 0$, there exists $\delta > 0$ such that any two points $x, y \in \mathbb{B}_\delta(\bar{x})$ with $x \in \Omega$ and any $z \in P_\Omega y$ satisfy*

$$\langle x - \bar{x}, y - z \rangle \leq \varepsilon \|x - \bar{x}\| \|y - z\|. \tag{2.30}$$

4. *If for all $\varepsilon > 0$ there exists $\delta > 0$ such that (2.29) holds for all $x, y \in \Omega \cap \mathbb{B}_\delta(\bar{x})$ and $v_x \in N_\Omega(x)$, then Ω is said to be super-regular at $\bar{x}$.*

5. *If Ω is closed, then $\Omega \subset \mathbb{R}^n$ is said to be prox-regular at a point $\bar{x} \in \Omega$ if there exists a neighborhood U of $\bar{x}$ such that the projection $P_C x$ is single-valued for all $x \in U$.*

Remark 2.4.2. *Let $\mathcal{H}$ be a Hilbert space. We refer to (Deutsch, 2001, Chapter 2, p. 21) for the definition of a Chebyshev set: a subset $\Omega \subset \mathcal{H}$ is* Chebyshev, *if P_Ω is a single-valued mapping for all points $x \in \mathcal{H}$. By (Deutsch, 2001, Theorem 12.7), in a finite dimensional inner product space $\mathbb{R}^n$, a nonempty subset $C \subset \mathbb{R}^n$ is closed and convex if and only if C is Chebyshev. With Definition 2.4.1, it is possible to give an alternative definition of convex sets in finite dimensional spaces. We recall the best approximation property for closed convex sets: by (Deutsch, 2001, Theorem 4.1), if a set $\Omega \subset \mathbb{R}^n$ is closed and convex , then for all $x \in \mathbb{R}^n$, $P_\Omega x = z$ if and only if*

$$\langle x - z, z - y \rangle \leq 0 \quad \text{for all } y \in \Omega. \tag{2.31}$$

For every $x \in \mathbb{R}^n$ there exists exactly one $z \in \Omega$ satisfying (2.31) if and only if Ω is a convex set. In other words, Ω $(0, \infty)$-subregular at every $x \in \Omega$ if and only if Ω is convex.

For a generalization to general Hilbert spaces, Deutsch closes his book with the question: "Must every Chebyshev set in (an infinite-dimensional) Hilbert space be convex? We believe that the answer is no."[2]

Remark 2.4.3. *The definition of (ε, δ)-subregularity was introduced in (Hesse and Luke, 2013) and is a generalization of the notion of (ε, δ)-regularity introduced in (Bauschke et al., 2013b, Definition 8.1). This regularity condition can be seen as a measure for violation of convexity of a set. By Cauchy-Schwarz, a trivial upper bound for ε is given by 1. Further, there is a monotinicity in ε with respect to δ: if a set Ω is (ε, δ)-subregular at a point $x \in \Omega$ and if Ω is (ε', δ')-subregular at x as well for $\delta' \leq \delta$, then $\varepsilon' \leq \varepsilon$. If a set Ω is (ε, δ)-subregular at a point $x \in \Omega$, and if Ω is $(\tilde{\varepsilon}, \tilde{\delta})$-subregular at x as well for $\tilde{\delta} \geq \delta$, then $\tilde{\varepsilon} \geq \varepsilon$.*

For the definition of super-regularity, we refer to (Lewis et al., 2009, Definition 4.3), while the definition stated in 2.4.1 (4) can be found in (Lewis et al., 2009, Proposition 4.4).

Similarly, the definition of Clarke regularity is cited from (Lewis et al., 2009, Definition 4.1).

The first appearance of prox-regularity was in (Poliquin et al., 2000, Definition 1.1). The way of stating it in Definition 2.4.1 (5) is actually due to the equivalence given by (Poliquin et al., 2000, Theorem 1.3 i)). Since prox-regularity is a local property, it is very natural to ask for counterexamples (see Figure 2.4).

Proposition 2.4.4 (relations of regularities)**.** *For a nonempty and closed subset $\Omega \subset \mathbb{R}^n$ the following chain of implications holds:*

1. *If Ω is prox-regular at $\bar{x}$, then Ω is super-regular at $\bar{x}$.*

2. *If Ω is super-regular at $\bar{x}$, then Ω is Clarke regular at $\bar{x}$.*

3. *If Ω is Clarke regular at $\bar{x}$, then Ω is (ε, δ)-subregular at $\bar{x}$.*

The converse does not hold.

Proof. The statement in (1) follows from (Lewis et al., 2009, Proposition 4.9). Claim number (2) is (Lewis et al., 2009, Corollary 4.5). To prove (3), we set $y = x$ in (2.30). Note that, for all $z \in P_\Omega x$, the vector $x - z$ is contained in $N_\Omega(x)$. Then Equation

---

[2](Deutsch, 2001, p. 306, Question)

(2.30) is exactly equivalent to (2.29). Counterexamples for the converse of (1) and (2) are shown in (Lewis et al., 2009, Example 4.6 and p. 494), while a counterexample for the converse of (3) is given in (Hesse and Luke, 2013, Remark 2.12). □
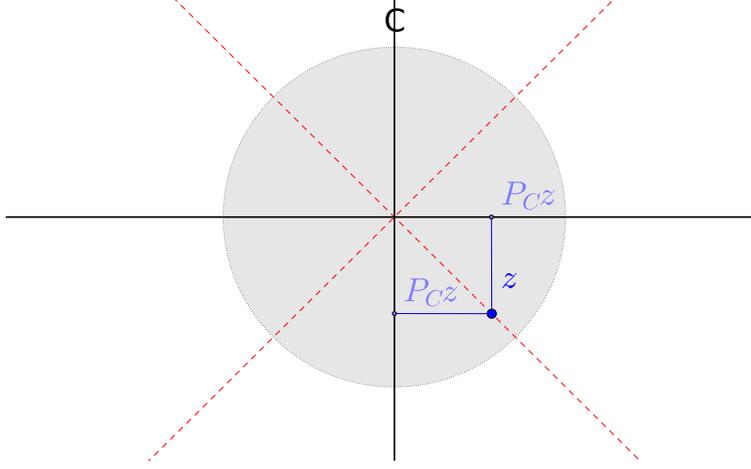
Figure 2.4: Let *C* be the union of the two black lines. It is not prox-regular at the intersection of the black lines. In particular, for any neighborhood of the point in the intersection, the projector onto *C* is not single-valued for all points on the red dashed lines. We illustrate this with the point *z* and its projection given by the end points of the blue lines.

Example 2.4.5. *At this point it is worth mentioning an important class of prox-regular sets. Let $\mathcal{M}$ be a smooth manifold. Then $\mathcal{M}$ is prox-regular everywhere (Lewis and Malick, 2008, Lemma 2.1).*

2.4.2 Regularity of Collections of Sets

Next, we define some notions of regularity of collections of sets that, together with (ε, δ)-subregularity, provide sufficient conditions for convergence of iterative methods used in the following chapters. Linear regularity, defined next, can be found in (Bauschke and Borwein, 1993, Definition 3.13). Local versions of this have appeared under various names in (Ioffe, 2000, Proposition 4), (Ngai and Théra, 2001, Section 3), and (Kruger, 2006, Equation (15)).

Definition 2.4.6 (linear regularity)**.** *A collection $(\Omega_1, \Omega_2, \ldots, \Omega_m) \subset \mathbb{R}^n$ of closed, non-empty sets $\Omega_j \subset \mathbb{R}^n$ is called* locally linearly regular *at $\bar{x} \in \cap_{j=1}^m \Omega_j$ on $\mathbb{B}_\delta(\bar{x})$ if there exist a $\kappa > 0$ and a $\delta > 0$ such that*

$$d_{\cap_{j=1}^m \Omega_j}(x) \leq \kappa \max_{i=1,\ldots,m} d_{\Omega_i}(x) \text{ for all } x \in \mathbb{B}_\delta(\bar{x}). \tag{2.32}$$

If (2.32) holds at $\bar{x}$ for every $\delta > 0$, the collection of sets is said to be linearly regular *there. The infimum over all κ such that (2.32) holds is called* modulus of regularity on $\mathbb{B}_\delta(\bar{x})$. *If the collection is linearly regular, one just speaks of the* modulus of regularity *(without mention of $\mathbb{B}_\delta(\bar{x})$).*

There is yet a stronger notion of regularity of collections of sets that we make use of called the *basic qualification condition for sets* in (Mordukhovich, 2006, Definition 3.2). For the purposes of this work, we refer to this as *strong regularity*.

Definition 2.4.7 (strong regularity)**.** *The collection (Ω_1, Ω_2) is strongly regular at $\bar{x}$ if*

$$N_{\Omega_1}(\bar{x}) \cap -N_{\Omega_2}(\bar{x}) = \{0\}. \tag{2.33}$$

It can be shown that strong regularity implies local linear regularity (Hesse and Luke, 2013). Any collection of finite dimensional affine subspaces with nonempty intersection is linearly regular (Bauschke and Borwein, 1996, Proposition 5.9 and Remark 5.10). Moreover, if Ω_1 and Ω_2 are affine subspaces,

$$\begin{aligned} &(\Omega_1, \Omega_2) \text{ is strongly regular at any } \bar{x} \in \Omega_1 \cap \Omega_2 \\ \iff \quad &\Omega_1^\perp \cap \Omega_2^\perp = \{0\} \quad \text{and} \quad \Omega_1 \cap \Omega_2 \neq \emptyset. \end{aligned} \tag{2.34}$$

In the case where Ω_1 and Ω_2 are affine subspaces we say that the collection is strongly regular without mention of any particular point in the intersection - as long as this is nonempty - since the collection is strongly regular at all points in the intersection.

3 Sparsity Optimization

In signal processing applications, it may be the case that signals satisfying certain linear constraints can be represented as vectors with just few nonzero entries in some generic basis. Of course, with the right change of basis, we can represent any signal as a sparse vector.

Consider for example the constant function $f : \mathbb{R} \to \mathbb{R}, x \mapsto 1$ and its Fourier transform

$$\mathcal{F}(f)(k) = \int_{-\infty}^{\infty} e^{2\pi i k x} \, d x = \delta(k).$$

Here, $\delta(k)$ denotes Dirac's delta function. Now, we translate this to applications with signals of finite length, say, $x = (1, \ldots, 1) \in \mathbb{R}^n$. Then $\mathcal{F}$ becomes the discrete Fourier transform F, and we get

$$F(x) = (n, 0, \ldots, 0).$$

As is shown in Theorem 3.1.4, finding these *sparse* vectors satisfying linear equations is in general NP-hard. The task of finding this sparse vectors is called *sparsity optimization*. With their preprint in 2004, the authors of (Candès and Tao, 2005) introduced a sufficient condition, called restricted isometry property, for the linear constraints such that a convex relaxation returns the correct solution. The work (Donoho, 2006) gave bounds for the number of linear constraints for recovery of the sparsest vector. It also gave this field of research its name "Compressed Sensing". Since then the field of Compressed Sensing has developed rapidly. In (Foucart and Rauhut, 2013), there is a first mathematical overview of this topic.

This chapter is based on (Hesse et al., 2014). Instead of applying a convex relaxation, we formulate the sparsity optimization problem as a *feasibility problem*. In other words, we seek points in the intersection of sets. The set of sparse vectors in a vector space is a nonconvex set. We study geometric properties of this set used in forthcoming chapters.

3.1 Definitions

To find a vector $x \in \mathbb{R}^n$ satisfying a set of linear constraints with the least possible number of nonzero entries can be formulated as a minimization problem. First, we introduce a notation for the number of nonzero entries of a vector.

Definition 3.1.1. *Denote the sign of a real number r λ by*

$$\text{sign}(\lambda) := \begin{cases} -1 & \text{if } \lambda < 0, \\ 0 & \text{if } \lambda = 0, \\ 1 & \text{if } \lambda > 0. \end{cases} \tag{3.1}$$

The ℓ_0-function of a real-valued vector $x \in \mathbb{R}^n$ is defined via

$$\ell_0(x) := \sum_{i=1}^{n} |\text{sign}(x_i)|. \tag{3.2}$$

The ℓ_0-function is thus the function that counts the number of nonzero entries in a vector. Instead of $\ell_0(x)$, some authors in the literature use the notation $\|x\|_0 := \ell_0(x)$. This reflects the fact that $\ell_0(x)$ can be written as the limit of ℓ_q-functions for $0 < q \leq 1$. It is not a norm since it violates the scalability of a norm. It is also not a convex function for all $n \geq 1$. To see this, let $x = 0, y = (1, 0, \ldots, 0)$, and $\lambda = \frac{1}{2}$. Then we have

$$1 = \ell_0(\lambda x + (1 - \lambda)y) = \ell_0((1/2, 0, \ldots, 0)) = 1 > \tfrac{1}{2} = \lambda \ell_0(x) + (1 - \lambda)\ell_0(y). \tag{3.3}$$
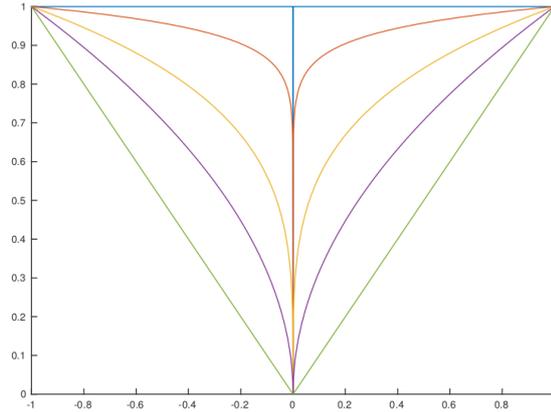
Figure 3.1: Graphs of *p*th powers of the functions ℓ_p for $p = 0$ (blue), $p = \frac{1}{16}$ (red), $p = \frac{1}{4}$ (yellow), $p = \frac{1}{2}$ (purple), and $p = 1$ (green).

Remark 3.1.2. *The convex hull of the ℓ_0-function is the constant zero function. To see this, we remind the reader of the definition of the convex hull in (2.14). Let $x \in \mathbb{R}^n$ and $\mu \geq 1$ be arbitrary. We define now $\lambda_0 := \frac{1}{\mu}, \lambda_1 := 1 - \lambda_0$, and $x_0 := \mu x, x_1 := 0$. We note that the equalities $\sum_{j=0}^{1} \lambda_j x_j = x$ and $\sum_{j=0}^{1} \lambda_j = 1$ hold. The value of $\ell_0(x)$ cannot exceed n. As a consequence, we have $\sum_{j=0}^{1} \lambda_j \ell_0(x_j) \leq \frac{n}{\mu}$. Since the number μ can be arbitrarily large, we see that the infimum over all λ_j of $\sum_{j=0}^{1} \lambda_j \ell_0(x_j)$ is zero.*

Definition 3.1.3. *Let $m \leq n$ and let $M \in \mathbb{R}^{m \times n}$ be a linear mapping of full rank. Further, let $p \in \mathbb{R}^m$ be an arbitrary vector. Then formulate the* compressed sensing problem *(Candès and Tao, 2005, Equation 1.3):*

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \quad \ell_0(x) \\ s.t. \quad Mx = p. \tag{3.4}$$

In applications, the vector p is usually a signal obtained by a measuring process. Due to this, p will sometimes be referred to as the "measurements".

The following theorem by (Natarajan, 1995) is the reason why relaxations or reformulations of (3.4) are necessary.

Theorem 3.1.4 ((Natarajan, 1995, Theorem 1)). *The compressed sensing problem (3.4) is NP-hard.*

The proof of Theorem 3.1.4 goes by finding examples of (3.4) that are equivalent to known NP-hard problems. In this case, a matrix and a vector of measurements is constructed such that finding the sparsest vector in the affine subspace is equivalent to finding a solution to the "exact 3-covering". The latter is NP-hard (Garey and Johnson, 1979).

As mentioned in the introductory part, we formulate a feasibility problem related to Problem (3.4). In (Blumensath and Davies, 2009), (Blumensath and Davies, 2010), and in (Beck and Teboulle, 2011), the authors make an a priori assumption s on the value of $\ell_0(\bar{x})$ of a solution $\bar{x}$ to (3.4). Then the authors use iterative hard thresholding to find a point $\bar{x}$ of sparsity s. We follow the same ansatz and make an a priori assumption s to the sparsity of a solution $\bar{x}$ to (3.4).

We formulate a *feasibility problem* closely related to (3.4). First, the sets involved in that feasibility problem need to be defined.

Definition 3.1.5. *Define for an integer s with $0 \leq s \leq n$ the set of vectors in $\mathbb{R}^n$ of sparsity at most s by*

$$A_s := \left\{ x \in \mathbb{R}^n \mid \ell_0(x) \leq s \right\}. \tag{3.5}$$

This set has been used throughout the literature, for example in (Bauschke et al., 2014b) and in (Hesse et al., 2014). We will name a collection of properties of A_s in Section 3.2. An intuitive example for $A_2 \subset \mathbb{R}^3$ is shown in Figure 3.2. Further, we define the set

$$B := \left\{ x \in \mathbb{R}^n \mid Mx = p \right\}. \tag{3.6}$$

The set B represents the set of vectors satisfying the linear constraints in Problem (3.4). Let us show some of its properties.

Lemma 3.1.6. *The set B defined by (3.6) is an affine subspace. Hence, it is a closed and convex set.*

Proof. To show that B is an affine subspace, it is sufficient to show that, for all $x, y \in B$ and for all $\lambda \in \mathbb{R}$, we have $\lambda x + (1 - \lambda)y \in B$. Because all points $x, y \in B$ satisfy
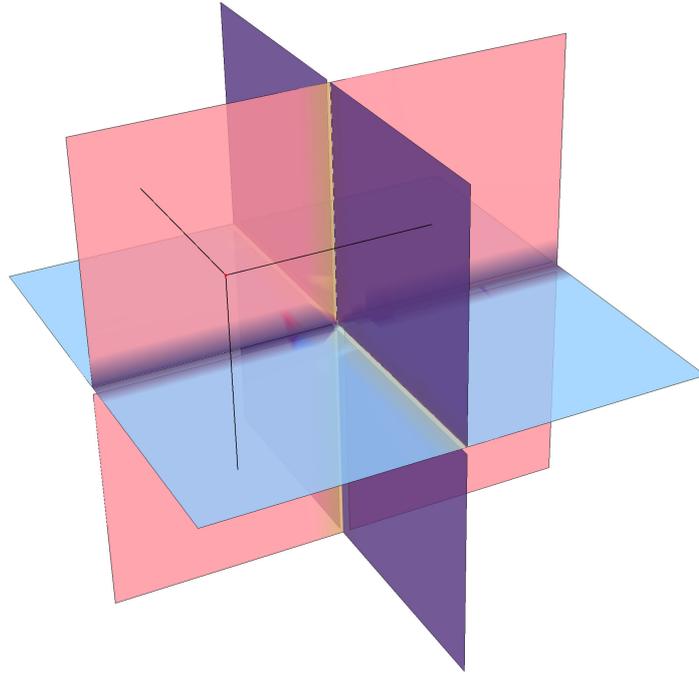
Figure 3.2: The set of 2-sparse vectors in $\mathbb{R}^3$ together with a point where its nearest point in A_2 is not unique

$Mx = My = p$, and since M is a linear mapping, this gives us

$$M(\lambda x + (1 - \lambda)y) = \lambda Mx + (1 - \lambda)My = \lambda p + (1 - \lambda)p = p, \qquad (3.7)$$

which is equivalent to $\lambda x + (1 - \lambda)y \in B$. $\qquad\qquad\square$

Now we can formulate our feasibility problem.

Definition 3.1.7. *Given a sparsity parameter $s \in \mathbb{N}$, define the* sparse-affine feasibility problem *by*

$$\textit{find } \bar{x} \in A_s \cap B. \qquad (3.8)$$

In Chapter 5, we analyze the performance of the alternating projections algorithm and the Douglas-Rachford applied to Problem (3.8). Their behavior heavily relies on the geometries of A_s and B. We describe these in the following section.

3.2 Regularity of Sparsity Sets

The following two sections are based on Sections II and III in (Hesse et al., 2014). Following (Bauschke et al., 2014b), we decompose the set A_s into a union of subspaces. For $x \in \mathbb{R}^n$, define the *sparsity subspace* associated with x by

$$\operatorname{supp}(x) := \left\{ y \in \mathbb{R}^n \mid y_j = 0 \text{ if } x_j = 0 \right\} \tag{3.9}$$

and the mapping

$$I : \mathbb{R}^n \to \{1, \ldots, n\}, \quad x \mapsto \left\{ j \in \{1, \ldots, n\} \mid x_j \neq 0 \right\}. \tag{3.10}$$

Define

$$\mathcal{J} := 2^{\{1,2,\ldots,n\}} \quad \text{and} \quad \mathcal{J}_s := \{ J \in \mathcal{J} \mid J \text{ has } s \text{ elements} \}. \tag{3.11}$$

The set A_s can be written as the union of all subspaces indexed by $J \in \mathcal{J}_s$ (Bauschke et al., 2014b, Equation (27d)),

$$A_s = \bigcup_{J \in \mathcal{J}_s} A_J \tag{3.12}$$

where $A_J := \operatorname{span} \{ e_j \mid j \in J \}$, and e_j is the jth standard unit vector in $\mathbb{R}^n$. For $x \in \mathbb{R}^n$, we define the index set of the s largest coordinates in absolute value by

$$C_s(x) := \left\{ J \in \mathcal{J}_s \,\middle|\, \min_{j \in J} |x_j| \geq \max_{j \notin J} |x_j| \right\}. \tag{3.13}$$

The next elementary result gives a compact presentation of the shortest distance of a point x to its projection onto the set A_s.

Lemma 3.2.1 ((Bauschke et al., 2014b, Lemma 3.4)). *Let $x \in A_s$ and assume $s \leq n - 1$. Then*

$$\min \left\{ d_{A_J}(x) \mid x \notin A_J, J \in \mathcal{J}_s \right\} = \min \left\{ |x_j| \mid j \in I(x) \right\}. \tag{3.14}$$

Using the above notation, the normal cone to the sparsity set A_s at $x \in A_s$ has the following closed-form representation. It was shown in (Bauschke et al., 2014b, Theorem 3.9). We also refer to (Luke, 2013, Proposition 3.6) and Proposition 8.2.7 for the general matrix representation.

$$\begin{aligned} N_{A_s}(x) &= \{ v \in \mathbb{R}^n \mid \ell_0(v) \leq n - s \} \cap (\operatorname{supp}(x))^\perp \\ &= \bigcup_{J \in \mathcal{J}_s,\, I(x) \subseteq J} A_J^\perp. \end{aligned} \tag{3.15}$$

The normal cone to the affine set B also has a simple closed form, namely, $N_B(x) = B^\perp$ (see, for example, (Mordukhovich, 2006, Proposition 1.5)). Let $y \in \mathbb{R}^n$ be a point such that $My = p$. Recall the definition of the Minkowski sum in (2.15). Note that $\ker M$ is the subspace parallel to B, i.e., $\ker M = B + \{-y\}$. This notation gives us explicit representations for the projectors onto A_s (see (Bauschke et al., 2014b, Proposition 3.6))

and B:

$$P_B x = x - M^\dagger(Mx - p) \quad \text{and} \quad P_{A_s} x = \bigcup_{J \in C_s(x)} P_{A_J} x, \tag{3.16}$$

where $M^\dagger$ is the Moore-Penrose inverse (see (5.2) for more details), and

$$(P_{A_J} x)_j = \begin{cases} x_j, & j \in J, \\ 0, & j \notin J. \end{cases} \tag{3.17}$$

Lemma 3.2.2. *The set of sparse vectors A_s is locally convex at points x with $\ell_0(x) = s$.*

Proof. Choose $x \in A_s$ with $\ell_0(x) = s$. Define $\delta < \min\{|x_j| \mid x_j \neq 0\}$ and choose y, z in the intersection of the ball $B_\delta(x)$ with A_s. Since the ball is a convex set, it only remains to prove that any convex combination of y and z is contained in A_s. But this is true since the index set

$$\mathbb{I}(x) = \{j \in \mathbb{N} \mid x_j \neq 0\} \tag{3.18}$$

is the same for y and z. Otherwise, a sparser vector would be contained in $B_\delta(x)$ since it is convex. This completes the proof. □

Next, we collect some facts about the projectors and reflectors of A_s and B. Because the set B is convex and closed, the projector onto B is single-valued. In the sense of Definition 2.1.1, we write $\bar{x} = P_B x$ instead of $\{\bar{x}\} = P_B x$.

Lemma 3.2.3 ((Hesse et al., 2014, Lemma III.2)). *Let A_s and B be defined by (3.5) and (3.6). Let $z \in A_s$ and $w \in B$. For any $\delta_z \in (0, \min\{|z_j| \mid j \in I(z)\})$ and for any $\delta_b \in (0, \infty)$, the following hold:*

(i) $P_B x \in \mathbb{B}_{\delta_w}(w)$ *for all* $x \in \mathbb{B}_{\delta_w}(w)$;

(ii) $P_{A_s} x \subset \mathbb{B}_{\delta_z/2}(z)$ *for all* $x \in \mathbb{B}_{\delta_z/2}(z)$;

(iii) $R_B x \in \mathbb{B}_{\delta_w}(w)$ *for all* $x \in \mathbb{B}_{\delta_w}(w)$;

(iv) $R_{A_s} x \subset \mathbb{B}_{\delta_z/2}(z)$ *for all* $x \in \mathbb{B}_{\delta_z/2}(z)$.

Proof. (i) This follows from the fact that the projector is nonexpansive[1] since B is convex and

$$\|P_B x - w\| = \|P_B x - P_B w\| \le \|x - w\|.$$

(ii) Let $x \in \mathbb{B}_{\delta_z/2}(z)$. We have

$$|x_j - z_j| = |x_j| \le \delta_z/2 \quad \text{for any} \quad j \in I^\circ(z) := \{j : z_j = 0\}.$$

Moreover, we have

$$|x_j - z_j| \le \delta_z/2 \quad \text{for all} \quad j \in I(z) := \{j : z_j \neq 0\},$$

---

[1]In fact, the projector is firmly nonexpansive as shown, for example, in (Zarantonello, 1971, Lemma 1.2).

and so $|x_j| > \delta_z/2$ for all $j \in I(z)$. Altogether, this means that $|x_j| > |x_i|$ for all $i \in I^\circ(z), j \in I(z)$.

Therefore, the indices of the nonzero elements of z exactly correspond to the indices of the $|I(z)|$-largest elements of x, where $|I(z)|$ denotes the cardinality of the set $I(z)$. Since $|I(z)| \leq s$, the projector of x does not need to be single-valued[2]. Nevertheless, for all $x^+ \in P_{A_s}(x)$, we have $z \in \text{supp}(x^+)$, where $\text{supp}(x^+)$ is defined by (3.9). Since $\text{supp}(x^+)$ is a subspace, x^+ is the orthogonal projection of x onto a subspace. Hence, by Pythagoras' theorem, we obtain

$$
\begin{aligned}
\|x - x^+\|_2^2 + \|x^+ - z\|_2^2 &= \|x - z\|_2^2 \\
\text{and} \quad \|x^+ - z\|_2 \leq \|x - z\|_2 &\leq \tfrac{\delta}{2}.
\end{aligned}
\tag{3.19}
$$

Thus, $P_{A_s} x \subset \mathbb{B}_{\delta_z/2}(z)$.

(iii) The reflector R_B is with respect to the affine subspace B containing w. The distance $d_B(R_B x)$ is equal to $d_B(x)$. The result follows.

(iv) As in the proof of (ii), for all $x \in \mathbb{B}_{\delta_z/2}$, we have $z \in \text{supp}(x^+)$ for each $x^+ \in P_{A_s}(x)$. In other words, the projector, and hence the corresponding reflector, is with respect to a subspace containing z. Thus, as in (iii), $\|R_{A_s} x - z\| = \|x - z\|$, though in this case only for $x \in \mathbb{B}_{\delta_z/2}$.

□

The next lemma shows that, around any point $\bar{x} \in A_s$, the set A_s is the union of subspaces in A_s containing $\bar{x}$. Hence, around any point $\bar{x} \in A_s \cap B$, the intersection $A_s \cap B$ can be described locally as the intersection of subspaces and the affine set B, each containing $\bar{x}$.

Lemma 3.2.4 ((Hesse et al., 2014, Lemma III.3)). *Let* $\bar{x} \in A_s \cap B$ *with* $0 < \ell_0(\bar{x}) \leq s$. *Then for all* $\delta < \min\{|\bar{x}_j| : \bar{x}_j \neq 0\}$, *we have*

$$
A_s \cap \mathbb{B}_\delta(\bar{x}) = \bigcup_{J \in \mathcal{J}_s,\ I(\bar{x}) \subseteq J} A_J \cap \mathbb{B}_\delta(\bar{x}),
\tag{3.20}
$$

and hence

$$
A_s \cap B \cap \mathbb{B}_\delta(\bar{x}) = \bigcup_{J \in \mathcal{J}_s,\ I(\bar{x}) \subseteq J} A_J \cap B \cap \mathbb{B}_\delta(\bar{x}).
\tag{3.21}
$$

If in fact $\ell_0(\bar{x}) = s$, *then there is a unique* $J \in \mathcal{J}_s$ *such that, for all* $\delta < \min\{|\bar{x}_j| : \bar{x}_j \neq 0\}$, *we have* $A_s \cap \mathbb{B}_\delta(\bar{x}) = A_J \cap \mathbb{B}_\delta(\bar{x})$ *and hence* $A_s \cap B \cap \mathbb{B}_\delta(\bar{x}) = A_J \cap B \cap \mathbb{B}_\delta(\bar{x})$.

Proof. If $s = n$, then the set A_s is equal to $\mathbb{R}^n$, and both statements are trivial. For the case $s \leq n - 1$, choose any $x \in \mathbb{B}_\delta(\bar{x}) \cap A_s$. From the definition of δ and Lemma 3.2.1 we have that, for any $J \in \mathcal{J}_s$, if $\bar{x} \notin A_J$, then $x \notin A_J$. By contraposition, therefore, $x \in A_J$ implies that $\bar{x} \in A_J$, hence, for each $x \in \mathbb{B}_\delta(\bar{x}) \cap A_s$, we have $x \in \mathbb{B}_\delta(\bar{x}) \cap A_{I(x)}$

---

[2]Consider the case $z = (1, 0, \ldots, 0)$, $x = (1, \delta/4, \delta/4, 0, \ldots, 0)$, and $s = 2$.

where $I(\bar{x}) \subseteq I(x) \in \mathcal{J}_s$. The intersection $\mathbb{B}_\delta(\bar{x}) \cap A_s$ is then the union over all such intersections as given by (3.20). Equation (3.21) is an immediate consequence of (3.20).

If in addition $\ell_0(\bar{x}) = s$, then the cardinality of $I(\bar{x})$ is s and by (Bauschke et al., 2014b, Lemma 3.5), we have $C_s(\bar{x}) = \{I(\bar{x})\}$, where $C_s(\bar{x})$ is given by (3.13). This means that if $\bar{x}$ has sparsity s, then there is exactly one subspace A_J with index set $J := I(\bar{x})$ in $\mathcal{J}_s$ containing $\bar{x}$. By Lemma 3.2.1,

$$d_{A_s \setminus A_J}(\bar{x}) = \min\left\{|\bar{x}_j| \,\middle|\, j \in J\right\} > \delta.$$

From this we conclude the equality $A_s \cap \mathbb{B}_\delta(\bar{x}) = A_J \cap \mathbb{B}_\delta(\bar{x})$ and hence

$$A_s \cap B \cap \mathbb{B}_\delta(\bar{x}) = A_J \cap B \cap \mathbb{B}_\delta(\bar{x}),$$

as claimed. □

Theorem 3.2.5 (regularity of A_s (Hesse et al., 2014, Theorem III.4))**.** *At any point $\bar{x} \in A_s \setminus \{0\}$ the set A_s is $(0, \delta)$-subregular at $\bar{x}$ for $\delta < \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$). On the other hand, the set A_s is not $(0, \delta)$-subregular at $\bar{x} \in A_s \setminus \{0\}$ for any $\delta \geq \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$. In contrast, the set A_s is $(0, \infty)$-subregular at 0.*

Proof. Choose any $x \in \mathbb{B}_\delta(\bar{x}) \cap A_s$ and any $v \in N_{A_s}(x)$. By the characterization of the normal cone in (3.15), there is some $J \in \mathcal{J}_s$ with $I(x) \subseteq J$ and $v \in A_J^\perp \subset N_{A_s}(x)$. As in the proof of Lemma 3.2.4, for any $\delta \in (0, \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\})$, we have $I(\bar{x}) \subseteq I(x)$. Hence, $\bar{x} - x \in A_J$ and $\langle v, \bar{x} - x \rangle = 0$. By the definition of (ε, δ)-regularity in Definition 2.4.1 1), A_s is $(0, \delta)$-subregular as claimed.

That A_s is not $(0, \delta)$-subregular at $\bar{x} \in A_s \setminus \{0\}$ for any $\delta \geq \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$) follows from the failure of Lemma 3.2.4 on balls larger than $\min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$. Indeed, suppose $\delta \geq \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$, then, by Lemma 3.2.1, there is a point $x \in \mathbb{B}_\delta(\bar{x}) \cap A_s$ with $x \in A_J \subset A_s$ but $\bar{x} \notin A_J$. Now we choose $v \in A_J^\perp \subset N_{A_s}(x)$. Because $\bar{x} \notin A_J$, then $\bar{x} - x \notin A_J$ and thus $|\langle v, \bar{x} - x \rangle| > 0$. Since $N_{A_s}(x)$ is a union of subspaces, the sign of v can be chosen so that $\langle v, \bar{x} - x \rangle > 0$, in violation of $(0, \delta)$-subregularity.

For the case $\bar{x} = 0$, Equation (3.15) yields $\langle v, x \rangle = 0$ for any $x \in A_s$ and $v \in N_{A_s}(x)$ since $\operatorname{supp}(x)^\perp \perp \operatorname{supp}(x)$. This completes the proof. □

Lemma 3.2.6. *Let $r, s, n \in \mathbb{N}$ with $r \leq s \leq n$. Then, for all $x \in \mathbb{R}^n$, we have*

$$P_{A_r}(P_{A_s}x) = P_{A_r}x. \tag{3.22}$$

Proof. By (3.16), we know that $P_{A_r}x$ is the set of vectors with the r largest absolute entries of x. Denote by E_r the set of r largest absolute entries of x and by E_s the set of s largest absolute entries of x. Then $E_r \subset E_s$ and, hence, $P_{A_r}(P_{A_s}x) = P_{A_r}x$. □

Finally, the above statements yield two useful properties of the ℓ_0-function.

Proposition 3.2.7. *The ℓ_0-function is subadditive and lower semicontinuous. Further, the equality $\ell_0(x) = \ell_0(-x)$ holds for all $x \in \mathbb{R}^n$.*

Proof. By Definition 2.1.7 based on (Rockafellar and Wets, 1998, Theorem 1.6), the lower semicontinuity of a function is equivalent to the lower level sets being closed subsets of $\mathbb{R}^n$. For all $s \in \mathbb{N}$, the lower level sets of ℓ_0 are given by

$$\text{lev}_{\leq s}\, \ell_0 = \{x \in \mathbb{R}^n \mid \ell_0 \leq s\}, \tag{3.23}$$

in other words, $\text{lev}_{\leq s}\, \ell_0 = A_s$. The latter is, as a union of finitely many subspaces, a closed set for all $x \in \mathbb{R}^n$.

The number of nonzero entries in a vector is independent of the sign of the entries, so $\ell_0(x) = \ell_0(-x)$.

To show subadditivity, we take $x, y \in \mathbb{R}^n$ with $\ell_0(x) = s_x$ and $\ell_0(y) = s_y$. If $I(x) \cap I(y) = \emptyset$, where $I(x)$ and $I(y)$ are given by (3.10), then $\ell_0(x + y) \leq s_x + s_y$ with equality if and only if $s_x + s_y \leq n$. If $I(x) \cap I(y) \neq \emptyset$, then in any case $\ell_0(x + y) < s_x + s_y$. Altogether, ℓ_0 is a subadditive function. □

3.3 Regularity of the Intersection

In (Hesse and Luke, 2013), it was shown that local linear regularity of intersections of sets is one of the ingredients for local linear convergence of the alternating projections algorithm. As will be shown in this section, the collection (A_s, B) has locally linearly regular intersection as long as the intersection is nonempty. We begin with a technical lemma.

Lemma 3.3.1 (linear regularity under unions (Hesse et al., 2014, Lemma III.5))**.** *Let* $(\Omega_1, \Omega_2, \ldots, \Omega_m, \Omega_{m+1})$ *be a collection of nonempty subsets of $\mathbb{R}^n$ with nonempty intersection. Let $\bar{x} \in \left(\cap_{j=1}^m \Omega_j\right) \cap \Omega_{m+1}$. Suppose that, for some $\delta > 0$, the pair (Ω_j, Ω_{m+1}) is locally linearly regular with modulus κ_j on $\mathbb{B}_\delta(\bar{x})$ for each $j \in \{1, 2, \ldots, m\}$. Then the collection $\left(\bigcup_{j=1}^m \Omega_j, \Omega_{m+1}\right)$ is locally linearly regular at $\bar{x}$ on $\mathbb{B}_\delta(\bar{x})$ with modulus $\bar{\kappa} = \max_j\{\kappa_j\}$.*

Proof. Denote $\Gamma := \bigcup_{j=1}^m \Omega_j$. First, note that, for all $x \in \mathbb{B}_\delta(\bar{x})$, we have

$$d_{\Gamma \cap \Omega_{m+1}}(x) = \min_j \left\{ d_{\Omega_j \cap \Omega_{m+1}}(x) \right\} \leq \min_j \left\{ \kappa_j \max\{d_{\Omega_j}(x), d_{\Omega_{m+1}}(x)\} \right\}, \tag{3.24}$$

where the inequality on the right follows from the assumption that (Ω_j, Ω_{m+1}) is locally linearly regular with modulus κ_j on $\mathbb{B}_\delta(x)$. Let $\bar{\kappa} \geq \max_j\{\kappa_j\}$. Then

$$\begin{aligned} d_{\Gamma \cap \Omega_{m+1}}(x) &\leq \bar{\kappa} \min_j \left\{ \max\{d_{\Omega_j}(x), d_{\Omega_{m+1}}(x)\} \right\} \\ &= \bar{\kappa} \max \left\{ \min_j\{d_{\Omega_j}(x)\}, d_{\Omega_{m+1}}(x) \right\}. \end{aligned} \tag{3.25}$$

This completes the proof. □

Theorem 3.3.2 (regularity of (A_s, B) (Hesse et al., 2014, Theorem III.6))**.** *Let A_s and B be defined by (3.5) and (3.6) with $A_s \cap B \neq \emptyset$. At any $\bar{x} \in A_s \cap B$ and for any $\delta \in$*

$(0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$, *the collection* (A_s, B) *is locally linearly regular on* $\mathbb{B}_{\delta/2}(\bar{x})$ *with modulus of regularity*

$$\bar{\kappa} = \max_{J \in \mathcal{J}_s, I(\bar{x}) \subseteq J} \{\kappa_J\}$$

where κ_J *is the modulus of regularity of the collection* (A_J, B).

Proof. For any $\bar{x} \in A_s \cap B$, we have $\bar{x} \in A_J \cap B$ for all $J \in \mathcal{J}_s$ with $I(\bar{x}) \subseteq J$. Thus, (A_J, B) is linearly regular with modulus of regularity κ_J (Bauschke and Borwein, 1996, Proposition 5.9 and Remark 5.10). Define

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

Then, by Lemma 3.3.1, the collection $(\overline{A_s}, B)$ is linearly regular at $\bar{x}$ with modulus of regularity $\bar{\kappa} := \max_{J \in \mathcal{J}_s,\ I(\bar{x}) \subseteq J}\{\kappa_J\}$. By Lemma 3.2.4, we obtain $A_s \cap \mathbb{B}_{\delta/2}(\bar{x}) = \overline{A_s} \cap \mathbb{B}_{\delta/2}(\bar{x})$ for any $\delta \in (0, \min\{|\bar{x}_j| \mid j \in I(\bar{x})\})$. Moreover, by Lemma 3.2.3(ii), for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$, we have $P_{A_s} x \subset \mathbb{B}_{\delta/2}(\bar{x})$, and thus $P_{A_s} x = P_{\overline{A_s}} x$. In other words, $d_{A_s}(x) = d_{\overline{A_s}}(x)$ for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$. Hence, the collection (A_s, B) is locally linearly regular on $\mathbb{B}_\delta(\bar{x})$ with modulus $\bar{\kappa}$. This completes the proof. □
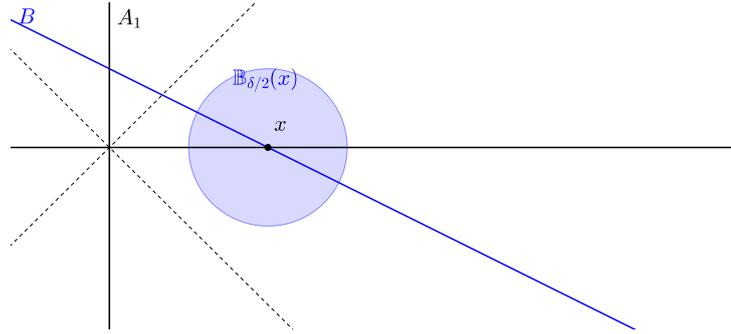
Figure 3.3: The intersection of the blue line B with the set A_1 is locally linearly regular. For example, the blue disc may be chosen as the corresponding neighborhood of a point in the intersection $A_1 \cap B$. The dashed lines is the subset of points of $\mathbb{R}^2$ that have *not* a unique projection in A_1.

As an example for local linear regularity, consider Figure 3.3.

Remark 3.3.3 ((Hesse et al., 2014, Remark III.7)). *Another example shows that the collection* (A_s, B) *does not need to be* linearly regular. *Consider the sparsity set* A_1, *the affine set* $B = \{(1, \tau, 0) \mid \tau \in \mathbb{R}\}$, *and the sequence of points* $\{x^k\}_{k \in \mathbb{N}}$ *defined by* $x^k = (0, k, 0)$. *Then* $A_1 \cap B = \{(1, 0, 0)\}$ *and* $\max\{d_{A_1}(x^k), d_B(x^k)\} = 1$ *for all* k *while* $d_{A_1 \cap B}(x^k) \to \infty$ *as* $k \to \infty$.

3.4 Second-Order Subdifferentials

This section is inspired by the works (Le, 2013) and (Mordukhovich and Rockafellar, 2012). In the former, the subdifferentials of the ℓ_0-function and of the rank-function are computed, while the latter gives an analysis of second-order subdifferentials. Our aim here is to extend the work of (Le, 2013) by one degree of derivation.

3.4.1 The Second-Order Subdifferential of the Counting Function

Lemma 3.4.1 ((Le, 2013, Theorem 1, Theorem 2)). *The ℓ_0-function is subdifferentially regular (Definition 2.1.11).*

Definition 3.4.2 (coderivatives (Rockafellar and Wets, 1998, Definition 8.33)). *Given a set-valued $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, define the* coderivative *at $(\overline{x}, \overline{y}) \in \mathrm{gph}(F)$ by*

$$D^* F(\overline{x}, \overline{y})(u) = \left\{ v \in \mathbb{R}^n \;\middle|\; (v, -u) \in N_{\mathrm{gph}(F)}(\overline{x}, \overline{y}) \right\}. \tag{3.26}$$

Our aim is to give a notation of a second-order derivative for set-valued mappings. With coderivatives of subdifferentials, this notation can be given.

Definition 3.4.3 ((Mordukhovich and Rockafellar, 2012, Definition 2.1)). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function which is finite at $\overline{x}$. Define the* second-order subdifferential

$$\partial^2 f(\overline{x}, \overline{y})(u) := (D^* \partial f)(\overline{x}, \overline{y})(u), u \in \mathbb{R}^n. \tag{3.27}$$

Now let the function $f : \mathbb{R}^n \to \mathbb{N}$ be the ℓ_0-function, i.e., $f(x) = \ell_0(x)$. Then, by Lemma 3.4.1, the different sorts of subdifferentials, namely, the regular and the limiting subdifferential of the ℓ_0-function, coincide. This leads to the subdifferential

$$\partial \ell_0(x) = \mathrm{supp}^{\perp}(x) := \left\{ y \in \mathbb{R}^n \;\middle|\; y_j = 0 \text{ if } x_j \neq 0 \right\}. \tag{3.28}$$

As a consequence, we get

$$\partial^2 \ell_0(\overline{x}, \overline{y})(u) = \left\{ v \in \mathbb{R}^n \;\middle|\; (v, -u) \in N_{\mathrm{gph}(\partial \ell_0)}((\overline{x}, \overline{y})) \right\}. \tag{3.29}$$

In the following, we compute the normal cones of the graph of the subdifferential of ℓ_0. We observe that every element (x, y) of $\mathrm{gph}(\partial \ell_0)$ has at most n nonzero components. This is due to the definition of $\mathrm{supp}^{\perp}(x)$ and the fact that x as well as y are elements of $\mathbb{R}^n$. Further, if x has k nonzero components, then y has at most $n - k$ nonzero components. The graph of the ℓ_0-function is thus the set of vectors in $\mathbb{R}^{2n}$ with sparsity at most n.

We write the space of sparse vectors $A_n \subset \mathbb{R}^n \times \mathbb{R}^n$ as a (of course not disjoint) union of subspaces. First, we define an index set

$$\mathbb{I}_k \in \left\{ \mathbb{I} \in 2^{\{1,\dots,n\}} \;\middle|\; |\mathbb{I}| = k \right\} \tag{3.30}$$

indexed by $k \in \{1, \ldots, n\}$. Now define the subspaces

$$A_{\mathbb{I}_k} = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid x_j = 0 \; \forall \, j \notin \mathbb{I}_k, y_l = 0 \; \forall \, l \in \mathbb{I}_k \right\}. \tag{3.31}$$

Then

$$A_n = \bigcup_{k=1}^{n} \bigcup_{\mathbb{I}_k} A_{\mathbb{I}_k}. \tag{3.32}$$

Lemma 3.4.4. *The normal cone of the graph of $\partial \ell_0$ is the set*

$$N_{\mathrm{gph}(\partial \ell_0)}(\overline{x}, \overline{y}) = \begin{cases} A_{\mathbb{I}_k}^{\perp} & \text{if } \ell_0((\overline{x}, \overline{y})) = n \text{ and } (\overline{x}, \overline{y}) \in A_{\mathbb{I}_k}, \\ \bigcup A_{\mathbb{I}_k}^{\perp} & \text{for all } A_{\mathbb{I}_k} \text{ with } (\overline{x}, \overline{y}) \in A_{\mathbb{I}_k}. \end{cases} \tag{3.33}$$

Proof. Consider the case that $\ell_0((\overline{x}, \overline{y})) = n$, and take an arbitrary sequence $\{(x, y)^k\}_{k \in \mathbb{N}}$ converging to $(\overline{x}, \overline{y})$. Then there exists a neighborhood of $(\overline{x}, \overline{y})$ and $k' \in \mathbb{N}$ such that, for every $k \geq k'$, the projection of $(x, y)^k$ onto $\mathrm{gph}(\partial \ell_0)$ lies in $\mathrm{supp}((\overline{x}, \overline{y}))$. But then any cone

$$\mathrm{cone}\left((x, y) - P_{\mathrm{gph}(\partial \ell_0)}(x, y) \right) \tag{3.34}$$

lies in $\mathrm{supp}((\overline{x}, \overline{y}))$. From this, we conclude that the limes superior of these cones is

$$\mathrm{supp}((\overline{x}, \overline{y}))^{\perp} = A_{(j,k)}. \tag{3.35}$$

The same argument applies for sparser points $(\overline{x}, \overline{y})$. The difference is that, for any subspace $A_{(j,k)}$, there is a sequence converging to $(\overline{x}, \overline{y})$ such that the projection of the elements of the sequence lies in $A_{(j,k)}$. We obtain the union of all $A_{(j,k)}^{\perp}$ with $(\overline{x}, \overline{y}) \in A_{(j,k)}$ as the Mordukhovich normal cone. $\square$

Theorem 3.4.5. *The second-order subdifferential of the ℓ_0-function is the set*

$$\partial^2 \ell_0((\overline{x}, \overline{y}))(u) = (D^* \partial \ell_0)((\overline{x}, \overline{y}))(u)$$

$$= \begin{cases} \left\{ v \mid (u, -v) \in \mathrm{supp}((\overline{x}, \overline{y}))^{\perp} \right\} & \text{if } \ell_0((\overline{x}, \overline{y})) = n, \\ \left\{ v \mid (u, -v) \in \bigcup A_{(j,k)}^{\perp} \text{ for all } A_{(j,k)} \text{ with } (\overline{x}, \overline{y}) \in A_{(j,k)} \right\} & \text{if } \ell_0((\overline{x}, \overline{y})) < n, \\ \varnothing & \text{else.} \end{cases}$$

Proof. The claim follows from Lemma 3.4.4. $\square$

3.4.2 The Inverse Second-Order Subdifferential

The goal of this section is to give an explicit formula for the inverse mapping as in Definition 2.1.1 of the second-order subdifferential operator $\partial^2 \ell_0((\overline{x}, \overline{y}))$ for some $(\overline{x}, \overline{y}) \in \mathrm{gph}(\partial \ell_0)$.

Lemma 3.4.6. *We have $\mathrm{dom}\, \partial^2 \ell_0((\overline{x}, \overline{y})) = \partial \ell_0(\overline{x})$.*

Proof. By Theorem 3.4.5, the second-order subdifferential of the ℓ_0-function is given by

$$\partial^2\ell_0((\overline{x},\overline{y}))(u) = (D^*\partial\ell_0)((\overline{x},\overline{y}))(u)$$
$$= \begin{cases} \left\{v \mid (u,-v) \in \text{supp}((\overline{x},\overline{y}))^\perp\right\} & \text{if } \ell_0((\overline{x},\overline{y})) = n, \\ \left\{v \mid (u,-v) \in \bigcup A_{(j,k)}^\perp \text{ for all } A_{(j,k)} \text{ with } (\overline{x},\overline{y}) \in A_{(j,k)}\right\} & \text{if } \ell_0((\overline{x},\overline{y})) < n, \\ \varnothing & \text{else.} \end{cases}$$

So, a necessary condition for $\partial^2\ell_0((\overline{x},\overline{y}))(u)$ to be nonempty is $u \in \text{supp}(\overline{x})^\perp$. By (3.28), this is equivalent to $u \in \partial\ell_0(\overline{x})$. $\qquad\square$

Corollary 3.4.7. *For the inverse of $\partial^2\ell_0((\overline{x},\overline{y}))$, we have*

$$\text{dom}\,\partial^2\ell_0((\overline{x},\overline{y}))^{-1} = \text{supp}(\overline{x}). \tag{3.36}$$

Proof. By (Rockafellar and Wets, 1998, Chapter 5.A), we have $\text{dom}\left(F^{-1}\right) = \text{range}(F)$ for a set-valued mapping F. But, for some $u \in \text{dom}\,\partial^2\ell_0((\overline{x},\overline{y}))$, we have $u \in \text{supp}(\overline{x})^\perp$ and $\overline{y} \in \text{supp}(\overline{x})^\perp$. Hence,

$$\partial^2\ell_0((\overline{x},\overline{y}))(u) \subset \text{supp}(u)^\perp, \tag{3.37}$$

and $\text{supp}(u)^\perp$ is a superset of $\text{supp}(\overline{x})$. Because $u \in \text{dom}\,\partial^2\ell_0((\overline{x},\overline{y}))$ is chosen arbitrarily, we get $\text{dom}\,\partial^2\ell_0((\overline{x},\overline{y}))^{-1} = \text{supp}(\overline{x})$. $\qquad\square$

3.4.3 Set-Valued Newton's Method

The definition of a second-order subdifferential could give rise to the minimization of the ℓ_0-function via a Newton-type method. We briefly note why this is suboptimal.

A naive approach would be an iteration $\left\{x^k\right\}_{k\in\mathbb{N}}$ with

$$x^{k+1} \in x^k - \partial^2\ell_0((x^k,y^k))^{-1}(\partial\ell_0(x^k))$$

with $y^k \in \partial\ell_0(x^k)$. But as we have seen in Corollary 3.4.7, the intersection of $\partial\ell_0(x^k)$ with $\text{dom}\,\partial^2\ell_0((x^k,y^k))^{-1}$ is $\{0\}$. Then $-\partial^2\ell_0((x^k,y^k))^{-1}(\partial\ell_0(x^k))$ is just the zero vector. As a consequence, the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ would be constant. Hence, the approach via this Newton-type method would not lead to any useful result.

Remark 3.4.8. *In the spirit of (Hiriart-Urruty, 2013, Theorem), we remark that every point of $\mathbb{R}^n$ is a local minimum of ℓ_0 since ℓ_0 is lower semicontinuous and takes only finitely many values. For all $x \in \mathbb{R}^n$, choosing $\varepsilon < \min\left\{|x_j| \mid x_j \neq 0\right\}$, we can find a neighborhood of x such that $\ell_0(x) \leq \ell_0(y)$ for all $y \in \mathbb{B}_\varepsilon(x)$. Hence, it is not surprising that a set-valued Newton's method stays at the same point in every iteration.*

4 Projection Methods

On several occasions in the previous chapters, we have named the method of alternating projections and the Douglas-Rachford algorithm. These algorithms are two representatives of the class of *projection methods*. We name all basic properties and convergence results of the two mentioned algorithms.

4.1 Properties of the Projector

Before we give theoretical results on projection methods, we discuss the properties of the projection operator defined in (2.17).

The following theorem is a classical result of convex analysis. It states that the projector onto a closed, convex set is a nonexpansive mapping.

Theorem 4.1.1 ((Cheney and Goldstein, 1959, Theorem 3)). *The projection operator for a closed, convex set $\Omega \subset \mathcal{H}$ in a Hilbert space $\mathcal{H}$ satisfies the Lipschitz condition*

$$\|P_\Omega x - P_\Omega y\| \leq \|x - y\|, \tag{4.1}$$

equality holding only if $\|x - P_\Omega x\| = \|y - P_\Omega y\|$.

Proof. By the best approximation property of convex sets, we have

$$\begin{aligned}
\langle P_\Omega x - P_\Omega y, P_\Omega y - y \rangle &\geq 0 \\
\text{and} \quad \langle P_\Omega y - P_\Omega x, P_\Omega x - x \rangle &\geq 0.
\end{aligned} \tag{4.2}$$

Then we have $\langle P_\Omega x - P_\Omega y, P_\Omega y - y \rangle + \langle P_\Omega y - P_\Omega x, P_\Omega x - x \rangle \geq 0$, which is equivalent to

$$\langle P_\Omega x - P_\Omega y, x - y \rangle \geq \langle P_\Omega x - P_\Omega y, P_\Omega x - P_\Omega y \rangle = \|P_\Omega x - P_\Omega y\|^2. \tag{4.3}$$

Applying the Cauchy-Schwarz inequality to (4.3) gives us

$$\|P_\Omega x - P_\Omega y\| \, \|x - y\| \geq \|P_\Omega x - P_\Omega y\|^2, \tag{4.4}$$

which is equivalent to $\|x - y\| \geq \|P_\Omega x - P_\Omega y\|$. Equality holds here only if $\langle P_\Omega x - P_\Omega y, P_\Omega y - y \rangle = \langle P_\Omega y - P_\Omega x, P_\Omega x - x \rangle = 0$ and if $P_\Omega y - P_\Omega x = \lambda (y - x)$ $\lambda \in \mathbb{R}$. From this, we conclude

$$\langle y - x, P_\Omega y - y \rangle = 0 \quad \text{and} \quad \langle y - x, P_\Omega x - x \rangle = 0.$$

Finally, computing

$$
\begin{aligned}
0 &= \langle P_\Omega x - P_\Omega y, P_\Omega y - y \rangle - \langle P_\Omega y - P_\Omega x, P_\Omega x - x \rangle \\
&\quad + \langle y - x, P_\Omega y - y \rangle + \langle y - x, P_\Omega x - x \rangle \\
&= \| P_\Omega x - x \|^2 - \| P_\Omega y - y \|^2,
\end{aligned}
\tag{4.5}
$$

we get the desired result. □

The next result is on preimages of the projector. Note that the underlying set C is not required to be convex.

Lemma 4.1.2. *Let $C \subset \mathbb{X}$ be a nonempty, closed subset and let $z \in C$. Then $P_C^{-1} z$ is a nonempty, convex subset of $\mathbb{X}$.*

Proof. Assume there exist $x, y \in P_C^{-1} z$ such that we find a $\tilde{\lambda} \in [0,1]$ together with $\tilde{z} \in C$ such that

$$
\| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - \tilde{z} \| < \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - z \|
$$

and $x, y \notin P_C^{-1} \tilde{z}$. Because $x, y \notin P_C^{-1} \tilde{z}$, we have $\tilde{\lambda} \in (0,1)$. By Pythagoras, this gives us

$$
\| x - \tilde{z} \|^2 = \| x - \tilde{\lambda} x - (1 - \tilde{\lambda}) y \|^2 + \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - \tilde{z} \|^2
\tag{4.6}
$$

and

$$
\| y - \tilde{z} \|^2 = \| y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y \|^2 + \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - \tilde{z} \|^2.
\tag{4.7}
$$

Define

$$
\tau := \underset{\lambda \in [0,1]}{\operatorname{argmin}} \| \lambda x + (1 - \lambda) y - z \|.
\tag{4.8}
$$

1. If $\tau \in \{0, 1\}$, then either x or y is the minimizer of the distance between the points on the line $\{ \lambda x + (1 - \lambda) y \mid \lambda \in [0,1] \}$ and z. Without loss of generality, assume $\| x - z \| \leq \| y - z \|$. This means that

$$
\langle y - \lambda x - (1 - \lambda) y, z - \lambda x - (1 - \lambda) y \rangle \geq 0 \quad \text{for all} \quad \lambda \in [0,1].
\tag{4.9}
$$

Then, by Pythagoras,

$$
\begin{aligned}
\| y - \tilde{z} \|^2 &= \| y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y \|^2 + \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - \tilde{z} \|^2 \\
&< \| y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y \|^2 + \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - z \|^2 \\
&\leq \| y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y \|^2 + \| \tilde{\lambda} x + (1 - \tilde{\lambda}) y - z \|^2 \\
&\quad + \langle y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y, \tilde{\lambda} x + (1 - \tilde{\lambda}) y - z \rangle \\
&= \| y - \tilde{\lambda} x - (1 - \tilde{\lambda}) y + \tilde{\lambda} x + (1 - \tilde{\lambda}) y - z \|^2 \\
&= \| y - z \|^2.
\end{aligned}
\tag{4.10}
$$

This is a contradiction since $\| y - \tilde{z} \| > \| y - z \|$.

2. If $\tau \in (0, 1)$, then we have (by Pythagoras again)

$$
\| x - z \|^2 = \| x - \tau x - (1 - \tau) y \|^2 + \| \tau x + (1 - \tau) y - z \|^2
\tag{4.11}
$$

and
$$\|y - z\|^2 = \|y - \tau x - (1 - \tau)y\|^2 + \|\tau x + (1 - \tau)y - z\|^2. \qquad (4.12)$$

Without loss of generality, assume $\tilde{\lambda} > \tau$. Then

$$\|x - \tilde{\lambda}x - (1 - \tilde{\lambda})y\| < \|x - \tau x - (1 - \tau)y\|. \qquad (4.13)$$

Further,
$$\|\tilde{\lambda}x + (1 - \tilde{\lambda})y - \tilde{z}\|^2 = \|x - \tilde{z}\|^2 - \|x - \tilde{\lambda}x - (1 - \tilde{\lambda})y\|^2. \qquad (4.14)$$

Observe that the angle between

$$\tilde{\lambda}x + (1 - \tilde{\lambda})y - \tilde{z} \quad \text{and} \quad \tilde{\lambda}x + (1 - \tilde{\lambda})y - x \qquad (4.15)$$

is larger than $\frac{\pi}{2}$ (see Figure 4.1 for a justification of this claim). This means that

$$\|\tilde{\lambda}x + (1 - \tilde{\lambda})y - z\|^2 < \|x - z\|^2 - \|x - \tilde{\lambda}x - (1 - \tilde{\lambda})y\|^2. \qquad (4.16)$$

Since $\|x - z\|^2 < \|x - \tilde{z}\|^2$, we see that

$$\|\tilde{\lambda}x + (1 - \tilde{\lambda})y - z\|^2 < \|\tilde{\lambda}x + (1 - \tilde{\lambda})y - \tilde{z}\|^2, \qquad (4.17)$$

which is a contradiction to the assumption. This proves the lemma. $\qquad\square$
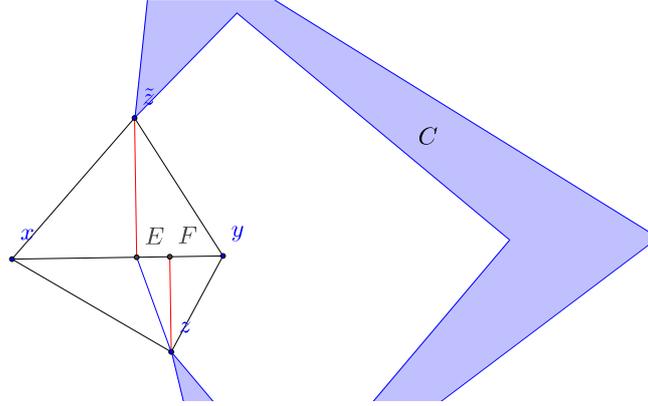
Figure 4.1: The point z represents the closest point in C to both x and y. The shortest distance from the line connecting x and y to the point $\tilde{z} \in C$, given by the line between E and $\tilde{z}$, is always larger than the distance from E to z.

The proof of Lemma 4.1.2 is very elementary. An alternative proof is as follows: by (Bauschke and Combettes, 2011, Example 20.12), the projector onto a nonempty subset C of $\mathcal{H}$ is a monotone operator. By (Kassay et al., 2009, Theorem 3.5), the inverse of a monotone operator is always convex-valued. The lemma then is a corollary of (Kassay et al., 2009, Theorem 3.5).

4.2 There and Back Again: An Alternating Projections' Tale

The idea of the method of alternating projections is as simple as it can get if one seeks to find the intersection of two closed sets: start at a given initial point x^0, then project onto the first set, and afterwards compute the projection onto the second set. Keep doing this procedure until a stopping criterion is fulfilled. Among these stopping criteria, we normally use the one where consecutive iterates are close to each other.

Researchers usually refer to John von Neumann's work as the initialization of the study of alternating projections. The result can, for example, be found in his book on functional operators (von Neumann, 1951, Theorem 13.7). There, von Neumann shows convergence of the sequence of projections onto linear subspaces to the projection onto the intersection of these subsapces[1].

Another article which has to be mentioned here is the one by Kaczmarz from 1937 (Kaczmarz, 1937). There, the author proves the convergence of the method of sucessive projections to the point in the intersection of n affine equations in an n-dimensional vector space, i.e., he proves convergence to the unique solution to a system of linear equations.

Research on this simple method was being pushed in the beginning of the 1990's. In the works (Combettes and Trussell, 1990), (Bauschke and Borwein, 1993), and in the article (Bauschke and Borwein, 1996), there has been a development in finding sufficient and necessary conditions for convergence.

It is remarkable that the first results, including the one by von Neumann, are restricted to finding the nonempty intersection of a family of convex sets. The key in the development away from convex sets lies in defining regularity conditions on the sets and on the intersections of the sets. With these conditions, convexity is no longer necessary (Lewis et al., 2009), (Hesse and Luke, 2013).

Definition 4.2.1 (alternating projections). *For two closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, the mapping*

$$T_{AP}x := P_{\Omega_1} P_{\Omega_2} x \tag{4.18}$$

is called the alternating projections operator. The corresponding alternating projections algorithm is given by the iteration

$$x^{k+1} \in T_{AP}x^k, \ k \in \mathbb{N}, \tag{4.19}$$

with x^0 given.

Other well-known algorithms, such as steepest descents for minimizing the average of squared distances between sets, can be formulated as instances of the alternating projections algorithm (Pierra, 1976), (Pierra, 1984). In Corollary 5.2.10, we show that, for problems with special linear structure, the alternating projections algorithm corresponds to projected gradients, based on (Hesse et al., 2014). We wish to remark that the method described in Definition 4.2.1 can be generalized to the case of more than two sets (see (Bauschke et al., 1997, Fact 1.1.1)).

---

[1]This is actually the central property of Dykstra's projection algorithm (Boyle and Dykstra, 1986), which coincides with the alternating projections in the case of linear subspaces.

4.2.1 Convex Results

The most widely studied case is the case where the sets which we project onto are closed, convex subsets of some vector space, which is in our case the Euclidean space $\mathbb{R}^n$. The reason for that lies in the fact that the projector (2.17) onto a closed and convex subset is always single-valued. Therefore, let Ω_1 and Ω_2 be two nonempty, closed, convex subsets of $\mathbb{R}^n$. At this point, we remark that in (Bauschke and Borwein, 1993) the authors give results for nonempty subsets of an arbitrary Hilbert space $\mathbb{X}$. However, we restrict ourselves to $\mathbb{X} = \mathbb{R}^n$. Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be closed and convex subsets. We formulate the feasibility problem

$$\text{find } \Omega_1 \cap \Omega_2. \tag{4.20}$$

In applied mathematics, the sets Ω_1 and Ω_2 are normally used to model a problem from physics, economics, chemistry, etc. Among these, the phase retrieval problem will be studied in more details in Chapter 9.

A crucial part of the analysis of an iterative method is the knowledge of its fixed points. The convergence results for projection methods always rely on certain regularity conditions. It is not always clear if these conditions are fulfilled in the settings of, for example, physical experiments.

If we borrow some notation from the regime of time-discrete dynamical systems, then we can declare different kinds of fixed points of the operator defined in (4.2.1). By a *stable* fixed point of an operator T, we denote a point x such that, for all y in a neighborhood of x, all iterates $T^k(y)$ stay in the same neighborhood of x. Otherwise, the fixed point will be called *unstable*. An example for an unstable fixed point is given in Figure 4.2.

Theorem 4.2.2 ((Cheney and Goldstein, 1959)). *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be closed convex subsets. Then $\bar{x} \in \Omega_1$ is a fixed point of the operator $P_{\Omega_1} P_{\Omega_2}$ if and only if $\bar{x}$ is a nearest point to Ω_2.*

Proof. The following proof can be found in (Cheney and Goldstein, 1959, Theorem 2). Suppose $y = P_{\Omega_2} x$ and $x = P_{\Omega_1} y$ and, to avoid trivialities, let $x \notin \Omega_2$ and $y \notin \Omega_1$. For arbitrary $z_1 \in \Omega_1$, we have $\langle x - z_1, x - y \rangle \leq 0$ as well as $\langle y - z_2, y - x \rangle \leq 0$ for an arbitrary $z_2 \in \Omega_2$. From this, we conclude

$$\langle z_1 - z_2, x - y \rangle \geq \langle x - y, x - y \rangle. \tag{4.21}$$

Using the Cauchy-Schwarz inequality, we obtain

$$\|z_1 - z_2\| \geq \|x - y\|. \tag{4.22}$$

To prove the converse, we suppose $\|P_{\Omega_2} x - x\| \leq \|P_{\Omega_2} z - z\|$ for all $z \in \Omega_1$. If $z = P_{\Omega_1} P_{\Omega_2} x$, then we get, by using the definition of the projection operator,

$$\|z - P_{\Omega_2} z\| \leq \|z - P_{\Omega_2} x\| \leq \|x - P_{\Omega_2} x\| \leq \|z - P_{\Omega_2} z\|. \tag{4.23}$$

By the uniqueness of the projection, since Ω_1 and Ω_2 are convex, we get $x = z$. □

Theorem 4.2.3 ((Bauschke and Borwein, 1993, Theorem 3.13)). *Let A, B be nonempty, convex, closed subsets of $\mathbb{R}^n$, and define $g := P_{A-B}0$. Let $(A, (B - g))$ have linearly regular intersection. Then any sequence generated by the AP-Operator converges at linear rate to a best approximation pair of A and B (Definition 2.3.7).*

4.2.2 Nonconvex Results
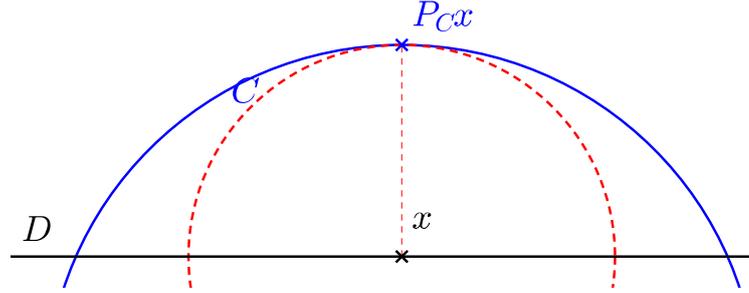
Figure 4.2: This example shows an unstable fixed point x of the alternating projections sequence. If the projection onto C is perturbed arbitrarily, then the sequence will move to the intersection of C and D. The dashed circle line illustrates that $P_C x$ is indeeed the projection of x onto C.

Lemma 4.2.4 (nonincreasing distance). *For two closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, let T_{AP} be given as in (4.18). For $x^0 \in \mathbb{R}^n$ chosen arbitrarily, generate the sequence as in (4.19) and define the sequence*

$$y^k := d_{\Omega_2}(x^k). \tag{4.24}$$

Then y^k is a monotonically decreasing sequence of nonnegative values.

Proof. Because each y^k is a distance, it is a positive real number. For all k, the value of y^k equals $\|x^k - P_{\Omega_2} x^k\|$. Then

$$\left\| P_{\Omega_2} x^k - P_{\Omega_1} P_{\Omega_2} x^k \right\| \leq \left\| x^k - P_{\Omega_2} x^k \right\|. \tag{4.25}$$

Hence,

$$\left\| P_{\Omega_1} P_{\Omega_2} x^k - P_{\Omega_2} P_{\Omega_1} P_{\Omega_2} x^k \right\| \leq \left\| P_{\Omega_2} x^k - P_{\Omega_1} P_{\Omega_2} x^k \right\|, \tag{4.26}$$

which is equivalent to $y^{k+1} = d_{\Omega_2}(x^{k+1}) \leq d_{\Omega_2}(x^k) = y^k$. $\qquad \square$

Remark 4.2.5. *The result of Lemma 4.2.4 solely bounds the distances y^k of the iterates x^k to the second set Ω_2. It gives no statement on the boundedness or convergence of the sequence $\{x^k\}_{k \in \mathbb{N}}$ at all. In fact, without additional assumptions on the sets Ω_1 and Ω_2, there exist*

examples of closed and convex sets where the sequence x^k is unbounded. Consider the case

$$\Omega_1 = \left\{ x \in \mathbb{R}^2 \mid x_2 = 0 \right\} \quad and \quad \Omega_2 = \left\{ x \in \mathbb{R}^2 \mid x_1 > 0, \ x_2 \geq \frac{1}{x_1} \right\}. \tag{4.27}$$

Then the sequence as given in (4.18) is unbounded.

Together with regularity conditions, we get local convergence results of the alternating projections.

Lemma 4.2.6 (local linear convergence of alternating projections (Hesse and Luke, 2013, Corollary 3.13)). *Let the collection (Ω_1, Ω_2) be locally linearly regular at $\bar{x} \in \Omega := \Omega_1 \cap \Omega_2$ with modulus of regularity κ on $\mathbb{B}_\delta(\bar{x})$. Further, let Ω_1 and Ω_2 be (ε, δ)-subregular at $\bar{x}$. For any $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$, generate the sequence $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ by alternating projections, that is, $x^{k+1} \in T_{AP}x^k$. Then*

$$d_\Omega(x^{k+1}) \leq \left(1 - \frac{1}{\kappa^2} + \varepsilon \right) d_\Omega(x^k). \tag{4.28}$$

Consequently, as long as $\frac{1}{\kappa^2} > \varepsilon$, the alternating projections sequence converges at linear rate.

4.3 Douglas-Rachford

The Douglas-Rachford algorithm was proposed by Lions and Mercier in (Lions and Mercier, 1979) for solving inclusions of the form $0 \in A + B$ where A and B are maximally monotone operators. An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone* (see (Bauschke and Combettes, 2011, Definition 20.1)) if we have

$$\langle z - w, x - y \rangle \geq 0 \quad \text{for all} \quad (x, z), (y, w) \in \text{gph}(A). \tag{4.29}$$

Citing (Bauschke and Combettes, 2011, Definition 20.20), an operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *maximally monotone* if there exists no monotone operator $C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ such that $\text{gph}(A)$ is a proper subset of $\text{gph}(C)$. Let $f : \mathbb{R}^n \to \overline{R}$ and $g : \mathbb{R}^n \to \overline{R}$ be proper, convex, and lower semicontinuous functions. For the case that $A = \partial f$ and $B = \partial g$, i.e., A and B are subdifferentials (2.1.9) of f and g we know, by (Moreau, 1965, Proposition 12b), that A and B are maximally monotone operators. If now, by (Bauschke and Combettes, 2011, Corollary 16.38 (iv)), $\text{ri dom } f \cap \text{ri dom } g \neq \emptyset$, then we have a sum rule for subdifferentials, i.e., $\partial f + \partial g = \partial(f + g)$. In that case, by Theorem 2.2.2, the Douglas-Rachford algorithm can be applied to satisfy necessary and sufficient optimality conditions for the sum $f + g$.

We define, for maximally monotone operators A and B, the resolvents of A and B via $J_A^\lambda := (\text{Id} - \lambda A)^{-1}$ and $J_B^\lambda := (\text{Id} - \lambda B)^{-1}$, respectively. Then the Douglas-Rachford iteration in terms of resolvents is given by (Lions and Mercier, 1979, Equation 10)

$$x^{k+1} = T_{DRop}x^k := J_A^\lambda \left(2J_B^\lambda - \text{Id} \right) x^k + \left(\text{Id} - J_B^\lambda \right) x^k. \tag{4.30}$$

Now, let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be nonempty and convex sets. Then ι_{Ω_1} and ι_{Ω_2} are lower semicontinuous, convex, and proper functions. The subdifferentials of the indicator functions are given by (cf. (Bauschke and Combettes, 2011, Example 16.12))

$$\partial \iota_{\Omega_1} = N_{\Omega_1}, \quad \partial \iota_{\Omega_2} = N_{\Omega_2}. \tag{4.31}$$

Inserting this into the resolvent $J_A^\lambda := (\text{Id} - \lambda A)^{-1}$ gives us $J_{N_{\Omega_1}}^\lambda = (\text{Id} - \lambda N_{\Omega_1})^{-1}$. Because N_{Ω_1} is a cone, we have $J_{N_{\Omega_1}}^{\lambda_1} = J_{N_{\Omega_1}}^1$ for all $\lambda > 0$. By (Bauschke and Combettes, 2011, Example 23.4), we know that $J_{N_{\Omega_1}}^1 = P_{\Omega_1}$. The same applies for Ω_2. Now, consider the iteration given in (4.30) with the resolvents of N_{Ω_1} and of N_{Ω_2}. Then, we obtain

$$\begin{aligned}
x^{k+1} &= J_A^\lambda \left(2J_B^\lambda - \text{Id} \right) x^k + \left(\text{Id} - J_B^\lambda \right) x^k \\
&= J_{N_{\Omega_1}}^1 \left(2J_{N_{\Omega_2}}^1 - \text{Id} \right) x^k + \left(\text{Id} - J_{N_{\Omega_2}}^1 \right) x^k \\
&= P_{\Omega_1} \left(2P_{\Omega_2} - \text{Id} \right) x^k + \left(\text{Id} - P_{\Omega_2} \right) x^k \\
&= \frac{1}{2} \left(2P_{\Omega_1} \left(2P_{\Omega_2} - \text{Id} \right) x^k - \left(2P_{\Omega_2} - \text{Id} \right) x^k + x^k \right) \\
&= \frac{1}{2} \left(\left(2P_{\Omega_1} - \text{Id} \right) \left(2P_{\Omega_2} - \text{Id} \right) x^k + x^k \right) \\
&= \frac{1}{2} \left(\left(R_{\Omega_1} R_{\Omega_2} + \text{Id} \right) x^k \right).
\end{aligned}$$

With this reformulation, we can define a Douglas-Rachford algorithm for feasibility problems.

Definition 4.3.1 (Douglas-Rachford). *For two closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, the mapping*

$$T_{DR} x := \frac{1}{2} \left(R_{\Omega_1} R_{\Omega_2} x + x \right), \tag{4.32}$$

where R_{Ω_1} is given by (2.19), *is called the Douglas-Rachford operator. The corresponding Douglas-Rachford algorithm for feasibility problems is the fixed point iteration $x^{k+1} \in T_{DR} x^k$, $k \in \mathbb{N}$, with x^0 given.*

In the past decades, there has been a broad research on applications of the Douglas-Rachford algorithm to different kinds of problems. For example, it can be applied to the determination of the shape of proteins (Aragón Artacho et al., 2013), road design (Bauschke and Koch, 2015), quantum channel construction (Drusvyatskiy et al., 2015), Cadzow denoising (Condat and Hirabayashi, 2015), and solutions to the phase retrieval (Bauschke et al., 2002).

4.3.1 General Convergence Results

In (Lions and Mercier, 1979, Proposition 2), it is shown that, for a firmly nonexpansive T_{DRop} as in (4.30), there is a weak convergence of the iteration. This includes also the case where Ω_1, Ω_2 are closed, convex sets. Further, in case that B is both coercive and Lipschitz, Douglas-Rachford was shown to converge linearly (Lions and Mercier, 1979, Proposition 4). Let $B = N_{\Omega_2}$ for Ω_2 closed and convex. In the following, we will show that the normal cone mapping is not necessarily Lipschitz continuous. Lipschitz continuity for set-valued mappings can be defined as follows (Dontchev and Rockafellar, 2014, Chapter 3.3): a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be Lipschitz continuous relative to a nonempty set D in $\mathbb{R}^m$ if $D \subset \text{dom}(S)$, if S is closed-valued on D, and if there exists $\kappa \geq 0$ such that

$$S(y') \subset S(y) + \kappa \left\| y' - y \right\| \mathbb{B} \quad \text{for all } y', y \in D. \tag{4.33}$$

As a counterexample, let Ω be the set $\left\{ x \in \mathbb{R}^2 \mid x_2 \geq 0 \right\}$. Then, for all points $x \in \text{int}(\Omega)$, we have $N_\Omega x = \{0\}$, while at the boundary, for example at 0, we have

$$N_\Omega 0 = \left\{ (0, -\lambda) \in \mathbb{R}^2 \mid \lambda \geq 0 \right\}.$$

Hence, there is no $\kappa \geq 0$ such that

$$\{(0, -\lambda) \mid \lambda \geq 0\} \subset \{0\} + \kappa \left\| x - (0,0) \right\| \mathbb{B}.$$

Hence, the result in (Lions and Mercier, 1979, Proposition 4) does not apply for Douglas-Rachford for feasibility. Convergence of the latter *with rates* has been, except for specific examples (Borwein and Sims, 2011), unknown until 2013. We cite the result in the following lemma.

Lemma 4.3.2 (local linear convergence of Douglas-Rachford (Hesse and Luke, 2013, Corollary 3.20)). *Let Ω_1, Ω_2 be two affine subspaces with $\Omega_1 \cap \Omega_2 \neq \emptyset$. The Douglas-Rachford algorithm converges to $\Omega_1 \cap \Omega_2$ for all $x^0 \in \mathbb{R}^n$ if and only if the collection (Ω_1, Ω_2) is strongly regular (2.33), in which case, convergence is linear.*

In (Phan, 2015, Theorem 4.7) it was shown, that the Douglas-Rachford for feasibility problems converges locally at a linear rate for super-regular closed sets Ω_1, Ω_2 (Definition 2.4.1(4)). The sets in (Phan, 2015) have to satisfy a weaker regularity condition than strong regularity: for $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, let $L = \text{aff}(\Omega_1 \cup \Omega_2)$. We say that $\{\Omega_1, \Omega_2\}$ is *affine hull regular* at $x \in \Omega_1 \cap \Omega_2$ if

$$N_{\Omega_1}(x) \cap (L - x) \cap (-(N_{\Omega_2}(x) \cap (L - x))) = \{0\}. \tag{4.34}$$

4.3.2 Convex Convergence Results

A remarkable fact about the Douglas-Rachford for feasibility problem operator is its behavior in the case of infeasibility. When Ω_1, Ω_2 are not convex, then the question of convergence is open. For Ω_1, Ω_2 closed and convex, we cite the following result:

Theorem 4.3.3 ((Bauschke et al., 2004, Theorem 3.13)). *Let Ω_1, Ω_2 be closed and convex subsets with $g = P_{\overline{\Omega_1 - \Omega_2}} 0$. Let $x^0 \in \mathbb{R}^n$ and let the sequence $\{x^k\}_{k \in \mathbb{N}}$ be generated by T_{DR}. Then the following hold.*

1. *$x^k - x^{k+1} \to g$.*

2. *If $\Omega_1 \cap \Omega_2 \neq \varnothing$, then the sequence of x^k converges to a fixed point of T_{DR}. Otherwise, $\left\| x^k \right\| \to \infty$.*

Definition 4.3.4 (shadows of Douglas-Rachford). *For two closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ and some $x^0 \in \mathbb{R}^n$, let the sequence $\{x^k\}_{k \in \mathbb{N}}$ be generated via*

$$x^{k+1} = T_{DR} x^k \tag{4.35}$$

as in Definition 4.3.1. The shadow sequence *$\{y^k\}_{k \in \mathbb{N}}$ of $\{x^k\}_{k \in \mathbb{N}}$ is defined as*

$$y^k := P_{\Omega_2} x^k. \tag{4.36}$$

In contrast to the alternating projections algorithm, the iterates of the Douglas-Rachford algorithm are not actually the points of interest – it is rather the shadows of the iterates that are relevant. This results in an occasional incongruence between the fixed points of Douglas-Rachford and the intersection that we seek.

Lemma 4.3.5 (fixed points of Douglas-Rachford (Bauschke et al., 2004, Corollary 3.9)). *Suppose that $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ are closed, convex, and such that $\Omega_1 \cap \Omega_2 \neq \varnothing$. Then*

$$\mathrm{Fix}(T_{DR}) = (\Omega_1 \cap \Omega_2) + N_D(0)$$

where $D := \overline{\Omega_2 - \Omega_1}$.

Remark 4.3.6. *Lemma 4.3.5 can be generalized to closed, convex neighborhoods in Ω_1 and Ω_2. Suppose there exist $x \in \Omega_1 \cap \Omega_2$ and $\varepsilon > 0$ such that $\Omega_1 \cap \mathbb{B}_\varepsilon(x)$ and $\Omega_2 \cap \mathbb{B}_\varepsilon(x)$ are convex sets. Define*

$$D_\varepsilon := \overline{(\Omega_2 \cap \mathbb{B}_\varepsilon(x)) - (\Omega_1 \cap \mathbb{B}_\varepsilon(x))}.$$

Then the inclusion $x + N_{D_\varepsilon}(0) \subset \mathrm{Fix}(T_{DR})$ holds.

4.3.3 The Linear Case

We give an auxiliary result that the Douglas-Rachford iteration applied to *linear subspaces* converges to its set of fixed points with linear rate. As the sparse feasibility problem locally reduces to finding the intersection of (affine) subspaces, by a translation to the origin, results for the case of subspaces will yield local linear convergence of Douglas-Rachford to fixed points associated with points $\bar{x} \in A_s \cap B$ such that $\ell_0(\bar{x}) = s$.

The idea of our proof is to show that the set of fixed points of the Douglas-Rachford algorithm applied to the subspaces A and B can always be written as the intersection of different subspaces $\widetilde{A}$ and $\widetilde{B}$, the collection of which is *strongly regular*. We then show

that the iterates of the Douglas-Rachford algorithm applied to the subspaces A and B are identical to those of the Douglas-Rachford algorithm applied to the subspaces $\widetilde{A}$ and $\widetilde{B}$. Linear convergence of Douglas-Rachford then follows directly from Lemma 4.3.2.

We recall that the set of fixed points of Douglas-Rachford in the case of two linear subspaces $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ is, by (Bauschke et al., 2004, Corollary 3.9) and (5.41), equal to

$$\text{Fix } T_{DR} = (A \cap B) + \left(A^\perp \cap B^\perp \right)$$

for $T_{DR} := \frac{1}{2} \left(R_A R_B + \text{Id} \right)$. For two linear subspaces $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$, define the enlargements $\widetilde{A} := A + \left(A^\perp \cap B^\perp \right)$ and $\widetilde{B} := B + \left(A^\perp \cap B^\perp \right)$. By definition of the Minkowski sum in Equation (2.15), these enlargements are given by

$$\widetilde{A} = \left\{ a + n \,\middle|\, a \in A, n \in A^\perp \cap B^\perp \right\} \tag{4.37a}$$

$$\text{and } \widetilde{B} = \left\{ b + n \,\middle|\, b \in B, n \in A^\perp \cap B^\perp \right\}. \tag{4.37b}$$

The enlargements $\widetilde{A}$ and $\widetilde{B}$ are themselves subspaces of $\mathbb{R}^n$ as the Minkowski sum of subspaces.

Lemma 4.3.7 ((Hesse et al., 2014, Lemma IV.3)). *The equation*

$$C := \left(A + \left(A^\perp \cap B^\perp \right) \right)^\perp \cap \left(B + \left(A^\perp \cap B^\perp \right) \right)^\perp = \{0\}$$

holds for any linear subspaces A and B of $\mathbb{R}^n$, and hence the collection $(\widetilde{A}, \widetilde{B})$ is strongly regular for any linear subspaces A and B.

Proof. Let v be an element of C. Because $C = \widetilde{A}^\perp \cap \widetilde{B}^\perp$, we know that

$$\langle v, \widetilde{a} \rangle = \langle v, \widetilde{b} \rangle = 0 \quad \text{for all } \widetilde{a} \in \widetilde{A}, \widetilde{b} \in \widetilde{B}. \tag{4.38}$$

Further, since $A \subset \widetilde{A}$ and $B \subset \widetilde{B}$, we have

$$\langle v, a \rangle = \langle v, b \rangle = 0 \quad \text{for all } a \in A, b \in B. \tag{4.39}$$

In other words, $v \in A^\perp$ and $v \in B^\perp$, so $v \in A^\perp \cap B^\perp$. On the other hand, $A^\perp \cap B^\perp \subset \widetilde{A}$ and $A^\perp \cap B^\perp \subset \widetilde{B}$, so we similarly have

$$\langle v, n \rangle = 0 \quad \text{for all } n \in A^\perp \cap B^\perp \tag{4.40}$$

because A and B are subspaces and $v \in C$. Hence, v is also an element of $\left(A^\perp \cap B^\perp \right)^\perp$. We conclude that v can only be zero. □

Lemma 4.3.8 ((Hesse et al., 2014, Lemma IV.4)). *Let A and B be linear subspaces, and let $\widetilde{A}$ and $\widetilde{B}$ be their corresponding enlargements defined by (4.37). Then the following statements hold.*

(i) $R_A d = -d$ for all $d \in A^\perp$.

(ii) $R_A x = R_{\widetilde{A}} x$ for all $x \in A + B$.

(iii) $R_{\widetilde{B}} a \in A + B$ for all $a \in A$.

(iv) $R_{\widetilde{A}} R_{\widetilde{B}} x = R_A R_B x$ for all $x \in \mathbb{R}^n$.

(v) For any $x \in \mathbb{R}^n$, the following equality holds:

$$\frac{1}{2} \left(R_{\widetilde{A}} R_{\widetilde{B}} + \mathrm{Id} \right) x = \frac{1}{2} \left(R_A R_B + \mathrm{Id} \right) x.$$

Proof. (i) To prove (i), let $d \in A^\perp$ be arbitrary. The projection $P_A d$ of d onto A is the orthogonal projection onto A. The orthogonal projection of $d \in A^\perp$ is the zero vector. This means that $R_A d = (2P_A - \mathrm{Id})d = -d$.

(ii) Note that $(A^\perp \cap B^\perp) = (A + B)^\perp$. Hence, $\widetilde{A} = A + (A + B)^\perp$. Now, by (Bauschke et al., 2006, Proposition 2.6), $P_{A+(A+B)^\perp} = P_A + P_{(A+B)^\perp}$. It follows that, for all $x \in A + B$, $P_{\widetilde{A}} x = P_A x$ and, consequently, $R_{\widetilde{A}} x = R_A x$, as claimed.

(iii) Let $a \in A$ and thus $a \in A + B$. We note that, by (ii) with A replaced by B, we have $R_B a = R_{\widetilde{B}} a$. Write a as a sum $b + v$ where $b = P_B a$ and $v = a - P_B a$. We note that $v \in A + B$ and so $-v \in A + B$. From (i) we conclude, since A in (i) can be replaced by B and $v \in B^\perp$, that $R_B v = -v$. Since $b \in B$, we have $R_B b = 2P_B b - b = b$ and so

$$R_{\widetilde{B}} a = R_B a = R_B b + R_B v = b - v \in A + B. \tag{4.41}$$

(iv) To see (iv), let $x \in \mathbb{R}^n$ be arbitrary. Define $D := A^\perp \cap B^\perp$. Then we can write $x = a + b + d$ with $a \in A$, $b \in B$, and $d \in D$. This expression does not have to be unique since A and B may have a nontrivial intersection. In any case, we have the identity $\langle b, d \rangle = \langle a, d \rangle = 0$. Since A and B are linear subspaces, the Douglas-Rachford operator is a linear mapping, which, together with parts (i)-(iii) of this lemma, yields

$$
\begin{aligned}
R_A R_B x &= R_A \left(R_B a + R_B b + R_B d \right) \\
&\overset{\text{(i).}}{=} R_A \left(R_B a + b - d \right) \\
&= R_A R_B a + R_A b + R_A(-d) \\
&\overset{\text{(i).}}{=} R_A R_B a + R_A b + d \\
&\overset{\text{(ii).}}{=} R_A R_{\widetilde{B}} a + R_{\widetilde{A}} b + d \\
&\overset{\text{(ii).}-\text{(iii).}}{=} R_{\widetilde{A}} R_{\widetilde{B}} a + R_{\widetilde{A}} b + d \\
&\overset{d \in \widetilde{A}}{=} R_{\widetilde{A}} \left(R_{\widetilde{B}} a + b + d \right) \\
&\overset{b, d \in \widetilde{B}}{=} R_{\widetilde{A}} \left(R_{\widetilde{B}} a + R_{\widetilde{B}} b + R_{\widetilde{B}} d \right) \\
&= R_{\widetilde{A}} R_{\widetilde{B}} x.
\end{aligned}
\tag{4.42}
$$

This proves (iv).

(v) Statement (v) is an immediate consequence of (iv), which completes the proof.

□

Proposition 4.3.9 ((Hesse et al., 2014, Proposition IV.5)). *Let A and B be linear subspaces, and let $\widetilde{A}$ and $\widetilde{B}$ be their corresponding enlargements defined by (4.37). The Douglas-Rachford iteration applied to the enlargements,*

$$x^{k+1} = \widetilde{T}_{DR}x^k := \frac{1}{2}\left(R_{\widetilde{A}}R_{\widetilde{B}} + \mathrm{Id}\right)x^k, \tag{4.43}$$

converges with linear rate to $\mathrm{Fix}\,\widetilde{T}_{DR}$ *for any starting point* $x^0 \in \mathbb{R}^n$.

Proof. By Lemma 4.3.7, we know that the zero vector is the only common element in $\left(A + \left(A^\perp \cap B^\perp\right)\right)^\perp$ and $\left(B + \left(A^\perp \cap B^\perp\right)\right)^\perp$. By Lemma 4.3.2 (Hesse and Luke, 2013, Corollary 3.20), the sequence

$$\widetilde{x}_{k+1} := \frac{1}{2}\left(R_{\widetilde{A}}R_{\widetilde{B}} + \mathrm{Id}\right)\widetilde{x}_k$$

converges linearly to the intersection $\widetilde{A} \cap \widetilde{B}$ for any starting point $\widetilde{x}_0 \in \mathbb{R}^n$. □

Combining these results, we obtain the following theorem confirming *linear* convergence of the Douglas-Rachford algorithm for subspaces. Convergence of the Douglas-Rachford algorithm for *strongly regular* affine subspaces was proved in (Hesse and Luke, 2013, Corollary 3.20) as a special case of a more general result (Hesse and Luke, 2013, Theorem 3.18) about linear convergence of the Douglas-Rachford algorithm for a strongly regular collection of a *super-regular set* (Lewis et al., 2009, Definition 4.3) and an affine subspace. Our result below shows that the iterates of the Douglas-Rachford algorithm for *linearly regular* affine subspaces (not necessarily strongly regular) converge linearly to the fixed point set. An analysis focused only on the affine case in (Bauschke et al., 2014a) also achieves linear convergence of the Douglas-Rachford algorithm.

Theorem 4.3.10 ((Hesse et al., 2014, Theorem IV.6)). *For any two affine subspaces $A, B \subset \mathbb{R}^n$ with nonempty intersection, the Douglas-Rachford iteration*

$$x^{k+1} = T_{DR}x^k := \frac{1}{2}\left(R_A R_B + \mathrm{Id}\right)x^k \tag{4.44}$$

converges for any starting point x^0 *to a point in the fixed point set with linear rate. Moreover, $P_B\overline{x} \in A \cap B$ for $\overline{x} = \lim_{k\to\infty} x^k$.*

Proof. Without loss of generality, by translation of the sets A and B by $-\overline{x}$ for $\overline{x} \in A \cap B$, we consider the case of subspaces. By Proposition 4.3.9, Douglas-Rachford applied to the enlargements $\widetilde{A} = A + \left(A^\perp \cap B^\perp\right)$ and $\widetilde{B} = B + \left(A^\perp \cap B^\perp\right)$, namely (4.43), converges to the intersection $\widetilde{A} \cap \widetilde{B}$ with linear rate for any starting point $x^0 \in \mathbb{R}^n$. By

43

(Bauschke et al., 2004, Corollary 3.9) and (2.34), the set of fixed points of the Douglas-Rachford algorithm (4.44) is

$$\text{Fix}_{T_{DR}} = (A \cap B) + \left(A^{\perp} \cap B^{\perp} \right) = \widetilde{A} \cap \widetilde{B}, \tag{4.45}$$

where the rightmost equality follows from repeated application of the identity $(\Omega_1 + \Omega_2)^{\perp} = (\Omega_1^{\perp} \cap \Omega_2^{\perp})$, the definition of set addition, and closedness of subspaces under addition. By Lemma 4.3.8(v) the iterates of (4.43) are the same as the iterates of (4.44). Thus, the iterates of the Douglas-Rachford algorithm applied to A and B converge to a point in the set of its fixed points with linear rate. Finally, by (Bauschke et al., 2004, Corollary 3.9), $P_B \bar{x} \in A \cap B$ for any $\bar{x} \in \text{Fix } T_{DR}$. $\qquad \square$

5 Alternating Projections and Sparsity

This chapter will be at the intersection of Chapter 3 and Chapter 4. On the one hand, using regularity conditions, we can prove local convergence of projection methods to solve (3.8). On the other hand, we will give sufficient conditions for global convergence of the alternating projections method applied to (3.8).

In (Bauschke et al., 2014b), using tools developed in (Bauschke et al., 2013a) and (Bauschke et al., 2013b), the method of alternating projections applied to (3.8) was shown to be locally linearly convergent with optimal rates in terms of the *Friedrichs angle*[1] b etween A_s and B, and an estimate of the radius of convergence was shown. Our approach, based on (Hesse and Luke, 2013) and (Hesse et al., 2014), is in line with (Luke, 2008) but does not rely on local firm nonexpansiveness of the fixed point mapping. It has the advantage of being general enough to be applied to *any* fixed point mapping, but the price one pays for this generality is in the rate estimates, which may not be optimal or easy to compute.

5.1 Local Linear Convergence of Alternating Projections

This section is based on (Hesse et al., 2014, Section III). The first result shows the local linear convergence of alternating projections to a solution of (3.8). This was also shown in (Bauschke et al., 2014b, Theorem 3.19) using very different techniques. The approach taken here, based on the modulus of regularity κ on $\mathbb{B}_\delta(x)$, is more general, that is, it can be applied to other nonconvex problems. The correspondence between the modulus of regularity and the angle of intersection is yet an open problem.

Theorem 5.1.1 (local linear convergence of alternating projections (Hesse et al., 2014, Theorem III.8))**.** *Let A_s and B be defined by (3.5) and (3.6) with nonempty intersection, and let $\bar{x} \in A_s \cap B$. Choose $0 < \delta < \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\}$. For $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$, the alternating projections iterates converge linearly to the intersection $A_s \cap B$ with rate $\left(1 - \frac{1}{\kappa^2}\right)$ where κ is the modulus of regularity of (A_s, B) on $\mathbb{B}_\delta(\bar{x})$ (Definition 2.4.6).*

Proof. By Lemma 3.2.3(i) and (ii), the projections P_B and P_{A_s} each map $\mathbb{B}_{\delta/2}(\bar{x})$ to itself. Hence, their composition maps $\mathbb{B}_{\delta/2}(\bar{x})$ to itself.

Finally, we show that we may apply Lemma 4.2.6. The set B is $(0, +\infty)$-subregular at every point in B (i.e., convex), and by Theorem 3.2.5 the sparsity set A_s is $(0, \delta)$-subregular at $\bar{x}$. Lastly, by Theorem 3.3.2 the pair (A_s, B) is locally linearly regular at $\bar{x}$ on $\mathbb{B}_\delta(\bar{x})$ for any $\delta \in (0, \min\left\{|\bar{x}_j| \,\middle|\, j \in I(\bar{x})\right\})$. The assertion then follows from Lemma 4.2.6 with $\varepsilon = 0$. $\qquad\square$

---

[1]See Definition 6.1.2.

Remark 5.1.2. *The above result does* not *need an exact a priori assumption on the sparsity s. If there is a solution $\bar{x} \in A_s \cap B$, then $\ell_0(\bar{x})$ can be smaller than s. Geometrically speaking, $\bar{x}$ is on a crossing of linear subspaces contained in A_s. It is worth noting that the assumptions are also* not *tantamount to local convexity. In the case that B is a subspace, the point 0 is trivially a solution to (3.8) (and to (3.4)). The set A_s is not convex on any neighborhood of 0. However, the assumptions of Theorem 5.1.1 hold, and alternating projections indeed converges locally linearly to 0, regardless of the size of the parameter s.*
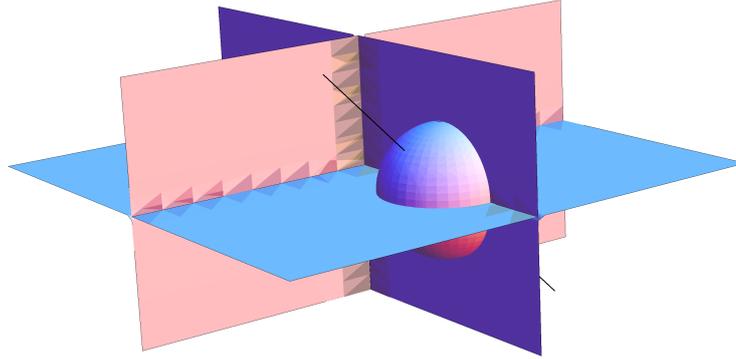
Figure 5.1: The set $A_2 \subset \mathbb{R}^3$ with an affine subspace (black line) and a neighborhood where the AP sequence converges to the intersection at a linear rate.

5.2 Global Convergence to Lower Level Sets

We clarify some notation on linear mappings that will be needed throughout this thesis.

Definition 5.2.1. *Let $M \in \mathbb{R}^{m \times n}$ be a linear mapping. From now on, assume that*

$$1) \quad m \leq n \text{ and}$$
$$2) \quad M \text{ is of full rank.} \tag{5.1}$$

The nullspace of M is denoted by ker *M, and $M^\dagger$ indicates the* Moore-Penrose *inverse (Moore, 1920), (Penrose, 1955) of M, which, because of the full rank assumption, becomes*

$$M^\dagger = M^\top \left(M M^\top \right)^{-1}. \tag{5.2}$$

The following definition is a generalization of the known concept of restricted isometry (Candès and Tao, 2005). Essentially, in the proofs including restricted isometry,

just the properties of lower semicontinuity, subadditivity, and the fact that $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}^n$ are used.

Definition 5.2.2. *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous (Definition 2.1.7) and subadditive function (Definition 2.1.6) satisfying $\inf_{x \in \mathbb{R}^n} \varphi(x) \neq -\infty$. Further, let $M : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping of full rank. The mapping M is said to satisfy the φ-RIP of order s if there exists $0 \leq \delta \leq 1$ such that*

$$(1 - \delta) \|x\|^2 \leq \left\| M^\dagger M x \right\|^2 \leq \|x\|^2 \quad \forall \, x \in \mathrm{lev}_{\leq s} \, \varphi. \tag{5.3}$$

The infimum δ_s of all such δ is the isometry constant for φ.

An important proporty of $M^\dagger M$ is that it is the orthogonal projection matrix onto the linear subspace $\ker(M)^\perp$. This means that the operator norm of $M^\dagger M$ is 1, and so we have, for all $x \in \mathbb{R}^n$, that

$$\|M^\dagger M x\|_2 \leq \|x\|_2. \tag{5.4}$$

This gives us the upper bound in (5.3) for free.

By Definition 2.1.7 based on (Rockafellar and Wets, 1998, Theorem 1.6), the lower semicontinuity of a function φ is equivalent to the set $\mathrm{lev}_{\leq s} \, \varphi$ being a closed subset of $\mathbb{R}^n$ for any s.

Now we define a minimization problem in a more general framework that can be specialized at two instances, namely, in Corollary 5.2.4 and Theorem 10.4.4.

$$\underset{x \in \mathrm{lev}_{\leq s} \, \varphi}{\text{minimize}} f(x) := \frac{1}{2} d_B(x)^2. \tag{5.5}$$

We can now prove one of the main results in this thesis.

Theorem 5.2.3 (global convergence of alternating projections to lower level sets). *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous and subadditive function satisfying $\inf_{x \in \mathbb{R}^n} \varphi(x) \neq -\infty$. Further, let $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}^n$. For a fixed $s > 0$, let the matrix $M^\dagger M$ satisfy (5.3) of order $2s$ with $\delta_{2s} \in [0, \frac{1}{2})$ for M in the definition of the affine set B given by (3.6). Then $B \cap \mathrm{lev}_{\leq s} \, \varphi$ is a singleton. Further, for any initial value $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by alternating projections (4.2.1) converges to $B \cap \mathrm{lev}_{\leq s} \, \varphi$ with $d_B(x^k) \to 0$ as $k \to \infty$ at a linear rate with constant bounded above by $\sqrt{\frac{\delta_{2s}}{1 - \delta_{2s}}}$.*

Proof. First, we show that $B \cap \mathrm{lev}_{\leq s} \, \varphi$ is a singleton. Assume there exist $x, y \in B \cap \mathrm{lev}_{\leq s} \, \varphi$. Then we have $Mx = My = p$. Since φ is subadditive, we know that $x - y \in \mathrm{lev}_{\leq 2s} \, \varphi$ and $M(x - y) = 0$. Hence, $\left\| M^\dagger M(x - y) \right\|^2 = 0$, which is a violation of (5.3) unless $x = y$. We conclude that $B \cap \mathrm{lev}_{\leq s} \, \varphi$ is a singleton.

To establish convergence, for the iterate x^k, define the mapping

$$q(x, x^k) := f(x^k) + \left\langle x - x^k, M^\dagger(Mx^k - p) \right\rangle + \frac{1}{2} \left\| x - x^k \right\|_2^2 \tag{5.6}$$

where f is the objective function defined in (5.5). By definition of the projector, the iterate x^{k+1} is a solution to the problem $\min\{q(x, x^k) \mid x \in \text{lev}_{\leq s}\, \varphi\}$. To see this, recall that, by the definition of the projection, $f(x^k) = \frac{1}{2}\|x^k - P_B(x^k)\|^2$. This yields

$$q(x, x^k) \overset{(3.16)}{=} \frac{1}{2}\left\|x^k - P_B(x^k)\right\|_2^2 + \left\langle x - x^k, x^k - P_B(x^k)\right\rangle + \frac{1}{2}\left\|x - x^k\right\|_2^2$$

$$= \frac{1}{2}\left\|x - x^k + x^k - P_B x^k\right\|_2^2 = \frac{1}{2}\left\|x - P_B x^k\right\|_2^2. \tag{5.7}$$

Now, by definition of the alternating projections sequence,

$$x^{k+1} \in P_{\text{lev}_{\leq s}\, \varphi} P_B(x^k) = P_{\text{lev}_{\leq s}\, \varphi}\left(x^k - (\text{Id} - P_B)x^k\right), \tag{5.8}$$

which, together with (5.7), yields

$$x^{k+1} \in \underset{x \in \text{lev}_{\leq s}\, \varphi}{\operatorname{argmin}}\left\{\left\|x - P_B x^k\right\|_2^2\right\} = \underset{x \in \text{lev}_{\leq s}\, \varphi}{\operatorname{argmin}}\{q(x, x^k)\}. \tag{5.9}$$

That is, x^{k+1} is a minimizer of $q(x, x^k)$ in $\text{lev}_{\leq s}\, \varphi$. Proceeding, we get

$$f(x^{k+1}) \overset{(5.5)}{=} \frac{1}{2}\left\|M^\dagger(Mx^{k+1} - p)\right\|_2^2$$

$$= \frac{1}{2}\left\|M^\dagger M(x^{k+1} - x^k) + M^\dagger(Mx^k - p)\right\|_2^2$$

$$= f(x^k) + \left\langle M^\dagger M(x^{k+1} - x^k), M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|M^\dagger M(x^{k+1} - x^k)\right\|_2^2$$

$$\leq f(x^k) + \left\langle M^\dagger M(x^{k+1} - x^k), M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$\overset{(5.2)}{=} f(x^k) + \left\langle M^\top(MM^\top)^{-1}M(x^{k+1} - x^k), M^\top(MM^\top)^{-1}(Mx^k - p)\right\rangle$$

$$+ \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$= f(x^k) + \left\langle MM^\top(MM^\top)^{-1}M(x^{k+1} - x^k), (MM^\top)^{-1}(Mx^k - p)\right\rangle$$

$$+ \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$= f(x^k) + \left\langle M(x^{k+1} - x^k), (MM^\top)^{-1}(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$= f(x^k) + \left\langle x^{k+1} - x^k, M^\top(MM^\top)^{-1}(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$= f(x^k) + \left\langle x^{k+1} - x^k, M^\dagger(Mx^k - p)\right\rangle + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_2^2$$

$$= q(x^{k+1}, x^k), \tag{5.10}$$

where the inequality in the middle follows from the fact that $M^\dagger M$ is an orthogonal

projection onto a subspace. Hence, $f(x^{k+1}) \leq q(x^{k+1}, x^k)$. But since x^{k+1} minimizes $q(x, x^k)$ over $\mathrm{lev}_{\leq s}\, \varphi$, we know that, for $\{\overline{x}\} = B \cap \mathrm{lev}_{\leq s}\, \varphi$,

$$q(x^{k+1}, x^k) \leq q(\overline{x}, x^k). \tag{5.11}$$

Moreover, by assumption (5.3) and by $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}^n$, we have

$$
\begin{aligned}
q(\overline{x}, x^k) =& f(x^k) + \left\langle \overline{x} - x^k, M^{\dagger}(Mx^k - p) \right\rangle + \frac{1}{2} \left\| \overline{x} - x^k \right\|_2^2 \\
\overset{(5.3)}{\leq}& f(x^k) + \left\langle \overline{x} - x^k, M^{\dagger}(Mx^k - p) \right\rangle + \frac{1}{2(1 - \delta_{2s})} \left\| M^{\dagger}M(\overline{x} - x^k) \right\|_2^2 \\
=& f(x^k) + \left\langle \overline{x} - x^k, M^{\dagger}(Mx^k - p) \right\rangle + \frac{1}{2(1 - \delta_{2s})} \left\| M^{\dagger}(p - Mx^k) \right\|_2^2 \\
\overset{(5.5)}{=}& \left(1 + \frac{1}{1 - \delta_{2s}}\right) f(x^k) + \left\langle \overline{x} - x^k, M^{\dagger}(Mx^k - p) \right\rangle \\
\overset{(5.2)}{=}& \left(1 + \frac{1}{1 - \delta_{2s}}\right) f(x^k) + \left\langle M^{\dagger}M(\overline{x} - x^k), M^{\dagger}(Mx^k - p) \right\rangle \\
\overset{(5.5)}{=}& \left(1 + \frac{1}{1 - \delta_{2s}}\right) 2f(x^k) - 2f(x^k) \\
=& \frac{\delta_{2s}}{1 - \delta_{2s}} f(x^k). \tag{5.12}
\end{aligned}
$$

When $0 \leq \delta_{2s} < \frac{1}{2}$, as assumed, we have $0 \leq \frac{\delta_{2s}}{1-\delta_{2s}} < 1$. Inequalities (5.10)-(5.12) then imply that $d_B(x^k) \to 0$ as $k \to \infty$ at a linear rate for $0 \leq \delta_{2s} < \frac{1}{2}$, with constant bounded above by $\sqrt{\frac{\delta_{2s}}{1-\delta_{2s}}} < 1$. Since the iterates x^k lie in $\mathrm{lev}_{\leq s}\, \varphi$, this proves convergence of the iterates to the intersection $\mathrm{lev}_{\leq s}\, \varphi \cap B$, that is, to $\overline{x}$, as claimed. $\qquad \square$

Next, we consider an application of Theorem 5.2.3. Namely, we apply it to Problem (3.8). Specifically, we assume that

$$M \quad \text{is of full rank and} \quad (1 - \delta_{2s}) \|x\|_2^2 \leq \left\| M^{\dagger}Mx \right\|_2^2 \quad \forall\, x \in A_{2s}. \tag{5.13}$$

Corollary 5.2.4 (global convergence of alternating projections in sparse affine feasibility (Hesse et al., 2014, Theorem III.15)). *For a fixed $s > 0$, let the matrix $M^{\dagger}M$ satisfy (5.13) with $\delta_{2s} \in [0, \frac{1}{2})$ for M in the definition of the affine set B given by (3.6). Then $B \cap A_s$ is a singleton. Further, for any initial value $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by alternating projections (Definition 4.2.1) converges to $B \cap A_s$ with $d_B(x^k) \to 0$ as $k \to \infty$ at a linear rate with constant bounded by $\sqrt{\frac{\delta_{2s}}{1-\delta_{2s}}}$.*

Proof. We note that the ℓ_0-function is, by Proposition 3.2.7, a lower semicontinuous and subadditive function. The lower level set of ℓ_0 is exactly equal to A_s. This means that property (5.13) is a specific instance of (5.3). Hence, the result follows from Theorem 5.2.3. $\qquad \square$

Corollary 5.2.4 shows that the values of the distance function g converge to zero at a linear rate. Together with the results established in Theorem 5.1.1, this actually shows linear convergence of the iterates to the intersection.

Theorem 5.2.5 (global linear convergence of alternating projections). *Under the assumptions of Corollary 5.2.4, for any initial point x^0 there exists a positive constant $\rho < 1$ such that, for $\bar{x} \in B \cap A_s$, the relation*

$$\left\|x^{k+1} - \bar{x}\right\| \leq \rho \left\|x^k - \bar{x}\right\| \tag{5.14}$$

holds for all $k \geq \tilde{k} \in \mathbb{N}$ with $\tilde{k}$ large enough.

Proof. By Corollary 5.2.4, we know that the squared distance of the iterates to the affine subspace B decreases at a linear rate to zero. This means that, for all $\varepsilon > 0$ and for all $x^0 \in \mathbb{R}^n$, there exists $\tilde{k} \in \mathbb{N}$ such that $\left\|x^k - P_B x^k\right\| \leq \varepsilon$ for all $k \geq \tilde{k}$. If we write A_s as a finite union of s-dimensional subspaces (see Equation (3.12)), i.e.,

$$A_s = \bigcup_{J \in \mathcal{J}} A_J,$$

we can define the set

$$D := \{ \mathrm{d}\,(B, A_J) \}_{J \in \mathcal{J}},$$

where $\mathrm{d}(B, A_J)$ is the distance between B and A_J (Definition 2.3.5). By Corollary 5.2.4, the intersection $A_s \cap B$ is equal to $\{\bar{x}\}$. This means that, for some $\bar{J} \in \mathcal{J}$, the distance $\mathrm{d}\left(B, A_{\bar{J}}\right)$ is equal to zero if and only if $\bar{x} \in A_{\bar{J}}$. Choose

$$\varepsilon < \tfrac{1}{2} \min \{ \lambda \in D \mid \lambda \neq 0 \}.$$

Then any point $x \in A_s$ satisfying $\|x - P_B x\| \leq \varepsilon$ cannot be an element of some A_J with $\mathrm{d}\,(B, A_J) > \varepsilon$. Hence, any point $x \in A_s$ satisfying $\|x - P_B x\| \leq \varepsilon$ is an element of $A_{\bar{J}}$ with $\bar{x} \in A_{\bar{J}}$.

For an arbitrary initial point x^0 given, choose $\tilde{k} \in \mathbb{N}$ such that the iterate $x^{\tilde{k}}$ satisfies $\left\|x^{\tilde{k}} - P_B x^{\tilde{k}}\right\| \leq \varepsilon$. Because, by Lemma 4.2.4, the distance of all iterates x^k, $k \geq \tilde{k}$, to B is less or equal to $\left\|x^{\tilde{k}} - P_B x^{\tilde{k}}\right\|$, we conclude that all iterates are elements of linear subspaces $A_{\bar{J}}$ with $\bar{x} \in A_{\bar{J}}$. In other words, at this instance, we have an alternating projections sequence between affine subspaces with a unique intersection $\bar{x}$. The linear convergence of the sequence $\{x^k\}_{k \geq \tilde{k}}$ to $\bar{x}$ follows from (Bauschke and Borwein, 1993). $\qquad \square$

After giving restrictions on the matrix $M^\dagger M$, we give a result using restricted isometry of the matrix M with respect to lower level sets itself. Following (Beck and Teboulle, 2011), where the authors consider the problem

$$\text{minimize } \tfrac{1}{2}\|Mx - p\|_2^2 \text{ subject to } x \in \mathrm{lev}_{\leq s}\, \varphi, \tag{5.15}$$

we present a sufficient condition for global linear convergence of the alternating projections algorithm for affine sparse feasibility. Though our presentation is modeled after (Beck and Teboulle, 2011), this work is predated by the nearly identical approach developed in (Blumensath and Davies, 2009) and (Blumensath and Davies, 2010) for $\varphi = \ell_0$. We also note that in light of Theorem 5.2.3, as well as in (Beck and Teboulle, 2011), the arguments presented do not use any structure that is particular to $\mathbb{R}^n$. Hence, the results can be extended to the problem of finding the intersection of the set of matrices with rank at most s and an affine subspace in the Euclidean space of matrices.

Key to the analysis of (Blumensath and Davies, 2009), (Blumensath and Davies, 2010), and (Beck and Teboulle, 2011) are the following well-known restrictions on the matrix M.

Definition 5.2.6. *The mapping $M : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the* restricted isometry property *of order s if there exists $0 \leq \delta \leq 1$ such that*

$$(1 - \delta)\|x\|_2^2 \leq \|Mx\|_2^2 \leq (1 + \delta)\|x\|_2^2 \quad \forall x \in \text{lev}_{\leq s} \, \varphi. \tag{5.16}$$

The infimum δ_s of all such δ is the restricted isometry constant.
The mapping $M : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the scaled/asymmetric restricted isometry property *(SRIP) of order (s, α) for $\alpha > 1$ if there exist $\nu_s, \mu_s > 0$ with $1 \leq \frac{\mu_s}{\nu_s} < \alpha$ such that*

$$\nu_s\|x\|_2^2 \leq \|Mx\|_2^2 \leq \mu_s\|x\|_2^2 \quad \forall x \in \text{lev}_{\leq s} \, \varphi. \tag{5.17}$$

The restricted isometry property (5.16) was introduced in (Candès and Tao, 2005) for the function $\varphi = \ell_0$, while the asymmetric version (5.17) first appeared in (Blumensath and Davies, 2009, Theorem 4). Clearly, (5.16) implies (5.17) since if a matrix M satisfies (5.16) of order s with restricted isometry constant δ_s, then it also satisfies (5.17) of order (s, β) for $\beta > \frac{1+\delta_s}{1-\delta_s}$.

To motivate the projected gradient algorithm given below, note that any solution to (3.8) is also a solution to

$$\text{find } \bar{x} \in S := \underset{x \in \text{lev}_{\leq s} \, \varphi}{\arg\min} \frac{1}{2}\|Mx - p\|_2^2. \tag{5.18}$$

Conversely, if $\text{lev}_{\leq s} \, \varphi \cap B \neq \varnothing$ and $\bar{x} \in S$, then $\bar{x}$ solves (3.8).

The condition (5.13) can be reformulated in terms of the scaled/asymmetric restricted isometry property (5.17), strong regularity of the range of $M^\top$, and the complement of each of the subspaces comprising A_{2s}. We remind the reader that

$$A_J := \text{span} \{e_i \mid i \in J\} \quad \text{for} \quad J \in \mathcal{J}_{2s} := \left\{ J \in 2^{\{1,2,\dots,n\}} \,\middle|\, J \text{ has } 2s \text{ elements} \right\}.$$

Proposition 5.2.7 (SRIP and strong regularity (Hesse et al., 2014, Proposition III.14))**.** *Let $M \in \mathbb{R}^{m \times n}$ with $m \leq n$ be of full rank. Then M satisfies (5.13) with $\delta_{2s} \in [0, \frac{\alpha - 1}{\alpha})$ for some fixed $s > 0$ and fixed $\alpha > 1$ if and only if $M^\dagger M$ satisfies the scaled/asymmetric restricted isometry property (5.17) of order $(2s, \alpha)$ with $\mu_{2s} = 1$ and $\nu_{2s} = (1 - \delta_{2s})$. Moreover, for*

M satisfying (5.13) *with* $\delta_{2s} \in [0, \frac{\alpha-1}{\alpha})$ *for some fixed* $s > 0$ *and fixed* $\alpha > 1$, *the collection* $(A_J^\perp, \mathrm{range}(M^\top))$ *is strongly regular (Definition 2.4.7) for all* $J \in \mathcal{J}_{2s}$, *that is,*

$$(\forall J \in \mathcal{J}_{2s}) \qquad A_J \cap \ker(M) = \{0\}. \tag{5.19}$$

Proof. The first statement follows directly from the definition of the scaled/asymmetric restricted isometry property.

For the second statement, note that if M satisfies the inequality (5.13) with $\delta_{2s} \in [0, \frac{\alpha-1}{\alpha})$ for some fixed $s > 0$ and fixed $\alpha > 1$, then the only element in A_{2s} satisfying $M^\dagger M x = 0$ is $x = 0$. Recall that $M^\dagger M$ is the projector onto the space orthogonal to the nullspace of M, that is, the projector onto the range of $M^\top$. Thus,

$$A_{2s} \cap [\mathrm{range}(M^\top)]^\perp = \{0\}. \tag{5.20}$$

Here, we have used the fact that the projection of a point x onto a subspace Ω is zero if and only if $x \in \Omega^\perp$. Now using the representation for A_{2s} given by (3.12), we have that (5.20) is equivalent to

$$A_J \cap \ker(M^\top) = \{0\} \quad \text{for all } J \in \mathcal{J}_{2s}. \tag{5.21}$$

But, by (2.34), this is equivalent to the strong regularity of $(A_J^\perp, \mathrm{range}(M^\top))$ for all $J \in \mathcal{J}_{2s}$. $\qquad\square$

Next, the projected gradient algorithm is defined. The goal is to reformulate the method of alternating projections in terms of the projected gradient to apply known results on the projected gradient to AP.

Definition 5.2.8 (projected gradients). *Given a closed set $A \subset \mathbb{R}^n$, a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, and a positive real number τ, the mapping*

$$T_{PG}(x; \tau) = P_A\left(x - \frac{1}{\tau}\nabla f(x)\right) \tag{5.22}$$

is called the projected gradient operator. The projected gradients algorithm is the fixed point iteration

$$x^{k+1} \in T_{PG}(x^k; \tau_k) = P_A\left(x^k - \frac{1}{\tau_k}\nabla f(x^k)\right), \ k \in \mathbb{N},$$

for x^0 given arbitrarily and a sequence of positive real numbers $(\tau_k)_{k \in \mathbb{N}}$.

In the context of linear least squares with a sparsity constraint, the projected gradient algorithm is equivalent to what is also known as the iterative hard thresholding algorithm (see, for instance, (Blumensath and Davies, 2009), (Blumensath and Davies, 2010), and (Kyrillidis and Cevher, 2014)), where the constraint $A = A_s$ and the projector given by (3.16) amount to a thresholding operation on the largest elements of the iterate.

With these definitions, we cite a result on convergence of the projected gradient algorithm applied to (5.18).

Theorem 5.2.9 (global convergence of projected gradients/iterative hard thresholding (Blumensath and Davies, 2010, Theorem 4) and (Beck and Teboulle, 2011, Theorem 3 and Corollary 1)). *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous and subadditive function satisfying $\inf_{x \in \mathbb{R}^n} \varphi(x) \neq -\infty$. Further, let $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}^n$. Let M satisfy (5.17) of order $(2s, 2)$. Further, for any given initial point x^0, let the sequence $\{x^k\}_{k \in \mathbb{N}}$ be generated by the projected gradient algorithm with $A = \text{lev}_{\leq s} \varphi$, $f(x) = \frac{1}{2} \|Mx - p\|_2^2$, and the constant step size $\tau \in [\mu_{2s}, 2\nu_{2s})$. Then the iterates converge to the unique global solution of (5.18), and $f(x^k) \to 0$ linearly as $k \to \infty$ with rate $\rho = \left(\frac{\tau}{\nu_{2s}} - 1 \right) < 1$, that is,*

$$f(x^{k+1}) \leq \rho f(x^k) \quad \text{for all } k \in \mathbb{N}. \tag{5.23}$$

Next, we specialize this theorem to alternating projections.

Corollary 5.2.10 (alternating projections in terms of projected gradients (Hesse et al., 2014, Corollary III.13)). *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous and subadditive function satisfying $\inf_{x \in \mathbb{R}^n} \varphi(x) \neq -\infty$. Further, let $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}^n$. Let the matrix M satisfy (5.17) of order $(2s, 2)$ with $\mu_{2s} = 1$ and $MM^\top = \text{Id}$. Then $\text{lev}_{\leq s} \varphi \cap B = \{\bar{x}\}$, i.e., the intersection is a singleton, and for any initial point x^0, the alternating projections sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by (4.19) applied to (3.8) converges to $\text{lev}_{\leq s} \varphi \cap B$. The values of $f(x^k) = \frac{1}{2} \|Mx^k - p\|_2^2$ converge to zero with linear rate $\rho = \left(\frac{1}{\nu_{2s}} - 1 \right) < 1$.*

Proof. For $f(x) = \frac{1}{2} \|Mx - p\|_2^2$, we have $\nabla f(x) = M^\top (Mx - p)$. The projected gradients iteration with constant step length $\tau = 1$ then takes the form

$$x^{k+1} \in P_{\text{lev}_{\leq s} \varphi} \left(x^k - \nabla f(x^k) \right) = P_{\text{lev}_{\leq s} \varphi} \left(x^k - M^\top (Mx^k - p) \right). \tag{5.24}$$

The projection onto the subspace B is given by (see (3.16))

$$P_B x = \left(\text{Id} - M^\top (MM^\top)^{-1} M \right) x + M^\top (MM^\top)^{-1} p. \tag{5.25}$$

Since $MM^\top = \text{Id}$, this simplifies to $x^k - M^\top (Mx^k - p) = P_B x^k$. Hence,

$$x^{k+1} \in P_{\text{lev}_{\leq s} \varphi} \left(x^k - \nabla f(x^k) \right) = P_{\text{lev}_{\leq s} \varphi} P_B x^k. \tag{5.26}$$

This shows that projected gradients 5.2.8 with unit step length applied to (5.18) with $A = \text{lev}_{\leq s} \varphi$ and $f(x) = \frac{1}{2} \|Mx - p\|_2^2$ is equivalent to the method of alternating projections 4.2.1 applied to (3.8).

To show convergence to a unique solution, we apply Theorem 5.2.9, for which we must show that the step length $\tau = 1$ lies in the nonempty interval $[\mu_{2s}, 2\nu_{2s})$. By assumption, M satisfies (5.17) of order $(2s, 2)$ with $\mu_{2s} = 1$. Hence, $\frac{1}{2} < \nu_{2s} \leq 1$, and $\tau = 1$ lies in the nonempty interval $[1, 2\nu_{2s})$. The assumptions of Theorem 5.2.9 are thus satisfied with $\tau = 1$, whence global convergence to the unique solution of (5.18), and consequently (3.8), immediately follows. $\square$

Remark 5.2.11. *Similarly to Corollary 5.2.4, the result in Corollary 5.2.10 can be adapted to sparse affine feasibility since the ℓ_0-function is lower semicontinuous, subadditive, and it satisfies $\ell_0(x) = \ell_0(-x)$ for all $x \in \mathbb{R}^n$.*

5.3 Restricted Isometry Constants and Dimensions

The next proposition will show the limits of Corollary 5.2.4. Namely, the restricted isometry property on $M^\dagger M$ in (5.13) gives explicit bounds for the dimension m in terms of n.

Proposition 5.3.1 ((Krahmer, 2014)). *Let $M : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping of full rank satisfying the inequalities*

$$(1 - \delta) \|x\|_2^2 \leq \left\| M^\top (MM^\top)^{-1} Mx \right\|_2^2 \leq \|x\|_2^2 \quad \forall x \in A_s \tag{5.27}$$

for some $\delta < \frac{1}{2}$. Then $m > \frac{n}{2}$.

Proof. If we look at the squared Frobenius norm of $M^\top (MM^\top)^{-1} M$, then we obtain

$$\left\| M^\top (MM^\top)^{-1} M \right\|_F^2 = \mathrm{Tr} \left(M^\top (MM^\top)^{-1} MM^\top (MM^\top)^{-1} M \right), \tag{5.28}$$

which is, by shifting the rightmost M to the left side, equal to $\mathrm{Tr}(\mathrm{Id}_m) = m$. Further, if we denote the jth column of $M^\top (MM^\top)^{-1} M$ by $M^\dagger M_j$, we have

$$m = \sum_{j=1}^{n} \left\| M^\dagger M_j \right\|_2^2 = \sum_{j=1}^{n} \left\| M^\dagger M e_j \right\|_2^2. \tag{5.29}$$

With (5.27), we have, since $e_j \in A_s$ for all j,

$$\sum_{j=1}^{n} \left\| M^\dagger M e_j \right\|_2^2 \geq \sum_{j=1}^{n} (1 - \delta) \|e_1\|_2^2 = (1 - \delta)n. \tag{5.30}$$

The result follows from $\delta < \frac{1}{2}$, which implies $m > \frac{n}{2}$. $\qquad\square$

Remark 5.3.2. *Speaking in terms of applications, the result in Proposition 5.3.1 states that, without any further assumptions on M and in order to achieve convergence in Corollary 5.2.4, we need at least more than $\frac{n}{2}$ measurements. This is not in line with the idea of compressed sensing, where the number ratio $\frac{m}{n}$ can be viewed as a compression level of data.*

The remaining part of this section will establish, with additional assumptions on M, a link between (5.13) and (5.17). Suppose that we have a matrix $M \in \mathbb{R}^{m \times n}$ with orthonormal rows, i.e., $MM^\top = \mathrm{Id}_m$. Let M satisfy a scaled restricted isometry property of order (s, α), as in (5.17) with $\varphi = \ell_0$. Then, we have positive constants $\nu_s, \mu_s > 0$

satisfying $1 \leq \frac{\mu_s}{\nu_s} < \alpha$. Now, let $c > 0$ and $\delta_s > 0$ be such that $\nu_s = \frac{1-\delta_s}{c}$ and $\mu_s = \frac{1-\delta_s}{c}$. Then (5.17) can be written as

$$\frac{1-\delta_s}{c} \|x\|_2^2 \leq \|Mx\|_2^2 \leq \frac{1+\delta_s}{c} \|x\|_2^2 \quad \text{for all } x \in A_s. \tag{5.31}$$

We present two lemmata whose proofs are an immediate generalization of results in (Needell and Tropp, 2009).

Definition 5.3.3. *Let $J \subset \{1, \ldots, n\}$ be an index set. Denote by M_J the matrix M restricted to the columns indexed by J.*

Lemma 5.3.4 ((Needell and Tropp, 2009, Proposition 3.1)). *Suppose M has scaled restricted isometry constants $\frac{1-\delta_s}{c}, \frac{1+\delta_s}{c}$. Let $I \in \mathcal{J}_t$ (see (3.11)) for $t \leq s$. Then the following inequalities hold for all $x \in \mathbb{R}^t$.*

1.

$$\frac{1-\delta_s}{c} \|x\|_2 \leq \left\| M_I^\top M_I x \right\|_2 \leq \frac{1+\delta_s}{c} \|x\|_2. \tag{5.32}$$

2.

$$\sqrt{\frac{1-\delta_s}{c}} \|x\|_2 \leq \left\| M_I^\top x \right\|_2 \leq \sqrt{\frac{1+\delta_s}{c}} \|x\|_2. \tag{5.33}$$

Proof. This proof is adapted from (Needell and Tropp, 2009, Proposition 3.1). Because of the equation (5.31), we know that the singular values of M_I are between $\sqrt{\frac{1-\delta_s}{c}}$ and $\sqrt{\frac{1+\delta_s}{c}}$. Then the singular values of $M_I^\top$ are the same as those of M_I. Further, the singular values of $M_I^\top M_I$ are the squares of those of M_I. This completes the proof. $\square$

Lemma 5.3.5 ((Needell and Tropp, 2009, Proposition 3.2)). *Suppose M has scaled restricted isometry constants $\frac{1-\delta_s}{c}, \frac{1+\delta_s}{c}$. Let S and T be disjoint sets of indices whose combined cardinality does not exceed s. Then we have*

$$\left\| M_S^\top M_T \right\| \leq \frac{\delta_s}{c}. \tag{5.34}$$

Proof. The proof is similar to the one in (Needell and Tropp, 2009). Write $R = S \cup T$. The matrix $M_S^\top M_T$ is a submatrix of $M_R^\top M_R - \frac{1}{c} \mathrm{Id}$. Since the spectral norm of a submatrix is bounded from above by the spectral norm of the matrix, and since the eigenvalues of $M_R^\top M_R$ lie between $\frac{1-\delta_s}{c}$ and $\frac{1+\delta_s}{c}$, we obtain

$$\left\| M_S^\top M_T \right\| \leq \left\| M_R^\top M_R - \tfrac{1}{c} \mathrm{Id} \right\| \leq \max \left\{ \frac{1+\delta_s}{c} - \frac{1}{c}, \frac{1}{c} - \frac{1-\delta_s}{c} \right\} = \frac{\delta}{c}. \tag{5.35}$$

$\square$

Let us combine Lemma 5.3.4 and Lemma 5.3.5. They allow us to give bounds of the RIP-constants of $M^\top M$ depending on those of M.

Proposition 5.3.6. *If a matrix M with $MM^\top = \mathrm{Id}_m$ satisfies (5.31) of order $2s$, then the matrix $M^\top M$ satisfies*

$$\left\| M^\top M x \right\|_2 \leq \frac{1 + \frac{n}{s}\delta_{2s}}{c} \tag{5.36}$$

for all s-sparse vectors x of unit length.

Proof. Let $x \in A_s$ and let $I := I(x)$ as in (3.10). Choose a partition $\{I, T_1, \ldots, T_k\}$ of $\{1, \ldots, n\}$, where $k \leq \frac{n}{s} - 1$, and where each of the T_j satisfies $|T_j| \leq s$. With this notation we get

$$
\begin{aligned}
\left\| M^\top M x \right\|_2 \quad &= \quad \left\| M^\top M_I x \right\|_2 \\
&= \quad \left\| M_I^\top M_I x + \sum_{j=1}^k M_{T_j}^\top M_I x \right\|_2 \\
&\leq \quad \left\| M_I^\top M_I x \right\|_2 + \sum_{j=1}^k \left\| M_{T_j}^\top M_I x \right\|_2 \\
\overset{\text{Prop. 5.3.5}}{\leq} \quad &\quad \left\| M_I^\top M_I x \right\|_2 + \sum_{j=1}^k \frac{\delta_{2s}}{c} \\
\overset{\text{Prop. 5.3.4}}{\leq} \quad &\quad \frac{1 + \delta_{2s}}{c} + k\frac{\delta_{2s}}{c} \\
&\leq \quad \frac{1 + \delta_{2s} + (\frac{n}{s} - 1)\delta_{2s}}{c} \\
&= \quad \frac{1 + \frac{n}{s}\delta_{2s}}{c}.
\end{aligned}
\tag{5.37}
$$

$\square$

5.4 Examples

In this short section, we give an example of a matrix satisfying the assumptions of Corollary 5.2.10, on the one hand. On the other hand, we name a lowdimensional example for a matrix M and a vector p such that there exist initial points for which alternating projections and Douglas-Rachford do not converge to the intersection $A_s \cap B$. This section can, with slight alterations, also be found in (Hesse et al., 2014, Section V.B).

5.4.1 Example of a matrix satisfying assumptions of Corollary 5.2.10

This example is first and foremost a confirmation that the assumptions on the restricted isometry property are accomplishable.

We take the matrix

$$
M = \frac{1}{\sqrt{8}}
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1
\end{pmatrix}.
\tag{5.38}
$$

The rows of M are pairwise orthogonal, which yields $MM^\top = \text{Id}_7$. We compute the constant δ in (5.16) for $s = 2$ to get a result for the recovery of 1-sparse vectors with alternating projections. Recall that s can be larger than the sparsest feasible solution (see Remark 5.1.2). In general, a normalized 2-sparse vector in $\mathbb{R}^8$ has the form

$$x = (\cos(\alpha), \sin(\alpha), 0, 0, 0, 0, 0, 0),$$

where the position of the sine and of the cosine in x are of course arbitrary. The squared norm of the product Mx is equal to

$$\|Mx\|_2^2 = \frac{1}{8} \sum_{j=1}^{7} |\cos(\alpha) + z_j \sin(\alpha)|^2,$$

where $z_j \in \{-1, 1\}$. We get

$$\frac{1}{8} \sum_{j=1}^{7} |\cos(\alpha) + z_j \sin(\alpha)|^2 = \frac{1}{8} \sum_{j=1}^{7} \cos(\alpha)^2 + 2z_j \sin(\alpha)\cos(\alpha) + \sin(\alpha)^2$$

$$= \frac{1}{8}(7 \pm \sin(2\alpha)) \in \left[\frac{3}{4}, 1\right].$$

This means that

$$\frac{3}{4}\|x\|_2^2 \le \|Mx\|_2^2 \le \|x\|_2^2 \quad \forall x \in A_2,$$

where A_2 is the set of 2-sparse vectors in $\mathbb{R}^8$. In other words, we can recover any 1-sparse vector with the method of alternating projections. We refer to Chapter 11 to show the performance of different algorithms applied to (3.8) within the setup presented here.

5.4.2 Counterexamples

The following example, discovered with help from MATLAB©'s Symbolic Math Toolbox™, shows some of the more interesting pathologies that one can see with these algorithms when not starting sufficiently close to a solution.

Let $n = 3, m = 2, s = 1$, and

$$M = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -1 \end{pmatrix}, \qquad p = \begin{pmatrix} -5 \\ 5 \end{pmatrix}. \tag{5.39}$$

The point $(0, 10, 0)^\top$ is the sparsest solution to the equation $Mx = p$, and the affine space B is

$$B = \begin{pmatrix} 0 \\ 10 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \qquad \text{with } \lambda \in \mathbb{R}.$$

If we take the initial point

$$x^0 = \begin{pmatrix} \frac{38894857328700073}{23768448754279301278063185 1008} \\ \frac{-29710560942850721475845458 0565}{11884224377139650639031592 5504} \\ \frac{-1188422437713940163629828887893}{23768448754279301278063185 1008} \end{pmatrix} \approx \begin{pmatrix} 0 \\ -\frac{5}{2} \\ -5 \end{pmatrix},$$

then $T_{DR} x^0 = x^0 + (-5, 0, 5)^\top$ and $T_{DR}^2 x^0 = x^0$.

Note that this example is different from the case in Theorem 5.5.1: in Theorem 5.5.1, we establish that if $s < \operatorname{rank}(M)$, then the fixed point set of T_{DR} is strictly larger than the solution set to problem (3.8). The specific case detailed here also satisfies $s < \operatorname{rank}(M)$. However, with the given x^0, we are not near the set of fixed points but in a *cycle* of T_{DR}.

If, on the other hand, we take the point $\hat{x}^0 = (-4, 0, 0)^\top$, then $P_B \hat{x}^0 = (-4, 2, -4)^\top$, and the set $P_{A_1} P_B \hat{x}^0$ is equal to $\{(-4, 0, 0)^\top, (0, 0, -4)^\top\}$. By projecting back onto the affine subspace, $P_B (0, 0, -4)^\top$ is again equal to $(-4, 2, -4)^\top$. This shows that the alternating projection iteration (4.18) is stuck at the points $(-4, 0, 0)^\top$ and $(0, 0, -4)^\top$, which are clearly not elements in the intersection $A_1 \cap B = \{(0, 10, 0)^\top\}$. This also highlights a manifestation of the multivaluedness of the projector P_{A_1}. In Figure 5.2, we show a visualization of this counterexample.
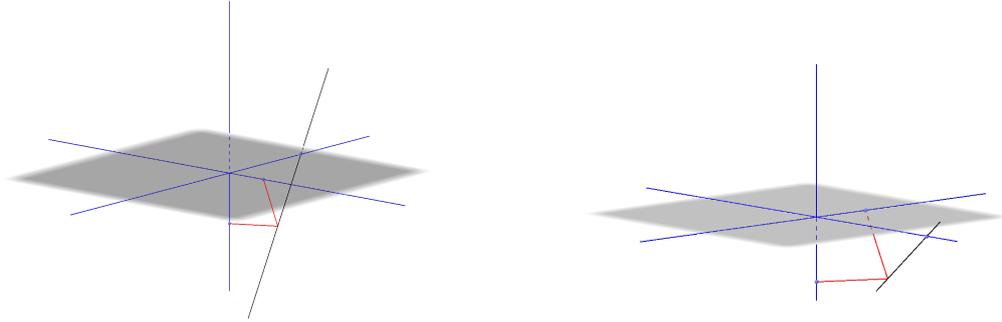
Figure 5.2: A visualization of the counterexample in (5.39): the black line represents the affine subspace B with an intersection with the set of sparse vectors, given by the blue lines. The red lines are connecting best approximation pairs between B and A_1.

5.5 Douglas-Rachford in Sparse-Affine Feasibility

To close this chapter projection methods and sparsity, we give two local results on Douglas-Rachford applied to sparse-affine feasibility. This section is based on (Hesse et al., 2014, Chapter IV).

Theorem 5.5.1 ((Hesse et al., 2014, Theorem IV.1))**.** *Let A_s and B be defined by (3.5) and (3.6), respectively, and suppose there exists a point $\bar{x} \in A_s \cap B$ with $\ell_0(\bar{x}) = s$. If $s <$*

$\mathrm{rank}(M)$. *Then, on all open neighborhoods $\mathcal{N}$ of $\bar{x} \in A_s \cap B$, there exist fixed points $z \in$ Fix T_{DR} with $z \notin A_s \cap B$.*

Proof. Let $\bar{x} \in A_s \cap B$ with $\ell_0(\bar{x}) = s$, and set $\delta < \min\{|\bar{x}_j| \mid \bar{x}_j \neq 0\}$. By Lemma 3.2.4, we have $A_s \cap B \cap \mathbb{B}_{\delta/2}(\bar{x}) = A_J \cap B \cap \mathbb{B}_{\delta/2}(\bar{x})$ for a unique $J := I(\bar{x}) \in \mathcal{J}_s$. Thus, on the neighborhood $\mathbb{B}_{\delta/2}(\bar{x})$, the feasibility problems

$$\text{find } x \in A_J \cap B \quad \text{and} \quad \text{find } x \in A_s \cap B$$

have the same set of solutions. We consider the Douglas-Rachford operators applied to these two feasibility problems, for which we introduce the following notation:

$$T_J := \tfrac{1}{2}\left(R_{A_J} R_B + \mathrm{Id}\right) \quad \text{and} \quad T_s := \tfrac{1}{2}\left(R_{A_s} R_B + \mathrm{Id}\right).$$

Our proof strategy is to show first that the operators T_J and T_s restricted to $\mathbb{B}_{\delta/2}(\bar{x})$ are identical. Hence, their fixed point sets intersected with $\mathbb{B}_{\delta/2}(\bar{x})$ are identical. We then show that under the assumption $s < \mathrm{rank}\,(M)$ the set Fix T_J is strictly larger than the intersection $A_J \cap B$, hence completing the proof.

To show that the operators T_J and T_s applied to points $x \in \mathbb{B}_{\delta/2}(\bar{x})$ are identical, note that, by Lemma 3.2.3 (ii) and (iv), we have $P_{A_s}(x) \subset \mathbb{B}_{\delta/2}(\bar{x})$ and $R_{A_s}(x) \subset \mathbb{B}_{\delta/2}(\bar{x})$ for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$. Moreover, by Lemma 3.2.4, we have $A_s \cap \mathbb{B}_{\delta}(\bar{x}) = A_J \cap \mathbb{B}_{\delta}(\bar{x})$ since $\ell_0(\bar{x}) = s$. Thus, for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$, we have $P_{A_s}(x) = P_{A_J}(x) \in \mathbb{B}_{\delta/2}(\bar{x})$ and $R_{A_s}(x) = R_{A_J}(x) \in \mathbb{B}_{\delta/2}(\bar{x})$. Also by Lemma 3.2.3, $R_B x \in \mathbb{B}_{\delta/2}(\bar{x})$ for $x \in \mathbb{B}_{\delta/2}(\bar{x})$. Altogether, this yields

$$T_s x = \tfrac{1}{2}\left(R_{A_s} R_B + \mathrm{Id}\right) x = \tfrac{1}{2}\left(R_{A_J} R_B + \mathrm{Id}\right) x = T_J x \in \mathbb{B}_{\delta/2}(\bar{x}) \tag{5.40}$$

for all $x \in \mathbb{B}_{\delta/2}(\bar{x})$. Hence, the operators T_s and T_J and their fixed point sets coincide on $\mathbb{B}_{\delta/2}(\bar{x})$.

Next, we derive an explicit characterization of Fix T_J. By (Bauschke et al., 2004, Corollary 3.9) and (2.34), we have:

$$\begin{aligned}
\text{Fix } T_J &= (A_J \cap B) + N_{A_J - B}(0) \\
&= (A_J \cap B) + \left(N_{A_J}(\bar{x}) \cap -N_B(\bar{x})\right) \\
&= (A_J \cap B) + \left(A_J^{\perp} \cap B^{\perp}\right).
\end{aligned} \tag{5.41}$$

The following equivalences show that $A_J^{\perp} \cap B^{\perp}$ is nontrivial if $s < \mathrm{rank}\,(M)$. Indeed,

$$\begin{aligned}
& & \mathrm{rank}(M) & > s \\
&\Leftrightarrow & \dim(\ker(M)^{\perp}) & > s \\
&\Leftrightarrow & n - s + \dim(\ker(M)^{\perp}) & > n \\
&\Leftrightarrow & \dim(A_J^{\perp}) + \dim(\ker(M)^{\perp}) & > n \\
&\Leftrightarrow & A_J^{\perp} \cap B^{\perp} & \neq \{0\}.
\end{aligned} \tag{5.42}$$

In other words, Fix T_J contains elements from the intersection $A_J \cap B$ and the nontrivial subspace $A_J^\perp \cap B^\perp$. This completes the proof. $\qquad\square$

Remark 5.5.2 ((Hesse et al., 2014, Remark IV.2)). *Inequality (5.42) shows that if* $\mathrm{rank}(M) > s$*, then the intersection $A_J \cap B$ is* not *strongly regular, or in other words, if $A_J \cap B$ is strongly regular, then* $\mathrm{rank}(M) \leq s$*. This was also observed in (Bauschke et al., 2014b, Remark 3.17) using tangent cones and* transversality*. The simple meaning of these results is that if the sparsity of a feasible point is less than the rank of the measurement matrix (the only interesting case in sparse signal recovery), then, since locally the affine feasibility problem is indistinguishable from simple linear feasibility at points $\bar{x} \in A_s$ with $\ell_0(\bar{x}) = s$, by Lemma 4.3.2, the Douglas-Rachford algorithm may fail to converge to the intersection on all balls around a feasible point. As we noted in the introduction of the Douglas-Rachford algorithm, however, it is not the fixed points of Douglas-Rachford themselves but rather their shadows that are of interest. This leads to positive convergence results, as shown in the next theorem.*

Theorem 5.5.3 ((Hesse et al., 2014, Theorem IV.7)). *Let A_s and B be defined by (3.5) and (3.6) with nonempty intersection and let $\bar{x} \in A_s \cap B$ 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 iterates converge with linear rate to $\mathrm{Fix}\, T_{DR}$. Moreover, for any $\hat{x} \in \mathrm{Fix}\, T_{DR} \cap \mathbb{B}_{\delta/2}(\bar{x})$, we have $P_B \hat{x} \in A_s \cap B$.*

Proof. By Lemma 3.2.4, we have $A_s \cap B \cap \mathbb{B}_\delta(\bar{x}) = A_J \cap B \cap \mathbb{B}_\delta(\bar{x})$ for a unique $J \in \mathcal{J}_s$. Thus, by (5.40), at all points in $\mathbb{B}_{\delta/2}(\bar{x})$, the Douglas-Rachford operator corresponding to A_s and B is equivalent to the Douglas-Rachford operator corresponding to A_J and B, whose intersection includes $\bar{x}$. Applying Theorem 4.3.10, shifting the subspaces appropriately, we see that the iterates converge to some point $\hat{x} \in \mathrm{Fix}\, T_{DR}$ with linear rate for all initial points $x^0 \in \mathbb{B}_{\delta/2}(\bar{x})$. The last statement follows from (5.40) and Theorem 4.3.10. $\qquad\square$

6 Angles, Polyhedral Sets, and Sparsity

As an interlude between the two main parts of this thesis, we discuss the connection of angles to the sparse-affine feasibility problem (3.8). The relation between convergence of the alternating projections and the angle between two subspaces can, for example, be found in (Deutsch, 2001, Theorem 9.8).

First, we name the different notions of angles between subspaces. Each of these notions has advantages and disadvantages regarding the description of geometric properties of the given sets, as will be pointed out. Then we study the behavior of alternating projections between the set of sparse vectors A_s (3.5) and a polyhedral set. Finally, we will give a necessary condition for alternating projections applied to (3.8) to converge for an arbitrary initial point.

6.1 Angles

Let U, V be two linear subspaces of $\mathbb{R}^n$ of dimensions $m_U \leq m_V \leq n$.

Definition 6.1.1 (Dixmier angle (Dixmier, 1949, Section 4))**.** *The Dixmier angle between U and V is the number in $[0, \frac{\pi}{2}]$ whose cosine is given by*

$$c_0(U, V) := \sup \left\{ \langle u, v \rangle \mid u \in U, v \in V, \|u\| = \|v\| = 1 \right\}. \tag{6.1}$$

Definition 6.1.2 (Friedrichs angle (Friedrichs, 1937, Equation 1.1))**.** *The Friedrichs angle between U and V is the number in $[0, \frac{\pi}{2}]$ whose cosine is given by*

$$c(U, V) := \sup \left\{ \langle u, v \rangle \; \middle| \; \begin{array}{l} u \in U \cap (U \cap V)^{\perp}, \\ v \in V \cap (U \cap V)^{\perp}, \|u\| = \|v\| = 1 \end{array} \right\}. \tag{6.2}$$

The Friedrichs angle has a crucial weakness: it cannot detect if subspaces may be subsets of each other. For example, consider for U the x-axis in $\mathbb{R}^3$ and for V the x-y-plane as a subset of $\mathbb{R}^3$. A well-defined angle between these sets would be of course zero, while the value of $c(U, V)$ will be zero due to the orthogonality conditions. This yields an angle of $\pm \frac{\pi}{2}$. This fact will also be formalized in Lemma 6.2.6. Note that by the notation $\langle \cdot, \cdot \rangle$ we always mean the Euclidean inner product.

Definition 6.1.3 (Principal angles (Jordan, 1875, Chapter IV, Eq. 60, pp. 122–130))**.** *For $k = 1, \ldots, m_U$, let u_k and v_k be the maximizing arguments of*

$$\max_{\substack{u \in U, v \in V \\ \|u\| = \|v\| = 1}} u^{\top} v \tag{6.3}$$

61

satisfying the equations $u_j^\top u = v_k^\top v = 0$ for all $1 \le j \le k - 1$ and for all $k \le m_U$, if u_j exists. Let the tuple $(\theta_1, \ldots, \theta_{m_U}) \in [0, \frac{\pi}{2}]^{m_U}$ be such that

$$\cos(\theta_k) = c_k := u_k^\top v_k \tag{6.4}$$

for all k. Then the tuple $(\theta_1, \ldots, \theta_{m_U})$ will be called Jordan's principal angles.

We give a relation between these three kinds of angles in the following proposition.

Proposition 6.1.4. *Let $U, V \subset \mathbb{R}^n$ be linear subspaces of dimensions $m_U \le m_V \le n$. Let $\theta_1, \ldots, \theta_{m_U}$ be the principal angles between U and V. Let $\nu = \dim(U \cap V)$. Then the following equations hold.*

i) $c_0(U, V) = \cos(\theta_1)$;

ii) $c(U, V) = \cos(\theta_{\nu+1})$.

Proof. The first principal angle θ_1 is, by definition, the number such that its cosine is the maximizing argument of

$$\max_{\substack{u \in U, v \in V \\ \|u\| = \|v\| = 1}} u^\top v$$

without any further restrictions. This is equivalent to the definition of the Dixmier angle and, thus, shows claim (i).

For claim (ii), we observe that the first ν principal angles can be obtained via taking an orthonormal basis $v_1, \ldots, v_\nu$ of $U \cap V$. Then, for $1 \le j \le \nu$, we get $c_j = v_j^\top v_j = 1$. Having Equation (6.3) in mind, the maximizing arguments $u_{\nu+1}, v_{\nu+1}$ in (6.4) for $k = \nu$ satisfy

$$u_{\nu+1} \in U \cap (U \cap V)^\perp \quad \text{and} \quad v_{\nu+1} \in V \cap (U \cap V)^\perp.$$

Hence, $u_{\nu+1}^\top v_{\nu+1} = c(U, V)$. $\square$

Remark 6.1.5. *We do not claim any novelty in Proposition 6.1.4. However, we think that it is important to point out that there are three different definitions of angles between subspaces that can be expressed in terms of the most general and earliest one, namely, Jordan's principle angles. Still, for convenience, we remain at the notion of the Friedrichs angle instead of writing "the $(\nu + 1)$st principle angle where $\nu = \dim(U \cap V)$".*

6.2 The Geometry of Polyhedral Sets

We turn our attention to the case where one seeks a sparsest vector not necessarily in an affine subspace B of $\mathbb{R}^n$ but more generally in a polyhedral set $\mathcal{C} \subset \mathbb{R}^n$. The main question in the beginning is the question of the fixed points of the alternating projections algorithm in this setting.

Given a matrix $M \in \mathbb{R}^{m \times n}$ and an arbitrary vector $p \in \mathbb{R}^m$, we define the set

$$\mathcal{C} := \{x \in \mathbb{R}^n \mid Mx \ge p\}. \tag{6.5}$$

The notation " $\geq$ " is to be read entrywise, i.e., if we denote by $m_j, j \in \{1, \ldots, m\}$, the lines of M, then $Mx \geq p$ means $m_j x \geq p_j$ for all $j \in \{1, \ldots, m\}$.

It is well known that polyhedral sets as defined in (6.5) are convex and closed.

Lemma 6.2.1. *Let C be defined as in (6.5). Then C is convex and closed.*

Proof. Let $x, y \in C$, and let $\lambda \in [0, 1]$ be arbitrary. Then

$$M(\lambda x + (1-\lambda)y) = \lambda Mx + (1-\lambda)My \geq \lambda p + (1-\lambda)p = p. \tag{6.6}$$

This proves convexity of C. The set C is closed because it is an intersection of m lower level sets of the shape $\{x \in \mathbb{R}^n \mid \langle -M_j, x \rangle \leq -p\}$. Because linear mappings are continuous, they are especially lower semicontinuous. This means that the lower level sets are closed sets. $\qquad\square$

The following definition formalizes what has been written in the introductory text of this chapter.

Definition 6.2.2. *Given a sparsity parameter s that satisfies $1 \leq s \leq \min\{m, n\}$, we define the sparse-polyhedral feasibility problem*

$$\text{find } \bar{x} \in A_s \cap C, \tag{6.7}$$

where C is defined via (6.5), and where A_s is defined via (3.5).

Lemma 6.2.3 ((Bauschke and Borwein, 1993)). *Let $V_1, V_2 \subset \mathbb{R}^n$ be affine subspaces. Let $g = P_{V_1 - V_2} 0$ be the gap-vector between V_1 and V_2. Then the intersection $V_1 \cap (V_2 - g)$ is linearly regular.*

Lemma 6.2.4 (alternating projections between parallel subspaces). *Let V_1 and V_2 be affine subspaces with gap vector g and such that $(V_1 - g) \subset V_2$. Then any alternating projections sequence $\{x^k\}_{k=0}^\infty$ with $x^0 \in \mathbb{R}^n$ arbitrary satisfies $x^k = x^{k+1}$ for all $k \geq 1$.*

Proof. Let $x^0 \in \mathbb{R}^n$ be arbitrary. Without loss of generality, assume that the alternatings projection sequence reads as $x^{k+1} = P_{V_2} P_{V_1} x^k$. Then $P_{V_1} x^0 \in V_1$, and we know that $P_{V_1} x^0 - g \in V_2$. Since the gap vector g is the shortest connection between V_1 and V_2, we have

$$P_{V_2} P_{V_1} x^0 = P_{V_1} x^0 - g. \tag{6.8}$$

By the same argument, i.e., the gap vector is the shortest conection between V_1 and V_2, we know that $P_{V_1} x^1 = x^1 + g = P_{V_1} x^0$. But then we are at a fixed point of the alternating projections operator by Theorem 4.2.2. $\qquad\square$

Corollary 6.2.5 (alternating projections between a subspace and a parallel face). *Let C be defined as in (6.5). Let F be a face of C, and let V be an affine subspace of $\mathbb{R}^n$. Denote by g the gap vector between C and V. If F is such that $(F - g) \subset V$, then every point $x \in F$ is a fixed point of the operator $P_C P_V$.*

Proof. Take any $x^0 \in F$. Because $(F - g) \subset V$, we know that $(\text{aff}(F) - g) \subset V$. The point x^0 is also contained in $\text{aff}(F)$. By Lemma 6.2.4, the point x^0 is a fixed point of alternating projections between F and V. Because $\mathcal{C}$ is convex, the projection onto $\mathcal{C}$ is single-valued, and because g is the gap, $P_V x^0 = x^0 - g$ and $P_{\mathcal{C}} P_V x^0 = x^0$. □

Lemma 6.2.6. *Let $\mathcal{C}$ be defined as in (6.5) and let V be an affine subspace. Let g denote the gap vector between $\mathcal{C}$ and V, and choose $p \in \mathcal{C}$ and $v \in V$ such that $p - v = g$. Let $F \subset \mathcal{C}$ be a face with $p \in F$. Define $B := \text{aff}(F)$. Then the cosine of the Friedrichs angle between F and V is 0 if and only if one the following cases occurs.*

1. $V \subset (B - g)$;

2. $(B - g) \subset V$;

3. $\big((B - g) \cap (V \cap (B - g))^{\perp}\big) \perp \big(V \cap (V \cap (B - g))^{\perp}\big).$

Proof. Recall the definition of the Friedrichs angle in 6.1.2 between the two sets V and $B - g$, i.e.,

$$c(V, (B - g)) := \sup \left\{ \langle u, v \rangle \;\middle|\; \begin{array}{c} u \in (B - g) \cap ((B - g) \cap V)^{\perp}, \\ v \in V \cap ((B - g) \cap V)^{\perp}, \|u\| = \|v\| = 1 \end{array} \right\}. \tag{6.9}$$

The claim now follows immediately from the definition. Suppose one of the sets is contained in the other one. Without loss of generality, let $(B - g) \subset V$. Then $(B - g) \cap ((B - g) \cap V)^{\perp} = \{0\}$, and the supremum over all u, v, giving us the value of $c(V, (B - g))$, is zero. If we have

$$\big((B - g) \cap (V \cap (B - g))^{\perp}\big) \perp \big(V \cap (V \cap (B - g))^{\perp}\big),$$

then naturally the value of $c(V, (B - g))$ is zero. □

Remark 6.2.7. *Lemma 6.2.6 characterizes the capability of the Friedrichs angle to detect orthogonality between subspaces. For the method of alternating projections and in fact for the Douglas-Rachford, it doesn't matter which of the three cases in Lemma 6.2.6 is at hand if we have a nonempty intersection $\mathcal{C} \cap V$. Since the Friedrichs angle is assumed to be π, both methods find the intersection after one step. If we have an empty intersection, i.e., $F \cap V = \emptyset$ with F and V as in Lemma 6.2.6, then the method of alternating projections still finds the correct point which is closest to the other set immediately. For Douglas-Rachford, on the other hand, the iteration will eventually diverge since we have an infeasible case of two closed and convex sets (see Theorem 4.3.3).*

Proposition 6.2.8. *Let $\mathcal{C}$ be defined as in (6.5), and let A_s be defined via (3.5). Write A_s as a union of linear subspaces as in Equation 3.12, i.e.,*

$$A_s = \bigcup_{J \in \mathcal{J}} A_J. \tag{6.10}$$

Define for every $J \in \mathcal{J}$ the gap vector

$$g_J := P_{\mathcal{C} - A_J}(0). \tag{6.11}$$

Then, for every $J \in \mathcal{J}$, the intersection $(\mathcal{C} - g_J) \cap A_J$ is linearly regular. Further, for every $J \in \mathcal{J}$, the intersection $(\mathcal{C} - g_J) \cap A_s$ is locally linearly regular.

Proof. Choose an arbitrary $J \in \mathcal{J}$ and an arbitrary $x \in \mathbb{R}^n$. Then, at the intersection $(\mathcal{C} - g_J) \cap A_s$, we have the intersection of finitely many faces of a convex polyhedron. Each of these faces is locally an affine subspace. Hence, their intersection with a linear subspace is linearly regular. Further, the intersection with a union of finitely many linear subspaces is locally linearly regular. □

6.3 Alternating Projections and Sparse-Polyhedral Feasibility

We want to give a complete anaysis of the behavior of alternating projections applied to problem (6.7), starting with preparatory lemmata.

Lemma 6.3.1. *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be closed subsets. If there exists $x \in \Omega_1$ such that $d_{\Omega_2}(x) = d_{\Omega_1}(P_{\Omega_2}x)$, then $x \in \mathrm{Fix}(P_{\Omega_1}P_{\Omega_2})$.*

Proof. This follows from the definitions of the projector (2.17) and fixed points (Definition 2.1.4). □

Lemma 6.3.2. *Let $(x, y) \in \mathcal{C} \times A_s$ be a local best approximation pair (Definition 2.3.7). Then $y \in \mathrm{StFix}(P_{A_s}P_{\mathcal{C}})$. Further, for all $x \in \mathrm{Fix}(P_{A_s}P_{\mathcal{C}})$, the point x is a local best approximation point to $\mathcal{C}$ in A_s. Not every local best approximation point to $\mathcal{C}$ in A_s is contained in $\mathrm{Fix}(P_{A_s}P_{\mathcal{C}})$.*

Proof. Let $(x, y) \in \mathcal{C} \times A_s$ be a local best approximation pair. Then $y \in P_{A_s}x$ and $x \in P_{\mathcal{C}}y$, and we have neighborhoods around x and y such that no point in these neighborhoods is closer to the other set than x and y, respectively. Hence, y is a stable fixed point. Now, let $x \in \mathrm{Fix}(P_{A_s}P_{\mathcal{C}})$. Then, for all $y \in A_s$ and since $x \in P_{A_s}P_{\mathcal{C}}x$, we know that $\|y - P_{\mathcal{C}}x\| \geq \|x - P_{\mathcal{C}}x\|$. We consider two cases:

1. If $x \in A_s \cap \mathcal{C}$, then x is by definition a local best approximation point.

2. If $x \notin A_s \cap \mathcal{C}$, then x has sparsity exactly equal to s since it is in the image of the projection onto A_s. Hence, there exists some $\delta > 0$ such that $A_s \cap \mathbb{B}_\delta(x) = A_J \cap \mathbb{B}_\delta(x)$ where $J = I(x)$ (see (3.10)). In other words, the set $A_J \cap \mathbb{B}_\delta(x)$ is convex. The vector $x - P_{\mathcal{C}}x$ is, since $x \in \mathrm{Fix}(P_{A_s}P_{\mathcal{C}})$, normal to A_J. Since the sets A_J and $\mathcal{C}$ are convex, there exists a separating hyperplane H between A_J and $\mathcal{C}$, which is normal to $x - P_{\mathcal{C}}x$. This hyperplane can be shifted such that $P_{\mathcal{C}}x \in H$. Because $x - P_{\mathcal{C}}x$ is also normal to A_J, we know that $P_{\mathcal{C}}x + A_J \subset H$, i.e., A_J is parallel to H, and $x - P_{\mathcal{C}}x$ is the gap between H and A_J. From this we conclude that any point $z \in \mathbb{B}_\delta(x) \cap A_J$ must satisfy $d_{\mathcal{C}}(z) \geq d_{\mathcal{C}}(x)$. Hence, x is a local best approximation point.

We give an example that shows that not every local best approximation point to $\mathcal{C}$ in A_s is a fixed point of $P_{A_s}P_{\mathcal{C}}$. Let A_s be the union of the x-axis and the $y-$axis in $\mathbb{R}^2$, and let $\mathcal{C} = \{(2,1)\}$. Then the point $(0,1)$ is a local best approximation point to $\mathcal{C}$ in A_s but

$$P_{A_s}P_{\mathcal{C}}(0,1) = (2,0) \neq (0,1).$$

This completes the proof. □

The following theorem gives an overview of the possibilities that can occur when dealing with alternating projections and the set of sparse vectors. A note on the alternating projections sequence in the following theorem: it may happen that the projection $P_{A_s}x$ of a point x onto the set A_s is not single-valued. Then it is important to have a look at *every* point in $P_{A_s}x$ because at one of them the sequence may continue in a good way. See Figure 6.1 for an example that illustrates this issue.
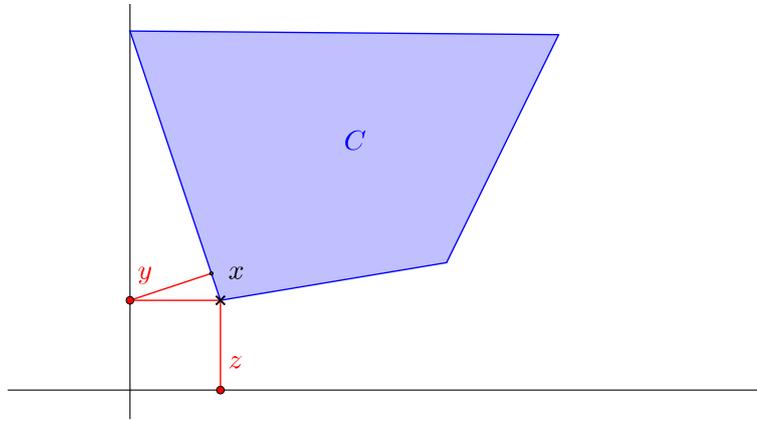
Figure 6.1: If we start at the point x in the graphic and project onto the set of 1-sparse vectors, then we can choose between the points y and z. If we choose z, then projecting back onto the polyhedron C results in x again, while choosing y would lead to an alternating projections sequence that converges to the intersection $C \cap A_1$. This example shows the necessity of different kinds of fixed points in Definition 2.1.4.

Theorem 6.3.3. *Let $\mathcal{C}$ be a polyhedral set as defined in (6.5), and let $s \in \mathbb{N}$ be such that $0 \leq s \leq \min\{m,n\}$. Let $\{x^n\}_{n\in\mathbb{N}}$ be a sequence generated by the alternating projections operator $P_{A_s}P_{\mathcal{C}}$, with x^0 arbitrary. Then the sequence $\{x^n\}_{n\in\mathbb{N}}$ has finitely many cluster points $\bar{x}_1, \ldots, \bar{x}_\nu$ satisfying $P_{\mathcal{C}}\bar{x}_1 = \cdots = P_{\mathcal{C}}\bar{x}_\nu$.*

Proof. First, we remind the reader of the result in Theorem 4.2.3 and that the set A_s can be written as a union of finitely many linear subspaces, i.e.,

$$A_s = \bigcup_{J \subset \{1,\ldots,n\}, \#J=s} A_J, \tag{6.12}$$

where A_J is the span of vectors of the standard basis of $\mathbb{R}^n$ indexed by J. Proceeding, we note that the polyhedron $\mathcal{C}$ has finitely many faces. These faces will be denoted by F_I with $I \in 2^{\{1,\dots,m\}}$ where I is an index set corresponding to lines of the matrix M, i.e.,

$$F_I := \mathcal{C} \cap B_I := \mathcal{C} \cap \{x \in \mathbb{R}^n \mid M_I x = p_I\}. \tag{6.13}$$

Further, define the sets

$$B_I := \operatorname{aff} F_I \quad \text{for all } I \in 2^{\{1,\dots,m\}}. \tag{6.14}$$

The dimension of F_I will be defined as the dimension of its affine hull B_I. For every index set $I \in 2^{\{1,\dots,m\}}$ and for every index set J as in (6.12), there exist, by Lemma 2.3.11, nearest points of the affine hull B_I to the subspace A_J. We observe that we have finitely many affine hulls of faces B_I of $\mathcal{C}$ and finitely many linear subspaces A_J. This gives us finitely many gap vectors g_{IJ} between the B_I and A_J as well as finitely many Friedrichs angles between the B_I and A_J, whose cosines are given by $c_{IJ} := c(B_I, A_J)$.

We analyze the sequence $\{x^n\}_{n \in \mathbb{N}}$ step by step. Let $x^0 \in \mathbb{R}^n$ be arbitrary and consider the kth iterate x^k.

1. Suppose now that, for some $k \geq 1$, $x^{k-1} \notin \operatorname{Fix} P_{A_s} P_{\mathcal{C}}$. Then x^k is exactly in one of the A_J in (6.12). This subspace A_J will be denoted by A_{J_k}. If $x^k \notin A_s \cap \mathcal{C}$, then we have $P_{\mathcal{C}} x^k \notin \operatorname{int}(\mathcal{C})$ for $k \neq 0$. In other words, $P_{\mathcal{C}} x^k$ satisfies at least one of the inequalities as equality, i.e., there exist $I \in 2^{\{1,\dots,m\}}$ such that $M x_I^k = p_I$. Denote by I_k the index set of the face of lowest dimension F_{I_k} such that $P_{\mathcal{C}} x^k \in F_{I_k}$. In each step of the alternating projections sequence, we have

$$P_{\mathcal{C}} x^k = P_{B_{I_k}} x^k. \tag{6.15}$$

In case of (6.15), define the gap vector $g_{I_k J_k} := P_{A_{I_k} - B_{J_k}} 0$. Assume that the cosine of the Friedrichs angle of $B_{J_k} - g_{I_k J_k}$ and A_{I_k} is not zero. Then we know that

$$d_{A_{J_k}} \left(P_{\mathcal{C}} x^k \right) < d_{\mathcal{C}} \left(x^k \right). \tag{6.16}$$

If we had equality, then, by Theorem 4.2.2, we would have a fixed point of alternating projections between $B_{J_k} - g_{I_k J_k}$ and A_{I_k} and, hence, a best approximation point to B_{J_k} in A_{I_k}. The only possiblity for the point x^k not to be a best approximation point between $\mathcal{C}$ and A_s is that there exists $x \in A_s$ with

$$\left\| x - P_{\mathcal{C}} x^k \right\| < \left\| P_{A_{J_k}} P_{\mathcal{C}} x^k - P_{\mathcal{C}} x^k \right\|. \tag{6.17}$$

Therefore, suppose that (6.16) holds. Further,

$$d_{A_s} \left(P_{\mathcal{C}} x^k \right) \leq d_{A_{J_k}} \left(P_{\mathcal{C}} x^k \right). \tag{6.18}$$

By Proposition 6.2.8, we have a linearly regular intersection of $\mathcal{C} - g_{I_k J_k}$ with A_{J_k}.

Further, we know by Theorem 4.2.3 that

$$
\begin{aligned}
d_{\mathcal{C}}\left(P_{A_{J_k}}P_{\mathcal{C}}x^k\right) - \|g_{I_kJ_k}\| &\leq \rho_k\left(d_{\mathcal{C}}\left(x^k\right) - \|g_{I_kJ_k}\|\right) \\
\Leftrightarrow \quad d_{\mathcal{C}}\left(P_{A_{J_k}}P_{\mathcal{C}}x^k\right) &\leq \rho_k\left(d_{\mathcal{C}}\left(x^k\right) - \|g_{I_kJ_k}\|\right) + \|g_{I_kJ_k}\|
\end{aligned}
\tag{6.19}
$$

with $\rho_k < 1$. In other words, in step k, the sequence approaches a local best approximation point to B_{J_k} in A_{I_k}. The constant ρ_k depends on the Friedrichs angle between A_{J_k} and B_{I_k}. Since there are only finitely many different Friedrichs angles between the affine hulls B_I of faces F_I and linear subspaces A_J, there are only finitely many different constants ρ_k. For simplicity, we define

$$
\bar{\rho} := \max\left\{\rho_k \mid \rho_k < 1,\ k \in \mathbb{N}\right\}.
\tag{6.20}
$$

If the cosine of the Friedrichs angle of $B_{J_k} - g$ and A_{I_k} is zero, then, by Lemma 6.2.6, either one of the spaces is contained in the other one or the spaces fulfill an orthogonality condition. In both cases, the sequence would attain a fixed point after one more step.

2. Let x^k be a fixed point of $P_{A_s}P_{\mathcal{C}}$. If $P_{A_s}P_{\mathcal{C}}x^k \subset \operatorname{Fix} P_{A_s}P_{\mathcal{C}}$, then, for every AP step, we can *choose* the next iterate among the points in $P_{A_s}P_{\mathcal{C}}x^k$ and, hence, obtain the subsequences $\{x^{n_{k_1}}\}_{k_1 \in \mathbb{N}}, \ldots, \{x^{n_{k_v}}\}_{k_v \in \mathbb{N}} \subset \{x^n\}_{n \in \mathbb{N}}$, where v is the cardinality of $P_{A_s}P_{\mathcal{C}}x^k$. Because $P_{A_s}P_{\mathcal{C}}x^k \subset \operatorname{Fix} P_{A_s}P_{\mathcal{C}}$ and since $\mathcal{C}$ is convex, we have $P_{\mathcal{C}}x = P_{\mathcal{C}}x^k$ for all $x \in P_{A_s}P_{\mathcal{C}}x^k$.

If $P_{A_s}P_{\mathcal{C}}x^k$ contains some y which is *not* a fixed point of $P_{A_s}P_{\mathcal{C}}$, then, as soon as $y = x^{k+1}$ is chosen as the next iterate, we have the case which is similar to the previous one.

What remains to be shown is that, by this procedure, the sequence actually converges. We know that the sequence $\{\delta^n\}_{n \in \mathbb{N}}$ with

$$
\delta^k := d_{\mathcal{C}}\left(x^n\right)
\tag{6.21}
$$

is nonnegative and monotonically decreasing (Lemma 4.2.4). This is already sufficient for the sequence $(\delta^n)_{n \in \mathbb{N}}$ to converge. We denote the limit by

$$
\delta := \lim_{n \to \infty} \delta^n.
\tag{6.22}
$$

Hence, there exists a sequence $\{\varepsilon^n\}_{n \in \mathbb{N}} \subset \mathbb{R}$ such that for all $n \in \mathbb{N}$, the term $d_{\mathcal{C}}\left(x^n\right)^2$ can be written as $\delta^2 + \varepsilon^n$ with $\varepsilon^n \geq 0$ and $\lim_{n \to \infty} \varepsilon^n = 0$. Now, let $n \in \mathbb{N}$ be such that ε^n is sufficiently small. By monotonicity of the distance of the iterates x^n to the set $\mathcal{C}$, there exists ε^{n+1} satisfying $0 \leq \varepsilon^{n+1} \leq \varepsilon^n$ such that

$$
d_{\mathcal{C}}\left(x^n\right)^2 = \delta^2 + \varepsilon^n \quad d_{\mathcal{C}}\left(x^{n+1}\right)^2 = \delta^2 + \varepsilon^{n+1}.
$$

We note that

$$\left\| x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 \leq \left\| x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 = \delta^2 + \varepsilon^n.$$

Together with convexity (see (2.31)) of the set $\mathcal{C}$, we can give a bound for the squared distance between $P_{\mathcal{C}} x^n$ and $P_{\mathcal{C}} x^{n+1}$:

$$
\begin{aligned}
\left\| x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 &= \left\| x^{n+1} - P_{\mathcal{C}} x^{n+1} \right\|_2^2 + \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 \\
&\quad + \underbrace{2 \left\langle x^{n+1} - P_{\mathcal{C}} x^{n+1}, P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\rangle}_{\leq 0 \text{ by Equation (2.31)}} \\
\Rightarrow \qquad \delta^2 + \varepsilon^n &\geq \delta^2 + \varepsilon^{n+1} + \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 \\
\Leftrightarrow \qquad \varepsilon^n - \varepsilon^{n+1} &\geq \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 \\
\Rightarrow \qquad \sqrt{\varepsilon^n} &\geq \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2.
\end{aligned}
$$

Since $\lim_{n\to\infty} \varepsilon^n = 0$, we conclude that $\lim_{n\to\infty} \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2^2 = 0$.

As already noted, there exist finitely many rates ρ_{IJ} in Equation (6.19), together with a smallest positive $\bar{\rho}$, defined in Equation (6.20). We distinguish between two cases:

1. If the number of iterates is such that $d_{\mathcal{C}}\left(x^{n+1}\right) = d_{\mathcal{C}}\left(x^n\right)$ is infinite, then this distance $d_{\mathcal{C}}\left(x^n\right)$ has to be equal to δ, and we have $P_{\mathcal{C}} x^n = P_{\mathcal{C}} x^{n+1}$. If $P_{\mathcal{C}} x^{n+2} \neq P_{\mathcal{C}} x^{n+1}$, then $d_{\mathcal{C}}\left(x^{n+2}\right) < d_{\mathcal{C}}\left(x^{n+1}\right)$, which cannot be true since $d_{\mathcal{C}}\left(x^{n+1}\right) = \delta$. Hence, $P_{\mathcal{C}} x^n = P_{\mathcal{C}} x^{n+m}$ for all $m \geq n$. Further, since $P_{A_s} P_{\mathcal{C}} x^n$ is of finite cardinality, there can only be finitely many cluster points of the sequence $\{x^n\}_{n\in\mathbb{N}}$. Their projection onto $\mathcal{C}$, however, is identical, which was claimed.

2. If the number of iterates is such that $d_{\mathcal{C}}\left(x^{n+1}\right) = d_{\mathcal{C}}\left(x^n\right)$ is finite, then we can choose $\bar{n}$ large enough such that $d_{\mathcal{C}}\left(x^{n+1}\right) < d_{\mathcal{C}}\left(x^n\right)$ for all $n \geq \bar{n}$. If we write $d_{\mathcal{C}}\left(x^n\right) = \delta + \tau^n$, then, by Equation (6.20), we know that $\tau^{n+1} \leq \rho\tau^n$ for all $n \geq \bar{n}$. To adapt our notation to (6.3), we can write

$$d_{\mathcal{C}}\left(x^n\right)^2 = \delta^2 + \varepsilon^n = \delta^2 + \tau^n(2\delta + \tau^n) \text{ ,i.e., } \varepsilon^n = \tau^n(2\delta + \tau^n).$$

We obtain the relation

$$
\begin{aligned}
d_{\mathcal{C}}\left(x^{n+1}\right)^2 &= \delta^2 + \varepsilon^{n+1} = \delta^2 + \tau^{n+1}(2\delta + \tau^{n+1}) \\
&\leq \delta^2 + \rho\tau^n(2\delta + \rho\tau^n) \leq \delta^2 + \rho\tau^n(2\delta + \tau^n) \\
&= \delta^2 + \rho\varepsilon^n.
\end{aligned}
$$

We conclude that $\varepsilon^{n+1} \leq \rho\varepsilon^n$ for all $n \geq \bar{n}$.

$$
\begin{aligned}
&\sum_{n=\bar{n}}^{\infty} \left\| P_{\mathcal{C}} x^{n+1} - P_{\mathcal{C}} x^n \right\|_2 \\
&\leq \sum_{n=\bar{n}}^{\infty} \sqrt{\varepsilon^n} \leq \sqrt{\varepsilon^n} \sum_{n=\bar{n}}^{\infty} \sqrt{\rho^n}
\end{aligned}
$$

$$= \sqrt{\varepsilon^n} \frac{1}{1 - \sqrt{\rho}}.$$

From this, we deduce that the sequence $\{P_{\mathcal{C}}x^n\}_{n\in\mathbb{N}}$ is bounded. Hence, it is a subset of a compact subset of $\mathbb{R}^n$, and there exists a cluster point $y \in \mathcal{C}$ of the sequence $\{P_{\mathcal{C}}x^n\}_{n\in\mathbb{N}}$. Further, this cluster point must satisfy $d_{A_s}(y) = \delta$. Finally, we show by contradiction that the cluster point y is unique. Assume there is another cluster point $y' \neq y \in \mathcal{C}$ of the sequence $\{P_{\mathcal{C}}x^n\}_{n\in\mathbb{N}}$. Then the distance between these points is $d := \|y - y'\|_2 > 0$. Since the sequence $\{\varepsilon^n\}_{n\in\mathbb{N}}$ converges monotonically to zero, we can find some $\bar{n} \in \mathbb{N}$ such that $\sqrt{\varepsilon^{\bar{n}}}\frac{1}{1-\sqrt{\rho}} \leq \frac{d}{2}$ and such that $P_{\mathcal{C}}x^{\bar{n}} - y < \frac{d}{4}$. Hence, there cannot be an infinite subsequence $\{P_{\mathcal{C}}x^k\}_{k\in I} \subset \{P_{\mathcal{C}}x^n\}_{n\in\mathbb{N}}$ having y' as a cluster point because there exist neighborhoods of y' such that there exists no k such that $P_{\mathcal{C}}x^k$ is contained in these neighborhoods. This contradicts the fact that y' is a cluster point. We conclude that $y \in \mathcal{C}$ is the only cluster point of $\{P_{\mathcal{C}}x^n\}_{n\in\mathbb{N}}$. As before, since $P_{A_s}P_{\mathcal{C}}x^n$ is of finite cardinality, there can only be finitely many cluster points of the sequence $\{x^n\}_{n\in\mathbb{N}}$.

This completes the proof. □

Remark 6.3.4. *The novelty in Theorem 6.3.3 is that we have shown convergence of the alternating projections in sparse polyhedral feasibility without any further restriction. In particular, if $\mathcal{C}$ is a compact set then classical convergence theory applies. In Theorem 6.3.3 we do not require any compactnes of the polyhedral set.*

We present examples for different cases in Theorem 6.3.3. A good example is the objective of finding the nearest point of a line L in three dimensions to the set of 1-sparse vectors $A_1 \subset \mathbb{R}^3$. Suppose that L has no common point with A_1. Then there exists a best approximation pair between L and A_1.

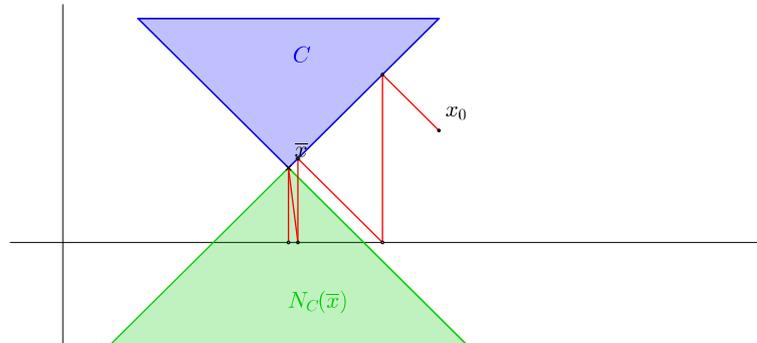Figure 6.2: An example of an alternating projections sequence that finds a best approximation pair between the set of 1-sparse vectors in $\mathbb{R}^2$ and the polyhedron C in finitely many steps. The proximal normal cone at $\bar{x}$ is shown in green.

The following can occur when the nearest point x in a polyhedron to the set of sparse vectors is a corner of the polyhedron. The alternating projection sequence behaves as if

it is finding the best approximation pair relative to one of the faces of the polyhedron. Then it may happen that the sequence contains points which are normal to another face of the polyhedron. The following projection onto the polyhedron will be in the intersection of these two faces. Figure 6.2 shows an example of this case.

Another example is as follows: take $\mathcal{C} = \{(1,1)\}$ and find the nearest point in the set of 1-sparse vectors. Generate the alternating projections sequence with an arbitrary initial point in $\mathbb{R}^2$. The projection $P_{\mathcal{C}}$ is always the point $(1,1)$, while $P_{A_1}(1,1) = \{(1,0),(0,1)\}$. Hence, we can choose a point in $P_{A_1}(1,1)$. So the AP-sequence can be any randomly chosen sequence of the points $(1,0)$ and $(0,1)$, which is not a best approximation pair (see Figure 6.3).
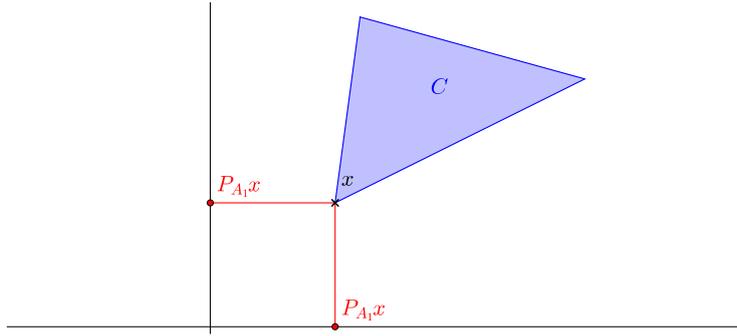
Figure 6.3: An example of a best approximation pair between the set of 1-sparse vectors in $\mathbb{R}^2$ and the polyhedron C where the pair is not unique.

Now we can formulate a necessary condition for the method of alternating projections to converge globally.

Theorem 6.3.5. *Let B and A_s be given as in (3.6) and (3.5) respectively. Suppose $A_s \cap B \neq \emptyset$. Let*

$$A_s = \bigcup_{J \in \mathcal{J}_s} A_J, \tag{6.23}$$

where $\mathcal{J}_s$ is given by (3.11). Denote by $(x_J, x_J^s) \in B \times A_s$ the best approximation pairs between B and A_J for every $J \in \mathcal{J}_s$. Then, for an arbitrary $x^0 \in \mathbb{R}^n$, the sequence generated by

$$x^{k+1} = T_{AP}x^k = P_{A_s}P_B x^k \tag{6.24}$$

converges to a point $\bar{x} \in A_s \cap B$ if and only if, for all $x_J \notin B \cap A_s$, we have

$$d_{A_s}(x_J) < d_{A_J}(x_J) \text{ for all } J \in \mathcal{J}_s. \tag{6.25}$$

Proof. First, we note that the set B is by definition a polyhedral set. By Theorem 6.3.3, the method of alternating projections converges at a linear rate to a best approximation pair. Now define

$$\delta := \min_{J \in \mathcal{J}_s} \mathrm{d}(A_J, B). \tag{6.26}$$

Suppose now that, for every best approximation pair (x_J, x_J^s), the property (6.25) is fulfilled. Then, after finitely many steps, say k, we have $d_B(x^k) < \delta$. From then on the sequence must converge to $A_s \cap B$ at a linear rate.

On the other hand, suppose property (6.25) is not fulfilled for some $I \in \mathcal{J}_s$. Then we can just consider the alternating projections sequence initiated at $x^0 = x_I$. Since $d_{A_s}(x_I) = d_{A_I}(x_I)$, we can choose $P_{A_s} x^k = x_I^s$. Then $x^{k+1} = x_I = x^k$, and the sequence will not converge at all to a solution of (3.8). This completes the proof. $\qquad\square$

As seen before, there are strong sufficient conditions for alternating projections to converge globally in the sparse affine feasibility problem. The condition (5.13) can also be formulated in terms of angles.

Lemma 6.3.6. *Property* (5.13) *holds true if and only if we have*

$$c\left(A_J, \ker(M)\right) \leq \sqrt{1 - \delta_{2s}} \quad \text{for all } J \in \mathcal{J}_s. \tag{6.27}$$

Proof. By (5.13), we have

$$(1 - \delta_{2s})\|x\|_2^2 \leq \|M^\dagger M x\|_2^2 \quad \text{for all } x \in A_{2s}. \tag{6.28}$$

In other words, $\cos(\alpha) \leq \sqrt{1 - \delta_{2s}}$ for all $J \in \mathcal{J}_s$. $\qquad\square$

To close this chapter, we consider an example for a setup where an alternating projections sequence will converge to the intersection of the set of sparse vectors.

Example 6.3.7. *Consider the alternating projections algorithm applied to the set of 2-sparse vectors in $\mathbb{R}^n$ and the one-dimensional affine space defined by*

$$B(\lambda) = \left\{ \begin{pmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \,\middle|\, \mu \in \mathbb{R} \right\}. \tag{6.29}$$

Apparently, the only solution to (3.8) is the point with its first entry equal to μ and all other entries equal to zero. For a thought experiment, we determine all best approximation pairs between B and the subspaces gained by the decomposition of A_2 (see Equation (3.12)).

First, we consider

$$A_2 = \bigcup_{i=1, i<j}^{n-1} \text{span}\left\{e_i, e_j\right\}, \tag{6.30}$$

where e_i are the standard unit vectors in $\mathbb{R}^n$. Let us determine the best approximation pairs between B and $\text{span}\left\{e_i, e_j\right\}$ for $i, j \neq 1$. The case where $i = 1$ is trivial due to the existing intersection. Without loss of generality, let $i = 2, j = 3$. A best approximation pair between B and $\text{span}\left\{e_i, e_j\right\}$ will be of the shape $((\lambda + x, x, \ldots, x), (0, x, x, 0, \ldots, 0))$. The projection

$P_B(0, x, x, 0, \ldots, 0)$ *can be computed via minimizing*

$$f(y) = (y + \lambda)^2 + (y - x)^2 + (y - x)^2 + \sum_{i=4}^{n} y^2, \tag{6.31}$$

which represents the squared Euclidean distance of $(0, x, x, 0, \ldots, 0)$ *to a point in B. The first-order necessary condition for optimality yields*

$$
\begin{aligned}
\frac{\mathrm{d}f}{\mathrm{d}y} &= 2\lambda + 2ny - 4x = 0 \\
\Leftrightarrow \quad y &= \frac{2x - \lambda}{n}.
\end{aligned}
\tag{6.32}
$$

Now, in order to have a best approximation pair, the equality

$$x = \frac{2x - \lambda}{n} \tag{6.33}$$

must hold. But (6.33) *is equivalent to* $x = \frac{-\lambda}{n-2}$. *The best approximation pair is thus*

$$\left(\left(\lambda - \frac{\lambda}{n-2}, \frac{-\lambda}{n-2}, \ldots, \frac{-\lambda}{n-2} \right)^\top, \left(0, \frac{-\lambda}{n-2}, \frac{-\lambda}{n-2}, 0, \ldots, 0 \right)^\top \right). \tag{6.34}$$

For all $\lambda \in \mathbb{R}_{>0}$ *and for* $n \geq 5$, *we have*

$$\left| \lambda - \frac{\lambda}{n-2} \right| > \left| \frac{-\lambda}{n-2} \right|. \tag{6.35}$$

In other words, the projection of

$$\left(\lambda - \frac{\lambda}{n-2}, \frac{-\lambda}{n-2}, \ldots, \frac{-\lambda}{n-2} \right)^\top \tag{6.36}$$

onto A_2 *is the point* $\left(\lambda - \frac{\lambda}{n-2}, \frac{-\lambda}{n-2}, 0, \ldots, 0 \right)^\top$. *From now on, the alternating projections sequence will converge to the solution of* (3.8). *Using Theorem 6.3.3, we know that any sequence of alternating projections between* A_2 *and a convex polyhedral set converges to best approximation pairs. But initiated at a best approximation pair, the sequence approaches the intersection* $A_2 \cap B$. *This shows that the affine subspace B is a prototype for an affine subspace where, independent of how it is shifted from the origin, alternating projections converges globally.*

7 Spectral Sets

With this chapter, we introduce a link between the analysis of sparsity optimization and the forthcoming analysis of matrices of a fixed rank. This chapter relies heavily on the fundamental work of Lewis and Sendov in (Lewis and Sendov, 2005). There the authors give relations between sets of vectors satisfying a symmetry condition and sets of matrices having these vectors as singular values. First, we introduce the so-called *transfer principle*, which has been analyzed in (Daniilidis et al., 2008) in case of prox-regularity. We extend one part of their analysis to weaker types of regularities.

7.1 Preparations

Definition 7.1.1. *Let G be a group, together with an action of G on a set C. The action may be both from the left and from the right. For some $x \in C$, we define the* orbit *of x under the group action of G,*

$$G \cdot x := \{g \cdot x \mid g \in G\}. \tag{7.1}$$

The two-sided orbit of G on C will be denoted by

$$G^2 \cdot C := \{g \cdot x \cdot h \mid g, h, \in G, x \in C\}. \tag{7.2}$$

The two-sided action can also be seen as an orbit of the group G^2 on the set C, which is why we chose the notation. For $x \in C$, we define furthermore the stabilizer *of x in G by*

$$\mathrm{Stab}_G(x) := \{g \in G \mid g \cdot x = x\}. \tag{7.3}$$

Theorem 7.1.2 (singular value decomposition (SVD) (Horn and Johnson, 1985, Theorem 7.3.5))**.** *If $X \in \mathbb{C}^{n \times n}$ has rank s, then it may be written in the form $X = U\Sigma V^*$ where $U \in U(n)$ and $V \in U(n)$ are unitary. The entries σ_{jk} of the matrix Σ satisfy $\sigma_{jk} = 0$ for all $j \neq k$ and*

$$\sigma_{11} \geq \cdots \geq \sigma_{ss} > \sigma_{s+1,s+1} = \cdots = \sigma_{nn} = 0.$$

The numbers σ_{jj} are the nonnegative square roots of eigenvalues of XX^*, and hence are uniquely determined. The columns of U are the eigenvectors of XX^*, and the columns of V are the eigenvectors of X^*X (arranged in the same order as the corresponding eigenvalues σ_{jj}^2). If XX^* has distinct eigenvalues, then U is determined up to a diagonal factor $D = \mathrm{diag}\left(\mathrm{e}^{\mathrm{i}\theta_1}, \ldots, \mathrm{e}^{\mathrm{i}\theta_n}\right)$ with all $\theta_j \in \mathbb{R}$; that is, if $X = U_1\Sigma_1 V_1^* = U_2\Sigma_2 V_2^*$, then $U_2 = U_1 D$. If U is given, then V is uniquely determined. If X is real, then U, Σ, and V may all be taken to be real.

Definition 7.1.3. *Denote by $S(n, -)$ the finite group of signed permutation matrices on $\mathbb{R}^n$. That is, for any $P \in S(n, -)$, every row and every column of P contains exactly one nonzero element. This nonzero element is ± 1.*

Definition 7.1.4 (singular value mapping)**.** *For $X \in \mathbb{C}^{n \times n}$, the decomposition $X = U \Sigma V^*$ in Theorem 7.1.2 is called a* singular value decomposition *of X. Define the* singular value mapping *by*

$$\sigma(X) := (\sigma_1(X), \ldots, \sigma_n(X)), \tag{7.4}$$

where, with $\sigma_j(X) = \sigma_{jj}$ as in Theorem 7.1.2, we have $\sigma_1(X) \geq \cdots \geq \sigma_n(X)$. Further, for some $x \in \mathbb{R}^n$, define

$$\sigma^{-1}(x) := \left\{ X \in \mathbb{C}^{n \times n} \mid \sigma(X) = Px \text{ for some } P \in S(n, -) \right\}. \tag{7.5}$$

7.2 The Transfer Principle

The transfer principle is a loose phrase for properties of functions and sets of matrices almost entirely relying on their action on its singular values. A set A of vectors is translated into a set of singular values by considering the set of diagonal matrices $\text{diag}(A)$ and then letting the orthogonal group act from left and right on $\text{diag}(A)$. This couples an orbit of the Lie group $O(n)$ to every matrix $Y \in \text{diag}(A)$, a set usually having more structure than $\text{diag}(A)$. This additional structure can be used to show properties of $O(n)^2 \cdot \text{diag}(A)$. We introduce all necessary notation to give our main result of this chapter.

Definition 7.2.1 (Lie group, (tom Dieck and Bröcker, 2003, Definition 1.1))**.** *A* Lie group *is a differentiable manifold G which is also a group such that the group multiplication is a differentiable map.*

Definition 7.2.2 (absolutely symmetric function and orthogonally invariant function, (Lewis and Sendov, 2005), (Daniilidis et al., 2008))**.**

- *A function $f : \mathbb{R}^n \to [-\infty, \infty]$ will be called* absolutely symmetric *if for all vectors $x \in \mathbb{R}^n$ and for all signed permutation matrices P, that is, every row and column of P contains exactly one nonzero entry of the form ± 1, we have*

$$f(x) = f(Px). \tag{7.6}$$

- *A set $C \subset \mathbb{R}^n$ will be called* absolutely symmetric, *if its indicator function is an absolutely symmetric function.*

- *A function $F : \mathbb{R}^{n \times n} \to [-\infty, \infty]$ will be called* orthogonally invariant *if we have*

$$F(UXV^\top) = F(X) \quad \text{for all} \quad X \in \mathbb{R}^{n \times n} \quad \text{and for all} \quad U, V \in O(n).$$

- *A subset $S \subset \mathbb{R}^{n \times n}$ will be denoted as a* spectral set *if there exists an orthogonally invariant function F such that $S = \text{dom}(F)$.*

In (Daniilidis et al., 2008) an orthogonally invariant function is also denoted as a *spectral function*. Note that a subset $S \subset \mathbb{R}^{n \times n}$ is a spectral subset if and only if we have $UXV^\top \in S$ for all $U, V \in O(n)$ and for all $X \in S$.

Definition 7.2.3 (singular value function, (Lewis and Sendov, 2005), Definition 5.2). *Given the singular value mapping σ (Definition 7.1.4), a* singular value function *is an extended-real-valued function defined on $\mathbb{R}^{n \times n}$ of the form $f \circ \sigma$ for an absolutely symmetric function $f : \mathbb{R}^n \to [-\infty, \infty]$.*

Proposition 7.2.4 ((Lewis and Sendov, 2005), Proposition 5.1). *Given σ as in Definition 7.1.4, a function $F : \mathbb{R}^{n \times n}$ is orthogonally invariant if and only if F is a singular value function (Definition 7.2.3).*

Theorem 7.2.5 ((Lewis and Sendov, 2005), Theorem 3.4). *The orbit*

$$O(n)^2 \cdot X = \{UXV \mid U, V \in O(n)\} \tag{7.7}$$

is a submanifold of the space $\mathbb{R}^{n \times n}$ with tangent space

$$T_{O(n)^2 \cdot X}(X) = \{XU^a - V^a X \mid U^a, V^a \text{ skew-symmetric}\} \tag{7.8}$$

and normal space

$$N_{O(n)^2 \cdot X}(X) = \left\{ Y \in \mathbb{R}^{n \times n} \mid X^\top Y \text{ and } XY^\top \text{ symmetric} \right\}. \tag{7.9}$$

Theorem 7.2.6 ((Lewis and Sendov, 2005), Theorem 7.1). *The limiting subdifferential of a singular value function $f \circ \sigma$ at a matrix $X \in \mathbb{R}^{n \times n}$ is given by the formula*

$$\partial(f \circ \sigma)(X) = \text{Stab}_{O(n)^2}(X)^2 \cdot \text{diag}\left(\partial f(\sigma(X))\right), \tag{7.10}$$

where

$$\text{Stab}_{O(n)^2}(X) = \left\{ (U, V) \in O(n)^2 \mid UXV^\top = X \right\}. \tag{7.11}$$

The sets of regular and horizon subgradients satisfy corresponding formulae[1].

Proposition 7.2.7. *Let A be an absolutely symmetric subset of $\mathbb{R}^n$ with (proximal or limiting) normal cone $N_A(x)$ at a point $x \in A$. Further, for $X \in \sigma^{-1}(A)$, let $U, W \in O(n)$ such that $U \text{diag}(\sigma(X)) W^\top = X$. Then the (proximal or limiting) normal cone at X is given by*

$$N_{\sigma^{-1}(A)}(X) = \text{Stab}_{O(n)^2}(X) \cdot \left(U \text{diag}(N_A(\sigma(X))) W^\top \right). \tag{7.12}$$

Proof. If the set A is absolutely symmetric, then the indicator function of $\sigma^{-1}(A)$ is a singular value function as in Definition 7.2.3. Hence, we can apply Theorem 7.2.6 to the indicator function of A and the result follows. □

---

[1] For a definition of a regular and a horizon subgradient, see (Rockafellar and Wets, 1998, Definition 8.3)

As a last statement in this chapter, we give the example of the transfer principle in prox-regularity. Note that prox-regularity is a local property. The following theorem does not indicate how the neighborhoods of the spectral set and the neighborhoods of the set of singular values are related to each other.

Theorem 7.2.8 ((Daniilidis et al., 2008), Theorem 9). *An absolutely symmetric set $A \subset \mathbb{R}^n$ is prox-regular if and only if $S := O(n)^2 \cdot \text{diag}\,(A) \subset \mathbb{R}^{n \times n}$ is prox-regular.*

7.3 Weaker Regularities

Theorem 7.3.1 (transfer principle for subregularity). *Let $A \subset \mathbb{R}^n$ be an absolutely symmetric, closed set and let $S := \sigma^{-1}(A)$ be as in Definition 7.1.4.*

1. *If S is (ε, δ)-subregular at $X \in S$ with $\delta > 0, \varepsilon < 1$, then A is (ε, δ)-subregular at $\sigma(X) \in A$.*

2. *If S is (ε, δ)-regular at $X \in S$ with $\delta > 0, \varepsilon < 1$, then A is (ε, δ)-regular at $\sigma(X) \in A$.*

3. *If S is Clarke regular at $X \in S$, then A is Clarke regular at $\sigma(X) \in A$.*

4. *If S is super-regular at $X \in S$, then A is super-regular at $\sigma(X) \in A$.*

Proof. As a preparation, choose

$$
\overline{Y} \in \mathbb{B}_{\tilde{\delta}}(X) \cap \left\{ Z \in S \;\middle|\; \begin{array}{c} X = U \,\text{diag}\,(\sigma(X))\, V^\top \Leftrightarrow Z = U \,\text{diag}\,(z)\, V^\top \\ \text{for some } z \in \mathbb{R}^n, U, V \in O(n) \end{array} \right\}. \tag{7.13}
$$

Now, let $\overline{U}, \overline{V} \in O(n)$ be such that $X = \overline{U} \,\text{diag}\,(\sigma(X))\, \overline{V}^\top$ and $\overline{Y} = \overline{U} \,\text{diag}\,(\overline{y})\, \overline{V}^\top$ for a vector $\overline{y} \in \mathbb{R}^n$.

If we have $\overline{U Y V}^\top = \text{diag}\,(\overline{y})$ then, by Proposition 7.2.7, we also have $\overline{U Q V}^\top = \text{diag}\,(\overline{q})$ for some $\overline{q} \in N_A^P(\overline{y})$ for all $\overline{Q} \in N_S^P(\overline{Y})$. We conclude

$$
\begin{aligned}
\langle \overline{q}, \sigma(X) - \overline{y} \rangle &= \text{tr}\left(\text{diag}\,(\overline{q})^\top \left(\text{diag}\,(\sigma(X)) - \text{diag}\,(\overline{y}) \right) \right) \\
&= \left\langle \overline{U}^\top \,\text{diag}\,(q)\, \overline{V}, \overline{U}^\top \,\text{diag}\,(\sigma(X))\, \overline{V} - \overline{U}^\top \,\text{diag}\,(\overline{y})\, \overline{V} \right\rangle \\
&= \left\langle \overline{Q}, X - \overline{Y} \right\rangle
\end{aligned} \tag{7.14}
$$

Because the Frobenius norm is orthogonally invariant, we have

$$
\begin{aligned}
\left\| \overline{Q} \right\|_F \left\| X - \overline{Y} \right\|_F &= \left\| \overline{U} \,\text{diag}\,(\overline{q})\, \overline{V}^\top \right\|_F \left\| U \,\text{diag}\,(\sigma(X))\, V^\top - U \,\text{diag}\,(\overline{y})\, V^\top \right\|_F \\
&= \left\| \text{diag}\,(\overline{q}) \right\|_F \left\| \text{diag}\,(\sigma(X)) - \text{diag}\,(\overline{y}) \right\|_F \tag{7.15} \\
&= \left\| \overline{q} \right\|_2 \left\| \sigma(X) - \overline{y} \right\|_2. \tag{7.16}
\end{aligned}
$$

Now, to show (1), let S be (ε, δ)-subregular at $X \in S$ with $\delta > 0$. Then, for all $Y \in \mathbb{B}_\delta(X) \cap S$ and for all $Q \in N_S^P(Y)$, we have

$$\langle Q, X - Y \rangle \leq \varepsilon \|Q\|_F \|X - Y\|_F. \tag{7.17}$$

Let $\bar{y} \in \mathbb{B}_\delta(\sigma(X)) \cap A$ and $\bar{q} \in N_A(\bar{y})$. Then there exists a matrix $\overline{Y}$ as in (7.13) such that $\sigma(\overline{Y}) = \bar{y}$, together with a matrix $Q \in N_S(\overline{Y})$ such that $\sigma(\overline{Q}) = \bar{q}$. This matrix $\overline{Y}$ satisfies (7.17). Using Equations (7.14) and (7.16), we obtain

$$\begin{aligned}
\langle \bar{q}, \sigma(X) - \bar{y} \rangle &= \langle \overline{Q}, X - \overline{Y} \rangle \\
&\leq \varepsilon \|\overline{Q}\|_F \|X - \overline{Y}\|_F = \varepsilon \|\bar{q}\|_2 \|\sigma(X) - \bar{y}\|_2.
\end{aligned}$$

Because $\overline{Y}$ is an arbitrary element of the intersection in (7.13), the above inequality yields the (ε, δ)-subregularity of A at $\sigma(X)$. This shows claim number (1).

To show (2), let S be (ε, δ)-regular at $X \in S$ with $\delta > 0, \varepsilon < 1$. Then, for all $Y, Z \in \mathbb{B}_\delta(X) \cap S$ and for all $Q \in N_S^P(Y)$, we have

$$\langle Q, Z - Y \rangle \leq \varepsilon \|Q\|_F \|Z - Y\|_F. \tag{7.18}$$

Let $\bar{y}, \bar{z} \in \mathbb{B}_\delta(\sigma(X)) \cap A$ and $\bar{q} \in N_A(\bar{y})$. Then there exist matrices $\overline{Y}, \overline{Z}$ as in (7.13) such that $\sigma(\overline{Y}) = \bar{y}, \sigma(\overline{Z}) = \bar{z}$, together with a matrix $Q \in N_S(\overline{Y})$ such that $\sigma(\overline{Q}) = \bar{q}$. As in the proof of (1), these matrices $\overline{Y}, \overline{Z}$ satisfy (7.18). Using Equations (7.14) and (7.16), we obtain

$$\begin{aligned}
\langle \bar{q}, \bar{z} - \bar{y} \rangle &= \langle \overline{Q}, \overline{Z} - \overline{Y} \rangle \\
&\leq \varepsilon \|\overline{Q}\|_F \|\overline{Z} - \overline{Y}\|_F = \varepsilon \|\bar{q}\|_2 \|\bar{z} - \bar{y}\|_2.
\end{aligned}$$

Because $\overline{Y}$ and $\overline{Z}$ are arbitrary elements of the intersection in (7.13), the above inequality yields the (ε, δ)-regularity of A at $\sigma(X)$. This shows claim number (2).

Using Definition 2.4.1, the proofs of (3) and (4) are similar to the one of (2), which is why we omit them at this point.

<div align="right">□</div>

8 The Set of Low Rank Matrices

This chapter puts the focus on the set of matrices of a fixed rank. First, we introduce the rank function together with some of its properties. Afterwards, the set of rank-constrained matrices will be defined.
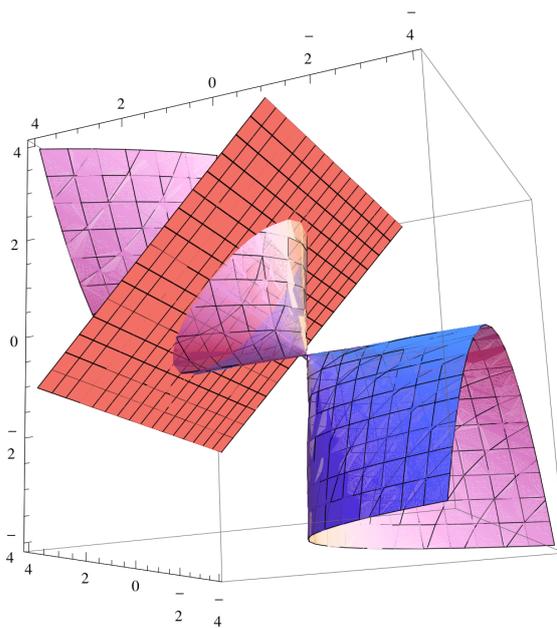
Figure 8.1: The set of 2-by-2 Hermitian matrices of rank one together with an affine subspace of matrices.

8.1 The Rank Function

As known from linear algebra, the rank of a matrix X is the largest number r of linearly independent columns of X. In order to give a proof of global convergence of the alternating projections in the matrix regime, similar to Corollary 5.2.4, we show that the rank function is lower semicontinuous and subadditive.

Lemma 8.1.1. *The rank function is subadditive.*

Proof. Take any two matrices $X, Y \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(X) = r_X$ and $\operatorname{rank}(Y) = r_Y$. If $m \leq n$, then there exist exactly r_X linearly independent columns of X and exactly r_Y

linearly independent columns of Y. This means that there exist at most $r_X + r_Y$ linearly independent columns of $X + Y$. If $m \geq n$, then the same argument applies for rows of X, Y, and $X + Y$. Altogether, we get rank $(X + Y) \leq$ rank $(X) +$ rank (Y). Since the matrices X and Y were chosen arbitrarily, the proof is complete. $\qquad\square$

Lemma 8.1.2. *The rank function is lower semicontinuous.*

Proof. We refer to Theorem 8.2.2, which says that between every matrix X of rank higher than $r \leq \min\{m, n\}$, there is a nonzero distance to the set of matrices of rank lower or equal to r. Hence, for every matrix X of rank $> r$ there is a neighborhood $\mathbb{B}_\varepsilon(X)$ having an empty intersection with the set of matrices of rank lower or equal to r. This makes the set of matrices of rank $> r$ an open subset of $\mathbb{R}^{m \times n}$. Hence, the lower levelsets of the rank function are closed sets, which is sufficient for the rank function to be lower semicontinuous by Definition 2.1.7. $\qquad\square$

8.2 Rank-Constrained Matrices

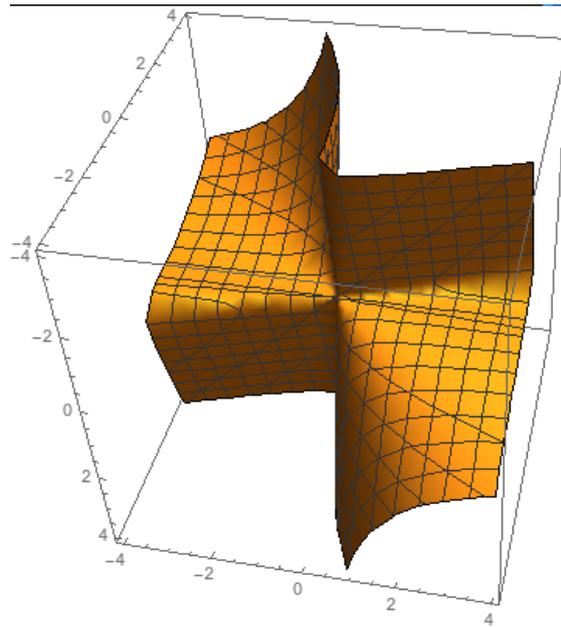We formalize the arguments used in Lemma 8.1.2 by giving the proper definitions.

Figure 8.2: A section of the set of 3-by-3 matrices of rank 2. In particular, the figure shows the zeros of $3xyz - z^2y - y^2z - x^3 = \det \begin{pmatrix} x & y & z \\ y & y & x \\ z & x & z \end{pmatrix}$.

Definition 8.2.1 (Set of rank-constrained matrices). *Let $r \leq m \leq n$, and define the set of matrices of rank at most r by*

$$S_{\leq r} := \left\{ X \in \mathbb{R}^{m \times n} \mid \operatorname{rank}(X) \leq r \right\}. \tag{8.1}$$

Further, define the set

$$S_{=r} := S_{\leq r} \setminus S_{\leq r-1} \tag{8.2}$$

as the set of matrices of rank exact r.

In contrast to $S_{\leq r}$, the set $S_{=r}$ is an embedded, smooth submanifold of the Euclidean space $\mathbb{R}^{mn}$, as will be shown in Theorem 8.2.6. By this, the set $S_{=r}$ satisfies useful regularity conditions, such as prox-regularity. We note that the set $S_{\leq r}$ can be defined for complex matrices equivalently.

Since for $X \in \mathbb{C}^{n \times n}$ and for all $\lambda \in \mathbb{C} \setminus \{0\}$ the equivalence

$$\operatorname{rank}(X) = r \quad \Leftrightarrow \quad \operatorname{rank}(\lambda X) = r \tag{8.3}$$

holds, we conclude that $S_{\leq r}$ is a cone.

The following theorem is known as the Eckart-Young theorem. We also refer to (Horn and Johnson, 1985, Example 7.4.1).

Theorem 8.2.2 (Eckart-Young theorem (Eckart and Young, 1936, Equation 17)). *Denote by $U(n)$ the group of unitary matrices in $\mathbb{C}^{n \times n}$. Let $A \in \mathbb{C}^{m \times n}$ be an arbitrary matrix and let $A = U \Sigma V^\top$ with $U \in U(m), V \in U(n), \Sigma \in \mathbb{R}^{m \times n}$ be its singular value decomposition. Then the matrix Σ can be written as $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, where $r = \min\{m, n\}$ and $\sigma_1 \geq \cdots \geq \sigma_n$. Then the projection of A onto the set of rank constrained matrices $S_{\leq r}$ is given by*

$$P_{S_{\leq r}} A = U \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) V^\top. \tag{8.4}$$

A useful issue of the set $S_{=r}$ is the fact that whenever we project a matrix $X \in \mathbb{R}^{m \times n}$ with rank greater or equal to r onto $S_{\leq r}$, then we actually project onto the set $S_{=r}$, as was shown in Theorem 8.2.2.

Proposition 8.2.3 (projection of Hermitian matrices). *Let $Z \in \mathbb{C}^{n \times n}$ be Hermitian, i.e., $Z^* = Z$. Then there exists a matrix $Z_r \in P_{S_{\leq r}} Z$ such that $Z_r = Z_r^*$. The problem of finding the matrix Z_r is equivalent to that of finding r eigenvectors of Z corresponding to r largest eigenvalues of Z.*

Proof. Let $Z = U \Sigma V^*$ be a singular value decomposition of Z. Then, by Theorem 7.1.2, the columns of U are the eigenvectors of $ZZ^* = Z^2$, and the columns of V are the eigenvectors of $Z^* Z = Z^2$. The eigenvectors of Z^2 are identical to those of Z. Further, the eigenvalues of a Hermitian matrix are real. Hence, the eigenvalues of Z differ from its singular values only by a factor ± 1. We conclude that the columns of U and the columns of V are the eigenvectors of Z, which means that we can choose $U = V$. Hence, $Z = U \Sigma U^*$. Let $u_1, \ldots, u_n$ be the columns of U. Let $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$ be as in Theorem 8.4. The columns of U are ordered in such a way that u_1 corresponds to

the largest absolute eigenvalue, u_2 to the second largest absolute eigenvalue and so on. Then $U\Sigma_r U^* \in P_{S_{\leq r}} Z$. Choose $Z_r = U\Sigma_r U^*$. Since $Z_r^* = (U\Sigma_r U^*)^* = U\Sigma_r U^* = Z_r$, we see that Z_r is Hermitian.

We observe that, for the matrix $D := \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{C}^{r \times r}$, we have

$$Z_r = (u_1, \ldots, u_r) D (u_1, \ldots, u_r)^*.$$

In other words, to obtain Z_r, it is necessary and sufficient to find the vectors $u_1, \ldots, u_r$.

<div align="right">□</div>

The set $S_{\leq r}$ is not convex since there are matrices whose projections onto $S_{\leq r}$ are not single-valued. The multivaluedness of the projector can be clarified here in the case of Hermitian matrices. If there is an eigenvalue λ with eigenspace of dimension larger than one, and if this eigenvalue is the largest absolute one, then we can choose eigenvectors v, w in that eigenspace. Consequently, the matrix $\lambda v w^*$ is contained in the rank one projection of the matrix. In order to stay in the space of Hermitian matrices, we have to *choose* the eigenvectors to be equal up to a complex scalar of modulus one.

Example 8.2.4 (multi-valuedness of the projection). *Note that the projection in (8.4) is not necessarily single-valued since the rth largest singular value need not to be unique. The identity matrix in $\mathbb{R}^{n \times n}$ is the most trivial example for a matrix whose projection onto $S_{\leq r}$ is not unique. In fact, for matrices in $\mathbb{R}^{n \times n}$, the set $P_{S_{\leq r}} \operatorname{Id}_n$ may be homeomorphic to cubes. For example, consider the case of $\operatorname{Id}_2 \in \mathbb{R}^{2 \times 2}$. Then we obtain*

$$P_{S_{\leq 1}} \operatorname{Id}_2 = \left\{ \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) \end{pmatrix} \;\middle|\; \alpha \in [0, 2\pi) \right\}. \tag{8.5}$$

To see this, we note that, because the singular values of Id_2 are both equal to 1, the Frobenius distance of Id_2 to $S_{\leq 1}$ has to be 1. Let X be an arbitrary matrix in the set proposed in Equation (8.5). We compute the distance of Id_2 to X.

$$
\begin{aligned}
\|\operatorname{Id}_2 - X\|_F &= \left\| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) \end{pmatrix} \right\|_F \\
&= \sqrt{(1 - \cos^2(\alpha))^2 + 2\cos^2(\alpha)\sin^2(\alpha) + (1 - \sin^2(\alpha))^2} \\
&= \sqrt{1 - 2\cos^2(\alpha) + \cos^4(\alpha) + 2\cos^2(\alpha)\sin^2(\alpha) + 1 - 2\sin^2(\alpha) + \sin^4(\alpha)} \\
&= \sqrt{\cos^4(\alpha) + 2\cos^2(\alpha)\sin^2(\alpha) + \sin^4(\alpha)} \\
&= \sqrt{\left(\cos^2(\alpha) + \sin^2(\alpha)\right)^2} \\
&= \sqrt{1} = 1.
\end{aligned}
$$

In other words, the set in Equation (8.5) is the projection of Id_2 onto $S_{\leq 1}$.

Lemma 8.2.5. *For any* $A \in \mathbb{C}^{m \times n}$ *we have*

$$\left\| P_{S_{\leq r}} A \right\|_F \leq \|A\|_F .$$

We give two different proofs.

1. *Proof.* The squared Frobenius norm of a matrix is equal to the sum of its squared singular values. This gives us

$$\left\| P_{S_{\leq r}} A \right\|_F^2 = \sum_{j=1}^{r} \sigma_j^2 \leq \sum_{j=1}^{n} \sigma_j^2 = \|A\|_F^2 .$$

   $\square$

2. *Proof.* The set of rank constrained matrices is a closed cone (by Equation (8.3)). In other words, for every matrix of rank r, we know that scalar multiples of this matrix are of rank at most r as well. In general, for any x in some Euclidean space and for any closed cone K in that space, we know that the points x, $P_K x$ and 0 give a right triangle such that

$$\|x - 0\|^2 = \|x - P_K x\|^2 + \|P_K x - 0\|^2 .$$

   This gives us $\|x\| \geq \|P_K x\|$, and this finishes the proof. $\square$

We consider now the local regularity conditions on the set $S_{\leq r}$. The set $S_{\leq r}$ is at all points X with rank $(X) = r$ a smooth manifold.

Theorem 8.2.6 ((Lee, 2003, Example 8.14)). *The set $S_{=r}$ is an embedded submanifold of dimension $(m + n)r - r^2$ in the space $\mathbb{R}^{m \times n}$.*

Since we will analyze the alternating projections algorithm applied to the rank-constrained regime in the next chapter, we need the explicit form of the normal cone to $S_{\leq r}$. On the one hand, we can obtain the expression by application of Proposition 7.2.7. On the other hand, an alternative description has been given in (Luke, 2013).

Proposition 8.2.7 ((Luke, 2013, Proposition 3.6)). *Suppose $r \leq n$. Then the (Mordukhovich) normal cone to the set of matrices of rank less or equal to r at a point X is the set*

$$N_{S_{\leq r}}(X) = \{V \mid \mathrm{range}(X) \perp \mathrm{range}(V), \ \mathrm{rank}\,(V) \leq n - r\}. \qquad (8.6)$$

Remark 8.2.8. *We have to clarify the notation in Proposition 8.2.7. The original (Luke, 2013, Proposition 3.6) states that*

$$N_{S_{\leq r}}(X) = \{V \mid \mathrm{range}(X) \cap \mathrm{range}(V) = \{0\}, \ \mathrm{rank}\,(V) \leq n - r\}.$$

In the proof of (Luke, 2013, Proposition 3.6), it is actually shown that instead of $\mathrm{range}(X) \cap \mathrm{range}(V) = \{0\}$, we have

$$\mathrm{range}(X) \perp \mathrm{range}(V).$$

This is slightly stronger than the previous formulation.

Further, we remind the reader that the normal cone at the set of sparse vectors is given by (3.15). With the help of Proposition 7.2.7, we can give an alternative formulation.

Lemma 8.2.9. *Let* $\mathrm{rank}\,(X) = r$ *and let* $N_{S_{\leq r}}(X)$ *be defined via Proposition 8.2.7. Further, let* $U, W \in O(n)$ *such that* $U \,\mathrm{diag}\,(\sigma(X))\,W^\top = X$. *Then we have*

$$N_{S_{\leq r}}(X) = \mathrm{Stab}_{O(n)}(X)^2 \cdot \left(U \,\mathrm{diag}\,(N_{A_r}(\sigma(X)))\,W^\top \right). \tag{8.7}$$

Proof. Let $X \in S_{\leq r}$ with $\mathrm{rank}\,(X) = r$. Let $x := \sigma(X)$ be the vector of singular values of X. Then, by (3.15), we know that $N_{A_r}(x) = \mathrm{supp}(x)^\perp$. The claim can then be rewritten as

$$N_{S_{\leq r}}(X) = \mathrm{Stab}_{O(n)}(X)^2 \cdot \left(U \,\mathrm{diag}\,(\mathrm{supp}(\sigma(X)))^\perp W^\top \right).$$

Now let $V \in N_{S_{\leq r}}(X)$. We reformulate the condition in (8.6) as

$$\mathrm{range}(X) \subset \ker(V) \quad \text{and} \quad \mathrm{range}(V) \subset \ker(X).$$

We show that there exists $(O_1, O_2) \in \mathrm{Stab}_{O(n)}(X)$ such that $V = O_1 \,\mathrm{diag}\,(v)\,O_2^\top$ for some $v \in N_{A_r}(x)$.

Let $y \in \mathrm{range}(X)$ so that $Vx = 0$. Let $O_1 \Sigma_V O_2^\top = V$ be an SVD of V. We know that y is in the span of those right singular vectors in O_2 and those left singular vectors in O_1 belonging to zero singular values of V. A similar relation holds for $v \in \mathrm{range}(V)$ being in the span of the left singular vectors and right singular vectors of X belonging to the zero singular values of X. Hence, we can choose O_1 and O_2 such that we have

$$O_1 \,\mathrm{diag}\,(\sigma(X))\,O_2^\top = X \text{ and } O_1 \,\mathrm{diag}\,(P\sigma(V))\,O_2^\top \tag{8.8}$$

where P is a permutation matrix moving the nonzero entries of $\sigma(V)$ into $\mathrm{supp}(x)^\perp$. This is possible because of the rank assumption on V, i.e., $\mathrm{rank}\,(V) \leq n - r$. This gives us $N_{S_{\leq r}}(X) \subseteq \left(U \,\mathrm{diag}\,(N_{A_r}(\sigma(X)))\,W^\top \right)^{\mathrm{Stab}_{O(n)}(X)}$.

Now let $V \in \mathrm{Stab}_{O(n)}(X)^2 \cdot \left(U \,\mathrm{diag}\,(N_{A_r}(\sigma(X)))\,W^\top \right)$. Then there exist $O_1, O_2 \in O(n)$ such that

$$O_1 \,\mathrm{diag}\,(\sigma(X))\,O_2^\top = X \quad \text{and} \quad O_1 \,\mathrm{diag}\,(v)\,O_2^\top = V$$

where v is a vector containing the singular values of V. In other words,

$$\begin{aligned}
\mathrm{tr}(X^\top V) &= \mathrm{tr}(O_2 \,\mathrm{diag}\,(\sigma(X))\,O_1^\top O_1 \,\mathrm{diag}\,(v)\,O_2^\top) \\
&= \mathrm{tr}(O_2 \,\mathrm{diag}\,(\sigma(X))\,\mathrm{diag}\,(v)\,O_2^\top) = 0.
\end{aligned}$$

Hence, the range of V is a subset of the kernel of X and vice versa. The condition $\mathrm{rank}\,(V) \leq n - s$ follows from the condition $v \in N_{A_r}(\sigma(X))$. □

Proposition 8.2.10 (tangent space, (Vandereycken, 2013, Proposition 2.1)). *Let $S_{\leq r}$ be the set of matrices in $\mathbb{R}^{m \times n}$ with rank at most r, let $X \in S_{\leq r}$ with $\text{rank}(X) = r$, and let $U\Sigma V^\top$ be an SVD of X. Then the tangent space $T_{S_{\leq r}}(X)$ has the form $\{UA^\top + BV^\top \mid A, B \in \mathbb{R}^{m \times r}\}$. This is equivalent to*

$$T_{S_{\leq r}}\bar{x} := \left\{ \begin{pmatrix} U & U_\perp \end{pmatrix} \begin{pmatrix} \mathbb{R}^{r \times r} & \mathbb{R}^{r \times (n-r)} \\ \mathbb{R}^{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{pmatrix} \begin{pmatrix} V & V_\perp \end{pmatrix}^\top \right\} \tag{8.9}$$

where the matrices $U_\perp$ and $V_\perp$ are such that $\begin{pmatrix} U & U_\perp \end{pmatrix} \in O(m)$ and $\begin{pmatrix} V & V_\perp \end{pmatrix} \in O(n)$.

Remark 8.2.11. *(Vandereycken, 2013) uses the notation $\tilde{U}\tilde{\Sigma}\tilde{V}^\top = \bar{x}$ with $\tilde{U} \in \mathbb{R}^{m \times r}, \tilde{\Sigma} \in \mathbb{R}^{r \times r}, \tilde{V} \in \mathbb{R}^{n \times r}$ and $\tilde{U}, \tilde{V}$ with orthonormal columns.*

Example 8.2.12. *Now we consider a simple example with the assumption that $n > m > 2r$: Take $\bar{x} = \begin{pmatrix} Id_r & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} U & U_\perp \end{pmatrix} = Id_m$, and $\begin{pmatrix} V & V_\perp \end{pmatrix} = Id_n$. Now we take the (affine) subspace*

$$C := \bar{x} + \lambda \begin{pmatrix} 0_{r \times r} & Id_r & 0_{r \times (n-2r)} \\ Id_r & 0_{r \times r} & 0_{r \times (n-2r)} \\ 0_{(m-2r) \times r} & 0_{(m-2r) \times r} & 0_{(m-2r) \times (n-2r)} \end{pmatrix}, \lambda \in \mathbb{R}. \tag{8.10}$$

This set is clearly a subset of the affine space $\bar{x} + T_{S_{\leq r}}\bar{x}$. Further, any matrix $y \in C$ is of rank at most $2r$ while the only matrix in C with rank less than $2r$ is $\bar{x}$ itself. This is an example of an affine subspace that intersects $S_{\leq r}$ at only one point and which is tangential to $S_{\leq r}$. By (Hesse, 2014, Theorem 5.19 c)), since $\max \{ \langle u, v \rangle \mid u \in T_{S_{\leq r}}(\bar{x}) \cap \mathbb{B}, v \in T_C(\bar{x}) \cap \mathbb{B} \} = 1$, we cannot have a locally linearly regular intersection $C \cap S_{\leq r}$. See Figure 8.3 for an example of the resulting slow convergence of alternating projections between C and $S_{\leq r}$.

8.3 Lifted Sets: A Dictionary

In Chapter 10, we will study the behavior of the method of alternating projections where one of the sets is the set of matrices of rank one. As a preparation of this, we establish a link between constraint sets in $\mathbb{C}^n$ and constraint sets of matrices in $\mathbb{C}^{n \times n}$. The motivation in our case is the fact that we can reformulate the physical problem of phase retrieval in $\mathbb{C}^n$ in terms of minimizing the rank of a matrix over an affine subspace of matrices in $\mathbb{C}^{n \times n}$.

Definition 8.3.1. *Given a finite collection of real-valued functions $f_1, \ldots, f_m$ on $\mathbb{C}^n$, define a set of constraints*

$$C := \{x \in \mathbb{C}^n \mid f_1(x) = \cdots = f_m(x) = 0\}. \tag{8.11}$$

Secondly, let $F_1, \ldots, F_M$ be a collection of real-valued functions over $\mathbb{C}^{n \times n}$, and let

$$\mathcal{C} := \{X \in \mathbb{C}^{n \times n} \mid F_1(x) = \cdots = F_m(x) = 0\}. \tag{8.12}$$

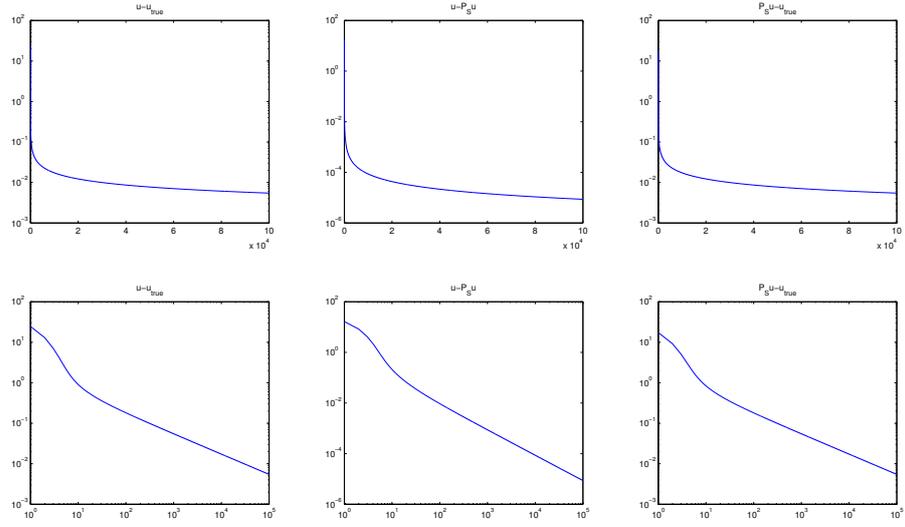*The set $\mathcal{C}$ is a lift of C if*

Figure 8.3: Convergence of alternating projections between the tangent space of $S_{\leq r}$ and $S_{\leq r}$ itself.

1. *for all Hermitian $X \in \mathcal{C} \cap S_{\leq 1}$ there exists $x \in C$ such that $xx^* = X$;*

2. *$xx^* \in \mathcal{C}$ for all $x \in C$.*

Remark 8.3.2. *For a given set $C \subset \mathbb{C}^n$, Definition 8.3.1 does not determine the lift of C uniquely. The idea is more to* translate *constraints, formulated in $\mathbb{C}^n$, into well-described constraints in the matrix space. Then an analysis of the intersection of the lift with the set of rank one matrices also gives further insight on C. In particular, for a subset $C \subset \mathbb{C}^n$, it would be fully sufficient to take $\mathcal{C} := \{xx^* \mid x \in C\}$. However, in order to benefit from additional structure of lifts of C, we can also take for example an affine hull of $\{xx^* \mid x \in C\}$.*

Proposition 8.3.3 (uniqueness up to a global phase)**.** *For all Hermitian $X \in \mathcal{C} \cap S_{\leq 1}$, the vector $x \in C$ such that $xx^* = X$ is determined up to global phase.*

Proof. If $X \in \mathcal{C} \cap S_{\leq 1}$ is Hermitian, then $X = xx^*$ for some $x \in \mathbb{C}^n$. Note that, for all $\theta \in [0, 2\pi]$, also $e^{i\theta} x$ satisfies

$$e^{i\theta} x \left(e^{i\theta} x \right)^* = x \, e^{i\theta} \, x \, e^{-i\theta} \, x^* = xx^* = X. \tag{8.13}$$

This means that from some rank one matrix xx^* the vector x is unique at least up to a global phase θ. Assume now that there exists $y \in \mathbb{C}^n$ such that $yy^* = xx^*$ and $y \neq e^{i\theta} x$ for all $\theta \in [0, 2\pi]$. Because $yy^* = xx^*$, we know that $x_j \overline{x_j} = y_j \overline{y_j}$ for all j. Hence, there exist $\theta_1, \ldots, \theta_n$ such that $y_j = e^{i\theta_j} x_j$ for all j. If $y \neq e^{i\theta} x$ for all $\theta \in [0, 2\pi]$, then there exists at least one k such that $\theta_k \neq \theta_j$ for all $j \neq k$. We look at one of those entries of yy^*

indexed by k and j with $\theta_j \neq \theta_k$:

$$yy^*_{jk} = y_j \overline{y_k} = \mathrm{e}^{\mathrm{i}\theta_j} x_j \, \mathrm{e}^{-\mathrm{i}\theta_k} \overline{x_k} = \mathrm{e}^{\mathrm{i}(\theta_j - \theta_k)} xx^*_{jk} \neq xx^*_{jk}. \tag{8.14}$$

This is a contradiction. Hence, there exists some $\theta \in [0, 2\pi]$ such that $y = \mathrm{e}^{\mathrm{i}\theta} x$. $\qquad \square$

8.3.1 Lifts of Linear Spaces and Cones

We focus now on affine subspaces and cones of $\mathbb{C}^n$. The first and easiest example would be the lift of an affine subspace of $\mathbb{C}^n$. Additionally, we consider two examples motivated by the physical application of phase retrieval that will be revisited in Definition 9.2.2.

Proposition 8.3.4 (lifts of affine spaces). *Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be a linear mapping and let $b \in \mathbb{C}^m$. Define the affine subspace*

$$B := \left\{ x \in \mathbb{C}^n \mid Ax = b \right\}. \tag{8.15}$$

Then the set

$$\mathcal{B} := \left\{ X \in \mathbb{C}^{n \times n} \mid AXA^* = bb^* \right\} \tag{8.16}$$

is an affine subspace, and it is a lift of B.

Proof. If there is some $x \in B$, then the rank one matrix xx^* satisfies the equation $Axx^*A^* = bb^*$.

The set $\mathcal{B}$ is an affine subspace of $\mathbb{C}^{n \times n}$ if for every $\lambda \in \mathbb{C}$ we have $\lambda x + (1 - \lambda)y \in \mathcal{B}$ for all $x, y \in \mathcal{B}$. Choose some arbitrary λ and some $X, Y \in \mathcal{B}$. Then

$$A(\lambda X + (1 - \lambda)Y)A^* = A\lambda XA^* + A(1 - \lambda)YA^*$$
$$= \lambda AXA^* + (1 - \lambda)AYA^* = bb^*.$$

$\square$

Lemma 8.3.5 (support constraints). *Let I be an index set in $\{1, \ldots, n\}$ and define*

$$C_s := \left\{ x \in \mathbb{C}^n \mid x_j = 0 \text{ if } j \notin I \right\}. \tag{8.17}$$

Then the set

$$\mathcal{C}_s := \left\{ X \in \mathbb{C}^{n \times n} \mid X_{jk} = 0 \text{ if } j \notin I \text{ or if } k \notin I \right\}$$

is a lift of C_s.

Proof. Let $X \in \mathcal{C}$ and $\mathrm{rank}(X) = 1$. Then X is Hermitian, and it can be written as some xx^* for $x \in \mathbb{C}^n$. Suppose there exists $j \notin I$. Then $X_{jk} = 0$ for all $1 \leq k \leq n$. If there exists k such that $X_{kk} \neq 0$, then this means that $x_j = 0$, and thus $x \in C_s$.

On the other hand, if $x \in C_s$, then, for all $j \notin I$, we have $x_j = 0$. This gives us $xx^*_{jk} = 0$ for all $1 \leq k \leq n$, which is all that is required to show that $xx^* \in \mathcal{C}_s$. $\qquad \square$

Proposition 8.3.6 (lift of a convex cone). *Let $A : \mathbb{C}^m \to \mathbb{C}^n$ be a linear mapping. Define the convex cone*

$$K = \{x \in \mathbb{C}^n \mid Ax \geq 0\}. \tag{8.18}$$

The set

$$\mathcal{K} := \{X \in \mathbb{C}^{n \times n} \mid AXA^* \geq 0\} \tag{8.19}$$

is a lift of K, and it is a convex cone.

Proof. Note that for every convex cone K we have $0 \in K$. This means that for every $x \in K$ we have $Axx^*A^* \geq 0$. On the other hand, if a positive semidefinite rank one matrix xx^* satisfies $Axx^*A^* \geq 0$, then either $x \in K$ or $-x \in K$. Let $X \in \tilde{K}$, then $AXA^* \geq 0$, and naturally, for all $\lambda \geq 0$, we have $A\lambda XA^* = \lambda AXA^* \geq 0$. On the other hand, let $\lambda \in [0,1]$ and X, Y. Then $A(\lambda X + (1-\lambda)Y)A^* = \lambda AXA^* + (1-\lambda)AYA^* \geq 0$. $\qquad\square$

Lemma 8.3.7 (positive real cone). *For*

$$C := \mathbb{R}^n_{\geq 0} = \{x \in \mathbb{C}^n \mid x \in \mathbb{R}^n, x_j \geq 0 \text{ for all } j\}, \tag{8.20}$$

the set $\mathcal{C} := \left\{X \in \mathbb{C}^{n \times n} \mid X \in \mathbb{R}^{n \times n}_{\geq 0}, X^* = X\right\}$ is a lift of C.

Proof. Let $X \in \mathcal{C}_s$ and $\operatorname{rank}(X) = 1$. Then X is Hermitian, and it can be written as some xx^* for $x \in \mathbb{C}^n$. Let j be such that $X_{jj} \neq 0$. If such a j does not exist, then X is the zero matrix. Choose $x_j = \sqrt{X_{jj}} > 0$. Because $x_j \overline{x_k} \in \mathbb{R}_{\geq 0}$ for all k, we conclude that $x_k \in \mathbb{R}_{\geq 0}$. $\qquad\square$

8.3.2 Quadratic Constraints

The examples so far gave subspaces as lifts of subspaces. An interesting example of the benefits of lifted sets is the following: suppose we are given an unitary mapping $A : \mathbb{C}^n \to \mathbb{C}^n$ and, for some $b \in \mathbb{R}^n_{\geq 0}$, the set

$$M := \left\{x \mid |\langle a_k, x\rangle|^2 = b_k \text{ for all } 1 \leq k \leq n\right\}, \tag{8.21}$$

where a_k are the complex conjugates of the rows of A. This is due to the fact that $\langle a_k, x\rangle = a_k^* x$. We reformulate this problem. We define for a vector $x \in \mathbb{C}^n$ the matrix $X := xx^*$, which is of rank one and positive semidefinite. Similarly, define for every $k = 1, \dots, m$ the matrix $A_k := a_k a_k^*$.

Proposition 8.3.8. *Define the mapping $\mathcal{A}_k$ via $\mathcal{A}_k(X) = \operatorname{Tr}\left(A_k^* X\right)$. The set*

$$\mathcal{B} := \{X \in \mathbb{C}^{n \times n} \mid \mathcal{A}_k(X) = b \text{ for all } 1 \leq k \leq n\} \tag{8.22}$$

is an affine subspace of $\mathbb{C}^{n \times n}$, and $\mathcal{B}$ is a lift of M.

Proof. The restrictions can be written as

$$|\langle a_k, x\rangle|^2 = \operatorname{Tr}\left(x^* a_k a_k^* x\right) = \operatorname{Tr}\left(a_k a_k^* xx^*\right) = \operatorname{Tr}\left(A_k X\right) = \mathcal{A}_k(X). \tag{8.23}$$

In other words, for every $x \in M$ there exists $X = xx^* \in \mathcal{B}$, and for every $X \in \mathcal{B} \cap S_{\leq 1}$ there exists $x \in \mathbb{C}^n$ such that $xx^* = X$. $\qquad\square$

If A is the discrete Fourier transform, then the above model can be applied to the phase retrieval problem (Chapter 9). The idea of reformulating quadratic constraints as linear equations in the matrix space comes from semidefinite programming. Its application to phase retrieval goes back to the article by Candès et. al in (Candès et al., 2011, Section 2.2). We discuss this matter in Section 10.1.

8.4 Second-Order Subdifferentials at the Set of Rank-Constrained Matrices
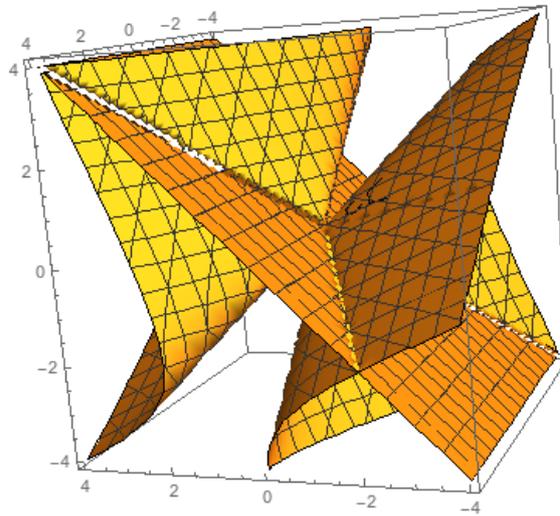
Figure 8.4: A section of the set of 5-by-5 matrices of rank 4. Displayed above is the zero
$$\text{set of } (x - z)(y - z)^2(x^2 - (y + z)^2 + x(y + 2z)) = \det \begin{pmatrix} x & y & z & x & z \\ y & x & z & y & x \\ z & z & y & z & y \\ x & y & z & z & x \\ z & x & y & x & z \end{pmatrix}.$$

In light of Section 3.4, we study the second-order subdifferential of the rank function in terms of coderivatives. This case is similar to the ℓ_0-function. In this section, we restrict ourselves to real-valued matrices.

We know that, by (Le, 2013, Proposition 2), based on (Lewis and Sendov, 2005), the

subdifferential of the rank function is given by

$$\partial \operatorname{rank}(A) = \ker(A) \otimes \ker(A^*). \tag{8.24}$$

In terms of graphs of functions, we obtain

$$\operatorname{gph}(\partial \operatorname{rank}(\cdot)) = \left\{ (A, Y) \in \mathbb{R}^{2m \times n} \mid Y \in \ker(A) \otimes \ker(A^*) \right\}. \tag{8.25}$$

If $\operatorname{rank}(A) = k$, then the set $\ker(A) \otimes \ker(A^*)$ is an $(n-k)(m-k)$-dimensional subspace of $\mathbb{R}^n \otimes \mathbb{R}^m$. The latter is canonically isomorphic to $\mathbb{R}^{m \times n}$, which means we can identify elements of $\ker(A) \otimes \ker(A^*)$ with $m \times n$-matrices over $\mathbb{R}$.

With the canonical isomorphism $\mathbb{R}^n \otimes \mathbb{R}^m \to \mathbb{R}^{m \times n}$, we conclude that $\ker(A) \otimes \ker(A^*)$ consists of linear combinations of matrices of the form

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \otimes \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} a_1 b_1 & \dots & a_n b_1 \\ \vdots & \ddots & \vdots \\ a_1 b_m & \dots & a_n b_m \end{pmatrix} =: M, \tag{8.26}$$

where

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \ker(A) \text{ and } \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \ker(A^*). \tag{8.27}$$

Then we have

$$A^* M = 0, \quad M A^* = 0, \quad A M^* = 0 \quad \text{for all } M \in \ker(A) \otimes \ker(A^*) \tag{8.28}$$

since $a \in \ker(A), b \in \ker(A^*)$. But the span of the columns of A^* is in the kernel of M, which means that the rank of M is less or equal to $m - k$. By that result, we can see that elements of $\operatorname{gph}(\partial \operatorname{rank}(\cdot))$ can be identified with matrices of $\mathbb{R}^{2m \times n}$ with rank less or equal to m.

To get the coderivative of the subdifferential of the rank function, by Definition 3.27, it remains to formulate the normal cone to the graph of the subdifferential of the rank function. Let $(A, M) \in S_{\leq m} \subset \mathbb{R}^{2m \times n}$ be such that $\operatorname{rank}(A) = k, \operatorname{rank}(M) = m - k$, and the rows of M are in the kernel of A. Then, for some matrix $X \in \mathbb{R}^{m \times n}$, we have the second-order subdifferential of the rank function given as

$$\partial^2 \operatorname{rank}(\cdot)(A, M)(X) = \left\{ Y \in \mathbb{R}^{m \times n} \mid (X, -Y) \in N_{S_{\leq m}}(A, M) \subset \mathbb{R}^{2m \times n} \right\}$$

$$= \begin{cases} Y \in S_{\leq m - \operatorname{rank}(X)} & \text{if } X \in \operatorname{range}(M), \\ \varnothing & \text{else.} \end{cases} \tag{8.29}$$

Remark 8.4.1. *Because $(X, -Y) \in N_{S_{\leq m}}(A, M) \subset \mathbb{R}^{2m \times n}$, the range of the matrix $(X^* Y^*)$ has trivial intersection with the range of $(A^* M^*)$. If we formulate a Newton-type method for*

the rank function as in Section 3.4.3, we would again see that every point is a fixed point of this Newton-type method. As with the ℓ_0-function, every matrix $X \in \mathbb{R}^{m \times n}$ is a local minimizer of the rank function.

9 Phase Retrieval

This short chapter briefly introduces the phase retrieval problem. We will establish its connections to projection methods and regularity conditions arising in feasibility problems. We refer the reader to (Luke et al., 2002) for a more detailed survey on phase retrieval and diffraction imaging. As it is stated in (Luke et al., 2002), abstractly, the phase retrieval problem reads as follows: Let $\mathcal{F}$ be the Fourier transform. Suppose we are given functions $a, b : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$, find the function $x : \mathbb{R}^2 \to \mathbb{C}$ satisfying $|\mathcal{F}x| = b$ and $|x| = a$.

This problem arises for example in microscopy (Drenth et al., 1975), (Miao et al., 1999), optical design (Farn, 1991), astronomy (Luke et al., 2002), holography (Fienup, 1980), (Bartels et al., 2015), and crystallography (Millane, 1990). Note that phase retrieval is an ill-posed inverse problem. To solve it, we usually incorporate a priori assumptions on the setup. One of these assumptions is the already mentioned amplitude a. We will name the most prominent examples of these assumptions in this chapter.

9.1 Problem Formulation

We first state the continuous problem of phase retrieval for the sake of completeness. At this point, we cite (Fienup and Wackerman, 1986) for the abstract problem formulation.

"Given the modulus $|F(u, v)|$ of the Fourier transform

$$
\begin{aligned}
F(u, v) &= |F(u, v)| \, e^{i \, \psi(u, v)} \\
&= \mathcal{F}(f(u, v)) \\
&= \int \int_{-\infty}^{\infty} f(u, v) \, e^{2\pi i (ux + vy)} \, d\,x \, d\,y
\end{aligned}
$$

of an object $f(x, y)$,

$$\text{reconstruct the object} \quad f(x, y) \tag{9.1}$$

or, equivalently, reconstruct the Fourier phase $\psi(u, v)$."[1] Note that, from now on, we just consider the discretized version of phase retrieval. The Fourier transform $\mathcal{F}$ becomes the discrete Fourier transform $F \in \mathbb{C}^{n \times n}$, x is a vector in $\mathbb{C}^n$, and $a, b \in \mathbb{R}_{\geq 0}^n$. One question arising in this context is the one for uniqueness of this inverse problem. If $|Fx| = b$, then we can multiply x by a factor $e^{i \, \varphi}$ for some $\varphi \in \mathbb{R}$ since F is linear. The implication is $|F(e^{i \, \varphi} x)| = |e^{i \, \varphi} Fx| = |Fx| = b$. This ambiguity is called *shift by a global phase*. We refer to (Hayes, 1982) for additional issues concerning uniqueness.

---

[1] (Fienup and Wackerman, 1986), p. 1897, Section 1, l.3–8

9.2 Phase Retrieval in Terms of Feasibility

In order to apply the theory shown and developed in this thesis to Problem (9.1), we formulate it as a feasibility problem. As a preparation, we introduce the sets involved in the upcoming formulation (9.4).

Definition 9.2.1 (magnitude constraint). *The measured data is by (9.1) given by $|Fx|^2 = b$. We define the set of all complex vectors x satisfying (9.1),*

$$M := \left\{ x \in \mathbb{C}^n \mid |Fx|^2 = b \right\}. \tag{9.2}$$

We refer to the set M as the magnitude constraint set.

Definition 9.2.2 (constraint sets). *Let $I \subset \{1, \dots, n\}$ be a set of indices and let $a \in \mathbb{R}^n_{\geq 0}$. We define the following subsets of $\mathbb{C}^n$:*

$$
\begin{aligned}
\textit{Support constraint:} && S &:= \left\{ x \in \mathbb{C}^n \mid x_j = 0 \text{ if } j \notin I \right\}. \\[4pt]
\textit{Support and real-valued constraint:} && S_* &:= \left\{ x \in \mathbb{R}^n \mid x_j = 0 \text{ if } j \notin I \right\}. \\[4pt]
\textit{Support and positive constraint:} && S_+ &:= \left\{ x \in \mathbb{R}^n_{\geq 0} \mid x_j = 0 \text{ if } j \notin I \right\}. \\[4pt]
\textit{Amplitude constraint:} && A &:= \left\{ x \in \mathbb{C}^n \mid |x| = a \right\}.
\end{aligned}
\tag{9.3}
$$

The phase retrieval problem can now be written as a feasibility problem. Let C be one of the constraint sets in (9.3), representing the a priori assumptions. Then we formulate

$$\text{Find } \bar{x} \in M \cap C. \tag{9.4}$$

It is self-evident that it is possible to use projection methods to solve (9.4). In order to apply these, we give the formulae of the projections onto the sets in (9.2) and (9.3).

Lemma 9.2.3 (magnitude projection, (Luke et al., 2002, Example 3.6)). *Let F be the discrete Fourier transform. For $x \in \mathbb{C}^n$, the projection onto the set M is given by*

$$P_M x = F^{-1} \left(b \odot Fx \odot \tfrac{1}{|Fx|} \right) \tag{9.5}$$

where $\odot$ denotes the entrywise product, and where $\frac{1}{|Fx|}$ is the vector with entries $\frac{1}{|Fx|_j}$.

Lemma 9.2.4 (projections onto constraints (Luke et al., 2002, Example 3.14), (Burke and Luke, 2003, Equation 9)). *The projections onto the constraint sets in (9.3) are given by*

$$(P_S x)_j = \begin{cases} x_j & \text{if } j \in I, \\ 0, & \text{otherwise,} \end{cases} \tag{9.6}$$

$$(P_{S_*} x)_j = \begin{cases} \operatorname{Re}(x_j) & \text{if } j \in I, \\ 0, & \text{otherwise,} \end{cases} \tag{9.7}$$

$$(P_{S_+}x)_j = \begin{cases} \max\{\mathrm{Re}(x_j), 0\} & \textit{if } j \in I, \\ 0, & \textit{otherwise,} \end{cases} \tag{9.8}$$

$$(P_A x)_j = \begin{cases} \frac{a_j x_j}{|x_j|} & \textit{if } x_j \neq 0, \\ \{e^{\mathrm{i}\varphi} a_j \mid \phi \in [0, 2\pi]\} & \textit{if } x_j = 0. \end{cases} \tag{9.9}$$

Further, the sets S, S_*, and S_+ are convex. The constraints S and S_* are linear subspaces of $\mathbb{C}^n$.

Lemma 9.2.5. *Let $x \mapsto P_M x$ be the magnitude projection. Then, for all $x \in \mathbb{C}^n$, we have $x^* P_M x \in \mathbb{R}$.*

Proof. Let $F \in \mathbb{C}^{n \times n}$ be the unitary discrete Fourier transform and let $b \in \mathbb{R}^n_+$ be the vector of magnitude measurements. Define $D := \mathrm{diag}\,(b)$. The operator P_M is given by

$$P_M x = F^* z, \text{ where } z \text{ is such that } z_j = \begin{cases} b_j \frac{(Fx)_j}{|(Fx)_j|} & \text{if } (Fx)_j \neq 0, \\ b_j & \text{otherwise.} \end{cases} \tag{9.10}$$

For a fixed $x \in \mathbb{C}^n$, we can write the projection onto the magnitude set M as

$$P_M x = F^* D F x, \tag{9.11}$$

which is just the matrix mapping x to $P_M x$. Because b consists of positive entries, the matrix D is positive definite, and hence the matrix $F^* D F$ is Hermitian and positive definite. This means that $x^* F^* D F x \geq 0$ with equality if and only if $x = 0$. This proves the claim. $\square$

9.3 Algorithms

We present a selection of methods to solve Problem (9.1).

Algorithm 9.3.1 (Gerchberg-Saxton (Gerchberg and Saxton, 1972, Figure 1)). *Let A be as defined in (9.3) and let M be the magnitude constraint (9.2). For a given initial point u^0, define the Gerchberg-Saxton algorithm by*

$$u^{k+1} \in P_A P_M u^k. \tag{9.12}$$

Algorithm 9.3.2 (Error Reduction (Fienup, 1982, Section II.)). *Let S be as defined in (9.3) and let M be the magnitude constraint (9.2). For a given initial point u^0, define the Error Reduction algorithm by*

$$u^{k+1} \in P_S P_M u^k. \tag{9.13}$$

Remark 9.3.3. *Note that Gerchberg-Saxton and Error Reduction are nothing but alternating projections between different constraint sets. The formulation of the Gerchberg-Saxton algorithm as a projection algorithm has been pointed out in (Levi and Stark, 1984).*

Algorithm 9.3.4 (HIO (Fienup, 1982, Section V.)). *Let I be an index set as in Definition 9.2.2, let u^0 be a given initial point, and let $\beta \in (0, 1]$. The Hybrid Input-Output (HIO) algorithm is given by*

$$\left(u^{k+1} \right)_j = \begin{cases} (P_M u^k)_j & \text{if } j \in I, \\ u_j^k - \beta (P_M u^k)_j & \text{if } j \notin I. \end{cases} \tag{9.14}$$

Algorithm 9.3.5 (DR-λ (Luke, 2005, Section 2.3)). *Let S_+ be as defined in (9.3), let M be the magnitude constraint (9.2), and let $\lambda \in (0, 1]$. For a given initial point u^0, define the DR-λ algorithm via*

$$u^{k+1} \in \frac{\lambda}{2} \left(R_{S_+} R_M + \text{Id} \right) u^k + (1 - \lambda) P_M u^k. \tag{9.15}$$

Remark 9.3.6. *In (Bauschke et al., 2002, Observation 5.10) it was shown that, for $\beta = 1$, the HIO is equivalent to the Douglas-Rachford algorithm (Definition 4.3.1).*

The DR-λ algorithm was introduced in (Luke, 2005) as Relaxed Averaged Alternating Reflections algorithm (RAAR). It is a relaxation of the Douglas-Rachford. The purpose of this relaxation lies in the behavior of Douglas-Rachford in case of infeasibility with convex sets: while the Douglas-Rachford iteration diverges in gap direction, the DR-λ has a dampening term to avoid this divergence.

10 Rank Minimization

A generalization of the problem of finding a sparsest vector satisfying an underdetermined system of linear equations is that of finding a matrix of least possible rank satisfying an underdetermined system of linear equations. Let $m, n \in \mathbb{N}$ be arbitrary and let $r \in \mathbb{N}$ be such that $0 \leq r \leq \min\{m, n\}$. Without loss of generality, we assume that $m \leq n$.

We define the main problem for this chapter,

$$\text{find } \bar{x} \in S_{\leq r} \cap \mathcal{B} \tag{10.1}$$

where $S_{\leq r}$ is as defined in (8.1), and $\mathcal{B}$ is an affine subspace of matrices. The problem will be denoted by affine-rank-constrained feasibility.

10.1 Motivation: Phase Lift

As a motivation for studying projection methods in a space of matrices, we consider a reformulation of the phase retrieval problem (Problem (9.1)) in terms of minimizing the rank of matrices with respect to a set of linear equations. The reformulation involves lifts of constraint sets as in Definition 8.3.1. We refer to this technique by the term *phase lift*. The idea is to go from the feasibility problem with quadratic constraints in a vector space to a feasibility problem with the set of matrices of rank one and with a linear contraint as in Equation (8.22). This approach goes back to (Candès et al., 2011, Section 2.2). We assume that there exists a signal $\bar{x} \in M$ where

$$M := \left\{ x \in \mathbb{C}^n \mid |\langle a_k, \bar{x} \rangle|^2 = b_k, k = 1, \ldots, n \right\} \tag{10.2}$$

is the set of magnitude constraints in (9.2). The vectors a_k are complex conjugate rows of the discrete Fourier transform, i.e., the entries $a_k(t)$ are proportional to terms of the form $w(t)e^{i2\pi\langle \omega_k, t \rangle}$. Following (Candès et al., 2011, Section 2.2), we define for a vector $x \in \mathbb{C}^n$ the matrix $X := xx^*$, which is of rank one and positive semidefinite. Similarly, define for every $k = 1, \ldots, m$ the matrix $A_k := a_k a_k^*$. Define, similarly to Proposition 8.3.8, the mapping $\mathcal{A}_k$ via $\mathcal{A}_k(X) = \text{Tr}\left(A_k^* X\right)$. Then, with $X = xx^*$, the restrictions in M can be written as

$$|\langle a_k, x \rangle|^2 = \text{Tr}\left(x^* a_k a_k^* x\right) = \text{Tr}\left(a_k a_k^* xx^*\right) = \text{Tr}\left(A_k X\right) = \mathcal{A}_k(X). \tag{10.3}$$

This results in a lift of the magnitude constraints as in (8.22):

$$\mathcal{B} := \left\{ X \in \mathbb{C}^{n \times n} \mid \mathcal{A}_k(X) = b \text{ for all } 1 \leq k \leq n \right\}. \tag{10.4}$$

The goal is now to find a positive semidefinite Hermitian matrix xx^* of rank one in the affine subspace $\mathcal{B}$. Then, by Proposition 8.3.3, the vector x can be determined uniquely up to a global factor $e^{2\pi i}$.

10.2 Projectors and their Compositions in Matrix Spaces

In Section 10.3, we will analyze a sequence of matrices generated by an alternating projections operator in the space $\mathbb{C}^{n \times n}$. As a preparation, we derive a collection of properties of the projectors involved, namely, the operators $P_{\mathcal{B}}, P_{S_{\leq 1}}$, and their composition $P_{S_{\leq 1}} P_{\mathcal{B}}$. We restrict ourselves to Hermitian matrices. At points where it is necessary, we mention why this restriction is still justified.

Especially when projecting onto the set $S_{\leq 1}$, we have to solve eigenvalue-eigenvector problems motivated by Proposition 8.2.3. Because of that, we recall several results from matrix analysis involving eigenvalues of matrices.

10.2.1 Projecting onto Lifts of Quadratic Constraints

First, we give an explicit formula for the projection onto $\mathcal{B}$ in terms of the unitary matrix A that we used to define $\mathcal{B}$. In the following, we assume that

$$a_k \text{ are the complex conjugated and transposed rows of } A. \tag{10.5}$$

We choose this way of writing the rows of A because with this notation the entries of the matrix vector product Ax have the shape $(Ax)_k = a_k^* x$. As in (8.22), we define the rank one matrix $A_k := a_k a_k^*$ and remember that

$$\mathcal{B} = \left\{ X \in \mathbb{C}^{n \times n} \mid \mathcal{A}_k(X) = b \text{ for all } 1 \leq k \leq n \right\}. \tag{10.6}$$

Because $\mathcal{B}$ is an affine subspace, there exists a linear mapping $\mathcal{A} : \mathbb{C}^{n \times n} \to \mathbb{R}^n$ such that

$$\mathcal{B} = \left\{ X \in \mathbb{C}^{n \times n} \mid \mathcal{A}(X) = b \right\}. \tag{10.7}$$

To compute the projection $P_{\mathcal{B}} X$ for arbitrary $X \in \mathbb{C}^{n \times n}$, we can use (3.16) and obtain

$$P_{\mathcal{B}} X = X - \mathcal{A}^* (\mathcal{A}\mathcal{A}^*)^{-1} (\mathcal{A}(X) - b). \tag{10.8}$$

The aim is to give explicit formulae for the expressions $\mathcal{A}, \mathcal{A}^*$, and $(\mathcal{A}\mathcal{A}^*)^{-1}$.

Proposition 10.2.1 (linear mapping in the lifted space)**.** *For the affine subspace $\mathcal{B}$ as in (10.7), the following holds for the linear mapping $\mathcal{A}$ and $\mathcal{B}$:*

1. $\mathcal{A}(X) = \begin{pmatrix} \mathcal{A}_1(X) \\ \vdots \\ \mathcal{A}_n(X) \end{pmatrix} = \begin{pmatrix} \mathrm{Tr}(A_1 X) \\ \vdots \\ \mathrm{Tr}(A_n X) \end{pmatrix}.$

2. $\mathcal{A}^*(x) = A^* \mathrm{diag}\,(x)\, A.$

3. $(\mathcal{A}\mathcal{A}^*)^{-1} = \mathrm{Id}_n.$

Proof.

1. This follows immediately from the fact that the space $\mathcal{B}$ is defined as all X satisfying $\mathrm{Tr}(A_j X) = b_j$ for all j. The composition of linear mappings, in this case the trace and multiplication with A_j, is again a linear mapping. This proves the claim.

2. The linear map $\mathcal{A}^*$ can be defined via the inner product: it is the linear map satisfying

$$w^* \mathcal{A}(X) = \langle w, \mathcal{A}(X) \rangle = \langle \mathcal{A}^*(w), X \rangle = \mathrm{Tr}\left((\mathcal{A}^*(w))^* X \right) \tag{10.9}$$

for all $w \in \mathbb{C}^n$ and for all $X \in \mathbb{C}^{n \times n}$.

We use for some arbitrary $w \in \mathbb{C}^n$ the identity (10.9) to determine the action of $\mathcal{A}^*$ onto w. First, we note

$$\mathcal{A}(X) = \begin{pmatrix} \mathrm{Tr}(a_1(a_1)^* X) \\ \dots \\ \mathrm{Tr}(a_n(a_n)^* X) \end{pmatrix}, \tag{10.10}$$

and further we have

$$a_k(a_k)^* = \begin{pmatrix} a_k^1 \overline{a_k^1} & \dots & a_k^1 \overline{a_k^n} \\ \vdots & \ddots & \vdots \\ a_k^n \overline{a_k^1} & \dots & a_k^n \overline{a_k^n} \end{pmatrix}. \tag{10.11}$$

Then

$$w^* \mathcal{A}(X) = \sum_{k=1}^n \overline{w}_k Tr(a_k(a_k)^* X) = \sum_{k=1}^n \overline{w}_k \sum_{j=1}^n (a_k(a_k)^* X)_{jj} \tag{10.12}$$

$$= \sum_{k=1}^n \overline{w}_k \sum_{j=1}^n \sum_{i=1}^n (a_k(a_k)^*)_{ji} X_{ij} \tag{10.13}$$

$$= \sum_{k=1}^n \overline{w}_k \sum_{j=1}^n \sum_{i=1}^n a_k^j \overline{a_k^i} X_{ij} \tag{10.14}$$

$$= \sum_{j=1}^n \sum_{i=1}^n \sum_{k=1}^n \overline{w}_k a_k^j \overline{a_k^i} X_{ij} = \mathrm{Tr}\left((\mathcal{A}^*(w))^* X \right) \tag{10.15}$$

The last term shows that $(\mathcal{A}^*(w))^*$ is of the shape

$$(\mathcal{A}^*(w))^* = \begin{pmatrix} \sum_{k=1}^n \overline{w}_k a_k^1 \overline{a_k^1} & \cdots & \sum_{k=1}^n \overline{w}_k a_k^1 \overline{a_k^n} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^n \overline{w}_k a_k^n \overline{a_k^1} & \cdots & \sum_{k=1}^n \overline{w}_k a_k^n \overline{a_k^n} \end{pmatrix}, \tag{10.16}$$

which gives us (the a_k^j are just complex numbers!)

$$\mathcal{A}^*(w) = \begin{pmatrix} \sum_{k=1}^n w_k a_k^1 \overline{a_k^1} & \cdots & \sum_{k=1}^n w_k a_k^1 \overline{a_k^n} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^n w_k a_k^n \overline{a_k^1} & \cdots & \sum_{k=1}^n w_k a_k^n \overline{a_k^n} \end{pmatrix}. \tag{10.17}$$

The above is equal to

$$\begin{aligned} \mathcal{A}^*(w) &= \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} w_1 \overline{a_1^1} & \cdots & w_1 \overline{a_1^n} \\ \vdots & \ddots & \vdots \\ w_n \overline{a_n^1} & \cdots & w_n \overline{a_n^n} \end{pmatrix} \\ &= \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix} \begin{pmatrix} \overline{a_1^1} & \cdots & \overline{a_1^n} \\ \vdots & \ddots & \vdots \\ \overline{a_n^1} & \cdots & \overline{a_n^n} \end{pmatrix} \\ &= A^* \text{diag}(w) A. \end{aligned}$$

3. First, we look for some $w \in \mathbb{C}^n$ at

$$\mathcal{A}\mathcal{A}^*(w) = \begin{pmatrix} \text{Tr}(a_1(a_1)^* \mathcal{A}^*(w)) \\ \cdots \\ \text{Tr}(a_n(a_n)^* \mathcal{A}^*(w)) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \sum_{i=1}^n a_1^j \overline{a_1^i} \sum_{k=1}^n w_k a_k^i \overline{a_k^j} \\ \vdots \\ \sum_{j=1}^n \sum_{i=1}^n a_n^j \overline{a_n^i} \sum_{k=1}^n w_k a_k^i \overline{a_k^j} \end{pmatrix}. \tag{10.18}$$

Now we look at the first entry of the resulting vector. The expression

$$\sum_{j=1}^n \sum_{i=1}^n a_1^j \overline{a_1^i} \sum_{k=1}^n w_k a_k^i \overline{a_k^j}$$

(remember, $(a_l)^* a_k = \sum_{i=1}^n a_k^i \overline{a_l^i} = \delta_{kl}$) turns to

$$\sum_{j=1}^n \sum_{i=1}^n a_1^j \overline{a_1^i} \sum_{k=1}^n w_k a_k^i \overline{a_k^j} = \sum_{k=1}^n w_k \sum_{j=1}^n a_1^j \overline{a_k^j} \sum_{i=1}^n a_k^i \overline{a_1^i} = \sum_{k=1}^n w_k \delta_{1k}^2 = w_1.$$

This means that

$$\mathcal{A}\mathcal{A}^*(w) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = w \tag{10.19}$$

for all $w \in \mathbb{C}^n$, and hence $\mathcal{A}\mathcal{A}^* = I_n$.

$\square$

Let us compute the expression $X - \mathcal{A}^*\left(\mathcal{A}\mathcal{A}^*\right)^{-1}\left(\mathcal{A}(X) - b\right)$ explicitly.

Proposition 10.2.2 (projection onto the affine subspace). *We have*

$$P_{\mathcal{B}}X = X - A^* \begin{pmatrix} (a_1)^* X a_1 - b_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (a_n)^* X a_n - b_n \end{pmatrix} A \tag{10.20}$$

for all $X \in \mathbb{C}^{n \times n}$.

Proof. The projection is given by

$$P_{\mathcal{B}}X = X - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\left(\mathcal{A}(X) - b\right). \tag{10.21}$$

From Proposition 10.2.1, 3., we know that we just have to compute $\mathcal{A}^*\left(\mathcal{A}(X) - b\right)$, which is, again by Proposition 10.2.1, given by

$$\mathcal{A}^*(\mathcal{A}(X) - b) = A^* \operatorname{diag}\left(\mathcal{A}(X) - b\right) A \tag{10.22}$$

$$= A^* \begin{pmatrix} \operatorname{Tr}(a_1(a_1)^* X) - b_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \operatorname{Tr}(a_n(a_n)^* X) - b_n \end{pmatrix} A \tag{10.23}$$

$$= A^* \begin{pmatrix} (a_1)^* X a_1 - b_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (a_n)^* X a_n - b_n \end{pmatrix} A. \tag{10.24}$$

$\square$

Proposition 10.2.3. *Given a Hermitian matrix $X \in \mathbb{C}^{n \times n}$, there exists a rank-one matrix $X_1 \in P_{S_{\leq 1}} P_{\mathcal{B}} X$ such that X_1 is Hermitian.*

Proof. By Proposition 10.2.2, the matrix $P_{\mathcal{B}}X$ is given by

$$X - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\left(\mathcal{A}(X) - b\right) = X - A^* \operatorname{diag}\left(v\right) A$$

where $v = \left(\operatorname{Tr}(a_1(a_1)^* X) - b_1, \dots, \operatorname{Tr}(a_n(a_n)^* X) - b_n\right)^\top$. Since the sum of Hermitian matrices is Hermitian again, we conclude that $P_{\mathcal{B}}X$ is Hermitian.

By Proposition 8.2.3, the projection $P_{S_{\leq 1}} P_{\mathcal{B}} X$ can be chosen as a Hermitian matrix again. This finishes the proof. $\square$

10.2.2 Rotating the Space of Complex Matrices

For our further analysis, we simplify the notation by applying a global change of basis to the space $\mathbb{C}^{n \times n}$. Instead of finding an element $x \in \mathbb{C}^n$ satisfying $a_j^* x x^* a_j = b_j$, as in Equation (10.4), we rotate the space of $n \times n$ complex matrices such that the problem formulation becomes

$$\text{find } \bar{x} \in \left\{ x \in \mathbb{C}^n \;\middle|\; e_j^* x x^* e_j = b_j, \; 1 \leq j \leq n \right\}, \tag{10.25}$$

where e_j is the jth standard unit vector. This is equivalent to

$$\text{find } \bar{x} \in \left\{ x \in \mathbb{C}^n \;\middle|\; x_j \bar{x}_j = b_j, \; 1 \leq j \leq n \right\}. \tag{10.26}$$

In other words, we have the restriction that the diagonal elements of xx^* have to be fixed. Define the affine subspace

$$\mathcal{B}_{\text{rot}} := \left\{ X \in \mathbb{C}^{n \times n} \;\middle|\; X_{jj} = b_j, \; 1 \leq j \leq n \right\}. \tag{10.27}$$

We verify the above statement by a proposition.

Proposition 10.2.4. *The linear mapping Ψ rotating the space such that $\Psi(a_k a_j^*) = e_k e_j^*$ for all k, j, satisfies*

$$\Psi \left\{ X \in \mathbb{C}^{n \times n} \;\middle|\; a_j^* X a_j = b_j, \; 1 \leq j \leq n \right\} = \left\{ X \in \mathbb{C}^{n \times n} \;\middle|\; X_{jj} = b_j, \; 1 \leq j \leq n \right\},$$

and it leaves the set of positive semidefinite rank one matrices invariant. The mapping Ψ is unitary.

Proof. We show that for all $X, Y \in \mathbb{C}^{n \times n}$ we have $\langle \Psi(X), \Psi(Y) \rangle = \langle X, Y \rangle$. First, note that the a_j form an orthonormal system of vectors as well as the e_j. Then, we have

$$\text{Tr}(a_i a_j^* a_k a_l^*) = a_l^* a_i a_j^* a_k = \begin{cases} 1 & \text{if } k = j \text{ and } i = l, \\ 0 & \text{otherwise.} \end{cases}$$

The same applies for the matrices $e_k e_j^*$. Hence, the matrices $\{a_k a_j^*\}_{k,j=1}^n$ are an orthonormal system of matrices. If Ψ maps the matrix $e_k e_j^*$ onto $a_k a_j^*$, then Ψ maps an orthonormal set of matrices onto an orthonormal set of matrices. We can write the matrices X and Y as linear combinations of the matrices $a_k a_j^*$, i.e., $X = \sum_{k,j=1}^n x_{kj} a_k a_j^*, Y = \sum_{k,j=1}^n y_{kj} a_k a_j^*$. Then we get

$$\text{Tr}(X^* Y) = \text{Tr}\left(\left(\sum_{k,j=1}^n x_{k,j} a_k a_j^* \right)^* \left(\sum_{k,j=1}^n y_{kj} a_k a_j^* \right) \right) = \sum_{k,j=1}^n \overline{x_{kj}} y_{kj}. \tag{10.28}$$

On the other hand,

$$\text{Tr}\left(\Psi(X)^*\Psi(Y)\right) = \text{Tr}\left(\left(\sum_{k,j=1}^{n} x_{kj}\Psi(a_k a_j^*)\right)^* \left(\sum_{k,j=1}^{n} y_{kj}\Psi(a_k a_j^*)\right)\right)$$

$$= \text{Tr}\left(\left(\sum_{k,j=1}^{n} x_{kj}e_k e_j^*\right)^* \left(\sum_{k,j=1}^{n} y_{ij}e_k e_j^*\right)\right) = \sum_{k,j=1}^{n} \overline{x_{kj}}y_{kj}.$$

This shows that Ψ is unitary. Now we have to show that the set of positive semidefinite rank one matrices is invariant under the action of Ψ. Given some vector $x \in \mathbb{C}^n$, we can write x as a linear combination of the basis vectors, namely $x = \sum_{k=1}^{n} x^k e_k$ and $x = \sum_{k=1}^{n} \tilde{x}^k a_k$. Then we obtain

$$xx^* = \left(\sum_{k=1}^{n} \tilde{x}^k a_k\right)\left(\sum_{k=1}^{n} \tilde{x}^k a_k\right)^* = \sum_{k,j=1}^{n} \tilde{x}^k\overline{\tilde{x}^j}a_k a_j^*.$$

$$\Rightarrow \Psi(xx^*) = \sum_{k,j=1}^{n} \tilde{x}^k\overline{\tilde{x}^j}\Psi(a_k a_j^*) = \sum_{k,j=1}^{n} \tilde{x}^k\overline{\tilde{x}^j}e_k e_j^*$$

$$= \left(\sum_{k=1}^{n} \tilde{x}^k e_k\right)\left(\sum_{k=1}^{n} \tilde{x}^k e_k\right)^*,$$

which is again a positive semidefinite rank one matrix. Finally, we have to show that

$$\Psi\left\{X \in \mathbb{C}^{n \times n} \;\middle|\; a_j^* Z a_j = b_j \;\forall j = 1, \ldots, n\right\} = \left\{X \in \mathbb{C}^{n \times n} \;\middle|\; X_{jj} = b_j \;\forall j = 1, \ldots, n\right\}.$$

If $X \in \left\{X \in \mathbb{C}^{n \times n} \;\middle|\; a_j^* X a_j = b_j \;\forall j = 1, \ldots, n\right\}$, then $X = \sum_{k,j} x_{kj}a_k a_j^*$ so that (remember that the set $\{a_1, \ldots, a_n\}$ is an orthonormal system of vectors.)

$$a_k^* X a_k = \sum_{k,j} x_{kj}a_k^* a_j a_j^* a_k = x_{kk} = b_k.$$

Then, for $\Psi(Z)$, we have

$$e_l^* \Psi(X) e_l = \sum_{k,j} x_{kj}e_l^* \Psi(a_k a_j^*) e_l = \sum_{k,j} x_{kj}e_l^* e_k e_j^* e_l = x_{ll}.$$

So $\Psi(X) \in \left\{X \in \mathbb{C}^{n \times n} \;\middle|\; Z_{jj} = b_j \;\forall j = 1, \ldots, n\right\}$. This shows the claim because Ψ is a unitary, and hence bijective, linear mapping. $\qquad\square$

Remark 10.2.5. *The above proposition is not only valid for the columns of the unitary discrete Fourier transform but also for columns of any unitary matrix U. This means that we are able to analyze the alternating projections algorithm for multiple measurements. Note that we do not assume the matrices in $\mathcal{B}$ to be Hermitian or positive semidefinite. We will see that this will be granted for free.*

10.2.3 Analysis of One Iteration of Alternating Projections

We make an initial guess $x^0 \in \mathbb{C}^n$ and start the iteration

$$X^{k+1} := P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}} X^k \tag{10.29}$$

with $X^0 := x^0 (x^0)^*$. The following lemmata investigate the behavior of the sequence of rank one matrices generated via this construction.

Lemma 10.2.6. *The set $\mathcal{B}_{rot}$ is an affine subspace of matrices. The projection of an arbitrary matrix $X \in \mathbb{C}^{n \times n}$ onto $\mathcal{B}_{rot}$ is given by*

$$P_{\mathcal{B}_{rot}} X = X + \mathrm{diag}\left((b_1 - X_{11}, \dots, b_n - X_{nn})^\top \right). \tag{10.30}$$

Proof. Let $\lambda \in \mathbb{C}$ be an arbitrary number and let $X, Y \in \mathcal{B}_{\mathrm{rot}}$. Then $X_{11} = Y_{11} = b_1, \dots, X_{nn} = Y_{nn} = b_n$. This means that $\lambda X_{11} + (1 - \lambda) Y_{11} = \lambda b_1 + (1 - \lambda) b_1 = b_1, \dots, \lambda X_{nn} + (1 - \lambda) Y_{nn} = \lambda b_n + (1 - \lambda) b_n = b_n$. This shows that $\mathcal{B}_{\mathrm{rot}}$ is an affine subspace. To show the projection, we use the best approximation property of convex sets. Let $X \in \mathbb{C}^{n \times n}$ and $Y \in \mathcal{B}_{\mathrm{rot}}$ be arbitrary. Define

$$D := \mathrm{diag}\left((b_1 - X_{11}, \dots, b_n - X_{nn})^\top \right).$$

Then

$$
\begin{aligned}
& \langle Y - X - D, X - X - D \rangle \\
={}& \langle Y - X - D, -D \rangle \\
={}& -\mathrm{Tr}\left((Y - X - D)^* D \right) \\
={}& -\mathrm{Tr}\left(\begin{pmatrix} b_1 - X_{11} - b_1 + X_{11} & \dots & Y_{1n} - X_{1n} \\ \vdots & \ddots & \vdots \\ Y_{n1} - X_{n1} & \dots & b_n - X_{nn} - b_n + X_{nn} \end{pmatrix}^* D \right) \\
={}& -\sum_{j=1}^n \overline{(b_j - X_{jj} - b_j + X_{jj})} \left(b_j - X_{jj} \right) = 0.
\end{aligned}
\tag{10.31}
$$

This shows that that the formula (10.30) indeed gives us the projection onto $\mathcal{B}_{\mathrm{rot}}$. $\qquad\square$

As explained, the affine subspace $\mathcal{B}$ is a lift (in the sense of Definition 8.3.1) of the set M. The amplitude set, defined in (9.3), is the analog for the affine space $\mathcal{B}_{\mathrm{rot}}$.

Lemma 10.2.7. *The affine space $\mathcal{B}_{rot}$ is a lift (in the sense of Definition 8.3.1) of A, defined in (9.3), with the vector a in (9.3) replaced by the vector whose entries are the positive square roots of the entries of b.*

Proof. The restrictions in $\mathcal{B}_{\mathrm{rot}}$ can be written as

$$|\langle e_k, x \rangle|^2 = x_k \overline{x_k} = b_k.$$

In other words, for every $x \in A$ there exists $X = xx^* \in \mathcal{B}_{\mathrm{rot}}$, and for every $X \in \mathcal{B}_{\mathrm{rot}} \cap S_{\leq 1}$ there exists $x \in \mathbb{C}^n$ such that $xx^* = X$. $\qquad\square$

Theorem 10.2.8 (Weyl's inequalities (Horn and Johnson, 1985, Theorem 4.3.7)). *Let $A, B \in \mathbb{C}^{n \times n}$ be Hermitian matrices, and let the eigenvalues of A, B, and $A + B$ be arranged in decreasing order. Then, for every pair of integers j, k such that $1 \leq j, k \leq n$ and $j + k \leq n + 1$, we have*

$$\lambda_{j+k-1}(A + B) \leq \lambda_j(A) + \lambda_k(B). \tag{10.32}$$

Further, for every pair of integers j, k such that $1 \leq j, k \leq n$ and $j + k \geq n + 1$, we have

$$\lambda_{j+k-n}(A + B) \geq \lambda_j(A) + \lambda_k(B). \tag{10.33}$$

Lemma 10.2.9 ((Horn and Johnson, 1985, Corollary 4.3.3)). *Let $A, B \in \mathbb{C}^{n \times n}$ be Hermitian. Assume that B is positive semidefinite, and that the eigenvalues of A and $A + B$ are arranged in decreasing order. Then*

$$\lambda_k(A) \leq \lambda_k(A + B) \text{ for all } k = 1, \ldots, n. \tag{10.34}$$

We note that for the eigenvalues of $P_{\mathcal{B}_{\text{rot}}} xx^*$ we have the trace condition, in other words, the sum of eigenvalues equals the sum of diagonal entries.

$$\sum_{j=1}^{n} b_j = \sum_{j=1}^{n} \lambda_j = \|x\|_2^2 + \sum_{j=1}^{n} (b_j - x_j \overline{x_j}). \tag{10.35}$$

Lemma 10.2.10 ((Horn and Johnson, 1985, Theorem 4.3.4)). *Let A be Hermitian and let $z \in \mathbb{C}^n$ be a given vector. If the eigenvalues of A and $A + zz^*$ are arranged in decreasing order, we have*

$$\lambda_k(A) \geq \lambda_{k+1}(A + zz^*) \geq \lambda_{k+2}(A) \; 1 \leq k \leq n - 2$$

and

$$\lambda_k(A + zz^*) \geq \lambda_{k+1}(A) \geq \lambda_{k+2}(A + zz^*) \; 1 \leq k \leq n - 2.$$

Corollary 10.2.11. *Let $0 \neq x \in \mathbb{C}^n$ be an arbitrary vector. Then the matrix*

$$X_0 := xx^* + \text{diag}\left((-x_1 \overline{x_1}, \ldots, -x_n \overline{x_n})^\top \right)$$

has exactly one positive eigenvalue λ_1, and this eigenvalue is also the largest absolute eigenvalue.

Proof. Because xx^* and X_0 are Hermitian matrices, all eigenvalues are real. Further, we note that the matrix X_0 has trace 0, and, hence the sum of eigenvalues of X_0 is zero. The eigenvalues of $\text{diag}\left((-x_1 \overline{x_1}, \ldots, -x_n \overline{x_n})^\top \right)$ are all less or equal to zero. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of X_0. With Lemma 10.2.10 we know that $\lambda_1 \left(\text{diag}\left((-x_1 \overline{x_1}, \ldots, -x_n \overline{x_n})^\top \right) \right) \geq \lambda_2$, in other words, $\lambda_2 \leq 0$. Because of $\sum_{j=1}^{n} \lambda_j = 0$, we see that $\lambda_1 = -\sum_{j=2}^{n} \lambda_j$. Hence, λ_1 has to be the largest absolute eigenvalue. If more than one x_i is different from zero, then it is also the unique largest absolute eigenvalue. $\square$

Now, because a diagonal matrix with real entries is Hermitian, and because the sum

of Hermitian matrices is Hermitian again, we see that $P_{\mathcal{B}_{\text{rot}}} xx^*$ is Hermitian. Because of the previous note, it follows that all iterates are Hermitian matrices. From Weyl's inequalities, Theorem 10.2.8, we can deduce a relation for the eigenvalues of the projection onto the affine subspace.

Lemma 10.2.12. *Let $x \in \mathbb{C}^n$ be arbitrary and let $b \in \mathbb{R}^n_{>0}$ be a vector with strictly positive entries. Let $P_{\mathcal{B}_{\text{rot}}} xx^*$ be the projection given by (10.30). Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $P_{\mathcal{B}_{\text{rot}}} xx^*$ in decreasing order. Then we have*

$$\max_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) \leq \lambda_1 \leq \max_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) + \|x\|_2^2 \tag{10.36}$$

and

$$\min_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) \leq \lambda_n \leq \min_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) + \|x\|_2^2. \tag{10.37}$$

Proof. We know that the eigenvalues of $\text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n})^\top \right)$ are just the quantities $b_j - x_j \overline{x_j}$, while the eigenvalues of xx^* are just one positive eigenvalue $\|x\|_2^2$ and 0 for all other eigenvalues. From (10.32) we get

$$\lambda_{1+1-1} \left(xx^* + \text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n})^\top \right) \right)$$

$$\leq \lambda_1(xx^*) + \lambda_1 \left(\text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n})^\top \right) \right),$$

which is equal to

$$\max_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) + \|x\|_2^2.$$

From (10.33) we get

$$\lambda_{n+1-n} \left(xx^* + \text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n})^\top \right) \right)$$

$$\geq \lambda_n(xx^*) + \lambda_1 \left(\text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n})^\top \right) \right),$$

which is equal to

$$0 + \max_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right).$$

The second claim follows analogously:

$$\lambda_{1+n-1} \leq \lambda_1(xx^*) + \lambda_n(\text{diag}\left((b_1 - x_1 \overline{z_1}, \ldots, b_n - x_n \overline{x_n}) \right))$$

$$= \min_{j \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right) + \|x\|_2^2$$

and

$$\lambda_{n+1-n} \geq \lambda_n(xx^*) + \lambda_1(\text{diag}\left((b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n}) \right))$$

$$= 0 + \min_{i \in \{1,\ldots,n\}} \left(b_j - x_j \overline{x_j} \right).$$

$\square$

Definition 10.2.13 ((Horn and Johnson, 1985, Definition 4.3.24)). *Let $\alpha = [\alpha_i] \in \mathbb{R}^n$ and $\beta = [\beta_i] \in \mathbb{R}^n$ be given. The vector β is said to* majorize *the vector α if*

$$\min\left\{\sum_{j=i}^{k} \beta_{i_j} : 1 \le i_1 < \cdots < i_k \le n\right\} \ge \min\left\{\sum_{j=i}^{k} \alpha_{i_j} : 1 \le i_1 < \cdots < i_k \le n\right\} \quad (10.38)$$

for all $k = 1, \ldots, n$ with equality for $k = n$. If we arrange the entries of α and β in increasing order, $\alpha_{j_1} \ge \cdots \ge \alpha_{j_n}$, $\beta_{m_1} \ge \cdots \ge \beta_{m_n}$, the defining inequalities can be restated in the equivalent form

$$\sum_{i=1}^{k} \beta_{m_i} \ge \sum_{i=1}^{k} \alpha_{j_i} \text{ for all } k = 1, \ldots, n \quad (10.39)$$

with equality for $k = n$.

Theorem 10.2.14 ((Horn and Johnson, 1985, Theorem 4.3.26)). *Let $A \in \mathbb{C}^{n \times n}$ be Hermitian. The vector of diagonal entries of A majorizes the vector of eigenvalues of A.*

Definition 10.2.15. *For a given vector $w \in \mathbb{R}^n$, denote by $w^{\ge}$ the vector whose entries are those of w in decreasing order. For example, for $w = (-2, 3, -5, 0, 1)$ we would have $w^{\ge} = (3, 1, 0, -2, -5)$.*

The next theorem describes the behavior of *one* step of alternating projections between $\mathcal{B}$ and $S_{\le 1}$. We assume that the initial point of the sequence of rank one matrices contains no zero entries. In Lemma 10.2.20, the case of rank one matrices with zero entries will be studied.

Theorem 10.2.16. *Let $xx^* \in \mathbb{C}^{n \times n}$ and $b = (b_1, \ldots, b_n) \in \mathbb{R}_+^n$ with $x_j \ne 0$ for all j. Then there exists a vector $(\mu_1, \ldots, \mu_n) \in \mathbb{R}_{>0}^n$ such that $v := (\mu_1 x_1, \ldots, \mu_n x_n)$ is an eigenvector of $xx^* + \mathrm{diag}\left(\left(b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n}\right)^{\top}\right)$. The largest absolute eigenvalue of $xx^* + \mathrm{diag}\left(\left(b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n}\right)^{\top}\right)$ corresponds to v.*

Proof. Note that, if xx^* is given, the vector $x \in \mathbb{C}^{n \times n}$ is determined up to global phase. Hence, without loss of generality, we can choose x such that $x_1 > 0$. By (10.30), we have

$$xx^* + \mathrm{diag}\left(\left(b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n}\right)^{\top}\right) = P_{\mathcal{B}_{\mathrm{rot}}} xx^*.$$

Let v be an eigenvector of $P_{\mathcal{B}_{\mathrm{rot}}} xx^*$. If v is an eigenvector of $P_{\mathcal{B}_{\mathrm{rot}}} xx^*$, then for all $\varphi \in [0, 2\pi]$ the vector $e^{i\varphi} v$ is an eigenvector of $P_{\mathcal{B}_{\mathrm{rot}}} xx^*$ with the same eigenvalue. Hence, we can assume that $v_1 \in \mathbb{R}_{\ge 0}$

Then, since $P_{\mathcal{B}_{\mathrm{rot}}} xx^*$ is Hermitian, the corresponding eigenvalue is real and there ex-

ists $\lambda \in \mathbb{R}$ such that $\lambda v = P_{\mathcal{B}_{\text{rot}}}(xx^*)v$, which is equivalent to

$$
\begin{aligned}
\lambda v_1 &= v^*x \cdot x_1 + (b_1 - x_1\overline{x_1}) \cdot v_1 \\
&\vdots \\
\lambda v_n &= v^*x \cdot x_n + (b_n - x_n\overline{x_n}) \cdot v_n.
\end{aligned}
\tag{10.40}
$$

Because $\lambda, v_1, x_1, b_1 - x_1\overline{x_1} \in \mathbb{R}$, we conclude that $v^*x \in \mathbb{R}$. The eigenvectors of $xx^* + \text{diag}\left((b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n})^\top\right)$ form an orthonormal system (since the matrix is Hermitian), so there has to exist at least one v such that $v^*x \neq 0$.

Now we show by contradiction that the eigenvector which belongs to the largest eigenvalue has to satisfy $v^*x \neq 0$. Denote by $b^{\geq}$ and $(b - x\overline{x})^{\geq}$ the ordered vectors as in Definition 10.2.15. Assume the eigenvector v corresponding to the largest eigenvalue satisfies $v^*x = 0$. Then we have

$$
\begin{aligned}
&\left(xx^* + \text{diag}\left((b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n})^\top\right)\right) v \\
&= xx^*v + \text{diag}\left((b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n})^\top\right) v,
\end{aligned}
\tag{10.41}
$$

which is, due to the orthogonality of x and v, equal to $(b_j - x_j\overline{x_j})v$ for some $(b_j - x_j\overline{x_j}) \leq \max_{j \in \{1,\ldots,n\}} (b_j - x_j\overline{x_j})$. Denote by $j_{\max}$ the index such that

$$
(b_{j_{\max}} - x_{j_{\max}}\overline{x_{j_{\max}}}) = \max_{j \in \{1,\ldots,n\}} (b_j - x_j\overline{x_j}).
$$

Further, denote by $j*$ the index such that $b_{j*} = \max\{b_j\}$. The vector of diagonal entries majorizes the vector of eigenvalues. Write the vector with entries $(b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n})$ as $(b - x\overline{x})$. The sum of eigenvalues is equal to the sum of diagonal entries of $xx^* + \text{diag}\left((b_1 - x_1\overline{x_1}, \ldots, b_n - x_n\overline{x_n})^\top\right)$, which is just $\sum_{j=1}^n b_j$. The largest eigenvalue is assumed to be equal to $b_{j_{\max}} - x_{j_{\max}}\overline{x_{j_{\max}}}$. Hence, we can use the majorization and the ordered vectors of eigenvalues and diagonal entries:

$$
\begin{aligned}
&& \sum_{j=2}^n (b_j)^{\geq} &\geq \|x\|_2^2 + \sum_{j=2}^n (b - x\overline{x})_i^{\geq} \\
\Leftrightarrow && \sum_{j=2}^n (b_j)^{\geq} - x_j\overline{x_j} &\geq x_{j*}\overline{x_{j*}} + \sum_{j=2}^n (b - x\overline{x})_i^{\geq} \\
\Leftrightarrow && b_{j_{\max}} - x_{j_{\max}}\overline{x_{j_{\max}}} &\geq x_{j*}\overline{x_{j*}} + b_{j*} - x_{j*}\overline{x_{j*}} \\
\Leftrightarrow && b_{j_{\max}} - x_{j_{\max}}\overline{x_{j_{\max}}} &\geq b_{j*}.
\end{aligned}
\tag{10.42}
$$

But the last inequality cannot be true because b_{j*} is the largest entry of b and must be strictly larger than $b_{j_{\max}} - x_{j_{\max}}$. Hence, the eigenvector v of $P_{\mathcal{B}_{\text{rot}}}xx^*$ corresponding to the largest eigenvalue has to satisfy $v^*x \neq 0$.

In case $v^*x \neq 0$, we see that

$$
v_1 = \frac{v^*x}{\lambda - b_1 + x_1\overline{x_1}}x_1, \ldots, v_n = \frac{v^*x}{\lambda - b_n + x_n\overline{x_n}}x_n.
$$

Using Equation (10.40), we see that if we have $\lambda - b_j + x_j\overline{x_j} = 0$ for some $j \in \{1, \ldots, n\}$

then x_j has to be zero, which is not possible by assumption.

We show now that the largest absolute eigenvalue is a positive one. Because of Corollary 10.2.11, we know that the largest absolute eigenvalue of the matrix $P_{\mathcal{B}_{\text{rot}}} xx^*$ is positive. With $b \in \mathbb{R}^n_{>0}$ and Lemma 10.2.9, we see that the eigenvalues of

$$P_{\mathcal{B}_{\text{rot}}} xx^* = xx^* + \text{diag}\left(\left(-x_1\overline{x_1}, \ldots, -x_n\overline{x_n}\right)^\top\right) + \text{diag}\left(b_1, \ldots, b_n\right)$$

are larger than the ones of $xx^* + \text{diag}\left(\left(-x_1\overline{x_1}, \ldots, -x_n\overline{x_n}\right)^\top\right)$. Hence, the largest eigenvalue of $P_{\mathcal{B}_{\text{rot}}} xx^*$ is positive and its absolute value is strictly larger than the absolute value of its smallest eigenvalue.

Due to Equation (10.36), if λ is the largest eigenvalue of $P_{\mathcal{B}_{\text{rot}}} xx^*$ then λ_1 is greater or equal to $\max_{j\in\{1,\ldots,n\}}\left(b_j - x_j\overline{x_j}\right)$. This means that $\frac{v^*x}{\lambda-b_j+x_j\overline{x_j}} > 0$. Further, because $\lambda \geq b_j - x_j\overline{x_j}$ for all j, we conclude that $v^*x > 0$. This gives us the existence of a vector $\mu \in \mathbb{R}^n_{>0}$ with $\mu_j = \frac{v^*x}{\lambda-b_j+x_j\overline{x_j}}$. $\qquad\square$

Lemma 10.2.17. *Let the assumptions of Theorem 10.2.16 hold. Then the eigenspace of the eigenvectors of $P_{\mathcal{B}_{\text{rot}}} xx^*$ corresponding to the largest eigenvalue λ_1 is of dimension 1.*

Proof. We give a proof by contradiction. Assume the eigenspace of $P_{\mathcal{B}_{\text{rot}}} xx^*$ corresponding to the largest eigenvalue λ_1 is of dimension 2 or higher. By Theorem 10.2.16, the eigenvectors and hence all elements v of the eigenspace satisfy $v_j = \mu_j x_j$ for some strictly positive scalar μ_j for all $j = 1, \ldots, n$. Because there have to be two orthonormal vectors v, w in the eigenspace, there exist real vectors μ^v and μ^w of positive scalars. We compute

$$0 = \langle v, w \rangle = \sum_{j=1}^n \overline{v_j} w_j = \sum_{j=1}^n \mu_j^v \mu_j^w \overline{x_j} x_j > 0,$$

which is a contradiction. $\qquad\square$

The consequence of Lemma 10.2.17 is that, if the assumptions of Theorem 10.2.16 are satisfied, then the projection of $P_{\mathcal{B}_{\text{rot}}} xx^*$ onto $S_{\leq 1}$ is always single-valued.

Proposition 10.2.18. *Given a Hermitian matrix $X \in \mathbb{C}^{n\times n}$, there exists a matrix $X_1 \in P_{S_{\leq 1}} P_{\mathcal{B}_{\text{rot}}} X$ such that X_1 is Hermitian. Consequently, the restriction of $P_{S_{\leq 1}} P_{\mathcal{B}_{\text{rot}}}$ to the space of Hermitian matrices has the same form as $P_{S_{\leq 1}} P_{\mathcal{B}_{\text{rot}}}$ in the space $\mathbb{C}^{n\times n}$.*

Proof. By Lemma 10.2.6, the matrix $P_{\mathcal{B}_{\text{rot}}} X$ is given by

$$X + \text{diag}\left(\left(b_1 - X_{11}, \ldots, b_n - X_{nn}\right)^\top\right).$$

Since the sum of Hermitian matrices is Hermitian again, we conclude that $P_{\mathcal{B}_{\text{rot}}} X$ is Hermitian.

By Proposition 8.2.3, the projection $P_{S_{\leq 1}} P_{\mathcal{B}_{\text{rot}}} X$ can be chosen Hermitian again. This finishes the proof. $\qquad\square$

10.2.4 Fixed Points of Alternating Projections in the Matrix Space

Lemma 10.2.19. *Let $C \subset \mathbb{C}^{n \times n}$ be a closed such that there exists a positive semidefinite, rank-one matrix vv^* that is a fixed point of $P_{S_{\leq 1}} P_C$. Then $\|v\|_2^2$ is the largest absolute eigenvalue of the matrix $P_C vv^*$.*

Proof. By Theorem 7.1.2, for a positive semidefinite Hermitian matrix $X \in \mathbb{C}^{n \times n}$, finding the projection onto $S_{\leq 1}$ is equivalent to finding the normalized eigenvector $v / \|v\|_2$ of X corresponding to the largest eigenvalue $\lambda_1(X)$. Then the matrix $P_{S_{\leq 1}} X$ is given by

$$\lambda_1(X) \frac{vv^*}{\|v\|_2^2}. \tag{10.43}$$

If vv^* is a fixed point of $P_{S_{\leq 1}} P_C$, then necessarily v has to be an eigenvector of $P_C vv^*$. By (10.43), we get the eigenvalue equation

$$vv^* = \lambda_1(P_C vv^*) \frac{vv^*}{\|v\|_2^2}.$$

This finishes the proof. $\qquad\square$

Lemma 10.2.20. *Let $b \in \mathbb{R}_{>0}^n$ and define the set*

$$\mathcal{I} := \left\{ I \subset \{1, \ldots, n\} \;\middle|\; \sum_{i \in I} b_i > b_j \; \forall j \notin I \right\}.$$

Denote for $I \in \mathcal{I}$ by $\sqrt{b}_I$ the vector with entries $\pm \sqrt{b_i}$, for $i \in I$, and zero otherwise. Define the set $B_{\mathcal{I}} := \left\{ Z \in \mathbb{R}^{n \times n} \;\middle|\; Z = \sqrt{b}_I \sqrt{b}_I^* \text{ for } I \in \mathcal{I} \right\}$. Then $B_{\mathcal{I}}$ is the set of fixed points of the operator $P_{S_{\leq 1}} P_{\mathcal{B}_{rot}}$.

Proof. Let v be a vector such that vv^* is a fixed point of the operator $P_{S_{\leq 1}} P_{\mathcal{B}_{rot}}$. This means that v is an eigenvector of the matrix $P_{\mathcal{B}_{rot}} vv^*$ with, by Lemma 10.2.19, eigenvalue $\|v\|_2^2$. Further, $\|v\|_2^2$ is also the largest absolute eigenvalue. In other words, we have

$$\|v\|_2^2 v_1 = \|v\|_2^2 v_1 + (b_1 - v_1^2) \cdot v_1$$

$$\vdots$$

$$\|v\|_2^2 v_n = \|v\|_2^2 v_n + (b_n - v_n^2) \cdot v_n,$$

which is equivalent to

$$0 = (b_1 - v_1^2) \cdot v_1$$

$$\vdots$$

$$0 = (b_n - v_n^2) \cdot v_n.$$

This can only hold if there is an index set $J \subset \{1, \ldots, n\}$ such that $v_j = 0$ for all $j \in J$, and if
$$b_i = v_i^2 \quad \text{for all} \quad i \in I := \{1, \ldots, n\} \setminus J.$$

Hence, $\|v\|_2^2$ is equal to $\sum_{i \in I} b_i$, and we see that this has to be the largest eigenvalue of $P_{\mathcal{B}_{\mathrm{rot}}} vv^*$ (by Lemma 10.2.19). Now, choose eigenvectors $\{w_l\}_{l=2,\ldots,n}$ of $P_{\mathcal{B}_{\mathrm{rot}}} vv^*$. In fact, every w_l can be chosen to be orthogonal to v because v is by assumption an eigenvector of the Hermitian matrix $P_{\mathcal{B}_{\mathrm{rot}}} vv^*$. Remember that, by (10.27), we have

$$P_{\mathcal{B}_{\mathrm{rot}}} vv^* = vv^* + \mathrm{diag}\left((b_1 - v_1^2, \ldots, b_n - v_n^2)^\top \right).$$

If $w_l^* v = 0$ for all $l = 2, \ldots, n$, we conclude that w_l must be an eigenvector of the matrix $\mathrm{diag}\left((b_1 - v_1^2, \ldots, b_n - v_n^2)^\top \right)$. If $\sum_{i \in I} b_i$ is the largest eigenvalue, then this is equivalent to $\sum_{i \in I} b_i > b_j - v_j^2 = b_j$ for all $j \in J$ (because the v_j are equal to zero). This shows that $\mathrm{Fix}(P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}}) \subset \mathcal{B}_{\mathcal{I}}$.

Now, let $Z \in \mathcal{B}_{\mathcal{I}}$. Then there exists a vector $z \in \mathbb{C}^n$ such that $Z = zz^*$ and we can choose $z = \sqrt{b_I}$ for an index set $I \in \mathcal{I}$ since $zz^* \in \mathcal{B}_{\mathcal{I}}$. Let $b_{I^\mathbb{C}}$ be the vector with entries

$$b_{I^\mathbb{C}} = \begin{cases} b_j & \text{if } j \notin I, \\ 0 & \text{if } j \in I, \end{cases}$$

where $I^\mathbb{C}$ is the complement of I in $\{1, \ldots, n\}$, i.e., $I^\mathbb{C} = \{1, \ldots, n\} \setminus I$. The projection of zz^* onto $\mathcal{B}_{\mathrm{rot}}$ can then be written as

$$P_{\mathcal{B}_{\mathrm{rot}}} zz^* = zz^* + \mathrm{diag}\left((b_1 - z_1^2, \ldots, b_n - z_n^2)^\top \right) = \sqrt{b_I}\left(\sqrt{b_I}\right)^* + \mathrm{diag}\left(b_{I^\mathbb{C}}\right). \quad (10.44)$$

To find the eigenvector v and the corresponding largest eigenvalue λ_1 of $P_{\mathcal{B}_{\mathrm{rot}}} zz^*$, we get again an eigenvector equation

$$\lambda_1 v_1 = z^* v v_1 + (b_{I^\mathbb{C}\,1}) \cdot v_1$$
$$\vdots$$
$$\lambda_1 v_n = z^* v v_n + (b_{I^\mathbb{C}\,n}) \cdot v_n.$$

For all $j \in I^\mathbb{C}$, the equation is reduced to $\lambda_1 v_j = (z^* v + b_j) v_j$, which implies $\lambda_1 = v^* z + b_j$ or $v_j = 0$. For all $i \in I$, we get $\lambda_1 v_i = z^* v v_i$, which implies $\lambda_1 = v^* z$ of $v_i = 0$.

The relation $b \in \mathbb{R}^n_{>0}$ implies

$$
\begin{aligned}
& v_i \neq 0 \quad \text{for some} \quad i \in I \quad \Rightarrow \quad v_j \neq 0 \quad \text{for all} \quad j \in I^C \\
& \text{and} \\
& v_j \neq 0 \quad \text{for some} \quad j \in I^C \quad \Rightarrow \quad v_i \neq 0 \quad \text{for all} \quad i \in I.
\end{aligned}
\tag{10.45}
$$

We study both cases.

1. If $v_j \neq 0$ for some $j \in I^C$ then we have $v^* z = 0$, by (10.45). Hence, the eigenvalue corresponding to this vector is b_j.

2. If $v_i \neq 0$ for some $i \in I$, then we have, by (10.45), $v^* z \neq 0$. Hence, the eigenvalue corresponding to this vector is $v^* z$. In this case, since $v_j = 0$ for all $j \in I^C$, the vector v has to be an eigenvector of the matrix zz^*. But the only eigenvector of zz^* is z itself. Hence, $v^* z = \sum_{i \in I} b_i$. Because $I \in \mathcal{I}$, the quantity $v^* z$ is larger than all other b_j for $j \in I^C$.

We conclude that the largest eigenvalue of $P_{\mathcal{B}_{\mathrm{rot}}} zz^*$ is $\lambda_1 = z^* z = \sum_{i \in I} b_i$. Hence, $z = v \in \mathrm{Fix}(P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}})$. This finishes the proof. $\qquad\square$

Theorem 10.2.21. *The alternating projections sequence generated by the operator $P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}}$, starting at some $xx^* \in \mathbb{C}^{n \times n}$, is bounded.*

Proof. Start at some xx^*. Then $xx^* \in \mathbb{B}_{\|xx^*\|_F + \left\| \sqrt{b}\sqrt{b}^* \right\|_F}$. Further, we have

$$
\mathbb{B}_{\left\| \sqrt{b}\sqrt{b}^* - xx^* \right\|_F} \left(\sqrt{b}\sqrt{b}^* \right) \subset \mathbb{B}_{\|xx^*\|_F + \left\| \sqrt{b}\sqrt{b}^* \right\|_F}.
$$

The projection onto $\mathcal{B}_{\mathrm{rot}}$ is nonexpansive, which gives us

$$
P_{\mathcal{B}_{\mathrm{rot}}} xx^* \in \mathbb{B}_{\left\| \sqrt{b}\sqrt{b}^* - xx^* \right\|_F} \left(\sqrt{b}\sqrt{b}^* \right),
$$

and consequently, $P_{\mathcal{B}_{\mathrm{rot}}} xx^* \in \mathbb{B}_{\|xx^*\|_F + \left\| \sqrt{b}\sqrt{b}^* \right\|_F}$. Because the projection onto $S_{\leq 1}$ satisfies $\left\| P_{S_{\leq 1}} Z \right\|_F \leq \|Z\|_F$ for all $Z \in \mathbb{C}^{n \times n}$, we have $P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}} xx^* \in \mathbb{B}_{\|xx^*\|_F + \left\| \sqrt{b}\sqrt{b}^* \right\|_F}$. This shows that the sequence is bounded. $\qquad\square$

Lemma 10.2.22. *Let $x \in \mathbb{C}^n$ with $x_j \neq 0$ for all j and let $b \in \mathbb{R}^n_{>0}$. Then*

$$
P_{S_{\leq 1} \cap \mathcal{B}_{rot}} xx^* = P_A x (P_A x)^*,
\tag{10.46}
$$

where P_A is given by (9.9).

Proof. The elements of the set of rank one matrices in $\mathcal{B}_{\mathrm{rot}}$ have the common restriction that the diagonal entries are the entries of b. For all $v \in \mathbb{C}^n$, the jth diagonal entry of the rank one positive semidefinite Hermitian matrix vv^* is given by $v_j \overline{v}_j$. Since for

all j, $v_j \overline{v_j} = b_j$ for all positive semidefinite rank one Hermitian matrices in $\mathcal{B}_{\mathrm{rot}}$, we have $v_j = b_j e^{\varphi_j i}$ for all positive semidefinite rank one matrices in $\mathcal{B}_{\mathrm{rot}}$. Hence, we can write any Hermitian rank one matrix matrix in $\mathcal{B}_{\mathrm{rot}}$ as outer product of the vector $\left(b_1 e^{\varphi_1 i}, \dots, b_n e^{\varphi_n i}\right)^\top$ with itself. Write $x = \left(|x_1| e^{i\psi_1}, \dots, |x_n| e^{i\psi_n}\right)^\top$. Then, by projecting onto the intersection of these sets, we seek

$$
\operatorname*{argmin}_{\varphi \in [0,2\pi]^n} \left\| xx^* - \left(b_1 e^{\varphi_1 i}, \dots, b_n e^{\varphi_n i}\right) \left(b_1 e^{\varphi_1 i}, \dots, b_n e^{\varphi_n i}\right)^* \right\|_F^2
$$

$$
= \operatorname*{argmin}_{\varphi \in [0,2\pi]^n} \sum_{j,k=1}^n |x_j x_k|^2 + b_j^2 b_k^2 - |x_j x_k| b_j b_k \left(e^{i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)} + e^{-i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)}\right)
$$

$$
= \operatorname*{argmin}_{\varphi \in [0,2\pi]^n} \sum_{j,k=1}^n -|x_j x_k| b_j b_k \left(e^{i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)} + e^{-i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)}\right)
$$

$$
= \operatorname*{argmax}_{\varphi \in [0,2\pi]^n} \sum_{j,k=1}^n e^{i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)} + e^{-i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)}
$$

Each term $e^{i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)} + e^{-i\left(\psi_j - \psi_k - \varphi_j + \varphi_k\right)}$ takes its value in $[-2,2]$ because it is the sum of a complex number of modulus one with its complex conjugate. If we have $\psi_j - \psi_k - \varphi_j + \varphi_k = 0$ for all j,k, then the maximum is attained. In other words, $\psi_j - \psi_k = \varphi_j - \varphi_k$, which is equivalent to $\varphi_j = \psi_j + c$ for some real constant c, is a sufficient condition for φ to be a maximizing argument. The constant c corresponds to the global phase shift. Hence, the matrix

$$
X_{\mathrm{proj}} := \left(b_1 e^{-(\psi_1 + c) i}, \dots, b_n e^{-(\psi_n + c) i}\right)^\top \left(b_1 e^{(\psi_1 + c) i}, \dots, b_n e^{(\psi_n + c) i}\right)
$$

is the projection of xx^* onto $S_{\leq 1} \cap \mathcal{B}_{\mathrm{rot}}$. We note that the vector $\left(b_1 e^{(\psi_1) i}, \dots, b_n e^{(\psi_n) i}\right)$ satisfies

$$
\left(b_1 e^{-(\psi_1 + c) i}, \dots, b_n e^{-(\psi_n + c) i}\right)^\top \left(b_1 e^{\psi_1 i}, \dots, b_n e^{\psi_n i}\right) = X_{\mathrm{proj}}
$$

and since $\left(b_1 e^{\psi_1 i}, \dots, b_n e^{\psi_n i}\right) \in P_A x$, this proves the claim. □

10.3 Alternating Projections in Affine-Rank-Constrained Feasibility

Now, we would like to study the behavior of the method of alternating projections between the affine space $\mathcal{B}$ (10.4) and the set $S_{\leq 1}$. We use the results derived in Section 10.2 to obtain a local convergence result in Subsection 10.3.1. In Subsection 10.3.2 we show how existing results can be used to solve (10.1).

10.3.1 Local Linear Convergence of Alternating Projections to $\mathcal{B} \cap S_{\leq 1}$

We use the result on local linear convergence of alternating projections to the intersection in the case of two transversally intersecting smooth manifolds in $\mathbb{R}^n$ to show local linear convergence of the AP sequence in the phase lift case. The first and most important fact is that in phase lift we are working in the complex setting, while theory of smooth manifolds is just defined on real spaces.

Definition 10.3.1. *Define for a vector $\varphi \in [0, 2\pi]^n$ of angles the* rotator

$$R_\varphi : \mathbb{C}^n \to \mathbb{C}^n, (x_1, \ldots, x_n) \mapsto (\mathrm{e}^{\mathrm{i}\,\varphi_1}\, x_1, \ldots, \mathrm{e}^{\mathrm{i}\,\varphi_n}\, x_n). \tag{10.47}$$

The map R_φ can also be written as $\mathrm{diag}\left(\left(\mathrm{e}^{\mathrm{i}\,\varphi_1}, \ldots, \mathrm{e}^{\mathrm{i}\,\varphi_n}\right)^\top\right).$

We use this mapping to rotate our current iterates from $\mathbb{C}^n$ into $\mathbb{R}^n$.

Proposition 10.3.2. *Let $x \in \mathbb{C}^n$ be such that $x_j \neq 0$ for all j and let $v \in \mathbb{C}^n$ be such that $vv^* \in P_{S_{\leq 1}} P_{\mathcal{B}_{rot}} xx^*$. Define the vector φ via $x = (\mathrm{e}^{\mathrm{i}\,\varphi_1}\,|x_1|, \ldots, \mathrm{e}^{\mathrm{i}\,\varphi_n}\,|x_n|)$. Then*

$$R_{-\varphi} v R_{-\varphi} v^* \in P_{S_{\leq 1}} P_{\mathcal{B}_{rot}} R_{-\varphi} x R_{-\varphi} x^*. \tag{10.48}$$

Proof. The mapping $R_{-\varphi}$ is unitary. Hence, we have for any eigenvector w with eigenvalue λ of a Hermitian matrix A

$$R_{-\varphi} A R^*_{-\varphi} (R_{-\varphi} w) = R_{-\varphi} A w = R_{-\varphi} \lambda w = \lambda R_{-\varphi} w.$$

Now we note that

$$
\begin{aligned}
P_{\mathcal{B}_{\mathrm{rot}}} R_{-\varphi} x R_{-\varphi} x^* &= R_{-\varphi} x R_{-\varphi} x^* + \mathrm{diag}\left(\left(b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n}\right)^\top\right) \\
&= R_{-\varphi} x R_{-\varphi} x^* + R_{-\varphi} \mathrm{diag}\left(\left(b_1 - x_1 \overline{x_1}, \ldots, b_n - x_n \overline{x_n}\right)^\top\right) R^*_{-\varphi} \\
&= R_{-\varphi} (P_{\mathcal{B}_{\mathrm{rot}}} xx^*) R^*_{-\varphi}.
\end{aligned}
$$

If v is an eigenvector of $P_{\mathcal{B}_{\mathrm{rot}}} xx^*$ with largest eigenvalue λ, then $R_{-\varphi} v$ is an eigenvector with largest eigenvalue λ of $P_{\mathcal{B}_{\mathrm{rot}}} R_{-\varphi} x R_{-\varphi} x^*$. But the latter is equivalent to $R_{-\varphi} v R_{-\varphi} v^* \in P_{S_{\leq 1}} P_{\mathcal{B}_{\mathrm{rot}}} R_{-\varphi} x R_{-\varphi} x^*$. This shows the claim. $\quad\square$

Remark 10.3.3. *We have seen in Theorem 10.2.16 that the phases of the entries in the eigenvectors corresponding to the largest eigenvalues do not change after an iteration of alternating projections. So we actually only have to focus on the changes of the absolute values of the entries of the vectors. The consequence of Proposition 10.3.2 is that we can analyze the convergence behavior of the AP sequence initiated at any $x \in \mathbb{C}^n$ with $x_j \neq 0$ for all j entirely in the real setting just by rotating it into $\mathbb{R}^n$. This means that we can apply convergence results for AP in real Euclidean spaces.*

Definition 10.3.4 (transversality (Lewis and Malick, 2008, Definiton 2.1))**.** *Suppose Ω_1 and Ω_2 are two C^k manifolds around a point $x \in \Omega_1 \cap \Omega_2$. We say that Ω_1 and Ω_2 are*

transverse *at x if*

$$T_{\Omega_1} + T_{\Omega_2} = \mathbb{R}^n. \tag{10.49}$$

Lemma 10.3.5. *Let $b \in \mathbb{R}_{>0}^n$ and let $\sqrt{b}$ be the vector with entries $+\sqrt{b_j}$ for all $1 \leq j \leq n$. Then the sets $\mathcal{B}_{rot} \cap \mathbb{R}^{n \times n}$ and $S_{\leq 1} \cap \mathbb{R}^{n \times n}$, defined via (10.27) and (8.1), respectively, intersect transversally at $\sqrt{b}\sqrt{b}^*$.*

Proof. The sets $\mathcal{B}_{\text{rot}} \cap \mathbb{R}^{n \times n}$ and $S_{\leq 1} \cap \mathbb{R}^{n \times n}$ intersect transversally if their tangent spaces span the space $\mathbb{R}^{n \times n}$. The set $\mathcal{B}_{\text{rot}} \cap \mathbb{R}^{n \times n}$ is an affine subspace defined by n linearly independent equations. Hence, it is of real dimension $n^2 - n$. The tangent space of $\mathcal{B}_{\text{rot}} \cap \mathbb{R}^{n \times n}$ at any point is the space $\mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$, which is the linear subspace parallel to $\mathcal{B}_{\text{rot}}$. The space $\mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$ can be written as the set of matrices with zero diagonals. Denote by $\sqrt{b}\sqrt{b}^* = U\Sigma U^*$ a singular value decomposition of $\sqrt{b}\sqrt{b}^*$. The set $S_{\leq 1} \cap \mathbb{R}^{n \times n}$ is a smooth manifold of dimension $2n - 1$. Its tangent space at the matrix $\sqrt{b}\sqrt{b}^*$ is given by (see (Vandereycken, 2013, Proposition 2.1))

$$T_{S_{\leq 1}} \sqrt{b}\sqrt{b}^* = U \begin{pmatrix} \mathbb{R} & \mathbb{R}^{1 \times (n-1)} \\ \mathbb{R}^{(n-1) \times 1} & 0^{n \times n} \end{pmatrix} U^*. \tag{10.50}$$

Now we show that this space contains n linearly indpendent elements that are not contained in $\mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$. The columns of the matrix U are the normalized vector with direction b and $n-1$ normalized vectors $u_2, \ldots, u_n$ orthogonal to b. Denote by w_j the conjugate transpose of the *j*th *row* of U and by e_j the *j*th standard unit vector in $\mathbb{R}^n$. Construct, for every $1 \leq j \leq n$, the matrix

$$Uw_je_1^*U^* = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \vdots & & \ddots & \\ 1 & 0 & \ldots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \ldots & 0 \end{pmatrix} U^* = \frac{1}{\|\sqrt{b}\|_2} \begin{pmatrix} 0 & \ldots & 0 \\ & \ddots & \\ & \sqrt{b}^* & \\ & \ddots & \\ 0 & \ldots & 0 \end{pmatrix} =: W^j.$$

Since $b \in \mathbb{R}_{>0}^n$, we observe that for every $1 \leq j \leq n$, the element W_{jj}^j is the only nonzero entry on the diagonal of W^j. Hence, the matrices W^j are linearly independent and $W_j \notin \mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$ for all j. Further, any linear combination of the W^j cannot be an element of $\mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$ as well. This means that there is an n-dimensional subspace, namely, the span of the matrices W^j, contained in $T_{S_{\leq 1}} \sqrt{b}\sqrt{b}^*$, which has only a trivial intersection with $\mathcal{B}_{\text{rot}} - \sqrt{b}\sqrt{b}^*$. This is sufficient for transversality. □

The next theorem gives us the key to show local linear convergence of a sequence of points generated by the alternating projections algorithm. We show afterwards that in our case the assumptions of this result of Lewis and Malick are satisfied.

Theorem 10.3.6 ((Lewis and Malick, 2008, Theorem 4.3)). *In the space $\mathbb{R}^n$, let Ω_1 and Ω_2 be two transverse manifolds around a point $\bar{x} \in \Omega_1 \cap \Omega_2$. If the initial point $x^0 \in \mathbb{R}^n$ is close*

to $\bar{x}$, then the method of alternating projections,

$$x^{k+1} = P_{\Omega_1} P_{\Omega_2} x^k \ (k = 0, 1, 2, \dots),$$

is well defined, and the distance $d_{\Omega_1 \cap \Omega_2}\left(x^k\right)$ from the iterate x^k to the intersection $\Omega_1 \cap \Omega_2$ decreases Q-linearly to zero. More precisely, given any constant c strictly larger than the cosine of the angle of intersection between the manifolds, if x^0 is close to $\bar{x}$, then the iterates satisfy

$$d_{\Omega_1 \cap \Omega_2}\left(x^{k+1}\right) \leq c d_{\Omega_1 \cap \Omega_2}\left(x^k\right) \ (k = 0, 1, 2, \dots).$$

Furthermore, x^k converges linearly to some point $x^* \in \Omega_1 \cap \Omega_2$: for some constant $\alpha > 0$,

$$\left\|x^k - x^*\right\| \leq \alpha c^k \ (k = 0, 1, 2, \dots).$$

Now we can proof one of our main results.

Theorem 10.3.7. *Given an arbitrary vector x with $x_j \neq 0$ for all j such that xx^* is sufficiently close to the intersection $S_{\leq 1} \cap \mathcal{B}_{rot}$, then the AP sequence in phase lift initiated at xx^* converges locally linearly to the matrix $P_A x P_A x^*$ where P_A is given by (9.9).*

Proof. By Theorem 10.2.16 and Proposition 10.3.2, we know that the iteration starting at xx^* converges to $P_A x P_A x^*$ if and only if the iteration starting at $R_{-\varphi} x R_{-\varphi} x^*$ converges to $R_{-\varphi} P_A x R_{-\varphi} P_A x^*$. The latter lives entirely in a real vector space, and it is an alternating projections iteration between two smooth manifolds. By Lemma 10.3.5, we know that these manifolds intersect transversally. Hence, Theorem 10.3.6 applies, and we conlude that we have local linear convergence of the AP iteration. □

10.3.2 Local Linear Convergence of Alternating Projections to $\mathcal{B} \cap S_{\leq r}$

Now, we give sufficient conditions for local convergence of the alternating projections sequence when applied to the sets $S_{\leq r}$ and an affine subspace $\mathcal{B} \subset \mathbb{C}^{n \times n}$. Here we use results from (Hesse, 2014). We bring these collections together to prove our main result of this section, Proposition 10.3.13.

Assumption 10.3.8. *Assume at this point, that there exists a solution $\overline{X}$ of $\mathcal{A}(X) = b$ with rank $\left(\overline{X}\right) = r$. Further, let the linear map $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ be such that*

$$\ker \mathcal{A} \cap S_{\leq 2r} = \{0\}.$$

Lemma 10.3.9. *If Assumption 10.3.8 holds, then the solution $\overline{X}$ to the linear system $\mathcal{A}(X) = b$ is unique.*

Proof. Assume that there exists a rank-r matrix X satisfying $\mathcal{A}(X) = b$ and $X \neq \overline{X}$. Then the matrix $Z := \overline{X} - X$ is a matrix of rank $2r$ satisfying $\mathcal{A}(Z) = 0$, which is a contradiction to Assumption 10.3.8. □

Theorem 10.3.10 ((Hesse, 2014), Theorem 5.19 c)). *Let $\Omega_1, \ldots, \Omega_m$ be nonempty and closed subsets of a finite dimensional Hilbert space. Then the collection $\{\Omega_1, \ldots, \Omega_m\}$ is locally linearly regular and $\bigcap_{j=1}^{m} \Omega_j = \{\overline{X}\}$ if and only if*

$$\bigcap_{j=1}^{m} T_{\Omega_j}(\overline{X}) = \{0\}. \tag{10.51}$$

Theorem 10.3.11 ((Hesse, 2014), Corollary 6.8). *Let Ω_1, Ω_2 be closed, nonempty, and super-regular. Let $\{\Omega_1, \Omega_2\}$ be locally linearly regular at $\bar{x} \in \Omega_1 \cap \Omega_2$. Then there is a $\delta > 0$ such that for all $x^0 \in \mathbb{B}_\delta(\bar{x}) \cap A$ any alternating projections sequence initiated at $\bar{x}$ converges with linear rate to $\Omega_1 \cap \Omega_2$.*

Lemma 10.3.12. *Let Assumption 10.3.8 hold. Then the tangent space $T_{S_{\leq r}}(\overline{X})$ at $S_{\leq r} \cap \mathcal{B} = \{\overline{X}\}$ satisfies $T_{S_{\leq r}}(\overline{X}) \cap \ker(\mathcal{A}) = 0$.*

Proof. Assume there exists $Y \in \ker(\mathcal{A}) \subset \mathbb{R}^{n \times n}$ with $y \neq 0$ and

$$Y \in T_{S_{\leq r}}(\overline{X}) = \left\{ UA^\top + BV^\top \,\middle|\, A, B \in \mathbb{R}^{n \times r} \right\}, \tag{10.52}$$

where $\overline{X} = U\Sigma V^\top$ is the singular value decomposition of $\overline{X}$. The elements of $T_{S_{\leq r}}(\overline{X})$ are of rank at most $2r$ due to the subadditivity of the rank function. Because of Assumption 10.3.8, there cannot be any matrix of rank $2r$ in the nullspace of $\mathcal{A}$ so that $Y \neq 0$ is a contradiction. $\square$

The previous lemma states that there is no common line in the nullspace of $\mathcal{A}$ and in the tangent space of the set of matrices of rank r. This means that there is an angle different from zero between the nullspace of $\mathcal{A}$ and $T_{S_{\leq r}}(\overline{X})$.

Proposition 10.3.13 (local linear convergence of alternating projections). *Let Assumption 10.3.8 hold. Then $S_{\leq r} \cap \mathcal{B} = \{\overline{X}\}$, and there exists $\varepsilon > 0$ such that the alternating projections sequence $X^{k+1} \in P_{S_{\leq r}} P_{\mathcal{B}} X^k$ initiated at an arbitrary point X^0 with $\|X^0 - \overline{X}\|$ converges to $\overline{X}$ at a linear rate.*

Proof. This follows from combining the results in Theorem 10.3.10, Theorem 10.3.11, and Lemma 10.3.12: Both sets $\mathcal{B}$ and $S_{\leq r}$ are closed and nonempty. With 10.3.12 we get a trivial intersection of the tangent spaces, which gives us local linear regularity at the intersection. Further, we know that both sets are by prox-regularity also super-regular. Then 10.3.11 gives the local linear convergence. $\square$

We bring our results into a context with existing conditions which are related to Assumption 10.3.8. It is not only possible to formulate a restricted isometry property for matrices in the case of sparse vectors, as done in Definition 5.2.6. This concept can be adapted to the case of rank constrained matrices.

Definition 10.3.14 ((Recht et al., 2010, Definition 3.1)). *Let* $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ *be a linear map. Without loss of generality, assume* $m \leq n$. *For every integer r with* $1 \leq r \leq m$, *define the r-restricted isometry constant to be the smallest number* $\delta_r(\mathcal{A})$ *such that*

$$(1 - \delta_r(\mathcal{A})) \|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \delta_r(\mathcal{A})) \|X\|_F^2. \tag{10.53}$$

holds for all matrices X of rank at most r.

Lemma 10.3.15 ((Recht et al., 2010, Theorem 3.2)). *Let* $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ *be a linear map satifying* (10.53). *Assume now that there exists a solution* $\overline{X}$ *of* $\mathcal{A}(X) = b$ *with* $\mathrm{rank}\,(\overline{X}) = r$. *Suppose that* $\delta_{2r}(\mathcal{A}) < 1$ *for some integer* $r \geq 1$. *Then* $\overline{X}$ *is the only matrix of rank at most r satisfying* $\mathcal{A}(X) = b$.

Remark 10.3.16. *We observe that Definition 10.3.14, together with Lemma 10.3.15, delivers Assumption 10.3.8. Since the latter is, due the lack of an upper bound, weaker than the condition* (10.53), *we chose to define Assumption 10.3.8 instead of referring to a restricted isometry property for rank constrained matrices.*

10.4 Global Convergence Revisited

Assume that we have several measurements $b^i, i = 1, \ldots, m$, to build the spaces

$$\mathcal{B}_i := \left\{ Z \in \mathbb{C}^{n \times n} \ \middle| \ (a_j^i)^* Z a_j^i = b_j^i \ \forall j = 1, \ldots, n \right\}.$$

Define the sets

$$\begin{aligned} \mathcal{B} &:= \mathcal{B}_1 \times \cdots \times \mathcal{B}_m \subset \mathbb{C}^{m(n \times n)}, \\ D &:= \left\{ (X_1, \ldots, X_m) \in \mathbb{C}^{m(n \times n)} \ \middle| \ X_j \in \mathbb{C}^{n \times n}, \ X_j = X_k \ \forall j, k = 1, \ldots, m \right\}, \\ S_{\leq 1}^m &:= \left\{ (X_1, \ldots, X_m) \in \mathbb{C}^{m(n \times n)} \ \middle| \ \mathrm{rank}(X_j) = 1 \ \forall j, k = 1, \ldots, m \right\}. \end{aligned} \tag{10.54}$$

Consider now the operator

$$P_{D \cap S_{\leq 1}^m} P_{\mathcal{B}}. \tag{10.55}$$

First, we need an explicit formula for the projector $P_{D \cap S_{\leq 1}^m}$, which will be given by the following two lemmata.

Lemma 10.4.1. *For all* $X = (X_1, \ldots, X_m) \in \mathbb{C}^{m(n \times n)}$ *with* $X_j \in \mathbb{C}^{n \times n}$ *for all* $1 \leq j \leq m$, *we have* $P_{S_{\leq 1}^m} P_D X \in D \cap S_{\leq 1}^m$.

Proof. The projection onto D is given by

$$P_D X = \begin{pmatrix} \frac{1}{m} \sum_{j=1}^m X_j \\ \vdots \\ \frac{1}{m} \sum_{j=1}^m X_j \end{pmatrix}.$$

To get the projection onto $S_{\leq 1}{}^m$, we compute the left singular vectors and right singular vectors, i.e., the eigenvectors of $P_D X (P_D X)^*$ and $(P_D X)^* P_D X$. Define

$$A := \frac{1}{m} \sum_{j=1}^{m} X_j \left(\frac{1}{m} \sum_{j=1}^{m} X_j \right)^*.$$

The matrix $P_D X (P_D X)^*$ can be written as

$$\begin{pmatrix} A & \cdots & A \\ \vdots & \ddots & \vdots \\ A & \cdots & A \end{pmatrix}$$

and

$$(P_D X)^* P_D X = \left(\frac{1}{m} \sum_{j=1}^{m} X_j \left(\sum_{j=1}^{m} X_j \right)^* \right).$$

Observe that $(P_D X)^* P_D X$ is just a scalar multiple of one of the $n \times n$ blocks of the matrix $P_D X (P_D X)^*$. If v is an eigenvector of $(P_D X)^* P_D X$, then it is also an eigenvector of an $n \times n$ block of the matrix $P_D X (P_D X)^*$. In other words, if $v \in \mathbb{C}^n$ is a right singular vector of $P_D X$, then $(v, \ldots, v)^\top \in \mathbb{C}^{mn}$ is a left singular vector of $P_D X$ with the same singular value. If now σ_1 is the largest singular value of $P_D X$ with left singular vector $(v, \ldots, v)^\top$ and right singular vector v, then we have

$$\begin{pmatrix} \sigma_1 v v^* \\ \vdots \\ \sigma_1 v v^* \end{pmatrix} \in P_{S_{\leq 1}{}^m} P_D X.$$

Note, that the left hand side is also contained in D. This shows the claim. $\qquad\square$

Lemma 10.4.2. *For all $X \in \mathbb{C}^{m(n \times n)}$, we have $P_{S_{\leq 1}{}^m} P_D X \in P_{D \cap S_{\leq 1}{}^m} X$.*

Proof. Note that, by Lemma 10.4.1, we have $P_{S_{\leq 1}{}^m} P_D X \in D \cap S_{\leq 1}{}^m$. We show that

$$\left\| X - P_{S_{\leq 1}{}^m} P_D X \right\|_F^2 = \left\| X - P_{D \cap S_{\leq 1}{}^m} X \right\|_F^2.$$

Because $P_{S_{\leq 1}{}^m} P_D X \in D$, and because D is an affine subspace, we have

$$\left\langle X - P_D X, P_D X - P_{S_{\leq 1}{}^m} P_D X \right\rangle = 0,$$

As a consequence, we obtain

$$\left\| X - P_{S_{\leq 1}{}^m} P_D X \right\|_F^2 = \left\| X - P_D X \right\|_F^2 + \left\| P_D X - P_{S_{\leq 1}{}^m} P_D X \right\|_F^2.$$

Further, we have, because $P_{D \cap S_{\leq 1}{}^m} X \in D$ as well,

$$\left\| X - P_{D \cap S_{\leq 1}{}^m} X \right\|_F^2 = \left\| X - P_D X \right\|_F^2 + \left\| P_D X - P_{D \cap S_{\leq 1}{}^m} X \right\|_F^2.$$

Because of the definition of the projection operator, we have

$$\left\| P_D X - P_{S_{\leq 1}{}^m} P_D X \right\|_F^2 \leq \left\| P_D X - P_{D \cap S_{\leq 1}{}^m} X \right\|_F^2,$$

and hence

$$\left\| X - P_{S_{\leq 1}{}^m} P_D X \right\|_F^2 \leq \left\| X - P_{D \cap S_{\leq 1}{}^m} X \right\|_F^2.$$

Because $P_{S_{\leq 1}{}^m} P_D X \in D \cap S_{\leq 1}{}^m$, we get equality and so $P_{S_{\leq 1}{}^m} P_D X \in P_{D \cap S_{\leq 1}{}^m} X$. $\qquad \square$

Proposition 10.4.3. *Let $X = (X_1, \ldots, X_m) \in \mathbb{C}^{m(n \times n)}$ be arbitrary with $X_j \in \mathbb{C}^{n \times n}$ for all $1 \leq j \leq m$. Then the relation*

$$P_{D \cap S_{\leq 1}{}^m} P_{\mathcal{B}} X = P_{S_{\leq 1}{}^m} P_D P_{\mathcal{B}} X \tag{10.56}$$

holds. Further, for $(Z_1, \ldots, Z_m) \in P_{D \cap S_{\leq 1}{}^m} P_{\mathcal{B}} X$ with $Z_j \in \mathbb{C}^{n \times n}$ for all $1 \leq j \leq m$, all blocks Z_j are Hermitian whenever X_j is Hermitian for all j.

Proof. The first claim follows immediately from Lemma 10.4.2. The projector $P_{S_{\leq 1}{}^m} P_D$ is composed by taking sums and scalar multiples of $n \times n$ blocks, which preserves hermiticity, together with rank one projections of $n \times n$ blocks, which preserves hermiticity as well. This proves the proposition. $\qquad \square$

In the following, we will apply Theorem 5.2.3 to the problem of rank minimization. We note that Theorem 5.2.3 is only formulated for real vector spaces. Hence, we restrict ourselves to the space $\mathbb{R}^{m(n \times n)}$.

Let $\mathcal{A} : \mathbb{R}^{m(n \times n)} \to \mathbb{R}^{mn}$ be the linear mapping such that

$$\mathcal{B} = \left\{ X \in \mathbb{R}^{m(n \times n)} \ \middle| \ \mathcal{A}(X) = b \right\}.$$

Assume now for our case that

$$(1 - \delta) \left\| X \right\|_F^2 \leq \left\| \mathcal{A}^\dagger \mathcal{A}(X) \right\|_F^2 \quad \forall X \in S_{\leq 2}{}^m. \tag{10.57}$$

We introduce the function

$$f(X) := \frac{1}{2} \left\| X - P_{\mathcal{B}} X \right\|_F^2. \tag{10.58}$$

We are now ready to prove a main result of this chapter.

Theorem 10.4.4 (global convergence of alternating projections in rank-affine feasibility). *Let $\mathcal{A}$ satisfy (10.57) with $\delta \in [0, \frac{1}{2})$. Then $\mathcal{B} \cap S_{\leq 1}{}^m \cap D$ is a singleton and, for any initial*

value $X^0 = (X_1^0, \ldots X_m^0) \in S_{\leq 1}{}^m \cap D$ with positive semidefinite and Hermitian $n \times n$ blocks X_j^0, the sequence $(X^k)_{k \in \mathbb{N}}$ generated by alternating projections with $X^{k+1} \in P_{S_{\leq 1}{}^m} P_D P_{\mathcal{B}} X^k$ converges to $\mathcal{B} \cap S_{\leq 1}{}^m \cap D$. Moreover, $d_{\mathcal{B}}(X^k) \to 0$ as $k \to \infty$ at a linear rate with the constant bounded by $\sqrt{\frac{\delta}{1-\delta}}$.

Proof. The proof is similar to the one of Corollary 5.2.4. We check if the assumptions of Theorem 5.2.3 are satisfied. As in the formulation of this theorem, we write a matrix $X \in \mathbb{R}^{m(n \times n)}$ as $X = (X_1, \ldots, X_m)$, where $X_j \in \mathbb{R}^{n \times n}$ for all $1 \leq j \leq m$. We define the function

$$\mathrm{mrank} : \mathbb{R}^{m(n \times n)} \to \mathbb{R}, \quad X = (X_1, \ldots, X_m) \mapsto \max_{j \in \{1,\ldots,m\}} \{\mathrm{rank}(X_j)\}. \qquad (10.59)$$

Since the function rank is, by Lemma 8.1.2, lower semicontinuous, then so is the function mrank as the maximum of lower semicontinuous functions. To see this directly, note that the lower levelsets of mrank are given by

$$\mathrm{lev}_{\leq r} \mathrm{mrank} = \left\{ (X_1, \ldots, X_m) \mid \mathrm{rank}(X_j) \leq r \quad \text{for all } 1 \leq j \leq m \right\}.$$

We observe that

$$\mathrm{lev}_{\leq r} \mathrm{mrank} = \bigcup_{j=1}^{m} \mathrm{lev}_{\leq r} \mathrm{rank}.$$

Hence, the set $\mathrm{lev}_{\leq r} \mathrm{mrank}$ is the m-fold direct product of the closed sets $\mathrm{lev}_{\leq r} \mathrm{rank}$, which is equivalent to $\mathrm{lev}_{\leq r} \mathrm{mrank}$ being a closed set. Then, by Definition 2.1.7, the function mrank is lower semicontinuous. Additionally, mrank satisfies $\mathrm{mrank}(X) = \mathrm{mrank}(-X)$ for all $X \in \mathbb{R}^{m(n \times n)}$. Further, observe that mrank is also subadditive since the rank function is subadditive (Lemma 8.1.1). Finally, note that $\mathrm{lev}_{\leq r} \mathrm{mrank} = S_{\leq 1}{}^m$.

Futher, define the function

$$\varphi : \mathbb{R}^{m(n \times n)} \to \mathbb{R}, \quad X \mapsto \mathrm{mrank}(X) + \iota_D(X).$$

Since D is a closed subspace, the indicator function ι_D is lower semicontinuous as well, and it satisfies $\iota_D(X) = \iota_D(-X)$ for all $X \in \mathbb{R}^{m(n \times n)}$. Observe also that ι_D is subadditive.

Altogether, we obtain the conclusion that φ satisfies the assumptions of Theorem 5.2.3. Since the condition (10.57) is sufficient for φ to satisfy the assumption in (5.3), we can apply Theorem 5.2.3 to get a single point in $S_{\leq 1}{}^m \cap D \cap \mathcal{B}$.

<div align="right">□</div>

11 Numerical Examples

11.1 Sparsity Optimization

We demonstrate the results in Chapter 5 on the following synthetic numerical examples. Their construction is similar to (Hesse et al., 2014, Chapter V.A). We construct a sparse object with 1311 uniform random positive and negative point-like sources in a 512-by-512 pixel field and randomly sample the Fourier transform of this object at a ratio of 1-to-8. This yields 32768 affine constraints. Local convergence results are illustrated in Figures 11.1 and 11.2. In these experiments, the initial points x^0 are selected by uniform random $(-\delta/2, \delta/2)$ perturbations of the true solution in order to satisfy the assumptions of Theorems 5.1.1 and 5.5.3. The alternating projections and Douglas-Rachford algorithms are shown in panels (a)-(b) of Figure 11.1 and (c)-(d) of Figure 11.2, respectively. We show both the step lengths per iteration as well as the *gap distance* at each iteration defined as

$$(\text{gap distance})^k := \|P_{A_s} x^k - P_B x^k\|. \tag{11.1}$$

Monitoring the gap allows one to ensure that the algorithm is indeed converging to a point of intersection instead of just a best approximation pair. We use the sparsity parameters $s \in \{1311, 1450\}$. We demonstrate the effect of overestimating the sparsity parameter, $s = 1450$.

The second collection of synthetic examples, shown in Figures 11.7 and 11.8, demonstrates global performance of the algorithms and illustrates the results in Theorem 5.1.1, Theorem 5.5.3, and Corollary 5.2.10. The solution is the vector

$$\bar{x} := (10, 0, 0, 0, 0, 0, 0, 0)^\top,$$

and the affine subspace is the one generated by the matrix in (5.38). This matrix fulfills the assumptions of Corollary 5.2.10, as shown in Section 5.4.1. For the cases Figure 11.7 (a) and Figure 11.8 (a), the initial point x^0 can be written as $x^0 \equiv \bar{x} + u$ where u is a vector with uniform random values from the interval $(-1, 1)$. The initial values hence fulfill the assumptions of Theorems 5.1.1 and 5.5.3. For Figure 11.7 (b) and Figure 11.8 (b), the initial point x^0 can again be written as $x^0 \equiv \bar{x} + u$ while u is now a vector with uniform random values from the interval $(-100, 100)$. As expected, the sequence of alternating projections converges to the true solution in Figure 11.8 (a). The case for Douglas-Rachford however, shown in Figure 11.8 (b), is not covered by our theory.
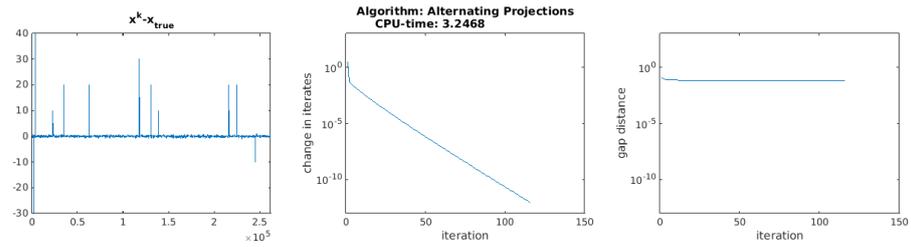
Figure 11.1: The plot shows the convergence of alternating projections in the case where the sparsity is underestimated, $s = 1300$.
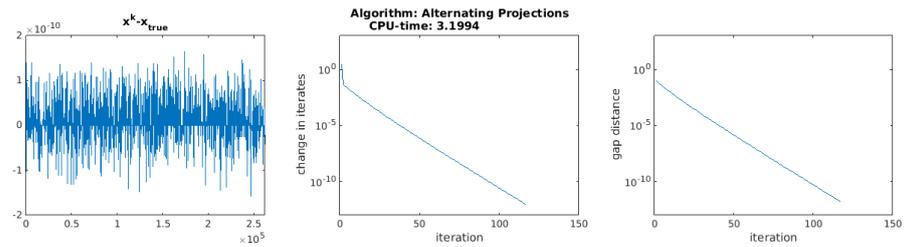

Figure 11.2: The plot shows the convergence of alternating projections in the case where the sparsity is exact, $s = 1311$.
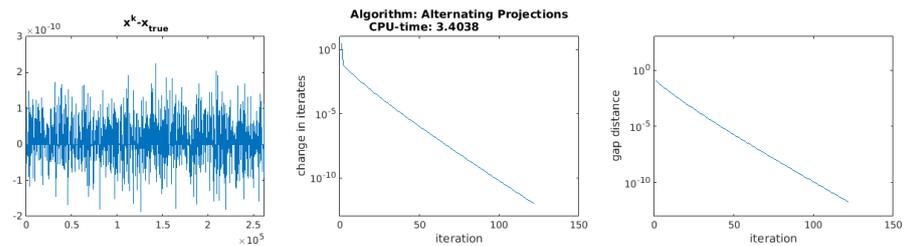

Figure 11.3: The plot shows the convergence of alternating projections in the case where the sparsity is overestimated, $s = 1450$. As expected, the convergence is robust to this uncertainty.
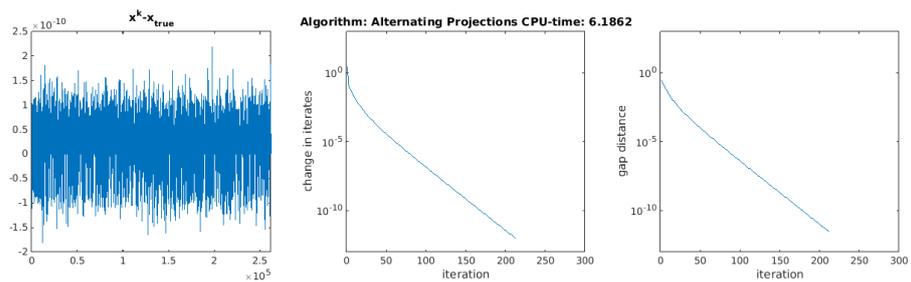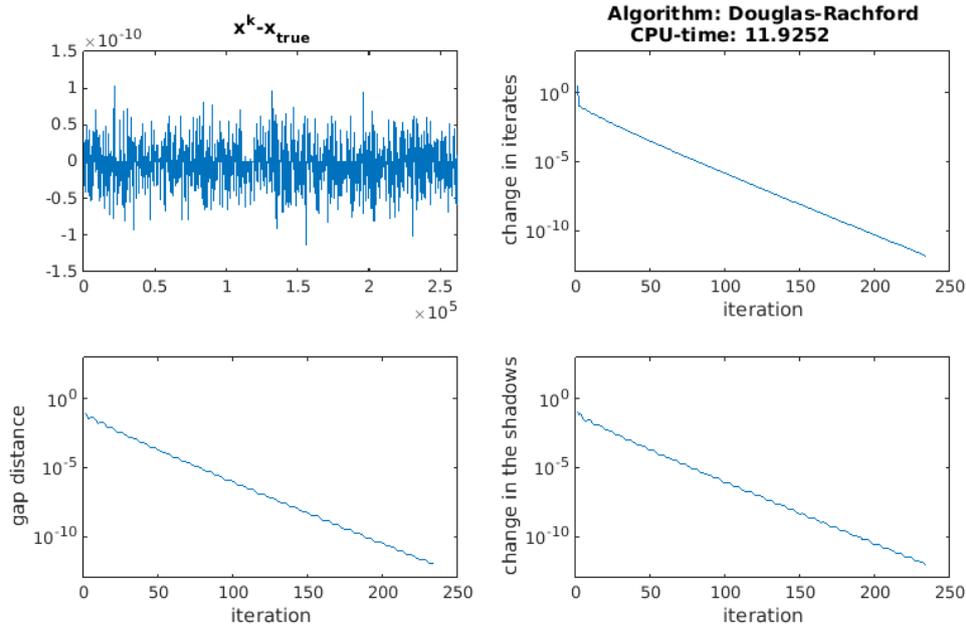

Figure 11.4: The plot shows the convergence of alternating projections in the case where the sparsity is overestimated, $s = 4000$. The speed of convergence is significantly slower than in Figure 11.3.

Figure 11.5: The plot shows the convergence of Douglas-Rachford in the case where the sparsity is exact, $s = 1311$.
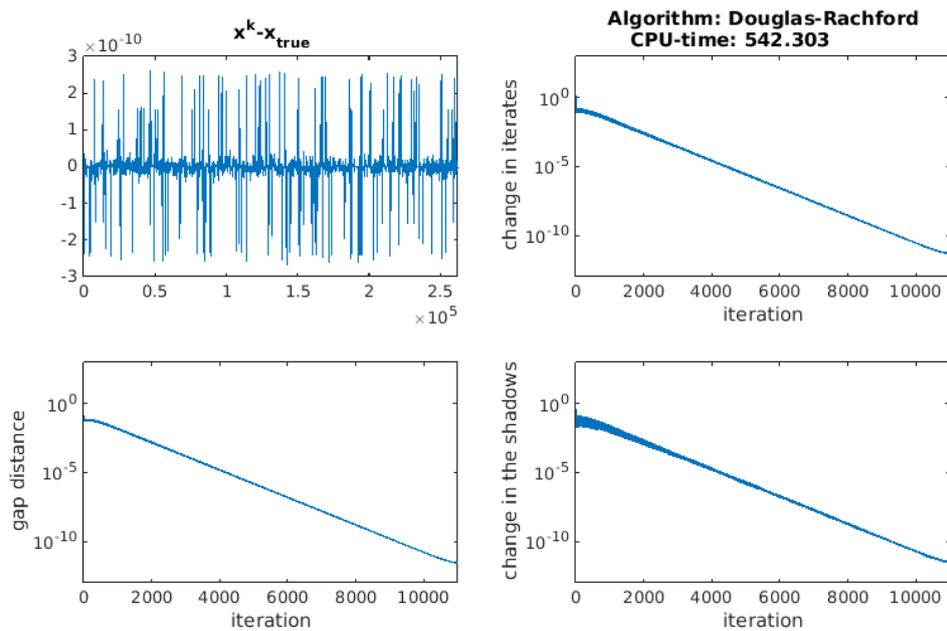
Figure 11.6: The plot shows the convergence of Douglas-Rachford in the case where the sparsity is overestimated, $s = 1450$. Note that, though the theory is not developed, there is a yet very slow convergence to the true solution.
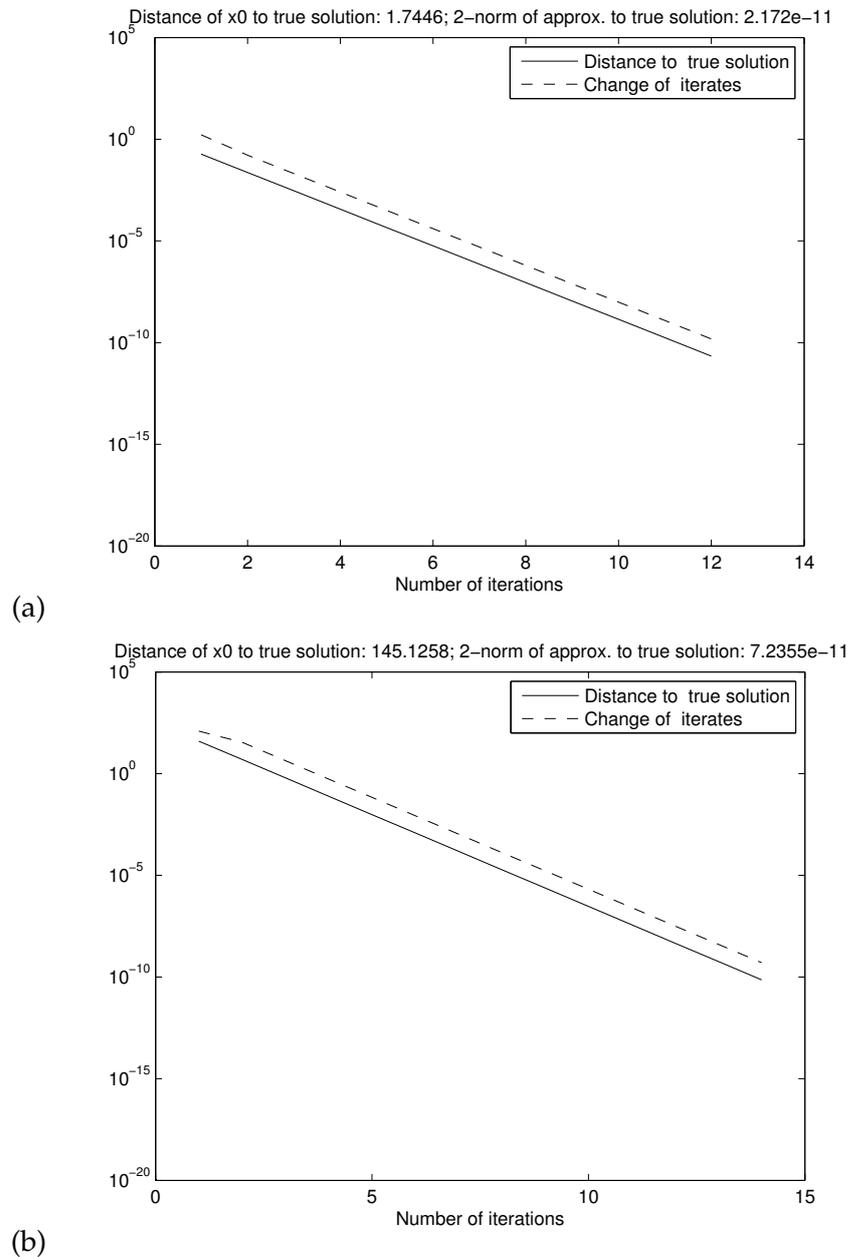
Distance of x0 to true solution: 1.7446; 2–norm of approx. to true solution: 2.172e−11

(a)

Distance of x0 to true solution: 145.1258; 2–norm of approx. to true solution: 7.2355e−11

(b)

Figure 11.7: Example with an affine subspace generated by the matrix from Section 5.4.1: (a) shows the local convergence as shown in Theorem 5.1.1, (b) is an example of global convergence of alternating projections as stated in Corollary 5.2.10.
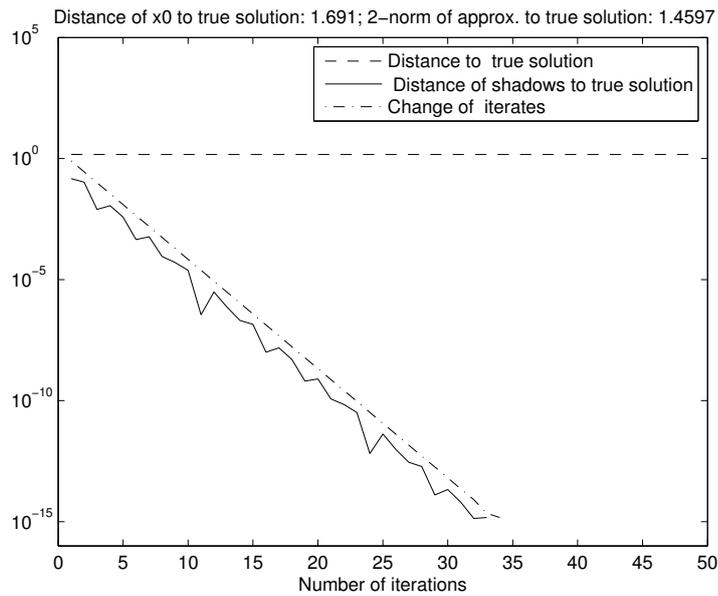
(a)
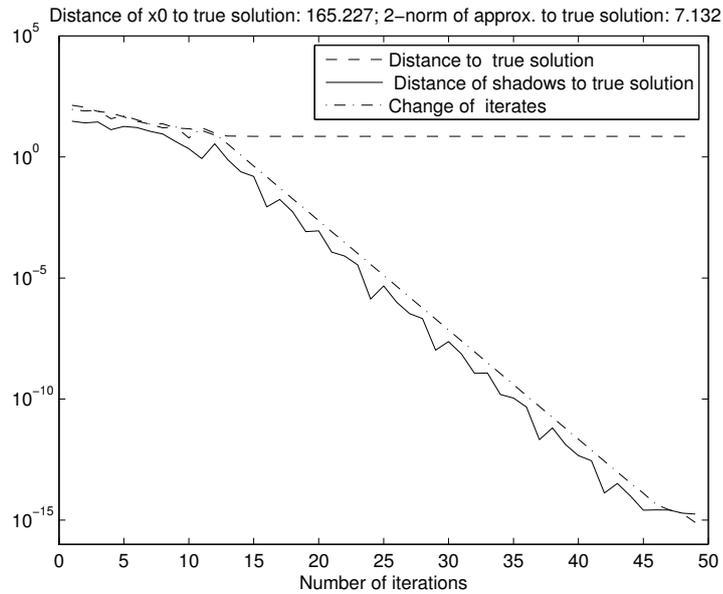
(b)

Figure 11.8: Example with an affine subspace generated by the matrix from Section 5.4.1: (a) is an example of local convergence of Douglas-Rachford to its fixed point set while the shadows converge to the intersection, as proven in Theorem 5.5.3. This example also shows that the iterates converge to a fixed point that is not in the intersection, as proven in Theorem 5.5.1. Plot (b) is an example where Douglas-Rachford appears to converge globally. This behavior is not covered by our theory.

11.2 Rank Minimization and Phase Retrieval

We turn our attention to the problem of phase retrieval. We demonstrate, why an implementation of rank minimization is not to be favored over the classical approaches. Evidently, computing a projection in the lifted space $\mathbb{C}^{n \times n}$ instead of $\mathbb{C}^n$ requires more computation time. This time increases by going into the space $\mathbb{C}^{m(n \times n)}$ by taking several measurements as in Theorem 10.4.4. Because of that, we compare the classical Error Reduction algorithm as in Definition 9.3.2 with a lifted version of Error Reduction. First, we show a technical lemma.

Lemma 11.2.1. *Let $\mathcal{C}_s$ be the lift of the support set in Lemma 8.3.5. Define now*

$$C_1 := \mathcal{C}_s \cap \mathbb{R}^{n \times n}. \tag{11.2}$$

Then, for all $X \in \mathbb{C}^{n \times n}$, we have $P_{S_{\leq 1} \cap C_1} X = P_{S_{\leq 1}} P_{C_1} X$.

Proof. We note that C_1 is a linear subspace of $\mathbb{C}^{n \times n}$. We show that, if $Y \in C_1$, then $P_{S_{\leq 1}} Y \in C_1$. By Theorem 8.2.2, we can choose $P_{S_{\leq 1}} Y \in \mathbb{R}^{n \times n}$. Let I be the index set corresponding to the support, as in Lemma 8.3.5. If $Y \in \mathcal{C}_s$ and if $j \notin I$, then both the jth row and the jth column are zero. Without loss of generality, we assume that $I = \{1, \dots, m\}$ with $m \leq n$. Hence, only the top left $m \times m$ block of Y is different from zero. Let $U \Sigma V^\top$ be a singular value decomposition of Y. Because only the top left block of Y is different from zero, only the first m entries in the left singular vectors $u_1, \dots, u_m$ of Y, and the right singular vectors $v_1, \dots, v_m$ respectively, can be different from zero. Only the top left $m \times m$ block of $P_{S_{\leq 1}} = \sigma_1(Y) \cdot u_1 v_1^*$ is different from zero. Hence, we have $P_{S_{\leq 1}} Y \in \mathcal{C}_s$. We conclude that, if $Y \in C_1$, then $P_{S_{\leq 1}} Y \in C_1$. Further, for all $Y \in C_1$, we have $Y - P_{S_{\leq 1}} Y \in C_1$.

Now, since C_1 is a linear subspace of $\mathbb{C}^{n \times n}$, the operator P_{C_1} is an orthogonal projection. Let $X \in \mathbb{C}^{n \times n}$. We get

$$\begin{aligned}
\left\| X - P_{S_{\leq 1} \cap C_1} X \right\|_F^2 &= \left\| X - P_{C_1} x \right\|_F^2 + \left\| P_{C_1} X - P_{S_{\leq 1} \cap C_1} X \right\|_F^2 \\
&\geq \left\| X - P_{C_1} x \right\|_F^2 + \left\| P_{C_1} X - P_{S_{\leq 1}} P_{C_1} X \right\|_F^2 \\
&= \left\| X - P_{S_{\leq 1}} P_{C_1} X \right\|_F^2.
\end{aligned}$$

But since $P_{S_{\leq 1}} P_{C_1} X \in S_{\leq 1} \cap C_1$, we conclude that

$$\left\| X - P_{S_{\leq 1} \cap C_1} X \right\|_F^2 = \left\| X - P_{S_{\leq 1}} P_{C_1} X \right\|_F^2.$$

In other words, $P_{S_{\leq 1} \cap C_1} X = P_{S_{\leq 1}} P_{C_1} X$. This finishes the proof. $\square$

With Lemma 11.2.1, we can formulate a sequential projection iteration between the sets C_1, $S_{\leq 1}$, and B as an alternating projections iteration between $C_1 \cap S_{\leq 1}$ and $\mathcal{B}$, namely, for X^0 given, we generate the sequence $\{X^k\}_{k \in \mathbb{N}}$ via

$$X^{k+1} = P_{S_{\leq 1}} P_{C_1} P_{\mathcal{B}} X^k.$$

We remind the reader that by going into the lifted space $\mathbb{C}^{n \times n}$, we are acting in a vector space of squared dimension. As a setup, we take a shrunken version of the Siemens star, a 32 by 32 pixel image with just $0, 1$ entries. In the lifted space, we obtain a 1024 by 1024 matrix. As inintial point, we generate a random 32 by 32 pixel image satisfying the magnitude constraint, defined as in Equation (9.2).
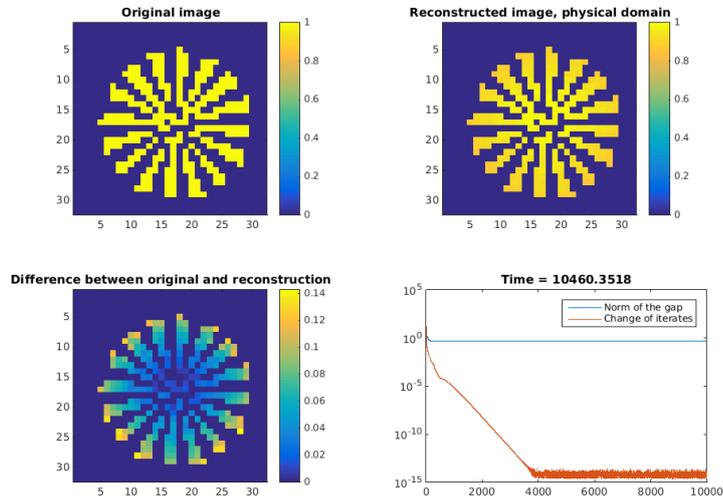

Figure 11.9: Alternating projections in phase lift with the Siemens Star using support and real constraints. The phases of the reconstructed image are all zero since we always project onto the real parts of the image. Also the phases of the projections onto the magnitude constraints in $\mathbb{C}^{32 \times 32}$ turn out to be zero.
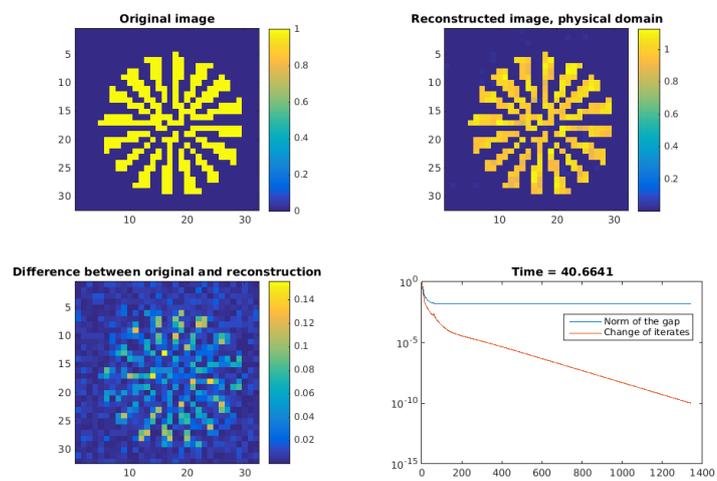
Figure 11.10: The alternating projections algorithm, also known as Error Reduction, in phase retrieval with the Siemens Star using support and real constraints.

12 Conclusion and Outlook

We split this review according to the two parts of this thesis into one for sparsity optimization and into one for rank minimization.

12.1 Sparsity Optimization

For the method of alternating projections in sparse affine feasibility, we have three kinds of results: first, there is Theorem 5.1.1 showing local linear convergence only with the restriction that the intersection $A_s \cap B$ is not empty. Secondly, there is Theorem 5.2.3 as the foundation for convergence results which requires strong conditions, namely a restricted isometry property, to be applied. Lastly, there is Theorem 6.3.5 giving necessary and sufficient conditions for global convergence of alternating projections. The question is now how to bring all these results together. We think that there exists a formulation of the conditions in Theorem 6.3.5 that are in the sense of Theorem 5.2.3. Further, encouraged by our numercial observations, we think that the assumptions in the latter can be weakened.

Conjecture 12.1.1. *If there is a condition on the linear mapping M such that the solution to the ℓ_1-relaxation of the compressed sensing problem (3.4) is also a solution to (3.1.3) then there exists a sparsity parameter s such that the method of alternating projections finds the true solution in $A_s \cap B$ for all initial points x^0.*

Also the behavior of the Douglas-Rachford algorithm in the sparse affine feasibility problem can be studied in a deeper way. In particular, we have seen in Theorem 5.5.3 that in the case of an exact a priori assumption on the sparsity parameter s, we have a local linear convergence of the iterates generated by the Douglas-Rachford algorithm to the true solution. In applications, we cannot expect that this assumption on s is always accurate. The experiment run to generate Figure 11.6 shows that Douglas-Rachford is very sensitive to the accuracy on s.

Also, we have an analysis on the outcome of the application of alternating projections to the sparse polyhedral feasibility problem. Without any further assumptions, we have shown that this sequence always converges to a fixed point. There are two possible ways for additional research:

1. Find sufficient conditions on the polyhedral set such that the sequence generated by alternating projections converges to the intersection or, if the latter is empty, to a best approximation pair attaining the minimal distance between $\mathcal{C}$ and A_s.

2. Is there a way to predict the outcome of the Douglas-Rachford algorithm if applied to the sparse polyhedral feasibility problem? If the intersection $\mathcal{C} \cap A_s$ is empty, then, as soon as the sequence generated by Douglas-Rachford approaches a best approximation pair, the sequence diverges in direction of the gap vector. But since A_s is not convex, this controlled divergence has to stop at some point.

12.2 Rank Minimization and Spectral Sets

We start by giving remarks on Chapter 7. In the spirit of (Daniilidis et al., 2008), we think that the transfer principle also applies for regularities that are weaker than prox-regularity.

Conjecture 12.2.1. *The transfer principle applies also to (ε, δ)-subregularity.*

As a point to start this analysis, we propose the investigation of (ε, δ)-subregularity of spectral sets. This would especially give rise to local regularity conditions of the set $S_{<r}$. This would give us formulations in the sense of Theorem 3.2.5. With the application of this formulation to Theorem 7.3.1, we would be able to prove the converse direction. We believe that it is necessary to apply differential geometric arguments to obtain this converse direction.

In the field of rank minimization, a possible point for further research would be a deeper analysis of fixed points of projection operators. This allows the derivation of sufficient conditions for global convergence of alternating projections and possible bridges between the lifted version and the classical alternating projections in the phase retrieval problem.

So far, Theorem 10.3.7 has shown that, eventually, the sequence of alternating projections just between the lifted magnitude set (in the proof it is the lifted amplitude set, but with respect to convergence, this is equivalent) and the set of rank one matrices converges to the lift of the magnitude projection given in Lemma 9.2.3. In terms of applications, it is important to expand this analysis to a more general setting. This requires a deeper knowledge of the regularities of $S_{\leq 1}$ and of the intersections.

Bibliography

Aragón Artacho, F., Borwein, J., and Tam, M. (2013). Douglas-rachford Feasibility Methods for Matrix Completion Problems. *Australian and New Zealand Industrial and Applied Mathematics Journal*, 55:299–326.

Bartels, M., Krenkel, M., Haber, J., Wilke, R. N., and Salditt, T. (2015). X-ray Holographic Imaging of Hydrated Biological Cells in Solution. *Physical Review Letters*, 114.

Bauschke, H. H., Bello Cruz, J. Y., Nghia, T. T. A., Phan, H. M., and Wang, X. (2014a). The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle. *Journal of Approximation Theory*, 185:63–79.

Bauschke, H. H., Borwein, J., and Lewis, A. S. (1997). The method of cyclic projections for closed convex sets in hilbert space. *Contemporary Mathematics*, 204:1–38.

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*. CMS Books in Mathematics. Springer-Verlag, New York.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2002). Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *Journal of the Optical Society of America A*, 19(7):1334–1345.

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(2):178–314.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. (2006). A strongly convergent reflection method for finding the projection onto the intersection of two closed convex sets in a Hilbert space. *Journal of Approximation Theory*, 141(1):63–69.

Bauschke, H. H. and Koch, V. (2015). Projection methods: Swiss Army knives for solving feasibility and best approximation problems with halfspaces. In Reich, S. and Zaslavski, A. J., editors, *Infinite Products and Their Applications*, volume 636 of *Contemporary Mathematics*. American Mathematical Society.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2013a). Restricted Normal Cones and the Method of Alternating Projections: Applications. *Set-Valued and Variational Analysis*, 21(3):475–501.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2013b). Restricted Normal Cones and the Method of Alternating Projections: Theory. *Set-Valued and Variational Analysis*, 21(3):431–473.

Bauschke, H. H., Luke, D. R., Phan, H. M., and Wang, X. (2014b). Restricted Normal Cones and Sparsity Optimization with Affine Constraints. *Foundations of Computational Mathematics*, 14(1):63–83.

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*, volume 49 of *Springer Optimization and Its Applications*, pages 33–48. Springer, New York.

Blumensath, T. and Davies, M. (2009). Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274.

Blumensath, T. and Davies, M. (2010). Normalised iterative hard thresholding; guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309.

Bonnesen, T. and Fenchel, W. (1934). *Theorie der konvexen Körper*. Springer-Verlag, Berlin Heidelberg New York.

Borwein, J. M. and Sims, B. (2011). The Douglas-Rachford Algorithm in the Absence of Convexity. 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*, volume 49 of *Springer Optimization and Its Applications*, pages 93–109. Springer, New York.

Boyle, J. and Dykstra, R. (1986). A Method for Finding Projections onto the Intersection of Convex Sets in Hilbert Spaces. *Advances in Order Restricted Statistical Inference, Lecture Notes in Statistics*, 37:28–47.

Burke, J. and Luke, D. R. (2003). Variational analysis applied to the problem of optical phase retrieval. *SIAM Journal on Control and Optimization*, 42(2):576–595.

Candès, E. J., Eldar, Y., Strohmer, T., and Voroninski, V. (2011). Phase Retrieval via Matrix Completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225.

Candès, E. J. and Tao, T. (2005). Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.

Cheney, E. W. and Goldstein, A. A. (1959). Proximity Maps for Convex Sets. *Proceedings of the American Mathematical Society*, 10(3):448–450.

Combettes, P. L. and Trussell, H. J. (1990). Method of Successive Projections for Finding a Common Point of Sets in Metric Spaces. *Journal of Optimization Theory and Applications*, 67(3):487–507.

Condat, L. and Hirabayashi, A. (2015). Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements. *Sampling Theory in Signal and Image Processing*, 14:17–47.

Daniilidis, A., Lewis, A. S., Malick, J., and Sendov, H. (2008). Prox-Regularity of Spectral Functions and Spectral Sets. *Journal of Convex Analysis*, 15(3):547–560.

Demanet, L. and Zhang, X. (2013). Eventual linear convergence of the Douglas Rachford iteration for basis pursuit. arXiv:1301.0542.

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.

Dixmier, J. (1949). Étude sur les variétés et les opérateurs de Julia, avec quelques applications. *Bulletin de la Société Mathématique de France*, 77:11–101.

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.

Dontchev, A. and Rockafellar, R. T. (2014). *Implicit Functions and Solution Mappings, 2nd ed.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York.

Douglas, J. and Rachford, H. H. (1956). On the numerical solution of heat conduction problems in two or three space variables. *Transactions of the American Mathematical Society*, 82:421–439.

Drenth, A., Huiser, A., and Ferwerda, H. (1975). The Problem of Phase Retrieval in Light and Electron Microscopy of Strong Objects. *Optica Acta*, 22(7):615–628.

Drusvyatskiy, D., Li, C.-K., Pelejo, D., Voronin, Y.-L., and Wolkowicz, H. (2015). Projection methods for quantum channel construction. *Quantum Information Processing*, to appear.

Eckart, C. and Young, G. (1936). The Approximation of one Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218.

Farn, M. W. (1991). New iterative algorithm for the design of phase-only gratings. *Proceedings SPIE 1555*, 34:34–42.

Fenchel, W. (1951). Convex Cones, Sets and Functions. Lecture Notes, Princeton University.

Fienup, J. (1980). Iterative method applied to image reconstruction and to computer-generated holograms. *Optical Engineering*, 19:297–305.

Fienup, J. (1982). Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769.

Fienup, J. and Wackerman, C. (1986). Phase-retrieval stagnation problems and solutions. *Journal of the Optical Society of America A*, 3(11):1897–1907.

Foucart, S. and Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel.

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.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Wh. H. Freeman & Co., New York.

Gerchberg, R. W. and Saxton, W. O. (1972). A Practical Algorithm for the Determination of the Phase from Image and Diffraction Plane Pictures. *Optik*, 35(2):237–246.

Gubin, L. G., Polyak, B. T., and Raik, E. (1967). The method of projections for finding the common point of convex sets. *Computational Mathematics and Mathematical Physics*, 7(6):1–24.

Hayes, M. (1982). The Reconstruction of a Multidimensional Sequence from the Phase or Magnitude of Its Fourier Transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 30(2):140–154.

Hesse, R. (2014). *Fixed point algorithms for nonconvex feasibility with applications*. Niedersächsische Staats- und Universitätsbibliothek Göttingen.

Hesse, R., Luke, D., and Neumann, P. (2014). Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility. *IEEE Transactions on Signal Processing*, 62(18):4868–4881.

Hesse, R. and Luke, D. R. (2013). Nonconvex Notions of Rregularity and Convergence of Fundamental Algorithms for Feasibility Problems. *SIAM Journal on Optimization*, 23(4):2397–2419.

Hiriart-Urruty, J.-B. (2013). When only global optimization matters. *Journal of Global Optimization*, 56(3):761–763.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*. Cambridge University Press.

Ioffe, A. D. (2000). Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501.

Jordan, C. (1875). Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France*, 3:103–174.

Kaczmarz, S. (1937). Angenäherte Auflösung von Systemen Linearer Gleichungen. *Bulletin de l'Académie des Sciences de Pologne*, 35:335–357.

Kassay, G., Pintea, C., and Szenkovits, F. (2009). On Convexity of Preimages of Monotone Operators. *Taiwanese Journal of Mathematics*, 13(2B):675–686.

Krahmer, F. (2014). Personal Communication.

Kruger, A. Y. (2006). About Regularity of Collections of Sets. *Set-Valued Analysis*, 14:187–206.

Kyrillidis, A. and Cevher, V. (2014). Matrix Recipes for Hard Thresholding Methods. *Journal of Mathematical Imaging and Vision*, 28(2):235–265.

Le, H. (2013). Generalized subdifferentials of the rank function. *Optimization Letters*, 7(4):731–743.

Lee, J. M. (2003). *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York.

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(9):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. *Mathematics of Operations Research*, 33(1):216–234.

Lewis, A. S. and Sendov, H. (2005). Nonsmooth Analysis of Singular Values, Part I: Theory. *Set-Valued Analysis*, 13:213–241.

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.

Luke, D. R. (2005). Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21(1):37–50.

Luke, D. R. (2008). Finding Best Approximation Pairs Relative to a Convex and Prox-Regular Set in a Hilbert Space. *SIAM Journal on Optimization*, 19(2):714–739.

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. (2015). Personal Communication.

Luke, D. R., Borwein, J., and Lyons, R. (2002). Optical Wavefront Reconstruction: Theory and Numerical Methods. *SIAM Review*, 44(2):169–224.

Miao, J., Charalambous, P., Kirz, J., and Sayre, D. (1999). Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400:342–344.

Millane, R. (1990). Phase retrieval in crystallography and optics. *Journal of the Optical Society of America A*, 7(3):394–411.

Moore, E. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26:394–395.

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

Mordukhovich, B. and Rockafellar, R. T. (2012). Second-order subdifferential calculus with applications to tilt stability in optimization. *SIAM Journal on Optimization*, 22(3):953–986.

Moreau, J.-J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299.

Natarajan, B. (1995). Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, 24(2):227–234.

Needell, D. and Tropp, J. (2009). Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321.

Ngai, H. V. and Théra, M. (2001). Metric Inequality, Subdifferential Calculus and Applications. *Set-Valued Analysis*, 9:187–216.

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51:406–413.

Phan, H. M. (2015). Linear convergence of the douglas–rachford method for two closed sets.

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

Pierra, G. (1984). Decomposition through formalization in a product space. *Mathematical Programming*, 28(1):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.

Recht, B., Fazel, M., and Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501.

Rockafellar, R. T. (1997). *Convex Analysis*. Princeton University Press.

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

tom Dieck, T. and Bröcker, T. (2003). *Representations of Compact Lie Groups*. Springer-Verlag, Berlin Heidelberg New York.

Vandereycken, B. (2013). Low-Rank Matrix Completion by Riemannian Optimization. *SIAM Journal on Optimization*, 23(2):1214–1235.

von Neumann, J. (1951). *Functional Operators, Volume 2: The Geometry of Orthogonal Spaces*. Princeton University Press.

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

Patrick Neumann, M.Sc.

Address	Institut für Numerische und Angewandte Mathematik
	Georg-August-Universität Göttingen
	Lotzestraße 16–18
	37083 Göttingen, Germany
Telephone	+49551394516
Email	Patrick.Neumann@mathematik.uni-goettingen.de
Homepage	http://www.uni-goettingen.de/de
	/patrick-neumann-msc/413677.html

Personal Details

Date of birth	March 27, 1987
Place of birth	Neuwied, Germany
Nationality	German

Academic Education

11/2011–06/2015	Doctoral student of mathematics at Göttingen University. Advisor: Prof. Dr. Russell Luke.
10/2009–10/2011	Master student of mathematics at Göttingen University, with physics as minor subject. Title of the thesis: Zeta Functions of Self-Similar Groups. Advisor: Prof. Dr. Laurent Bartholdi.
10/2006–09/2009	Bachelor student of mathematics at Göttingen University, with theoretical physics as minor subject. Title of the thesis: Durchmesser und Wachstum endlicher Heisenberg-Gruppen. Advisor: Prof. Dr. Andreas Thom.

Internship

08/2009–09/2009   Internship in the area of component optimization of machine parts at Hauni Maschinenbau AG in Hamburg, Germany.

Education

08/1997–03/2006   Secondary School "Werner-Heisenberg-Gymnasium", Neuwied.

Research Experience

11/2011–05/2015   Doctoral student at the Institute for Numerical and Applied Mathematics, Göttingen University.

08/2014–05/2015   Member of "Project C2 – Inverse scattering problems without phase" as part of the DFG Sonderforschungsbereich 755.

11/2011–06/2014   Member of the "DFG Research Training Group 1023: Identification in Mathematical Models" at the Institute for Numerical and Applied Mathematics, Göttingen University.

Publications   Robert Hesse, D. Russell Luke, and Patrick Neumann. 2014 Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility, *IEEE Transactions on Signal Processing*, 62(18): 4868–4881.

Participation in Workshops and Conferences

2015   "SFB 755 Winter School 2015 Nanoscale Photonic Imaging" in Teistungen, Germany, January 15–16, 2015.

2014   Workshop "Mathematical Signal Processing and Phase Retrieval" at Akademie Waldschlösschen, Gleichen-Reinhausen, Germany, September 01–03, 2014.

"SIAM Conference on Imaging Science 2014" at Hong Kong Baptist University, China, May 09–12, 2014.

2013   "Final Colloquium of the Research Training Group 1023" at Göttingen University, Germany, November 29–30, 2013.

"Statistical Issues in Compressive Sensing" at Göttingen University, Germany, November 11–13, 2013.

"Gene Golub SIAM Summer School 2013 on Matrix Functions and Matrix Equations" at Fudan University, Shanghai, China, July 22 – August 09, 2013.

Workshop "Signal Processing with Adaptive Sparse Structured Representations" at École Polytechnique Fédérale de Lausanne, Switzerland, July 08–11, 2013

"International Symposium Nanoscale Photonic Imaging" at Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, April 04–05, 2013.

2012   Annual Goslar Workshop of the Research Training Group 1023 in Goslar, Germany, October 17–19, 2012.

Workshop "Advances in Image Processing" at Akademie Waldschlösschen Gleichen-Reinhausen, Germany, September 04–06, 2012.

"Spring School on Analysis 2012: Variational Analysis and its Applications" in Paseky nad Jizerou, Czech Republic, April 22–28, 2012.

2010   Workshop "Espaces métriques singuliers et théorie des groupes" at Centre international de rencontres mathématiques, Marseille, France, May 26–28, 2010.

2008   Workshop "Groups and Infinite Graphs geometric group theory, group actions on graphs, infinite graph theory" at Erwin Schrödinger International Institute, Vienna, Austria, August 25–27, 2008.

Teaching Experience

2014	"Introduction to Optimization", tutorial.
2013	"Introduction to Optimization", teaching assistant.
2012/2013	"Seminar on Industrial Mathematics", "Numerical mathematics I", tutorial.
2012	"Mathematics for Computer scientists II", tutorial.
2011/2012	"Mathematics for Computer scientists I", tutorial.
2011	"Numbers and Number Theory", tutorial.
2010/2011	"Discrete Mathematics", tutorial.
2010	"Analytic Geometry and Linear Algebra II", tutorial.
2009/2010	"Analysis I", tutorial.
2009	"Analytic Geometry and Linear Algebra II", tutorial.