The Douglas–Rachford Operator in the Possibly Inconsistent Case: Static Properties and Dynamic Behaviour

by

Walaa M. Moursi

B.Sc. (with distinction), Mansoura University, Egypt, 1999
M.Sc., Mansoura University, Egypt, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE COLLEGE OF GRADUATE STUDIES
(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Okanagan)

December 2016

© Walaa M. Moursi, 2016
Abstract

The problem of finding a minimizer of the sum of two convex functions — or, more generally, that of finding a zero of the sum of two maximally monotone operators — is of central importance in variational analysis and optimization. Perhaps the most popular method of solving this problem is the Douglas–Rachford splitting method. Surprisingly, little is known about the behaviour of the algorithm in the general inconsistent case, i.e., when the sum problem has no zeros.

This thesis provides a comprehensive study of the Douglas–Rachford operator and the corresponding algorithm. First, a novel framework of the normal problem that copes with the possibly inconsistent situation is introduced and studied. We present a systematic study of the range of the operator and displacement map, which provides precise information on the minimal displacement vector needed to define the normal problem. A new Fejér monotonicity principle as well as new identities for the Douglas–Rachford operator are developed. A systematic and detailed study of affine nonexpansive operators and their asymptotic behaviour is presented. In the light of the new analysis, we significantly advance the understanding of the inconsistent case by providing a complete proof of the full weak convergence of the shadow sequence to a best approximation solution in the convex feasibility setting. In fact, a more general sufficient condition for weak convergence in the general case is presented. Under some extra assumptions, we were able to prove strong and linear convergence. Our results are illustrated through numerous examples. We conclude with a list of open problems which serve as promising directions of future research.
Preface

The research work presented in this thesis is based on the nine published papers [13], [14], [15] by Heinz Bauschke and Walaa Moursi; [26] by Heinz Bauschke, Radu Boţ, Warren Hare and Walaa Moursi; [31], [38] by Heinz Bauschke, Warren Hare and Walaa Moursi; [36], [33] by Heinz Bauschke, Minh Dao and Walaa Moursi; [107] by Sarah Moffat, Walaa Moursi and Xianfu Wang; and the two submitted manuscripts [39] by Heinz Bauschke and Walaa Moursi; and [110] by Walaa Moursi.

The results in these papers appear in the thesis as follows: Chapter 4 is based on the work in [39] and [26]; Chapter 5 is based on the work in [15], [39], [26] and [31]; Chapter 6 is based on the work in [36], [13] and [33]; Chapter 7 is based on the work in [31], [15], and [39]; Chapter 8 is based on the work in [15], [39] and [110]; Chapter 9 is based on the work in [15], [39] and [26]; Chapter 11 is based on the work in [31] and [38]; Chapter 12 and Chapter 13 are based on the work in [38]; Chapter 14 is based on the work in [13]; Chapter 15 is based on the work in [15]; and Chapter 16 is based on the work in [36].

For all co-authored papers, each author contributed equally.
# Table of Contents

Abstract ............................................................... ii
Preface ................................................................. iii
Table of Contents ..................................................... iv
List of Figures ......................................................... viii
Glossary of Notation and Symbols ................................. x
Acknowledgements .................................................... xii
Dedication .............................................................. xiii

Chapter 1: Introduction ............................................... 1
  1.1 The goal of the thesis ........................................... 1
  1.2 Contributions in this thesis ................................... 2

Chapter 2: Convex analysis: Overview ............................. 5
  2.1 Convex sets and projection operators ....................... 5
  2.2 Convex functions .............................................. 7
  2.3 Subdifferentiability and conjugation ....................... 8

Chapter 3: (Firmly) nonexpansive mappings and monotone operators ........................................... 10
  3.1 Nonexpansive and firmly nonexpansive operators .......... 10
  3.2 Monotone operators: Basic definitions and facts ........... 11
  3.3 Resolvents of monotone operators .......................... 13
  3.4 Sums of monotone operators: maximality and range of the sum ........................................... 16
  3.5 Examples of linear monotone operators ..................... 20
# Chapter 4: Attouch–Théra duality and paramonotonicity

**Overview** 29

**Duality for monotone operators** 29

**Solution mappings $K$ and $Z$** 33

**Paramonotonicity** 38

**Projection operators and solution sets** 41

**Subdifferential operators** 43

# Chapter 5: Zeros of the sum and the Douglas–Rachford operator

**Overview** 47

**Basic properties and facts** 47

**Useful identities for the Douglas–Rachford operator** 49

**Douglas–Rachford operator and Attouch–Théra duality** 51

**The Douglas–Rachford operator and solution sets** 55

**From PDEs to maximally monotone operators** 62

**Why do we need to work with monotone operators and not just subdifferential operators?** 65

# Chapter 6: On Fejér monotone sequences and nonexpansive mappings

**Overview** 67

**Fejér monotonicity: New principles** 67

# Chapter 7: Nonexpansive mappings and the minimal displacement vector

**Overview** 74

**Auxiliary results** 74

**The displacement map and the minimal displacement vector** 76

# Chapter 8: Asymptotic behaviour of nonexpansive mappings

**Overview** 79

**Iterating nonexpansive operators** 80

**Affine nonexpansive operators** 84

8.3.1 **Iterating an affine nonexpansive operator** 84

8.3.2 **Strong and linear convergence of affine nonexpansive operators** 90

8.3.3 **Some algorithmic consequences** 95

**Further results** 97

# Chapter 9: The Douglas–Rachford algorithm: convergence analysis
# TABLE OF CONTENTS

9.1 Overview ................................................. 101
9.2 Convergence of Douglas–Rachford algorithm ................. 102
9.3 A new proof of the Lions–Mercier–Svaiter theorem .......... 103
9.4 The Douglas–Rachford algorithm in the affine case ........ 106
9.5 Eckstein–Ferris–Pennanen–Robinson duality and algorithms 110
9.6 Brief history .............................................. 113

Chapter 10: On the order of the operators ......................... 115
10.1 Overview .............................................. 115
10.2 Connection between \( \text{Fix} T_{(A,B)} \) and \( \text{Fix} T_{(B,A)} \) ......... 115
10.3 Iterates of \( T_{(A,B)} \) vs. iterates of \( T_{(B,A)} \) .................. 117

Chapter 11: Generalized solutions for the sum of two maximally
monotone operators ........................................ 126
11.1 Overview .............................................. 126
11.2 A motivation from Linear Algebra .......................... 127
11.3 Perturbation calculus ..................................... 128
11.4 Perturbations of the Douglas–Rachford operator .......... 129
11.5 The \( w \)-perturbed problem ............................. 131
11.6 The normal problem ...................................... 133
11.7 Examples ................................................ 135

Chapter 12: On the range of the Douglas–Rachford operator: Theory140
12.1 Overview .............................................. 140
12.2 Near convexity and near equality .......................... 141
12.3 Attouch–Théra duality and the normal problem ............ 144
12.4 Main results .............................................. 145
12.5 Some infinite-dimensional observations ..................... 151

Chapter 13: On the range of the Douglas–Rachford operator: Ap-
lications ...................................................... 155
13.1 Overview .............................................. 155
13.2 On the minimal displacement vector \( v_{(A,B)} \) .............. 156
13.3 Applications to subdifferential operators ..................... 157
13.4 Application to firmly nonexpansive mappings ............... 159

Chapter 14: The Douglas–Rachford algorithm in the possibly in-
consistent case .............................................. 161
14.1 Overview .............................................. 161
14.2 The convex feasibility setting ............................. 162
<table>
<thead>
<tr>
<th>14.3 Convergence analysis in the inconsistent case</th>
<th>165</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 15: The Douglas–Rachford algorithm in the affine-affine feasibility case</td>
<td>173</td>
</tr>
<tr>
<td>15.1 Overview</td>
<td>173</td>
</tr>
<tr>
<td>15.2 The Douglas–Rachford operator for two affine subspaces</td>
<td>174</td>
</tr>
<tr>
<td>Chapter 16: The Douglas–Rachford algorithm in the affine-convex case</td>
<td>181</td>
</tr>
<tr>
<td>16.1 Overview</td>
<td>181</td>
</tr>
<tr>
<td>16.2 Convergence results</td>
<td>181</td>
</tr>
<tr>
<td>16.3 Spingarn’s method</td>
<td>184</td>
</tr>
<tr>
<td>Chapter 17: Conclusion</td>
<td>187</td>
</tr>
<tr>
<td>17.1 Main results</td>
<td>187</td>
</tr>
<tr>
<td>17.2 Side results</td>
<td>188</td>
</tr>
<tr>
<td>17.3 Future work</td>
<td>189</td>
</tr>
<tr>
<td>Bibliography</td>
<td>191</td>
</tr>
<tr>
<td>Index</td>
<td>205</td>
</tr>
</tbody>
</table>
List of Figures

Figure 8.1 The solid curve represents $\lim_{n \to \infty} (T - v)^n x$, the dashed dotted curve represents $\lim_{n \to \infty} (v + T)^n x$, and the dashed curve represents $\lim_{n \to \infty} T^n x + nv$, when $\alpha = 0.5$ and $\beta = 1$.

Figure 8.2 A GeoGebra [78] snapshot that illustrates Example 8.39. The first few terms of the sequence $(T^n x + nv)_{n \in \mathbb{N}}$ (blue points) are depicted.

Figure 10.1 A GeoGebra [78] snapshot. Two closed convex sets in $\mathbb{R}^2$, $U$ is a linear subspace (green line) and $V$ (the ball). Shown are also the first five terms of the sequences $(T^n_{(A,B)} R A x_0)_{n \in \mathbb{N}}$ (red points) and $(T^n_{(B,A)} x_0)_{n \in \mathbb{N}}$ (blue points) in each case.

Figure 10.2 A GeoGebra [78] snapshot. Two closed convex sets in $\mathbb{R}^2$, $U$ is the halfspace (cyan region) and $V$ (the ball). Shown are also the first five terms of the sequences $(T^n_{(A,B)} R A x_0)_{n \in \mathbb{N}}$ (red points) and $(T^n_{(B,A)} x_0)_{n \in \mathbb{N}}$ (blue points) in each case.

Figure 11.1 A diagram that illustrates Remark 11.12.

Figure 11.2 A GeoGebra [78] snapshot that illustrates Example 11.24.

Figure 11.3 A GeoGebra [78] snapshot that illustrates Example 11.25.

Figure 12.1 A GeoGebra [78] snapshot that illustrates Example 12.15.
LIST OF FIGURES

Figure 14.1 A GeoGebra [78] snapshot that illustrates Fact 14.6(ii). Two intersecting lines in $\mathbb{R}^3$, $U$ the blue line and $V$ the red line. The first few iterates of the sequences $(T^n x)_{n \in \mathbb{N}}$ (red points) and $(P_U T^n x)_{n \in \mathbb{N}}$ (blue points) are also depicted. ........................................... 164

Figure 14.2 A GeoGebra [78] snapshot that illustrates Theorem 14.11. Two nonintersecting polyhedral sets in $\mathbb{R}^2$, $U$ and $V$. The first few iterates of the sequences $(T^n x)_{n \in \mathbb{N}}$ (red points) and $(P_U T^n x)_{n \in \mathbb{N}}$ (blue points) are also depicted. Shown is the minimal displacement vector $v$ as well. .................. 167

Figure 15.1 Two nonintersecting affine subspaces $U$ (blue line) and $V$ (purple line) in $\mathbb{R}^3$. Shown are also the first few iterates of $(T^n x_0)_{n \in \mathbb{N}}$ (red points) and $(P_U T^n x_0)_{n \in \mathbb{N}}$ (blue points). .............. 178

Figure 16.1 A GeoGebra [78] snapshot that illustrates Example 16.2. ........................................... 183

Figure 16.2 A GeoGebra [78] snapshot that illustrates Corollary 16.6. Three nonintersecting closed convex sets, $C_1$ (the blue triangle), $C_2$ (the red polygon) and $C_3$ (the green circle), are shown along with their translations forming the generalized intersection. The first few terms of the sequence $(e(P_U T^n x))_{n \in \mathbb{N}}$ (yellow points) are also depicted. Here $e : U \rightarrow \mathbb{R}^2 : (x,x,x) \mapsto x$. ............... 186
# Glossary of Notation and Symbols

## Real line:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{N} )</td>
<td>The set of natural numbers ( {0, 1, 2, 3, \ldots } )</td>
</tr>
<tr>
<td>( \mathbb{Z} )</td>
<td>The set of integers ( {\ldots, -2, -1, 0, 1, 2, \ldots } )</td>
</tr>
<tr>
<td>( \mathbb{R} )</td>
<td>The set of real numbers</td>
</tr>
<tr>
<td>( \mathbb{R}_+ )</td>
<td>The set of positive real numbers ( [0, +\infty[ )</td>
</tr>
<tr>
<td>( \mathbb{R}_{++} )</td>
<td>The set of strictly positive real numbers ( ]0, +\infty[ )</td>
</tr>
<tr>
<td>( \mathbb{R}_- )</td>
<td>The set of negative real numbers ( ]-\infty, 0] )</td>
</tr>
<tr>
<td>( \mathbb{R}_{--} )</td>
<td>The set of strictly negative real numbers ( ]-\infty, 0[ )</td>
</tr>
</tbody>
</table>

## Sets:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{C} )</td>
<td>p. 5 Closure of a set C</td>
</tr>
<tr>
<td>intC</td>
<td>p. 5 Interior of a set C</td>
</tr>
<tr>
<td>riC</td>
<td>p. 5 Relative interior of a set C</td>
</tr>
<tr>
<td>sriC</td>
<td>p. 5 Strong relative interior of a set C</td>
</tr>
<tr>
<td>convC</td>
<td>p. 5 Convex hull of a set C</td>
</tr>
<tr>
<td>affC</td>
<td>p. 5 Affine hull of a set C</td>
</tr>
<tr>
<td>coneC</td>
<td>p. 5 Conic hull of a set C</td>
</tr>
<tr>
<td>( C^\perp )</td>
<td>p. 5 Orthogonal complement of a set C</td>
</tr>
<tr>
<td>parC</td>
<td>p. 6 Linear space parallel to a set C</td>
</tr>
<tr>
<td>( C^\ominus )</td>
<td>p. 5 Polar cone of a set C</td>
</tr>
<tr>
<td>( C^\oplus )</td>
<td>p. 5 Dual cone of a set C</td>
</tr>
<tr>
<td>recC</td>
<td>p. 5 Recession cone of a set C</td>
</tr>
<tr>
<td>( N_C )</td>
<td>p. 5 Normal cone operator of a set C</td>
</tr>
<tr>
<td>( P_C )</td>
<td>p. 6 Orthogonal projection onto a set C</td>
</tr>
<tr>
<td>ball(( x; r ))</td>
<td>p. 6 Closed ball with centre ( x ) and radius ( r )</td>
</tr>
</tbody>
</table>
Glossary of Notation and Symbols

Functions:

\( \Gamma(X) \) p. 8 The set of all proper convex lower semi-continuous functions defined from \( X \) to \( ]-\infty, +\infty[ \)

\( \operatorname{argmin} f \) p. 8 The set of minimizers of a function \( f \)

\( \operatorname{epi} f \) p. 7 Epigraph of a function \( f \)

\( i_C \) p. 8 Indicator function of a set \( C \)

\( \operatorname{dom} f \) p. 7 Domain of a function \( f \)

\( \operatorname{Prox}_f \) p. 8 Proximal mapping of a function \( f \)

\( \partial f \) p. 8 Subdifferential operator a function of \( f \)

\( f^* \) p. 9 Fenchel conjugate of a function of \( f \)

\( f' \) p. 9 Reversal of a function of \( f \)

Set-valued operators:

\( N_C \) p. 5 Normal cone operator of a set \( C \)

\( A : X \rightrightarrows X \) p. 11 \( A \) is a set valued operator on \( X \)

\( A^{-1} \) p. 11 Inverse of an operator \( A \)

\( A^\ominus \) p. 11 The reversal of an operator \( A \)

\( A^{-\ominus} \) p. 11 The reversal of the inverse of an operator \( A \)

\( \operatorname{dom} A \) p. 11 Domain of an operator \( A \)

\( \operatorname{ran} A \) p. 11 Range of an operator \( A \)

\( \operatorname{gra} A \) p. 11 Graph of an operator \( A \)

\( \zer A \) p. 11 Set of zeros of an operator \( A \)

\( J_{A} \) p. 13 Resolvent of an operator \( A \)

\( R_{A} \) p. 13 Reflected resolvent of an operator \( A \)

\( \ker L \) p. 17 Kernel of a linear operator \( L \)

\( A \Box B \) p. 34 Parallel sum of the operators \( A \) and \( B \)

Single-valued operators:

\( M^T \) Transpose of a matrix \( M \)

\( T : X \rightarrow X \) \( T \) is a single-valued operator on \( X \) defined everywhere

\( T(C) \) Image of a set \( C \) by an operator \( T \)

\( T^{-1}(C) \) Inverse image of a set \( C \) by an operator \( T \)

\( T|_S \) p. 55 The restriction of the map \( T \) to the set \( S \)

\( \operatorname{Fix} T \) p. 13 The set of fixed points of an operator \( T \)

\( \ker L \) p. 17 Kernel of a linear operator \( L \)

\( \| L \| \) p. 90 Norm of a linear operator \( L \)
Acknowledgements

First and foremost I thank my supervisors, Professor Heinz Bauschke and Professor Warren Hare. Their broad knowledge, remarkable supervision and incredible research abilities make this acknowledgment a challenge to write. I do not have enough good words to express my gratitude to their brilliant mentorship and unconditional guidance. Their support, encouragement and patience go beyond the work in this thesis. They never hesitate to put time and effort towards enhancing and polishing my academic skills, as a researcher, as a doctoral student, as a mentor, and as a teacher.

I deeply thank Professor Shawn Wang, the Graduate Program Coordinator at UBC (Okanagan campus), for his continuous support and for the remarkable research experience I gained from our collaboration. I also thank Professor Yong Gao for agreeing to serve on my thesis committee and for his constructive comments.

I thank Dr. Hung Phan and Dr. Minh Dao for sharing their knowledge and experience in computational techniques.

I recognize the support of Unit 5 members and my colleagues in the COCANA lab at UBC (Okanagan campus), who had made this journey possible and cheerful. In particular, I thank Sarah Moffat, Julie Nutini, Chayne Planiden and Liangjin Yao. Special thanks to Mohamed Yafia.

I thank UBC (Okanagan campus) for providing me with financial support over the years of my doctoral studies.

Finally, I thank my husband Ahmad, whose support always leaves me speechless; my daughter Hala, whose love and belief in me are the ultimate encouragement I have; my son Yaseen, who witnessed this journey since he was a day old with love and patience; and my parents, who have been and will always be my role models.
Dedication

To Ahmad, Hala, Yaseen and my parents.
Chapter 1

Introduction

The problem of finding a zero of the sum of two maximally monotone operators is of fundamental importance in optimization and variational analysis. Indeed, a classical problem in optimization is to find a minimizer of the sum of two proper convex lower semicontinuous functions. This problem can be modelled as

\[ \text{find } x \in X \text{ such that } 0 \in (A + B)x, \quad (1.1) \]

where \( A \) and \( B \) are maximally monotone operators on \( X \), namely the subdifferential operators of the functions under consideration. (For detailed discussions on problem (1.1) and the connection to optimization problems we refer the reader to [12], [42], [46], [50], [55], [58], [123], [125], [127], [128], [137], [140], [141], [142], and the references therein.)

Due to its general convergence results, the Douglas–Rachford algorithm has become a very popular splitting technique to solve the sum problem (1.1) (see, e.g., [64], [65] and [67] for applications of the method) provided that a solution exists.

As we shall explain below, the goal of this thesis is to explore the behaviour of the algorithm when the sum problem has no solution.

1.1 The goal of the thesis

The Douglas–Rachford algorithm was first introduced in [70] to numerically solve certain types of heat equations. Let \( T = T_{(A,B)} \) be the Douglas–Rachford operator associated with the ordered pair \((A,B)\), let \( J_A \) be the resolvent of \( A \) (see Definitions 5.2 and 3.13), and let \( \text{zer}(A + B) \) denote the set of zeros of \( A + B \). In their seminal work [96] (see also [95]), Lions and Mercier extended the algorithm to be able to find a zero of the sum of two, not necessarily linear and possibly multivalued, maximally monotone operators. Let \( X \) be a real Hilbert space and let \( x \in X \). Lions and Mercier proved that the governing sequence \((T^n x)_{n \in \mathbb{N}}\) converges weakly to a fixed point of \( T \), and that if \( A + B \) is maximally monotone, then the
1.2 Contributions in this thesis

weak cluster points of the shadow sequence \((J_{AT^n}x)_{n \in \mathbb{N}}\) are solutions of (1.1). In [132], weak convergence of the shadow sequence, regardless of the maximal monotonicity of \(A + B\), was established. The convergence proofs critically rely on the assumption that the sum problem is consistent, i.e., \(\text{zer}(A + B) \neq \emptyset\). Nonetheless, not every sum problem admits a solution: Suppose that \(A\) and \(B\) are normal cone operators of nonempty closed convex subsets of \(X\). It is clear that \(\text{zer}(A + B)\) is equal to the intersection of the two sets — however, this intersection may be empty, in which case the sum problem does not have any solution.

In this context, the behaviour of the algorithm in the inconsistent setting, i.e., when the set of zeros of the sum is empty, was not fully understood and our main motivation in this thesis is to better understand what the algorithm does in the absence of zeros of the sum. In fact, it is a fundamental problem in optimization to understand the problem in the case when there is no solution and the corresponding behaviour of associated optimization algorithms — perhaps the most famous occurrence is the birth of the least-squares method for solving inconsistent linear systems due to Gauss. The inconsistent case often results in considerable technical difficulties: For instance, Newton’s method for finding zeros applied to a quadratic function without zeros may lead to chaotic behaviour, see [104, Problem 7-a on page 72].

The goal of the work in this thesis is to understand the behaviour of the Douglas–Rachford algorithm in the possibly inconsistent case, to introduce new concepts that cope with the new situation, to develop a novel analysis that, unlike the analysis used in the consistent setting, does not rely on the existence of solutions of the sum problem and to explore the possible algorithmic consequences.

1.2 Contributions in this thesis

In this section, we summarize the main contributions of the research that constitutes this thesis, in the same order they appear in the sequel.

Chapter 2 and Chapter 3 present standard material and basic facts from convex analysis and monotone operator theory.

Our new results start in Chapter 4. We systematically study Attouch–Théra duality for the sum problem. We provide new results related to Passty’s parallel sum, to Eckstein and Svaiter’s extended solution set, and to Combettes’ fixed point description of the set of primal solutions. Furthermore, paramonotonicity is revealed to be a key property, as it allows for the recovery of all primal solutions given just one arbitrary dual solution.
1.2. Contributions in this thesis

In Chapter 5, we start by reviewing the bridge between the Douglas–Rachford splitting operator and the zeros of the sum. We provide a new description of the fixed point set of the Douglas–Rachford splitting operator (see Theorem 5.14). We also use the Douglas–Rachford operator to formulate new results on the relative geometry of the primal and dual solutions (in the sense of Attouch–Théra duality).

Chapter 6 presents new Fejér monotonicity principles. In particular, we point out that Lemma 6.3 is of critical importance in our main proofs. It also presents a powerful generalization of the classical Fejér monotonicity principle (see [12, Theorem 5.5]).

Chapter 7 and Chapter 8 are dedicated to providing a detailed study of the behaviour of (firmly) nonexpansive operators whose fixed point set is possibly empty. We provide a precise description of the connection of the fixed point sets of both the inner and outer shifts of nonexpansive operators. Under the additional assumption that the operator is affine, we are able to get stronger conclusions. As a consequence, we are able to provide new characterizations of strong and linear convergence of iterates of affine nonexpansive operators that were previously known only in the case of linear operators.

Chapter 9 focuses on the convergence analysis of the Douglas–Rachford algorithm in the consistent case. We provide a new proof of the weak convergence of the shadow sequence. We apply the new asymptotic results of Chapter 8 to the Douglas–Rachford algorithm when used to find a zero of the sum of two affine relations. This allows for stronger conclusions including strong and linear convergence and more precise information about the limits of the governing and shadow sequences.

Chapter 10 contains a collection of results that explore the connection between the two Douglas–Rachford operators associated with the sum problem. We show that the reflectors of the operators involved in the sum act as bijections between the sets of fixed points of the two Douglas–Rachford operators. Moreover, in some special cases, we are able to relate the two sequences obtained by iterating the two operators. This allows for new conclusions regarding linear rates of convergence, as well as new sufficient conditions for finite convergence.

In Chapter 11, we introduce the concept of the normal problem associated with finding a zero of the sum of two maximally monotone operators. This new concept is a milestone in our analysis. If the original problem admits solutions, then the normal problem returns this same set of solutions. The normal problem may yield solutions when the original problem does not admit any; furthermore, it has attractive variational and duality properties.
1.2. Contributions in this thesis

The normal problem proves to be useful in studying the behaviour of not only the Douglas–Rachford algorithm, but also other splitting techniques, namely the forward–backward algorithm, in the possibly inconsistent case (see [110]).

In Chapter 12, we focus on the ranges of the Douglas–Rachford operator and the corresponding displacement map. Under mild assumptions, we are able to provide explicit formulae for both ranges in terms of the domains and the ranges of the operators considered in the sum problem. Our formulae use near equality to describe the ranges, and we tighten our conclusions by providing a counterexample that shows that near equality is optimal. These results are important because the problem of finding a point in $\text{zer}(A + B)$ has a solution if and only if the Douglas–Rachford operator has a fixed point; equivalently if 0 is in the range of the displacement mapping. It also provides information on finding the minimal displacement vector that defines the normal problem introduced in Chapter 11.

Applications of the results in Chapter 12 are given in Chapter 13. We emphasize the case when the two operators in the sum problem are subdifferential operators — a typical setting in optimization.

Algorithmic consequences of our analysis start to appear in Chapter 14. We advance the understanding of the inconsistent case significantly by providing a complete proof of the full weak convergence in the convex feasibility setting when $A$ and $B$ are normal cone operators of nonempty closed convex subsets of $X$. This is an important instance of the Douglas–Rachford algorithm, as it has often been applied in the context of feasibility problems in both convex and nonconvex settings (see, e.g., [1], [2], [43], [45], [58], [76], [77], [87], [88], and [94]). We also provide some sufficient conditions that guarantee the convergence of the shadow sequence in more general settings.

In Chapter 15, we prove the strong convergence of the shadow sequence when the sets are affine subspaces that do not necessarily intersect. We precisely identify the limit to be the best approximation solution closest to the starting point. Furthermore, we prove the rate of convergence to be linear when the sum of the two subspaces is closed.

In Chapter 16, we explore what happens when only one set is an affine subspace. Whether or not the two shadow sequences converge depends on the order of the operators used in the definition of the Douglas–Rachford operator. Using the Pierra product space technique, we also apply our results to solve a least squares problem involving three or more sets.

Finally, in Chapter 17, we give a brief summary of our results and provide a list of possible avenues of future research.
Chapter 2

Convex analysis: Overview

The underlying space for most of the results in this thesis is a real Hilbert space. Nonetheless, few results hold more generally in Banach spaces. Unless otherwise specified, throughout this thesis

\[ X \] is a real Hilbert space,

with an inner product \( \langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{R} \). The induced norm is denoted by \( \| \cdot \| \). We start by recalling the basic definitions and facts we need from convex analysis.

2.1 Convex sets and projection operators

Definition 2.1 (convex set). A subset \( C \) of \( X \) is convex if \((\forall \alpha \in [0, 1]) \alpha C + (1 - \alpha)C \subseteq C\), or equivalently \((\forall (x, y) \in C \times C) |x, y| \subseteq C\).

Definition 2.2 (cone). A subset \( C \) of \( X \) is a cone if \( \mathbb{R}_+ C = C \).

Example 2.3 (normal cone operator). Let \( C \) be a nonempty convex subset of \( X \) and let \( x \in X \). The normal cone of \( C \) at \( x \) is the operator

\[
N_C : X \ni x \mapsto \begin{cases} \{ u \in X \mid \sup_{c \in C} \langle c - x, u \rangle \leq 0 \}, & \text{if } x \in C; \\ \emptyset, & \text{otherwise.} \end{cases}
\]

Let \( C \) be a subset of \( X \). The convex hull of \( C \), denoted by \( \text{conv } C \) is the smallest convex set containing \( C \). The affine hull of \( C \) denoted by \( \text{aff } C \) is the smallest affine set containing \( C \). The conic hull of \( C \) denoted by \( \text{cone } C \) is the smallest cone containing \( C \). The closure of \( C \) is denoted by \( \overline{C} \). The interior of \( C \) is denoted by \( \text{int } C \). The relative interior of \( C \) is defined by \( \text{ri } C = \{ x \in \text{aff } C \mid \exists \epsilon > 0 \text{ such that ball}(x; \epsilon) \cap \text{aff } C \subseteq C \} \). The strong relative interior of \( C \), denoted by \( \text{sri } C \), is the interior with respect to the closed affine hull of \( C \). The recession cone of \( C \) is \( \text{rec } C = \{ x \in X \mid x + C \subseteq C \} \). The polar cone of \( C \) is \( C^\circ = \{ u \in X \mid \sup_{c \in C} \langle c, u \rangle \leq 0 \} \). The dual cone of \( C \) is \( C^\circ = -C^\circ \). The orthogonal complement of \( C \) is \( C^\perp = \)
2.1. Convex sets and projection operators

\{ y \in X \mid (\forall x \in X) \langle x, y \rangle = 0 \}. When C is an affine subspace the linear space parallel to C is \( \text{par } C = C - C \). Let \( u \in X \) and let \( r > 0 \). We use ball\((x; r)\) to denote the closed ball in X centred at \( u \) with radius \( r \). Further notation is developed as necessary during the course of this thesis.

**Definition 2.4 (projection).** Let \( C \) be a nonempty closed convex subset of \( X \) and let \( x \in X \). The projection (this is also known as the closest point mapping) of \( x \) onto \( C \) is the unique point in \( C \) denoted by \( P_C x \) that satisfies

\[
\| x - P_C x \| = \inf_{c \in C} \| x - c \|. \tag{2.2}
\]

**Fact 2.5 (characterization of projection).** Let \( C \) be a nonempty closed convex subset of \( X \), let \( x \in X \) and let \( p \in X \). Then

\[ p = P_C x \iff p \in C \text{ and } (\forall y \in C) \langle y - p, x - p \rangle \leq 0. \tag{2.3} \]

*Proof.* See [12, Theorem 3.14].

The following useful translation formula could be readily verified.

**Fact 2.6.** Let \( S \) be a nonempty subset of \( X \), and let \( y \in X \). Then

\[ (\forall x \in X) P_{y+S} x = y + P_S (x - y). \tag{2.4} \]

*Proof.* See [12, Proposition 3.17].

**Fact 2.7.** Let \( C \) be a closed affine subspace of \( X \) and let \( x \in X \). Then, the following hold:

(i) \( P_C \) is affine.

If \( C \) is a closed linear subspace of \( X \), then we additionally have:

(ii) \( P_C \) is linear.

(iii) \( P_{C^\perp} = \text{Id} - P_C \).

*Proof.* (i): See [12, Corollary 3.20(ii)]. (ii)&(iii): See [12, Corollary 3.22(iii)&(v)] respectively.

**Lemma 2.8.** Suppose that \( C \) is an affine subspace and that \( w \in (\text{par } C)^\perp \). Then

\[ (\forall \alpha \in \mathbb{R})(\forall x \in X) \ P_C (x + \alpha w) = P_C x. \tag{2.5} \]
2.2 Convex functions

Proof. Let \( a \in C \). Then \( C = a + Par C \). Applying Fact 2.6 and Fact 2.7(ii) we have \((\forall a \in \mathbb{R})(\forall x \in X) P_C(x + aw) = P_{a + Par C}(x + aw) = a + P_{Par C}(x + aw - a) = a + P_{Par C}(x - a) + aP_{Par C}w = a + P_{Par C}(x - a) = P_{a + Par C}x = P_Cx. \)

The following facts regarding projection operators will be used in the sequel.

Fact 2.9 (Zarantonello 1971). Let \( C \) be a nonempty closed convex subset of \( X \). Then \( \text{ran} (Id - P_C) = (\text{rec} C)^\ominus \).

Proof. See [139, Theorem 3.1].

Fact 2.10. Let \( U \) and \( V \) be nonempty closed convex subsets of \( X \) such that \( U \perp V \). Then \( U + V \) is convex and closed, and \( P_{U + V} = P_U + P_V \).

Proof. See [21, Proposition 2.6].

2.2 Convex functions

Convex functions are major tools in optimization. The pleasant properties of convex functions are usually associated with the concept of lower semi-continuity. We start by recalling the following definitions.

Definition 2.11. Let \( f : X \to [-\infty, +\infty] \). Then \( f \) is lower semicontinuous if for every sequence \((x_n)_{n \in \mathbb{N}} \) in \( X \), we have

\[
x_n \to x \quad \text{implies that} \quad f(x) \leq \liminf_{n \to \infty} f(x_n). \tag{2.6}
\]

Definition 2.12. Let \( f : X \to [-\infty, +\infty] \). Then \( f \) is convex if \((\forall x \in X) (\forall y \in X) (\forall \alpha \in [0,1]) \)

\[
f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y). \tag{2.7}
\]

Geometrically, a convex function is characterized by having its epigraph

\[
\text{epi } f = \{(x, \alpha) \in X \times \mathbb{R} \mid f(x) \leq \alpha\} \tag{2.8}
\]

be a convex subset of \( X \times \mathbb{R} \).

Let \( f : X \to [-\infty, +\infty] \). Then \( f \) is proper if \(-\infty \not\in f(X) \) and \((\exists x \in X) \) such that \( f(x) < +\infty \). The domain of \( f \) is \( \text{dom } f = \{x \in X \mid f(x) < +\infty\} \).
2.3. Subdifferentiability and conjugation

**Definition 2.13.** Let $f : X \to ]-\infty, +\infty]$ be proper and let $x \in X$. Then $x$ is a minimizer of $f$ if $f(x) = \inf f(X)$. The set of minimizers of $f$

$$\{ x \in X \mid f(x) = \inf f(X) \}, \quad (2.9)$$

is denoted by argmin $f$.

Note that if argmin $f \neq \emptyset$, then (2.9) becomes $\{ x \in X \mid f(x) = \min f(X) \}$.

The set of proper convex lower semicontinuous functions defined on $X$ shall be denoted by $\Gamma(X)$.

**Example 2.14.** Let $C$ be a subset of $X$ and let $\iota_C$ be the indicator function associated with $C$ defined by

$$\iota_C : X \to ]-\infty, +\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \not\in C. \end{cases} \quad (2.10)$$

Then $\iota_C$ is proper if and only if $C$ is nonempty and $\iota_C$ is lower semicontinuous if and only if $C$ is closed. Moreover $\text{epi } \iota_C = C \times \mathbb{R}_+$ and hence $\iota_C$ is convex if and only if $C$ is convex.

**Proof.** See [12, Examples 1.25 and 8.3].

**Definition 2.15.** Let $f \in \Gamma(X)$ and let $x \in X$. Then Prox$_f$ is the unique point in $X$ that satisfies

$$\min_{y \in X} \left( f(y) + \frac{1}{2} \|x - y\|^2 \right) = f(\text{Prox}_f x) + \frac{1}{2} \|x - \text{Prox}_f x\|^2. \quad (2.11)$$

The operator Prox$_f : X \to X$ is the proximity operator or the proximal mapping of $f$.

### 2.3 Subdifferentiability and conjugation

**Definition 2.16 (subdifferential operator).** Let $f : X \to ]-\infty, +\infty]$ be proper. The subdifferential of $f$ is the set-valued operator

$$\partial f : X \rightrightarrows X : x \mapsto \{ u \in X \mid (\forall y \in X) \langle y - x, u \rangle + f(x) \leq f(y) \}. \quad (2.12)$$

Let $x \in X$. Then $f$ is subdifferentiable at $x$ if $\partial f(x) \neq \emptyset$. Moreover, if a proper convex function $f$ is differentiable at $x$, then $\partial f(x) = \{ \nabla f(x) \}$, by [137, Theorem 2.4.4(i)].
Example 2.17. Let $C$ be a nonempty convex subset of $X$. Then $\partial C = N_C$.

Proof. See [12, Example 16.12]. ■

Fact 2.18 (Fermat’s rule). Let $f : X \to ]-\infty, +\infty]$ be proper. Then

$$\text{argmin } f = \text{zer } \partial f = \{ x \in X \mid 0 \in \partial f(x) \}. \quad (2.13)$$

Proof. See [12, Theorem 16.2]. ■

Definition 2.19 (conjugate function). Let $f : X \to ]-\infty, +\infty]$. The conjugate (this is also known as Fenchel conjugate) of $f$ is

$$f^* : X \to ]-\infty, +\infty] : u \mapsto \sup_{x \in X} (\langle x, u \rangle - f(x)) \quad (2.14)$$

and the biconjugate of $f$ is $f^{**} = (f^*)^*$.

Fact 2.20. Let $f \in \Gamma(X)$, let $x \in X$ and let $u \in X$. Then the following are equivalent:

(i) $u \in \partial f(x)$.

(ii) $f(x) + f^*(u) = \langle x, u \rangle$.

(iii) $x \in \partial f^*(u)$.

Proof. See [123, Theorem 23.5] or [12, Theorem 16.23]. ■

Fact 2.21. Let $f \in \Gamma(X)$. Then the following hold:

(i) $\text{dom } \partial f$ is a dense subset of $\text{dom } f$.

(ii) $\text{Prox}_f = (\text{Id} + \partial f)^{-1}$.

Proof. See [12, Corollary 16.29 & Proposition 16.34]. ■

Let $f : X \to ]-\infty, +\infty]$. Then $f^\prime : X \to ]-\infty, +\infty] : x \mapsto f(-x)$.

Fact 2.22 (Fenchel duality). Let $f$ and $g$ be functions in $\Gamma(X)$ such that $0 \in \text{sri}(\text{dom } f - \text{dom } g)$. Then

$$\inf (f + g)(X) = -\min (f^* + g^{*\prime})(X). \quad (2.15)$$

Proof. See [12, Proposition 15.13]. ■
Chapter 3

(Firmly) nonexpansive mappings and monotone operators

Most of the material in this chapter is standard. Facts without explicit references may be found in [12], [81], or [82]. However, we point out that Example 3.36 and the results in Section 3.5 are new and appear in [38] and [39] respectively.

3.1 Nonexpansive and firmly nonexpansive operators

Definition 3.1. Let $T: X \to X$. Then

(i) $T$ is nonexpansive if $(\forall x \in X)(\forall y \in X)$

$$\|Tx - Ty\| \leq \|x - y\|. \tag{3.1}$$

(ii) $T$ is firmly nonexpansive if $(\forall x \in X)(\forall y \in X)$

$$\|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2. \tag{3.2}$$

Clearly,

$$T \text{ is firmly nonexpansive } \Rightarrow T \text{ is nonexpansive.} \tag{3.3}$$

Fact 3.2. Let $T: X \to X$. Then the following are equivalent:

(i) $T$ is firmly nonexpansive.

(ii) $\text{Id} - T$ is firmly nonexpansive.

(iii) $2T - \text{Id}$ is nonexpansive.

(iv) $(\forall x \in X)(\forall y \in X) \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle.$
(v) \( (\forall x \in X) \ (\forall y \in X) \ <Tx - Ty, (\text{Id} - T)x - (\text{Id} - T)y> \geq 0 \).

Proof. See [81, Theorem 12.1] or [82, Section 11 in Chapter 1]. ■

3.2 Monotone operators: Basic definitions and facts

Let \( A : X \rightrightarrows X \) be an arbitrary possibly set-valued operator, i.e., \( (\forall x \in X) \ Ax \subseteq X \). The graph of \( A \), denoted \( \text{gra} \ A \), is defined as

\[
\text{gra} \ A = \{(x,u) \in X \times X \mid u \in Ax\}. \tag{3.4}
\]

The domain of \( A \), denoted \( \text{dom} \ A \), is defined as

\[
\text{dom} \ A = \{x \in X \mid Ax \neq \emptyset\}. \tag{3.5}
\]

The range of \( A \), denoted \( \text{ran} \ A \), is defined as

\[
\text{ran} \ A = \{u \in X \mid \exists x \in X \text{ such that } u \in Ax\}. \tag{3.6}
\]

The set of zeros of \( A \) is written as

\[
\text{zer} \ A = A^{-1}(0) = \{x \in X \mid 0 \in Ax\}. \tag{3.7}
\]

The inverse operator, denoted \( A^{-1} : X \rightrightarrows X \) is defined in terms of its graph by \( \text{gra} A^{-1} = \{(u,x) \mid (x,u) \in \text{gra} A\} \). We say that \( A \) is at most single-valued if \( (\forall x \in X) \ Ax \) is at most a singleton. We say that \( A : X \rightrightarrows X \) is a linear relation (respectively affine relation) if \( \text{gra} A \) is a linear subspace (respectively affine subspace) of \( X \times X \). For further information of affine relations we refer the reader to [25]. It will be quite convenient to introduce the notation

\[
A^\Box = (\text{-Id} \circ A \circ (\text{-Id})\). \tag{3.8}
\]

An easy calculation shows that \( (A^{-1})^\Box = (A^\Box)^{-1} \), which motivates the notation

\[
A^{-\Box} = (A^{-1})^\Box = (A^\Box)^{-1}. \tag{3.9}
\]

(This is similar to the linear-algebraic notation \( A^{-T} \) for invertible square matrices.)

Definition 3.3. Let \( A : X \rightrightarrows X \). Then
3.2. Monotone operators: Basic definitions and facts

(i) $A$ is monotone if
\[
\forall (x,u) \in \text{gr } A (\forall (y,v) \in \text{gr } A) \quad \langle x - y, u - v \rangle \geq 0.
\] (3.10)

(ii) $A$ is strictly monotone if
\[
(\forall (x,u) \in \text{gr } A)(\forall (y,v) \in \text{gr } A) \quad x \neq y \Rightarrow \langle x - y, u - v \rangle > 0.
\] (3.11)

(iii) $A$ is uniformly monotone with modulus $\phi : \mathbb{R}_+ \to [0, +\infty]$ if $\phi$ is increasing, vanishes only at 0, and
\[
(\forall (x,u) \in \text{gr } A)(\forall (y,v) \in \text{gr } A) \quad \langle x - y, u - v \rangle \geq \phi(\|x - y\|).
\] (3.12)

(iv) $A$ is maximally monotone if it is monotone and it is impossible to properly enlarge the graph of $A$ while keeping monotonicity.

The following fact is easy to verify.

**Fact 3.4.** Let $T : X \to X$ be nonexpansive and let $\alpha \in [-1, 1]$. Then $\text{Id} + \alpha T$ is maximally monotone.

*Proof.* See [12, Example 20.26]. ■

The following deep facts are known in the literature of monotone operators. The proofs are beyond the scope of the thesis.

**Fact 3.5.** Let $f \in \Gamma(X)$. Then the following hold:

(i) $\partial f$ is maximally monotone.

(ii) $(\partial f)^{-1} = \partial f^*$. 


**Fact 3.6.** Let $A : X \rightrightarrows X$ be maximally monotone such that $\text{gra } A$ is convex. Then $\text{gra } A$ is actually affine.

*Proof.* See [23, Theorem 4.2]. ■

**Fact 3.7.** Let $A : \mathbb{R} \rightrightarrows \mathbb{R}$. Then $A$ is maximally monotone if and only if $(\exists f \in \Gamma(X))$ such that $A = \partial f$. 

12
3.3. Resolvents of monotone operators

Proof. See [125, Exercise 12.26] or [12, Corollary 22.19].

Fact 3.8. Let \( A : X \rightrightarrows X \) be maximally monotone with bounded domain. Then \( A \) is surjective.

Proof. See [12, Corollary 21.21].

The following fact can be readily verified.

Fact 3.9. Let \( A : X \rightrightarrows X \) be maximally monotone and let \( x \in X \). Then \( Ax \) is closed and convex.

Proof. See [12, Proposition 20.31].

Example 3.10. Suppose that \( C \) is nonempty closed convex subset of \( X \). Then, in view of Example 2.17, Example 2.14 and Fact 3.5(i), \( N_C \) is maximally monotone.

3.3 Resolvents of monotone operators

Definition 3.11 (resolvent and reflected resolvent). Let \( A : X \rightrightarrows X \). The resolvent of \( A \) is the operator

\[
J_A = (\text{Id} + A)^{-1}, \tag{3.13}
\]

and the reflected resolvent is

\[
R_A = 2J_A - \text{Id}. \tag{3.14}
\]

Fact 3.12 (monotonicity and firm nonexpansiveness). Let \( D \) be a nonempty subset of \( X \), let \( T : D \to X \), let \( A : X \rightrightarrows X \), and suppose that \( A = T^{-1} - \text{Id} \). Then the following hold:

(i) \( T = J_A \).

(ii) \( A \) is monotone if and only if \( T \) is firmly nonexpansive.

(iii) \( A \) is maximally monotone if and only if \( T \) is firmly nonexpansive and \( D = X \).

Proof. See [72, Theorem 2].

Let \( T : X \to X \). The set of fixed points of \( T \) is \( \text{Fix} T = \{ x \in X \mid Tx = x \} \).
### 3.3. Resolvents of monotone operators

**Fact 3.13.** Suppose that $A$ is maximally monotone. Then the following hold:

(i) $J_A$ is firmly nonexpansive and $\text{dom} \ J_A = X$.

(ii) $R_A$ is nonexpansive.

(iii) We have the following inverse resolvent identity

$$J_A + J_{A^{-1}} = \text{Id}. \quad (3.15)$$

(iv) $\text{zer} \ A = \text{Fix} \ J_A = \{x \in X \mid J_Ax = x\}$.

**Proof.** (i): This follows from [105, Corollary on page 344] and [124, Proposition 1(c)]. Alternatively use Fact 3.12. (ii): Combine (i) and Fact 3.2. (iii): This follows from [125, Lemma 12.14]. (iv): Clear. ■

**Example 3.14.** The following are examples of resolvents of monotone operators.

(i) Let $f \in \Gamma(X)$ and suppose that $A = \partial f$. Then $J_A = \text{Prox}_f$, the proximal point mapping of the function $f$ (see Definition 2.15).

(ii) Let $C$ be a nonempty closed convex subset of $X$ and suppose that $A = N_C$. Then $J_A = \text{Prox}_C = P_C$, the projection onto the set $C$.

**Proof.** See [12, Examples 23.3 & 23.4]. ■

We obtain the following useful lemma.

**Lemma 3.15.** Let $C$ be a nonempty closed convex subset of $X$ and let $c \in C$ satisfy that $\|c\| = \|P_C0\|$. Then $c = P_C0$.

**Proof.** In view of Example 3.14(ii) and Fact 3.13(i) we learn that $P_C$ is firmly nonexpansive. Therefore we have

$$\|c - P_C0\|^2 = \|P_Cc - P_C0\|^2 \quad (3.16a)$$

$$\leq \|c - 0\|^2 - \|\text{Id} - P_C\|c - (\text{Id} - P_C)0\|^2 \quad (3.16b)$$

$$= \|c\|^2 - \|c - P_Cc + P_C0\|^2 = \|c\|^2 - \|P_C0\|^2 = 0. \quad (3.16c)$$

■
3.3. Resolvents of monotone operators

**Proposition 3.16.** Suppose that $A: X \to X$ is continuous, linear, and single-valued such that $A$ and $-A$ are monotone, and $A^2 = -\alpha \text{Id}$, where $\alpha \in \mathbb{R}^+$. Then

$$J_A = \frac{1}{1 + \alpha} (\text{Id} - A) \quad \text{and} \quad R_A = \frac{1 - \alpha}{1 + \alpha} \text{Id} - \frac{2}{1 + \alpha} A. \quad (3.17)$$

**Proof.** We have

$$J_A J_{-A} = (\text{Id} + A)^{-1}(\text{Id} - A)^{-1} = \left( (\text{Id} - A)(\text{Id} + A) \right)^{-1} \quad (3.18a)$$

$$= (\text{Id} - A^2)^{-1} = (\text{Id} + \alpha \text{Id})^{-1} \quad (3.18b)$$

$$= \frac{1}{1 + \alpha} \text{Id}. \quad (3.18c)$$

It follows that $J_A = (1 + \alpha)^{-1}(J_{-A})^{-1} = (1 + \alpha)^{-1}(\text{Id} - A)$ and hence that

$$R_A = 2J_A - \text{Id} = \frac{2}{1 + \alpha}(\text{Id} - A) - \text{Id} = \frac{1 - \alpha}{1 + \alpha} \text{Id} - \frac{2}{1 + \alpha} A, \quad (3.19)$$

as claimed. ■

**Example 3.17.** Suppose that $X = \mathbb{R}^2$ and that $A: \mathbb{R}^2 \to \mathbb{R}^2$: $(x, y) \mapsto (-y, x)$ is the rotator by $\pi/2$. Then $A^2 = -\text{Id}$; consequently, by Proposition 3.16, $J_A = (1/2)(\text{Id} - A)$ and $R_A = -A$.

We recall the following very useful Minty parametrization.

**Fact 3.18 (Minty parametrization).** Let $A: X \rightrightarrows X$ be maximally monotone. Then $M: X \to \text{gra } A: x \mapsto (J_A x, J_{-A} x)$ is a continuous bijection, with continuous inverse $M^{-1}: \text{gra } A \to X: (x, u) \mapsto x + u$; consequently,

$$\text{gra } A = M(X) = \{ (J_A x, x - J_A x) \mid x \in X \}. \quad (3.20)$$

**Proof.** See [105]. ■

**Lemma 3.19.** Let $A: X \rightrightarrows X$. Then $\text{dom } A = \text{ran } J_A$.

**Proof.** Indeed, $\text{dom } A = X \cap \text{dom } A = \text{dom } \text{Id} \cap \text{dom } A = \text{dom } (\text{Id} + A) = \text{ran } (\text{Id} + A)^{-1} = \text{ran } J_A$. ■

Applying Lemma 3.19 to $A^{-1}$ and using (3.15), we obtain

$$\text{ran } A = \text{dom } A^{-1} = \text{ran } J_{A^{-1}} = \text{ran } (\text{Id} - J_A). \quad (3.21)$$

The following fact could be directly verified.
3.4 Sums of monotone operators: maximality and range of the sum

**Fact 3.20 (resolvent calculus).** Let \( A : X \rightrightarrows X \) be maximally monotone and let \( w \in X \). Then the following hold:

(i) Suppose that \( B = w + A \). Then \( J_B = J_A(\cdot - w) \).

(ii) Suppose that \( B = A(\cdot + w) \). Then \( J_B = -w + J_A(\cdot + w) \)

**Proof.** See [12, Proposition 23.15(ii)&(iii)]. ■

The following lemma shall be useful in our work.

**Lemma 3.21.** Let \( C \) be a nonempty closed convex subset of \( X \) and let \( w \in X \). Then the following hold:

(i) \( \text{N}_C(\cdot - w) = \text{N}_{w+C} \).

(ii) \( J_{-w+C} = J_{N_C}(\cdot + w) = P_C(\cdot + w) \).

(iii) \( J_{N_C(\cdot - w)} = w + J_{N_C}(\cdot - w) = w + P_C(\cdot - w) \).

**Proof.** (i): One can easily verify that \( (\forall w \in X) \) we have \( \iota_C(\cdot - w) = \iota_{w+C} \). Therefore \( \text{N}_C(\cdot - w) = \partial \iota_C(\cdot - w) = \partial \iota_{w+C} = \text{N}_{w+C} \). (ii)&(iii): Apply Fact 3.20(i)&(ii) with \( w \) replaced by \(-w\) and use Example 3.14(ii). ■

3.4 Sums of monotone operators: maximality and range of the sum

In this section we review further notions of monotonicity that allow for stronger conclusions, as we shall demonstrate.

**Definition 3.22 (paramonotone operator).** An operator \( C : X \rightrightarrows X \) is **paramonotone**, if it is monotone and we have

\[
\begin{aligned}
(x, u) \in \text{gra} \, C \\
(y, v) \in \text{gra} \, C \\
\langle x - y, u - v \rangle = 0
\end{aligned}
\] \( \Rightarrow \) \( \{ (x, v), (y, u) \} \subseteq \text{gra} \, C. \) \tag{3.22}

**Remark 3.23.** Paramonotonicity has proven to be a very useful property for finding solution of variational inequalities by iteration; see, e.g., [89], [57], [54], [112], and [84].
3.4. Sums of monotone operators: maximality and range of the sum

Let \( L : X \to X \) be linear. Recall that the kernel of \( L \) is \( \ker L = \{ x \in X \mid Lx = 0 \} \). We now provide examples of paramonotone operators below.

Example 3.24. Each of the following is paramonotone.

(i) \( \partial f \), where \( f \in \Gamma(X) \) (see [89, Proposition 2.2] or [12, 22.3(i)]).

(ii) \( C : X \mapsto X \), where \( C \) is strictly monotone.

(iii) \( \mathbb{R}^n \to \mathbb{R}^n : x \mapsto Lx + b \), where \( L \in \mathbb{R}^{n \times n} \), \( b \in \mathbb{R}^n \), \( L_+ = \frac{1}{2}L + \frac{1}{2}L^T \), \( \ker L_+ \subseteq \ker L \), and \( L_+ \) is positive semidefinite [89, Proposition 3.1].

Example 3.25. Suppose that \( X = \mathbb{R}^2 \) and that \( A : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (y, -x) \). Then one can easily verify that \( A \) and \( -A \) are maximally monotone but not paramonotone by [89, Section 3] (or [32, Theorem 4.9]).

For further examples, and detailed discussion on paramonotone operators see [89]. It is straightforward to check that for \( C : X \mapsto X \), we have

\[
C \text{ is paramonotone } \iff \text{ } C^{-1} \text{ is paramonotone} \\
\iff \text{ } C^{\circ} \text{ is paramonotone} \\
\iff \text{ } C^{-\circ} \text{ is paramonotone}. 
\]

Definition 3.26 (3* monotone operator). (See [51, page 166].) Let \( A : X \mapsto X \) be monotone. Then \( A \) is 3* monotone (this is also known as rectangular) if

\[
(\forall x \in \text{dom } A)(\forall v \in \text{ran } A) \inf_{(z,w) \in \text{gra } A} \langle x - z, v - w \rangle > -\infty. 
\]

When \( C \) is a continuous linear monotone operator, then \( C \) is paramonotone if and only if \( C \) is 3* monotone; see [22, Section 4].

Fact 3.27. Let \( f \in \Gamma(X) \). Then \( \partial f \) is 3* monotone.

Proof. See [51, page 167].

Corollary 3.28. Let \( A : \mathbb{R} \mapsto \mathbb{R} \) be monotone. Then \( A \) is 3* monotone and paramonotone.

Proof. In view of Fact 3.7, there exists \( f \in \Gamma(\mathbb{R}) \) such that \( \text{gra } A \subseteq \text{gra } \partial f \). The result now follows from Fact 3.27 and Example 3.24(i).

It is clear that sum of two maximally monotone operators is monotone. This sum is maximally monotone if an appropriate constraint qualification is imposed, as we see now.
3.4. Sums of monotone operators: maximality and range of the sum

**Fact 3.29.** Let $A : X \rightrightarrows X$ and $B : X \rightrightarrows X$ be maximally monotone such that one of the following conditions holds:

(i) $\text{dom } B = X$.

(ii) $\text{dom } A \cap \text{int } \text{dom } B \neq \emptyset$.

(iii) $0 \in \text{int}(\text{dom } A - \text{dom } B)$.

(iv) $\text{dom } A$ and $\text{dom } B$ are convex and $0 \in \text{sri}(\text{dom } A - \text{dom } B)$.

Then $A + B$ is maximally monotone.

**Proof.** See [12, Corollary 24.4].

Before we proceed further, we need to recall the following definitions and facts.

**Definition 3.30 (near convexity).** Let $X$ be finite-dimensional and let $D$ be a subset of $X$. Then $D$ is *nearly convex* if there exists a convex subset $C$ of $X$ such that $C \subseteq D \subseteq C$.

**Fact 3.31.** Let $X$ be finite-dimensional and let $A : X \rightrightarrows X$ be maximally monotone. Then $\text{dom } A$ and $\text{ran } A$ are nearly convex.

**Proof.** See [125, Theorem 12.41].

**Definition 3.32 (near equality).** (See [29, Definition 2.3].) Let $X$ be finite-dimensional and let $C$ and $D$ be subsets of $X$. We say that $C$ and $D$ are *nearly equal* if

$$C \simeq D : \iff \overline{C} = \overline{D} \quad \text{and} \quad \text{ri } C = \text{ri } D.$$  

(3.25)

**Fact 3.33.** Let $X$ be finite-dimensional and let $D$ be a nonempty nearly convex subset of $X$, say $C \subseteq D \subseteq \overline{C}$, where $C$ is a convex subset of $X$. Then

$$D \simeq \overline{D} \simeq \text{ri } D \simeq \text{conv } D \simeq \text{ri conv } D \simeq C.$$  

(3.26)

In particular, $\overline{D}$ and $\text{ri } D$ are convex and nonempty.

**Proof.** See [29, Lemma 2.7].

**Fact 3.34.** Suppose that $X$ is finite-dimensional. Let $A : X \rightrightarrows X$ and $B : X \rightrightarrows X$ be maximally monotone such that $\text{ri } \text{dom } A \cap \text{ri } \text{dom } B \neq \emptyset$. Then $A + B$ is maximally monotone.
3.4. Sums of monotone operators: maximality and range of the sum

Proof. See [125, Corollary 12.44].

Fact 3.35 (Brezis–Haraux). Let $A : X \rightrightarrows X$ and $B : X \rightrightarrows X$ be monotone operators such that $A + B$ is maximally monotone and one of the following conditions holds:

(i) $A$ and $B$ are $3^*$-monotone.

(ii) $\text{dom } A \subseteq \text{dom } B$ and $B$ is $3^*$-monotone.

Then

$$\text{ran}(A + B) = \overline{\text{ran } A + \text{ran } B} \quad \text{and} \quad \text{int ran}(A + B) = \text{int}(\text{ran } A + \text{ran } B).$$

(3.27)

If $X$ is finite-dimensional, then

$$\text{ran}(A + B)$$

is nearly convex and

$$\text{ran}(A + B) \simeq \text{ran } A + \text{ran } B.$$  (3.28)

Proof. See [51, Theorems 3 and 4] for the proof of (3.27) and [29, Theorem 3.13] for the proof of (3.28).

Example 3.36 (and later Proposition 12.17 below) illustrate that the results of Fact 3.35 are sharp in the sense that actual equality fails.

Example 3.36. Suppose that $X = \mathbb{R}^2$ and let $f : \mathbb{R}^2 \to ]-\infty, +\infty[: (\xi_1, \xi_2) \mapsto \max \{g(\xi_1), |\xi_2|\}$, where $g(\xi_1) = 1 - \sqrt{\xi_1}$ if $\xi_1 \geq 0$, $g(\xi_1) = +\infty$ if $\xi_1 < 0$.

Set $A = \partial f^*$. Then $A$ is $3^*$-monotone and $2A = A + A$ is maximally monotone, yet

$$2 \text{ran } A = \text{ran } A + \text{ran } A \not\subseteq \text{ran } A + \text{ran } A. \quad (3.29)$$

Proof. First notice that by Fact 3.27 $A$ is $3^*$-monotone. Moreover, since $A$ is maximally monotone, it follows from Fact 3.33 that $\text{ri } \text{dom } A$ is nonempty and convex. Since $\text{ri } \text{dom } A \cap \text{ri } \text{dom } A = \text{ri } \text{dom } A \neq \emptyset$, Fact 3.34 implies that $A + A = 2A$ is maximally monotone. Using [12, Proposition 16.24] and [123, example on page 218] we know that

$$\text{ran } \partial f^* = \text{ran } (\partial f)^{-1} = \text{dom } \partial f$$

$$= \{ (\xi_1, \xi_2) \mid \xi_1 > 0 \} \cup \{ (0, \xi_2) \mid |\xi_2| \geq 1 \}$$

(3.30a)

and

$$\text{ran } (A + A) = \{ (\xi_1, \xi_2) \mid \xi_1 > 0 \} \cup \{ (0, \xi_2) \mid |\xi_2| \geq 2 \}$$

$$\neq \{ (\xi_1, \xi_2) \mid \xi_1 \geq 0 \} = \text{ran } A + \text{ran } A.$$  (3.31a)
3.5 Examples of linear monotone operators

In this section we present numerous examples of linear monotone operators that are partly motivated by applications in partial differential equations; see, e.g., [80] and [133]. We also pay attention to the tridiagonal Toeplitz matrices and Kronecker products. These examples not only are helpful in understanding the connection between the classical Douglas–Rachford setting (see Definition 5.1 below) and modern monotone operator theory but also they allow explicit forms which should turn out to be useful in applications.

Let \( M \in \mathbb{R}^{n \times n} \). Then we have the following equivalences:

\[
\begin{align*}
M \text{ is monotone} & \iff \frac{M + M^T}{2} \text{ is positive semidefinite} \quad (3.32a) \\
& \iff \text{the eigenvalues of } \frac{M + M^T}{2} \text{ lie in } \mathbb{R}_+. \quad (3.32b)
\end{align*}
\]

**Lemma 3.37.** Let

\[
M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbb{R}^{2 \times 2}.
\]

Then \( M \) is monotone if and only if \( \alpha \geq 0 \), \( \delta \geq 0 \) and \( 4\alpha \delta \geq (\beta + \gamma)^2 \).

**Proof.** Indeed, the principal minors of \( M + M^T \) are \( 2\alpha \), \( 2\delta \) and \( 4\alpha \delta - (\beta + \gamma)^2 \); by [102, (7.6.12) on page 566]. \( \blacksquare \)

Note that if \( M = M^T \), then \( M \) is monotone if and only if the eigenvalues of \( M \) lie in \( \mathbb{R}_+ \). If \( M \neq M^T \), then some information about the location of the (possibly complex) eigenvalues of \( M \) is available:

**Lemma 3.38.** Let \( M \in \mathbb{R}^{n \times n} \) be monotone, and let \( \{\lambda_k\}_{k=1}^n \) denote the set of eigenvalues of \( M \). Then\(^1\) \( \text{Re}(\lambda_k) \geq 0 \) for every \( k \in \{1, \ldots, n\} \).

**Proof.** Write \( \lambda = \alpha + i\beta \), where \( \alpha \) and \( \beta \) belong to \( \mathbb{R} \) and \( i = \sqrt{-1} \) and assume that \( \lambda \) is an eigenvalue of \( M \) with (nonzero) eigenvector \( w = u + iv \), where \( u \) and \( v \) are in \( \mathbb{R}^n \). Then \( (M - \lambda \text{Id})w = 0 \iff ((M - \alpha \text{Id}) - i\beta \text{Id})(u + iv) = 0 \iff (M - \alpha \text{Id})u + \beta v = 0 \) and \( (M - \alpha \text{Id})v - \beta u = 0 \). Hence

\[
\begin{align*}
\langle u, (M - \alpha \text{Id})u \rangle + \beta \langle u, v \rangle &= 0, \quad (3.34a) \\
\langle v, (M - \alpha \text{Id})v \rangle - \beta \langle v, u \rangle &= 0. \quad (3.34b)
\end{align*}
\]

\(^1\)We use \( \text{Re}(z) \) to refer to the real part of the complex number \( z \).
3.5. Examples of linear monotone operators

Adding (3.34) yields \( \langle u, (M - \alpha \text{Id})u \rangle + \langle v, (M - \alpha \text{Id})v \rangle = 0 \); equivalently, 
\[
\langle u, Mu \rangle + \langle v, Mv \rangle - \alpha \|u\|^2 - \alpha \|v\|^2 = 0.
\]
Solving for \( \alpha \) yields
\[
\text{Re}(\lambda) = \alpha = \frac{\langle u, Mu \rangle + \langle v, Mv \rangle}{\|u\|^2 + \|v\|^2} \geq 0, \tag{3.35}
\]
as claimed.

The converse of Lemma 3.38 is not true in general, as we demonstrate in the following example.

Example 3.39. Let \( \xi \in \mathbb{R} \setminus [-2, 2] \), and set
\[
M = \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix}. \tag{3.36}
\]
Then \( M \) has 1 as its only eigenvalue (with multiplicity 2), \( M \) is not monotone by Lemma 3.37, and \( M \) is not symmetric.

Proposition 3.40. Consider the tridiagonal Toeplitz matrix
\[
M = \begin{pmatrix} \beta & \gamma & 0 \\ \alpha & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots \\ \lambda_k & \gamma & \alpha \end{pmatrix} \in \mathbb{R}^{n \times n}. \tag{3.37}
\]
Then \( M \) is monotone if and only if \( \beta \geq |\alpha + \gamma| \cos(\pi/(n + 1)) \).

Proof. Note that
\[
\frac{1}{2}(M + M^T) = \begin{pmatrix} 
\beta & \frac{1}{2}(\alpha + \gamma) & 0 \\
\frac{1}{2}(\alpha + \gamma) & \ddots & \ddots \\
0 & \ddots & \ddots & \frac{1}{2}(\alpha + \gamma) \\
\end{pmatrix}. \tag{3.38}
\]
By (3.32a), \( M \) is monotone \( \iff \frac{1}{2}(M + M^T) \) is positive semidefinite. If \( \alpha + \gamma = 0 \), then \( \frac{1}{2}(M + M^T) = \beta \text{Id} \) and therefore \( \frac{1}{2}(M + M^T) \) is positive semidefinite \( \iff \beta \geq 0 = |\alpha + \gamma| \). Now suppose that \( \alpha + \gamma \neq 0 \). It follows from [102, Example 7.2.5] that the eigenvalues of \( \frac{1}{2}(M + M^T) \) are
\[
\lambda_k = \beta + (\alpha + \gamma) \cos \left( \frac{k\pi}{n+1} \right), \tag{3.39}
\]
for \( k = 0, 1, \ldots, n \).
3.5. Examples of linear monotone operators

where \( k \in \{1, \ldots, n\} \). Consequently,

\[
\{\lambda_k\}_{k=1}^n \subseteq \mathbb{R}_+ \iff \beta \geq \left| (\alpha + \gamma) \cos \left( \frac{\pi}{n+1} \right) \right|. \tag{3.40}
\]

The characterization of monotonicity of \( M \) now follows from (3.32b).

**Proposition 3.41.** Let

\[
M = \begin{pmatrix} \beta & \gamma & 0 \\ \alpha & \ddots & \ddots \\ \vdots & \ddots & \ddots \\ 0 & \ldots & \alpha & \beta \end{pmatrix} \in \mathbb{R}^{n \times n}. \tag{3.41}
\]

Then exactly one of the following holds:

(i) \( \alpha \gamma = 0 \) and \( \det(M) = \beta^n \). Consequently, \( M \) is invertible \( \iff \beta \neq 0 \), in which case

\[
[M^{-1}]_{i,j} = (-\alpha)^\max\{i-j,0\} (-\gamma)^\max\{j-i,0\} \beta^{\min\{j-i,0\}}. \tag{3.42}
\]

(ii) \( \alpha \gamma \neq 0 \). Set \( r = \frac{1}{2\alpha} (-\beta + \sqrt{\beta^2 - 4\alpha \gamma}) \), \( s = \frac{1}{2\alpha} (-\beta - \sqrt{\beta^2 - 4\alpha \gamma}) \) and

\[\Lambda = \{ \beta + 2\gamma \sqrt{\alpha/\gamma} \cos(k\pi/(n+1)) \mid k \in \{1,2,\ldots,n\} \}. \]

Then \( rs \neq 0 \). Moreover, \( M \) is invertible \( \iff 0 \notin \Lambda \), in which case

\[
r \neq s \Rightarrow [M^{-1}]_{i,j} = -\frac{\gamma^{-1} (r^{\min\{i,j\}} - s^{\min\{i,j\}}) (r^{n+1} s^{\max\{i,j\}} - r^{\max\{i,j\}} s^{n+1})}{\alpha(r-s)(r^{n+1} - s^{n+1})}, \tag{3.43a}
\]

\[
r = s \Rightarrow [M^{-1}]_{i,j} = -\frac{\gamma^{-1} \min\{i,j\} (n+1 - \max\{i,j\})}{\alpha(n+1)} r^{i+j-1}. \tag{3.43b}
\]

Alternatively, define the recurrence relations

\[
u_{n+1} = 0, \quad v_1 = 1, \quad v_k = -\frac{1}{\alpha} (\beta v_{k-1} + \gamma v_{k+1}), \quad k \leq n - 1. \tag{3.44b}
\]

Then

\[
[M^{-1}]_{i,j} = -\frac{u_{\min\{i,j\}} v_{\max\{i,j\}}}{\alpha v_0} \left( \frac{\gamma}{\alpha} \right)^{i-1}. \tag{3.45}
\]

\(2\)In the special case, when \( \beta = 0 \), this is equivalent to saying that \( M \) is invertible \( \iff n \) is even.
3.5. Examples of linear monotone operators

Proof. (i): \( \alpha \gamma = 0 \Leftrightarrow \alpha = 0 \) or \( \gamma = 0 \), in which case \( M \) is a (lower or upper) triangular matrix. Hence \( \det(M) = \beta^n \), and the characterization follows. The formula in (3.42) is easily verified. (ii): Note that \( 0 \in \{ r, s \} \Leftrightarrow \beta \in \{ \pm \sqrt{\beta^2 - 4\alpha \gamma} \} \Leftrightarrow \beta^2 = \beta^2 - 4\alpha \gamma \Leftrightarrow \alpha \gamma = 0 \). Hence \( rs \neq 0 \). Moreover, it follows from [102, Example 7.2.5] that \( \Lambda \) is the set of eigenvalues of \( M \); therefore, \( M \) is invertible \( \Leftrightarrow 0 \notin \Lambda \). The formulae (3.43) follow from [136, Remark 2 on page 110]. The recurrence formulae defined in (3.44) and [136, Theorem 2] yield (3.45).

Remark 3.42. Concerning Proposition 3.41, using [134, Section 2 on page 44] we also have the alternative formulae

\[
\begin{align*}
\text{if } r \neq s & \Rightarrow [M^{-1}]_{i,j} = \begin{cases} 
-\frac{1}{\gamma} s^{-i} r^{-i} s^{-n+j-1} r^{-n+j-1}, & j \geq i; \\
-\frac{1}{\alpha} s^{-i} r^{-i} s^{-n+1} r^{-n+1}, & j \leq i,
\end{cases} \\
\text{if } r = s & \Rightarrow [M^{-1}]_{i,j} = \begin{cases} 
-\frac{i}{\gamma} \left( 1 - \frac{j}{n+1} \right) r^{j-i+1}, & j \geq i; \\
-\frac{i}{\alpha} \left( 1 - \frac{i}{n+1} \right) r^{j-i-1}, & j \leq i.
\end{cases}
\end{align*}
\]

Using the binomial expansion, (3.46a), and a somewhat tedious calculation which we omit here and provided that \( r \neq s \), one can show that \([M^{-1}]_{i,j}\) is equal to

\[
\frac{2S_1 S_2}{(2\alpha)^{\min\{i,j\}}(2\gamma)^{\min\{i,j\}} S_3},
\]

where

\[
\begin{align*}
S_1 &= \left[ \frac{\min\{i,j\}}{2} \right] \sum_{m=0}^{\min\{i,j\}} \left( \frac{\min\{i,j\}}{2m+1} \right) (-\beta)^{\min\{i,j\}} - (2m+1)(\beta^2 - 4\alpha \gamma)^m, \\
S_2 &= \left[ \frac{(n+1-\max\{i,j\})}{2} \right] \sum_{m=0}^{(n+1-\max\{i,j\})} \left( \frac{n-\max\{i,j\}}{2m+1} + 1 \right) (-\beta)^{n-\max\{i,j\}} - 2m(\beta^2 - 4\alpha \gamma)^m, \\
S_3 &= \left[ \frac{(n+1)}{2} \right] \sum_{m=0}^{(n+1)} \left( \frac{n+1}{2m+1} \right) (-\beta)^{n-2m}(\beta^2 - 4\alpha \gamma)^m.
\end{align*}
\]

23
3.5. Examples of linear monotone operators

Example 3.43. Let $\beta \geq 2$, set

$$M = \begin{pmatrix} \beta & -1 & 0 \\ -1 & \ddots & \ddots \\ \ddots & \ddots & -1 \\ 0 & -1 & \beta \end{pmatrix} \in \mathbb{R}^{n \times n},$$  \hspace{1cm} (3.49)

set $r = \frac{1}{2}(\beta + \sqrt{\beta^2 - 4})$ and set $s = \frac{1}{2}(\beta - \sqrt{\beta^2 - 4})$. Then $M$ is monotone and invertible. Moreover,

$$(r - s) \neq 0 \Rightarrow [M^{-1}]_{ij} = \frac{r_{\min} \{i, j\} - s_{\min} \{i, j\} (r^{n+1} s_{\max} \{i, j\} - r_{\max} \{i, j\} s^{n+1})}{(r - s)(r^{n+1} - s^{n+1})},$$  \hspace{1cm} (3.50a)

$$(r = s) \Rightarrow [M^{-1}]_{ij} = \frac{\min \{i, j\} (n + 1 - \max \{i, j\})}{n + 1}. \hspace{1cm} (3.50b)$$

Alternatively, define the recurrence relations

$$u_0 = 0, \quad u_1 = 1, \quad u_k = \beta u_{k-1} - u_{k-2}, \quad k \geq 2, \hspace{1cm} (3.51a)$$
$$v_{n+1} = 0, \quad v_n = 1, \quad v_k = \beta v_{k+1} - v_{k+2}, \quad k \leq n - 1. \hspace{1cm} (3.51b)$$

Then

$$[M^{-1}]_{ij} = -\frac{u_{\min} \{i, j\} v_{\max} \{i, j\}}{v_0}. \hspace{1cm} (3.52)$$

Proof. The monotonicity of $M$ follows from Proposition 3.40 by noting that $\beta \geq 2 > 2 \cos(\pi/(n + 1))$. The same argument implies that

$$0 \notin \Lambda = \left\{ \beta - 2 \cos \left( \frac{k\pi}{n + 1} \right) \middle| k \in \{1, 2, \ldots, n\} \right\}. \hspace{1cm} (3.53)$$

Hence $M$ is invertible by Proposition 3.41(ii). Note that $\beta = 2 \Leftrightarrow \beta^2 - 4 = 0 \Leftrightarrow r = s = 1$. Now apply Proposition 3.41(ii). \text{ } \blacksquare

Let $M_1 = [\alpha_{i,j}]_{i,j=1}^n \in \mathbb{R}^{n \times n}$ and $M_2 = [\beta_{i,j}]_{i,j=1}^n \in \mathbb{R}^{n \times n}$. Recall that the Kronecker product of $M_1$ and $M_2$ (see [93, page 407] or [102, Exercise 7.6.10]) is defined by the block matrix

$$M_1 \otimes M_2 = [\alpha_{i,j} M_2] \in \mathbb{R}^{n^2 \times n^2}. \hspace{1cm} (3.54)$$

Lemma 3.44. Let $M_1$ and $M_2$ be symmetric matrices in $\mathbb{R}^{n \times n}$. Then $M_1 \otimes M_2 \in \mathbb{R}^{n^2 \times n^2}$ is symmetric.
3.5. Examples of linear monotone operators

Proof. Using [102, Exercise 7.8.11(a)] or [93, Proposition 1(e) on page 408] we have \((M_1 \otimes M_2)^T = M_1^T \otimes M_2^T = M_1 \otimes M_2\). ■

The following fact is very useful in the conclusion of the upcoming results.

**Fact 3.45.** Let \(M_1\) and \(M_2\) be in \(\mathbb{R}^{n \times n}\), with eigenvalues \(\{\lambda_k \mid k \in \{1, \ldots, n\}\}\) and \(\{\mu_k \mid k \in \{1, \ldots, n\}\}\), respectively. Then the eigenvalues of \(M_1 \otimes M_2\) are \(\{\lambda_j \mu_k \mid j, k \in \{1, \ldots, n\}\}\).

Proof. See [93, Corollary 1 on page 412] or [102, Exercise 7.8.11(b)]. ■

**Corollary 3.46.** Let \(M_1\) and \(M_2\) in \(\mathbb{R}^{n \times n}\) be monotone such that \(M_1\) or \(M_2\) is symmetric. Then \(M_1 \otimes M_2\) is monotone.

Proof. According to (3.32), it is suffices to show that all the eigenvalues of \(M_1 \otimes M_2 + (M_1 \otimes M_2)^T\) are nonnegative. Suppose first that \(M_1\) is symmetric. Then using [93, Proposition 1(e)&(c)] we have \(M_1 \otimes M_2 + (M_1 \otimes M_2)^T = M_1 \otimes M_2 + M_1 \otimes M_2^T = M_1 \otimes (M_2 + M_2^T)\). Since \(M_2\) is monotone, it follows from (3.32) that all the eigenvalues of \(M_2 + M_2^T\) are nonnegative. Now apply Fact 3.45 to \(M_1\) and \(M_2 + M_2^T\) to learn that all the eigenvalues of \(M_1 \otimes M_2 + (M_1 \otimes M_2)^T\) are nonnegative, hence \(M_1 \otimes M_2\) is monotone by (3.32b). A similar argument applies if \(M_2\) is monotone. ■

Note that the assumption that at least one matrix is symmetric is critical in Corollary 3.46, as we show in the next example.

**Example 3.47.** Suppose that

\[
M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.
\] (3.55)

Then \(M\) is monotone, with eigenvalues \(\{\pm i\}\), but not symmetric. However,

\[
M \otimes M = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}
\] (3.56)

is a symmetric matrix with eigenvalues \(\{\pm 1\}\) by Fact 3.45. Therefore \(M \otimes M\) is not monotone by (3.32).
3.5. Examples of linear monotone operators

**Proposition 3.48.** Let \( M \in \mathbb{R}^{n \times n} \) be symmetric. Then \( \text{Id} \otimes M \) is monotone \iff\ M \otimes \text{Id} \) is monotone \iff\ M is monotone, in which case we have

\[
J_{\text{Id}_n \otimes M} = \text{Id}_n \otimes J_M \quad \text{and} \quad J_{M \otimes \text{Id}_n} = J_M \otimes \text{Id}_n. \tag{3.57}
\]

**Proof.** In view of Fact 3.45 the sets of eigenvalues of \( \text{Id} \otimes M \), \( M \otimes \text{Id} \), and \( M \) coincide. It follows from Lemma 3.44 that \( \text{Id} \otimes M \) and \( M \otimes \text{Id} \) are symmetric. Now apply (3.32b) and use the monotonicity of \( M \). To prove (3.57), we use [93, Proposition 1(c) on page 408] to learn that

\[
J_{\text{Id}_n \otimes M} = \text{Id}_n \otimes (\text{Id}_n \otimes M)^{-1} = \text{Id}_n \otimes (\text{Id}_n \otimes M)^{-1} = \text{Id}_n \otimes (\text{Id}_n \otimes M)^{-1} = \text{Id}_n \otimes J_M. \tag{3.58}
\]

Therefore, we learn from [93, Corollary 1(b) on page 408] that

\[
J_{\text{Id}_n \otimes M} = \text{Id}_n \otimes J_M. \tag{3.57}
\]

The other identity in (3.57) is proved similarly. \( \blacksquare \)

**Corollary 3.49.** Let \( \beta \in \mathbb{R} \). Set

\[
M_{[\beta]} = \begin{pmatrix}
\beta & -1 & 0 \\
-1 & \ddots & \ddots \\
0 & \ddots & -1 \\
0 & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
\end{pmatrix} \in \mathbb{R}^{n^2 \times n^2}, \tag{3.59}
\]

and let \( M_{\downarrow} \) and \( M_{\uparrow} \) be block matrices in \( \mathbb{R}^{n^2 \times n^2} \) defined by

\[
M_{\downarrow} = \begin{pmatrix}
M_{[\beta]} & 0_n & 0_n \\
0_n & \ddots & \ddots \\
0_n & \ddots & 0_n \\
0_n & \ddots & \ddots & 0_n \\
0_n & \ddots & \ddots & \ddots & 0_n \\
0_n & \ddots & \ddots & \ddots & \ddots & 0_n \\
\end{pmatrix} \tag{3.60}
\]

and

\[
M_{\uparrow} = \begin{pmatrix}
\beta \text{Id}_n & -\text{Id}_n & 0_n \\
-\text{Id}_n & \ddots & \ddots \\
0_n & \ddots & -\text{Id}_n \\
0_n & \ddots & \ddots & -\text{Id}_n \\
0_n & \ddots & \ddots & \ddots & -\text{Id}_n \\
0_n & \ddots & \ddots & \ddots & \ddots & -\text{Id}_n \\
\end{pmatrix}. \tag{3.61}
\]

Then \( M_{\downarrow} = \text{Id}_n \otimes M_{[\beta]} \) and \( M_{\uparrow} = M_{[\beta]} \otimes \text{Id}_n \). Moreover, \( M_{\downarrow} \) is monotone \iff\ \( M_{\uparrow} \) is monotone \iff\ \( M_{[\beta]} \) is monotone \iff\ \( \beta \geq 2 \cos(\pi/(n+1)) \), in which case

\[
J_{M_{\downarrow}} = \text{Id}_n \otimes M_{[\beta - 1]}^{-1} \quad \text{and} \quad J_{M_{\uparrow}} = M_{[\beta + 1]}^{-1} \otimes \text{Id}_n \tag{3.62}
\]
Moreover the following hold:

Let \( \mathbf{x} = (\mathbf{M}_t, \mathbf{M}_r) \) follow by applying (3.57).

Proposition 3.50. Douglas–Rachford algorithm — see the Section 5.6 for details.

**Proof.** It is straightforward to verify that \( \mathbf{M}_t = \mathbf{I} \otimes \mathbf{M}[\beta] \) and \( \mathbf{M}_r = \mathbf{M}[\beta] \otimes \mathbf{I} \). It follows from Proposition 3.40 that \( \mathbf{M}[\beta] \) is monotone \( \iff \beta \geq 2 \cos \left( \frac{\pi}{n+1} \right) \). Now combine with Proposition 3.48. To prove (3.62) note that \( \mathbf{I} + \mathbf{M}[\beta] = \mathbf{M}[\beta+1] \), and therefore \( J_{\mathbf{M}[\beta]} = \mathbf{M}[\beta+1]^{-1} \). The conclusion follows by applying (3.57).

The above matrices play a key role in the original design of the Douglas–Rachford algorithm — see the Section 5.6 for details.

**Proposition 3.50.** Let \( n \in \{2, 3, \ldots\} \), let \( \mathbf{M} \in \mathbb{R}^{n \times n} \) and consider the block matrix

\[
\mathbf{M} = \begin{pmatrix}
\mathbf{M} & -\mathbf{I} \otimes \mathbf{I} \\
-\mathbf{I} \otimes \mathbf{M} & \ddots & \ddots \\
& \ddots & -\mathbf{I} \otimes \mathbf{M} \\
0 & -\mathbf{I} \otimes \mathbf{M} & \mathbf{M}
\end{pmatrix}.
\]  

(3.63)

Let \( \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \), where \( x_i \in \mathbb{R}^n, i \in \{1, 2, \ldots, n\} \). Then

\[
\langle \mathbf{x}, \mathbf{Mx} \rangle = \langle x_1, (\mathbf{M} - \mathbf{I})x_1 \rangle + \sum_{k=2}^{n-1} \langle x_k, (\mathbf{M} - 2\mathbf{I})x_k \rangle + \langle x_n, (\mathbf{M} - \mathbf{I})x_n \rangle + \sum_{i=1}^{n-1} \| x_i - x_{i+1} \|^2.
\]

(3.64)

Moreover the following hold:

(i) Suppose that \( n = 2 \). Then \( \mathbf{M} - \mathbf{I} \) is monotone \( \iff \mathbf{M} \) is monotone.

(ii) \( \mathbf{M} - 2\mathbf{I} \) is monotone \( \Rightarrow \mathbf{M} \) is monotone.

(iii) \( \mathbf{M} \) is monotone \( \Rightarrow \mathbf{M} - 2(1 - \frac{1}{n}) \mathbf{I} \) is monotone \( \Rightarrow \mathbf{M} \) is monotone.

**Proof.** We have

\[
\langle \mathbf{x}, \mathbf{Mx} \rangle = \langle (x_1, x_2, \ldots, x_n), (\mathbf{M}x_1 - x_2, -x_1 + \mathbf{M}x_2 - x_3, \ldots, -x_{n-1} + \mathbf{M}x_n) \rangle
\]

\[
= \langle x_1, \mathbf{M}x_1 \rangle - \langle x_1, x_2 \rangle - \langle x_1, x_2 \rangle + \langle x_2, \mathbf{M}x_2 \rangle - \langle x_2, x_3 \rangle
\]

\[
- \langle x_2, x_3 \rangle + \langle x_3, \mathbf{M}x_3 \rangle - \cdots - \langle x_2, x_3 \rangle + \langle x_n, \mathbf{M}x_n \rangle - \langle x_{n-1}, x_n \rangle
\]

\[
= \langle x_1, \mathbf{M}x_1 \rangle - 2\langle x_1, x_2 \rangle + \langle x_2, \mathbf{M}x_2 \rangle - 2\langle x_2, x_3 \rangle
\]

\[
- \cdots - 2\langle x_{n-1}, x_n \rangle
\]

\[
= \langle x_1, \mathbf{M}x_1 \rangle + \| x_1 - x_2 \|^2 - \| x_1 \|^2 - \| x_2 \|^2 + \langle x_2, \mathbf{M}x_2 \rangle + \| x_2 - x_3 \|^2
\]
3.5. Examples of linear monotone operators

\[ -\|x_2\|^2 - \|x_3\|^2 + \ldots + \|x_{n-1} - x_n\|^2 \]
\[ -\|x_{n-1}\|^2 - \|x_n\|^2 + \langle x_n, Mx_n \rangle \]
\[ = \langle x_1, Mx_1 \rangle - \|x_1\|^2 + \langle x_2, Mx_2 \rangle - 2\|x_2\|^2 + \ldots + \langle x_n, Mx_n \rangle \]
\[ -\|x_n\|^2 + \ldots + \|x_1 - x_2\|^2 + \|x_2 - x_3\|^2 + \ldots + \|x_{n-1} - x_n\|^2 \]
\[ = \langle x_1, (M - \text{Id})x_1 \rangle + \left( \sum_{k=2}^{n-1} \langle x_k, (M - 2 \text{Id})x_k \rangle \right) + \langle x_n, (M - \text{Id})x_n \rangle \]
\[ + \sum_{i=1}^{n-1} \|x_i - x_{i+1}\|^2. \]

(i) \(\Rightarrow\): Apply (3.64) with \(n = 2\). \(\Leftarrow\): Let \(y \in \mathbb{R}^2\). Applying (3.64) to the point \(x = (y, y) \in \mathbb{R}^4\), we get \(0 \leq \langle x, Mx \rangle = 2\langle y, (M - \text{Id})y \rangle\).

(ii): This is clear from (3.64). (iii): Let \(y \in \mathbb{R}^n\). Applying (3.64) to the point \(x = (y, y, \ldots, y) \in \mathbb{R}^{n^2}\) yields \(0 \leq \langle x, Mx \rangle = 2\langle y, (M - \text{Id})y \rangle + (n - 2)\langle y, (M - 2 \text{Id})y \rangle = \langle y, (nM - 2(n - 1) \text{Id})y \rangle\). Therefore, \(M - 2(1 - \frac{1}{n}) \text{Id}\) is monotone.

The converse of Proposition 3.50(ii) is not true in general, as we illustrate now.

Example 3.51. Set

\[ M = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \]  \hspace{1cm} (3.65)

and let \(M\) be as defined in Proposition 3.50. Then one verifies easily that \(M\) is monotone while \(M - 2 \text{Id}\) is not.

We now show that the converse of the implications in Proposition 3.50(iii) are not true in general.

Example 3.52. Set \(n = 2\), set \(M = \frac{1}{2} \text{Id} \in \mathbb{R}^{2 \times 2}\), and let \(M\) be as defined in Proposition 3.50. Then \(M\) is monotone but \(M - 2(1 - \frac{1}{2}) \text{Id} = -\frac{1}{2} \text{Id}\) is not monotone, and

\[ M = \frac{1}{2} \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 \\ -2 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix}. \]  \hspace{1cm} (3.66)

Note the \(M\) is symmetric and has eigenvalues \([-\frac{1}{2}, \frac{3}{2}]\), hence \(M\) is not monotone by (3.32).
Chapter 4

Attouch–Théra duality and paramonotonicity

4.1 Overview

In this chapter, we systematically study Attouch–Théra duality for the problem of finding a zero of the sum of two maximally monotone operators. Let us now summarize our main results.

- We observe a curious convexity property of the intersection of two sets involving the graphs of $A$ and $B$ (see Theorem 4.17). This relates to Passty’s work on the parallel sum as well as to Eckstein and Svaiter’s work on the Kuhn–Tucker set (a.k.a. extended solution set, see [63] and [74, Section 2.1]).

- We reveal the importance of paramonotonicity: in this case, it is possible to recover all primal solutions from one dual solution (see Theorem 4.28).

- We generalize the best approximation results by Bauschke–Combettes-Luke from normal cone operators to paramonotone operators with a common zero (see Corollary 4.41).

Except for facts with explicit references, this chapter is based on results that appear in [26] and [39].

4.2 Duality for monotone operators

In this section, we review and slightly refine the basic results on Attouch–Théra duality. Throughout this chapter we assume

\begin{equation}
A \text{ and } B \text{ are maximally monotone operators on } X. \tag{4.1}
\end{equation}

Definition 4.1 (primal problem). The primal problem, for the ordered pair $(A, B)$, is to find a zero of $A + B$. 

4.2. Duality for monotone operators

At first, it looks strange to define the primal problem with respect to the (ordered) pair \((A, B)\). The reason we must do this is to associate a unique dual problem. (The ambiguity arises because addition is commutative.) Now since \(A\) and \(B\) form a pair of maximally monotone operators, so do \(A^{-1}\) and \(B^{-\ominus}\). We thus define the dual pair

\[
(A, B)^* = (A^{-1}, B^{-\ominus}).
\]  

(4.2)

The biduality

\[
(A, B)^{**} = (A, B)
\]  

(4.3)

holds, since \((A^{-1})^{-1} = A\), \((B^\ominus)^\ominus = B\), and \((B^\ominus)^{-1} = (B^{-1})^\ominus\).

We are now in a position to formulate the dual problem.

**Definition 4.2 (Attouch–Théra dual problem).** The (Attouch–Théra) dual problem, for the ordered pair \((A, B)\), is to find a zero of \(A^{-1} + B^{-\ominus}\).

This duality was systematically studied by Attouch and Théra [3]; although, it is worth noting that it was also touched upon earlier by Mercier [101, page 40]. (See [108] for the special case of variational inequalities, and also [56] for work on Toland duality.) Additional relevant work can be found in [52] and [66]. In view of (4.3), it is clear that the primal problem is precisely the dual of the dual problem, as expected. One central aim of this chapter is to understand the interplay between the primal and dual solutions that we formally define next.

**Definition 4.3 (primal and dual solutions).** The primal solutions are the solutions to the primal problem and analogously for the dual solutions. We shall abbreviate these sets by

\[
Z_{(A, B)} = Z = (A + B)^{-1}(0),
\]  

(4.4)

and

\[
K_{(A, B)} = K = (A^{-1} + B^{-\ominus})^{-1}(0),
\]  

(4.5)

respectively.

As observed by Attouch and Théra in [3, Corollary 3.2], one has:

\[
Z \neq \emptyset \iff K \neq \emptyset.
\]  

(4.6)

Let us make this simple but important equivalence a little more precise. In order to do so, we define

\[
(\forall z \in X) \quad K_z = (Az) \cap (-Bz)
\]  

(4.7)
4.2. Duality for monotone operators

and

\[(\forall k \in X)\quad Z_k = (A^{-1}k) \cap (-B^{-\ominus}k) = (A^{-1}k) \cap (B^{-1}(-k)).\] (4.8)

As the next proposition illustrates, these objects are intimately tied to primal and dual solutions defined in Definition 4.3. This result is elementary and implicitly contained in [3] and [101].

**Proposition 4.4.** Let \(z \in X\) and \(k \in X\). Then the following hold.

(i) \(K_z\) and \(Z_k\) are closed convex (possibly empty) subsets of \(X\).

(ii) \( k \in K_z \Leftrightarrow z \in Z_k \).

(iii) \( z \in Z \Leftrightarrow K_z \neq \emptyset \).

(iv) \( \bigcup_{z \in Z} K_z = K \).

(v) \( Z \neq \emptyset \Leftrightarrow K \neq \emptyset \).

(vi) \( k \in K \Leftrightarrow Z_k \neq \emptyset \).

(vii) \( \bigcup_{k \in K} Z_k = Z \).

**Proof.** (i): Because \(A\) and \(B\) are maximally monotone, Fact 3.9 implies that the sets \(Az\) and \(Bz\) are closed and convex. Hence \(K_z\) is also closed and convex. We see analogously that \(Z_k\) is closed and convex as well.

(ii): This is easily verified from the definitions.

(iii): Indeed, \( z \in Z \Leftrightarrow 0 \in (A + B)z \Leftrightarrow (\exists a^* \in Az \cap (-Bz)) \Leftrightarrow (\exists a^* \in K_z) \Leftrightarrow K_z \neq \emptyset \).

(iv): Take \( k \in \bigcup_{z \in Z} K_z \). Then there exists \( z \in Z \) such that \( k \in K_z = Az \cap (-Bz) \). Hence \( z \in A^{-1}k \) and \( z \in (-B)^{-1}k = B^{-1}(-\text{Id})^{-1}k = B^{-1}(-k) \). Thus \( z \in A^{-1}k \) and \( -z \in B^{-\ominus}k \). Hence \( 0 \in (A^{-1} + B^{-\ominus})k \) and so \( k \in K \). The reverse inclusion is proved analogously.

(v): Combine (iii) and (iv).

(vi)&(vii): The proofs are analogous to the ones of (iii)&(iv).

Let us provide some examples illustrating these notions.

**Example 4.5.** Suppose that \(X = \mathbb{R}^2\), and that we consider the rotators by \(\pm \pi/2\), i.e.,

\[
A : \mathbb{R}^2 \to \mathbb{R}^2 : (x_1, x_2) \mapsto (x_2, -x_1)
\] (4.9)

and

\[
B : \mathbb{R}^2 \to \mathbb{R}^2 : (x_1, x_2) \mapsto (-x_2, x_1).
\] (4.10)
4.2. Duality for monotone operators

Note that $B = -A = A^{-1} = A^*$, where $A^*$ denote the adjoint operator. Hence $A + B = 0$, $Z = X$, and $(\forall z \in Z) K_z = \{ Az \} = \{-Bz\}$. Furthermore, $A^{-1} = B$ while the linearity of $B$ implies that $B^{-\ominus} = B^{-1} = -B = A$. Therefore, $(A,B)^* = (B,A)$. Hence $K = Z$, while $(\forall k \in K) Z_k = \{ A^{-1}k \} = \{ Bk \}$.

**Example 4.6.** Suppose that $X = \mathbb{R}^2$, that $A$ is the normal cone operator of $\mathbb{R}^2_+$, and that $B: X \to X: (x_1,x_2) \mapsto (-x_2,x_1)$ is the rotator by $\pi/2$. As already observed in Example 4.5, we have $B^{-1} = -B$ and $B^{-\ominus} = B^{-1} = -B$. A routine calculation yields

$$Z = \mathbb{R}_+ \times \{0\}; \quad (4.11)$$

thus, since $B$ is single-valued,

$$(\forall z = (z_1,0) \in Z) \quad K_z = \{-Bz\} = \{(0,-z_1)\}. \quad (4.12)$$

Thus,

$$K = \bigcup_{z \in Z} K_z = \{0\} \times \mathbb{R}_- \quad (4.13)$$

and so

$$(\forall k = (0,k_2) \in K) \quad Z_k = \{-B^{-\ominus}k\} = \{Bk\} = \{(k_2,0)\}. \quad (4.14)$$

The dual problem is to find a zero of $A^{-1} + B^{-\ominus}$, i.e., a zero of the sum of the normal cone operator of the negative orthant and the rotator by $-\pi/2$.

**Example 4.7 (convex feasibility).** Suppose that $A = N_U$ and $B = N_V$, where $U$ and $V$ are closed convex subsets of $X$ such that $U \cap V \neq \emptyset$. Then clearly $Z = U \cap V$. Using Fact 14.2(i) below (see also [18, Fact A1]), we deduce that $(\forall z \in Z) K_z = N_{U \cap V}(0) = K$. Note that we do know at least one dual solution: $0 \in K$. Thus, by Proposition 4.4(ii)&(vii), $(\forall k \in K) Z_k = Z$.

**Remark 4.8.** The preceding examples give some credence to the conjecture that

$$\begin{cases} z_1 \in Z \\ z_2 \in Z \\ z_1 \neq z_2 \end{cases} \quad \Rightarrow \quad \text{either } K_{z_1} = K_{z_2} \text{ or } K_{z_1} \cap K_{z_2} = \emptyset. \quad (4.15)$$

Note that (4.15) is trivially true whenever $A$ or $B$ is at most single-valued. While this conjecture fails in general (see Example 4.9 below), it does, however, hold true for the large class of paramonotone operators (see Theorem 4.28).
4.3. Solution mappings $K$ and $Z$

Example 4.9. Suppose that $X = \mathbb{R}^2$, and set $U = \mathbb{R} \times \mathbb{R}_+$, $V = \mathbb{R} \times \{0\}$, and $R: X \to X: (x_1, x_2) \mapsto (-x_2, x_1)$. Now suppose that $A = N_U + R$ and that $B = N_V$. Then $\text{dom} \ A = U$ and $\text{dom} \ B = V$; hence, $\text{dom} \ (A + B) = U \cap V = V$. Let $x = (\xi, 0) \in V$. Then $Ax = \{0\} \times [-\infty, \xi]$ and $Bx = \{0\} \times \mathbb{R}$. Hence $Ax \subseteq Bx$, $(A + B)x = \{0\} \times \mathbb{R}$ and therefore $Z = V$. Furthermore, $K_x = Ax \cap (Bx) = Ax$. Now take $y = (\eta, 0) \in V = Z$ with $\xi < \eta$. Then $K_x = Ax \subseteq Ay = K_y$ and thus (4.15) fails.

Proposition 4.10 (common zeros). $\text{zer} \ A \cap \text{zer} \ B \neq \emptyset \iff 0 \in K$.

Proof. Suppose first that $z \in \text{zer} \ A \cap \text{zer} \ B$. Then $0 \in Az$ and $0 \in Bz$, so $0 \in Az \cap (Bz) = K_z \subseteq K$. Now assume that $0 \in K_z$ for some $z \in Z$ and so $0 \in Az \cap (Bz)$. Therefore, $0 \in \text{zer} \ A \cap \text{zer} \ B$.

Example 4.11. Suppose that $B = A$. Then $Z = \text{zer} \ A$, and $\text{zer} \ A \neq \emptyset \iff 0 \in K$.

Proof. Since $2A$ is maximally monotone and $A + A$ is a monotone extension of $2A$, we deduce that $A + A = 2A$. Hence $Z = \text{zer} \ (2A) = \text{zer} \ A$ and the result follows from Proposition 4.10.

The following result, observed first by Passty, is very useful. For the sake of completeness, we include its short proof.

Proposition 4.12 (Passty 1986). Let $x \in X$ and suppose that, for every $i \in \{0, 1\}$, $w_i \in Ay_i \cap B(x - y_i)$. Then $\langle y_0 - y_1, w_0 - w_1 \rangle = 0$.

Proof. (See [115, Lemma 14] and [115, Proposition 14]). Since $A$ is monotone, $0 \leq \langle y_0 - y_1, w_0 - w_1 \rangle$. On the other hand, since $B$ is monotone, $0 \leq \langle (x - y_0) - (x - y_1), w_0 - w_1 \rangle = \langle y_1 - y_0, w_0 - w_1 \rangle$. Altogether, $\langle y_0 - y_1, w_0 - w_1 \rangle = 0$.

Corollary 4.13. Suppose that $z_1$ and $z_2$ belong to $Z$, that $k_1 \in K_{z_1}$, and that $k_2 \in K_{z_2}$. Then $\langle k_1 - k_2, z_1 - z_2 \rangle = 0$.

Proof. Apply Proposition 4.12 ($x, B$) replaced by $(0, B^\ominus)$.

4.3 Solution mappings $K$ and $Z$

In this section we focus on the solution mappings between primal and dual solutions. We now interpret the families of sets $(K_z)_{z \in X}$ and $(Z_k)_{k \in X}$ as set-
valued operators by setting

\[ K: X \ni x \mapsto K_x \quad \text{and} \quad Z: X \ni x \mapsto Z_k. \tag{4.16} \]

Let us record some basic properties of these fundamental operators.

**Proposition 4.14.** The following hold.

(i) \( \text{gr} K = \text{gr} A \cap \text{gr} (-B) \) and \( \text{gr} Z = \text{gr} A^{-1} \cap \text{gr} (-B^{-1}) \).

(ii) \( \text{dom} K = Z, \ \text{ran} K = K, \ \text{dom} Z = K, \ \text{and} \ \text{ran} Z = Z. \)

(iii) \( \text{gr} K \) and \( \text{gr} Z \) are closed sets.

(iv) The operators \( K, -K, Z, -Z \) are monotone.

(v) \( K^{-1} = Z. \)

*Proof.* (i): This is clear from the definitions.

(ii): This follows from Proposition 4.4.

(iii): Since \( A \) and \( B \) are maximally monotone, the sets \( \text{gr} A \) and \( \text{gr} B \) are closed. Hence, by (i), \( \text{gr} K \) is closed and similarly for \( \text{gr} Z \).

(iv): Since \( \text{gr} K \subseteq \text{gr} A \) and \( A \) is monotone, we see that \( K \) is monotone. Similarly, since \( B \) is monotone and \( \text{gr} (-K) \subseteq \text{gr} B \), we obtain the monotonicity of \( -K \). The proofs for \( \pm Z \) are analogous.

(v): Clear from Proposition 4.4(ii). \[ \square \]

In Proposition 4.4(i), we observed the closedness and convexity of \( K_x \) and \( Z_k \). In view of Proposition 4.4(iii)&(vii), the sets of primal and dual solutions are both unions of closed convex sets. It would seem that we cannot a priori deduce convexity of these solution sets because unions of convex sets need not be convex. However, not only are \( Z \) and \( K \) indeed convex, but so are \( \text{gr} Z \) and \( \text{gr} K \). This surprising result, which is basically contained in works by Passty [115] and by Eckstein and Svaiter [74, 75], is best stated by using the parallel sum, a notion systematically explored by Passty in [115]. See also [97].

**Definition 4.15 (parallel sum).** (see [97], [115], or [12, Section 24.4]). The parallel sum of \( A \) and \( B \) is

\[ A \square B = (A^{-1} + B^{-1})^{-1}. \tag{4.17} \]
Let us recall that the infimal convolution of \( f \) and \( g \), denoted by \( f \square g \), is defined by
\[
f \square g : X \to \mathbb{R} : x \mapsto \inf_{y \in X} (f(y) + g(x - y)).
\]
(4.18)

The notation we use for the parallel sum (see [12, Section 24.4]) is nonstandard but highly convenient in view of the following fact.

**Fact 4.16.** Let \( f \) and \( g \) be in \( \Gamma(X) \) such that \( 0 \in \text{sri}(\text{dom} f^* - \text{dom} g^*) \). Then \( f \square g \in \Gamma(X) \) and
\[
\partial(f \square g) = (\partial f) \square (\partial g).
\]
(4.19)

**Proof.** See [115, Theorem 28], [109, Proposition 4.2.2], or [12, Proposition 15.7(i) and Proposition 24.27]. □

The proof of the following result is contained in the proof of [116, Theorem 21], although Passty stated a much weaker conclusion. For the sake of completeness, we present his proof.

**Theorem 4.17.** For every \( x \in X \), the set
\[
\left( \text{gr} A \right) \cap ((x, 0) - \text{gr}(-B)) = \{ (y, w) \in \text{gr} A \mid (x - y, w) \in \text{gr} B \}
\]
(4.20)
is convex.

**Proof.** (See also [116, Proof of Theorem 21].) The identity (4.20) is easily verified. To tackle convexity, for every \( i \in \{0, 1\} \) take \( (y_i, w_i) \) from the intersection (4.20); equivalently,
\[
(\forall i \in \{0, 1\}) \quad w_i \in A y_i \cap B (x - y_i).
\]
(4.21)

By Proposition 4.12,
\[
\langle y_0 - y_1, w_0 - w_1 \rangle = 0.
\]
(4.22)

Now let \( t \in [0, 1] \), set \( (y_t, w_t) = (1 - t)(y_0, w_0) + t(y_1, w_1) \), and take \( (a, a^*) \in \text{gr} A \). Using (4.21) and the monotonicity of \( A \) in (4.23c), we obtain
\[
\langle y_t - a, w_t - a^* \rangle
\]
\[
= \langle (1 - t)(y_0 - a) + t(y_1 - a), (1 - t)(w_0 - a^*) + t(w_1 - a^*) \rangle
\]
\[
= (1 - t)^2 \langle y_0 - a, w_0 - a^* \rangle + t^2 \langle y_1 - a, w_1 - a^* \rangle
\]
\[
+ (1 - t)t \left( \langle y_0 - a, w_1 - a^* \rangle + \langle y_1 - a, w_0 - a^* \rangle \right)
\]
\[
\geq (1 - t)t \left( \langle y_0 - a, w_1 - a^* \rangle + \langle y_1 - a, w_0 - a^* \rangle \right).
\]
(4.23a)
(4.23b)
(4.23c)
4.3. Solution mappings $K$ and $Z$

Thus, using again monotonicity of $A$ and recalling (4.22), we obtain

$$
\langle y_0 - a, w_1 - a^* \rangle + \langle y_1 - a, w_0 - a^* \rangle \\
= \langle y_0 - a, w_1 - w_0 \rangle + \langle y_0 - a, w_0 - a^* \rangle \\
+ \langle y_1 - a, w_0 - w_1 \rangle + \langle y_1 - a, w_1 - a^* \rangle \tag{4.24a}
$$

$$
= \langle y_1 - y_0, w_0 - w_1 \rangle \\
+ \langle y_0 - a, w_0 - a^* \rangle + \langle y_1 - a, w_1 - a^* \rangle \tag{4.24b}
$$

$$
\geq \langle y_1 - y_0, w_0 - w_1 \rangle \tag{4.24c}
$$

$$
= 0. \tag{4.24d}
$$

Combining (4.23) and (4.24), we obtain

$$
\langle y_t - a, w_t - a^* \rangle \geq 0. \tag{4.24e}
$$

Since $(a, a^*)$ is an arbitrary element of $\text{gr} A$ and $A$ is maximally monotone, we deduce that $(y_t, w_t) \in \text{gr} A$. A similar argument yields $(x - y_t, w_t) \in \text{gr} B$. Therefore, $(y_t, w_t)$ is an element of the intersection (4.20).

Before returning to the objects of interest, we record Passty’s [115, Theorem 21] as a simple corollary.

**Corollary 4.18 (Passty 1986).** For every $x \in X$, the set $(A \Box B)x$ is convex.

**Proof.** Let $x \in X$. Since the mapping $(y, w) \mapsto w$ is linear and the set

$$
\{ (y, w) \in \text{gr} A \mid (x - y, w) \in \text{gr} B \}
$$

is convex (Theorem 4.17), we deduce that

$$
\{ w \in X \mid (\exists y \in X) \ w \in Ay \cap B(x - y) \}
$$

is convex. (4.25)

On the other hand, a direct computation or [115, Lemma 2] implies that $(A \Box B)x = \bigcup_{y \in X} (Ay \cap B(x - y))$. Altogether, $(A \Box B)x$ is convex.

**Corollary 4.19.** For every $x \in X$, the set $(\text{gr} A) \cap ((x, 0) + \text{gr}(-B))$ is convex.

**Proof.** On the one hand, $- \text{gr}(-B^\circ) = \text{gr}(-B)$. On the other hand, $B^\circ$ is maximally monotone. Altogether, Theorem 4.17 (applied with $B^\circ$ instead of $B$) implies that $(\text{gr} A) \cap ((x, 0) - \text{gr}(-B^\circ)) = (\text{gr} A) \cap ((x, 0) + \text{gr}(-B))$ is convex.

**Remark 4.20.** Using Theorem 4.17 and Corollary 4.19 we learn that the intersections $(\text{gr} A) \cap \pm (\text{gr}(-B))$ are convex. This somewhat resembles works by Martínez-Legaz (see [100, Theorem 2.1]) and by Zălinescu [138], who encountered convexity when studying the sum/difference $(\text{gr} A) \pm (\text{gr}(-B))$. 

36
4.3. Solution mappings $K$ and $Z$

**Corollary 4.21 (convexity).** The sets $\operatorname{gr} Z$ and $\operatorname{gr} K$ are convex; consequently, $Z$ and $K$ are convex.

**Proof.** Combining Proposition 4.14(i) and Corollary 4.19 (with $x = 0$), we obtain the convexity of $\operatorname{gr} K$. Hence $\operatorname{gr} Z$ is convex by Proposition 4.14(v). It thus follows that $Z$ and $K$ are convex as images of convex sets under linear transformations. ■

**Remark 4.22.** Since $Z = (A^{-1} \Box B^{-1})(0)$ and $K = (A \Box B^\ominus)(0)$, the convexity of $Z$ and $K$ also follows from Corollary 4.18.

**Lemma 4.23.** Suppose that $B : X \Rightarrow X$ is linear. Then the following hold:

(i) $B^\ominus = B$ and $B^{-\ominus} = B^{-1}$.

(ii) $(A^{-1}, B^{-\ominus}) = (A^{-1}, B^{-1})$.

(iii) $K = (A \Box B)(0)$.

**Proof.** This is straightforward from the definitions. ■

**Remark 4.24 (Connection to Eckstein and Svaiter’s “extended solution set”).** In [74, Section 2.1], Eckstein and Svaiter considered the Kuhn–Tucker set (a.k.a. extended solution set) (for the primal problem), defined by

$$S_{(A,B)} = \{(z,w) \in X \times X \mid -w \in Bz, w \in Az\}. \quad (4.26)$$

It is clear that $\operatorname{gr} Z^{-1} = \operatorname{gr} K = S_{(A,B)}$. Unaware of Passty’s work, they proved in [74, Lemma 1 and Lemma 2] (in the present notation) that $Z = \operatorname{ran} Z$, and that $\operatorname{gr} Z$ is closed and convex. Their proof is very elegant and completely different from the above Passty-like proof. In their 2009 follow-up paper [75], Eckstein and Svaiter generalize the notion of the Kuhn–Tucker set to three or more operators; their corresponding proof of convexity in [75, Proposition 2.2] is more direct and along Passty’s lines. Note that the extension to three or more operators in [75] is a product space reformulation of the corresponding result in [74] and that such results are significantly extended in [52] and [66].

**Remark 4.25 (Convexity of $Z$ and $K$).** If $Z$ is nonempty and a constraint qualification holds, then $A + B$ is maximally monotone (see [12, Section 24.1]) and therefore $Z = \operatorname{zer}(A + B)$ is convex. It is somewhat surprising that $Z$ is always convex even without the maximal monotonicity of $A + B$. 

37
One may inquire whether or not $Z$ is also closed, which is another standard property of zeros of maximally monotone operators. The next example illustrates that $Z$ may fail to be closed.

**Example 4.26 (Z need not be closed).** Suppose that $X = \ell_2$, the real Hilbert space of square-summable sequences. In [24, Example 3.17], the authors provide a monotone discontinuous linear at most single-valued operator $S$ on $X$ such that $S$ is maximally monotone and its adjoint $S^*$ is a maximally monotone single-valued extension of $-S$. Hence $\text{dom } S$ is not closed. Now assume that $A = S$ and $B = S^*$. Then $A + B$ is operator that is zero on the dense proper subspace $Z = \text{dom}(A + B) = \text{dom } S$ of $X$. Thus $Z$ fails to be closed. Furthermore, in the language of Passty’s parallel sums (see Remark 4.22), this also illustrates that the parallel sum need not map a point to a closed set.

**Remark 4.27.** We finish this section with some concluding remarks.

(i) We do not know whether or not such counterexamples can reside in finite-dimensional Hilbert spaces when $\text{dom } A \cap \text{dom } B \neq \emptyset$. On the one hand, in view of the forthcoming Corollary 4.30(i), any counterexample must feature at least one operator that is not paramonotone, which means that the operators cannot be simultaneously subdifferential operators of functions in $\Gamma(X)$. On the other hand, one has to avoid the situation when $A + B$ is maximally monotone, which happens when $\text{ri } \text{dom } A \cap \text{ri } \text{dom } B \neq \emptyset$. This means that neither is one of the operators allowed to have full domain, nor can they simultaneously have relatively open domains, which excludes the situation when both operators are maximally monotone linear relations (i.e., maximally monotone operators with graphs that are linear subspaces, see [23]).

(ii) We note that $K$ and $Z$ are in general not maximally monotone. Indeed if $Z$, say, is maximally monotone, then Corollary 4.21 and Fact 3.6 imply that $\mathbf{gr } Z$ is actually affine (i.e., a translate of a subspace) and so are $Z$ and $K$ (as range and domain of $Z$). However, the set $Z$ of Example 4.7 need not be an affine subspace (e.g., when $U$, $V$ and $Z$ coincide with the closed unit ball in $X$).

### 4.4 Paramonotonicity

In this section we provide new results that underline the importance of paramonotonicity in the understanding of the zeros of the sum.
4.4. Paramonotonicity

**Theorem 4.28.** Suppose that $A$ and $B$ are paramonotone. Then $(\forall z \in Z) K_z = K$ and $(\forall k \in K) Z_k = Z$.

**Proof.** Suppose that $z_1$ and $z_2$ belong to $Z$ and that $z_1 \neq z_2$. Take $k_1 \in K_{z_1} = Az_1 \cap (-Bz_1)$ and $k_2 \in K_{z_2} = Az_2 \cap (-Bz_2)$. By Corollary 4.13,

$$\langle k_1 - k_2, z_1 - z_2 \rangle = 0.$$  (4.27)

Since $A$ and $B$ are paramonotone, we have $k_2 \in Az_1$ and $-k_2 \in Bz_1$; equivalently, $k_2 \in K_{z_1}$. It follows that $K_{z_2} \subseteq K_{z_1}$. Since the reverse inclusion follows in the same fashion, we see that $K_{z_1} = K_{z_2}$. In view of Proposition 4.4(iv), $K_{z_1} = K$, which proves the first conclusion. Since $A$ and $B$ are paramonotone so are $A^{-1}$ and $B^{-\odot}$ by (3.23). Therefore, the second conclusion follows from what we already proved (applied to $A^{-1}$ and $B^{-\odot}$). ■

**Remark 4.29 (recovering all primal solutions from one dual solution).**

Suppose that $A$ and $B$ are paramonotone and we know one (arbitrary) dual solution, say $k_0 \in K$. Then

$$Z_{k_0} = A^{-1}k_0 \cap (B^{-1}(-k_0))$$  (4.28)

reverses the set $Z$ of all primal solutions, by Theorem 4.28. If $A = \partial f$ and $B = \partial g$, where $f$ and $g$ belong to $\Gamma(X)$, then, since $(\partial f)^{-1} = \partial f^*$ and $(\partial g)^{-1} = \partial g^*$, we obtain a formula well-known in Fenchel duality, namely,

$$Z = \partial f^*(k_0) \cap \partial g^*(-k_0).$$  (4.29)

We shall revisit this setting in more detail in Section 4.6. In striking contrast, the complete recovery of all primal solutions from one dual solution is generally impossible when at least one of the operators is no longer paramonotone — see, e.g., Example 4.6 where one of the operators is a normal cone operator.

**Corollary 4.30.** Suppose $A$ and $B$ are paramonotone. Then the following hold.

(i) $Z$ and $K$ are closed.

(ii) $\text{gr } K$ and $\text{gr } Z$ are the rectangles $Z \times K$ and $K \times Z$, respectively.

(iii) $(Z - Z) \perp (K - K)$.

(iv) $\text{span } (K - K) = X \Rightarrow Z$ is a singleton.
4.4. Paramonotonicity

(v) \( \text{span} (Z - Z) = X \Rightarrow K \text{ is a singleton.} \)

Proof. (i): Combine Theorem 4.28 and Proposition 4.4(i).
(ii): Clear from Theorem 4.28.
(iii): Combine Corollary 4.13 with Theorem 4.28.
(iv): In view of (iii), we have that
\[
0 = \langle Z - Z, K - K \rangle = \langle Z - Z, \text{span} (K - K) \rangle = \langle Z - Z, X \rangle
\]
\[\Rightarrow Z - Z = \{0\} \iff Z \text{ is a singleton.} \]
(v): This is verified analogously to the proof of (iv).

Corollary 4.31. Let \( U \) be a closed affine subspace of \( X \), suppose that \( A = N_U \) and that \( B \) is paramonotone. Then the following hold:

(i) \( Z = U \cap (B^{-1}((\text{par } U)^\perp)) \subseteq U. \)

(ii) \( (\forall z \in Z) \ K = (-Bz) \cap (\text{par } U)^\perp \subseteq (\text{par } U)^\perp. \)

(iii) \( K \perp (Z - Z). \)

Proof. Since \( A = N_C = \partial C \), it is paramonotone by Example 3.24(i).
(i): Let \( x \in X \). Then \( x \in Z \iff 0 \in Ax + Bx = (\text{par } U)^\perp + Bx \iff [x \in U \text{ and there exists } y \in X \text{ such that } y \in (\text{par } U)^\perp \text{ and } y \in Bx] \iff [x \in U \text{ and there exists } y \in X \text{ such that } x \in B^{-1}y \text{ and } y \in (\text{par } U)^\perp] \iff x \in U \cap B^{-1}((\text{par } U)^\perp). \)
(ii): Let \( z \in Z \). Applying Remark 4.29 to \( (A^{-1}, B^{-\circ}) \) yields \( K = (-Bz) \cap (Az) = (-Bz) \cap (\text{par } U)^\perp. \)

(iii): By (i) \( Z - Z \subseteq U - U = \text{par } U \). Now use (ii).

Theorem 4.32. Suppose that \( A \) and \( B \) are paramonotone and that \( \text{zer } A \cap \text{zer } B \neq \emptyset \). Then the following hold:

(i) \( Z = (\text{zer } A) \cap (\text{zer } B) \text{ and } 0 \in K. \)

(ii) \( K \perp (Z - Z). \)

If, in addition, \( A \) or \( B \) is single-valued, then we also have:

(iii) \( K = \{0\}. \)
4.5. Projection operators and solution sets

**Proof.** (i): Since \( \text{zer } A \cap \text{zer } B \neq \emptyset \), it follows from Proposition 4.10 that \( 0 \in K \). Now apply Remark 4.29 to get \( Z = A^{-1}(0) \cap B^{-1}(0) = (\text{zer } A) \cap (\text{zer } B) \).

(ii): Combine (i) and Corollary 4.30(iii). (iii): Let \( C \in \{A, B\} \) be single-valued. Using (i) we have \( Z \subseteq \text{zer } C \). Suppose that \( C = A \) and let \( z \in Z \).

We use Remark 4.29 applied to \((A^{-1}, B^{-1})\) to learn that \( K = (Az) \cap (Bz) \). Therefore \( \{0\} \subseteq K \subseteq Az \subseteq A(\text{zer } A) = \{0\} \). A similar argument applies if \( C = B \).

**Remark 4.33.** The conclusion of Theorem 4.32(i) generalizes the setting of convex feasibility problems. Indeed, suppose that \( A = N_U \) and \( B = N_V \), where \( U \) and \( V \) are nonempty closed convex subsets of \( X \) such that \( U \cap V \neq \emptyset \). Then \( Z = U \cap V = \text{zer } A \cap \text{zer } B \).

The assumptions that \( A \) and \( B \) are paramonotone are critical in the conclusion of Theorem 4.32(i) as we illustrate now.

**Example 4.34.** Suppose that \( X = \mathbb{R}^2 \), that \( U = \mathbb{R} \times \{0\} \), that \( A = N_U \) and that \( B : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (-y, x) \), is the counterclockwise rotator in the plane by \( \pi/2 \). Then one verifies that \( \text{zer } A = U \), \( \text{zer } B = \{(0, 0)\} \), \( Z = \text{zer } (A + B) = U \); however \( (\text{zer } A) \cap (\text{zer } B) = \{(0, 0)\} \neq U = Z \). Note that \( A \) is paramonotone by Example 3.24(i) while \( B \) is not paramonotone by Example 3.25.

In the light of Example 4.7 we learn that if neither \( A \) nor \( B \) is single-valued, then the conclusion of Theorem 4.32(iii) may fail.

**Remark 4.35 (paramonotonicity is critical).** Various results in this section — e.g., Theorem 4.28, Remark 4.29, Corollary 4.30(ii)–(v) — fail if the assumption of paramonotonicity is omitted. To generate these counterexamples, assume that \( A \) and \( B \) are as in Example 4.5 or Example 4.6.

4.5 Projection operators and solution sets

**Theorem 4.36.** Suppose that \( A \) and \( B \) are paramonotone, that \( (z_0, k_0) \in Z \times K \), and that \( x \in X \). Then the following hold.

(i) \( Z + K \) is convex and closed.

(ii) \( P_{Z+K}(x) = P_Z(x - k_0) + P_K(x - z_0) \).

(iii) If \( (Z - Z) \perp K \), then \( P_{Z+K}(x) = P_Z(x) + P_K(x - z_0) \).
4.5. Projection operators and solution sets

(iv) If $Z \perp (K - K)$, then $P_{Z+K}(x) = P_Z(x - k_0) + P_K(x)$.

Proof. (i): The convexity and closedness of $Z$ and $K$ follows from Corollary 4.21 and Corollary 4.30(i). By Corollary 4.30(iii),

$$ (Z - z_0) \perp (K - k_0). \quad (4.30) $$

Using Fact 2.10,

$$ Z + K - z_0 - k_0 \text{ is convex and closed,} $$

and $P_{Z+K-z_0-k_0} = P_{Z-z_0} + P_{K-k_0}$. \quad (4.31)

Hence $Z + K$ is convex and closed. (ii): Using (4.31), Fact 2.10, and Fact 2.6, we obtain

$$ P_{Z+K}x = P_{(z_0+k_0)+(Z+K-z_0-k_0)}x \quad (4.32a) $$

$$ = z_0 + k_0 + P_{(Z-z_0)+(K-k_0)}(x - (z_0 + k_0)) \quad (4.32b) $$

$$ = z_0 + P_{Z-z_0}((x - k_0) - z_0) + k_0 + P_{K-k_0}((x - z_0) - k_0) \quad (4.32c) $$

$$ = P_Z(x - k_0) + P_K(x - z_0). \quad (4.32d) $$

(iii): Using Fact 2.10 and Fact 2.6, we have

$$ P_{Z+K}x = P_{z_0+(Z+K-z_0)}x \quad (4.33a) $$

$$ = z_0 + P_{(Z-z_0)+K}(x - z_0) \quad (4.33b) $$

$$ = z_0 + P_{Z-z_0}(x - z_0) + P_K(x - z_0) \quad (4.33c) $$

$$ = P_Zx + P_K(x - z_0). \quad (4.33d) $$

(iv): Argue analogously to the proof of (iii). \hfill \blacksquare

Remark 4.37. Suppose that $A$ and $B$ are paramonotone and that $0 \in K$. Then Corollary 4.30(iii) implies that $(Z - Z) \perp K - \{0\} = K$ and we thus may employ Theorem 4.36(ii) (with $k_0 = 0$) or Theorem 4.36(iii) to obtain the formula for $P_{Z+K}$.

However, if $(Z - Z) \perp K$, then the next two examples show—in strikingly different ways since $Z$ is either large or small—that we cannot conclude that $0 \in K$:

Example 4.38. Fix $u \in X$ and suppose that $(\forall x \in X) \ Ax = u$ and $B = -A$. Then $A$ and $B$ are paramonotone, $A + B \equiv 0$, and hence $Z = X$. Furthermore, $K = \{u\}$. Thus if $u \neq 0$, then $K \not\perp X = (Z - Z)$.  

42
Example 4.39. Let $U$ and $V$ be closed convex subsets of $X$ such that
\[ 0 \notin U \cap V \text{ and } U - V = X. \tag{4.34} \]
(For example, suppose that $X = \mathbb{R}$ and set $U = V = [1, +\infty[.)$ Now assume that $(A, B) = (N_U, N_V)^\ast$. In view of Example 4.7, $K = U \cap V$ and $Z = N_{U - V}(0) = N_X(0) = \{0\}$. Hence $Z$ is a singleton and thus $Z - Z = \{0\} \perp K$ while $0 \notin K$.

Theorem 4.40. Suppose that $A$ and $B$ are paramonotone, let $k_0 \in K$, and let $x \in X$. Then the following hold.

(i) $J_A P_{Z + k}(x) = P_{Z}(x - k_0)$.

(ii) If $(Z - Z) \perp K$, then $J_A \circ P_{Z + k} = P_Z$.

Proof. Take an arbitrary $z_0 \in Z$. (i): Set $z = P_{Z}(x - k_0)$. Using Theorem 4.36(ii) and Theorem 4.28, we have
\[ P_{Z + k}x - z = P_{Z + k}x - P_{Z}(x - k_0) = P_{k}(x - z_0) \in K = K_z \subseteq Ax. \tag{4.35} \]
Hence $P_{Z + k}x \in (\text{Id} + A)z \iff z = J_A P_{Z + k}x \iff P_{Z}(x - k_0) = J_A P_{Z + k}x$.

(ii): This time, let us set $z = P_{Z}x$. Using Theorem 4.36(iii) and Theorem 4.28, we have
\[ P_{Z + k}x - z = P_{Z + k}x - P_{Z}x = P_{k}(x - z_0) \in K = K_z \subseteq Ax. \tag{4.36} \]
Hence $P_{Z + k}x \in (\text{Id} + A)z \iff z = J_A P_{Z + k}x \iff P_{Z}x = J_A P_{Z + k}x$. ■

Corollary 4.41. Suppose that $A$ and $B$ are paramonotone, and that $0 \in K$. Then
\[ P_Z = J_A P_{Z + k}. \tag{4.37} \]

4.6 Subdifferential operators

In this section, we assume that
\[ A = \partial f \text{ and } B = \partial g, \tag{4.38} \]
where $f$ and $g$ belong to $\Gamma(X)$. Recall that (see Fact 3.5(ii))
\[ (\partial f)^{-1} = \partial f^\ast. \tag{4.39} \]
Let $f \in \Gamma(X)$ and let $g \in \Gamma(X)$.
Corollary 4.42 (subdifferential operators). The following hold:

(i) \( Z = (\partial f^* \square \partial g^*)(0) \).

(ii) \( K = (\partial f \square \partial g^*)(0) \).

(iii) Suppose that \( \text{argmin } f \cap \text{argmin } g \neq \emptyset \). Then \( Z = \partial f^*(0) \cap \partial g^*(0) \).

(iv) Suppose that \( 0 \in \text{sri}(\text{dom } f - \text{dom } g) \). Then \( Z = \partial (f^* \square g^*)(0) \).

(v) Suppose that \( 0 \in \text{sri}(\text{dom } f^* + \text{dom } g^*) \). Then \( K = \partial (f \square g^*)(0) \).

Proof. Note that \( A \) and \( B \) are paramonotone by Example 3.24(i). (i): Using (4.39) and Remark 4.22 we have
\[
Z = (A + B)^{-1}(0) = ((\partial f)^{-1} + (\partial g)^{-1})^{-1}(0) = ((\partial f^*)^{-1} + (\partial g^*)^{-1})^{-1}(0) = (\partial f^* \square \partial g^*)(0).
\]

(ii): Observe that \( (\partial g)^{-1} = (\partial g)^{-1} \neq (\partial g^*)^{-1} \). Therefore using Remark 4.22 we have
\[
K = (\partial f)^{-1} + (\partial g^*)^{-1}(0) = (\partial f)^{-1} + (\partial g^*)^{-1}(0) = (\partial f \square g^*)(0).
\]

(iii): Using Theorem 4.32(i), Fermat’s rule (see Fact 2.18) and (4.39) we have
\[
Z = (\text{zer } A) \cap (\text{zer } B) = \text{argmin } f \cap \text{argmin } g = (\partial f)^{-1}(0) \cap (\partial g)^{-1}(0) = \partial f^*(0) \cap \partial g^*(0).
\]

(iv): Combine (i) and [12, Proposition 24.27] applied to the functions \( f^* \) and \( g^* \).

(v): Combine (ii) and [12, Proposition 24.27] applied to the functions \( f \) and \( g^* \).

We now turn to applications to best approximation. Consider the primal problem

\[
\text{minimize } f(x) + g(x),
\]

the associated Fenchel dual problem

\[
\text{minimize } f^*(x^*) + g^*(-x^*),
\]

the primal and dual optimal values

\[
\mu = \inf(f + g)(X) \text{ and } \mu^* = \inf(f^* + g^*)(X).
\]

Note that
\[
\mu \geq -\mu^*.
\]

Following [47] and [48], we say that total duality holds if \( \mu = -\mu^* \in \mathbb{R} \), the primal problem (4.40) has a solution, and the dual problem (4.41) has a solution.

Theorem 4.43 (total duality). Suppose that \( A = \partial f \) and \( B = \partial g \), where \( f \) and \( g \) belong to \( \Gamma \). Then \( Z \neq \emptyset \iff \text{total duality holds, in which case } Z \text{ coincides with the set of solutions to the primal problem (4.40)}.\]
Moreover, dual problem (4.41). Thus, Take if dom If
\[ \mu = f(z) + g(z). \] (4.44)

Proof. Observe that (\( \partial f \))^{-1} = \partial f^* and that (\( \partial g \))^-\( \partial g^* \) = \partial (g^*)^\circ = \partial (g^*)^\circ.

\[ \Rightarrow: \] Suppose that \( Z \neq \emptyset \), and let \( z \in Z \). Then \( 0 \in \partial f(z) + \partial g(z) \subseteq \partial (f + g)(z) \). Hence \( z \) solves the primal problem (4.40), and

\[ \mu = f(z) + g(z). \] (4.44)

Take \( k \in K = K_z = (\partial f)_{\cap} (z) \cap (-\partial g)(z) \). First, we note that \( 0 \in (\partial f)^{-1}(k) + (\partial g)^{-\partial}(k) = \partial f^*(k) + \partial g^*(k) \subseteq \partial (f^* + g^*)(k) \) and so \( k \) solves the Fenchel dual problem (4.41). Thus,

\[ \mu^* = f^*(k) + g^*(k). \] (4.45)

Moreover, \( k \in \partial f(z) \) and \( -k \in \partial g(z) \), i.e., \( f(z) + f^*(k) = \langle z, k \rangle \) and \( g(z) + g^*(-k) = \langle z, -k \rangle \). Adding these equations gives \( 0 = f(z) + f^*(k) + g(z) + g^*(-k) = \mu + \mu^* \). This verifies total duality.

\[ \Leftarrow: \] Suppose we have total duality. Then there exists \( x \in \text{dom} f \cap \text{dom} g \) and \( x^* \in \text{dom} f^* \cap \text{dom} g^* \) such that

\[ f(x) + g(x) = \mu = -\mu^* = -f^*(x^*) - g^*(x^*) \in \mathbb{R}. \] (4.46)

Hence

\[ 0 = (f(x) + f^*(x^*)) + (g(x) + g^*(-x^*)) \geq \langle x, x^* \rangle + \langle x, -x^* \rangle = 0. \]

Therefore, using convex analysis and Proposition 4.4,

\[ (x^* \in \partial f(x) \text{ and } -x^* \in \partial g(x)) \iff x^* \in K_x \iff x \in Z_{x^*.} \] (4.47)

Hence \( x \in Z \).

Note that \( Z = \text{zer} (\partial f + \partial g) \subseteq \text{zer} \partial (f + g) \) since \( \text{gr} (\partial f + \partial g) \subseteq \text{gr} \partial (f + g) \). Hence \( Z \) is a subset of the set of primal solutions. Conversely, if \( x \) is a primal solution and \( x^* \) is a dual solution, then (4.46) holds and the rest of the proof of \( \Leftarrow \) shows that \( x \in Z \). Altogether, \( Z \) coincides with the set of primal solutions.

\[ \text{Remark 4.44 (sufficient conditions).} \] On the one hand,

\[ \text{the primal problem has at least one solution} \] (4.48)

if \( \text{dom} f \cap \text{dom} g \neq \emptyset \) and one of the following holds (see [12, Corollary 11.15]):

(i) \( f \) is supercoercive; (ii) \( f \) is coercive and \( g \) is bounded below; (iii) \( 0 \in \text{sr}(\text{dom} f^* + \text{dom} g^*) \) (by Fact 2.22 and since (4.40) is the Fenchel dual problem of (4.41)). On the other hand,

\[ \text{the sum rule } \partial (f + g) = \partial f + \partial g \text{ holds} \] (4.49)
whenever one of the following is satisfied (see [12, Corollary 16.38]): (i) (Attouch-Brezis condition) $\mathbb{R}_{++}(\text{dom } f - \text{dom } g)$ is a closed linear subspace; (ii) $\text{dom } f \cap \text{int dom } g \neq \emptyset$; (iii) $\text{dom } g = X$; (iv) $X$ is finite-dimensional and $\text{ri dom } f \cap \text{ri dom } g \neq \emptyset$. If both (4.48) and (4.49) hold, then $Z \neq \emptyset$ and $Z$ coincides with the set of primal solutions.
Chapter 5

Zeros of the sum and the Douglas–Rachford operator

5.1 Overview

Recall that the sum problem for two maximally monotone operators $A$ and $B$ is to

\[ \text{find } x \in X \text{ such that } 0 \in Ax + Bx. \tag{5.1} \]

When $\text{zer}(A + B) \neq \emptyset$, one approach to solve the problem is the Douglas–Rachford splitting technique. In this chapter we focus on the basic properties of the Douglas–Rachford splitting operator and the connection to the primal and dual sets of solutions. Our main results in this chapter are:

- We provide a new description of the fixed point set of the Douglas–Rachford splitting operator (see Theorem 5.14); that refines Combettes description of $Z = (A + B)^{-1}(0)$.

- Using the Douglas–Rachford operator we can formulate new results, on the relative geometry of the primal and dual solutions (in the sense of Attouch–Théra duality) (see Theorem 5.19).

- We sketch the connection between the historical Douglas–Rachford algorithm and the powerful extension provided by Lions and Mercier in [96] (see Section 5.6).

Except for facts with explicit references, this chapter is based on results that appear in [13], [26], [31] and [39].

5.2 Basic properties and facts

Definition 5.1. The Douglas–Rachford splitting operator [96] for the ordered pair of operators $(A, B)$ is defined by

\[ T_{(A,B)} = \frac{1}{2} (\text{Id} + R_BR_A) = \text{Id} - J_A + J_B R_A. \tag{5.2} \]
Let us record some useful properties of $T_{(A,B)}$.

**Fact 5.2.** The following hold:

(i) **(Lions and Mercier).** $T_{(A,B)}$ is firmly nonexpansive.

(ii) **(Eckstein).** $T_{(A,B)} = T_{(A^{-1},B^{-\ominus})}$.

(iii) **(Combettes).** $Z = J_{A}(\text{Fix } T_{(A,B)})$.

(iv) $K = J_{A^{-1}}(\text{Fix } T_{(A,B)})$.

(v) $\text{Fix } T_{(A,B)} = \text{Fix } R_{B}R_{A}$.

**Proof.** (i): See [96, Lemma 1], [71, Corollary 4.2.1 on page 139], [72, Corollary 4.1], or Theorem 5.9(i). (ii): See [71, Lemma 3.6 on page 133] or [26, Proposition 2.16]. (iii): See [61, Lemma 2.6(iii)]. (iv): Apply (iii) to the dual pair $(A^{-1},B^{-\ominus})$ and use (ii) or Corollary 5.12 below. (v): Clear from the definition.

It is clear from the definition and Fact 5.2(i) that $\text{Id} - T_{A,B}$ is also firmly nonexpansive. In fact, we note in passing that $\text{Id} - T_{A,B}$ is itself a Douglas–Rachford splitting operator:

**Proposition 5.3.** $\text{Id} - T_{(A,B)} = T_{(A,B^{-1})}$.

**Proof.** Using (3.15), we obtain

$$T_{(A,B)} + T_{(A,B^{-1})} = \text{Id} - J_{A} + J_{B}R_{A} + \text{Id} - J_{A} + J_{B^{-1}}R_{A}$$

$$= 2\text{Id} - 2J_{A} + (J_{B} + J_{B^{-1}})R_{A}$$

$$= 2\text{Id} - 2J_{A} + R_{A}$$

$$= \text{Id},$$

and the conclusion follows.

We will simply use $T$ instead of $T_{(A,B)}$ provided there is no cause for confusion.

**Fact 5.4 (Eckstein 1989).**

$$\text{gr}(T) = \{(a + a^*, a^* + b) \mid (a,a^*) \in \text{gr } A, (b,b^*) \in \text{gr } B, a - b = a^* + b^*\}.$$  

(5.4)
5.3 Useful identities for the Douglas–Rachford operator

Proof. See [71, Proposition 4.1]. □

Corollary 5.5. We have

\[ \text{gr}(\text{Id} - T) = \{ (a + a^*, a - b) \mid (a, a^*) \in \text{gr} A, (b, b^*) \in \text{gr} B, a - b = a^* + b^* \}; \]  

consequently,

\[ \text{ran}(\text{Id} - T) = \{ a - b \mid (a, a^*) \in \text{gr} A, (b, b^*) \in \text{gr} B, a - b = a^* + b^* \} \]  

\[ \subseteq (\text{dom} A - \text{dom} B) \cap (\text{ran} A + \text{ran} B). \]  

5.3 Useful identities for the Douglas–Rachford operator

We start with some elementary identities which are easily verified directly.

Lemma 5.6. Let \((a, b, z) \in X^3\). Then the following hold:

(i) \( \langle z, z - a + b \rangle = \|z - a + b\|^2 + \langle a, z - a \rangle + \langle b, 2a - z - b \rangle \).

(ii) \( \langle z, a - b \rangle = \|a - b\|^2 + \langle a, z - a \rangle + \langle b, 2a - z - b \rangle \).

(iii) \( \|z\|^2 = \|z - a + b\|^2 + \|b - a\|^2 + 2\langle a, z - a \rangle + 2\langle b, 2a - z - b \rangle \).

Lemma 5.7. Let \((a, b, x, y, a^*, b^*, u, v) \in X^8\). Then

\[ \langle (a, b) - (x, y), (a^*, b^*) - (u, v) \rangle = \langle a - b, a^* \rangle + \langle x, u \rangle - \langle x, a^* \rangle - \langle a - b, u \rangle + \langle b, a^* + b^* \rangle + \langle y, v \rangle - \langle y, b^* \rangle - \langle b, u + v \rangle. \]  

The following result will be useful.

Lemma 5.8. Let \(x \in X\). Then the following hold:

(i) \( x - Tx = J_Ax - J_BRAx = J_{A^{-1}}x + J_{B^{-1}}RAx \).

(ii) \( (J_Ax, J_BRAx, J_{A^{-1}}x, J_{B^{-1}}RAx) \) lies in \( \text{gra}(A \times B) \).
Proof. (i): The first identity is a direct consequence of (5.2). In view of (3.15) $I_A x - J_B R_A x = J_A x - (x - J_A x) - (J_B + J_B^{-1}) R_A x = R_A x - R_A x = 0$, which proves the second identity. (ii): Use (5.31) below and Fact 3.18 applied to $A \times B$ at $(x, R_A x) \in X \times X$.

The next theorem is a direct consequence of the key identities presented in Lemma 5.6.

**Theorem 5.9.** Let $x \in X$ and let $y \in X$. Then the following hold:

1. $\langle Tx - Ty, x - y \rangle = \|Tx - Ty\|^2 + \langle J_A x - J_A y, J_A^{-1} x - J_A^{-1} y \rangle$
   
   \begin{align*}
   &+ \langle J_B R_A x - J_B R_A y, J_B^{-1} R_A x - J_B^{-1} R_A y \rangle.
   \end{align*}

2. $\|J_A x - J_A y\|^2 + \|J_A^{-1} x - J_A^{-1} y\|^2 - \|J_A x - J_A^{-1} y\|^2

   = \langle (\text{Id} - T)x - (\text{Id} - T)y, x - y \rangle$

   \begin{align*}
   &+ \langle J_A x - J_A y, J_A^{-1} x - J_A^{-1} y \rangle
   &+ \langle J_B R_A x - J_B R_A y, J_B^{-1} R_A x - J_B^{-1} R_A y \rangle.
   \end{align*}

3. $\|x - y\|^2 = \|Tx - Ty\|^2 + \|J_A x - J_A^{-1} y\|^2

   \begin{align*}
   &+ 2 \langle J_A x - J_A y, J_A^{-1} x - J_A^{-1} y \rangle
   &+ 2 \langle J_B R_A x - J_B R_A y, J_B^{-1} R_A x - J_B^{-1} R_A y \rangle.
   \end{align*}

4. $\|J_A x - J_A y\|^2 + \|J_A^{-1} x - J_A^{-1} y\|^2 - \|J_A x - J_A^{-1} y\|^2

   = \langle (\text{Id} - T)x - (\text{Id} - T)y, x - y \rangle$

   \begin{align*}
   &+ 2 \langle J_A x - J_A y, J_A^{-1} x - J_A^{-1} y \rangle
   &+ 2 \langle J_B R_A x - J_B R_A y, J_B^{-1} R_A x - J_B^{-1} R_A y \rangle.
   \end{align*}

5. $\|J_A x - J_A y\|^2 + \|J_A x - J_A^{-1} y\|^2

   \begin{align*}
   \leq & \|J_A x - J_A y\|^2 + \|J_A^{-1} x - J_A^{-1} y\|^2.
   \end{align*}

Proof. (i)–(iii): Use Lemma 5.6(i)–(iii) respectively, with $z = x - y$, $a = J_A x - J_A y$ and $b = J_B R_A x - J_B R_A y$, (3.20) and (5.2). (iv) It follows from the (3.15) that

\begin{align*}
\|x - y\|^2 &= \|J_A x - J_A y + J_A^{-1} x - J_A^{-1} y\|^2 \\
&= \|J_A x - J_A y\|^2 + \|J_A^{-1} x - J_A^{-1} y\|^2 \\
&+ 2 \langle J_A x - J_A y, J_A^{-1} x - J_A^{-1} y \rangle.
\end{align*}
Applying (5.8) to \((Tx, Ty)\) instead of \((x, y)\) yields
\[
\|Tx - Ty\|^2 = \|J_ATx - J_A Ty\|^2 + \|J_A^{-1}Tx - J_A^{-1}Ty\|^2 \\
+ 2\langle J_ATx - J_A Ty, J_A^{-1}Tx - J_A^{-1}Ty \rangle.
\] (5.9)

Now combine (5.8), (5.9) and (iii) to obtain (iv). (v): In view of (3.20), the monotonicity of \(A\) and \(B\) implies \(\langle J_A Tx - J_A Ty, J_A^{-1}Tx - J_A^{-1}Ty \rangle \geq 0\) and \(\langle J_B R_A x - J_B R_A y, J_B^{-1} R_A x - J_B^{-1} R_A y \rangle \geq 0\). Now use (iv). □

**5.4 Douglas–Rachford operator and Attouch–Théra duality**

We start with some useful identities involving resolvents and reflected resolvents.

**Proposition 5.10.** Let \(C: X \rightrightarrows X\) be maximally monotone. Then the following hold.

(i) \(R_{C^{-1}} = -R_C\).

(ii) \(J_{C^\ominus} = J_{C}^\ominus\).

(iii) \(R_{C^\ominus} = \text{Id} - 2J_{C}^\ominus\).

**Proof.** (i): By (3.15), we have \(R_{C^{-1}} = 2J_{C^{-1}} - \text{Id} = 2(\text{Id} - J_C) - \text{Id} = \text{Id} - 2J_C = -(2J_C - \text{Id}) = -R_C\).

(ii): Indeed,
\[
J_{C^\ominus} = \left( \text{Id} + (\text{Id}) \circ C \circ (-\text{Id}) \right)^{-1} \\
= \left( (-\text{Id}) \circ (\text{Id} + C) \circ (-\text{Id}) \right)^{-1} \\
= (-\text{Id})^{-1} \circ (\text{Id} + C)^{-1} \circ (-\text{Id})^{-1} \\
= (-\text{Id}) \circ J_C \circ (-\text{Id}) \\
= J_{C}^\ominus.
\] (5.10a)

(iii): Using (3.15) and (ii), we have that \(R_{C^\ominus} = 2J_{C^\ominus} - \text{Id} = 2(\text{Id} - J_{C^\ominus}) - \text{Id} = \text{Id} - 2J_{C}^\ominus\). □

Recall that the Peaceman–Rachford operator for the ordered pair \((A, B)\) is \(R_B R_A\).
Corollary 5.11 (Peaceman–Rachford operator is self-dual). (See Eckstein’s [71, Lemma 3.5 on page 125].) The Peaceman–Rachford operators for \((A, B)\) and \((A, B)^* = (A^{-1}, B^{\ominus})\) coincide, i.e., we have self-duality in the sense that
\[ R_B R_A = R_{B^{\ominus}} R_{A^{-1}}. \] (5.11)
Consequently,
\[ (\forall \lambda \in [0, 1]) \quad (1 - \lambda) \text{Id} + \lambda R_B R_A = (1 - \lambda) \text{Id} + \lambda R_{B^{\ominus}} R_{A^{-1}}. \] (5.12)
Proof. Using Proposition 5.10(i)&(iii), we obtain
\[ R_{B^{\ominus}} R_{A^{-1}} = (\text{Id} - 2J_B^{\subseteq})(-R_A) = -R_A + 2J_B R_A = 2(J_B - \text{Id})R_A = R_B R_A. \] (5.13)
which proves (5.11). Now (5.12) follows immediately from (5.11). □

Corollary 5.12 (Douglas–Rachford operator is self-dual). (See Eckstein’s [71, Lemma 3.6 on page 133].) For the Douglas–Rachford operator
\[ T_{(A, B)} = \frac{1}{2} \text{Id} + \frac{1}{2} R_B R_A \] (5.14)
we have
\[ T_{(A, B)} = J_B R_A + \text{Id} - J_A = T_{(A^{-1}, B^{\ominus})}. \] (5.15)
Proof. The left equality is a simple expansion while self-duality is (5.12) with \(\lambda = \frac{1}{2}\). □

Remark 5.13 (backward-backward operator is not self-dual). In contrast to Corollary 5.11, the backward–backward operator is not self-dual: indeed, using (3.15) and Proposition 5.10(iii), we deduce that
\[ J_{B^{\ominus}} J_{A^{-1}} = (\text{Id} - J_B^{\subseteq})(\text{Id} - J_A) = \text{Id} - J_A + J_B(J_A - \text{Id}) = (J_B - \text{Id})(J_A - \text{Id}). \] (5.16)
Thus if \(A \equiv 0\) and \(\text{dom } B\) is not a singleton (equivalently, \(J_A = \text{Id}\) and \(\text{ran } J_B\) is not a singleton), then \(J_{B^{\ominus}} J_{A^{-1}} \equiv (J_B - \text{Id})0 \equiv J_B0 \neq J_B = J_B J_A\).

We now provide a new description of the fixed point set of the Douglas–Rachford splitting operator (see Theorem 5.14); this refines Combettes’ description of \((A + B)^{-1}(0)\).
Theorem 5.14. The mapping
\[
\Psi : \text{gr } K \rightarrow \text{Fix } T : (z, k) \mapsto z + k
\] (5.17)
is a well-defined bijection that is continuous in both directions, with \(\Psi^{-1} : x \mapsto (J_Ax, x - J_Ax)\).

Proof. Take \((z, k) \in \text{gr } K\). Then \(k \in K = (Az) \cap (-Bz)\). Now \(k \in Az \iff z + k \in (\text{Id} + A)z \iff z = J_A(z + k)\). Similarly, \(k \in (-Bz) \iff -k \in Bz \iff z - k \in (\text{Id} + B)z \iff z = J_B(z - k)\). Set \(x = z + k\). Then \(J_Ax = J_A(z + k) = z\) and hence \(R_Ax = 2J_Ax - x = 2z - (z + k) = z - k\). Thus,
\[
Tx = x - J_Ax + J_BR_Ax = z + k - z + J_B(z - k) = k + z = x,
\] (5.18)
i.e., \(x \in \text{Fix } T\). It follows that \(\Psi\) is well-defined.

Let us now show that \(\Psi\) is surjective. To this end, take \(x \in \text{Fix } T\). Set \(z = J_Ax\) as well as \(k = (\text{Id} - J_A)x = x - z\). Clearly,
\[
x = z + k.
\] (5.19)
Now \(z = J_Ax \iff x \in (\text{Id} + A)z = z + Az \iff k = x - z \in Az\). Thus,
\[
k \in Az.
\] (5.20)
We also have \(R_Ax = 2J_Ax - x = 2z - (z + k) = z - k\); hence, \(x = Tx = x - J_Ax + J_BR_Ax \iff J_Ax = J_BR_Ax \iff z = J_B(z - k) \iff z - k \in (\text{Id} + B)z = z + Bz \iff k \in -Bz\).
\[
(5.21)
\]
Altogether, \(k \in (Az) \cap (-Bz) = K = (z, k) \in \text{gr } K\). Hence \(\Psi\) is surjective.

In view of Fact 3.18 and since \(\text{gr } K \subseteq \text{gr } A\), it is clear that \(\Psi\) is injective and the result follows. \(\blacksquare\)

The following result is a straightforward consequence of Theorem 5.14.

Corollary 5.15. We have
\[
(\forall z \in Z ) \quad K_z = J_A^{-1}(J_A^{-1}z \cap \text{Fix } T)
\] (5.22)
and
\[
(\forall k \in K) \quad Z_k = J_A(J_A^{-1}k \cap \text{Fix } T).
\] (5.23)
5.4. Douglas–Rachford operator and Attouch–Théra duality

Proof. Let \((z,k) \in Z \times X\). Then \(k \in K_z \iff (z,k) \in \text{gra} K \iff (\exists f \in \text{Fix} T)\) such that \(z = J_A f\) and \(k = J_{A^{-1}} f \iff (\exists f \in \text{Fix} T)\) such that \(f \in J_A^{-1} z\) and \(k = J_{A^{-1}} f \iff k \in J_{A^{-1}} (J_A^{-1} z \cap \text{Fix} T)\). Similarly, one can prove \((5.23)\). ■

Recall that the so-called Kuhn–Tucker set (see Section 4.3) is defined by
\[
S = S_{(A,B)} = \{(z,k) \in X \times X \mid - k \in Bz, k \in Az\} \subseteq Z \times K. \tag{5.24}
\]

In the following result we reveal the importance of paramonotonicity. Indeed we prove in Corollary 5.16(v) that the fixed point set of the Douglas–Rachford splitting operator is a rectangle, see Corollary 5.16(v) below.

**Corollary 5.16.** Recalling Fact 3.18, let \(M: X \to \text{gra} A : x \mapsto (J_A x, J_{A^{-1}} x)\). Then the following hold:

(i) \(S = M(\text{Fix} T) = \{(J_A \times J_{A^{-1}})(y,y) \mid y \in \text{Fix} T\}\).

(ii) \(\text{Fix} T = M^{-1}(S) = \{z + k \mid (z,k) \in S\}\).

(iii) (Eckstein and Svaiter). \(S\) is closed and convex.

If \(A\) and \(B\) are paramonotone, then we additionally have:

(iv) \(S = Z \times K\).

(v) \(\text{Fix} T = Z + K\).

(vi) \(Z = J_A(Z + K)\).

(vii) \(K = J_{A^{-1}}(Z + K) = (\text{Id} - J_A)(Z + K)\).

Proof. (i)&(ii): This is Theorem 5.14. (iii): See [74, Lemma 2]. Alternatively, since \(\text{Fix} T\) is closed, and \(M\) and \(M^{-1}\) are continuous, we deduce the closedness from (i). The convexity was proved in Proposition 4.14(ii) and Remark 4.24. (iv): Combine Remark 4.24 and Corollary 4.30(ii). (v)–(vii): Combine Corollary 4.30(ii) with Theorem 5.14. ■

Specializing the result of Corollary 4.41 to normal cone operators, in view of Corollary 5.16(v), we recover the consistent case of [20, Corollary 3.9].

**Example 5.17.** Suppose that \(A = N_U\) and \(B = N_V\), where \(U\) and \(V\) are closed convex subsets of \(X\) such that \(U \cap V \neq \emptyset\). Then \(Z = U \cap V\), \(K = N_{U \cap V}(0)\), and
\[
P_Z = P_U P_{Z+K} = P_U P_{\text{Fix} T}. \tag{5.25}
\]

Proof. Combine Example 4.7, Corollary 5.16(v), and Corollary 4.41. ■
5.5 The Douglas–Rachford operator and solution sets

Lemma 5.18. Suppose that $A$ and $B$ are paramonotone. Let $k \in K$ be such that $(\forall z \in Z) \ J_A(z + k) = P_Z(z + k)$. Then $k \in (Z - Z)^\perp$.

Proof. By Corollary 5.16(v), $\text{Fix } T = Z + K$. Let $z_1$ and $z_2$ be in $Z$. It follows from Theorem 5.14 that $(\forall z \in Z) \ J_A(z + k) = z$. Therefore,

$$(\forall i \in \{1, 2\}) \ z_i + k \in \text{Fix } T \text{ and } z_i = J_A(z_i + k) = P_Z(z_i + k). \quad (5.26)$$

Furthermore, the Projection Theorem (see Fact 2.5) yields

$$\langle k, z_1 - z_2 \rangle = \langle z_1 + k - z_1, z_1 - z_2 \rangle \quad (5.27a)$$

$$= \langle z_1 + k - P_Z(z_1 + k), P_Z(z_1 + k) - z_2 \rangle \geq 0. \quad (5.27b)$$

On the other hand, interchanging the roles of $z_1$ and $z_2$ yields $\langle k, z_2 - z_1 \rangle \geq 0$. Altogether, $\langle k, z_1 - z_2 \rangle = 0$. ■

The next results relate the Douglas–Rachford operator to orthogonal properties of primal and dual solutions. These results extend and sharpen our results in Chapter 4.

Let $S$ be a subset of $X$ and let $T : X \to X$. We use $T|_S$ to refer to the restriction of the map $T$ to the set $S$.

Theorem 5.19. Suppose that $A$ and $B$ are paramonotone. Then the following are equivalent:

(i) $J_A P_{\text{Fix } T} = P_Z$.

(ii) $J_A|_{\text{Fix } T} = P_Z|_{\text{Fix } T}$.

(iii) $K \perp (Z - Z)$.

Proof. “(i)⇒(ii)” : This is obvious. “(ii)⇒(iii)” : Let $k \in K$ and let $z \in Z$. Then $\text{Fix } T = Z + K$ by Corollary 5.16(v); hence, $z + k \in \text{Fix } T$. Therefore $J_A(z + k) = P_Z(z + k)$. Now apply Lemma 5.18. “(iii)⇒(i)” : This follows from Theorem 4.40(ii) and Corollary 5.16(v). ■

As a consequence of Theorem 5.19 we now refine Example 5.17.

Example 5.20. Let $U$ be a closed affine subspace of $X$, suppose that $A = N_U$ and that $B$ is paramonotone such that $Z \neq \emptyset$. Then

$$P_Z = J_A P_{\text{Fix } T} = P_U P_{\text{Fix } T}. \quad (5.28)$$
5.5. The Douglas–Rachford operator and solution sets

\textit{Proof.} Combine Corollary 4.31(iii) and Theorem 5.19. □
Recall that Proposition 4.10 implies
\[ \text{zer } A \cap \text{zer } B \neq \emptyset \iff 0 \in K. \hspace{1cm} (5.29) \]

\textbf{Theorem 5.21.} Suppose that \( A \) and \( B \) are paramonotone and that \( \text{zer } A \cap \text{zer } B \neq \emptyset \). Then the following hold:

(i) \( J_A P_{\text{Fix } T} = P_Z \).

(ii) We have the implication \( A \) or \( B \) is single-valued \( \Rightarrow \text{Fix } T = (\text{zer } A) \cap (\text{zer } B) \). \hspace{1cm} (5.30)

\textit{Proof.} (i): Combine Corollary 4.41 and Corollary 5.16(v). (ii): Combine Corollary 5.16(v) with Theorem 4.32(i)&(iii). □

In view of (5.29) and Theorem 5.21(i), when \( A \) and \( B \) are paramonotone, we have the implication \( 0 \in K \Rightarrow J_A P_{\text{Fix } T} = P_Z \). However the converse implication is not true, as we show in the next example.

\textbf{Example 5.22.} Suppose that \( a \in X \setminus \{0\} \), that \( A = \text{Id} - 2a \) and that \( B = \text{Id} \). Then \( Z = \{a\} \), \( (A^{-1}, B^{-1}) = (\text{Id} + 2a, \text{Id}) \), hence \( K = \{-a\} \), \( Z \setminus Z = \{0\} \) and therefore \( K \perp (Z \setminus Z) \) which implies that \( J_A P_{\text{Fix } T} = P_Z \) by Theorem 5.19, but \( 0 \not\in K \).

If neither \( A \) nor \( B \) is single-valued, then the conclusion of Theorem 5.21(ii) may fail as we now illustrate.

\textbf{Example 5.23.} Suppose that \( X = \mathbb{R}^2 \), that \( U = \mathbb{R} \times \{0\} \), that \( V = \text{ball}((0,1);1) \), that \( A = N_U \) and that \( B = N_V \). By Example 4.7, \( Z = U \cap V = \{(0,0)\} \) and \( K = N_{T^*T}(0) = \mathbb{R}_+ \cdot (0,1) \neq \{(0,0)\} \). Therefore \( \text{Fix } T = \mathbb{R}_+ \cdot (0,1) \neq \{(0,0)\} = U \cap V = \text{zer } A \cap \text{zer } B \).

Working in \( X \times X \), we recall that (see [12, Proposition 23.16])
\[ A \times B \text{ is maximally monotone } \text{ and } J_{A \times B} = J_A \times J_B. \hspace{1cm} (5.31) \]

\textbf{Corollary 5.24.} Let \( x \in X \) and let \( (z,k) \in S \). Then
\[
\|(J_A T x, J_{A^{-1}} T x) - (z,k)\|^2 = \|J_A T x - z\|^2 + \|J_{A^{-1}} T x - k\|^2 \hspace{1cm} (5.32a)
\leq \|J_A x - z\|^2 + \|J_{A^{-1}} x - k\|^2 \hspace{1cm} (5.32b)
= \|(J_A x, J_{A^{-1}} x) - (z,k)\|^2. \hspace{1cm} (5.32c)
\]
5.5. The Douglas–Rachford operator and solution sets

Proof. It follows from Theorem 5.14 that \( z + k \in \text{Fix } T \), \( J_A(z + k) = z \) and \( J_{A^{-1}}(z + k) = k \). Now combine with Theorem 5.9(v) with \( y \) replaced by \( z + k \). ■

We recall, as consequence of Fact 3.7 and Example 3.24(i), that when \( X = \mathbb{R} \), the operators \( A \) and \( B \) are paramonotone. Using Corollary 5.16(iv), we then have \( S = Z \times K \).

Lemma 5.25. Suppose that \( X = \mathbb{R} \). Let \( x \in X \) and let \( (z,k) \in Z \times K \). Then the following hold:

(i) \( |J_A Tx - z|^2 \leq |J_Ax - z|^2 \).

(ii) \( |J_{A^{-1}} Tx - k|^2 \leq |J_{A^{-1}}x - k|^2 \).

Proof. (i): Set
\[
q(x,z) = |J_A Tx - z|^2 - |J_Ax - z|^2. \tag{5.33}
\]
If \( x \in \text{Fix } T \) we get \( q(x,z) = 0 \). Suppose that \( x \in \mathbb{R} \setminus \text{Fix } T \). Since \( T \) is firmly nonexpansive we have that \( \text{Id} - T \) is firmly nonexpansive (see Fact 3.2(ii)), hence monotone by Fact 3.4. Therefore \( (\forall x \in \mathbb{R} \setminus \text{Fix } T)(\forall f \in \text{Fix } T) \) we have
\[
(x - Tx)(x - f) = ((\text{Id} - T)x - (\text{Id} - T)f)(x - f) > 0. \tag{5.34}
\]
Notice that (5.33) can be rewritten as
\[
q(x,z) = (J_A Tx - J_Ax)((J_A Tx - z) + (J_Ax - z)). \tag{5.35}
\]
We argue by cases.

Case 1: \( x < Tx \).
It follows from (5.34) that
\[
(\forall f \in \text{Fix } T) \ x < f. \tag{5.36}
\]
On the one hand, since \( J_A \) is firmly nonexpansive, we have \( J_A \) is monotone and therefore \( J_A Tx - J_Ax \geq 0 \). On the other hand, it follows from Fact 5.2(iii) that \( (\exists f \in \text{Fix } T) \) such that \( z = J_Af = J_ATf \). Using (5.36) and the fact that \( J_A \) is monotone we conclude that \( J_Ax - z = J_Ax - J_Af \leq 0 \). Moreover, since \( J_A \) and \( T \) are firmly nonexpansive operators on \( \mathbb{R} \), we have \( J_A \circ T \) is firmly nonexpansive hence monotone and therefore (5.36) implies that \( J_A Tx - z = J_A Tx - J_A Tf \leq 0 \). Combining with (5.35) we conclude that (i) holds.
5.5. The Douglas–Rachford operator and solution sets

Case 2: \( x > T x \).
The proof follows similar to Case 1.

(ii) Apply the results of (i) to \( A^{-1} \) and use (3.15). ■

In view of (5.24) one might conjecture that Corollary 5.24 holds when we replace \( S \) by \( Z \times K \). The following example gives a negative answer to this conjecture. It also illustrates that when \( X \neq \mathbb{R} \), the conclusion of Lemma 5.25 could fail.

Example 5.26. Suppose that \( X = \mathbb{R}^2 \), that \( A \) is the normal cone operator of \( \mathbb{R}^2_+ \), and that \( B : X \to X : (x_1,x_2) \mapsto (-x_2,x_1) \) is the rotator by \( \pi/2 \). Then Fix \( T = \mathbb{R}_+ \cdot (1,-1) \), \( Z = \mathbb{R}_+ \times \{0\} \) and \( K = \{0\} \times \mathbb{R}_- \). Moreover, \( (\exists x \in \mathbb{R}^2) (\exists (z,k) \in Z \times K) \) such that

\[
\| (J_A Tx, J_{A^{-1}} Tx) - (z,k) \|^2 - \| (J_A x, J_{A^{-1}} x) - (z,k) \|^2 > 0
\]

and \( \| J_A Tx - z \|^2 - \| J_A x - z \|^2 > 0 \).

Proof. Let \((x_1,x_2) \in \mathbb{R}^2 \). Using Example 3.17 we have

\[
J_B(x_1,x_2) = \left( \frac{1}{2}(x_1 + x_2), \frac{1}{2}(-x_1 + x_2) \right)
\]

and

\[
R_B(x_1,x_2) = (x_2,-x_1) = -B(x_1,x_2).
\]

Hence, \( R_B^{-1} = (-B)^{-1} = B \). Using (5.2) we conclude that \((x_1,x_2) \in \text{Fix } T \iff (x_1,x_2) \in \text{Fix } R_B R_A \). Hence

\[
\text{Fix } T = \{ (x_1,x_2) \in \mathbb{R}^2 \mid (x_1,x_2) = R_B R_A(x_1,x_2) \} \tag{5.39a}
\]

\[
= \{ (x_1,x_2) \in \mathbb{R}^2 \mid R_B^{-1}(x_1,x_2) = 2J_A(x_1,x_2) - (x_1,x_2) \} \tag{5.39b}
\]

\[
= \{ (x_1,x_2) \in \mathbb{R}^2 \mid B(x_1,x_2) + (x_1,x_2) = 2P_{R_B^2}(x_1,x_2) \} \tag{5.39c}
\]

\[
= \{ (x_1,x_2) \in \mathbb{R}^2 \mid (x_1 - x_2, x_1 + x_2) = 2P_{R_B^2}(x_1,x_2) \}. \tag{5.39d}
\]

We argue by cases.

Case 1: \( x_1 \geq 0 \) and \( x_2 \geq 0 \). Then \((x_1,x_2) \in \text{Fix } T \iff (x_1 - x_2, x_1 + x_2) = 2P_{R_B^2}(x_1,x_2) = (2x_1,2x_2) \iff x_1 = -x_2 \) and \( x_1 = x_2 \iff x_1 = x_2 = 0 \).

Case 2: \( x_1 < 0 \) and \( x_2 < 0 \). Then \((x_1,x_2) \in \text{Fix } T \iff (x_1 - x_2, x_1 + x_2) = 2P_{R_B^2}(x_1,x_2) = (0,0) \iff x_1 = x_2 \) and \( x_1 = -x_2 \iff x_1 = x_2 = 0 \), which contradicts that \( x_1 < 0 \) and \( x_2 < 0 \).

Case 3: \( x_1 \geq 0 \) and \( x_2 < 0 \). Then \((x_1,x_2) \in \text{Fix } T \iff (x_1 - x_2, x_1 + x_2) = 2P_{R_B^2}(x_1,x_2) = (2x_1,0) \iff x_1 = -x_2 \).
Lemma 5.28. Suppose that $U$ is a linear subspace of $X$ and that $A = P_U$. Then $A$ is maximally monotone,

\[ J_A = J_{P_U} = \frac{1}{2}(\text{Id} + P_{U^\perp}), \quad \text{and} \quad R_A = P_{U^\perp} = \text{Id} - A. \]  

5.5. The Douglas–Rachford operator and solution sets

Case 4: $x_1 < 0$ and $x_2 \geq 0$. Then $(x_1, x_2) \in \text{Fix } T \iff (x_1 - x_2, x_1 + x_2) = 2P_{R^2}(x_1, x_2) = (0, 2x_2) \iff x_1 = x_2$, which never occurs since $x_1 < 0$ and $x_2 \geq 0$. Altogether we conclude that $\text{Fix } T = R_+ \cdot (1, -1)$, as claimed.

Using Fact 5.2(iii)&(iv) we have $Z = J_A(\text{Fix } T) = R_+ \times \{0\}$, and $K = J_{A^\perp}(\text{Fix } T) = (\text{Id} - J_A)(\text{Fix } T) = P_{R^2}(\text{Fix } T) = \{0\} \times R_-.$

Now let $a > 0$, let $x = (a, 0)$, set $z = (2a, 0) \in Z$ and set $k = (0, -a) \in K$. Notice that $Tx = x - P_{R^2} + \frac{1}{2}(\text{Id} - B)x = (a, 0) - (a, 0) + \frac{1}{2}((a, 0) - (0, a)) = (\frac{1}{2}a, -\frac{1}{2}a)$. Hence, $J_Ax = P_{R^2}(a, 0) = (a, 0)$, $J_{A^\perp}x = P_{R^2}(a, 0) = (0, 0)$, $J_ATx = P_{R^2}(\frac{1}{2}a, -\frac{1}{2}a) = (\frac{1}{2}a, 0)$, and $J_{A^\perp}x = P_{R^2}(\frac{1}{2}a, -\frac{1}{2}a) = (0, -\frac{1}{2}a)$. Therefore

\[
\| (J_ATx, J_{A^\perp}x) - (z, k) \|^2 - \| (J_Ax, J_{A^\perp}x) - (z, k) \|^2 = \| J_ATx - z \|^2 + \| J_{A^\perp}x - k \|^2 - \| J_Ax - z \|^2 - \| J_{A^\perp}x - k \|^2
\]
\[
= \| (\frac{1}{2}a, 0) - (2a, 0) \|^2 + \| (0, -\frac{1}{2}a) - (0, -a) \|^2
\]
\[
- \| (a, 0) - (2a, 0) \|^2 - \| (0, 0) - (0, -a) \|^2
\]
\[
= \frac{9}{4}a^2 + \frac{1}{4}a^2 - a^2 - a^2 = \frac{5}{4}a^2 > 0.
\]

Similarly one can verify that $\| J_ATx - z \|^2 - \| J_Ax - z \|^2 = \frac{5}{4}a^2 > 0$. ■

Throughout the rest of this section, we assume that

\[ A : X \rightrightarrows X \text{ and } B : X \rightrightarrows X \text{ are maximally monotone linear relations; } \]
equivalently, by [27, Theorem 2.1(xviii)], that

\[ J_A \text{ and } J_B \text{ are linear operators from } X \text{ to } X. \] (5.40)

This additional assumption leads to stronger conclusions.

Lemma 5.27. \( \text{Id} - T = J_A - 2J_BJ_A + J_B. \)

Proof: Let $x \in X$. Then indeed $x - Tx = J_Ax - J_BR_Ax = J_Ax - J_B(2J_Ax - x) = J_Ax - 2J_BJ_Ax + J_Bx$. ■

Lemma 5.28. Suppose that $U$ is a linear subspace of $X$ and that $A = P_U$. Then $A$ is maximally monotone,

\[ J_A = J_{P_U} = \frac{1}{2}(\text{Id} + P_{U^\perp}), \quad \text{and} \quad R_A = P_{U^\perp} = \text{Id} - A. \] (5.41)
5.5. The Douglas–Rachford operator and solution sets

Proof. Let \((x, y) \in X \times X\). Then

\[ y = J_A x \iff x = y + P_U y. \tag{5.42} \]

Now assume \(y = J_A x\). Since \(P_U\) is linear, (5.42) implies that
\[ P_{U^\perp} x = P_{U^\perp} y. \]
Moreover, \(y = x - P_U y = \frac{1}{2}(x + x - 2P_U y) = \frac{1}{2}(x + y + P_U y - 2P_U y) = \frac{1}{2}(x + y - P_U y) = \frac{1}{2}(x + P_{U^\perp} y) = \frac{1}{2}(x + P_{U^\perp} x). \]

Next, \(R_A = 2I_A - \text{Id} = (\text{Id} + P_{U^\perp}) - \text{Id} = P_{U^\perp}. \)

We say that a linear relation \(A\) is skew (see [25]) if \((\forall (a, a^*) \in \text{gra } A)\)
\(\langle a, a^* \rangle = 0. \)

Lemma 5.29. Suppose that \(A: X \to X\) and \(B : X \to X\) are both skew, and \(A^2 = B^2 = -\text{Id}.\) Then \(\text{Id} - T = \frac{1}{2}(\text{Id} - BA).\)

Proof. It follows from Proposition 3.16 that \(R_A = A\) and \(R_B = B.\) Therefore (5.2) implies that \(\text{Id} - T = \frac{1}{2}(\text{Id} - R_B R_A) = \frac{1}{2}(\text{Id} - BA). \)

Example 5.30. Suppose that \(A\) and \(B\) are skew. Let \(x \in X\) and let \(y \in X.\) Then the following hold:

(i) \(\langle Tx - Ty, x - y \rangle = \|Tx - Ty\|^2.\)
(ii) \(\langle (\text{Id} - T)x - (\text{Id} - T)y, x - y \rangle = \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.\)
(iii) \(\|x - y\|^2 = \|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.\)
(iv) \(\|J_A x - J_A y\|^2 + \|J_{A^\perp} x - J_{A^\perp} y\|^2 = \|J_A Tx - J_A Ty\|^2 \]
\(- \|J_{A^\perp} Tx - J_{A^\perp} Ty\|^2 = \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.\)
(v) \(\|x\|^2 = \|Tx\|^2 + \|x - Tx\|^2.\)
(vi) \(\langle Tx, x - Tx \rangle = 0.\)

Proof. (i)–(iv): Apply Theorem 5.9, and use (3.20) as well as the skewness of \(A\) and \(B.\)
(v): Apply (iii) with \(y = 0.\)
(vi): We have \(2\langle Tx, x - Tx \rangle = \|x\|^2 - \|Tx\|^2 - \|x - Tx\|^2.\) Now apply (v).

Suppose that \(U\) is a closed affine subspace of \(X.\) One can easily verify that
\[
(\forall x \in X) (\forall y \in X) \quad \langle P_U x - P_U y, (\text{Id} - P_U)x - (\text{Id} - P_U)y \rangle = 0. \tag{5.43}
\]
5.5. The Douglas–Rachford operator and solution sets

Example 5.31. Suppose that $U$ and $V$ are closed affine subspaces of $X$ such that $U \cap V \neq \emptyset$, that $A = N_U$, and that $B = N_V$. Let $x \in X$, and let $(z,k) \in Z \times K).$ Then

\[
\| (P_U x, (\text{Id} - P_U) x) - (z,k) \|^2 - \| (P_U T x, (\text{Id} - P_U) T x) - (z,k) \|^2 = \| x - (z+k) \|^2 - \| T x - (z+k) \|^2 \quad (5.44a)
\]

\[
= \| P_U x - z \|^2 + \| \text{Id} - P_U \| x - k \|^2 \quad (5.44b)
\]

\[
= \| P_U x - P_U(z+k) \|^2 + \| \text{Id} - P_U \| x - (\text{Id} - P_U)(z+k) \|^2 + 2 \langle P_U x - P_U(z+k), \text{Id} - P_U \| x - (\text{Id} - P_U)(z+k) \rangle \quad (5.44c)
\]

\[
= \| P_U x - P_U(z+k) + \text{Id} - P_U \| x - (\text{Id} - P_U)(z+k) \|^2 \quad (5.44d)
\]

\[
= \| x - (z+k) \|^2. \quad (5.44e)
\]

Applying (5.46) with $x$ replaced by $T x$ yields

\[
\| (P_U T x, (\text{Id} - P_U) T x) - (z,k) \|^2 = \| T x - (z+k) \|^2. \quad (5.47)
\]

Combining (5.46) and (5.47) yields (5.44b). It follows from (5.43) and Theorem 5.9(iii) applied with $(A,B,Y)$ replaced by $(N_U, N_V, z+k)$ that

\[
\| x - (z+k) \|^2 - \| T x - T(z+k) \|^2 = \| x - T x - ((z+k) - T(z+k)) \|^2, \quad (5.48)
\]

which in view of (5.45), proves (5.44c).

Now we turn to (5.44d). Let $w \in U \cap V$. Then $U = w + \text{par } U$ and $V = w + \text{par } V$. Suppose momentarily that $w = 0$. In this case, $\text{par } U = U$ and $\text{par } V = V$. Using [30, Proposition 3.4(i)], we have

\[
T = T_{(U,V)} = P_V P_U + P_U P_{U^\perp}. \quad (5.49)
\]
Therefore
\[ x - Tx = P_U x + P_U \perp x - P_V P_U x - P_V \perp P_U x \]
\[ = (\text{Id} - P_V)P_U x + (\text{Id} - P_V \perp)P_U x \]
\[ = P_V \perp P_U x + P_V P_U \perp x. \] (5.50a)

Using (5.50c) we have
\[ \|x - Tx\|^2 = \|P_V \perp P_U x + P_V P_U \perp x\|^2 \]
\[ = \|P_U x - P_V P_U x + P_V x - P_V P_U x\|^2 \]
\[ = \|P_U x\|^2 + \|P_V x\|^2 + 4\|P_V P_U x\|^2 \]
\[ + 2\langle P_U x, P_V x\rangle - 4\langle P_U x, P_V P_U x\rangle - 4\langle P_V x, P_V P_U x\rangle \]
\[ = \|P_U x\|^2 + \|P_V x\|^2 - 2\langle P_U x, P_V x\rangle = \|P_U x - P_V x\|^2. \] (5.51e)

Now, if \( w \neq 0 \), by Proposition 9.11 below we have \( Tx = w + T_{(\text{par} U, \text{par} V)}(x - w) \). Therefore, (5.51) yields
\[ \|x - Tx\|^2 = \|(x - w) - T_{(\text{par} U, \text{par} V)}(x - w)\|^2 \]
\[ = \|P_{\text{par} U}(x - w) - P_{\text{par} V}(x - w)\|^2 \]
\[ = \|w + P_{\text{par} U}(x - w) - (w + P_{\text{par} V}(x - w))\|^2 \]
\[ = \|P_U x - P_V x\|^2, \] (5.52d)
where the last equality follows from Fact 2.6.

5.6 From PDEs to maximally monotone operators

In this section we briefly show the connection between the original Douglas–Rachford algorithm introduced in [70] (see also [69], [103] and [117] for variations of this method) to solve certain types of heat equations and the general algorithm introduced by Lions and Mercier in [96]. At times, we follow in our presentation [61] which also contains additional historical comments.

Suppose that \( \Omega \) is a bounded square region in \( \mathbb{R}^2 \). Consider the Dirichlet problem for the Poisson equation: Given \( f \) and \( g \), find \( u : \Omega \to \mathbb{R} \) such that
\[ \Delta u = f \text{ on } \Omega \text{ and } u = g \text{ on } \text{bdry } \Omega, \] (5.53)
5.6. From PDEs to maximally monotone operators

where \( \Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \) is the Laplace operator and \( \text{bdry} \Omega \) denotes the boundary of \( \Omega \). Discretizing \( u \) followed by converting it into a long vector \( y \) (see [102, Example 7.6.2 & Problem 7.6.9]) we obtain the system of linear equations

\[
L_{\rightarrow} y + L_{\uparrow} y = -b. \tag{5.54}
\]

Here \( L_{\rightarrow} \) and \( L_{\uparrow} \) denote the horizontal (respectively vertical) positive definite discretization of the negative Laplacian over a square mesh with \( n^2 \) points at equally spaced intervals (see [102, Problem 7.6.10]). We have

\[
L_{\rightarrow} = \text{Id} \otimes M \quad \text{and} \quad L_{\uparrow} = M \otimes \text{Id}, \tag{5.55}
\]

where

\[
M = \begin{pmatrix}
2 & -1 & 0 \\
-1 & \ddots & \ddots \\
\ddots & \ddots & -1 \\
0 & -1 & 2
\end{pmatrix} \in \mathbb{R}^{n \times n}. \tag{5.56}
\]

To see the connection to monotone operators, set \( A = L_{\rightarrow} \) and \( B : L_{\uparrow} + b : y \mapsto L_{\uparrow} y + b \). Then \( A \) and \( B \) are affine and strictly monotone. The problem then reduces to

\[
\text{find } y \in \mathbb{R}^{n^2} \text{ such that } Ay + By = 0, \tag{5.57}
\]

and the algorithm proposed by Douglas and Rachford in [70] becomes

\[
y_{n+1/2} + Ay_n + By_{n+1/2} - y_n = 0, \tag{5.58a}
\]

\[
y_{n+1} - y_{n+1/2} - Ay_n + Ay_{n+1} = 0. \tag{5.58b}
\]

Consequently,

\[
(5.58a) \iff (\text{Id} + B)(y_{n+1/2}) = (\text{Id} - A)y_n \iff y_{n+1/2} = J_B(\text{Id} - A)y_n, \tag{5.59a}
\]

\[
(5.58b) \iff (\text{Id} + A)y_{n+1} = Ay_n + y_{n+1/2} \iff y_{n+1} = J_A(Ay_n + y_{n+1/2}). \tag{5.59b}
\]

Substituting (5.59a) into (5.59b) to eliminate \( y_{n+1/2} \) yields

\[
y_{n+1} = J_A(Ay_n + J_B(\text{Id} - A)y_n). \tag{5.60}
\]

To proceed further, we must show that

\[
(\text{Id} - A)J_A = R_A \tag{5.61a}
\]
5.6. From PDEs to maximally monotone operators

\[ AJ_A = \text{Id} - J_A. \] (5.61b)

Indeed, note that \( \text{Id} - A = 2 \text{Id} - (\text{Id} + A) \), therefore multiplying by \( J_A = (\text{Id} + A)^{-1} \) from the right yields \( (\text{Id} - A)J_A = (2 \text{Id} - (\text{Id} + A))J_A = 2J_A - \text{Id} = R_A \). Hence \( J_A - AJ_A = J_A - (\text{Id} - J_A) \); equivalently, \( AJ_A = \text{Id} - J_A \).

Now consider the change of variable \( (\forall n \in \mathbb{N}) \quad x_n = (\text{Id} + A)y_n \), (5.62)

which is equivalent to \( y_n = J_Ax_n \). Substituting (5.60) into (5.62), and using (5.61), yield

\[
\begin{align*}
x_{n+1} &= (\text{Id} + A)y_{n+1} = (\text{Id} + A)(Ay_n + J_B(\text{Id} - A)y_n) \\
&= Ay_n + J_B(\text{Id} - A)y_n \\
&= AJ_Ax_n + J_B(\text{Id} - A)J_Ax_n \\
&= x_n - J_Ax_n + J_BR_Ax_n \\
&= (\text{Id} - J_A + J_BR_A)x_n,
\end{align*}
\]

which is the Douglas–Rachford update formula (5.2).

We point out that \( J_A = J_{L^*} \), and using [12, Proposition 23.15(ii)] we have \( J_B = J_{L^*}J_{L^*} = J_{L^*} - J_{L^*}b \). To calculate \( J_A \) and \( J_B \) apply Corollary 3.49 to get

\[
J_A = \text{Id}_n \otimes J_M \quad \text{and} \quad J_B = J_M \otimes \text{Id}_n - (J_M \otimes \text{Id}_n)(b). \] (5.64)

For instance, when \( n = 3 \), the above calculations yield

\[
J_M = \begin{pmatrix}
\frac{8}{21} & \frac{1}{7} & \frac{1}{8} \\
\frac{1}{7} & \frac{3}{7} & \frac{1}{7} \\
\frac{1}{8} & \frac{1}{7} & \frac{21}{7}
\end{pmatrix}, \quad \text{(5.65)}
\]

\[
\begin{pmatrix}
\frac{8}{21} & \frac{1}{7} & \frac{1}{8} \\
\frac{1}{7} & \frac{3}{7} & \frac{1}{7} \\
\frac{1}{8} & \frac{1}{7} & \frac{21}{7}
\end{pmatrix}
\]

\[
\text{Id}_3 \otimes J_M = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}, \quad \text{(5.66)}
\]

64
5.7 Why do we need to work with monotone operators and not just subdifferential operators?

As shown earlier, the Douglas–Rachford operator $T$ is firmly nonexpansive (see Fact 5.2(i)) which, in view of Fact 3.12, asserts that $T = J_C$, i.e., $T$ is a resolvent of some maximally monotone operator $C$. Hence the Douglas–Rachford algorithm is a particular instance of the proximal point method, where in the latter $C$ is a subdifferential operator of some $f \in \Gamma(X)$. (The results by Rockafellar [124] also allow for additional parameters and summable errors.) Interestingly, in [71] Eckstein demonstrated that even if $A$ and $B$ are subdifferential operators whose resolvents in this case are proximal mappings (see Example 3.14) the corresponding Douglas–Rachford operator is just a resolvent that is not a proximal mapping (see also [40] for detailed discussion and examples of this case). For the sake of completeness we include the following example that appears in [126].

Example 5.32. Suppose that $X = \mathbb{R}^2$, that $A = N_{\mathbb{R}(1,0)}$ and that $B = N_{\mathbb{R}(1,1)}$. Simple calculations yield

$$T = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

which implies that $T = J_C$ where

$$C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$ 

Note that $T$ is not symmetric, and therefore is not a proximal mapping by [40, Corollary 2.5].
5.7. Why do we need to work with monotone operators and not just subdifferential operators?

Therefore, even if we are interested in solving only optimization problems, where the resolvents of the operators (which in this case are subdifferential operators) are proximal mappings, the corresponding Douglas–Rachford splitting operator could possibly be just a resolvent of a monotone operator which is not a proximal mapping. Therefore, in view of Fact 3.12, we do actually need to work with general monotone operators and not just subdifferential operators.
Chapter 6

On Fejér monotone sequences and nonexpansive mappings

6.1 Overview

Let $C$ be a nonempty closed convex subset of $X$. A sequence $(x_n)_{n \in \mathbb{N}}$ in $X$ is called Fejér monotone (see [11], [59] and [60]) with respect to $C$ if

$$\forall c \in C \forall n \in \mathbb{N} \quad \|x_{n+1} - c\| \leq \|x_n - c\|. \quad (6.1)$$

In other words, each point in a Fejér monotone sequence is not further from any point in $C$ than its predecessor.

The notion of Fejér monotonicity has proven to be a fruitful concept in fixed point theory and optimization. It has shown to be an efficient tool to analyze various iterative algorithms in convex optimization.

We summarize our main results in this chapter as follows:

- We present new conditions sufficient for convergence of Fejér monotone sequences (see Lemma 6.3 and Theorem 6.11) and for the convergence of the sequence (see Lemma 6.4 and Proposition 6.5).

- We provide a mild generalization of Ostrowski’s Theorem [113, Theorem 26.1] (see Lemma 6.8).

Except for facts with explicit references, this chapter is based on results that appear in [36], [13], and [33].

6.2 Fejér monotonicity: New principles

We start by recalling some pleasant properties of Fejér monotone sequences.

**Fact 6.1.** Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $X$ that is Fejér monotone with respect to a nonempty closed convex subset $C$ of $X$. Then the following hold:
6.2. Fejér monotonicity: New principles

(i) The sequence \((x_n)_{n \in \mathbb{N}}\) is bounded.

(ii) For every \(c \in C\), the sequence \((\|x_n - c\|)_{n \in \mathbb{N}}\) converges.

(iii) The set of strong cluster points of \((x_n)_{n \in \mathbb{N}}\) lies in a sphere of \(X\).

(iv) The sequence \((P_C x_n)_{n \in \mathbb{N}}\) converges strongly to a point in \(C\).

(v) If \(\text{int } C \neq \emptyset\), then \((x_n)_{n \in \mathbb{N}}\) converges strongly to a point in \(X\).

(vi) If \(C\) is a closed affine subspace of \(X\), then \((\forall n \in \mathbb{N}) P_C x_n = P_C x_0\).

(vii) Every weak cluster point of \((x_n)_{n \in \mathbb{N}}\) that belongs to \(C\) must be \(\lim_{n \to \infty} P_C x_n\).

(viii) The sequence \((x_n)_{n \in \mathbb{N}}\) converges weakly to some point in \(C\) if and only if all weak cluster points of \((x_n)_{n \in \mathbb{N}}\) lie in \(C\).

(ix) If all weak cluster points of \((x_n)_{n \in \mathbb{N}}\) lie in \(C\), then \((x_n)_{n \in \mathbb{N}}\) converges weakly to \(\lim_{n \to \infty} P_C x_n\).

Proof. (i)&(ii): [12, Proposition 5.4]. (iii): Clear from (ii). (iv): [12, Proposition 5.7]. (v): [12, Proposition 5.10]. (vi): [12, Proposition 5.9(i)]. (vii): This follows from [12, Corollary 5.11]. (viii): This follows from [12, Theorem 5.5]. (ix): Combine (viii) with (vii). \(\blacksquare\)

The following result was first presented in [7, Theorem 6.2.2(ii)]; for completeness, we include its short proof.

Lemma 6.2. Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\) that is Fejér monotone with respect to a nonempty closed convex subset \(C\) of \(X\). Let \(w_1\) and \(w_2\) be weak cluster points of \((x_n)_{n \in \mathbb{N}}\). Then \(w_1 - w_2 \in (C - C)^\perp\).

Proof. Let \((c_1, c_2) \in C \times C\). Using Fact 6.1(ii), set \(L_i = \lim_{n \to \infty} \|x_n - c_i\|\), for \(i \in \{1, 2\}\). Note that

\[\|x_n - c_1\|^2 = \|x_n - c_2\|^2 + \|c_1 - c_2\|^2 + 2\langle x_n - c_2, c_2 - c_1 \rangle.\] (6.2)

Now suppose that \(x_{k_n} \to w_1\) and \(x_{l_n} \to w_2\). Taking the limit in (6.2) along the two subsequences \((k_n)_{n \in \mathbb{N}}\) and \((l_n)_{n \in \mathbb{N}}\) yields \(L_1 = L_2 + \|c_2 - c_1\|^2 + 2\langle w_1 - c_2, c_2 - c_1 \rangle\) and \(L_1 = L_2 + \|c_2 - c_1\|^2 + 2\langle w_2 - c_2, c_2 - c_1 \rangle\). Subtracting the last two equations yields \(2\langle c_2 - c_1, w_1 - w_2 \rangle = 0\). \(\blacksquare\)

The next novel result on Fejér monotone sequences is of critical importance in our analysis. (When \((u_n)_{n \in \mathbb{N}} = (x_n)_{n \in \mathbb{N}}\) one obtains a well-known result; see [12, Theorem 5.5].)
Proof. Then \((\forall e \in E) \langle u_n - e, u_n - x_n \rangle \to 0. \) (6.3)

Then \((u_n)_{n \in \mathbb{N}}\) converges weakly to some point in \(E.\)

**Proof.** It follows from (6.3) that \((\forall (e_1, e_2) \in E \times E) \)

\[
\langle e_2 - e_1, u_n - x_n \rangle = \langle u_n - e_1, u_n - x_n \rangle - \langle u_n - e_2, u_n - x_n \rangle \to 0. \quad (6.4)
\]

Next, let us assume that \(e_1\) and \(e_2\) are two (possibly different) weak cluster points of \((u_n)_{n \in \mathbb{N}}.\) Obtain four subsequences \((x_{k_n})_{n \in \mathbb{N}}, (x_{l_n})_{n \in \mathbb{N}}, (u_{k_n})_{n \in \mathbb{N}}\)

and \((u_{l_n})_{n \in \mathbb{N}}\) such that \(x_{k_n} \to y_1, x_{l_n} \to y_2, u_{k_n} \to e_1\) and \(u_{l_n} \to e_2.\) Taking the limit in (6.4) along these subsequences we have \(\langle e_2 - e_1, e_1 - y_1 \rangle = 0 = \langle e_2 - e_1, e_2 - y_2 \rangle,\) hence

\[
\|e_2 - e_1\|^2 = \langle e_2 - e_1, y_2 - y_1 \rangle. \quad (6.5)
\]

Since \(\{e_1, e_2\} \subseteq E,\) we conclude, in view of Lemma 6.2 that \(\langle e_2 - e_1, y_2 - y_1 \rangle = 0.\) By (6.5), \(e_1 = e_2.\)

**Lemma 6.4.** Let \(A\) be a closed linear subspace of \(X,\) let \(C\) be a nonempty closed convex subset of \(A,\) and let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X.\) Suppose that \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(C,\) i.e., \((\forall n \in \mathbb{N}) (\forall c \in C) \|x_{n+1} - c\| \leq \|x_n - c\|,\) and that all the weak cluster points of \((P_A x_n)_{n \in \mathbb{N}}\) lie in \(C.\) Then \((P_A x_n)_{n \in \mathbb{N}}\) converges weakly to some point in \(C.\)

**Proof.** Since \((x_n)_{n \in \mathbb{N}}\) is bounded (by Fact 6.1(ii)) and \(P_A\) is (firmly) non-expansive we learn that \((P_A x_n)_{n \in \mathbb{N}}\) is bounded and by assumption, its weak cluster points lie in \(C \subseteq A.\) Now let \(c_1\) and \(c_2\) be in \(C.\) On the one hand, the Fejér monotonicity of \((x_n)_{n \in \mathbb{N}}\) implies the convergence of the sequences \((\|x_n - c_1\|^2)_{n \in \mathbb{N}}\) and \((\|x_n - c_2\|^2)_{n \in \mathbb{N}}\) by Fact 6.1(ii). On the other hand, expanding and simplifying yield \(\|x_n - c_1\|^2 - \|x_n - c_2\|^2 = \|x_n\|^2 + \|c_1\|^2 - 2\langle x_n, c_1 \rangle - \|x_n\|^2 - \|c_2\|^2 + 2\langle x_n, c_2 \rangle = \|c_1\|^2 - 2\langle x_n, c_1 - c_2 \rangle - \|c_2\|^2,\) which in turn implies that \((\langle x_n, c_1 - c_2 \rangle)_{n \in \mathbb{N}}\) converges. Since \(c_1 \in A\) and \(c_2 \in A\) we have

\[
\langle x_n, c_1 - c_2 \rangle = \langle x_n, P_A c_1 - P_A c_2 \rangle = \langle x_n, P_A (c_1 - c_2) \rangle = \langle P_A x_n, c_1 - c_2 \rangle. \quad (6.6)
\]
Now assume that \((P_Ax_n)_{n \in \mathbb{N}}\) and \((P_{A}\bar{x}_n)_{n \in \mathbb{N}}\) are subsequences of \((P_Ax_n)_{n \in \mathbb{N}}\) such that \(P_Ax_n \to c_1\) and \(P_{A}\bar{x}_n \to c_2\). By the uniqueness of the limit in (6.6) we conclude that \(\langle c_1, c_1 - c_2 \rangle = \langle c_2, c_1 - c_2 \rangle\) or equivalently \(\|c_1 - c_2\|^2 = 0\), hence \((P_Ax_n)_{n \in \mathbb{N}}\) has a unique weak cluster point which completes the proof. ■

The next result can be seen as a finite-dimensional variant of Lemma 6.4 (where \(A\) is a closed linear subspace) and Fact 6.1(viii) (where \(A = X\)).

**Proposition 6.5.** Suppose that \(X\) is finite-dimensional, let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\) that is Fejér monotone with respect to a nonempty closed convex subset \(C\) of \(X\), and let \(A\) be a closed convex subset of \(X\) such that \(C \subseteq A\). Suppose that all cluster points of \((P_Ax_n)_{n \in \mathbb{N}}\) lie in \(C\). Then \((P_Ax_n)_{n \in \mathbb{N}}\) converges; in fact,

\[
\lim_{n \to \infty} P_Ax_n = \lim_{n \to \infty} P_Cx_n. \tag{6.7}
\]

**Proof.** Set \(c^* = \lim_{n \to \infty} P_Cx_n\) (see Fact 6.1(iv)). By Fact 6.1(i), \((x_n)_{n \in \mathbb{N}}\) is bounded, hence so is \((P_Ax_n)_{n \in \mathbb{N}}\) because \(P_A\) is nonexpansive. Note that by assumption, all cluster points of \((P_Ax_n)_{n \in \mathbb{N}}\) lie in \(C\). Let \(c\) be an arbitrary cluster point of \((P_Ax_n)_{n \in \mathbb{N}}\). Then there exist a subsequence \((x_{k_n})_{n \in \mathbb{N}}\) of \((x_n)_{n \in \mathbb{N}}\) and a point \(x \in X\) such that \(x_{k_n} \to x\) and \(P_Ax_{k_n} \to P_Ax = c \in C\). It follows from [12, Proposition 28.5] that \(P_Cx_{k_n} \to P_Cx = P_Ax = c\). On the other hand, by assumption we have \(P_Cx_{k_n} \to c^*\). Altogether, \(c = c^*\) and the result follows. ■

Our next result decouples Fejér monotonicity into two properties in the case when the underlying set can be written as the sum of a set and a cone.

**Proposition 6.6.** Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\), let \(E\) be a nonempty subset of \(X\) and let \(K\) be a nonempty convex cone of \(X\). Then the following are equivalent:

(i) \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(E + K\).

(ii) \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(E\) and \((\forall n \in \mathbb{N})\ x_{n+1} \in x_n + K^\oplus\), where \(K^\oplus = \{u \in X : \inf \langle u, K \rangle = 0\}\).

**Proof.** Set

\[
(\forall x \in X) (\forall n \in \mathbb{N}) \quad \Delta_n(x) = \|x_n - x\|^2 - \|x_{n+1} - x\|^2. \tag{6.8}
\]

Then for every \(e \in E\) and \(k \in K\), we have

\[
\Delta_n(e + k) = \|x_n - e\|^2 + \|k\|^2 - 2 \langle x_n - e, k \rangle
\]
6.2. Fejér monotonicity: New principles

\[
\begin{align*}
- (\|x_{n+1} - e\|^2 + \|k\|^2 - 2 \langle x_{n+1} - e, k \rangle) &= \Delta_n(e) + 2 \langle x_{n+1} - x_n, k \rangle. \\
&= \Delta_n(e) + 2 \langle x_{n+1} - x_n, k \rangle. 
\end{align*}
\]

(6.9a) (6.9b)

Assume first that (i) holds. Then \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(E\) because \(E \subseteq E + K\). Let \((e, k) \in E \times K\) and \(n \in \mathbb{N}\). Using (6.8),

\[
0 \leq \Delta_n(e + k) = \Delta_n(e) + 2 \langle x_{n+1} - x_n, k \rangle.
\]

(6.10)

Since \(K\) is a cone, this shows that \(2 \inf \langle x_{n+1} - x_n, R_{++}k \rangle \geq -\Delta_n(e) > -\infty\). Hence \(\langle x_{n+1} - x_n, k \rangle \geq 0\). It follows that \(x_{n+1} - x_n \in K^0\). Conversely, if (ii) holds, then (6.9) immediately yields (i).

The following consequence of Proposition 6.6 shows that Proposition 6.6 is a generalization of Fact 6.1(vi).

**Corollary 6.7.** Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\), and let \(C\) be a closed affine subspace of \(X\), say \(C = c + Y\), where \(Y\) is a closed linear subspace of \(X\). Then \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(C\) if and only if (\(\forall n \in \mathbb{N}\)) \(\|x_{n+1} - c\| \leq \|x_n - c\|\) and \(x_{n+1} \in x_n + Y^\perp\), in which case \((P_C x_n)_{n \in \mathbb{N}}\) is a constant sequence.

We continue with the following lemma, which is a minor generalization of a theorem of Ostrowski (see [113, Theorem 26.1]) whose proof we follow.

**Lemma 6.8.** Let \((Y, d)\) be a metric space, and let \((x_n)_{n \in \mathbb{N}}\) be a sequence in a compact subset \(C\) of \(Y\) such that \(d(x_n, x_{n+1}) \to 0\). Then the set of cluster points of \((x_n)_{n \in \mathbb{N}}\) is a compact connected subset of \(C\).

**Proof.** Denote the set of cluster points of \((x_n)_{n \in \mathbb{N}}\) by \(S\) and assume to the contrary that \(S = A \cup B\) where \(A\) and \(B\) are nonempty closed subsets of \(X\) and \(A \cap B = \emptyset\). Then

\[
\delta = \inf_{(a, b) \in A \times B} d(a, b) > 0.
\]

(6.11)

By assumption on \((x_n)_{n \in \mathbb{N}}\), there exists \(n_0 \in \mathbb{N}\) such that \((\forall n \geq n_0)\) we have \(d(x_n, x_{n+1}) \leq \delta/3\). Let \(a \in A\). Then there exists \(m > n_0\) such that \(d(x_m, a) < \delta/3\). Because \((x_n)_{n > m}\) has a cluster point in \(B\), there exists a smallest integer \(k > m\) such that \(d(x_k, B) < 2\delta/3\). Then \(d(x_{k-1}, B) \geq 2\delta/3\) and hence \(d(x_k, B) \geq d(x_{k-1}, B) - d(x_{k-1}, x_k) \geq 2\delta/3 - \delta/3 = \delta/3\). Thus \(\delta/3 \leq d(x_k, B) < 2\delta/3\). Repeating this argument yields a subsequence \((x_{k_n})_{n \in \mathbb{N}}\) of \((x_n)_{n \in \mathbb{N}}\) such that \((\forall n \in \mathbb{N}\) \(\delta/3 \leq d(x_{k_n}, B) < 2\delta/3\). Let \(x\) be a cluster point of \((x_{k_n})_{n \in \mathbb{N}}\). It follows that

\[
\delta/3 \leq d(x, B) \leq 2\delta/3
\]

(6.12)
Obviously, \( x \notin B \). Hence \( x \in A \), and therefore (recall (6.11)) \( \delta \leq d(x, B) \leq 2\delta /3 < \delta \), which is absurd. 

To proceed to the next results we need to recall the following definition.

**Definition 6.9 (asymptotically regular sequence).** Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\). Then \((x_n)_{n \in \mathbb{N}}\) is asymptotically regular if \(x_n - x_{n+1} \to 0\).

An immediate consequence of Lemma 6.8 is the classical Ostrowski result.

**Corollary 6.10 (Ostrowski).** Suppose that \(X\) is finite-dimensional and let \((x_n)_{n \in \mathbb{N}}\) be a bounded sequence in \(X\) such that \((x_n)_{n \in \mathbb{N}}\) is asymptotically regular. Then the set of cluster points of \((x_n)_{n \in \mathbb{N}}\) is compact and connected.

We are now in position to prove the following key result which can be seen as a variant of Fact 6.1(v).

**Theorem 6.11 (a new sufficient condition for convergence).** Suppose that \(X\) is finite-dimensional and that \(C\) is a nonempty closed convex subset of \(X\) of co-dimension 1, i.e.,

\[
\text{codim } C = \text{codim}(\text{aff } C - \text{aff } C) = 1. \tag{6.13}
\]

Let \((x_n)_{n \in \mathbb{N}}\) be a sequence that is Fejér monotone with respect to \(C\) and asymptotically regular. Then \((x_n)_{n \in \mathbb{N}}\) is actually convergent.

**Proof.** By Fact 6.1(i), \((x_n)_{n \in \mathbb{N}}\) is bounded. Denote by \(S\) the set of cluster points of \((x_n)_{n \in \mathbb{N}}\). Since \(x_n - x_{n+1} \to 0\), Corollary 6.10 implies that \(S\) is connected. Moreover, \(S\) lies in a sphere of \(X\) due to Fact 6.1(iii). On the other hand, by combining Lemma 6.2 and (6.13), \(S\) lies in a line of \(X\). Altogether \(S\) is a connected subset of a sphere that lies on a line. We deduce that \(S\) is a singleton. 

We conclude with two examples illustrating that the assumptions on asymptotic regularity and co-dimension 1 are important.

**Example 6.12.** Suppose that \(X = \mathbb{R}^2\), set \(C = \{0\} \times \mathbb{R}\), and \((\forall n \in \mathbb{N})\) \(x_n = ((-1)^n, 0)\). Then \(\text{codim } C = 1\) and \((\forall c \in C) (\forall n \in \mathbb{N}) \|x_n - c\| = \|x_{n+1} - c\|\), hence \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(C\). However, \((x_n)_{n \in \mathbb{N}}\) does not converge. This does not contradict Theorem 6.11 because \(\|x_n - x_{n+1}\| = 2 \nrightarrow 0\).
Example 6.13. Suppose that $X = \mathbb{R}^2$, set $C = \{(0,0)\} \subseteq X$, and $(\forall n \in \mathbb{N}) \theta_n = \sum_{k=1}^{n} (1/k)$ and $x_n = \cos(\theta_n)(1,0) + \sin(\theta_n)(0,1)$. Then $(x_n)_{n \in \mathbb{N}}$ is asymptotically regular and Fejér monotone with respect to $C$. However, the set of cluster points of $(x_n)_{n \in \mathbb{N}}$ is the unit sphere because the harmonic series diverges. Again, this does not contradict Theorem 6.11 because $\text{codim} C = 2 \neq 1$. 
Chapter 7

Nonexpansive mappings and the minimal displacement vector

7.1 Overview
The goal of this chapter is to provide information on the connection between the inner and outer shifts of an operator. These results are heavily used in subsequent chapters. We summarize the main results of this chapter as follows:

- We provide a characterization of the existence of fixed points of affine operators (see Theorem 7.3).
- We explore the connection between the inner and outer shifts, by the minimal displacement vector (see Proposition 7.8), of a nonexpansive operator and the corresponding sets of fixed points (see Proposition 7.8).

Except for facts with explicit references, this chapter is based on results that appear in [31], [15] and [39].

7.2 Auxiliary results
We start with the following definition and auxiliary results.

Definition 7.1. Let \( w \in X \). For a single-valued operator \( T \) we define the inner shift and outer shift by \( w \) at \( x \in X \) by

\[
T_w x = T(x - w) \quad \text{and} \quad wT x = -w + T x,
\]

respectively.

Lemma 7.2. Let \( T : X \to X \) and let \( w \in X \). Then the following hold:
7.2. Auxiliary results

(i) $\text{Fix}(T - w) = -w + \text{Fix}(w + T) = -w + \text{Fix}(-wT)$.

(ii) $w \in \text{ran}(\text{Id} - T) \iff \text{Fix}(w + T) \neq \emptyset \iff \text{Fix}(T - w) \neq \emptyset$.

Proof. (i): Let $x \in X$. Then $x \in \text{Fix}(T - w) \iff x = T(x + w) \iff x + w = w + T(x + w) \iff x + w \in \text{Fix}(w + T) \iff x = -w + \text{Fix}(w + T)$.

(ii): $w \in \text{ran}(\text{Id} - T) \iff (\exists x \in X) \text{ such that } w = x - Tx \iff (\exists x \in X) \text{ such that } x = w + Tx \iff \text{Fix}(w + T) \neq \emptyset$. Now combine with (i). ■

Working with affine operators, we arrive at the following very useful result.

**Theorem 7.3.** Let $L : X \to X$ be linear, let $b \in X$, set $T : X \to X : x \mapsto Lx + b$. Then the following are equivalent:

(i) $\text{Fix } T \neq \emptyset$.

(ii) $b \in \text{ran}(\text{Id} - L)$.

(iii) There exists a point $a \in X$ such that $b = a - La$ and

$$(\forall x \in X) \quad Tx = L(x - a) + a. \quad (7.2)$$

(iv) There exists a point $a \in X$ such that $b = a - La$ and

$$\text{Fix } T = a + \text{Fix } L. \quad (7.3)$$

Moreover we have the implication:

$$\text{Fix } T \neq \emptyset \implies (\exists a \in X) \text{ such that } b = a - La \text{ and } \quad (\forall x \in X) \ P_{\text{Fix } T}x = a + P_{\text{Fix } L}(x - a) = P_{(\text{Fix } L)^{\perp}}a + P_{\text{Fix } L}x. \quad (7.4)$$

Proof. “(i)$\iff$(ii)”: $\text{Fix } T \neq \emptyset \iff (\exists y \in X) \ y = Ly + b \iff b \in \text{ran}(\text{Id} - L)$.

“(ii)$\implies$(iii)”: The existence of $a$ is a direct consequence of (ii) whereas (7.2) follows from the linearity of $L$. “(iii)$\implies$(iv)” : Let $y \in X$. Then $y \in \text{Fix } T \iff y - a \in \text{Fix } L \iff y \in a + \text{Fix } L$. “(iv)$\implies$(i)” : This is clear since $0 \in \text{Fix } L \neq \emptyset$. Now we turn to the implication (7.4). It follows from the equivalence of (i) and (iv) that (7.3) holds, which as well proves the existence of $a$. The first identity follows from combining (7.3) and Fact 2.6 applied with $(y, S)$ replaced by $(a, \text{Fix } L)$. It follows from the linearity of $P_{\text{Fix } L}$ (see Fact 2.7(iii)) that $a + P_{\text{Fix } L}(x - a) = a + P_{\text{Fix } L}x - P_{\text{Fix } L}a = P_{(\text{Fix } L)^{\perp}}a + P_{\text{Fix } L}x$. ■
### 7.3 The displacement map and the minimal displacement vector

In this section, we collect new results concerning nonexpansive and firmly nonexpansive operators whose fixed point sets could possibly be empty.

**Fact 7.4 (minimal displacement vector).** Let $T : X \to X$ be nonexpansive. Then $\text{ran}(\text{Id} - T)$ is convex; consequently, the minimal displacement vector $v = P_{\text{ran}(\text{Id} - T)}0$ is the unique element in $\text{ran}(\text{Id} - T)$ such that

$$\|v\| = \inf_{x \in X} \|x - Tx\|.$$  \hfill (7.5)

**Proof.** See [6], [53] or [116].  

The next example is readily verified.

**Example 7.5 (Fix$(v + T)$ vs. Fix$(T - v)$).** Let $C$ be a nonempty closed convex subset of $X$ and suppose that $T = \text{Id} - P_C$. Then $T$ is firmly nonexpansive and $v = P_C0$. Let $x \in X$. Then $x \in \text{Fix}(v + T) \Leftrightarrow P_Cx = v$, while $x \in \text{Fix}(T - v) \Leftrightarrow P_C(x + v) = v$.

**Lemma 7.6.** Let $T_1 : X \to X$ and $T_2 : X \to X$ be nonexpansive. Set $v_1 = P_{\text{ran}(\text{Id} - T_1T_2)}0$ and $v_2 = P_{\text{ran}(\text{Id} - T_2T_1)}0$. Then $\|v_1\| = \|v_2\|$.  

**Proof.** By definition of $v_1$, there exists a sequence $(x_n)_{n \in \mathbb{N}}$ in $X$ such that $\|x_n - T_1T_2x_n\| \to \|v_1\|$. Hence $(\forall n \in \mathbb{N}) \|v_2\| \leq \|(T_2x_n) - T_2T_1(T_2x_n)\| \leq \|x_n - T_1T_2x_n\|$ and thus $\|v_2\| \leq \|v_1\|$. Analogously, $\|v_1\| \leq \|v_2\|$.  

Unless stated otherwise, throughout the rest of this chapter we assume that

$$T \text{ is a nonexpansive operator on } X,$$  \hfill (7.6) 

and that

$$v = P_{\text{ran}(\text{Id} - T)}0.$$  \hfill (7.7) 

In view of (7.7) and Lemma 7.2(ii) we have

$$v \in \text{ran}(\text{Id} - T) \Leftrightarrow \text{Fix}(T - v) \neq \emptyset \Leftrightarrow \text{Fix}(v + T) \neq \emptyset.$$  \hfill (7.8)

The following example provides scenarios where $v \in \text{ran}(\text{Id} - T)$.

**Example 7.7.** Suppose that one of the following holds:
7.3. The displacement map and the minimal displacement vector

(i) $X$ is finite-dimensional and $T : X \to X$ is an affine operator.

(ii) $U$ and $V$ are nonempty closed convex subsets of $X$, $T$ is defined as in (14.8) below, and one of the following conditions holds:

(a) $V$ is bounded.

(b) $U$ and $V$ are polyhedral subsets\(^3\) of $X$.

Then

$$v \in \text{ran} \left( \text{Id} - T \right).$$

Proof. (i): $\text{ran}(\text{Id} - T)$ is a finite-dimensional affine subspace, hence closed. Therefore, $\text{ran}(\text{Id} - T) = \overline{\text{ran}(\text{Id} - T)}$ and (7.7) holds. (ii): It follows from the Browder–Göde–Kirk fixed point theorem (see [12, Theorem 4.19]) applied to the case (ii)(a), and [17, Theorem 5.6.1] applied to the case (ii)(b), that $\text{Fix} P_V P_U \neq \emptyset$. By Example 11.24 and Proposition 11.10 below, $v \in \text{ran}(\text{Id} - T)$ as claimed. \(\blacksquare\)

**Proposition 7.8.** Suppose that $v \in \text{ran}(\text{Id} - T)$ and let $y_0 \in \text{Fix}(v + T)$. Then the following hold\(^4\):

(i) $y_0 - R_+ v \subseteq \text{Fix}(v + T)$.

(ii) $\text{Fix}(v + T) - R_+ v = \text{Fix}(v + T)$.

(iii) $-R_+ v \subseteq \text{rec}(\text{Fix}(v + T))$.

(iv) $]-\infty, 1] \cdot v + \text{Fix}(T_+ v) \subseteq \text{Fix}(v + T)$. In particular it holds that $\text{Fix}(T_+ v) \subseteq \text{Fix}(v + T)$.

Proof. (i): First we use induction to show that

$$\left( \forall n \in \mathbb{N} \right) \ y_0 - nv \in \text{Fix}(v + T). \tag{7.9}$$

Clearly when $n = 0$ the base case holds true. Now suppose that for some $n \in \mathbb{N}$ it holds that $y_0 - nv \in \text{Fix}(v + T)$, i.e.,

$$y_0 - nv = v + T(y_0 - nv). \tag{7.10}$$

Using (7.7) and (7.10) we have

$$\|v\| \leq \|(\text{Id} - T)(y_0 - (n + 1)v)\| = \|y_0 - (n + 1)v - T(y_0 - (n + 1)v)\|$$

\(^3\)A subset of $X$ is polyhedral if it is a finite intersection of closed half-spaces.

\(^4\)Let $x \in X$. Then $R_+ x = \{rx \mid r \in [0, +\infty[\}$. 

77
7.3. The displacement map and the minimal displacement vector

\[
\begin{align*}
&= \| y_0 - nv - v - T(y_0 - (n + 1)v) \| \\
&= \| T(y_0 - nv) - T(y_0 - (n + 1)v) \| \leq \| v \|.
\end{align*}
\]

Consequently all the inequalities above are equalities and we conclude that \( \| v \| = \| y_0 - (n + 1)v - T(y_0 - (n + 1)v) \| \). It follows from (7.7) and Lemma 3.15 that

\[
y_0 - (n + 1)v - T(y_0 - (n + 1)v) = v. \tag{7.11}
\]

That is, \( y_0 - (n + 1)v = v + T(y_0 - (n + 1)v) \), which proves (7.9). Now using [12, Corollary 4.15] we learn that \( \text{Fix}(v + T) \) is convex, which when combined with (7.9) yields (i).

(ii): On the one hand, using (i) one concludes that \( \text{Fix}(v + T) - \mathbb{R}_+v \subseteq \text{Fix}(v + T) \). On the other hand, \( \text{Fix}(v + T) = \text{Fix}(v + T) - 0 \cdot v \subseteq \text{Fix}(v + T) - \mathbb{R}_+v \).

(iii): This follows directly from Proposition 7.8(ii).

(iv): Using Lemma 7.2(i) and Proposition 7.8(ii) we have \( [-\infty, 1] \cdot v + \text{Fix}(T_{-v}) = v - \mathbb{R}_+v + \text{Fix}(T_{-v}) = \text{Fix}(v + T) - \mathbb{R}_+v = \text{Fix}(v + T) \). In particular, we have \( \text{Fix}(T_{-v}) = 0 \cdot v + \text{Fix}(T_{-v}) \subseteq \text{Fix}(v + T) \).
Chapter 8

Asymptotic behaviour of nonexpansive mappings

8.1 Overview

This chapter is devoted to applications in fixed point theory. Throughout this chapter, we assume that

\[ T : X \rightarrow X \text{ is nonexpansive} \]

and that

\[ v = P_{\text{ran}(Id - T)}0. \]

Our main results in this chapter are:

- We prove the Fejér monotonicity of \((T^n x + nv)_{n \in \mathbb{N}}\) with respect to \(\text{Fix}(T - v)\) and \(\text{Fix}(v + T)\) (see Proposition 8.4).

- We compare the sequences \((-vT)^n x)_{n \in \mathbb{N}}, \((T-v)^n x)_{n \in \mathbb{N}}\) and \((T^nx + nv)_{n \in \mathbb{N}}\) when \(T\) is an affine nonexpansive operator and \(v \in \text{ran}(Id - T)\). We prove that the three sequences coincide (see Theorem 8.8). Surprisingly, when we drop the assumption of \(T\) being affine, the sequences can be dramatically different (see Example 8.9).

- We extend the well-known result in Fact 8.12 from linear to affine operators (see Theorem 8.20).

- A new characterization of strongly convergent iterations of affine nonexpansive operators is presented (see Theorem 8.20). We also discuss when the convergence is linear in Section 8.3.2.

Except for facts with explicit references, this chapter is based on results that appear in [15], [39] and [110].

\(^5\)In light of Fact 7.4 the vector \(v\) is unique and well-defined.
8.2 Iterating nonexpansive operators

In this section, we explore the asymptotic behaviour of nonexpansive mappings. The following two results are well-known.

**Fact 8.1 (Pazy 1971).** Exactly one of the following holds:

(i) Fix $T = \emptyset$ and $(\forall x \in X) \|T^n x\| \to +\infty.$

(ii) Fix $T \neq \emptyset$ and $(\forall x \in X) \ (T^n x)_{n\in\mathbb{N}}$ is bounded.

*Proof.* See [116, Corollary 6].

**Fact 8.2.** Let $T : X \to X$ be firmly nonexpansive. Then

$$(\forall x \in X) \quad v = \lim_{n \to \infty} T^n x - T^{n+1} x = - \lim_{n \to \infty} \frac{T^n x}{n}. \quad (8.1)$$

*Proof.* See [53, Corollary 1.5], [6, Corollary 2.3], and [116].

Before we proceed, we state the following lemma that relates the iterates of inner and outer shifts of a single-valued operator.

**Lemma 8.3.** Let $S : X \to X$, let $w \in X$, let $x \in X$ and let $n \in \mathbb{N}$. Then the following hold:

$$(S-w)^n x = -w + (w+S)^n (x+w) = -w + (-wS)^n(x+w). \quad (8.2)$$

*Proof.* We proceed by induction. The conclusion is clear when $n = 0$. Now assume that for some $n \in \mathbb{N}$ it holds that $(S-w)^n x = -w + (w+S)^n (x+w)$. Then $(S-w)^{n+1} x = S((S-w)^n x + w) = S(-w + (w+S)^n (x+w) + w) = -w + w + S((w+S)^n (x+w)) = -w + (w + S)^{n+1}(x+w)$, as claimed.

**Proposition 8.4.** Suppose that $v \in \text{ran}(\text{Id} - T)$ and let $y_0 \in \text{Fix}(v + T)$. Then the following hold:

(i) $(\forall n \in \mathbb{N}) \ T^n y_0 = y_0 - nv.$

(ii) For every $x \in X$, the sequence $(T^n x + nv)_{n\in\mathbb{N}}$ is Féjer monotone with respect to both $\text{Fix}(v + T)$ and $\text{Fix}(T-v)$.

---

6Let $x \in X$. Then $\mathbb{R}_+ x = \{rx \mid r \in [0, +\infty[\}$. 

80
8.2. Iterating nonexpansive operators

(iii) Suppose that \( x_0 \in \text{Fix}(T - v) \) and set \((\forall n \in \mathbb{N}) \) \( x_n = T^n x_0 \). Then \( x_n = x_0 - n v \) and \((x_n)_{n \in \mathbb{N}} \) lies in \( \text{Fix}(T - v) \).

Proof. (i): We use induction. Clearly \( y_0 - 0 v = y_0 = T^0 y_0 \). Now suppose that for some \( n \in \mathbb{N} \) it holds \( T^n y_0 = y_0 - n v \). Using Proposition 7.8(i) we have \( T^{n+1} y_0 = T(y_0 - n v) = -v + y_0 - n v = y_0 - (n + 1)v \).

(ii): Let \( x \in X \) and let \( y \in \text{Fix}(v + T) \). Then using (i) we have for every \( n \in \mathbb{N} \),

\[
\| T^{n+1} x + (n + 1)v - y \| = \| T^{n+1} x - (y - (n + 1)v) \| = \| T^{n+1} x - T^{n+1} y \| \\
\leq \| T^n x - T^n y \| = \| T^n x - (y - n v) \| \\
= \| T^n x + n v - y \|. \tag{8.3}
\]

The statement for \( \text{Fix}(T - v) \) follows from Proposition 7.8(iv).

(iii): Combine Proposition 7.8(iv) and (i) to get that \( x_n = x_0 - n v \). Now by Lemma 7.2(i) \( x_0 + v \in \text{Fix}(v + T) \). Using Proposition 7.8(i) we have \((\forall n \in \mathbb{N}) \) \( x_0 + v - n v \in \text{Fix}(v + T) \) or equivalently by Lemma 7.2(i) \( x_0 - n v \in -v + \text{Fix}(v + T) = \text{Fix}(T - v) \). \( \blacksquare \)

**Proposition 8.5.** Suppose that \( X = \mathbb{R} \), that \( \text{Fix} T = \emptyset \) and that \( v \in \text{ran}(v - T) \). Let \( x \in \mathbb{R} \) and set \((\forall n \in \mathbb{N}) \) \( y_n = T^n x + n v \). Then the following hold:

(i) \( (y_n)_{n \in \mathbb{N}} \) converges.

(ii) \( \mathbb{R} \to \mathbb{R} : x \mapsto \lim_{n \to \infty} (T^n x + n v) \) is nonexpansive.

(iii) Suppose that \( T \) is firmly nonexpansive. Then \( \mathbb{R} \to \mathbb{R} : x \mapsto \lim_{n \to \infty} (T^n x + n v) \) is firmly nonexpansive.

Proof. (i): In view of Proposition 8.4(ii) the sequence \((y_n)_{n \in \mathbb{N}} \) is Féjer monotone with respect to \( \text{Fix}(v + T) \). Now by Proposition 7.8(i) we know that \( \text{Fix}(v + T) \) contains an unbounded interval. Since \( X = \mathbb{R} \) we conclude that \( \text{int} \text{Fix}(v + T) \neq \emptyset \). It follows from [12, Proposition 5.10] that \((y_n)_{n \in \mathbb{N}} \) converges.

(ii): Let \( y \in \mathbb{R} \). Then

\[
\lim_{n \to \infty} (T^n x + n v) - \lim_{n \to \infty} (T^n y + n v) = \lim_{n \to \infty} (T^n x + n v - T^n y - n v) \\\n= \lim_{n \to \infty} |T^n x - T^n y| \leq \lim_{n \to \infty} |x - y|
\]

81
8.2. Iterating nonexpansive operators

\[ |x - y|. \quad (8.4) \]

(iii): It follows from [12, Proposition 4.2(iv)] that an operator is firmly nonexpansive if and only if it is nonexpansive and monotone. Therefore, in view of (ii), we need to check monotonicity. Without loss of generality let \( y \in \mathbb{R} \) such that \( x \leq y \). Since \( T \) is firmly nonexpansive, hence monotone, one can verify that \((\forall n \in \mathbb{N}) T^n x \leq T^n y \) and therefore \((\forall n \in \mathbb{N}) T^n x + nv \leq T^n y + nv \). Now take the limit as \( n \to \infty \). \( \blacksquare \)

Example 8.6 below shows that the assumptions that \( \text{Fix} T = \emptyset \) and \( v \in \text{ran}(\text{Id} - T) \) (equivalently; \( v \neq 0 \)) in Proposition 8.5 are important.

Example 8.6. Suppose that \( X = \mathbb{R} \), that \( T = -\text{Id} \) and that \( x \neq 0 \). Then the sequence \((T^n x)_{n \in \mathbb{N}} = (x, -x, x, -x, \ldots)\) is not convergent.

When \( X = \mathbb{R} \), it follows from Proposition 8.5(i) that the sequence \((T^n x + nv)_{n \in \mathbb{N}}\) converges. In view of Proposition 8.4(ii) the sequence \((T^n x + nv)_{n \in \mathbb{N}}\) is Féjer monotone with respect to \( \text{Fix}(v + T) \) which might suggest that the limit lies in \( \text{Fix}(v + T) \). We show in the following example that this is not true in general.

Example 8.7. Suppose that \( X = \mathbb{R} \) and that

\[
T : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 
  x - \alpha, & \text{if } x \leq \alpha; \\
  0, & \text{if } \alpha < x \leq \beta; \\
  x - \beta, & \text{if } x > \beta,
\end{cases} \quad (8.5)
\]

where \( 0 < \alpha < \beta \). Then \( T \) is firmly nonexpansive but not affine, \( v = \alpha \), \( \text{Fix}(v + T) = [-\infty, \alpha] \), \( \text{Fix}(T - v) = [-\infty, 0] \), and

\[
T^n + nv : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 
  x, & \text{if } x \leq \alpha; \\
  \alpha, & \text{if } \alpha < x \leq \beta; \\
  x - n(\beta - \alpha), & \text{if } x > \beta \text{ and } n \leq \lfloor x/\beta \rfloor; \\
  \min \{ \alpha, x - \lfloor x/\beta \rfloor \beta \} + \lfloor x/\beta \rfloor \alpha, & \text{if } x > \beta \text{ and } n > \lfloor x/\beta \rfloor. 
\end{cases} \quad (8.6)
\]

Consequently,

\[
\lim_{n \to \infty} (T^n + nv) : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 
  x, & \text{if } x \leq \alpha; \\
  \alpha, & \text{if } \alpha < x \leq \beta; \\
  \min \{ \alpha, x - \lfloor x/\beta \rfloor \beta \} + \alpha \lfloor x/\beta \rfloor, & \text{if } x > \beta. 
\end{cases} \quad (8.7)
\]
Therefore for every $x_0 \in \mathbb{R}$ the sequence $(T^n x_0 + n v)_{n \in \mathbb{N}}$ is eventually constant. However, if the starting point $x_0$ lies in the interval $]\beta, +\infty[$, then $\lim_{n \to \infty} T^n x_0 + n v = \min\{\alpha, x - \left\lfloor \frac{x}{\beta} \right\rfloor \beta\} + \alpha \left\lfloor \frac{x}{\beta} \right\rfloor \notin \text{Fix}(v + T)$.

Proof. Clearly

$$\text{Id} - T = P_{[\alpha, \beta]} : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} \alpha, & \text{if } x \leq \alpha; \\ x, & \text{if } \alpha < x \leq \beta; \\ \beta, & \text{if } x > \beta. \end{cases} \quad (8.8)$$

Therefore, $\text{ran}(\text{Id} - T) = [\alpha, \beta]$, and consequently $v = \alpha$. Moreover

$$(\forall x \in \mathbb{R}) \quad x \geq Tx + \alpha \geq T^2 x + 2\alpha \geq \cdots \geq T^n x + n\alpha \geq \cdots. \quad (8.9)$$

It is clear from Example 7.5 that

$$\text{Fix}(v + T) = ]-\infty, \alpha]. \quad (8.10)$$

The statement for $\text{Fix}T_{v}$ then follows from combining (8.10) and Lemma 7.2(i). The convergence of the sequence follows from Example 7.5 or Proposition 8.5(i). Now we prove (8.6). We claim that

$$T^n : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} x - n\alpha, & \text{if } x \leq \alpha; \\ (1 - n)\alpha, & \text{if } \alpha < x \leq \beta; \\ x - n\beta, & \text{if } x > \beta \text{ and } n \leq \left\lfloor x/\beta \right\rfloor; \\ \min \{\alpha, x - \left\lfloor \frac{x}{\beta} \right\rfloor \beta\} + \left(\left\lfloor \frac{x}{\beta} \right\rfloor - n\right)\alpha, & \text{if } x > \beta \text{ and } n > \left\lfloor x/\beta \right\rfloor. \end{cases} \quad (8.11)$$

Using induction it is easy to verify the cases when $x \leq \alpha$ and when $\alpha < x \leq \beta$. Now we focus on the case when $x > \beta$. Set

$$K = \left\lfloor x/\beta \right\rfloor \text{ and } r = x - K\beta, \quad (8.12)$$

and note that $x = K\beta + r$, $K \in \{1, 2, 3, \ldots\}$ and $0 \leq r < \beta$. In view of (8.9), if $n \in \{0, 1, 2, 3, \ldots, K\}$ we get $T^n x = x - n\beta = (K - n)\beta + r$. In particular,

$$T^K x = x - \left\lfloor x/\beta \right\rfloor \beta = r. \quad (8.13)$$

If $n > K$ we examine two cases. Case 1: $0 \leq r \leq \alpha$. It follows from (8.13) and (8.5) that $(\forall n \geq K) T^n x = r + (K - n)\alpha$. Case 2: $\alpha < r < \beta$.
8.3 Affine nonexpansive operators

Note that $T^{K+1}x = 0$, therefore using (8.13) and (8.5) we have $(\forall n > K) T^n x = (K + 1 - n) a = a + (K - n) a$, which proves (8.11). Now (8.6) follows from (8.11) because $v = a$. Letting $n \to \infty$ in (8.6) yields (8.7). Note that $\min\{a, x - \lceil \frac{x}{\beta} \rceil \beta \} \geq 0$ and $\lceil \frac{x}{\beta} \rceil \beta \geq 1$. By considering cases ($K = 1$ and $K \geq 1$), (8.7) implies that $\lim_{n \to \infty} (T^n x_0 + nv) = \min\{a, x - \lceil \frac{x}{\beta} \rceil \beta \} + \lceil \frac{x}{\beta} \rceil \beta > a \not\in [-\infty, a] = \text{Fix}(v + T)$. ■

8.3 Affine nonexpansive operators

In this section, we investigate properties of affine nonexpansive operators and their corresponding inner and outer normal shifts. This additional assumption allows for stronger results than those obtained in the previous section. Various examples that illustrate our theory are provided.

8.3.1 Iterating an affine nonexpansive operator

In this section, we compare the sequences $((\tau v T)^n x)_{n \in \mathbb{N}}, ((T - v)^n x)_{n \in \mathbb{N}}$, and $(T^n x + nv)_{n \in \mathbb{N}}$ when $T$ is an affine nonexpansive operator and $v = P_{\text{ran}(\text{Id} - T)}(0 \in \text{ran}(\text{Id} - T))$. We prove that the three sequences coincide (see Theorem 8.8). Surprisingly, when we drop the assumption of $T$ being affine, the sequences can be dramatically different (see Example 8.9).

**Theorem 8.8.** Let $L : X \to X$ be linear and nonexpansive, let $b \in X$, and suppose that $T : X \to X : x \mapsto Lx + b$. Suppose also that $v \in \text{ran}(\text{Id} - T)$, and let $x \in X$. Then the following hold:

(i) $v = P_{\text{Fix}L}(-b) \in \text{Fix} L = (\text{ran}(\text{Id} - L))^\perp$, and $v \neq 0 \Leftrightarrow b \not\in \text{ran}(\text{Id} - L)$.

(ii) $(\forall n \in \mathbb{N}) T^n x = L^n x + \sum_{k=0}^{n-1} L^k b$.

(iii) $(\forall n \in \mathbb{N}) T^n x + nv = L^n x + \sum_{k=0}^{n-1} L^k P_{\text{ran}(\text{Id} - L)} b$.

(iv) $(\forall n \in \mathbb{N}) (T - v)^n x = T^n x + nv$.

(v) $(\forall n \in \mathbb{N}) (\tau v)^n x = (v + T)^n x$.

(vi) $\text{Fix}(T - v) = -v + \text{Fix}(T - v) = -v + \text{Fix}(\tau v T) = -v + \text{Fix}(v + T) = \text{Fix}(v + T)$.
(vii) Fix(T₀) = Fix(v + T) = Rv + Fix(v + T) = Rv + Fix(T₀). Consequently v lies in the lineality space\(^7\) of Fix(T₀) = Fix(v + T).

**Proof.** (i): Note that ran(Id − T) = ran(Id − L) − b and hence ran(Id − T) = ran(Id − L) − b. Therefore, using Fact 2.6 we have

\[
v = P_{\overline{\text{ran}}(\text{Id} - T)}0 = P_{-b + \overline{\text{ran}}(\text{Id} - L)}0 = -b + P_{\overline{\text{ran}}(\text{Id} - L)}(0 - (-b)) \tag{8.14a}
\]

\[
= -b + P_{\overline{\text{ran}}(\text{Id} - L)}b. \tag{8.14b}
\]

Using [12, Fact 2.18(iv)] and [19, Lemma 2.1], we learn that \((\overline{\text{ran}}(\text{Id} - L))^\perp = \ker(\text{Id} - L^*) = \text{Fix } L^* = \text{Fix } L\), and hence

\[
v = (\text{Id} - P_{\overline{\text{ran}}(\text{Id} - L)})(-b) = P_{(\overline{\text{ran}}(\text{Id} - L))^\perp}(-b) = P_{\text{Fix } L}(-b). \tag{8.15}
\]

Note that \(v \neq 0 \iff b \not\in \text{ran}(\text{Id} - L).\)

(ii): We prove this by induction. When \(n = 0\) the conclusion is obviously true. Now suppose that for some \(n \in \mathbb{N}\) it holds that

\[T^n x = L^n x + \sum_{k=0}^{n-1} L^k b. \tag{8.16}\]

Then \(T^{n+1} x = T(T^n x) = T(L^n x + \sum_{k=0}^{n-1} L^k b) = L(L^n x + \sum_{k=0}^{n-1} L^k b) + b = L^{n+1} x + \sum_{k=0}^{n} L^k b\), as claimed.

(iii): Note that \(b = P_{\overline{\text{ran}}(\text{Id} - L)}b + \text{Fix } L b\). Using (i) and Theorem 8.8 (ii) yield

\[
T^n x + nv = L^n x + \sum_{k=0}^{n-1} (L^k b + v) = L^n x + \sum_{k=0}^{n-1} (L^k b + L^k v)
\]

\[
= L^n x + \sum_{k=0}^{n-1} (L^k b - L^k P_{\text{Fix } L} b) = L^n x + \sum_{k=0}^{n-1} L^k (\text{Id} - P_{\text{Fix } L} b)
\]

\[
= L^n x + \sum_{k=0}^{n-1} L^k P_{\overline{\text{ran}}(\text{Id} - L)} b.
\]

(iv): We prove this by induction. Note that by (i), \(v \in \text{Fix } L\), hence \(L v = v\). When \(n = 0\) we have \((T_0)^0 x = x = T^0 x + 0 \cdot v\). Now suppose that for some \(n \in \mathbb{N}\) it holds that \((T_0)^n x = T^n x + nv\). Then \((T_0)^{n+1} x = T(T^n x + nv) = T(T^n x + nv + v) = L(T^n x) + L((n + 1)v) + b = T^{n+1} x + (n + 1)v\).

---

\(^7\) For the definition and a detailed discussion of the lineality space, we refer the reader to [123, page 65].
8.3. Affine nonexpansive operators

(v) We use induction again. The base case is obviously true. Now suppose that for some \( n \in \mathbb{N} \) it holds that \((v + T)^n x = T^n x + n v\). Then \((v + T)^{n+1} x = v + T(v + T)^n x = v + T(T^n x + n v) + b = v + LT^n x + n v + b = LT^n x + b + (n + 1)v = T^{n+1} x + (n + 1)v\). Now combine with (iv).

(vi): It follows from (v) applied with \( n = 1 \) that \( T^{-v} = v + T \). Now apply Lemma 7.2(i).

(vii): Using (vi) and the assumption that \( T \) is an affine operator, we have \( \text{Fix}(T^{-v}) = \text{Fix}(v + T) \) is an affine subspace. Now let \( y_0 \in \text{Fix}(T^{-v}) = \text{Fix}(v + T) \). Using Proposition 7.8(i) we have \(-R + v \subseteq \text{Fix}(v + T) - y_0 = \text{par Fix}(v + T)\) and therefore \( Rv \subseteq \text{par Fix}(v + T) \). Hence \( y_0 + Rv \subseteq \text{Fix}(v + T) \) which yields \( \text{Fix}(v + T) + Rv \subseteq \text{Fix}(v + T) \). Since the opposite inclusion is obviously true we conclude that (vii) holds.

Suppose \( T \) is nonexpansive but not affine. Theorem 8.8 might suggest that, for every \( x \in X \), the sequences \((T^n x + n v)_{n \in \mathbb{N}}\), \(((T^{-v})^n x)_{n \in \mathbb{N}}\) and \(((v + T)^n x)_{n \in \mathbb{N}}\) coincide, and consequently \((T^n x + n v)_{n \in \mathbb{N}}\) is a sequence of iterates of a nonexpansive operator. Interestingly, this is not the case as we illustrate now.

**Example 8.9.** Suppose that \( X = \mathbb{R} \) and let \( \beta > 0 \). Suppose that

\[
T: \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 
  x - \beta, & x \leq \beta; \\
  \alpha (x - \beta), & x > \beta,
\end{cases}
\]

where \( 0 < \alpha < 1 \). Then \( \text{Fix} T = \emptyset, v = \beta \), for every \( n \in \mathbb{N} \)

\[
(T^{-v})^n : \mathbb{R} \to \mathbb{R} : x \mapsto \alpha^n \max \{x, 0\} + \min \{x, 0\},
\]

\[
(v + T)^n : \mathbb{R} \to \mathbb{R} : x \mapsto \alpha^n \max \{x - \beta, 0\} + \min \{x, \beta\},
\]

and

\[
T^n + n v : \mathbb{R} \to \mathbb{R} :
\]

\[
x \mapsto \begin{cases} 
  x, & \text{if } x \leq \beta; \\
  \alpha^n x - \left( \frac{\alpha(1 - \alpha^n)}{1 - \alpha} \right) \beta + n \beta, & \text{if } x > \beta, n < q(x); \\
  \alpha^{q(x)} x - \left( \frac{\alpha(1 - \alpha^{q(x)})}{1 - \alpha} \right) \beta + q(x) \beta, & \text{if } x > \beta, n \geq q(x),
\end{cases}
\]

where \( q(x) : \mathbb{R} \to \mathbb{N} : x \mapsto \left\lfloor \log_{\alpha} \frac{\beta}{\alpha \beta + (1 - \alpha) \beta} \right\rfloor \).
8.3. Affine nonexpansive operators

Consequently,

$$\lim_{n \to \infty} (T - v)^n x = \min \{x, 0\}, \quad (8.21)$$

and

$$\lim_{n \to \infty} (v + T)^n x = \min \{x, \beta\}, \quad (8.22)$$

Moreover, there is no operator $S : \mathbb{R} \to \mathbb{R}$ such that for every $x \in \mathbb{R}$ and for every $n \in \mathbb{N}$ we have $S^n x = T^n x + nv$.

Proof. Considering cases, we easily check that

$$\text{Id} - T : \mathbb{R} \to \mathbb{R} : x \mapsto (1 - \alpha) \max \{x, \beta\} + \alpha \beta \geq \beta > 0. \quad (8.24)$$

Hence $\text{Fix } T = \emptyset$ and $v = \beta$ as claimed. Moreover, using (8.24) one can verify that

$$\lim_{n \to \infty} (T - v)^n x = \min \{x, 0\}, \quad (8.25)$$

We also verify that

$$\lim_{n \to \infty} (v + T)^n x = \min \{x, \beta\}, \quad (8.26)$$

We now prove (8.18) by induction. Let $x \in \mathbb{R}$.

Clearly when $n = 0$ the base case holds true. Now suppose that for some $n \in \mathbb{N}$ (8.18) holds. If $x \leq 0$, then $(T - v)^n x = x \leq 0$, and therefore, (8.26) implies that $(T - v)^{n+1} x = T - v((T - v)^n x) = T - v x = x$. Similarly we have $x > 0 \Rightarrow \alpha^n x = (T - v)^n x > 0$, and consequently (8.26) implies that $(T - v)^{n+1} x = T - v((T - v)^n x) = T - v(\alpha^n x) = \alpha^{n+1} x$. The proof of (8.19) follows from combining (8.18) and Lemma 8.3. Now we turn to (8.20). We consider two cases.

Case 1: $x \leq \beta$. It is obvious using the definition of $T$ that $(\forall n \in \mathbb{N}) T^n x = x - n\beta$.

Case 2: $x > \beta$. Let $n \in \mathbb{N}$ be such that $T^n x > \beta$. By (8.25) and (8.17) we have

$$T^{n+1} x = \alpha^{n+1} x - (\alpha^{n+1} + \alpha^n + \cdots + \alpha)\beta = \alpha^{n+1} x - \frac{\alpha(1 - \alpha^{n+1})}{1 - \alpha} \beta$$
8.3. Affine nonexpansive operators

\[= \alpha^{n+1} \left( \frac{(1-\alpha)x + \alpha \beta}{1-\alpha} \right) - \frac{\alpha}{1-\alpha} \beta. \quad (8.27)\]

In view of (8.25), there exists a unique integer, say, \(q(x) \in \{1, 2, \ldots\}\) that satisfies \(T^{q(x)-1}x > \beta\) and \(T^{q(x)}x \leq \beta\). Since \(0 < \alpha < 1\), using (8.27) we have

\[T^{q(x)}x \leq \beta \iff \alpha^{q(x)} \left( \frac{(1-\alpha)x + \alpha \beta}{1-\alpha} \right) - \frac{\alpha}{1-\alpha} \beta \leq \beta \]
\[
\iff \alpha^{q(x)} \left( \frac{(1-\alpha)x + \alpha \beta}{1-\alpha} \right) \leq \frac{\beta}{1-\alpha} \iff \alpha^{q(x)} ((1-\alpha)x + \alpha \beta) \leq \beta \]
\[
\iff \alpha^{q(x)} \leq \frac{\beta}{(1-\alpha)x + \alpha \beta} \iff q(x) \geq \log_{\alpha} \frac{\beta}{\alpha \beta + (1-\alpha)x}.
\]

Consequently, \(q(x) = \lceil \log_{\alpha} \frac{\beta}{\alpha \beta + (1-\alpha)x} \rceil \). At this point, since \(T^{q(x)}x \leq \beta\), we must have \((\forall n \geq q(x)) T^n x = T^{q(x)}x - (n - q(x))\beta\), which proves (8.20). The formulae (8.21), (8.22) and (8.23) are direct consequences of (8.18), (8.19) and (8.20), respectively. To prove the last claim note that if \(S : \mathbb{R} \to \mathbb{R}\) is such that for every \(n \in \mathbb{N}\) we have \(S^n = T^n + nv\), then setting \(n = 1\) must yield

\[S = v + T : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} x, & x \leq \beta; \\ \alpha (x - \beta) + \beta, & x > \beta. \end{cases} \quad (8.28)\]

Now compare (8.19) and (8.20).
8.3. Affine nonexpansive operators

Figure 8.1: The solid curve represents \( \lim_{n \to \infty} (T - v)^n x \), the dashed dotted curve represents \( \lim_{n \to \infty} (v + T)^n x \), and the dashed curve represents \( \lim_{n \to \infty} T^n x + n v \), when \( \alpha = 0.5 \) and \( \beta = 1 \).

Figure 8.1 provides a Maple [99] plot of the functions defined by (8.21), (8.22) and (8.23) that illustrates that they are pairwise distinct.

8.3.2 Strong and linear convergence of affine nonexpansive operators

In this section, we focus on strong convergence of iterates of affine nonexpansive operators (see Theorem 8.20 and Corollary 8.21). We also present mild conditions sufficient to obtain linear rate of convergence (see Theorem 8.23). Let \( Y \) be a Banach space. We shall use \( B(Y) \) to denote the set of bounded linear operators on \( Y \). Let \( L \in B(Y) \). The operator norm of \( L \) is \( \|L\| = \sup_{\|y\| \leq 1} \|Ly\| \).

Definition 8.10 (asymptotic regularity of operators). Let \( T : X \to X \). Then \( T \) is asymptotically regular if \( (\forall x \in X) \ T^n x - T^{n+1} x \to 0 \), i.e., \( (\forall x \in X) \ (T^n x)_{n \in \mathbb{N}} \) is asymptotically regular.

We start by collecting various results that will be useful in the sequel.
8.3. Affine nonexpansive operators

Fact 8.11. Let $T : X \to X$. Then
\[
\begin{align*}
T \text{ firmly nonexpansive} & \quad \text{Fix } T \neq \emptyset \\
\Rightarrow & \quad T \text{ asymptotically regular. (8.29)}
\end{align*}
\]

Proof. See [53, Corollary 1.1] or [12, Corollary 5.16(ii)]. \hfill ■

Fact 8.12. Let $L : X \to X$ be linear and nonexpansive, and let $x \in X$. Then
\[ L^n x \to P_{\text{Fix } L} x \iff L^n x - L^{n+1} x \to 0. \]  (8.30)

Proof. See [5, Proposition 4], [6, Theorem 1.1], [19, Theorem 2.2] or [12, Proposition 5.27]. (We mention in passing that in [5, Proposition 4] the author proved the result for general odd nonexpansive mappings in Hilbert spaces and in [6, Theorem 1.1], the authors generalize the result to Banach spaces.) \hfill ■

Definition 8.13. Let $Y$ be a real Banach space, let $(y_n)_{n \in \mathbb{N}}$ be a sequence in $Y$ and let $y_\infty \in Y$. Then $(y_n)_{n \in \mathbb{N}}$ converges to $y_\infty$, denoted $y_n \to y_\infty$, if $\|y_n - y_\infty\| \to 0$; $(y_n)_{n \in \mathbb{N}}$ converges $\mu$-linearly\(^8\) to $y_\infty$ if $\mu \in [0, 1]$ and there exists $M \geq 0$ such that
\[ (\forall n \in \mathbb{N}) \quad \|y_n - y_\infty\| \leq M\mu^n; \]  (8.31)
and finally $(y_n)_{n \in \mathbb{N}}$ converges linearly to $y_\infty$ if there exists $\mu \in [0, 1]$ and $M \geq 0$ such that (8.31) holds.

Example 8.14 (convergence vs. pointwise convergence of continuous linear operators). Let $Y$ be a real Banach space, let $(L_n)_{n \in \mathbb{N}}$ be a sequence in $B(Y)$, and let $L_\infty \in B(Y)$. Then:

(i) $(L_n)_{n \in \mathbb{N}}$ converges or converges uniformly to $L_\infty$ in $B(Y)$ if $L_n \to L_\infty$ (in $B(Y)$).

(ii) $(L_n)_{n \in \mathbb{N}}$ converges pointwise to $L_\infty$ if $(\forall y \in Y) L_n y \to L_\infty y$ (in $Y$).

Remark 8.15. It is easy to see that the convergence of a sequence of continuous linear operators implies pointwise convergence; however, the converse is not true (see [92, Example 4.9-2]).

\(^8\) This is also known as $R$-linear (or root linear) convergence; see [111, Appendix 2 on page 620]. We use here a more concise notation for convenience to expose directly the rate of convergence. By [28, Remark 3.7], this is equivalent to $(\exists M > 0)(\exists N \in \mathbb{N})(\forall n \geq N) \|y_n - y_\infty\| \leq M\mu^n$.
8.3. Affine nonexpansive operators

The following two results are part of the folklore; however, we have not been able to find precise references and thus include the short proofs for the reader’s convenience.

**Lemma 8.16.** Let $Y$ be a real Banach space, let $(L_n)_{n \in \mathbb{N}}$ be a sequence in $B(Y)$, let $L_\infty \in B(Y)$, and let $\mu \in [0, 1]$. Then

$$(\forall y \in Y) \ L_n y \to L_\infty y \ \mu\text{-linearly (in } Y) \iff L_n \to L_\infty \ \mu\text{-linearly (in } B(Y)).$$

(8.32)

**Proof.** Let $y \in Y$. “$\Rightarrow$”: Because $L_n y \to L_\infty y \ \mu\text{-linearly}$, there exists $M_y \geq 0$ such that $$(\forall n \in \mathbb{N}) \ \| (L_n - L_\infty) y \| \leq \mu^n M_y;$$ equivalently,

$$\left\| \left( \frac{L_n - L_\infty}{\mu^n} \right) y \right\| = \left\| \left( \frac{L_n - L_\infty}{\mu^n} \right) \mu y \right\| \leq M_y. \quad (8.33)$$

It follows from the Uniform Boundedness Principle (see [92, 4.7-3]) applied to the sequence $\left( \frac{L_n - L_\infty}{\mu^n} \right)_{n \in \mathbb{N}}$ that $(\exists M \geq 0) (\forall n \in \mathbb{N}) \ \| (L_n - L_\infty)/\mu^n \| \leq M$; equivalently, $\| L_n - L_\infty \| \leq M\mu^n$, as required. “$\Leftarrow$”: Since $L_n \to L_\infty \ \mu\text{-linearly}$, we have $(\exists M \geq 0) (\forall n \in \mathbb{N}) \ \| L_n - L_\infty \| \leq M\mu^n$. Therefore, $(\forall n \in \mathbb{N}) \ \| L_n y - L_\infty y \| \leq \| L_n - L_\infty \| \| y \| \leq M\| y \| \mu^n$. ■

**Lemma 8.17.** Suppose that $X$ is finite-dimensional, let $(L_n)_{n \in \mathbb{N}}$ be a sequence of linear nonexpansive operators on $X$ and let $L_\infty : X \to X$. Then the following are equivalent:

(i) $$(\forall x \in X) \ L_n x \to L_\infty x.$$ 

(ii) $L_n \to L_\infty$ pointwise (in $X$), and $L_\infty$ is linear and nonexpansive.

(iii) $L_n \to L_\infty$ (in $B(X)$).

**Proof.** The implications “(i)$\Rightarrow$(ii)” and “(iii)$\Rightarrow$(i)” are easy to verify. We now prove “(ii)$\Rightarrow$(iii)”: Suppose that $(x_n)_{n \in \mathbb{N}}$ is a sequence in $X$ such that $(\forall n \in \mathbb{N}) \ \| x_n \| = 1$ and

$$\| L_n - L_\infty \| - \| L_n x_n - L_\infty x_n \| \to 0. \quad (8.34)$$

After passing to a subsequence and relabeling if necessary, we can and do assume that $x_n \to x_\infty$ because $X$ is finite-dimensional. Since $L_\infty$ and $(L_n)_{n \in \mathbb{N}}$ are linear and nonexpansive, we have $\| L_\infty \| \leq 1$ and $(\forall n \in \mathbb{N}) \ \| L_n \| \leq 1$. Using the triangle inequality, we have

$$\| L_n x_n - L_\infty x_n \| = \| (L_n - L_\infty)(x_n - x_\infty) \| + (L_n - L_\infty)x_\infty \| \leq \| L_n - L_\infty \| \| x_n - x_\infty \| + \| L_\infty \| \| x_\infty \| \leq \| L_n - L_\infty \| \| x_n - x_\infty \| + \| x_\infty \|.$$
8.3. Affine nonexpansive operators

\[ x_\infty + \|(L_n - L_\infty)x_\infty\| \leq 2\|x_n - x_\infty\| + \|(L_n - L_\infty)x_\infty\| \rightarrow 0 + 0 = 0. \]

Now combine with (8.34). ■

**Corollary 8.18.** Suppose that \( X \) is finite-dimensional, let \( L : X \rightarrow X \) be linear, and let \( L_\infty : X \rightarrow X \) be such that \( L^n \rightarrow L_\infty \) pointwise. Then \( L^n \rightarrow L_\infty \) linearly.

**Proof.** Combine Lemma 8.17 and [35, Theorem 2.12(i)]. ■

**Lemma 8.19.** Let \( L : X \rightarrow X \) be linear, let \( b \in X \), set \( T : X \rightarrow X : x \mapsto Lx + b \), and suppose that \( \text{Fix } T \neq \emptyset \). Then there exists a point \( a \in X \) such that \( b = a - La \), and

\[ (\forall n \in \mathbb{N}) (\forall x \in X) \quad T^n x = a + L^n(x - a). \quad (8.35) \]

**Proof.** The existence of \( a \) follows from Theorem 7.3. By telescoping, we have

\[ \sum_{k=0}^{n-1} L^k b = \sum_{k=0}^{n-1} L^k(a - La) = a - L^n a. \quad (8.36) \]

Consequently, Theorem 8.8(ii) and (8.36) yield \( T^n x = L^n x + a - L^n a = a + L^n(x - a) \). ■

The following result extends Fact 8.12 from the linear to the affine case.

**Theorem 8.20.** Let \( L : X \rightarrow X \) be linear and nonexpansive, let \( b \in X \), set \( T : X \rightarrow X : x \mapsto Lx + b \), and suppose that \( \text{Fix } T \neq \emptyset \). Then the following are equivalent:

(i) \( L \) is asymptotically regular.

(ii) \( L^n \rightarrow P_{\text{Fix } L} \) pointwise.

(iii) \( T^n \rightarrow P_{\text{Fix } T} \) pointwise.

(iv) \( T \) is asymptotically regular.

**Proof.** Let \( x \in X \) and note that in view of Theorem 7.3 there exists \( a \in X \) such that \( b = a - La \). “(i)⇔(ii)”: This is Fact 8.12. “(ii)⇒(iii)” and (7.4) we have \( T^n x = L^n(x - a) + a \rightarrow P_{\text{Fix } L}(x - a) + a = P_{\text{Fix } T}x. \) “(iii)⇒(iv)” and (8.35) we have \( T^n x - T^{n+1} x \rightarrow P_{\text{Fix } T}x - P_{\text{Fix } T}x = 0. \) “(iv)⇒(i)” and (8.36) we have \( L^n x - L^{n+1} x = T^n(x + a) - T^{n+1}(x + a) \rightarrow 0, \) which finishes the proof. ■
8.3. Affine nonexpansive operators

Corollary 8.21. Let \( L : X \to X \) be linear and nonexpansive, let \( b \in X \), suppose that \( T : X \to X : x \mapsto Lx + b \) and that \( v \in \text{ran}(\text{Id} - T) \). Let \( x \in X \). Then \( \text{Fix}(v + T) \neq \emptyset \). Moreover the following are equivalent:

(i) \( L \) is asymptotically regular.

(ii) \( L^n x \to P_{\text{Fix}L}x \).

(iii) \( T^n x + nv = (v + T)^n x = (T - v)^n x \to P_{\text{Fix}(v + T)x} \).

(iv) \( T - v = v + T \) is asymptotically regular.

(v) \((T^n x + nv)_{n \in \mathbb{N}}\) is asymptotically regular.

Proof: The fact that \( \text{Fix}(v + T) \neq \emptyset \) follows from (7.8). Note that \( v + T = L + b + v \). Now apply Theorem 8.20 with \((T,b)\) replaced by \((v + T,v + b)\) and use Theorem 8.8(iv)&(v). Note that in view of Theorem 8.8(iv)&(v) the asymptotic regularity of \( v + T = T - v \) is equivalent to the asymptotic regularity of the sequence \((T^n x + nv)_{n \in \mathbb{N}}\). \( \blacksquare \)

We now turn to linear convergence.

Lemma 8.22. Suppose that \( X \) is finite-dimensional, and let \( L : X \to X \) be linear and nonexpansive. Then the following are equivalent:

(i) \( L \) is asymptotically regular.

(ii) \( L^n \to P_{\text{Fix}L} \) pointwise (in \( X \)).

(iii) \( L^n \to P_{\text{Fix}L} \) (in \( \mathcal{B}(X) \)).

(iv) \( L^n \to P_{\text{Fix}L} \) linearly pointwise (in \( X \)).

(v) \( L^n \to P_{\text{Fix}L} \) linearly (in \( \mathcal{B}(X) \)).

Proof: “(i)\(\Leftrightarrow\)(ii)”: This follows from Fact 8.12. “(ii)\(\Leftrightarrow\)(iii)”: Combine Fact 8.12 and Lemma 8.17. “(iii)\(\Rightarrow\)(v)”: Apply Corollary 8.18 with \( L_\infty \) replaced by \( P_{\text{Fix}L} \). “(v)\(\Rightarrow\)(iii)”: This is obvious. “(iv)\(\Leftrightarrow\)(v)”: Apply Lemma 8.16 to the sequence \((L^n)_{n \in \mathbb{N}}\) and use Fact 8.12. \( \blacksquare \)

Theorem 8.23. Let \( L : X \to X \) be linear and nonexpansive, let \( b \in X \), set \( T : X \to X : x \mapsto Lx + b \), suppose that \( \text{Fix} T \neq \emptyset \), and let \( \mu \in ]0,1[ \). Then the following are equivalent:

(i) \( T^n \to P_{\text{Fix}T} \) \( \mu \)-linearly pointwise (in \( X \)).
8.3. Affine nonexpansive operators

(ii) \( L^n \rightarrow P_{\text{Fix}L} \mu\text{-linearly pointwise (in } X) \).

(iii) \( L^n \rightarrow P_{\text{Fix}L} \mu\text{-linearly (in } B(X)) \).

Proof. In view of Theorem 7.3 there exists \( a \in X \) such that \( b = a - La \).

“(i)\(\Leftrightarrow\)(ii)”: It follows from Lemma 8.19 and (7.4) that

\[
T^n x - P_{\text{Fix}T} x = a + L^n (x - a) - (a + P_{\text{Fix}L} (x - a)) = L^n (x - a) - P_{\text{Fix}L} (x - a).
\]


Corollary 8.24. Let \( L : X \rightarrow X \) be linear and nonexpansive, let \( b \in X \), suppose that \( T : X \rightarrow X : x \mapsto Lx + b \) and that \( v \in \text{ran}(\text{Id} - T) \). Let \( x \in X \) and let \( \mu \in ]0,1[ \). Then the following are equivalent:

(i) \( T^n x + nv = (v + T)^n x = (T - v)^n x \rightarrow P_{\text{Fix}(v + T)} x \mu\text{-linearly}. \)

(ii) \( L^n x \rightarrow P_{\text{Fix}L} x \mu\text{-linearly}. \)

(iii) \( L^n \rightarrow P_{\text{Fix}L} \mu\text{-linearly (in } B(X)) \).

Proof. Note that \( \text{Fix}(v + T) \neq \emptyset \) by (7.8). Now apply Theorem 8.23 with \( (T, b) \) replaced by \( (v + T, v + b) \) and use Theorem 8.8(iv)&(v). ■

Corollary 8.25. Suppose that \( X \) is finite-dimensional. Let \( L : X \rightarrow X \) be linear, nonexpansive and asymptotically regular, let \( b \in X \), set \( T : X \rightarrow X : x \mapsto Lx + b \) and suppose that \( \text{Fix} T \neq \emptyset \). Then \( T^n \rightarrow P_{\text{Fix}T} \) pointwise linearly.

Proof. It follows from Fact 8.12 that \( L^n \rightarrow P_{\text{Fix}L} \) pointwise. Consequently, by Corollary 8.18, \( L^n \rightarrow P_{\text{Fix}L} \) linearly. Now apply Theorem 8.23. ■

Corollary 8.26. Suppose that \( X \) is finite-dimensional. Let \( L : X \rightarrow X \) be linear, nonexpansive and asymptotically regular, let \( b \in X \) and set \( T : X \rightarrow X : x \mapsto Lx + b \). Let \( x \in X \). Then \( v \in \text{ran}(\text{Id} - T) \) and

\[
T^n x + nv = (v + T)^n x = (T - v)^n x \rightarrow P_{\text{Fix}(v + T)} x \text{ linearly.} \tag{8.37}
\]

Proof. Since \( X \) is finite-dimensional we learn that \( \text{ran}(\text{Id} - T) \) is a closed affine subspace of \( X \), hence \( v \in \text{ran}(\text{Id} - T) \); equivalently \( \text{Fix}(v + T) \neq \emptyset \) by (7.8). Now apply Theorem 8.23 with \( (T, b) \) replaced by \( (v + T, v + b) \) and use Theorem 8.8(iv)&(v). ■
8.3. Affine nonexpansive operators

8.3.3 Some algorithmic consequences

In this section, we make use of the following useful fact that is well known in analysis.

**Fact 8.27.** Suppose that \((a_n)_{n \in \mathbb{N}}\) is a decreasing sequence of nonnegative real numbers such that \(\sum_{n=0}^{\infty} a_n < +\infty\). Then

\[ na_n \to 0. \quad (8.38) \]

**Proof.** See [90, Section 3.3, Theorem 1]. □

To make further progress we impose now additional assumptions on \(T\).

**Definition 8.28 (averaged operator).** Let \(T : X \to X\). Then \(T\) is averaged if there exist a nonexpansive operator \(R : X \to X\) and a constant \(\alpha \in [0, 1]\) such that \(T = (1 - \alpha) \text{Id} + \alpha R\); equivalently, (see [12, Proposition 4.25])

\[ \|Tx - Ty\|^2 + \frac{1-\alpha}{\alpha} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2. \quad (8.39) \]

Clearly,

\(T\) is averaged with \(\alpha = 1/2 \iff T\) is firmly nonexpansive. \( (8.40) \)

Averaged operators have proven to be a useful class in fixed point theory and optimization; see [6] and [61]. The following result is well known when \(T\) is firmly nonexpansive. We include a simple proof, when \(T\) is averaged, for the sake of completeness.

**Proposition 8.29.** Suppose that \(T\) is averaged and that \(v \in \text{ran}(\text{Id} - T)\). Let \(x \in X\). Then the following hold:

(i) \(\sum_{n=0}^{\infty} \|T^n x - T^{n+1} x - v\|^2 < +\infty\).

(ii) \(T^n x - T^{n+1} x \to v\), equivalently; the sequence \((T^n x + nv)_{n \in \mathbb{N}}\) is asymptotically regular.

**Proof.** It follows from [61, Lemma 2.1] that \((\exists \alpha \in [0, 1])\) such that \((\forall x \in X)(\forall y \in X)\)

\[ \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \frac{\alpha}{1 - \alpha} \left(\|x - y\|^2 - \|Tx - Ty\|^2\right). \quad (8.41) \]

Moreover, Proposition 8.4(ii) implies that \((T^n x + nv)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(\text{Fix}(v + T)\). Now let \(n \in \mathbb{N}\) and let \(y_0 \in \text{Fix}(v + T)\). Using
Proposition 8.4(iii) we learn that \( T^n y_0 = y_0 - nv \). It follows from (8.41) applied with \((x, y)\) replaced by \((T^n x, T^n y_0)\) that

\[
\| T^n x - T^{n+1} x - v \|^2 = \|(I - T) T^n x - (I - T) T^n y_0 \|^2 \\
\leq \frac{\alpha}{1 - \alpha} \left( \| T^n x - T^n y_0 \|^2 - \| T^{n+1} x - T^{n+1} y_0 \|^2 \right).
\]

8.3. Affine nonexpansive operators

(i): This follows from (8.42) by telescoping. (ii): This is a direct consequence of (i).

Lemma 8.30. Let \( L : X \to X \) be linear, nonexpansive and asymptotically regular, let \( b \in X \), and suppose that \( T : X \to X : x \mapsto Lx + b \) and that \( v \in \text{ran}(I - T) \). Let \( x \in X \). Then the sequence \( \| T^n x - T^{n+1} x - v \| \) is a decreasing sequence of nonnegative real numbers that converges to 0.

Proof. Let \( n \in \mathbb{N} \). It follows from Theorem 8.8(iv)\&(v) that \( T^n x + nv = (v + T)^n x \). Moreover, since \( L \) is nonexpansive so is \( v + T \). Now

\[
\| T^n x - T^{n+1} x - v \| = \| T^n x + nv - (T^{n+1} x + (n + 1)v) \| \\
= \| (v + T)^n x - (v + T)^{n+1} x \| \\
\leq \| (v + T)^n x - (v + T)^n x \| \\
= \| T^n x + (n - 1)v - (T^n x + nv) \| \\
= \| T^n x - T^n x - v \|.
\]

The claim about convergence follows from Proposition 8.29(ii).

Theorem 8.31. Let \( L : X \to X \) be linear and averaged, let \( b \in X \), and suppose that \( T : X \to X : x \mapsto Lx + b \) and that \( v \in \text{ran}(I - T) \). Let \( x \in X \) and set

\[
(\forall n \in \mathbb{N}) \quad x_n = T^n x + n(T^{n^2} x - T^{n^2+1} x). 
\]

Then \( x_n \to P_{\text{Fix}(v + T)x} \).

Proof. We have

\[
\| x_n - (v + T)^n x \| = \| T^n x + n(T^{n^2} x - T^{n^2+1} x) - (T^n x + nv) \| \\
= n \| T^{n^2} x - T^{n^2+1} x - v \| \\
= \sqrt{n^2} \| T^{n^2} x - T^{n^2+1} x - v \| \to 0,
\]

96
where the limit follows from combining Proposition 8.29(i), Lemma 8.30 and Fact 8.27 applied with $a_n$ replaced by $\|T^n x - T^{n+1} x - v\|^2$. It follows from Corollary 8.21 that $(v + T)^n x \to P_{\text{Fix}(v + T)} x$, hence the conclusion follows.

8.4 Further results

Let $x$ and $y$ be in $X$. It is clear that $(\forall n \in \mathbb{N}) \| T^n x - T^{n+1} x \| \leq \| T^n x - T^n y \|$ hence $(\| T^n x - T^n y \|)_{n \in \mathbb{N}}$ is bounded. The following question is thus natural:

Under which conditions on $T$ must $(T^n x - T^n y)_{n \in \mathbb{N}}$ converge weakly?

(8.46)

In this section we present some sufficient conditions that guarantee that (8.46) holds. We first note that, in view of Example 8.6, (8.46) will impose some restriction on $T$:

**Fact 8.32.** Suppose that $\text{Fix} T \neq \emptyset$ and let $x \in X$. Then $(T^n x)_{n \in \mathbb{N}}$ is weakly convergent if and only if $T^n x - T^{n+1} x \rightharpoonup 0$; if this is the case, then $(T^n x)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix} T$.

*Proof.* See [6, Theorem 1.2].

**Lemma 8.33.** Suppose that $\text{Fix}(v + T) \neq \emptyset$. Then

$$(\forall x \in X)(\forall y \in X) \quad (T^n x - T^n y)_{n \in \mathbb{N}} \text{ is weakly convergent} \quad (8.47a)$$

if and only if

$$(\forall x \in X) \quad (T^n x + nv)_{n \in \mathbb{N}} \text{ is weakly convergent.} \quad (8.47b)$$

*Proof.* Indeed, if (8.47a) holds, then (8.47b) follows by choosing $y \in \text{Fix}(v + T)$ and recalling Proposition 8.4(i). Conversely, assume that (8.47b) holds. Then $(T^n x + nv)_{n \in \mathbb{N}}$ and $(T^n y + nv)_{n \in \mathbb{N}}$ are weakly convergent, and so is their difference which yields (8.47a).

**Remark 8.34.** Suppose that $X = \mathbb{R}$, that $v \neq 0$, and that $\text{Fix}(v + T) \neq \emptyset$. Then by Proposition 8.5(i) the sequence $(T^n x + nv)_{n \in \mathbb{N}}$ is convergent, which gives a mild sufficient condition for (8.46).
8.4. Further results

Recall that $T$ is asymptotically regular at $x$ if $T^n x - T^{n+1} x \to 0$ and that
$T$ is asymptotically regular if it is asymptotically regular at every point.

**Theorem 8.35.** Suppose that $T$ is affine, say $T: x \to Lx + b$, where $L$ is linear
and nonexpansive, and $b \in X$. Suppose furthermore that $L$ is asymptotically
regular, and let $x$ and $y$ be points in $X$. Then

$$T^n x - T^n y = L^n (x - y) \to P_{\text{Fix}(L)}(x - y). \quad (8.48)$$

**Proof.** Using Theorem 8.8(iii), we have $(\forall n \in \mathbb{N})$ $T^n x - T^n y = L^n x - L^n y = L^n (x - y)$. The asymptotic regularity assumption yields
$L^n (x - y) - L^{n+1} (x - y) \to 0$. Using Fact 8.12, we see that altogether
$T^n x - T^n y = L^n (x - y) \to P_{\text{Fix}(L)}(x - y)$. ■

Amazingly, on the real line, averagedness is another sufficient condition
(see Remark 8.34) for (8.46):

**Theorem 8.36.** Suppose that $X = \mathbb{R}$ and that $T$ is averaged. Let $x$ and $y$ be in $\mathbb{R}$.
Then the sequence $(T^n x - T^n y)_{n \in \mathbb{N}}$ is convergent.

**Proof.** Set $(\forall n \in \mathbb{N}) a_n = T^n x - T^n y$. We must show that $(a_n)_{n \in \mathbb{N}}$ is convergent. From (8.39), there exists $\alpha \in ]0, 1[$ such that

$$(\forall n \in \mathbb{N}) \quad a_{n+1}^2 + \frac{1-\alpha}{\alpha} (a_n - a_{n+1})^2 \leq a_n^2. \quad (8.49)$$

Set $\beta = 1 - 2\alpha$ and note that $0 \leq |\beta| < 1$. By viewing (8.49) as a quadratic
inequality in $a_{n+1}$, we learn that

$$(\forall n \in \mathbb{N}) \quad |a_{n+1}| \leq |a_n| \quad \text{and} \quad a_{n+1} \text{ lies between } a_n \text{ and } \beta a_n. \quad (8.50)$$

If some $a_n = 0$, then $a_n \to 0$ and we are done. So assume that $a_n \neq 0$
for every $n \in \mathbb{N}$. If $(a_n)_{n \in \mathbb{N}}$ changes sign only finitely many times, then
$(a_n)_{n \in \mathbb{N}}$ is eventually always positive or negative. Since $(|a_n|)_{n \in \mathbb{N}}$
is decreasing, we deduce that $(a_n)_{n \in \mathbb{N}}$ is convergent. Finally, we assume that
$(a_n)_{n \in \mathbb{N}}$ changes signs frequently. If $n \in \mathbb{N}$ and sgn$(a_{n+1}) = -\text{sgn}(a_n)$, then $|a_{n+1}| \leq |\beta||a_n|$; since this occurs infinitely many times, it follows that
$a_n \to 0$. ■

**Theorem 8.37.** Suppose that $X$ is finite-dimensional, that $T$ is averaged, that
Fix($v + T$) $\neq \emptyset$, and that codim Fix($v + T$) $\leq 1$. Then for every $(x, y) \in X \times X$, the sequence $(T^n x - T^n y)_{n \in \mathbb{N}}$ is convergent.
8.4. Further results

Proof. In view of Lemma 8.33, we let $x \in X$ and must show that $(T^n x + n v)_{n \in \mathbb{N}}$ is convergent. Set $C = \text{Fix}(v + T)$ and $(\forall n \in \mathbb{N})$ $x_n = T^n x + n v$. By Proposition 8.4(ii), $(x_n)_{n \in \mathbb{N}}$ is Fejér monotone with respect to $C$. Suppose first that codim $C = 0$. Then $\text{int} C \neq \emptyset$ and we are done by Fact 6.1(v). Now assume that codim $C = 1$. By Proposition 8.29(ii), $(x_n)_{n \in \mathbb{N}}$ is asymptotically regular. Altogether, by Theorem 6.11, $(x_n)_{n \in \mathbb{N}}$ is convergent. $\blacksquare$

Corollary 8.38. Let $x \in X$. Suppose that $X = \mathbb{R}^2$, that $T$ is averaged, that $v \neq 0$, and that $\text{Fix}(v + T) \neq \emptyset$. Then for every $(x, y) \in X \times X$, the sequence $(T^n x - T^n y)_{n \in \mathbb{N}}$ is convergent.

Proof. Because $v \neq 0$, Proposition 7.8(ii) implies that dim $\text{Fix}(v + T) \geq 1$, i.e., codim $\text{Fix}(v + T) \leq \dim(X) - 1 = 1$. The result now follows from Theorem 8.37. $\blacksquare$

Although it is not clear if Corollary 8.38 remains true when dim $(X) \geq 3$, we conclude with an example which numerically illustrates that the answer to this question may be positive.

Example 8.39. Suppose that $X = \mathbb{R}^3$ and let $A$ and $B$ be two closed balls in $X$ such that $A \cap B = \emptyset$. Set $T = \frac{1}{2}(\text{Id} + R_B R_A)$. Then $T$ is firmly nonexpansive and hence averaged. (In fact, $T$ is the Douglas–Rachford operator (see Definition 5.1) associated with the sets $A$ and $B$.) It follows from Fact 14.4 below that $A \cap (v + B) + N_{A-B} v \subseteq \text{Fix}(v + T) \subseteq v + A \cap (v + B) + N_{A-B} v$. Furthermore, Example 12.15 below implies that $A - B$ is a closed ball and $v$ lies on the boundary of $A - B$. Consequently, $A \cap (v + B)$ is a singleton and $N_{A-B} v$ is a ray. Hence Fix $(v + T)$ is ray and therefore dim Fix $(v + T) = 1$ and so codim Fix $(v + T) = 2$. Even though Theorem 8.37 is not applicable here, we still conjecture that $(T^n x + n v)_{n \in \mathbb{N}}$ converges (see Figure 8.2 below).
8.4. Further results

Figure 8.2: A GeoGebra [78] snapshot that illustrates Example 8.39. The first few terms of the sequence \((T^n x + n v)_{n \in \mathbb{N}}\) (blue points) are depicted.

---

100
Chapter 9

The Douglas–Rachford algorithm: convergence analysis

9.1 Overview

The results in this chapter are about the convergence of the Douglas–Rachford algorithm in the consistent case, i.e., when \( \text{zer}(A + B) \neq \emptyset \). In this chapter we assume that \( A \) and \( B \) are maximally monotone operators on \( X \)

\[
T = T_{(A,B)} = \text{Id} - J_A + J_BR_A.
\]

We also assume that

\[
Z = \text{zer}(A + B) \neq \emptyset \quad \text{and} \quad \text{Fix} T \neq \emptyset.
\]

Let \( x \in X \). The governing sequence and the shadow sequence generated by the Douglas–Rachford algorithm are \( (T^n x)_{n \in \mathbb{N}} \) and \( (J_A T^n x)_{n \in \mathbb{N}} \) respectively.

It is known that the governing sequence \( (T^n x)_{n \in \mathbb{N}} \) converges weakly to a point in \( \text{Fix} T \) and the shadow sequence \( (J_A T^n x)_{n \in \mathbb{N}} \) converges weakly to a point in \( \text{zer}(A + B) \) (see Fact 9.2 below).

We summarize our main results in this chapter as follows:

- We present a new proof for Svaiter’s result [132] concerning the weak convergence of the shadow sequence (see Theorem 9.4) in the consistent case. Our proof is in the spirit of the techniques used in the original paper by Lions and Mercier [96].

- The main algorithmic result (Theorem 9.7) is derived in Section 9.4. We provide precise information on the behaviour of the Douglas–Rachford algorithm in the affine case.

- We also provide an application to the parallel splitting method that can be used to find a zero of the sum of more than two operators (see Proposition 9.14).

- We sketch some algorithmic consequences and conclude by commenting on the applicability of our work to a more general duality framework.

Except for facts with explicit references, this chapter is based on results that appear in [13], [26], and [39].

9.2 Averaged operators, Krasnosel’skii–Mann iteration and convergence of Douglas–Rachford algorithm

Fact 9.1 (Krasnosel’skii–Mann). Suppose that $T : X \to X$ is averaged such that $\text{Fix} T \neq \emptyset$. Let $x \in X$. Then the following hold:

(i) $(T^n x)_{n \in \mathbb{N}}$ is Fejér monotone with respect to $\text{Fix} T$.

(ii) $(T^n x - T^{n+1} x)_{n \in \mathbb{N}}$ converges strongly to 0.

(iii) $(T^n x)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix} T$.

Proof. See [91], [98] or [12, Theorem 5.14]. □

Fact 9.2 (convergence of Douglas–Rachford algorithm). Let $A : X \rightrightarrows X$ and $B : X \rightrightarrows X$ be maximally monotone such that $\text{zer}(A + B) \neq \emptyset$ and let $x \in X$. Then $\text{Fix} T \neq \emptyset$ and there exists $\bar{x} \in \text{Fix} T$ such that the following hold:

(i) $J_A \bar{x} \in Z$.

(ii) $(T^n x - T^{n+1} x)_{n \in \mathbb{N}} = (J_A T^n x - J_B R_A T^n x)_{n \in \mathbb{N}}$ converges strongly to 0.

(iii) $(T^n x)_{n \in \mathbb{N}}$ converges weakly to $\bar{x}$.

(iv) $(J_A T^n x)_{n \in \mathbb{N}}$ converges weakly to $J_A \bar{x}$.

(v) $(J_B R_A T^n x)_{n \in \mathbb{N}}$ converges weakly to $J_A \bar{x}$.

(vi) Suppose that $B = N_V$, where $V$ is a closed affine subspace of $X$. Then $(P_Y T^n x)_{n \in \mathbb{N}}$ converges weakly to $J_A \bar{x}$.
(vii) Suppose that \( \text{int Fix } T \neq \emptyset \). Then \((J_AT^n x)_{n \in \mathbb{N}}\) and \((J_{BR}A T^n x)_{n \in \mathbb{N}}\) converge strongly to \( J_A \bar{x} \).

(viii) Suppose that one of the following hold:

(a) \( A \) is uniformly monotone on every nonempty bounded subset of \( \text{dom } A \).

(b) \( B \) is uniformly monotone on every nonempty bounded subset of \( \text{dom } B \).

Then \( \text{zer}(A + B) \) is a singleton. Moreover, \((J_AT^n x)_{n \in \mathbb{N}}\) and \((J_{BR}A T^n x)_{n \in \mathbb{N}}\) converge strongly to the unique point in \( Z \).

Proof. See [96], [62, Theorem 2.1] or [12, Section 25.2]. ■

9.3 A new proof of the Lions–Mercier–Svaiter theorem

Parts of the following two results are implicit in [132]; however, our proofs are different.

Proposition 9.3. Let \( x \in X \). Then the following hold:

(i) \( T^n x - T^{n+1} x = J_AT^n x - J_{BR}A T^n x = J_{A^{-1}} T^n x + J_{B^{-1}} R_A T^n x \to 0 \).

(ii) The sequence \((J_AT^n x, J_{BR}A T^n x, J_{A^{-1}} T^n x, J_{B^{-1}} R_A T^n x)_{n \in \mathbb{N}}\) is bounded and lies in \( \text{gra}(A \times B) \).

Suppose that \((a, b, a^*, b^*)\) is a weak cluster point of the sequence

\[ (J_AT^n x, J_{BR}A T^n x, J_{A^{-1}} T^n x, J_{B^{-1}} R_A T^n x)_{n \in \mathbb{N}}. \]  (9.1)

Then:

(iii) \( a - b = a^* + b^* = 0 \).

(iv) \( \langle a, a^* \rangle + \langle b, b^* \rangle = 0 \).

(v) \( (a, a^*) \in \text{gra } A \) and \( (b, b^*) \in \text{gra } B \).

(vi) For every \( x \in X \), the sequence \((J_AT^n x, J_{A^{-1}} T^n x)_{n \in \mathbb{N}}\) is bounded and its weak cluster points lie in \( S \).
Proof. (i): Apply Lemma 5.8(i) with $x$ replaced by $T^n x$. The claim of the strong limit follows from combining Fact 5.2(i) and [6, Corollary 2.3] or [12, Theorem 5.14(ii)].

(ii): The boundedness of the sequence follows from the weak convergence of $(T^n x)_{n \in \mathbb{N}}$ (see Fact 9.2(iii)) and the nonexpansiveness of the resolvents and reflected resolvents of monotone operators (see [12, Corollary 23.10(i) and (ii)]). Now apply Lemma 5.8(ii) with $x$ replaced by $T^n x$.

(iii): In view of (iii) we have $\langle a, a^* \rangle + \langle b, b^* \rangle = \langle a, a^* + b^* \rangle = \langle a, 0 \rangle = 0$. (v): Let $((x, y), (u, v)) \in \text{gra}(A \times B)$ and set

$$a_n = J_A T^n x, a_n^* = J_{A^{-1}} T^n x, b_n = J_B R_A T^n x, b_n^* = J_{B^{-1}} R_A T^n x.$$  \hfill (9.2)

Applying Lemma 5.7 with $(a, b, a^*, b^*)$ replaced by $(a_n, b_n, a_n^*, b_n^*)$ yields

$$\langle a_n - b_n, a_n^* - (u, v) \rangle - \langle a_n, a_n^* - (u, v) \rangle = \langle b_n, b_n^* \rangle + \langle x, u \rangle - \langle x, a_n^* \rangle - \langle a_n - b_n, u \rangle + \langle b_n, b_n^* \rangle + \langle y, v \rangle - \langle y, b_n^* \rangle - \langle b_n, u + v \rangle.$$ \hfill (9.3)

By (5.31), $A \times B$ is monotone. In view of (9.2), (9.3) and Proposition 9.3(ii), we deduce that

$$\langle a_n - b_n, a_n^* \rangle + \langle x, u \rangle - \langle x, a_n^* \rangle - \langle a_n - b_n, u \rangle + \langle b_n, a_n^* \rangle + \langle y, v \rangle - \langle y, b_n^* \rangle - \langle b_n, u + v \rangle \geq 0.$$ \hfill (9.4)

It follows from taking the limit in (9.4) along a subsequence and using (9.2), Proposition 9.3(i), (iii) and (iv) that

$$0 \leq \langle x, u \rangle - \langle x, a^* \rangle + \langle y, v \rangle - \langle y, b^* \rangle - \langle b, u + v \rangle \quad \text{(9.5a)}$$

$$= \langle x, u \rangle - \langle x, a^* \rangle + \langle y, v \rangle - \langle y, b^* \rangle - \langle a, u \rangle - \langle b, v \rangle + \langle a, a^* \rangle + \langle b, b^* \rangle \quad \text{(9.5b)}$$

$$= \langle a, a^* \rangle - \langle a, b^* \rangle - \langle b, b^* \rangle - \langle x, a^* \rangle - \langle b, v \rangle + \langle b, b^* \rangle \quad \text{(9.5c)}$$

$$= \langle (a, b) - (x, y), (a^*, b^*) - (u, v) \rangle.$$ \hfill (9.5d)

By maximality of $A \times B$ (see (5.31)), we deduce that $((a, b), (a^*, b^*))$ lies in $\text{gra}(A \times B)$. Therefore, $(a, a^*) \in \text{gra} A$ and $(b, b^*) \in \text{gra} B$. (vi): The boundedness of the sequence follows from (ii). Now let $(a, b, a^*, b^*)$ be a weak cluster point of $(J_A T^n x, J_B R_A T^n x, J_{A^{-1}} T^n x, J_{B^{-1}} R_A T^n x)_{n \in \mathbb{N}}$. By (v) we know that $(a, a^*) \in \text{gra} A$ and $(b, b^*) = (a, b^*) \in \text{gra} B$, which in view of (iv) implies $a^* \in A a$ and $-a^* = b^* \in B b = B a$, hence $(a, a^*) \in S$, as claimed (see (5.24)).

\[\blacksquare\]
### 9.3. A new proof of the Lions–Mercier–Svaiter theorem

**Theorem 9.4.** Let \( x \in X \) and let \((z, k) \in S\). Then the following hold:

(i) For every \( n \in \mathbb{N} \),

\[
\| (J A T^{n+1} x, J A^{-1} T^{n+1} x) - (z, k) \|^2 \\
= \| J A T^{n+1} x - z \|^2 + \| J A^{-1} T^{n+1} x - k \|^2 \\
\leq \| J A T^n x - z \|^2 + \| J A^{-1} T^n x - k \|^2 \\
= \| (J A T^n x, J A^{-1} T^n x) - (z, k) \|^2.
\]  

(ii) The sequence \((J A T^n x, J A^{-1} T^n x)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \( S \).

(iii) The sequence \((J A T^n x, J A^{-1} T^n x)_{n \in \mathbb{N}}\) converges weakly to some point in \( S \).

**Proof.** (i): Apply Corollary 5.24 with \( x \) replaced by \( T^n x \). (ii): This follows directly from (i). (iii): Combine Proposition 9.3(vii), (ii), Corollary 5.16(iii) and Fact 6.1(viii).

We are now ready for the main result of this section, i.e., an alternative shorter proof of the Lions–Mercier–Svaiter result.

**Corollary 9.5. (Lions–Mercier–Svaiter).** \((J A T^n x)_{n \in \mathbb{N}}\) converges weakly to some point in \( Z \).

**Proof.** This follows from Theorem 9.4(iii) and (5.24); see also Lions and Mercier’s [96, Theorem 1] and Svaiter’s [132, Theorem 1].

In our final result we show that when \( X = \mathbb{R} \), the Fejér monotonicity of the sequence \((J A T^n x, J A^{-1} T^n x)_{n \in \mathbb{N}}\) with respect to \( S \) can be decoupled to yield Fejér monotonicity of \((J A T^n x)_{n \in \mathbb{N}}\) and \((J A^{-1} T^n x)_{n \in \mathbb{N}}\) with respect to \( Z \) and \( K \), respectively.

**Lemma 9.6.** Suppose that \( X = \mathbb{R} \). Let \( x \in X \) and let \((z, k) \in Z \times K\). Then the following hold:

(i) The sequence \((J A T^n x)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \( Z \).

(ii) The sequence \((J A^{-1} T^n x)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \( K \).

**Proof.** Apply Lemma 5.25 with \( x \) replaced by \( T^n x \).

We point out that the conclusion of Lemma 9.6 does not hold when \( \dim X \geq 2 \), see [30, Section 5 & Figure 1].
9.4 The Douglas–Rachford algorithm in the affine case

In this section we assume that

\[ A : X \rightrightarrows X \] and \[ B : X \rightrightarrows X \] are maximally monotone and affine,

and that

\[ Z = \{ x \in X \mid 0 \in Ax + Bx \} \neq \emptyset. \tag{9.7} \]

Since the resolvents \( J_A \) and \( J_B \) are affine (see [27, Theorem 2.1(xix)]), so is \( T \). The results of this section are of historical interest and they extend the original setting of the Douglas–Rachford method from affine operators defined on a finite-dimensional space to possibly infinite-dimensional settings.

**Theorem 9.7.** Let \( x \in X \). Then the following hold:

(i) \( T^n x \to P_{\text{Fix} T} x \) and \( J_AT^n x \to J_A P_{\text{Fix} T} x \in Z. \)

(ii) Suppose that \( A \) and \( B \) are paramonotone such that \( K \perp (Z - Z) \) (as is the case when \(^9\) \( A \) and \( B \) are paramonotone and \( (\text{zer } A) \cap (\text{zer } B) \neq \emptyset \)). Then \( J_AT^n x \to P_Z x. \)

(iii) Suppose that \( X \) is finite-dimensional. Then \( T^n x \to P_{\text{Fix} T} x \) linearly and \( J_AT^n x \to J_A P_{\text{Fix} T} x \) linearly.

**Proof.** (i): Note that in view of Fact 5.2(iii) and (9.7) we have \( \text{Fix } T \neq \emptyset \). Moreover Fact 5.2(i) and Fact 8.11 imply that \( T \) is asymptotically regular. It follows from Theorem 8.20 that (i) holds. (ii): Use (i) and Theorem 5.19. (iii): The linear convergence of \( (T^n x)_{n \in \mathbb{N}} \) follows from Corollary 8.25. The linear convergence of \( (J_AT^n x)_{n \in \mathbb{N}} \) is a direct consequence of the linear convergence of \( (T^n x)_{n \in \mathbb{N}} \) and the fact that \( J_A \) is (firmly) nonexpansive. \( \blacksquare \)

**Remark 9.8.** Theorem 9.7 generalizes the convergence results for the original Douglas–Rachford algorithm [70] from particular symmetric matrices/affine operators on a finite-dimensional space to general affine relations defined on possibly infinite dimensional spaces, while keeping strong and linear convergence of the iterates of the governing sequence \( (T^n x)_{n \in \mathbb{N}} \) and identifying the limit to be \( P_{\text{Fix} T} x. \) Paramonotonicity coupled with common zeros yields convergence of the shadow sequence \( (J_AT^n x)_{n \in \mathbb{N}} \) to \( P_Z x. \)

\(^9\)See Theorem 4.32(ii).
The assumption that both operators are paramonotone is critical for the conclusion in Theorem 9.7(ii), as shown below.

**Example 9.9.** Suppose that \( X = \mathbb{R}^2 \), that

\[
A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and that} \quad B = N_{\{0\} \times \mathbb{R}}.
\] (9.8)

Then \( T : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto \frac{1}{2} (x - y) \cdot (1, -1) \), Fix \( T = \mathbb{R} \cdot (1, -1) \), \( Z = \{0\} \times \mathbb{R} \), \( K = \{0\} \), hence \( K \perp (Z - Z) \), and \( (\forall (x, y) \in \mathbb{R}^2) (\forall n \geq 1) T^n(x, y) = T(x, y) = \frac{1}{2} (x - y, y - x) \in \text{Fix} \, T \), however \( (\forall (x, y) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R}) \) \( (\forall n \geq 1) \)

\[
(0, y - x) = J_A T^n(x, y) \neq P_Z(x, y) = (0, y).
\] (9.9)

Note that \( A \) is not paramonotone by Example 3.25.

**Proof.** We have

\[
J_A = (\text{Id} + A)^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},
\] (9.10)

and

\[
R_A = 2J_A - \text{Id} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.
\] (9.11)

Moreover, by [12, Example 23.4],

\[
J_B = P_{\{0\} \times \mathbb{R}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.
\] (9.12)

Consequently

\[
T = \text{Id} - J_A + J_B R_A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}
\]

\[
= \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},
\] (9.13)

i.e.,

\[
T : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto \frac{x - y}{2}(1, -1).
\] (9.14)

Now let \((x, y) \in \mathbb{R}^2\). Then \((x, y) \in \text{Fix} \, T \iff (x, y) = (\frac{x - y}{2}, -\frac{x - y}{2}) \iff x = \frac{x - y}{2} \text{ and } y = -\frac{x - y}{2} \iff x + y = 0\), hence \( \text{Fix} \, T = \mathbb{R} \cdot (1, -1) \) as claimed. It follows from Fact 5.2(iii) that \( Z = J_A(\text{Fix} \, T) = \)
9.4. The Douglas–Rachford algorithm in the affine case

\[ \mathbb{R} \cdot J_A(1, -1) = \mathbb{R} \cdot \frac{1}{2}(0, 2) = \{0\} \times \mathbb{R}, \] as claimed. Now let \((x, y) \in \mathbb{R}^2\). By (9.14) we have \(T(x, y) = \frac{x-y}{2}(1, -1) \in \text{Fix} T\), hence \((\forall n \geq 1) T^n(x, y) = T(x, y) = \frac{x-y}{2}(1, -1)\). Therefore, \((\forall n \geq 1) J_A T^n(x, y) = J_A T(x, y) = J_A \left( \frac{x-y}{2}(1, -1) \right) = (0, y - x) \neq (0, y) = P_Z(x, y)\) whenever \(x \neq 0\).

In Example 9.10 below, we show that the assumption that \(A\) and \(B\) have common zeros (equivalently, in view of Theorem 4.32, that \(K \perp (Z-Z)\)) is critical in Theorem 9.7(ii) — it cannot not be relaxed to assuming merely (9.7).

**Example 9.10 (when \(K \not\perp (Z-Z)\)).** Let \(u \in X \setminus \{0\}\). Suppose that \(A : X \to X : x \mapsto u\) and \(B : X \to X : x \mapsto -u\). Then \(A\) and \(B\) are paramonotone, \(A + B \equiv 0\) and therefore \(Z = X\). Moreover, by Remark 4.29 \((\forall z \in Z = X) K = (Az) \cap (-Bz) = \{u\} \not\perp (Z-Z) = X\). Note that \(\text{Fix} T = Z + K = X + \{u\} = X\) and \(J_A : X \to X : x \mapsto x - u\). Consequently

\[
(\forall x \in X) (\forall n \in \mathbb{N}) \quad J_A T^n x = J_A P_{\text{Fix} T} x = J_A x = x - u \neq x = P_Z x. \quad (9.15)
\]

Suppose that \(U\) and \(V\) are nonempty closed convex subsets of \(X\). Then

\[
T_{U,V} = T_{(\mathbb{N}u, \mathbb{N}V)} = \text{Id} - P_U + P_V(2P_U - \text{Id}). \quad (9.16)
\]

**Proposition 11.** Suppose that \(U\) and \(V\) are closed linear subspaces of \(X\). Let \(w \in X\). Then \(w + U\) and \(w + V\) are closed affine subspaces of \(X\), \((w + U) \cap (w + V) \neq \emptyset\) and \((\forall n \in \mathbb{N})\)

\[
T_{w+U,w+V}^n = T_{U,V}^n (\cdot - w) + w. \quad (9.17)
\]

**Proof.** Let \(x \in X\). We proceed by induction. The case \(n = 0\) is clear. We now prove the case when \(n = 1\), i.e.,

\[
T_{w+U,w+V} = T_{U,V} (\cdot - w) + w. \quad (9.18)
\]

Indeed, using Fact 2.6 \(T_{w+U,w+V} x = (\text{Id} - P_{w+U} + P_{w+V}(2P_{w+U} - \text{Id})) x = x - w - P_U(x-w) + w + P_V(2P_{w+U}x-x-w) = x - w - P_U(x-w) + w + P_V(2w + 2P_U(x-w) - x - w) = (x - w) - P_U(x-w) + P_V(2P_U(x-w) - (x-w)) + w = (\text{Id} - P_U + P_V R_U)(x-w) + w = T_{U,V} (x-w) + w\). We now assume that (9.17) holds for some \(n \in \mathbb{N}\). Applying (9.18) with \(x\) replaced by \(T_{w+U,w+V}^n x\) yields

\[
T_{w+U,w+V}^n x = T_{w+U,w+V} (T_{w+U,w+V}^{n-1} x) = T_{U,V} (T_{w+U,w+V}^{n-1} x - w) + w
\]
9.4. The Douglas–Rachford algorithm in the affine case

\[ T_{U,V}^{n}(T_{U,V}^{n}(x-w)+w)+w=T_{U,V}^{n+1}(x-w)+w; \quad (9.19) \]

hence (9.17) holds for all \( n \in \mathbb{N} \).

Example 9.12 (Douglas–Rachford in the affine feasibility case). (see also [30, Corollary 4.5]) Suppose that \( U \) and \( V \) are closed linear subspaces of \( X \). Let \( w \in X \) and let \( x \in X \). Suppose that \( A = N_{w+U} \) and that \( B = N_{w+V} \). Then \( T_{w+U,w+V}x = Lx + b \), where \( L = T_{U,V} \) and \( b = w - T_{U,V}w \). Moreover,

\[ T_{w+U,w+V}^{n}x \to P_{\text{Fix} T_{w+U,w+V}}x \quad (9.20) \]

and

\[ I_{A} T_{w+U,w+V}^{n}x = P_{w+U} T_{w+U,w+V}^{n}x \to P_{Z}x = P_{(w+V) \cap (w+U)x}. \quad (9.21) \]

Finally, if \( U + V \) is closed (as is the case when \( X \) is finite-dimensional), then the convergence is linear with rate \( c_{F}(U,V) < 1 \), where \( c_{F}(U,V) \) is the cosine of the Friedrich’s angle\(^\text{10}\) between \( U \) and \( V \).

Proof. Using (9.17) with \( n = 1 \) and the linearity of \( T_{U,V} \) we have

\[ T_{w+U,w+V} = T_{U,V}(\cdot-w)+w=T_{U,V}+w-T_{U,V}w. \quad (9.22) \]

Hence \( L = T_{U,V} \) and \( b = w - T_{U,V}w \), as claimed. To obtain (9.20) and (9.21), use Theorem 9.7(i) and Theorem 9.7(ii), respectively. The claim about the linear rate follows by combining [30, Corollary 4.4] and Theorem 8.23 with \( T \) replaced by \( T_{w+U,w+V} \) and \((L,b) \) replaced by \((T_{U,V},w-T_{U,V}w)\).

Remark 9.13. When \( X \) is infinite-dimensional, it is possible to construct an example (see [30, Section 6]) of two linear subspaces \( U \) and \( V \) where \( c_{F}(U,V) = 1 \), and the rate of convergence of \( T \) is not linear.

The following result complements Combettes’s work [62, Section 2.2] which deals with more general operators; here, we obtain a strong convergence result in the affine setting.

\(^{10}\) Suppose that \( U \) and \( V \) are closed linear subspaces of \( X \). The cosine of the Friedrichs angle is \( c_{F}(\text{par } U, \text{par } V) = \sup_{u \in \text{par } U \cap W \cap \text{ball}(0,1)} |\langle u, v \rangle| < 1 \), where \( W = \text{par } U \cap \text{par } V \).
Proposition 9.14 (parallel splitting). Let \( m \in \{2, 3, \ldots \} \), and let \( B_i : X \rightrightarrows X \) be maximally monotone and affine, \( i \in \{1, 2, \ldots, m\} \), such that \( \text{zer}(\sum_{i=1}^{m} B_i) \neq \emptyset \). Set \( \Delta = \{(x_1, \ldots, x_m) \in X^m \mid x \in X\} \), set \( A = N_\Delta \), set \( B = \times_{i=1}^{m} B_i \), set \( T = T(A, B) \), let \( j : X \to X^m : x \mapsto (x, x, \ldots, x) \), and let \( e : X^m \to X : (x_1, x_2, \ldots, x_m) \mapsto \frac{1}{m} (\sum_{i=1}^{m} x_i) \). Let \( x \in X^m \). Then

\[
\Delta^\perp = \left\{ (u_1, \ldots, u_m) \in X^m \mid \sum_{i=1}^{m} u_i = 0 \right\},
\]

(9.23)

\[
Z = Z_{(A, B)} = j(\text{zer}(\sum_{i=1}^{m} B_i)) \subseteq \Delta
\]

(9.24)

and

\[
K = K_{(A, B)} = (-B(Z)) \cap \Delta^\perp \subseteq \Delta^\perp.
\]

(9.25)

Moreover, the following hold:

(i) \( T^n x \to P_{\text{Fix} T} x \) and \( P_\Delta T^n x \to P_\Delta P_{\text{Fix} T} x \).

(ii) Suppose that \( X \) is finite-dimensional. Then \( T^n x \to P_{\text{Fix} T} x \) linearly and \( J_A T^n x = P_\Delta T^n x \to P_\Delta P_{\text{Fix} T} x \) linearly.

(iii) Suppose that \( B_i : X \rightrightarrows X, i \in \{1, 2, \ldots, m\} \), are paramonotone. Then \( B \) is paramonotone and \( J_A T^n x = P_\Delta T^n x \to P_\text{Fix} T x \). Consequently, \( e(J_A T^n x) = e(P_\Delta T^n x) \to e(P_\text{Fix} T x) \in Z \).

Proof. Both (9.23) and (9.24) follow from [12, Proposition 25.5(i)&(vi)]. On the other hand (9.25) follows from Corollary 4.31(iii) applied to \( (A, B) \). (i): Apply Theorem 9.7(i) to \( (A, B) \). (ii): Apply Theorem 9.7(iii) to \( (A, B) \). (iii): Let \( (x, u), (y, v) \) be in \( \text{gra } B \). On the one hand, \( \langle x - y, u - v \rangle = 0 \iff \sum_{i=1}^{m} (x_i - y_i, u_i - v_i) = 0 \). \( (x_i, u_i), (y_i, v_i) \) are in \( \text{gra } B_i, i \in \{1, \ldots, m\} \). On the other hand, since \( \langle \forall i \in \{1, \ldots, m\} \rangle B_i \) are monotone we learn that \( \langle \forall i \in \{1, \ldots, m\} \rangle (x_i - y_i, u_i - v_i) \geq 0 \). Altogether, \( \langle \forall i \in \{1, \ldots, m\} \rangle (x_i - y_i, u_i - v_i) = 0 \). Now use that paramonotonicity of \( B_i \) to deduce that \( (x_i, v_i), (y_i, u_i) \) are in \( \text{gra } B_i, i \in \{1, \ldots, m\} \); equivalently, \( (x, v), (y, u) \) in \( \text{gra } B \). Finally, apply Example 5.20.

9.5 Eckstein–Ferris–Pennanen–Robinson duality and algorithms

In this last section, we sketch some algorithmic consequences and then conclude by commenting on the applicability of our work to a more general
9.5. Eckstein–Ferris–Pennanen–Robinson duality and algorithms

duality framework. Recall that

\[ T = \frac{1}{2} \Id + \frac{1}{2} R_B R_A = \Id - J_A + J_B R_A, \quad (9.26) \]

and that the set of primal solutions \( Z \) coincides with \( J_A(\text{Fix} T) \) (see Fact 5.2(iii)). This explains the interest in finding fixed points of \( T \).

Moreover, if the nearest primal solution is of interest (e.g., the problem of finding the projection onto the intersection of two nonempty closed convex sets), then the following result may be helpful:

**Theorem 9.15 (abstract algorithm).** Suppose that \( A \) and \( B \) are paramonotone. Let \( (x_n)_{n \in \mathbb{N}} \) be a sequence such that \( (x_n)_{n \in \mathbb{N}} \) converges (weakly or strongly) to \( x \in \text{Fix} T \) and \( (J_Ax_n)_{n \in \mathbb{N}} \) converges (weakly or strongly) to \( J_Ax \). Then the following hold.

(i) \( (\forall k \in K) \ J_Ax = P_Z(x - k) \).

(ii) If \( (Z - Z) \perp K \), then \( J_Ax = P_Zx \).

**Proof.** Combine Corollary 5.16(v) with Theorem 4.40.

We provide three examples.

**Example 9.16 (Douglas–Rachford algorithm).** Suppose that \( A \) and \( B \) are paramonotone and that the sequence \( (x_n)_{n \in \mathbb{N}} \) is generated by \( (\forall n \in \mathbb{N}) \ x_{n+1} = T x_n \). The hypothesis in Theorem 9.15 is satisfied, and the convergence of the sequences is with respect to the weak topology [132]. See also [9] for a much simpler proof and [12, Theorem 25.6] for a powerful generalization.

**Example 9.17 (Halpern-type algorithm).** Suppose that \( A \) and \( B \) are paramonotone and that the sequence \( (x_n)_{n \in \mathbb{N}} \) is generated by \( (\forall n \in \mathbb{N}) \ x_{n+1} = (1 - \lambda_n)T x_n + \lambda_n y \), where \( (\lambda_n)_{n \in \mathbb{N}} \) is a sequence of parameters in \([0,1]\) and \( y \in X \) is given. Under suitable assumptions on \( (\lambda_n)_{n \in \mathbb{N}} \), it is known (see [85], [135]) that \( x_n \to x = P_{\text{Fix} T}y \) with respect to the norm topology. Since \( J_A \) is (firmly) nonexpansive, it is clear that the hypothesis of Theorem 9.15 holds. Furthermore, \( J_Ax_n \to J_Ax = J_A P_{\text{Fix} T}y \). Thus, if \( k_0 \in K \), then \( J_Ax_n \to P_Z(y - k_0) \) by Theorem 4.40(i). And if \( (Z - Z) \perp K \), then \( J_Ax_n \to P_Zy \) by Theorem 4.40(ii).

**Example 9.18 (Haugazeau-type algorithm).** This is similar to Example 9.17 in that \( x_n \to x = P_{\text{Fix} T}y \) with respect to the norm topology and where \( y \in X \) is given. For the precise description of the (somewhat complicated) update
formula for \((x_n)_{n \in \mathbb{N}}\), we refer the reader to [12, Section 29.2] or [11]; see also [86]. Once again, we have \(J_{A_n} \rightarrow J_{A} = J_{A} P_{\text{Fix}T} \) and thus, if \(k_0 \in K\), then \(J_{A_n} \rightarrow P_{Z}(y - k_0)\) by Theorem 4.40(i). And if \((Z - Z) \perp K\), then \(J_{A_n} \rightarrow P_{Z}y\) by Theorem 4.40(ii). Consequently, in the context of Example 5.17, we obtain \(P_{U} \rightarrow P_{U \cap V}\); in fact, this is [21, Theorem 3.3], which is the main result of [21].

Turning to Eckstein–Ferris–Pennanen–Robinson duality, let us assume the following:

- \(Y\) is a real Hilbert space (and possibly different from \(X\));
- \(C\) is a maximally monotone operator on \(Y\);
- \(L: X \rightarrow Y\) is continuous and linear.

Around the turn of the millennium, Eckstein and Ferris [73], Pennanen [118] as well as Robinson [121] considered the problem of finding zeros of

\[ A + L^* CL. \]  \hspace{1cm} (9.27)

This framework is more flexible than the Attouch–Théra framework, which corresponds to the case when \(Y = X\) and \(L = \text{Id}\). (For an even more general framework, see [66].) Note that just as Attouch–Théra duality relates to classical Fenchel duality in the subdifferential case (see Section 4.6), the Eckstein–Ferris–Pennanen–Robinson duality pertains to the classical Fenchel–Rockafellar duality for the problem of minimizing \(f + h \circ L\) when \(f \in \Gamma_X\) and \(h \in \Gamma_Y\), and \(A = \partial f\) and \(C = \partial h\).

The results in the previous sections can be used in the Eckstein–Ferris–Pennanen–Robinson framework thanks to items (ii) and (iii) of the following result, which allows us to set \(B = L^* CL\).

**Proposition 9.19.** The following hold.

(i) If \(C\) is paramonotone, then \(L^* CL\) is paramonotone.

(ii) (Pennanen) If \(\mathbb{R}_{++}(\text{ran } L - \text{ dom } C)\) is a closed subspace of \(Y\), then \(L^* CL\) is maximally monotone.

(iii) If \(C\) is paramonotone and \(\mathbb{R}_{++}(\text{ran } L - \text{ dom } C)\) is a closed subspace of \(Y\), then \(L^* CL\) is maximally monotone and paramonotone.
9.6. Brief history

Proof. (i): Take $x_1$ and $x_2$ in $X$, and suppose that $x_1^* \in \text{CL}x_1$ and $x_2^* \in \text{CL}x_2$. Then there exist $y_1^* \in \text{CL}x_1$ and $y_2^* \in \text{CL}x_2$ such that $x_1^* = L^*y_1^*$ and $x_2^* = L^*y_2^*$. Thus,

$$\langle x_1 - x_2, x_1^* - x_2^* \rangle = \langle x_1 - x_2, L^*y_1^* - L^*y_2^* \rangle = \langle Lx_1 - Lx_2, y_1^* - y_2^* \rangle \geq 0$$

(9.28) because $C$ is monotone. Hence $L^*\text{CL}$ is monotone. Now suppose furthermore that $\langle x_1 - x_2, x_1^* - x_2^* \rangle = 0$. Then $\langle Lx_1 - Lx_2, y_1^* - y_2^* \rangle = 0$ and the paramonotonicity of $C$ yields $y_2^* \in \text{CL}(Lx_1)$ and $y_1^* \in \text{CL}(Lx_2)$. Therefore, $x_2^* = L^*y_2^* \in L^*\text{CL}x_1$ and $x_1^* = L^*y_1^* \in L^*\text{CL}x_2$.

(ii): See [118, Corollary 4.4.(c)].

(iii): Combine (i) and (ii).

9.6 Brief history

The Douglas–Rachford algorithm has its roots in the 1956 paper [70] as a method for solving a system of linear equations, where the matrices are symmetric and positive semi-definite. In 1969 Lieutaud (see [95]) extended their method to deal with (possibly nonlinear) maximally monotone operators that are defined everywhere. Lions and Mercier, in their paper [96] from 1979, presented a broad and powerful generalization to its current form, i.e., to handle the sum of any two maximally monotone operators that are possibly nonlinear, possibly set-valued and not necessarily defined everywhere. (For details on this connection we refer the reader to [95] and [61].) In their seminal work, they showed that, for every $x \in X$, $(T^n x)_{n \in \mathbb{N}}$ converges weakly to a point in Fix $T$ and that the bounded shadow sequence $(J_A T^n x)_{n \in \mathbb{N}}$ has all its weak cluster points in zer$(A + B)$ provided that $A + B$ was maximally monotone. (Note that resolvents are not weakly continuous in general; see [139] or [12, Example 4.12].) In their joint work from 1992, Eckstein and Bertsekas proved that $J_A(\text{Fix } T) \subseteq \text{zer}(A + B)$ (see [72, Theorem 5]). Later on, in 2006, Combettes refined the results by Eckstein and Bertsekas by providing the first characterization of the set of zeros of the sum to be precisely the shadows of the fixed points of $T$, namely $J_A(\text{Fix } T)$ (see [61, Lemma 2.6(iii)]). In the finite-dimensional setting, together with the earlier results by Lions and Mercier [96], the work by Eckstein and Bertsekas and later by Combettes, asserts the convergence of the shadow sequence to a solution of the sum (without assuming maximality). Working with the shadow sequence, we point out that explicit proof of weak convergence of the shadow sequence under
additional assumptions and strong convergence results in special cases can be found in [62]. The first proof that the weak cluster points of the shadow sequence are zeros of $A + B$ (without assuming the maximality of the sum) in the convex feasibility setting appeared in [18, Fact 5.9] in 2002. Building on [8] and [74], Svaiter provided a complete answer in 2011 (see [132]) demonstrating that $A + B$ does not have to be maximally monotone and that the shadow sequence $(J_{A^T}x)_{n \in \mathbb{N}}$ in fact does converge weakly to a point in $\text{zer}(A + B)$. (He used Theorem 9.4; however, his proof differs from ours which is more in the style of the original paper by Lions and Mercier [96].) Nonetheless, when $Z = \emptyset$, the complete understanding of $(J_{A^T}x)_{n \in \mathbb{N}}$ remains open — to the best of our knowledge, Theorem 14.10 below is currently the most powerful result available.
Chapter 10

On the order of the operators

10.1 Overview

By definition, the Douglas–Rachford splitting operator associated with the ordered pair of operators \((A, B)\) is dependent on the order of the operators \(A\) and \(B\), even though the sum problem remains unchanged when interchanging \(A\) and \(B\). The goal of this chapter is to investigate the connection between the operators \(T_{(A,B)}\) and \(T_{(B,A)}\). Our main results can be summarized as follows.

- We show that \(R_A\) is an isometric bijection from the fixed points set of \(T_{(A,B)}\) to that of \(T_{(B,A)}\), with inverse \(R_B : \text{Fix } T_{(B,A)} \rightarrow \text{Fix } T_{(A,B)}\) (see Theorem 10.2).

- When \(A\) is an affine relation, we have \((\forall n \in \mathbb{N}) R_A T^n_{(A,B)} = T^n_{(B,A)} R_A\). In particular when \(A = N_U\) where \(U\) is a closed affine subspace of \(X\), we have \((\forall n \in \mathbb{N}) T^n_{(A,B)} = R_A T^n_{(B,A)} R_A\) and \(T^n_{(B,A)} = R_A T^n_{(A,B)} R_A\) (see Proposition 10.5(i) and Theorem 10.8(i)).

- Our results connect to the recent linear and finite convergence results (see Remark 10.11) for the Douglas–Rachford algorithm (see [1], [2], [37], [34], [87] and [88]).

The results in this chapter are mainly based on the work published in [14].

10.2 Connection between \(\text{Fix } T_{(A,B)}\) and \(\text{Fix } T_{(B,A)}\)

In view of Definition 4.3, one easily verifies that

\[
Z_{(B,A)} = (B + A)^{-1}(0) = Z \quad \text{(10.1)}
\]

and

\[
K_{(B,A)} = (B^{-1} + A^{-\ominus})^{-1}(0) = -K. \quad \text{(10.2)}
\]
10.2. Connection between $\text{Fix } T_{(A,B)}$ and $\text{Fix } T_{(B,A)}$

We further recall (see ?? and ??) that

$$Z = J_A(\text{Fix } T_{(A,B)}) \quad \text{and} \quad K = (\text{Id} - J_A)(\text{Fix } T_{(A,B)}),$$

and we will make use of the following useful lemma.

**Lemma 10.1.** We have

$$R_A T_{(A,B)} - T_{(B,A)} R_A = 2J_A T_{(A,B)} - J_A - J_A R_B R_A. \quad (10.4)$$

**Proof.** Using (5.2) we have

$$R_A T_{(A,B)} - T_{(B,A)} R_A = 2J_A T_{(A,B)} - T_{(A,B)} - R_A + J_B R_A - J_A R_B R_A$$

$$= 2J_A T_{(A,B)} - \text{Id} + J_A - J_B R_A - 2J_A$$

$$+ \text{Id} + J_B R_A - J_A R_B R_A$$

$$= 2J_A T_{(A,B)} - J_A - J_A R_B R_A. \quad \blacksquare$$

We are now ready for the first main result in this chapter.

**Theorem 10.2.** $R_A$ is an isometric bijection from $\text{Fix } T_{(A,B)}$ to $\text{Fix } T_{(B,A)}$, with isometric inverse $R_B$. Moreover, we have the following commutative diagram:

$$\begin{array}{ccc}
\text{Fix } T_{(A,B)} & \xrightarrow{R_A} & \text{Fix } T_{(B,A)} \\
\downarrow & & \downarrow \\
S_{(A,B)} & \xrightarrow{(I_d, \text{Id} - J_A) \circ \Delta} & \text{Id} \times (\text{Id}) & \xrightarrow{(I_d, \text{Id} - J_B) \circ \Delta} & S_{(B,A)} \\
\end{array}$$

Here $S_{(A,B)} = \{(z, -w) \in X \times X \mid -w \in Bz, w \in Az\}$ is the Kuhn–Tucker set\(^{11}\) for the pair $(A, B)$, and $\Delta: X \to X \times X: x \mapsto (x, x)$. In particular, we have

$$R_A: \text{Fix } T_{(A,B)} \to \text{Fix } T_{(B,A)}: z + k \mapsto z - k,$$

where $(z, k) \in S_{(A,B)}$.

\(^{11}\)For further information on the Kuhn–Tucker set, we refer the reader to [74, Section 2.1].
10.3. Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$

Proof. Let $x \in X$ and note that (5.2) implies that $\text{Fix } T_{(A,B)} = \text{Fix } R_BR_A$ and $\text{Fix } T_{(B,A)} = \text{Fix } R_AR_B$. Now $x \in \text{Fix } T_{(A,B)} \iff x = R_BR_Ax \iff R_Ax = R_AR_Bx \iff R_Ax \in \text{Fix } R_AR_B = \text{Fix } T_{(B,A)}$, which proves that $R_A$ maps $T_{(A,B)}$ into $\text{Fix } T_{(B,A)}$. By interchanging $A$ and $B$ one sees that $R_B$ maps $\text{Fix } T_{(B,A)}$ into $\text{Fix } T_{(A,B)}$. We now prove the claim that $R_A$ maps $\text{Fix } T_{(A,B)}$ onto $\text{Fix } T_{(B,A)}$. To this end, let $y \in 
\text{Fix } T_{(B,A)}$ and note that $R_By \in \text{Fix } T_{(A,B)}$ and $R_AR_By = y$, which proves that $R_A$ maps $\text{Fix } T_{(A,B)}$ onto $\text{Fix } T_{(B,A)}$. The same argument holds for $R_B$. Finally since $(\forall x \in \text{Fix } T_{(A,B)}) R_BGR_Ax = x$, this proves that $R_A$ is a bijection from $\text{Fix } T_{(A,B)}$ to $\text{Fix } T_{(B,A)}$ with the desired inverse. To prove that $R_A : \text{Fix } T_{(A,B)} \to \text{Fix } T_{(B,A)}$ is an isometry note that $(\forall x \in \text{Fix } T_{(A,B)}) (\forall y \in \text{Fix } T_{(B,A)})$ we have $\|x - y\| = \|R_BR_Ax - R_BR_By\| \leq \|R_AR_Ax - R_AR_Ay\| \leq \|x - y\|$.

We now turn to the diagram. The correspondence of $\text{Fix } T_{(A,B)}$ and $\text{Fix } T_{(B,A)}$ follows from our earlier argument. On the other hand, the correspondences of $\text{Fix } T_{(A,B)}$ and $\text{S}_{(A,B)}$, and $\text{Fix } T_{(B,A)}$ and $\text{S}_{(B,A)}$ follow from combining Remark 4.24 and Theorem 5.14 applied to $T_{(A,B)}$ and $T_{(B,A)}$ respectively. The fourth correspondence is obvious from the definition of $\text{S}_{(A,B)}$ and $\text{S}_{(B,A)}$. To prove (10.5) we let $y \in \text{Fix } T_{(A,B)}$ and recall that in view of Theorem 5.14 and Remark 4.24, that $y = z + k$ where $(z,k) \in \text{S}_{(A,B)}$ and $R_A(z + k) = (J_A - (\text{Id} - J_A))(z + k) = J_A(z + k) - (\text{Id} - J_A)(z + k) = z - k$, which completes the proof of the theorem. 

Remark 10.3. In view of Remark 4.24, Theorem 5.14 and Corollary 5.16(\#), when $A$ and $B$ are paramonotone (as is always the case when $A$ and $B$ are subdifferential operators of functions in $\Gamma(X)$), we can replace $\text{S}_{(A,B)}$ and $\text{S}_{(B,A)}$ by, respectively, $Z \times K$ and $Z \times (-K)$.

10.3 Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$

Lemma 10.4. Suppose that $A$ is an affine relation. Then

(i) $J_A$ is affine and $J_AR_A = 2J_A^2 - J_A = R_AJ_A$.

If $A = N_U$, where $U$ is a closed affine subspace of $X$, then we have additionally:

(ii) $P_U = J_A = J_AR_A = R_AR_A$ and $(\text{Id} - J_A)R_A = J_A - \text{Id}$.

(iii) $R_A^2 = \text{Id}$, $R_A = R_A^{-1}$, and $R_A : X \to X$ is an isometric bijection.

Proof. (i): The fact that $J_A$ is affine follows from [27, Theorem 2.1(xix)]. Hence $J_AR_A = J_A(2J_A - \text{Id}) = 2J_A^2 - J_A = R_AJ_A$. 

117
10.3. Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$

(ii): It follows from Example 3.14(ii) that $P_U = J_A$. Now using (i) we have $R_AJ_A = J_AR_A = 2P^2_U - P_U = 2P_U - P_U = P_U = J_A$. To prove the last identity note that by (i) we have $(\text{Id} - J_A)R_A = R_A - J_AR_A = 2P_U - \text{Id} + P_U = P_U - \text{Id}.$

(iii): Because $R_A$ is affine, it follows from (ii) that $R^2_A = R_A(2J_A - \text{Id}) = 2R_AJ_A - R_A = 2P_U - 2(P_U - \text{Id}) = \text{Id}.$ Finally let $x, y \in X$. Since $R_A$ is nonexpansive we have $\|x - y\| = \|R^2_Ax - R^2_Ay\| \leq \|R_Ax - R_Ay\| \leq \|x - y\|$, hence all the inequalities become equalities which completes the proof. ■

We now turn to the iterates of the Douglas–Rachford algorithm.

**Proposition 10.5.** Suppose that $A$ is an affine relation. Then the following hold:

(i) $(\forall n \in \mathbb{N})$ we have $R_AT_{(A,B)}^n = T_{(B,A)}^nR_A$.

(ii) $R_AZ = J_A \text{Fix } T_{(B,A)}$ and $R_AK = (J_A - \text{Id})(- \text{Fix } T_{(B,A)})$.

If $B$ is an affine relation, then we additionally have:

(iii) $T_{(A,B)}R_BR_A = R_BR_AT_{(A,B)}$.

(iv) $4(T_{(A,B)}T_{(B,A)} - T_{(B,A)}T_{(A,B)}) = R_BR_A^2 - R_A^2R_B$. Consequently,

$$T_{(A,B)}T_{(B,A)} = T_{(B,A)}T_{(A,B)} \iff R_BR_A^2 = R_A^2R_B.$$ \hspace{1cm} (10.6)

(v) If $R_A^2 = R_B^2 = \text{Id}$, then $T_{(A,B)}T_{(B,A)} = T_{(B,A)}T_{(A,B)}$.

**Proof.** (i): It follows from (10.4), Lemma 10.4(i) and (5.2) that

$$R_AT_{(A,B)} - T_{(B,A)}R_A = 2J_AT_{(A,B)} - J_A - J_AR_BR_A$$ \hspace{1cm} (10.7a)

$$= J_A(2T_{(A,B)} - \text{Id}) - J_AR_BR_A$$ \hspace{1cm} (10.7b)

$$= J_A(2(\frac{1}{2}(\text{Id} + R_BR_A)) - \text{Id}) - J_AR_BR_A$$ \hspace{1cm} (10.7c)

$$= J_AR_BR_A - J_AR_BR_A = 0,$$ \hspace{1cm} (10.7d)

which proves the claim when $n = 1$. The general proof follows by induction.

(ii): Using (10.3), Lemma 10.4(i) and Theorem 10.2, we have

$$R_AZ = R_AJ_A(\text{Fix } T_{(A,B)}) = J_AR_A(\text{Fix } T_{(A,B)}) = J_A(\text{Fix } T_{(B,A)}).$$ \hspace{1cm} (10.8)
By the inverse resolvent identity (3.15) we easily deduce that \( R_{A^{-1}} = -R_A \). Therefore using Lemma 10.4(i) applied to \( A^{-1} \) and Theorem 10.2, we obtain

\[
R_A K = -R_{A^{-1}}J_{A^{-1}}(\text{Fix} T_{(A,B)}) = -J_{A^{-1}}R_{A^{-1}}(\text{Fix} T_{(A,B)}) = -J_{A^{-1}}(-R_A \text{ Fix} T_{(A,B)}) = (J_A - \text{Id})(- \text{Fix} T_{(B,A)}).
\] (10.9a, 10.9b)

(iii): Note that \( T_{(A,B)} \) and \( T_{(B,A)} \) are affine. It follows from (5.2) that

\[
T_{(A,B)}R_B R_A = T_{(A,B)}(2T_{(A,B)} - \text{Id}) = 2T_{(A,B)}^2 - T_{(A,B)} = (2T_{(A,B)} - \text{Id})T_{(A,B)} = R_B R_A T_{(A,B)}.
\] (10.10a, 10.10b)

(iv) We have

\[
4(T_{(A,B)}T_{(B,A)} - T_{(B,A)}T_{(A,B)}) = 4\left(\frac{1}{2}(\text{Id} + R_B R_A)\right)^2(\text{Id} + R_A R_B) - \frac{1}{2}(\text{Id} + R_A R_B)^2(\text{Id} + R_B R_A)
\] (10.11a)

\[
= \text{Id} + R_B R_A + R_A R_B + R_B R_A^2 R_B - (\text{Id} + R_A R_B + R_B R_A + R_A R_B^2 R_A).
\] (10.11b)

(v): This is a direct consequence of (iv).

**Remark 10.6.** In passing, we point out that the assumption in Proposition 10.5(v) is equivalent to saying that \( A = N_U \) and \( B = N_V \) where \( U \) and \( V \) are closed affine subspaces of \( X \). Indeed, \( R_A^2 = \text{Id} \Leftrightarrow J_{A} = J_{A}^2 \) and therefore we conclude that \( \text{ran} J_A = \text{Fix} J_A \). Combining with [139, Theorem 1.2] yields that \( J_A \) is a projection, hence \( A \) is an affine normal cone operator using Example 3.14(ii).

With regards to Proposition 10.5(i), one may inquire whether the conclusion still holds when \( R_A \) is replaced by \( R_B \). We now give an example illustrating that the answer to this question is negative.

**Example 10.7.** Suppose that \( X = \mathbb{R}^2 \), that \( U = \mathbb{R} \times \{0\} \), that \( V = \{0\} \times \mathbb{R}_+ \), that \( A = N_U \) and that \( B = N_V \). Then \( A \) is linear, hence \( R_A T_{(A,B)} = T_{(B,A)}R_A \), however \( R_B T_{(A,B)} \neq T_{(B,A)}R_B \) and \( R_B T_{(B,A)} \neq T_{(A,B)}R_B \).

**Proof.** To prove the identity \( R_A T_{(A,B)} = T_{(B,A)}R_A \) apply Proposition 10.5(i) with \( n = 1 \). Now let \((x,y) \in \mathbb{R}^2 \) and set for every \( x \in \mathbb{R} \), \( x^+ = \max\{x,0\} \) and \( x^- = \min\{x,0\} \). Elementary calculations show that \( R_A(x,y) = (x,-y) \) and \( R_B(x,y) = (-x,|y|) \). Consequently, (5.2) implies that \( T_{(A,B)}(x,y) = (0,y^+) \) and \( T_{(B,A)}(x,y) = (0,y^-) \). Therefore,

\[
R_B T_{(A,B)}(x,y) = (0,y^+),
\] (10.12a)
10.3. 

\[ T_{(A,B)}R_B(x, y) = (0, 0), \]  
\[ R_B T_{(B,A)}(x, y) = (0, -y^+), \]  
\[ T_{(A,B)}R_B(x, y) = (0, |y|). \]  

(10.12b)  
(10.12c)  
(10.12d)

The conclusion follows from comparing equations (10.12a)–(10.12d).

We are now ready for our second main result.

**Theorem 10.8 (When A is normal cone of closed affine subspace).** Suppose that \( U \) is a closed affine subspace and that \( A = N_U \). Then the following hold:

(i) \( \forall n \in \mathbb{N} \) \( R_A T^n_{(B,A)} = T^n_{(A,B)} R_A \), \( T^n_{(B,A)} = R_A T^n_{(B,A)} R_A \) and \( T^n_{(A,B)} = R_A T^n_{(B,A)} R_A \).

(ii) \( R_A : \text{Fix} \, T_{(B,A)} \to \text{Fix} \, T_{(A,B)}, Z = J_A(\text{Fix} \, T_{(B,A)}) \) and \( K = (J_A - \text{Id})(\text{Fix} \, T_{(B,A)}) \).

(iii) Suppose that \( V \) is a closed affine subspace of \( X \) and that \( B = N_V \). Then \( T_{(A,B)} R_A R_B = R_A R_B T_{(B,A)} \) and \( T_{(A,B)} T_{(B,A)} = T_{(B,A)} T_{(A,B)} \).

**Proof.** (i): Let \( n \in \mathbb{N} \). It follows from Proposition 10.5(i) and Lemma 10.4(iii) that \( T^n_{(A,B)} = R_A T^n_{(B,A)} = R_A T^n_{(B,A)} R_A \). Hence \( T^n_{(A,B)} R_A = R_A T^n_{(B,A)} R_A R_A = R_A T^n_{(B,A)} \).

(ii): The statement for \( R_A \) follows from combining Theorem 10.2 and Lemma 10.4(iii). In view of (10.3), Lemma 10.4(ii) and Theorem 10.2 one learns that \( Z = J_A(\text{Fix} \, T_{(A,B)}) = J_A R_A(\text{Fix} \, T_{(A,B)}) = J_A(\text{Fix} \, T_{(B,A)}) \). Finally, (10.3), Lemma 10.4(iii) and (ii), and Theorem 10.2 imply that

\[ K = (\text{Id} - J_A)(\text{Fix} \, T_{(A,B)}) = (\text{Id} - J_A) R_A(\text{Fix} \, T_{(A,B)}) = (J_A - \text{Id})(\text{Fix} \, T_{(B,A)}). \]  
\[ = (J_A - \text{Id}) \text{Fix} \, T_{(B,A)}. \]  

(10.13a)  
(10.13b)

(iii): In view of (i) applied to \( A \) and \( B \) we have \( T_{(A,B)} R_A R_B = R_A T_{(B,A)} R_B = R_A R_B T_{(A,B)} \). The second identity now follows from combining Proposition 10.5(v) and Lemma 10.4(iii) applied to both \( A \) and \( B \).
10.3. *Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$*

Figure 10.1: A GeoGebra [78] snapshot. Two closed convex sets in $\mathbb{R}^2$, $U$ is a linear subspace (green line) and $V$ (the ball). Shown are also the first five terms of the sequences $(T^n_{(A,B)} R_A x_0)_{n \in \mathbb{N}}$ (red points) and $(T^n_{(B,A)} x_0)_{n \in \mathbb{N}}$ (blue points) in each case.

Figure 10.2: A GeoGebra [78] snapshot. Two closed convex sets in $\mathbb{R}^2$, $U$ is the halfspace (cyan region) and $V$ (the ball). Shown are also the first five terms of the sequences $(T^n_{(A,B)} R_A x_0)_{n \in \mathbb{N}}$ (red points) and $(T^n_{(B,A)} x_0)_{n \in \mathbb{N}}$ (blue points) in each case.
10.3. Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$

Figure 10.1 illustrates Theorem 10.8(i) while Figure 10.2 illustrates the failure of this result when the subspace is replaced by a cone.

The conclusion of Theorem 10.8(iii) may fail when we assume that $A$ or $B$ is an affine, but not a normal cone operator, as we illustrate next.

**Example 10.9.** Suppose that $X = \mathbb{R}^2$, that $U = \mathbb{R} \times \{0\}$, that

$$A = N_U \text{ and that } B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (10.14)$$

Then $B$ is linear and maximally monotone but not a normal cone operator and

$$\frac{1}{2} \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} = T_{(A,B)} T_{(B,A)} \neq T_{(B,A)} T_{(A,B)} = \frac{1}{2} \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}. \quad (10.15)$$

**Proof.** Maximal monotonicity of $B$ follows from [12, Example 20.19 and Example 20.29]. Moreover, since $B$ is nonzero and single-valued it cannot be a normal cone operator. Now on the one hand, using Lemma 10.4(ii) we have

$$J_A = P_U = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ hence } R_A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (10.16)$$

On the other hand one readily verifies that

$$J_B = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \text{ hence } R_B = \frac{1}{3} \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}. \quad (10.17)$$

Therefore,

$$T_{(A,B)} = \frac{1}{2}(\text{Id} + R_B R_A) = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}, \quad (10.18)$$

and

$$T_{(B,A)} = \frac{1}{2}(\text{Id} + R_A R_B) = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}, \quad (10.19)$$

which completes the proof. ■

**Corollary 10.10.** Suppose that $U$ is an affine subspace, that $A = N_U$ and that $Z \neq \varnothing$. Let $x$ and $y$ be in $X$. Then the following hold:

(i) $$(\forall n \in \mathbb{N}) \ J_A T_{(B,A)}^n x = J_A T_{(A,B)}^n R_A x. \quad (10.20)$$

Furthermore, $(J_A T_{(B,A)}^n x)_{n \in \mathbb{N}}$ converges weakly to a point in $Z$.  

122
10.3. Iterates of \( T_{(A,B)} \) vs. iterates of \( T_{(B,A)} \)

(ii) \( \| T_{(A,B)} x - T_{(A,B)} y \| = \| T_{(B,A)} R_A x - T_{(B,A)} R_A y \| \leq \| R_A x - R_A y \|. \)

Proof. (i): It follows from applying Lemma 10.4(iii), Proposition 10.5(i) and Lemma 10.4(ii) that

\[
J_A T_{(B,A)}^n x = J_A T_{(B,A)}^n R_A R_A x = J_A R_A T_{(A,B)}^n R_A x = J_A T_{(A,B)}^n R_A x, \quad (10.20)
\]
as claimed. The convergence of the sequence \( (J_A T_{(B,A)}^n x)_{n \in \mathbb{N}} \) follows from [12, Theorem 25.6].

(ii): Apply Lemma 10.4(iii) with \((x, y)\) replaced by \((T_{(A,B)} x, T_{(A,B)} y)\), Proposition 10.5(i) with \(n = 1\), and use nonexpansiveness of \( T_{(B,A)} \).

\[\blacksquare\]

Remark 10.11.

(i) The results of Theorem 10.8 and Corollary 10.10 are of interest when the Douglas–Rachford method is applied to find a zero of the sum of more than two operators in which case one can use a parallel splitting method (see [12, Proposition 25.7]), where one operator is the normal cone operator of the diagonal subspace in a product space.

(ii) A second glance at the proof of Theorem 10.8(i) reveals that the result remains true if \( J_B \) is replaced by any operator \( Q_B : X \to X \) (and \( R_B \) is replaced by \( 2Q_B - \text{Id} \), of course). This is interesting because in [1], [2], [87] and [88], \( Q_B \) is chosen to be a selection of the (set-valued) projector onto a set \( V \) that is not convex. Hence the generalized variant of Theorem 10.8(i) then guarantees that the orbits of the two Douglas–Rachford operators are related via

\[
(\forall n \in \mathbb{N}) \quad T_{(B,A)}^n = R_A T_{(A,B)}^n R_A. \quad (10.21)
\]

(iii) As a consequence of (ii) and Lemma 10.4(iii), we see that if linear convergence is guaranteed for the iterates of \( T_{(A,B)} \), then the same holds true for the iterates of \( T_{(B,A)} \) provided that \( U \) is a closed affine subspace, \( V \) is a nonempty closed set, \( A = N_U \) and \( J_B \) is a selection of the projection onto \( V \). This is not particularly striking when we compare to sufficient conditions that are already symmetric in \( A \) and \( B \) (such as, e.g., \( \text{ri} \; U \cap \text{ri} \; V \neq \emptyset \) in [34] and [119]); however, this is a new insight when the sufficient conditions are not symmetric (as in, e.g., [1], [43], [87] and [88]).

(iv) A comment similar to (iii) can be made for finite convergence results; see [130] and [37] for nonsymmetric Slater’s type \((U \cap \text{int} \; V)\) sufficient conditions.
10.3. Iterates of $T_{(A,B)}$ vs. iterates of $T_{(B,A)}$

We now turn to the Borwein–Tam method [44].

**Proposition 10.12.** Suppose that $U$ is an affine subspace of $X$, that $A = N_U$, and set

$$T_{[A,B]} = T_{(A,B)}T_{(B,A)}. \quad (10.22)$$

Then the following holds:

(i) $T_{[A,B]} = R_AR_{[B,A]}R_A = (T_{(A,B)}R_A)^2 = (R_AT_{(B,A)})^2$.

(ii) Suppose that $V$ is an affine subspace and that $B = N_V$. Then $T_{[A,B]} = T_{[B,A]}$. Consequently $T_{[A,B]} = (R_BT_{(A,B)} = (T_{(B,A)}R_B)^2 = \frac{1}{2}(T_{(A,B)} + T_{(B,A)}), \text{ and } T_{[A,B]}$ is firmly nonexpansive.

**Proof.** (i): Using (10.22) and Theorem 10.8(i) with $n = 1$ we obtain

$$T_{[A,B]} = R_AR_{[B,A]}R_AT_{(B,A)}R_A = R_AT_{(B,A)}R_A = (T_{(A,B)}R_A)^2 = (R_AT_{(B,A)})^2. \quad (10.23)$$

(ii): The identity $T_{[A,B]} = T_{[B,A]}$ follows from Theorem 10.8(iii) and (10.22). Now combine with (i) with $A$ and $B$ switched, and use [44, Remark 4.1]. The fact that $T_{[A,B]}$ (hence $T_{[B,A]}$) is firmly nonexpansive follows from the firm nonexpansiveness of $T_{(A,B)}$ and $T_{(B,A)}$ and the fact that the class of firmly nonexpansive operators is closed under convex combinations (see [12, Example 4.31]).

Following [44], the Borwein–Tam method specialized to two nonempty closed convex subsets $U$ and $V$ of $X$, iterates the operator $T_{[A,B]}$ of (10.22), where $A = N_U$ and $B = N_V$. We conclude with an example that shows that if $A$ or $B$ is not an affine normal cone operator, then $T_{[A,B]}$ and $T_{[B,A]}$ need not be firmly nonexpansive.

**Example 10.13.** Suppose that $X = \mathbb{R}^2$, that $U = \mathbb{R}_+ \cdot (1,1)$, that $V = \mathbb{R} \times \{0\}$, that $A = N_U$ and that $B = N_V$. Then neither $T_{[A,B]}$ nor $T_{[B,A]}$ is firmly nonexpansive.

**Proof.** Applying Example 3.14(ii) to $A$ and $B$ respectively we verify that $J_A = P_U : \mathbb{R}^2 \to \mathbb{R}^2 : (x,y) \mapsto \frac{1}{2} \max\{0,(x+y)\} \cdot (1,1)$ and $J_B = P_V : \mathbb{R}^2 \to \mathbb{R}^2 : (x,y) \mapsto (x,0)$. Consequently

$$R_A : \mathbb{R}^2 \to \mathbb{R}^2 : (x,y) \mapsto (\max\{-x,y\}, \max\{x,-y\}) \quad (10.24)$$

12See [30, Proposition 3.5] for the case when $U$ and $V$ are linear subspaces.
and
\[ R_B : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (x, -y). \] (10.25)

This implies that
\[ R_A R_B : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto \begin{cases} (-x, y), & x - y \leq 0; \\ (-y, x), & x - y > 0, \end{cases} \] (10.26)

and
\[ R_B R_A : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto \begin{cases} (-x, y), & x + y \leq 0; \\ (y, -x), & x + y > 0. \end{cases} \] (10.27)

Let \((x, y) \in \mathbb{R}^2\). Using (5.2) we verify that
\[ T_{A,B}(x, y) = \left( \frac{1}{2}(x + y)^+, y - \frac{1}{2}(x + y)^+ \right) \] and \(T_{B,A}(x, y) = \left( \frac{1}{2}(x - y)^+, y + \frac{1}{2}(x - y)^+ \right)\).

Now let \(\alpha > 0\), let \(x = (-2\alpha, 2\alpha)\) and let \(y = (0, 0)\). A routine calculation shows that \(T_{A,B}x = T_{A,B}T_{B,A}(-2\alpha, 2\alpha) = (\alpha, \alpha)\) and \(T_{A,B}y = T_{A,B}T_{B,A}(0, 0) = (0, 0)\), hence
\[ \langle T_{A,B}x - T_{A,B}y, (\text{Id} - T_{A,B})x - (\text{Id} - T_{A,B})y \rangle = \langle (\alpha, \alpha), (-3\alpha, \alpha) \rangle = -2\alpha^2 < 0. \] (10.28)

Applying similar argument to \(T_{B,A}\) with \(x = (-2\alpha, -2\alpha)\) and \(y = (0, 0)\) shows that
\[ \langle T_{B,A}x - T_{B,A}y, (\text{Id} - T_{B,A})x - (\text{Id} - T_{B,A})y \rangle = \langle (-3\alpha, -\alpha), (\alpha, -\alpha) \rangle = -2\alpha^2 < 0. \] (10.29)

It then follows from Fact 3.2(v) that neither \(T_{A,B}\) nor \(T_{B,A}\) is firmly nonexpansive. □
Chapter 11

Generalized solutions for the sum of two maximally monotone operators

11.1 Overview

A common theme in mathematics is to define generalized solutions to deal with problems that potentially do not have solutions. A classical example is the introduction of least squares solutions via the normal equations associated with a possibly infeasible system of linear equations.

Our goal in this chapter is to define a normal problem associated with the original problem of finding a zero of the sum of two maximally monotone operators that has attractive and useful properties. Similarly to the complete extension of classical linear equations via normal equations (see Section 11.2), our proposed approach achieves the following:

- If the original problem has a solution, then so does the normal problem and the sets of solutions to these problems coincide.
- The normal problem may have a solution even if the original problem does not have any.
- The solutions of the normal problem have a variational interpretation as minimal displacement solutions related to the Douglas–Rachford splitting operator.
- The normal problem interacts well with Attouch–Théra duality.

New results presented in this chapter are published in [31] and [38].
11.2 A motivation from Linear Algebra

A classical problem rooted in Linear Algebra and of central importance in the natural sciences is to solve a system of linear equations, say

\[ Ax = b. \]  \hspace{1cm} (11.1)

However, it may occur (due to noisy data, for instance) that (11.1) does not have a solution. An ingenious approach to cope with this situation, dating back to Carl Friedrich Gauss and his famous prediction of the asteroid Ceres (see [41, Subsection 1.1.1] and [102, Epilogue in Section 4.6]) in 1801, is to consider the normal equation associated with (11.1), namely

\[ A^*Ax = A^*b, \]  \hspace{1cm} (11.2)

where \( A^* \) denotes the transpose of \( A \). The normal equation (11.2) has extremely useful properties:

- If the original system (11.1) has a solution, then so does the associated system (11.2); furthermore, the sets of solutions of these two systems coincide in this case.

- The associated system (11.2) always has a solution.

- The solutions of the normal equations have a variational interpretation as least squares solutions: they are the minimizers of the function \( x \mapsto \| Ax - b \|^2 \).

From now on, throughout the rest of this chapter, we assume that

\[ A \text{ and } B \text{ are maximally monotone operators on } X. \]  \hspace{1cm} (11.3)

Recall that

\[ (A, B)^* = (A^{-1}, B^{-\ominus}), \]  \hspace{1cm} (11.4)

that (see Definition 4.1) the primal problem associated with the (ordered) pair \((A, B)\) is to determine the set of zeros of the sum,

\[ Z = Z_{(A, B)} = (A + B)^{-1}(0), \]  \hspace{1cm} (11.5)

and that the (Attouch–Théra) dual problem (see Definition 4.2) associated with the pair \((A, B)\) is to determine the set of zeros of the sum,

\[ K = K_{(A, B)} = (A^{-1} + B^{-1})^{-\ominus}(0), \]  \hspace{1cm} (11.6)

where \( A^\ominus = (- \text{Id}) \circ A \circ (- \text{Id}) \).
11.3 Perturbation calculus

Recall that (see Section 4.3) the dual and primal solution mappings associated with \((A, B)\) are

\[
K: X \ni x \mapsto (Ax) \cap (-Bx) \quad (11.7)
\]

and

\[
Z: X \ni x \mapsto (A^{-1}x) \cap (-B^{-\square}x), \quad (11.8)
\]

respectively. Note that the primal solution mapping \(Z\) of \((A, B)\) is the dual solution mapping of \((A, B)^*\) and analogously for \(K\).

The importance of these mappings stems from the following result (see [3] or Proposition 4.14),

\[
\text{dom } K = Z, \text{ ran } K = K, \text{ dom } Z = K, \text{ ran } Z = Z, \text{ and } Z = K^{-1}. \quad (11.9)
\]

which shows that the solutions mappings relate the sets of solutions \(Z\) and \(K\) to each other.

**Definition 11.1 (shift operator and corresponding inner/outer perturbations).** Let \(w \in X\). We define\(^{13}\) the associated shift operator

\[
S_w: X \to X: x \mapsto x - w, \quad (11.10)
\]

and we extend \(S_w\) to deal with subsets of \(X\) by setting \((\forall C \subseteq X) S_w(C) = \bigcup_{c \in C} \{S_w(c)\}\). We define the corresponding inner and outer perturbations of \(A\) by

\[
A_w = A \circ S_w: X \ni x \mapsto A(x - w), \quad (11.11)
\]

and

\[
wA = S_w \circ A: X \ni x \mapsto Ax - w. \quad (11.12)
\]

Observe that if \(w \in X\), then the operators \(A_w\) and \(A_w\) are maximally monotone, with domains \(\text{dom } S_{-w}(\text{dom } A) = w + \text{dom } A\) and \(\text{dom } A\), respectively.

**Lemma 11.2 (perturbation calculus).** Let \(w \in X\). Then the following hold:

(i) \(A_w^{-1} = -w(A^{-1})\).

(ii) \((wA)^{-1} = (A^{-1})_{-w}\).

\(^{13}\)We point out that Definition 11.1 extends our earlier definition of inner and outer shifts of single-valued operators (see Definition 7.1) to possibly set-valued operators.
11.4. Perturbations of the Douglas–Rachford operator

(iii) \((A_w)^\ominus = (A^\ominus)_{-w}\).

(iv) \((wA)^\ominus = -w(A^\ominus)\).

(v) \((A_w)^{-\ominus} = w(A^{-\ominus})\).

(vi) \((wA)^{-\ominus} = (A^{-\ominus})_w\).

Proof. Let \((x, y) \in X^2\).

(i): \(y \in (A_w)^{-1}x \iff x \in A_w y = A(y - w) \iff y - w \in A^{-1}x \iff y \in A^{-1}x + w = -w(A^{-1})x\).

(ii): \(y \in (wA)^{-1}x \iff x \in wAy \iff x + w \in Ay \iff y \in A^{-1}(x + w) = (A^{-1})_{-w}x\).

(iii): \((A_w)^\ominus = -A_w(-x) = -A(-x - w) = A^\ominus(x + w) = (A^\ominus)_{-w}\).

(iv): \((A_w)^\ominus x = -A_w(-x) = -(A(-x - w)) = A^\ominus x - (-w) = -w(A^\ominus)x\).

(v): Using (i) and (iv), we see that \((A_w)^{-\ominus} = ((A_w)^{-1})^\ominus = (-w(A^{-1}))^\ominus = w(A^{-\ominus})\).

(vi): Using (ii) and (iii), we see that \((wA)^{-\ominus} = ((wA)^{-1})^\ominus = ((A^{-1})_{-w})^\ominus = (A^{-\ominus})_w\).

As an application, we record the following result which will be useful later.

Corollary 11.3 (dual of inner-outer perturbation). Let \(w \in X\). Then

\[ (wA, B_w)^\star = ((wA)^{-1}, (B_w)^{-\ominus}) = ((A^{-1})_{-w}, w(B^{-\ominus})) \]  \hspace{1cm} (11.13)

Proof. Combine (11.4) with Lemma 11.2(ii)&(v).

11.4 Perturbations of the Douglas–Rachford operator

We now turn to the Douglas–Rachford operator (see Definition 5.1) defined by

\[ T = T_{(A,B)} = J_B R_A + \text{Id} - J_A. \]  \hspace{1cm} (11.14)

Proposition 11.4. Let \(w \in X\). Then the following hold:

(i) If \(x \in \text{Fix}(-wT)\), then \(x - w - J_A x \in wA J_A x \cap (-B_w J_A x)\).

(ii) If \(y \in wAz \cap (-B_w z)\), then \(x = w + y + z \in \text{Fix}(-wT)\) and \(z = J_A x\).

Proof. If \(x \in X\), then \(x - w - J_A x \in wA J_A x\).

(i): Since \(x \in \text{Fix}(-wT)\), we have \(x - Tx = w\); equivalently, \(J_A x - w = J_B R_A x\). Hence

\[ 2J_A x - x = R_A x \in (B + \text{Id})(J_A x - w) = B_w J_A x + J_A x - w \]  and thus

\[ -(x - w - J_A x) \in B_w J_A x \].

129
11.4. Perturbations of the Douglas–Rachford operator

(ii): Since \( y \in wAz \cap (-B_wz) = (Az - w) \cap (-B(z - w)) \), we have \( z = J_Ax \) and \( z - w = J_B(-y + z - w) \). Hence \( R_Ax = 2J_Ax - x = 2z - (w + y + z) = z - w - y \) and so \( J_BR_Ax = J_B(z - w - y) = z - w \). Thus, \( x - Tx = J_Ax - J_BR_Ax = z - (z - w) = w \).

\[ \square \]

Corollary 11.5. Let \( w \in X \). Then

\[ \text{Fix}(-wT) = w + \bigcup_{z \in X} (z + wAz \cap (-B_wz)). \]  

(11.15)

Proposition 11.6. Let \( w \in X \). Then

\[ T_{(wA, B_w)} = T_{-w} \]  

(11.16)

and

\[ \text{Fix}(T_{-w}) = -w + \text{Fix}(-wT) = \bigcup_{z \in X} \left( z + ((Az - w) \cap (-B(z - w))) \right). \]  

(11.17)

Proof. Let \( x \in X \). Using Fact 3.20, we obtain \( J_{B_w}x = J_B(x - w) + w \) and \( J_{wA}x = J_A(x + w) \). Consequently, \( R_{wA}x = 2J_A(x + w) - x \). It thus follows with (11.14) that

\[ T_{(wA, B_w)}x = x - J_{wA}x + J_{B_w}R_{wA}x \]  

(11.18a)

\[ = x - J_A(x + w) + J_B(2J_A(x + w) - x - w) + w \]  

(11.18b)

\[ = (x + w) - J_A(x + w) + J_B(R_A(x + w)) \]  

(11.18c)

\[ = T(x + w) = (T_{-w})x, \]  

(11.18d)

and so (11.16) holds. Next, \( x \in \text{Fix}(T_{-w}) \Leftrightarrow x = T(x + w) \Leftrightarrow x + w = w + T(x + w) \Leftrightarrow x + w \in \text{Fix}(-wT) \), and have thus verified the left identity in (11.17). Finally to see the right identity in (11.17), use Corollary 11.5. \[ \square \]

We now obtain a generalization of Theorem 5.14, which corresponds to the case when \( w = 0 \).

Proposition 11.7. Let \( w \in X \) and define

\[ K_w : \text{gr} \rightarrow (Ax - w) \cap (-B(x - w)). \]  

(11.19)

Then

\[ \Psi_w : \text{gr}K_w \rightarrow \text{Fix}(-wT) : (z, k) \mapsto z + k + w \]  

(11.20)

is a well-defined bijection that is continuous in both directions, with \( \Psi_w^{-1} : x \mapsto (J_Ax, x - J_Ax - w) \).
11.5. The $w$-perturbed problem

**Proof.** For the pair $(wA, B_w)$, the dual solution mapping is $K_w$ and by (11.16) the Douglas–Rachford operator is $T_{-w}$. Applying Theorem 5.14 in this context, we obtain

$$
\Phi: \text{gr } K_w \rightarrow \text{Fix } T_{-w}: (z, k) \mapsto z + k
$$

(11.21)

is continuous in both directions with $\Phi^{-1}: x \mapsto (J_{wA}x, x - J_{wA}x) = (J_A(x + w), x - J_A(x + w))$. Furthermore, $S_{-w}$ is a bijection from $\text{Fix}(T_{-w})$ to $\text{Fix}(T_w)$ by (11.17). This shows that $\Psi_w = S_{-w} \circ \Phi$ as claimed. ■

11.5 The $w$-perturbed problem

**Definition 11.8 ($w$-perturbed problem).** Let $w \in X$. The $w$-perturbation of $(A, B)$ is $(wA, B_w)$. The $w$-perturbed problem associated with the pair $(A, B)$ is to determine the set of zeros

$$
Z_w = (wA + B_w)^{-1}(0) = \{ x \in X \mid w \in Ax + B(x - w) \}.
$$

(11.22)

Note that the $w$-perturbed problem of $(A, B)$ is precisely the primal problem of $(wA, B_w)$, i.e., of the $w$-perturbation of $(A, B)$.

**Proposition 11.9 (Douglas–Rachford operator of the $w$-perturbation).** Let $w \in X$. Then the Douglas–Rachford operator of the $w$-perturbation $(wA, B_w)$ of $(A, B)$ is

$$
T_{(wA,B_w)} = T_{-w}.
$$

(11.23)

**Proof.** This follows from (11.16) of Proposition 11.6. ■

**Proposition 11.10.** Let $w \in X$. Then

$$
Z_w = J_{wA}(\text{Fix}(T_{-w})) = J_A(w + \{ x \in X \mid x = T(x + w) \}).
$$

(11.24)

Furthermore, the following are equivalent:

(i) $Z_w \neq \varnothing$.

(ii) $\text{Fix}(T_{-w}) \neq \varnothing$.

(iii) $w \in \text{ran}(\text{Id} - T)$.

(iv) $w \in \text{ran}(A + B_w)$.
11.5. The \( w \)-perturbed problem

**Proof.** The identity (11.24) follows by combining ?? with Proposition 11.9. This also yields the equivalence of (i) and (ii). Let \( x \in X \). Then \( x \in Z_w \iff 0 \in wA_x + B_wx \iff w \in Ax + B_wx \), and we deduce the equivalence of (i) and (iv). Finally, \( x \in \text{Fix}(T_w) \iff x = T(x + w) \iff w \in (\text{Id} - T)(x + w) \), which yields the equivalence of (ii) and (iii). ■

The equivalence of (i) and (iii) yields the following key result on which \( w \)-perturbations have nonempty solution sets.

**Corollary 11.11.** \( \{ w \in X \mid Z_w \neq \emptyset \} = \text{ran}(\text{Id} - T) \).

**Remark 11.12 (Attouch–Théra dual of the perturbed problem).** Consider the given pair of monotone operators \((A, B)\). We could either first perturb and then take the Attouch–Théra dual or start with the Attouch–Théra dual and then perturb. It turns out that the order of these operations does not matter — up to a horizontal shift of the graphs. Indeed, for every \( x \in X \), we have

\[
(w(A^{-1}) + (B^{-\ominus})_w)x = A^{-1}x - w + B^{-\ominus}(x - w) \quad (11.25a)
\]

\[
= A^{-1}((x - w) + w) + B^{-\ominus}(x - w) - w \quad (11.25b)
\]

\[
= (A^{-1})_w(x - w) + w(B^{-\ominus})(x - w) \quad (11.25c)
\]

\[
= ((A^{-1})_{-w} + w(B^{-\ominus}))(x - w). \quad (11.25d)
\]

Hence \( \text{gr}(w(A^{-1}) + (B^{-\ominus})_w) = (w, 0) + \text{gr}((A^{-1})_w + w(B^{-\ominus})) \), which gives rise to the following diagram:
11.6 The normal problem

We are now in a position to define the normal problem.

**Definition 11.13 (minimal displacement vector and the normal problem).**

The vector

\[ v_{(A,B)} = v = P_{\text{ran}(\text{Id} - T)}^0 \]  \hspace{1cm} (11.26)

is the **minimal displacement vector** of \((A,B)\). The **normal problem** associated with \((A,B)\) is the \(v_{(A,B)}\)-perturbed problem of \((A,B)\). The set of **primal normal solutions** (normal solutions for simplicity) is

\[ Z_v = Z_{(v_{A,B}v)} \]  \hspace{1cm} (11.27)

and the set of **dual normal solutions** is

\[ K_v = K_{(v_{A,B}v)} = Z_{((v_{A})^{-1}(B_{v})^{-\ominus})}. \]  \hspace{1cm} (11.28)

**Lemma 11.14.** The following hold:

(i) \[ Z_v = J_{vA}(\text{Fix}(T_{-v})) = J_{(-v+A)}(\text{Fix}(T_{-v})) = J_{A}(\text{Fix}(T_{-v}) + v) = J_{A}(\text{Fix}(v + T)). \]

(ii) \[ K_v = (\text{Id} - J_{vA})(\text{Fix}(T_{-v})) = (\text{Id} - J_{(-v+A)})(\text{Fix}(T_{-v})). \]
11.6. The normal problem

(iii) \( K_v \neq \emptyset \iff Z_v \neq \emptyset \iff v \in \text{ran}(\text{Id} - T) \).

Proof. (i): Apply (10.3) to the normal pair \((vA, B_v)\) and use (11.23) with \( w \) replaced by \( v \) and (7.1). Now apply Fact 3.20(i). The last equality follows from Lemma 7.2(i). (ii): Apply (10.3) to the normal pair \((vA, B_v)\), then use (11.23) with \( w \) replaced by \( v \) and Fact 3.20(ii). (iii): The first equivalence follows from applying Proposition 4.4(v) to the normal pair \((vA, B_v)\). Now apply Corollary 11.11. 

Remark 11.15 (new notions are well-defined). The notions presented in Definition 11.13 are well-defined: indeed, since \( T \) is firmly nonexpansive (Fact 5.2(i)), it is also nonexpansive and the existence and uniqueness of \( v_{(A,B)} \) follows from Fact 7.4.

Remark 11.16 (new notions extend original notions). Suppose that for the original problem \((A, B)\), we have \( Z = Z_0 = (A + B)^{-1}(0) \neq \emptyset \). By Corollary 11.11, \( 0 \in \text{ran}(\text{Id} - T) \) and so \( v_{(A,B)} = 0 \). Hence the normal problem coincides with the original problem, as do the associated sets of solutions.

Remark 11.17 (normal problem may or may not have solutions). If the set of original solutions \( Z \) is empty, then the set of normal solutions may be either nonempty (see Example 11.24) or empty (see Example 11.25).

The original problem of finding a zero of \( A + B \) is clearly symmetric in \( A \) and \( B \). We now present a statement about the magnitude of the corresponding minimal displacement vectors:

Proposition 11.18. \( \|v_{(A,B)}\| = \|v_{(B,A)}\| \).

Proof. It follows from Fact 5.2(i) that

\[
\text{Id} - T_{(A,B)} = \frac{1}{2} \left( \text{Id} - R_B R_A \right) \quad \text{and} \quad \text{Id} - T_{(B,A)} = \frac{1}{2} \left( \text{Id} - R_A R_B \right).
\]

(11.29)

Thus, using Lemma 7.6, we see that

\[
\|v_{(A,B)}\| = 2 \|P_{\text{ran}(\text{Id} - R_B R_A)}0\| = 2 \|P_{\text{ran}(\text{Id} - R_A R_B)}0\| = \|v_{(B,A)}\|.
\]

In view of Definition 11.13 and Proposition 11.18, the magnitude of the vector \( v_{(A,B)} \) is actually a measure of how far the original problem is from the normal problem. This magnitude is the same for the pairs \((A, B)\) and \((B, A)\).
11.7. Examples

Remark 11.19 \((v_{(A,B)} \neq v_{(B,A)}) may occur\). We will see in the sequel examples where \(v_{(A,B)} \neq 0\) but (i) \(v_{(B,A)} = -v_{(A,B)}\) (see Remark 11.23); or (ii) \(v_{(B,A)} = v_{(A,B)}\) (see Example 11.27); or (iii) the enclosed angle between \(v_{(B,A)}\) and \(v_{(A,B)}\) takes any value (see Example 11.26 and Example 11.28).

Remark 11.20 (self-duality: \(v_{(A,B)} = v_{(A^{-1},B^{-1})}\)). Since, by Corollary 5.12, we have \(T_{(A,B)} = T_{(A^{-1},B^{-1})}\), it is clear that \(v_{(A,B)} = v_{(A^{-1},B^{-1})}\). It follows from Remark 11.12 that the operations of perturbing by \(v_{(A,B)}\) and taking the Attouch–Théra dual commute, up to a shift.

11.7 Examples

Proposition 11.21. \((\text{Id} - J_B)A^{-1}(0) \subseteq \text{ran}(\text{Id} - T) \subseteq \text{dom} A - \text{dom} B.\)

Proof. The right inclusion follows from (12.21). To tackle the left inclusion, suppose that \(z \in A^{-1}(0)\) and set \(w = z - J_Bz\). Then \(w = z - J_Bz \in B(J_Bz) = B(z - w) + 0 \subseteq Az + B_w(z)\). Hence, by Proposition 11.10, \(w \in \text{ran}(\text{Id} - T)\). \(\blacksquare\)

Proposition 11.22 (normal cone operators). Suppose that \(A = N_U\) and \(B = N_V\), where \(U\) and \(V\) are nonempty closed convex subsets of \(X\). Then

\[ v_{(A,B)} = P_{U-V}0 \quad (11.30) \]

and the set of normal solutions is

\[ U \cap (v_{(A,B)} + V) = \text{Fix}(P_UP_V). \quad (11.31) \]

Proof. Since \(A^{-1}(0) = U\) and \(J_B = P_V\), Proposition 11.21 yields \(C = \{ u - P_Vu \mid u \in U \} \subseteq \text{ran}(\text{Id} - T) \subseteq U - V\); hence,

\[ C \subseteq \text{ran}(\text{Id} - T) \subseteq U - V. \quad (11.32) \]

Set \(g = P_{U-V}0\). By [10, Theorem 4.1], there exists a sequence \((u_n)_{n \in \mathbb{N}}\) in \(U\) such that \(u_n - P_Vu_n \to g\). It follows that \((u_n - P_Vu_n)_{n \in \mathbb{N}}\) lies in \(C\) and hence that \(g \in C\). Therefore \(P_C0 = P_{\text{ran}(\text{Id} - T)}0 = P_{U-V}0\) and we obtain (11.30). (For an alternative proof, see [20, Theorem 3.5].)

Let \(x \in X\). Then \(x\) is a normal solution if and only if

\[ g \in N_ux + N_V(x - g). \quad (11.33) \]
11.7. Examples

Assume first that (11.33) holds. Then $x \in U$ and $x - g \in V$. Hence

$x \in U \cap (g + V) = \operatorname{Fix}(P_U P_V)$ by [10, Lemma 2.2]. Conversely, assume $x \in U \cap (g + V) = \operatorname{Fix}(P_U P_V)$. Then $x - g \in V$, $x \in U$, $P_V x = x - g$ and $P_U(x - g) = x$. Hence $N_U(x) \supseteq \mathbb{R}^- g$ and $N_V(x - g) \supseteq \mathbb{R}^+ g$; consequently, $N_U(x) + N_V(x - g) \supseteq \mathbb{R}^- g + \mathbb{R}^+ g = \mathbb{R}g \ni g$ and (11.33) holds.

Remark 11.23. Proposition 11.22 is consistent with the theory dealing with inconsistent feasibility problems (see [10]). Note that it also yields the formula

$$v_{(A,B)} = -v_{(B,A)} \quad (11.34)$$

in this particular context.

Example 11.24 (no original solutions but normal solutions exist). Suppose that $A$ and $B$ are as in Proposition 11.22, that $U \cap V = \emptyset$, and that $V$ is also bounded. Then $U \cap (v_{(A,B)} + V) = \operatorname{Fix}(P_U P_V) \neq \emptyset$ by the Browder–Göhde–Kirk fixed point theorem (see [12, Theorem 4.19]). So, the original problem has no solution but there exist normal solutions.

![Figure 11.2: A GeoGebra [78] snapshot that illustrates Example 11.24.](image)

Example 11.25 (neither original nor normal solutions exist). Suppose that $X = \mathbb{R}^2$, that $A$ and $B$ are as in Proposition 11.22, that

$$U = \{(x, y) \in \mathbb{R}^2 \mid \beta + \exp(x) \leq y\}, \quad (11.35)$$

where $\beta \in \mathbb{R}^+$, and that $V = \mathbb{R} \times \{0\}$. Then $v_{(A,B)} = (\beta, 0)$ yet $\operatorname{Fix}(P_U P_V) = \emptyset$. 

136
Example 11.26. Suppose that $X = \mathbb{R}^2$, let $L : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the rotator by $\pi/2$, let $a^* \in \mathbb{R}^2$ and $b^* \in \mathbb{R}^2$. Suppose that $(\forall x \in \mathbb{R}^2)$ $Ax = Lx + a^*$ and $Bx = -Lx - b^*$. Now let $x \in X$ and let $w \in X$. Then $0 = Ax + B(x - w) - w = Lx + a^* - L(x - w) - b^* - w$ and so $(\text{Id} - L)w = a^* - b^*$, i.e., $w = \frac{1}{2}(\text{Id} + L)(a^* - b^*)$ by Example 3.17. It follows that

$$v_{(A,B)} = \frac{1}{2}(\text{Id} + L)(a^* - b^*). \quad (11.36)$$

An analogous argument yields

$$v_{(B,A)} = \frac{1}{2}(\text{Id} - L)(b^* - a^*). \quad (11.37)$$

Setting $d^* = a^* - b^*$, we have $4 \langle v_{(A,B)}, v_{(B,A)} \rangle = \langle Ld^* + d^*, Ld^* - d^* \rangle = \|Ld^*\|^2 - \|d^*\|^2 = 0$. and $v_{(A,B)} + v_{(B,A)} = Ld^*$. Thus if $d^* \neq 0$, i.e., $a^* \neq b^*$, then

$$v_{(A,B)} \neq 0 \text{ and } v_{(A,B)} \perp v_{(B,A)}. \quad (11.38)$$

Example 11.27. Suppose that there exists $a^*$ and $b^*$ in $X$ such that $\text{gr } A = X \times \{a^*\}$ and $\text{gr } B = X \times \{b^*\}$. By (12.21), $\emptyset \neq \text{ran}(\text{Id} - T) \subseteq \{a^* + b^*\}$. Hence $v_{(A,B)} = a^* + b^*$ and analogously $v_{(B,A)} = a^* + b^*$. Thus, if $a^* + b^* \neq 0$, we have

$$v_{(A,B)} \neq 0 \text{ and } v_{(A,B)} = v_{(B,A)} \quad (11.39)$$

In the previous examples we constructed cases where

$$\frac{\langle v_{(A,B)}, v_{(B,A)} \rangle}{\|v_{(A,B)}\|\|v_{(B,A)}\|} \in \{-1,0,1\}. \quad (11.40)$$

We now show that this quotient can take on any value in $[-1, 1]$.
11.7. Examples

Example 11.28 (angle between $v_{(A,B)}$ and $v_{(B,A)}$). Suppose that $S$ is a linear subspace of $X$ such that \{0\} \subsetneq S \subsetneq X$. Let $\theta \in \mathbb{R}$, let $u \in S$, and let $v \in S^\perp$ such that $\|u\| = \|v\| = 1$. Set $a = \sin(\theta)u$, and set $b = \cos(\theta)u$. Suppose that $A = NS_{+a}$ and that $B = NS + b$. Then $D = \text{dom} A – \text{dom} B = S + a – S = S + a$, and $R = \text{ran} A + \text{ran} B = S^\perp + S^\perp + b = S^\perp + b$. Consequently, $- D = S – a$. Clearly, $(D \cap R) = D \cap R = \{b + a\}$, whereas $(-D) \cap R = (-D) \cap R = \{b – a\}$. Therefore, $v_{(A,B)} = b + a$, and $v_{(B,A)} = b – a$. By Proposition 11.18 $\|v_{(A,B)}\| = \|v_{(B,A)}\| = 1$. Moreover, since $a \perp b$

\[
\langle v_{(A,B)}, v_{(B,A)} \rangle = \langle b + a, b – a \rangle = \|b\|^2 – \|a\|^2 \quad (11.41a)
\]

\[
= \cos^2(\theta) – \sin^2(\theta) = \cos(2\theta). \quad (11.41b)
\]

Proposition 11.29. Suppose that there exists continuous linear monotone operators $L$ and $M$ on $X$, and vectors $a^*$ and $b^*$ in $X$ such that $(\forall x \in X)$ $Ax = Lx + a^*$ and $Bx = Mx + b^*$. Consider the problem

\[
\text{minimize } \|w\|^2 \text{ subject to } \quad (w, x) \in X \times X \text{ and } (\text{Id} + M)w – (L + M)x = a^* + b^*. \quad (11.42)
\]

Let $(w, x) \in X \times X$. Then $(w, x)$ solves (11.42) $\iff w = v_{(A,B)}$ and $x$ is a normal solution $\iff w = P_{J_M(\text{ran}(A + B))}0$ and $x \in (A + B)^{-1}(\text{Id} + M)w$.

Proof. Let $(w, x) \in X \times X$. Then $w = Ax + Bw x \iff (\text{Id} + M)w = (L + M)x = a^* + b^* \iff (\text{Id} + M)w = (L + M)x + a^* + b^* \iff w = J_M((L + M)x + a^* + b^*) = J_M(A + B)x$. The conclusion thus follows from Proposition 11.10.

It is nice to recover a special case of our original motivation given in Section 11.2:

Example 11.30 (classical least squares solutions). Suppose that $X = \mathbb{R}^n$, let $M \in \mathbb{R}^{n \times n}$ be such that $M + M^*$ is positive semidefinite, and let $b \in \mathbb{R}^n$. Suppose that $(\forall x \in \mathbb{R}^n)$ $A = M$ and $Bx = -b$ so that the original problem is to find $x \in \mathbb{R}^n$ such that $Mx = b$. Then $v_{(A,B)} = P_{\text{ran} M}(b) – b$ and the normal solutions are precisely the least squares solutions.

Proof. We will use Proposition 11.29. The constraint in (11.42) turns into $(\text{Id} + 0)w = (0 + M)x = 0 + (-b)$, i.e., $w = Mx – b$ so that the optimization problem in (11.42) is

\[
\text{minimize } \|Mx – b\|^2. \quad (11.43)
\]
11.7. Examples

Hence the normal solutions in our sense are precisely the classical least squares solutions. Furthermore, \( v = P_{\text{ran}(A+B)}0 = P_{-b+b_{\text{ran}}(M)}(0) = P_{\text{ran}M}(b) - b \).
Chapter 12

On the range of the Douglas–Rachford operator: Theory

12.1 Overview

As seen previously, the Douglas–Rachford splitting method is one of the most popular methods of solving the problem of finding a zero of the sum of two maximally monotone operators. Surprisingly, little is known about the range of the Douglas–Rachford operator.

The main goal of this chapter is to analyze the range of \( T_{(A,B)} \) and of the displacement mapping \( \text{Id} - T_{(A,B)} \). Recall that (see Corollary 5.5)

\[
\text{ran}(\text{Id} - T_{(A,B)}) \subseteq (\text{dom } A - \text{dom } B) \cap (\text{ran } A + \text{ran } B). \tag{12.1}
\]

It is natural to inquire whether or not this is a mere inclusion or an inequality. In general, this inclusion is strict — sometimes even extremely so in the sense that \( \text{ran}(\text{Id} - T_{(A,B)}) \) may be a singleton while \( (\text{dom } A - \text{dom } B) \cap (\text{ran } A + \text{ran } B) \) may be the entire space; see Example 12.8. This likely has discouraged efforts to obtain a better description of these ranges. However, and somewhat surprisingly, we are able to obtain — under fairly mild assumptions on \( A \) and \( B \) — simple yet elegant formulae for the ranges of \( \text{Id} - T_{(A,B)} \) and \( T_{(A,B)} \).

Our main findings are summarized as follows:

- We prove that for 3* monotone operators a very pleasing formula can be found that reveals the range to be nearly equal to a simple set involving the domains and ranges of the underlying operators. More precisely we obtain the simple and elegant formulae (see Theorem 12.10 and Corollary 12.11)

\[
\text{ran}(\text{Id} - T_{(A,B)}) \simeq (\text{dom } A - \text{dom } B) \cap (\text{ran } A + \text{ran } B). \tag{12.2}
\]
12.2 Near convexity and near equality

and

\[ \text{ran } T_{(A,B)} \simeq (\text{dom } A - \text{ran } B) \cap (\text{ran } A + \text{dom } B), \] (12.3)

where “\( \simeq \)” denotes to the near equality discussed in Definition 3.32.

- Using the correspondence between maximally monotone operators and firmly nonexpansive mappings (see Fact 3.12), we are able to reformulate our results for firmly nonexpansive mappings (see Corollary 13.7).

- We provide some partial results when \( X \) is possibly infinite-dimensional (see Section 12.5).

Except for facts with explicit references, all the new results in this chapter are published in [38].

12.2 Near convexity and near equality

In this section we review and extend the notions of near convexity (see Definition 3.30) and near equality (see Definition 3.32), and we also present some new results. Unless otherwise stated, throughout this chapter

\[ X \text{ is a finite-dimensional real Hilbert space.} \]

Remark 12.1. In fact, it is the calculus of relative interiors of nearly convex sets along with the finite-dimensional version of the Brezis–Haraux theorem (see [4, Section 6.5]) that prompted us to work in finite-dimensional space. Moreover, we are unaware of any generalization of the relative interior that would be always nonempty for nonempty closed convex subsets of a general infinite-dimensional Hilbert space.

Fact 12.2. Let \( C \) and \( D \) be nearly convex subsets of \( X \). Then

\[ C \simeq D \iff \overline{C} = \overline{D}. \] (12.4)

Proof. See [29, Proposition 2.12(i)&(ii)]. \[ \blacksquare \]

Fact 12.3. Let \( (C_i)_{i \in I} \) be a finite family of nearly convex subsets of \( X \), and let \( (\lambda_i)_{i \in I} \) be a finite family of real numbers. Then \( \sum_{i \in I} \lambda_i C_i \) is nearly convex and

\[ \text{ri}(\sum_{i \in I} \lambda_i C_i) = \sum_{i \in I} \lambda_i \text{ ri } C_i. \]
12.2. Near convexity and near equality

Proof. See [29, Lemma 2.13].

Fact 12.4. Let \((C_i)_{i \in I}\) be a finite family of nearly convex subsets of \(X\), and let \((D_i)_{i \in I}\) be a family of subsets of \(X\) such that \(C_i \simeq D_i\) for every \(i \in I\). Then \(\sum_{i \in I} C_i \simeq \sum_{i \in I} D_i\).

Proof. See [29, Theorem 2.14].

Fact 12.5. Let \((C_i)_{i \in I}\) be a finite family of convex subsets of \(X\). Suppose that \(\bigcap_{i \in I} \text{ri } C_i \neq \emptyset\). Then \(\bigcap_{i \in I} \text{ri } C_i = \bigcap_{i \in I} \text{ri } C_i\) and \(\text{ri } \bigcap_{i \in I} C_i = \bigcap_{i \in I} \text{ri } C_i\).

Proof. See [123, Theorem 6.5].

Most of the following results are known. For different proofs see also [49, Theorem 2.1], [106] and [107].

Lemma 12.6. Let \(C\) and \(D\) be nearly convex subsets of \(X\) such that \(\text{ri } C \cap \text{ri } D \neq \emptyset\). Then the following hold:

(i) \(C \cap D\) is nearly convex.

(ii) \(C \cap D \simeq \text{ri } C \cap \text{ri } D\).

(iii) \(\text{ri } (C \cap D) = \text{ri } C \cap \text{ri } D\).

(iv) \(\overline{C \cap D} = \overline{C} \cap \overline{D}\).

Proof. (i): Since \(C\) and \(D\) are nearly convex, by Fact 3.33, \(\text{ri } C\) and \(\text{ri } D\) are convex. Consequently,

\[
\text{ri } C \cap \text{ri } D \text{ is convex,}
\]

and clearly

\[
\text{ri } C \cap \text{ri } D \subseteq C \cap D.
\]

By Fact 3.33 we have \(\text{ri } C \simeq C\) and \(\text{ri } D \simeq D\). Hence, \(\text{ri } (\text{ri } C) = \text{ri } C\) and \(\text{ri } (\text{ri } D) = \text{ri } D\). Therefore,

\[
\text{ri } (\text{ri } C) \cap \text{ri } (\text{ri } D) = \text{ri } C \cap \text{ri } D \neq \emptyset.
\]

Using (12.7) and Fact 12.5 applied to the convex sets \(\text{ri } C\) and \(\text{ri } D\) yield \(\text{ri } (C \cap D) = \text{ri } C \cap \text{ri } D\); hence

\[
\overline{\text{ri } C \cap \text{ri } D} = \overline{\text{ri } C} \cap \overline{\text{ri } D}.
\]
12.2. Near convexity and near equality

Since \( \text{ri } C \simeq C \) and \( \text{ri } D \simeq D \) by Fact 3.33, we have \( \overline{\text{ri } C} = \overline{C} \) and \( \overline{\text{ri } D} = \overline{D} \). Combining with (12.6) and (12.8), we obtain

\[
\text{ri } C \cap \text{ri } D \subseteq C \cap D \subseteq \overline{\text{ri } C} \cap \overline{\text{ri } D} = \overline{\text{ri } C \cap \text{ri } D},
\]

(12.9)

which in turn yields (i) in view of (12.5). (ii): Use (i) and Fact 3.33 applied to the convex set \( \text{ri } C \cap \text{ri } D \) and the nearly convex set \( C \cap D \). (iii): Using (ii), Fact 12.5 applied to the convex sets \( \text{ri } C \) and \( \text{ri } D \) and (12.7) we have

\[
\text{ri}(C \cap D) = \text{ri}(\text{ri } C \cap \text{ri } D) = \text{ri}(\text{ri } C) \cap \text{ri}(\text{ri } D) = \text{ri } C \cap \text{ri } D,
\]

(12.10)

as required. (iv): Since \( C \) and \( D \) are nearly convex, it follows from Fact 3.33 that \( \overline{\text{ri } C} = \overline{C} \) and \( \overline{\text{ri } D} = \overline{D} \). Combining with (ii) and (12.8) we have

\[
\overline{C \cap D} = \overline{\text{ri } C \cap \text{ri } D} = \overline{\text{ri } C} \cap \overline{\text{ri } D} = \overline{C \cap D},
\]

(12.11)

as claimed. \[\blacksquare\]

**Corollary 12.7.** Let \( C_1 \) and \( C_2 \) be nearly convex subsets of \( X \), and let \( D_1 \) and \( D_2 \) be subsets of \( X \) such that \( C_1 \simeq D_1 \) and \( C_2 \simeq D_2 \). Suppose that \( \text{ri } C_1 \cap \text{ri } C_2 \neq \emptyset \). Then

\[
C_1 \cap C_2 \simeq D_1 \cap D_2.
\]

(12.12)

**Proof.** Let \( i \in \{1, 2\} \). Since \( C_i \simeq D_i \), by Definition 3.32 we have

\[
\text{ri } C_i = \text{ri } D_i \quad \text{and} \quad \overline{C_i} = \overline{D_i}.
\]

(12.13)

Hence,

\[
\text{ri } D_1 \cap \text{ri } D_2 = \text{ri } C_1 \cap \text{ri } C_2 \neq \emptyset.
\]

(12.14)

Moreover, since \( C_i \) is nearly convex, it follows from Fact 3.33 that \( \overline{\text{ri } C_i} = \overline{C_i} \) and \( \overline{\text{ri } C_i} \) is convex. Therefore

\[
\text{ri } C_i = \text{ri } D_i \subseteq D_i \subseteq \overline{C_i} = \overline{\text{ri } C_i}.
\]

(12.15)

Hence \( D_i \) is nearly convex. Applying Lemma 12.6 (iii) to the two sets \( C_1 \) and \( C_2 \) implies that

\[
\text{ri}(C_1 \cap C_2) = \text{ri } C_1 \cap \text{ri } C_2.
\]

(12.16)

Similarly we have

\[
\text{ri}(D_1 \cap D_2) = \text{ri } D_1 \cap \text{ri } D_2.
\]

(12.17)

143
Using (12.14) and Lemma 12.6(i), applied to the sets $C_1$ and $C_2$ we have $C_1 \cap C_2$ is nearly convex. Similarly, $D_1 \cap D_2$ is nearly convex. By Fact 3.33 we have

$$C_1 \cap C_2 \simeq \text{ri}(C_1 \cap C_2) \quad \text{and} \quad D_1 \cap D_2 \simeq \text{ri}(D_1 \cap D_2).$$

(12.18)

Hence $C_1 \cap C_2 = \text{ri}(C_1 \cap C_2)$ and $D_1 \cap D_2 = \text{ri}(D_1 \cap D_2)$. Combining with (12.16), (12.14) and (12.17) yield

$$C_1 \cap C_2 = \text{ri}(C_1 \cap C_2) = \text{ri}C_1 \cap \text{ri}C_2 = \text{ri}D_1 \cap \text{ri}D_2$$

$$= \text{ri}(D_1 \cap D_2) = D_1 \cap D_2.$$  

(12.19)

Now, Fact 12.2 applied to the nearly convex sets $C_1 \cap C_2$ and $D_1 \cap D_2$ implies that $C_1 \cap C_2 \simeq D_1 \cap D_2$.  

\[\square\]

12.3 Attouch–Théra duality and the normal problem

This section provides a review of the Attouch–Théra duality\(^{14}\) [3] and the associated normal problem. From now on, throughout the rest of this chapter, we assume that

$$A : X \rightrightarrows X \quad \text{and} \quad B : X \rightrightarrows X \quad \text{are maximally monotone},$$

and that $T = T_{(A,B)}$ refers to the Douglas–Rachford splitting operator for two operators $A$ and $B$, defined in (5.2).

For convenience of the reader, let us recall that (see Corollary 5.12) $T$ is self-dual, i.e.,

$$T_{A^{-1},B^{-1}} = T_{A,B}.$$  

(12.20)

Also recall (see Corollary 5.5) that

$$\text{ran}(\text{Id} - T_{(A,B)}) = \{a - b \mid (a, a^*) \in \text{gr} A, (b, b^*) \in \text{gr} B, a - b = a^* + b^*\}$$

$$\subseteq (\text{dom} A - \text{dom} B) \cap (\text{ran} A + \text{ran} B).$$  

(12.21)

Let $w \in X$ be fixed. For the operator $A$, the inner and outer shifts associated with $A$ are defined by (see Definition 11.1)

$$A_w : X \rightrightarrows X : x \mapsto A(x - w),$$

(12.23)

\(^{14}\)See the discussion in Chapter 4 as well as [66], [73], [118], and [121].

144
12.4. Main results

\[ \ul{w}A : X \rightharpoonup X : x \mapsto Ax - w. \quad (12.24) \]

Notice that \( \ul{w}A \) and \( wA \) are maximally monotone, with \( \text{dom} \, \ul{w}A = \text{dom} \, A + \ul{w} \) and \( \text{dom} \, wA = \text{dom} \, A \).

Finally, recall that Proposition 11.18 implies

\[ \| v_{(A,B)} \| = \| v_{(B,A)} \|. \quad (12.25) \]

We now explore how \( \text{ran}(\text{Id} - T) \) is related to the set \( (\text{dom} \, A - \text{dom} \, B) \cap (\text{ran} \, A + \text{ran} \, B) \). We will prove that when the operators \( A \) and \( B \) are “sufficiently nice”, we have

\[ \text{ran}(\text{Id} - T_{(A,B)}) \simeq (\text{dom} \, A - \text{dom} \, B) \cap (\text{ran} \, A + \text{ran} \, B). \quad (12.26) \]

In general, (12.26) may fail spectacularly as we will now illustrate.

**Example 12.8.** Suppose that \( X = \mathbb{R}^2 \), and that

\[ A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad B = -A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (12.27) \]

Then

\[ \text{ran}(\text{Id} - T_{(A,B)}) = \{0\} \subsetneq \mathbb{R}^2 \]

\[ = (\text{dom} \, A - \text{dom} \, B) \cap (\text{ran} \, A + \text{ran} \, B). \quad (12.29) \]

**Proof.** Recall that \( \text{dom} \, A = \text{dom} \, B = \text{ran} \, A = \text{ran} \, B = \mathbb{R}^2 \), consequently

\[ (\text{dom} \, A - \text{dom} \, B) \cap (\text{ran} \, A + \text{ran} \, B) = \mathbb{R}^2. \]

On the other hand, one checks that \( R_A : (x,y) \mapsto (y,-x) = B \) and \( R_B : (x,y) \mapsto (-y,x) = A \). Hence \( R_B R_A = \text{Id} \) and therefore

\[ \text{Id} - T_{(A,B)} = \frac{1}{2} (\text{Id} - R_B R_A) \equiv 0. \]

12.4 Main results

In this section we present the main results of this chapter. Keeping the notation of Section 12.3, we set additionally

\[ D = D_{(A,B)} = \text{dom} \, A - \text{dom} \, B \quad \text{and} \quad R = R_{(A,B)} = \text{ran} \, A + \text{ran} \, B. \quad (12.30) \]

We start by proving some auxiliary results.

**Lemma 12.9.** The following hold:
12.4. Main results

(i) The sets $D$ and $R$ are nearly convex.

(ii) $\text{ri } D \cap \text{ri } R \neq \emptyset$.

(iii) $D \cap R$ is nearly convex.

(iv) $\text{ri}(D \cap R) = \text{ri } D \cap \text{ri } R$.

(v) $\overline{D \cap R} = \overline{\text{ri } D \cap \text{ri } R}$.

(vi) $\overline{D \cap R} = \overline{D \cap R}$.

Proof. (i): Combine Fact 3.31 and Fact 12.3. (ii): Since $B$ is maximally monotone, the Minty parametrization (3.20) implies that $X = \text{dom } B + \text{ran } B$. Hence by (i) and Fact 12.3

$$0 \in X = \text{ri } X = \text{ri}(\text{ran } A + \text{ran } B - (\text{dom } A - \text{dom } B)) = \text{ri } R - \text{ri } D.$$ (12.31)

Hence, $\text{ri } D \cap \text{ri } R \neq \emptyset$, as claimed. (Note that we did not use the maximal monotonicity of $A$ in this proof.) (iii): Combine (i), (ii) and Lemma 12.6(i). (iv): Combine (i), (ii) and Lemma 12.6(iii). (v): Combine Lemma 12.6(ii), (i) and (ii). (vi): Combine (i), (ii) and Lemma 12.6(iv).

Theorem 12.10. Suppose that $A$ and $B$ satisfy one of the following:

(i) $(\forall w \in \text{ri } D \cap \text{ri } R) \text{ri}(\text{ran } A + \text{ran } B) \subseteq \text{ri } (A + Bw)$.

(ii) $A$ and $B$ are $3^*$ monotone.

(iii) $\text{dom } B + \text{ri } D \cap \text{ri } R \subseteq \text{dom } A$ and $A$ is $3^*$ monotone.

Then

$$\text{ran}(\text{Id} - T_{(A,B)}) \simeq D \cap R.$$ (12.32)

Furthermore, the following implications hold:

$(\exists C \in \{A,B\}) \text{ dom } C = X$ and $C$ is $3^*$ monotone

$\Rightarrow \text{ran}(\text{Id} - T_{(A,B)}) \simeq R,$ (12.33)

$(\exists C \in \{A,B\}) \text{ ran } C = X$ and $C$ is $3^*$ monotone

$\Rightarrow \text{ran}(\text{Id} - T_{(A,B)}) \simeq D,$ (12.34)

and

$$\text{ri}(D \cap R) = D \cap R \Rightarrow \text{ran}(\text{Id} - T_{(A,B)}) = D \cap R.$$ (12.35)
Proof. First we show that
\[(\forall w \in \text{ri } D) \quad A + B_w \text{ is maximally monotone.} \quad (12.36)\]

Notice that \((\forall w \in X) \text{ dom } B_w = \text{ dom } B + w\). Let \(w \in \text{ri } D = \text{ri}(\text{dom } A - \text{ dom } B)\). Then \(\text{ri dom } A \cap \text{ri dom } B_w \neq \emptyset\). Using Fact 3.34, we conclude that \(A + B_w\) is maximally monotone, which proves (12.36). Now, suppose that (i) holds. Then \((\forall w \in D \cap \text{ri } R) w \in \text{ri ran}(A + B_w) \subseteq \text{ ran}(A + B_w)\).

Combining with Proposition 11.10 we conclude that \((\forall w \in D \cap \text{ri } R) w \in \text{ ran}(\text{Id} - T_{(A,B)})\). Hence \(\text{ri } D \cap \text{ri } R \subseteq \text{ ran}(\text{Id} - T_{(A,B)})\). (12.37)

It follows from Lemma 12.9(v) that \(\text{ri } D \cap \text{ri } R = D \cap R\). Altogether, \( D \cap R \subseteq \text{ ran}(\text{Id} - T_{(A,B)})\). (12.38)

It follows from (12.22) that \(\text{ ran}(\text{Id} - T_{(A,B)}) \subseteq D \cap R\). Therefore, \(D \cap R = \text{ ran}(\text{Id} - T_{(A,B)})\). (12.39)

Since \(T_{(A,B)}\) is firmly nonexpansive, hence nonexpansive, it follows from [12, Example 20.26] that \(\text{Id} - T_{(A,B)}\) is maximally monotone, and therefore \(\text{ran}(\text{Id} - T_{(A,B)})\) is nearly convex by Fact 3.31. Using Lemma 12.9(ii)&(iii), we know that \(\text{ri } D \cap \text{ri } R \neq \emptyset\) and \(D \cap R\) is nearly convex. Therefore, using (12.39) and Fact 12.2 applied to the nearly convex sets \(D \cap R\) and \(\text{ran}(\text{Id} - T_{(A,B)})\), we get \(\text{ran}(\text{Id} - T_{(A,B)}) \simeq D \cap R\). Now we show that each of the conditions (ii) and (iii) imply (i). Let \(w \in \text{ri } D \cap \text{ri } R\), and notice that (ii) implies that \(B_w\) is \(3^*\) monotone, whereas (iii) implies that \(\text{dom } B_w = \text{ dom } B + w \subseteq \text{ dom } A\). Using (12.36) and Fact 3.35 applied to \(A\) and \(B_w\) we have \(\forall w \in \text{ri } D \cap \text{ri } R\)

\[w \in \text{ri } R = \text{ri(ran } A + \text{ran } B)\]
\[= \text{ri(ran } A + \text{ran } B_w) = \text{ri(ran } A + B_w\).

That is, (i) holds, and consequently (12.32) holds.

We now turn to the implication (12.33). Observe first that \(D = X\). If \(A\) is \(3^*\) monotone and \(\text{dom } A = X\), then clearly (iii) holds. Thus, it remains to consider the case when \(B\) is \(3^*\) monotone and \(\text{dom } B = X\). Then \(B_w\) is \(3^*\) monotone and \(\text{dom } A \subseteq X = \text{ dom } B_w\). As before, we obtain \(w \in \text{ri } R = \text{ri(ran } A + \text{ran } B) = \text{ri(ran } A + \text{ran } B_w) = \text{ri(ran } A + B_w\). Hence
(i) holds, which completes the proof of (12.33). To prove the implication (12.34), first notice that \((\exists C \in \{A, B\}) \text{ ran } C = X \text{ and } C \text{ is } 3^* \text{ monotone} \iff (\exists C \in \{A^{-1}, B^{-1}\}) \text{ dom } C = X \text{ and } C \text{ is } 3^* \text{ monotone.}
\) Therefore using Remark 11.20 and (12.33) applied to the operators \(A^{-1}\) and \(B^{-1}\)
\((\exists C \in \{A, B\}) \text{ ran } C = X \text{ and } C \text{ is } 3^* \text{ monotone} \Rightarrow \text{ ran(Id } - T_{(A,B)} = \text{ ran(Id } - T_{(A^{-1},B^{-1})} = R_{(A^{-1},B^{-1})} = \text{ ran } A^{-1} + \text{ ran } B^{-1} = \text{ dom } A - \text{ dom } B = D, \) which proves (12.34).

Now suppose that \(\text{ri}(D \cap R) = D \cap R.\) It follows from (12.32) and (12.22) that
\[ D \cap R = \text{ri}(D \cap R) = \text{ri ran(Id } - T_{(A,B)} \subseteq \text{ ran(Id } - T_{(A,B)} \subseteq D \cap R. \] (12.41)
Hence all the inclusions become equalities, which proves (12.35).

**Corollary 12.11 (range of the Douglas–Rachford operator).** Suppose that \(A\) and \(B\) satisfy one of the following:

(i) \((\forall w \in \text{ri } D_{(A,B^{-1})} \cap \text{ri } R_{(A,B^{-1})}) \text{ ri ran } A + \text{ dom } B \subseteq \text{ ri ran } (A + B^{-1}).\)

(ii) \(A\) and \(B\) are \(3^*\) monotone.

(iii) \(\text{ran } B + \text{ri } D_{(A,B^{-1})} \cap \text{ri } R_{(A,B^{-1})} \subseteq \text{ dom } A \text{ and } A \text{ is } 3^* \text{ monotone.}\)

Then
\[ \text{ran } T_{(A,B)} \simeq (\text{dom } A - \text{ ran } B) \cap (\text{ran } A + \text{ dom } B). \] (12.42)
Furthermore, the following implications hold:

\((\exists C \in \{A, B^{-1}\}) \text{ dom } C = X \text{ and } C \text{ is } 3^* \text{ monotone} \Rightarrow \text{ ran } T_{(A,B)} \simeq \text{ ran } A + \text{ dom } B \) (12.43)

and
\[ \text{ri } (D_{(A,B^{-1})} \cap R_{(A,B^{-1})}) = D_{(A,B^{-1})} \cap R_{(A,B^{-1})} \]
\[ \Rightarrow \text{ ran } T_{(A,B)} = (\text{dom } A - \text{ ran } B) \cap (\text{ran } A + \text{ dom } B). \] (12.44)

**Proof.** Using Proposition 5.3, we know that \(T_{(A,B)} = \text{ Id } - T_{(A,B^{-1})}.\) The result thus follows by applying Theorem 12.10 to \((A,B^{-1}).\)

The assumptions in Theorem 12.10 are critical. Example 12.8 shows that when neither \(A\) nor \(B\) is \(3^*\) monotone, the conclusion of the theorem fails. Now we show that the conclusion of Theorem 12.10 fails even if one of the operators is a subdifferential operator.
12.4. Main results

Example 12.12. Suppose that $X = \mathbb{R}^2$, set $C = \mathbb{R} \times \{0\}$, and suppose that
\[
A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad B = N_C.
\] (12.45)

Then $\text{Id} - T_{(A,B)} = J_A - P_C R_A$. Notice that $P_C : (x, y) \mapsto (x, 0)$, $J_A : (x, y) \mapsto \frac{1}{2}(x + y, -x + y)$ and consequently $R_A : (x, y) \mapsto (y, -x)$. Hence
\[
\text{ran}(\text{Id} - T_{(A,B)}) = \mathbb{R} \cdot (1, -1) \subseteq \mathbb{R}^2 = (\text{dom } A - \text{dom } B) \cap (\text{ran } B + \text{ran } A).
\] (12.46)

Corollary 12.13. Suppose that $A$ and $B$ satisfy one of the following:

(i) $\forall w \in \text{ri } D \cap \text{ri } R$ \quad $\text{ri}(\text{ran } A + \text{ran } B) \subseteq \text{ri ran } (A + B_w)$.

(ii) $A$ and $B$ are $3^*$ monotone.

(iii) $\text{dom } B + \text{ri } D \cap \text{ri } R \subseteq \text{dom } A$ and $A$ is $3^*$ monotone.

(iv) $\exists C \in \{A, B\}$ \quad $\text{dom } C = X$ and $C$ is $3^*$ monotone.

Furthermore, suppose that $D$ and $R$ are affine subspaces. Then $\text{ran}(\text{Id} - T_{(A,B)}) = D \cap R$.

Proof. Since $\text{ri } D = D$ and $\text{ri } R = R$, Lemma 12.9(iv) yields $D \cap R = \text{ri } D \cap \text{ri } R = \text{ri } (D \cap R)$. Now apply (12.35). \ ■

Corollary 12.14. Suppose that $X = \mathbb{R}$. Then $\text{ran}(\text{Id} - T_{(A,B)}) \simeq D \cap R$.

Proof. Indeed, it follows from e.g. [12, Corollary 22.19] and Fact 3.27 that $A$ and $B$ are $3^*$ monotone. Now apply Theorem 12.10(ii). \ ■

We now construct an example where $\text{ran}(\text{Id} - T_{(A,B)})$ properly lies between $\text{ri } (D \cap R)$ and $\overline{D \cap R}$. This illustrate that Theorem 12.10 is optimal in the sense that near equality cannot be replaced by actual equality.

Example 12.15. Suppose that $\text{dim } X \geq 2$, let $u$ and $v$ be in $X$ with $u \neq v$, let $r$ and $s$ be in $[0, +\infty[$, set $U = \text{ball}(u; r)$ and $V = \text{ball}(v; s)$, and suppose that $A = N_U$ and $B = N_V$. Then $D \cap R = \text{ball}(u - v; r + s)$ and
\[
\text{ran}(\text{Id} - T_{(A,B)}) = \text{int } \text{ball}(u - v; r + s)
\]
\[
\cup \left\{ \left(1 - \frac{r+s}{\|u-v\|}\right)(u - v), \left(1 + \frac{r+s}{\|u-v\|}\right)(u - v) \right\} \quad ; \quad (12.47)
\]
consequently,
\[ \text{ri}(D \cap R) \subseteq \text{ran}(\text{Id} - T_{(A,B)}) \subseteq D \cap R. \]  
(12.48)

Moreover,
\[ v_{(A,B)} = \max \left\{ (r + s) - \|v - u\|, 0 \right\} \cdot \frac{v - u}{\|v - u\|}. \]  
(12.49)

**Proof.** It follows from Fact 3.27 that $A$ and $B$ are 3*-monotone. Using Fact 3.8, we have $\text{ran} A = \text{ran} B = X$, hence $R = X$ and $D \cap R = D = U - V$. First notice that
\[ D \cap R = D = U - V = \text{ball}(u - v; r + s). \]  
(12.50)

We claim that
\[ (\forall w \in D \setminus \text{ri} D) \quad U \cap (V + w) \quad \text{is a singleton.} \]  
(12.51)

Since $D = U - V$, we have $(\forall w \in D) \ U \cap (V + w) \neq \emptyset$. Now let $w \in D \setminus \text{ri} D$ and assume to the contrary that $\{y, z\} \subseteq U \cap (V + w)$ with $y \neq z$. Then $\{y - w, z - w\} \subseteq V$, and $(\forall \lambda \in ]0, 1[)$
\[ \lambda y + (1 - \lambda)z \in \text{int} \ U \quad \text{and} \quad \lambda y + (1 - \lambda)z - w \in \text{int} \ V. \]  
(12.52)

It follows from Fact 12.3 and the above inclusions that $w \in \text{int} U - \text{int} V = \text{ri} U - \text{ri} V = \text{ri} D$, which is absurd. Therefore (12.51) holds. Now, let $w \in D \setminus \text{ri} D$ and notice that $V + w = \text{ball}(v + w; r)$. Using (12.51) we have $U \cap (V + w) = \text{dom}(A + B_w)$ is a singleton. Consequently, $\text{ran}(A + B_w)$ is the line passing through the origin parallel to the line passing through $u$ and $v + w$, and by Proposition 11.10, we have $w \in \text{ran}(\text{Id} - T_{(A,B)}) \iff w \in \text{ran}(A + B_w) \iff w = \lambda(u - v - w)$ for some $\lambda \in \mathbb{R} \setminus \{-1\}$ $\iff w = \frac{\lambda}{1 + \lambda}(u - v)$ with $\lambda \in \mathbb{R} \setminus \{-1\}$ (since $u \neq v$), or equivalently,
\[ w = \alpha(u - v), \quad \text{where} \quad \alpha \in \mathbb{R} \setminus \{1\}. \]  
(12.53)

Finally notice that $w$ is on the boundary of $U - V$. Therefore, using (12.50) and (12.53) we must have $\|w - (u - v)\| = r + s \iff |\alpha - 1||u - v|| = r + s \iff \alpha = 1 \pm \frac{r + s}{\|u - v\|}$, which means that only two points on the boundary of $D$ are included in $\text{ran}(\text{Id} - T_{(A,B)})$. Moreover, if $\|u - v\| > r + s$, then $0 \in \text{int} \text{ball}(u - v; s + r)$, hence $v_{(A,B)} = 0$. Else if $\|u - v\| \leq r + s$, using [12, Proposition 28.10] we get $v_{(A,B)} = \left(1 - \frac{r + s}{\|u - v\|}\right)(u - v)$, which completes the proof.

\[\blacksquare\]
12.5 Some infinite-dimensional observations

In this section, we provide some results that remain true in a possibly infinite-dimensional settings. We also provide various examples and counterexamples. We assume henceforth that

\( H \) is a (possibly infinite-dimensional) real Hilbert space. (12.54)

A pleasing identity arises when the we are dealing with normal cone operators of closed subspaces.

**Proposition 12.16.** Let \( U \) and \( V \) be closed linear subspaces of \( H \), and suppose that \( A = N_U \) and \( B = N_V \). Then \( \text{ran}(\text{Id} - T_{(A,B)}) = (U + V) \cap (U^\perp + V^\perp) \).

**Proof.** Since \( \text{gr} N_U = U \times U^\perp \) and \( \text{gr} N_V = V \times V^\perp \), the result follows from (12.21).

\[ \square \]

**Proposition 12.17.** Let \( U \) and \( V \) be closed linear subspaces of \( H \) such that

\[ U^\perp \cap V = \{0\}, \]

and suppose that \( A = N_U \) and \( B = P_V \). Then the following hold:

(i) \( U^\perp \cap P_V^{-1}(P_V(U^\perp) \setminus P_V(U)) \subseteq (D \cap R) \setminus \text{ran}(\text{Id} - T_{(A,B)}) \).

(ii) \( (U^\perp + V) \cap (V + P_V^{-1}(P_V(U^\perp) \setminus P_V(U))) \subseteq (\text{ran} A + \text{ran} B) \setminus \text{ran}(A + B) \).
12.5. Some infinite-dimensional observations

Consequently, if $P_V(U^\perp) \setminus P_V(U) \neq \emptyset$, then

$$\text{ran}(\text{Id} - T_{(A,B)}) \subsetneq D \cap R$$

(12.56)

and

$$\text{ran}(A + B) \subsetneq \text{ran} A + \text{ran} B.$$ 

(12.57)

**Proof.** It is clear that $D \cap R = (U - H) \cap (U^\perp + V) = U^\perp + V$. Notice that (i) and (ii) trivially hold when $P_V(U^\perp) \setminus P_V(U) = \emptyset$. Now suppose that $P_V(U^\perp) \setminus P_V(U) \neq \emptyset$. It follows from (11.22) and (12.55) that

\[
(\forall w \in U^\perp \subseteq U^\perp + V) \quad Z_w \neq \emptyset \iff (\exists u \in U) \quad P_V u - P_V w \in U^\perp - w = U^\perp \\
\iff (\exists u \in U) \quad P_V u - P_V w = 0 \iff P_V w \in P_V(U). 
\]

(12.58)

Now let $w \in U^\perp \subseteq U^\perp + V = D \cap R$ such that $P_V w \notin P_V(U)$. Then (12.58) implies that $Z_w = \emptyset$, hence by Corollary 11.11 $w \notin \text{ran}(\text{Id} - T_{(A,B)})$, which proves (i) and consequently (12.56). To complete the proof we need to show that (ii) holds. Notice that $(\forall u^\perp \in U^\perp) \quad u^\perp + P_V u^\perp \in U^\perp + V = \text{ran} A + \text{ran} B$. It follows from (12.55) that

\[
u^\perp + P_V u^\perp \in \text{ran}(A + B) \iff (\exists u \in U = \text{dom}(A + B)) \quad u^\perp + P_V u^\perp \\
\in U^\perp + P_V u \\
\iff (\exists u \in U) \quad P_V u^\perp - P_V u \in U^\perp - u^\perp = U^\perp \\
\iff (\exists u \in U) \quad P_V u^\perp = P_V u.
\]

(12.59)

Now, let $u^\perp \in U^\perp$ such that $P_V u^\perp \notin P_V(U)$. Then using (12.59) $w = u^\perp + P_V u^\perp \notin \text{ran}(A + B)$. Notice that by construction $w \in U^\perp + V = \text{ran} A + \text{ran} B$. ■

**Remark 12.18.** Notice that in Proposition 12.17 both $A$ and $B$ are linear relations, maximally monotone and $3^\ast$ monotone operators. Consequently, the sets $D$ and $R$ are linear subspaces of $H$. When $H$ is finite-dimensional, Corollary 12.13 and [51, footnote on page 174] imply that $\text{ran}(\text{Id} - T) = D \cap R$ and $\text{ran}(A + B) = \text{ran} A + \text{ran} B$. Thus, if (12.56) or (12.57) holds, then $H$ is necessarily infinite-dimensional.

We now provide a concrete example in $l^2(\mathbb{N})$ where both (12.56) and (12.57) hold. This illustrates again the requirement of the closure in Fact 3.35.
Proposition 12.19. Suppose that $H = \ell^2(\mathbb{N})$, let $p \in \mathbb{R}_{++}$, and let $(\alpha_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that

$$\alpha_n \to 0, \quad \sum_{n=0}^{\infty} \alpha_n^{2p-2} < +\infty \quad \text{and} \quad \sum_{n=0}^{\infty} \alpha_n^{2p-4} = +\infty. \quad (12.60)$$

Set

$$U = \{ x = (x_n)_{n \in \mathbb{N}} \in H \mid x_{2n+1} = -\alpha_n x_{2n} \} \quad (12.61)$$

and

$$V = \{ x = (x_n)_{n \in \mathbb{N}} \in H \mid x_{2n} = 0 \}, \quad (12.62)$$

and suppose that $A = N_U$ and $B = P_V$. Then $P_V(U^\perp) \setminus P_V(U) \neq \emptyset$ and hence

$$\text{ran}(\text{Id} - T_{(A,B)}) \subsetneq D \cap R \quad \text{and} \quad \text{ran}(A + B) \subsetneq \text{ran} A + \text{ran} B. \quad (12.63)$$

Proof. It is easy to check that $U^\perp = \{ x = (x_n) \in H \mid x_{2n+1} = \alpha_n^{-1} x_{2n} \}$. Hence $U^\perp \cap V = \{0\}$. Let $w \in H$ be defined as $(\forall n \in \mathbb{N})$ $w_{2n} = \alpha_n^p$ and $w_{2n+1} = \alpha_n^{p-1}$. Clearly $w \in U^\perp$. We claim that $P_V w \notin P_V U$. Suppose this is not true. Then $(\exists u \in U)$ such that $P_V w = P_V u$. Hence $(\forall n \in \mathbb{N}) u_{2n+1} = (P_V u)_{2n+1} = (P_V w)_{2n+1} = w_{2n+1} = \alpha_n^{p-1}$. Consequently, $(\forall n \in \mathbb{N}) u_{2n} = -\alpha_n^{p-2}$, which is absurd since it implies that $\sum_{n=0}^{\infty} u_{2n}^2 = \sum_{n=0}^{\infty} \alpha_n^{2p-4} = +\infty$, by (12.60). Therefore, $P_V(U^\perp) \setminus P_V(U) \neq \emptyset$. Using Proposition 12.17 we conclude that (12.63) holds.

The next example is a special case of Proposition 12.19.

Example 12.20. Suppose that $H = \ell^2(\mathbb{N})$, let $(\alpha_n)_{n \in \mathbb{N}} = (1/(n+1))_{n \in \mathbb{N}}$, let $p \in [3/2, 5/2]$ and let $U, V, A$ and $B$ be as defined in Proposition 12.19. Since $2p - 2 > 1$ and $2p - 4 \leq 1$, we see that (12.60) holds. From Proposition 12.19 we conclude that $\text{ran}(\text{Id} - T_{(A,B)}) \subsetneq D \cap R$ and $\text{ran}(A + B) \subsetneq \text{ran} A + \text{ran} B$.

When $A$ or $B$ has additional structure, it may be possible to traverse between $\text{ran}(A + B)$ and $\text{ran}(\text{Id} - T_{(A,B)})$ as we illustrate now.

Proposition 12.21. Let $A : H \rightrightarrows H$ and $B : H \rightrightarrows H$ be maximally monotone. Then the following hold:

(i) If $B : H \to H$ is linear, then $\text{ran}(\text{Id} - T_{(A,B)}) = J_B(\text{ran}(A + B))$ and $\text{ran}(A + B) = (\text{Id} + B) \text{ran}(\text{Id} - T_{(A,B)})$. 

153
12.5. Some infinite-dimensional observations

(ii) If \( A : H \to H \) is linear and \( \text{Id} - A \) is invertible, then \( \text{ran}(\text{Id} - T_{(A,B)}) = (\text{Id} - A)^{-1}(\text{ran}(A + B)) \) and \( \text{ran}(A + B) = (\text{Id} - A)\text{ran}(\text{Id} - T_{(A,B)}) \).

(iii) If \( A : H \to H \) and \( B : H \to H \) are linear and \( A^* = -A \), then \((\forall \lambda \in [0,1])\) \( \text{ran}(\text{Id} - T_{(A,B)}) = J_{\lambda A^* + (1 - \lambda)B}(\text{ran}(A + B)) \).

Proof. Let \( w \in X \). It follows from Proposition 11.10 that

\[
w \in \text{ran}(\text{Id} - T_{(A,B)}) \iff (\exists x \in H) \text{ such that } w \in Ax + B(x - w). \tag{12.64}
\]

(i): It follows from (12.64) that \( w \in \text{ran}(\text{Id} - T_{(A,B)}) \iff (\exists x \in H) \text{ such that } w \in Ax + Bx - Bw \iff (\exists x \in H) (\text{Id} + B)w = w + Bw \in (A + B)x \iff (\exists x \in H) w \in J_B((A + B)x) \iff w \in J_B(\text{ran}(A + B)). \) Using [27, Theorem 2.1(ii)&(iv)] we learn that \( J_B \) is a bijection, hence invertible, and \( \text{ran}(A + B) = J_B^{-1}\text{ran}(\text{Id} - T_{(A,B)}) = (\text{Id} + B)\text{ran}(\text{Id} - T_{(A,B)}) \), as claimed.

(ii): It follows from (12.64) that \( w \in \text{ran}(\text{Id} - T_{(A,B)}) \iff (\exists x \in H) \text{ such that } w - Aw \in A(x - w) + B(x - w) = (A + B)(x - w) \iff (\exists x \in H) (\text{Id} - A)w \in (A + B)(x - w) \iff (\exists x \in H) w \in (\text{Id} - A)^{-1}((A + B)(x - w)) \iff w \in (\text{Id} - A)^{-1}(\text{ran}(A + B)). \) Since \( \text{Id} - A \) is invertible, we learn that \( \text{Id} - A \) is a bijection and \( \text{ran}(A + B) = (\text{Id} - A)\text{ran}(\text{Id} - T_{(A,B)}) \), as claimed.

(iii): It follows from (12.64) that

\[
w \in \text{ran}(\text{Id} - T_{(A,B)}) \\
\iff (\exists x \in H) \text{ such that } (\forall \lambda \in [0,1]) w - \lambda Aw + (1 - \lambda)Bw \\
= A(x - \lambda w) + B(x - w + (1 - \lambda)w) \in (A + B)(x - \lambda w) \tag{12.65a} \\
\iff (\text{Id} + \lambda A^* + (1 - \lambda)B)w \in \text{ran}(A + B) \tag{12.65b} \\
\iff w \in J_{\lambda A^* + (1 - \lambda)B}(\text{ran}(A + B)). \tag{12.65c}
\]
Chapter 13

On the range of the Douglas–Rachford operator: Applications

13.1 Overview

In this chapter we present some applications of the results in Chapter 12. As is the case in Chapter 12 we assume that

\[ X \text{ is a finite-dimensional real Hilbert space.} \]

We also follow the same notation as in Chapter 12. Our main results can be summarized as:

- We apply of the results of Chapter 12 to obtain more accurate formulae for the minimal displacement vector (see Proposition 13.1).

- When \( A = \partial f \) and \( B = \partial g \) are subdifferential operators, which is the key setting in convex optimization, formulae (12.2) and (12.3) can be written as (see Corollary 13.4)

\[
\text{ran}(\text{Id} - T_{(\partial f, \partial g)}) \simeq (\text{dom } f - \text{dom } g) \cap (\text{dom } f^* + \text{dom } g^*)
\]

(13.1)

and

\[
\text{ran} T_{(\partial f, \partial g)} \simeq (\text{dom } f - \text{dom } g^*) \cap (\text{dom } f^* + \text{dom } g).
\]

(13.2)

- Using the correspondence between maximally monotone operators and firmly nonexpansive mappings (see Fact 3.12), we are able to reformulate our results for firmly nonexpansive mappings (see Corollary 13.7).

Except for facts with explicit references, all the new results in this chapter are published in [38].
13.2 On the minimal displacement vector $v_{(A,B)}$

In this section, we focus on $v_{(A,B)}$.

**Proposition 13.1.** Suppose that $A$ and $B$ satisfy one of the following:

(i) $(\forall w \in \text{ri } D \cap \text{ri } R) \quad \text{ri}(\text{ran } A + \text{ran } B) \subseteq \text{ri ran } (A + B_w)$.

(ii) $A$ and $B$ are $3^*$ monotone.

(iii) $\text{dom } B + \text{ri } D \cap \text{ri } R \subseteq \text{dom } A$ and $A$ is $3^*$ monotone.

(iv) $(\exists C \in \{A, B\}) \quad \text{dom } C = X$ and $C$ is $3^*$ monotone.

Then $\text{ran}(\text{Id} - T_{(A,B)}) = \overline{D \cap R} = \overline{D \cap R}$ and $v_{(A,B)} = \text{P}_{\overline{D \cap R}}(0)$.

**Proof.** Combine Theorem 12.10, Lemma 12.9(vi), and (11.26).

Using the symmetric hypotheses of Theorem 12.10, we obtain the following result:

**Lemma 13.2.** Suppose that both $A$ and $B$ are $3^*$ monotone, or that $(\exists C \in \{A, B\})$ such that $\text{dom } C = X$ and $C$ is $3^*$ monotone. Then the following hold:

(i) If $D$ is a linear subspace of $X$, then $\text{ran}(\text{Id} - T_{(A,B)}) \simeq \text{ran}(\text{Id} - T_{(B,A)})$ and $v_{(A,B)} = v_{(B,A)}$.

(ii) If $R$ is a linear subspace of $X$, then $\text{ran}(\text{Id} - T_{(A,B)}) \simeq -\text{ran}(\text{Id} - T_{(B,A)})$ and $v_{(A,B)} = -v_{(B,A)}$.

(iii) If $\text{dom } A = X$ or $\text{dom } B = X$, then $\text{ran}(\text{Id} - T_{(A,B)}) \simeq \text{ran}(\text{Id} - T_{(B,A)}) \simeq R$, and $v_{(A,B)} = v_{(B,A)} = \text{P}_{\overline{R}}(0)$.

(iv) If $\text{dom } A$ or $\text{dom } B$ is bounded, then $\text{ran}(\text{Id} - T_{(A,B)}) \simeq -\text{ran}(\text{Id} - T_{(B,A)}) \simeq D$, and $v_{(A,B)} = -v_{(B,A)} = \text{P}_{D}(0)$.

**Proof.** Observe first that

$$\text{ran}(\text{Id} - T_{(B,A)}) \simeq (-D) \cap R$$  \hspace{1cm} (13.3)

by (12.30). (i): Since $D = -D$, Theorem 12.10 and (13.3) yield

$$\text{ran}(\text{Id} - T_{(A,B)}) \simeq \text{ran}(\text{Id} - T_{(B,A)}),$$  \hspace{1cm} (13.4)

and the conclusion follows from (11.26). (ii): Let $u \in X$. Since $R = -R$, we obtain the equivalences $u \in D \cap R \Leftrightarrow -u \in -D$ and $-u \in R$.
Applications to subdifferential operators

\[ \Leftrightarrow -u \in (-D) \cap R \Leftrightarrow u \in -((-D) \cap R). \]  Hence, \( D \cap R = -((-D) \cap R). \) Consequently, \( D \cap R = -((-D) \cap R) \) and \( \text{ri}(D \cap R) = \text{ri}((-D) \cap R). \) Applying Theorem 12.10, in view of (13.3), to the pair \((A, B)\) and the pair \((B, A)\), we conclude \( \text{ran}(\text{Id} - T_{(A,B)}) \simeq -\text{ran}(\text{Id} - T_{(B,A)}). \) Thus \( \text{ran}(\text{Id} - T_{(A,B)}) = -\text{ran}(\text{Id} - T_{(B,A)}), \) and the result follows from (11.26).

(iii): The hypothesis implies \( D = X = \overline{X} = \overline{D}. \) Now combine with (i) and Proposition 13.1(ii). (iv): The assumption that either \( \text{dom} A \) or \( \text{dom} B \) is bounded implies that \( \text{ran} A = X \) (respectively \( \text{ran} B = X \)) (see [12, Corollary 21.21]). Hence \( R = X = \overline{X} = \overline{R}. \) Now combine with (ii) and Proposition 13.1(ii).

Example 13.3. Suppose that \( X = \mathbb{R}. \) It follows from Proposition 11.18 that \( v_{(A,B)} = \pm v_{(B,A)}. \)

13.3 Applications to subdifferential operators

We now turn to subdifferential operators.

Corollary 13.4. Let \( f : X \to [-\infty, +\infty] \) and \( g : X \to [-\infty, +\infty] \) be proper lower semicontinuous convex functions. Then the following hold:

(i) \( \text{ran}(\text{Id} - T_{(\partial f, \partial g)}) \simeq (\text{dom} f - \text{dom} g) \cap (\text{dom} f^* + \text{dom} g^*). \)

(ii) \( \text{ran} T_{(\partial f, \partial g)} \simeq (\text{dom} f - \text{dom} g^*) \cap (\text{dom} f^* + \text{dom} g). \)

Proof. It is well known that (see [12, Corollary 16.29]) \( \overline{\text{dom}} f = \overline{\text{dom}} \partial f. \) Since \( f \) is convex, so is \( \text{dom} f. \) Moreover, by Fact 3.31 \( \text{dom} \partial f \) is nearly convex. Therefore, applying Fact 12.2 to the sets \( \text{dom} f \) and \( \text{dom} \partial f \) we conclude that \( \text{dom} \partial f \simeq \text{dom} f. \) Using [12, Proposition 16.24], and the previous conclusion applied to \( f^*, \) we have \( \text{ran} \partial f = \text{dom}(\partial f)^{-1} = \text{dom} \partial f^* \simeq \text{dom} f^*. \) Altogether,

\[ \text{dom} \partial f \simeq \text{dom} f \quad \text{and} \quad \text{ran} \partial f \simeq \text{dom} f^*. \]  \hspace{1cm} (13.5)

Applying Fact 12.4 with \( C_1 = \text{dom} f, \ C_2 = -\text{dom} g, \ D_1 = \text{dom} \partial f, \ D_2 = -\text{dom} \partial g, \) we conclude that

\[ \text{dom} \partial f - \text{dom} \partial g \simeq \text{dom} f - \text{dom} g. \]  \hspace{1cm} (13.6)

One shows similarly that

\[ \text{ran} \partial f + \text{ran} \partial g \simeq \text{dom} f^* + \text{dom} g^*. \]  \hspace{1cm} (13.7)

157
13.3. Applications to subdifferential operators

It follows from the maximal monotonicity of $\partial f$ and $\partial g$ and Lemma 12.9(ii) that $(\text{dom } \partial f - \text{dom } \partial g) \cap (\text{ran } \partial f + \text{ran } \partial g) \neq \emptyset$. Applying Corollary 12.7 with $C_1 = \text{dom } \partial f - \text{dom } \partial g, C_2 = \text{ran } \partial f + \text{ran } \partial g, D_1 = \text{dom } f - \text{dom } g,$ and $D_2 = \text{dom } f^* + \text{dom } g^*$, we conclude that

$$(\text{dom } \partial f - \text{dom } \partial g) \cap (\text{ran } \partial f + \text{ran } \partial g) \simeq (\text{dom } f - \text{dom } g) \cap (\text{dom } f^* + \text{dom } g^*). \quad (13.8)$$

To complete the proof, notice that by Fact 3.27 $\partial f$ and $\partial g$ are $3^*$ monotone operators. Therefore, by Theorem 12.10, we have

$$\text{ran}(\text{Id} - T_{(\partial f, \partial g)}) \simeq (\text{dom } \partial f - \text{dom } \partial g) \cap (\text{ran } \partial f + \text{ran } \partial g). \quad (13.9)$$

Combining (13.8) and (13.9) we conclude that (i) holds true. To prove (ii), combine Corollary 12.11, (13.5) and Corollary 12.7.

Corollary 13.5. Let $f \in \Gamma(X)$ and suppose that $V$ is a nonempty closed convex subset of $X$. Suppose that $A = \partial f$ and $B = N_V$. Then the following hold:

(i) $T_{(A,B)} = J_{N_V} R_{\partial f} + \text{Id} - J_{\partial f} = P_V(2 \text{Prox}_f - \text{Id}) + \text{Id} - \text{Prox}_f$.

(ii) $\text{ran}(\text{Id} - T_{(A,B)}) \simeq (\text{dom } f - V) \cap (\text{dom } f^* + (\text{rec } V)^\ominus)$.

(iii) $\text{ran } T_{(A,B)} \simeq (\text{dom } f - (\text{rec } V)^\ominus) \cap (\text{dom } f^* + V)$.

Consequently, if $V$ is a linear subspace we may add to this list the following items:

(iv) $\text{ran}(\text{Id} - T_{(A,B)}) \simeq (\text{dom } f + V) \cap (\text{dom } f^* + V^\perp)$.

(v) $\text{ran } T_{(A,B)} \simeq (\text{dom } f + V^\perp) \cap (\text{dom } f^* + V)$.

Proof. Since $\text{ran } N_V$ is nearly convex and $(\text{rec } V)^\ominus$ is convex, it follows from (3.21), Fact 2.9 and Fact 12.2 that

$$\text{ran } N_V \simeq (\text{rec } V)^\ominus. \quad (13.10)$$

(i): This follows from (12.20) and the fact that $J_{N_V} = P_V$ and $J_{\partial f} = \text{Prox}_f$.


(iii): Combine (13.5), (13.10), Corollary 12.11, Fact 12.4 and Corollary 12.7.

(iv)&(v): It follows from [12, Proposition 6.22 and Corollary 6.49] that $\text{rec } V = V$ and $(\text{rec } V)^\ominus = V^\perp$. Combining this with (ii) and (iii), we obtain (iv) and (v), respectively.
Corollary 13.6 (two normal cone operators). Let $U$ and $V$ be two nonempty closed convex subsets of $X$, and suppose that $A = N_U$ and that $B = N_V$. Then the following hold:

(i) $\text{ran}(\text{Id} - T_{(A,B)}) \simeq (U - V) \cap ((\text{rec } U)^\varnothing + (\text{rec } V)^\varnothing)$.

(ii) $\text{ran} T_{(A,B)} \simeq (U - (\text{rec } V)^\varnothing) \cap ((\text{rec } U)^\varnothing + V)$.

(iii) $v_{(A,B)} = P_{\overline{D}}(0)$.

Proof. Clearly $\text{dom } A = U$ and $\text{dom } B = V$. It follows from (13.10) that $\text{ran } N_U \simeq (\text{rec } U)^\varnothing$ and $\text{ran } N_V \simeq (\text{rec } V)^\varnothing$. Therefore, Fact 12.4 implies that $R \simeq (\text{rec } U)^\varnothing + (\text{rec } V)^\varnothing$. (13.11)

Now (i) follows from combining (13.11) and Theorem 12.10, and (ii) follows from combining (13.10) applied to the sets $U$ and $V$, Fact 12.4 and Corollary 12.11. It remains to show (iii) is true. Set $v = P_{\overline{D}}(0) = P_{\overline{D}}(0)$.

On the one hand, by definition of $v$ and Proposition 13.1(ii), we have $v_{(A,B)} \in \overline{D} \cap \overline{R} \subseteq \overline{D}$ and hence

$$\|v\| \leq \|v_{(A,B)}\|.$$ (13.12)

On the other hand, using [20, Corollary 2.7] we have $v \in (P_U - \text{Id})(V) \cap (\text{Id} - P_V)(U) \subseteq (\text{rec } U)^\varnothing \cap (\text{rec } V)^\varnothing$. Therefore, using (13.10) and that $0 \in (\text{rec } U)^\varnothing$ we have $v \in (\text{rec } U)^\varnothing \cap (\text{rec } V)^\varnothing \subseteq (\text{rec } V)^\varnothing \subseteq (\text{rec } U)^\varnothing + (\text{rec } V)^\varnothing \subseteq \overline{R}$. Hence,

$$v \in \overline{D} \cap \overline{R}.$$ (13.13)

Combining (13.12), (13.13) and Proposition 13.1(ii) yields $v = v_{(A,B)}$. ■

13.4 Application to firmly nonexpansive mappings

We now restate the main result from the perspective of fixed point theory.

Corollary 13.7. Let $T_1 : X \to X$ and $T_2 : X \to X$ be firmly nonexpansive such that each $T_i$ satisfies

$$(\forall x \in X)(\forall y \in X) \inf_{z \in X} (T_ix - T_iz, (y - T_iz) - (z - T_iz)) > +\infty, \quad (13.14)$$

and set $T = T_2(2T_1 - \text{Id}) + \text{Id} - T_1$. Then

$$\text{ran } T \simeq (\text{ran } T_1 - \text{ran } (\text{Id} - T_2)) \cap (\text{ran } (\text{Id} - T_1) + \text{ran } T_2), \quad (13.15)$$
13.4. Application to firmly nonexpansive mappings

and

\[ \text{ran}(\text{Id} - T) \simeq (\text{ran} T_1 - \text{ran} T_2) \cap (\text{ran}(\text{Id} - T_1) + \text{ran}(\text{Id} - T_2)). \]  \hspace{1cm} (13.16)

**Proof.** Using Fact 3.12 we conclude that there exist maximally monotone operators \( A : X \rightrightarrows X \) and \( B : X \rightrightarrows X \) such that

\[ T_1 = J_A \quad \text{and} \quad T_2 = J_B. \]  \hspace{1cm} (13.17)

Moreover, it follows from [27, Theorem 2.1(xvii)] and (13.14) that \( A \) and \( B \) are \( 3^* \) monotone. By (5.2), we conclude that \( T = T_{(A,B)} \). Using Corollary 12.11, Lemma 3.19 and (3.21) we have

\[ \text{ran} T \simeq (\text{dom} A - \text{ran} B) \cap (\text{ran} A + \text{dom} B) \]
\[ = (\text{ran}(\text{Id} - T_1) - \text{ran} T_2) \cap (\text{ran} T_1 + \text{ran}(\text{Id} - T_2)). \]

That is, (13.15) holds true. Similarly, one can prove (13.16) by combining Theorem 12.10, Lemma 3.19 and (3.21). \( \blacksquare \)
Chapter 14

The Douglas–Rachford algorithm in the possibly inconsistent case

14.1 Overview

Throughout this chapter, we assume that

\[ A \text{ and } B \text{ are maximally monotone operators on } X. \quad (14.1) \]

Let \( T = T_{(A,B)} \) be the Douglas–Rachford operator associated with \((A,B)\) and let \( x_0 \in X \). Recall that when \( Z = (A + B)^{-1}(0) \neq \emptyset \) the governing sequence \((T^n x_0)_{n \in \mathbb{N}}\) produced by the Douglas–Rachford operator converges weakly to a point in \( \text{Fix } T \) and the shadow sequence \((J_A T^n x_0)_{n \in \mathbb{N}}\) converges weakly to a point in \( Z = (A + B)^{-1}(0) \) (see Fact 9.2). In this chapter we consider the inconsistent case when the set \( Z = (A + B)^{-1}(0) \) is possibly empty. Our main results in this chapter can be summarized as follows:

- We advance the understanding of the inconsistent case significantly by providing a complete proof of the full weak convergence in the convex feasibility setting, when \((A,B) = (N_U, N_V)\), where \( U \) and \( V \) are nonempty closed convex subsets of \( X \) (see Theorem 14.11). This is remarkable because we do not have to have prior knowledge about the gap vector \( v \); the shadow sequence is simply \((J_A T^n x_0)_{n \in \mathbb{N}} = (P_U T^n x_0)_{n \in \mathbb{N}}\).

- In fact, while the general case remains open (especially because of Remark 14.16(ii) and Example 14.17), a more general sufficient condition for weak convergence in the general case is presented (see Theorem 14.10 and Corollary 14.14).

Except for facts with explicit references, all the new results in this chapter appear in [13].
14.2 The convex feasibility setting

We recall that the minimal displacement vector (see Fact 7.4)

\[ v = P_{\text{ran}([\text{Id} - T])0} \]  

is unique and well-defined. Recall also that the normal problem (see Chapter 11) associated with the ordered pair \((A, B)\) is to

find \(x \in X\) such that \(0 \in vAx + Bv x = Ax - v + B(x - v)\). \hspace{1cm} (14.3)

and that

\[ Z_v = Z_{(vA, Bv)} \quad \text{and} \quad K_v = K_{((vA)^{-1}, (Bv)^{-1})}, \]  

(14.4)

denote the primal normal and dual normal solutions, respectively. It follows from Corollary 11.11 that

\[ Z_v \neq \emptyset \Leftrightarrow v \in \text{ran}(\text{Id} - T). \]  

(14.5)

14.2 The convex feasibility setting

We now turn to the convex feasibility problem. Throughout the rest of this section we assume that

\[ U \text{ and } V \text{ are nonempty closed convex subsets of } X \]  

(14.6)

and that

\[ A = N_U, \quad B = N_V \quad \text{and} \quad v = P_{\text{ran}([\text{Id} - T])0}. \]  

(14.7)

Note that, in view of Proposition 11.22, the above formula for \(v\) coincide with (14.2). We shall use \(R_U\) (respectively \(R_V\)) to denote the reflector by the \(U\) (respectively \(V\)) defined by \(R_U = 2P_U - \text{Id}\). One can see that the convex feasibility problem is a special case of the sum problem. Indeed, when \(A = N_U\) and \(B = N_V\), the sum problem is equivalent to the convex feasibility problem: Find \(x \in U \cap V\). In this case, using Example 3.14(ii),

\[ T = T_{(N_U, N_V)} = \text{Id} - P_U + P_V R_U. \]  

(14.8)

**Definition 14.1 (best approximation pair).** Suppose that \(U\) and \(V\) are nonempty closed convex subsets of \(X\) and let \((\bar{u}, \bar{v}) \in U \times V\). Then \((\bar{u}, \bar{v})\) is a best approximation pair relative to \(U\) and \(V\) if

\[ \|\bar{u} - \bar{v}\| = \inf \{\|u - v\| \mid u \in U, v \in V\}. \]  

(14.9)
14.2. The convex feasibility setting

In the following we collect some facts whose detailed proofs appear in [20].

**Fact 14.2.** Suppose that \( e \in U \cap (v + V) \) and that \( y \in N_{U - V} v \) and set \( f = e - v \in (U - v) \cap V \). Then the following hold:

(i) \( N_{U - V} v = (N_U e) \cap (-N_V f) \).

(ii) \( P_U (e + y) = e \).

(iii) \( P_V (f - y) = f \).

*Proof.* See [20, Proposition 2.4]. ■

**Fact 14.3.** The following hold:

(i) \( v \in (P_U - \text{Id})(V) \cap (\text{Id} - P_V)(U) \subseteq (\text{rec } U)^\oplus \cap (\text{rec } V)^\ominus \).

(ii) Suppose that \( S \in \{U, V\} \) is a closed affine subspace. Then

\[
\begin{align*}
v & \in \text{par}(S - S)^\perp.
\end{align*}
\]

(14.10)

*Proof.* See [20, Corollary 2.7 and Remark 2.8(ii)]. ■

**Fact 14.4.** The set \( \text{Fix}(v + T) \) is closed and convex. Moreover

\[
U \cap (v + V) + N_{U - V} v \subseteq \text{Fix}(v + T) \subseteq v + U \cap (v + V) + N_{U - V} v.
\]

(14.11)

*Proof.* See [20, Theorem 3.5]. ■

**Fact 14.5.** Suppose that \( U \cap V \neq \emptyset \). Then \( v = 0, \)

\[
\text{Fix } T = U \cap V + N_{U - V} 0 \quad \text{and} \quad P_U (\text{Fix } T) = U \cap V.
\]

(14.12)

*Proof.* See [20, Corollary 3.9]. ■

**Fact 14.6.** Let \( x \in X \). Then the following hold:

(i) \( T^n x - T^{n+1} x = P_U T^n x - P_V R_U T^n x \rightarrow v \) and \( P_U T^n x - P_V P_U T^n x \rightarrow v \).

(ii) If \( U \cap V \neq \emptyset \) then \( (T^n x)_{n \in \mathbb{N}} \) converges weakly to a point in \( \text{Fix } T = U \cap V + N_{U - V} v \); otherwise \( \|T^n x\| \rightarrow \infty \).
14.2. The convex feasibility setting

(iii) Exactly one of the following alternatives hold:

(a) \( U \cap (v + V) = \emptyset, \|P_U T^n x\| \to \infty \) and \( \|P_V P_U T^n x\| \to \infty \).

(b) \( U \cap (v + V) \neq \emptyset \), the sequences \((P_U T^n x)_{n \in \mathbb{N}}\) and \((P_V P_U T^n x)_{n \in \mathbb{N}}\) are bounded and their weak cluster points belong to \( U \cap (v + V) \) and \((U - v) \cap V\), respectively; in fact, the weak cluster points of the sequences \(((P_V R_U T^n x, P_U T^n x))_{n \in \mathbb{N}}\) and \(((P_V P_U T^n x, P_U T^n x))_{n \in \mathbb{N}}\) are best approximation pairs relative to \((U, V)\).

Proof. See [20, Theorem 13.3].

Figure 14.1: A GeoGebra [78] snapshot that illustrates Fact 14.6(ii). Two intersecting lines in \( \mathbb{R}^3 \), \( U \) the blue line and \( V \) the red line. The first few iterates of the sequences \((T^n x)_{n \in \mathbb{N}}\) (red points) and \((P_U T^n x)_{n \in \mathbb{N}}\) (blue points) are also depicted.

**Fact 14.7.** Suppose that \( V \) is a closed affine subspace. Let \( x \in X \) Then \( P_U T^n x - P_V T^n x \to v \).

Proof. See [20, Theorem 3.17].

In passing we recall that, in the inconsistent convex feasibility setting, when \( U \cap V = \emptyset \) (equivalently \( \text{Fix } T = \emptyset \) by Fact 5.2(iii)), the governing sequence is known to satisfy \( \|T^n x_0\| \to +\infty \) (see Fact 8.1(i)) while the shadow sequence \((P_U T^n x_0)_{n \in \mathbb{N}}\) remains bounded with weak cluster points being the best approximation pairs relative to \( U \) and \( V \) provided they exist.
14.3 Convergence analysis in the inconsistent case

(see Fact 14.6(iii)(b)). Unlike the method of alternating projections, which employs the operator $P_VP_U$, the Douglas–Rachford method is not fully understood in the inconsistent case.

14.3 Convergence analysis in the inconsistent case

Recall the following important fact.

Fact 14.8 (Eckstein and Bertsekas (1992)). Suppose that $\text{zer}(A + B) = \emptyset$ and let $x \in X$. Then $\|T^n x\| \to +\infty$.

Proof. See [72, Corollary 6.2]. Alternatively, note that $\text{zer}(A + B) = \emptyset \iff \text{Fix } T = \emptyset$. Now apply Fact 8.1(i). ■

We start the following useful consequence of Theorem 5.9(iii).

Corollary 14.9. Let $x \in X$ and let $y \in X$. Then the following hold:

\[
\sum_{n=0}^{\infty} \| (\text{Id} - T)T^n x - (\text{Id} - T)T^n y \|^2 < +\infty, \quad (14.13a)
\]
\[
\sum_{n=0}^{\infty} \langle J_A T^n x - J_A T^n y, J_{A^{-1}} T^n x - J_{A^{-1}} T^n y \rangle \geq 0, \quad (14.13b)
\]
\[
\sum_{n=0}^{\infty} \langle J_B T^{n-1} R_A T^n x - J_B T^{n-1} R_A T^n y, J_{B^{-1}} R_A T^n x - J_{B^{-1}} R_A T^n y \rangle \geq 0. \quad (14.13c)
\]

Consequently,

\[
(\text{Id} - T)T^n x - (\text{Id} - T)T^n y \to 0, \quad (14.14a)
\]
\[
\langle J_A T^n x - J_A T^n y, J_{A^{-1}} T^n x - J_{A^{-1}} T^n y \rangle \to 0, \quad (14.14b)
\]
\[
\langle J_B R_A T^n x - J_B R_A T^n y, J_{B^{-1}} R_A T^n x - J_{B^{-1}} R_A T^n y \rangle \to 0. \quad (14.14c)
\]

Proof. Let $n \in \mathbb{N}$. Applying (3.20), to the points $T^n x$ and $T^n y$, we learn that $\{ (J_A T^n x, J_{A^{-1}} T^n x), (J_A T^n y, J_{A^{-1}} T^n y) \} \subseteq \text{gra } A$, hence, by monotonicity of $A$ we have $\langle J_A T^n x - J_A T^n y, J_{A^{-1}} T^n x - J_{A^{-1}} T^n y \rangle \geq 0$. Similarly $\langle J_B R_A T^n x - J_B R_A T^n y, J_{B^{-1}} R_A T^n x - J_{B^{-1}} R_A T^n y \rangle \geq 0$. Now (14.13) and (14.14) follow from Theorem 5.9(iii) by telescoping. ■

We are now ready for our main result.
14.3. Convergence analysis in the inconsistent case

Theorem 14.10 (shadow convergence). Suppose that \( x \in X \), that the sequence \((J_AT^n)x_{n \in \mathbb{N}}\) is bounded and its weak cluster points lie in \( Z_v \), that \( Z_v \subseteq \text{Fix}(v+T) \) and that \((\forall n \in \mathbb{N}) (\forall y \in Z_v) J_AT^ny = y \). Then the shadow sequence \((J_AT^n)x_{n \in \mathbb{N}}\) converges weakly to some point in \( Z_v \).

Proof. Let \( y \in Z_v \subseteq \text{Fix}(v+T) \). Using (14.14b) and Proposition 8.4(i) we have

\[
\langle J_AT^n x - y, T^n x + nv - J_AT^n x \rangle = (J_AT^n x - y, T^n x - J_AT^n x - (y - nv - y)) = (J_AT^n x - J_AT^n y, (\text{Id} - J_A)T^n x - (\text{Id} - J_A)T^n y) \tag{14.15a}
\]

\[
\rightarrow 0. \tag{14.15b}
\]

Note that Proposition 8.4(ii) implies that \((T^n x + nv)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \( \text{Fix}(v+T) \) and consequently with respect to \( Z_v \). Now apply Lemma 6.3 with \( E \) replaced by \( Z_v \), \((u_n)_{n \in \mathbb{N}}\) replaced by \((J_AT^n x)_{n \in \mathbb{N}}\), and \((x_n)_{n \in \mathbb{N}}\) replaced by \((T^n x + nv)_{n \in \mathbb{N}}\). \( \blacksquare \)

As a powerful application of Theorem 14.10, we obtain the following striking result on the convergence of the Douglas–Rachford algorithm for feasibility problems. It illustrates that, even in the inconsistent case, the shadow sequence \((P_UT^n x)_{n \in \mathbb{N}}\) behaves extremely well because it converges to a normal solution without prior knowledge of the minimal displacement vector.

Theorem 14.11 (possibly inconsistent feasibility problem). Suppose that \( U \) and \( V \) are nonempty closed convex subsets of \( X \), that \( A = N_U \), that \( B = N_V \), that \( v = P_{\mathcal{R}_{\text{ran}}(\text{Id}-T)} 0 \) and that \( U \cap (v+V) \neq \emptyset \). Let \( x \in X \). Then \((P_UT^n x)_{n \in \mathbb{N}}\) converges weakly to some point in \( Z_v = U \cap (v+V) \).

Proof. It follows from Fact 14.6(iii)(b) that \((P_UT^n x)_{n \in \mathbb{N}}\) is bounded and its weak cluster points lie in \( U \cap (v+V) \). Moreover Fact 14.4 implies that \( Z_v = U \cap (v+V) \subseteq U \cap (v+V) + N_{U\cap U^\perp T}(v) \subseteq \text{Fix}(v+T) \). Finally, Proposition 8.4(i) and Fact 14.2(ii) imply that \((\forall y \in \text{Fix}(v+T)) (\forall n \in \mathbb{N}) P_UT^ny = P_U(y - nv) = y \), hence all the assumptions of Theorem 14.10 are satisfied and the result follows. \( \blacksquare \)

Remark 14.12. Suppose that \( v \in \text{ran}(\text{Id}-T) \). More than a decade ago, it was shown in [20] that when \( A = N_U \) and \( B = N_V \), where \( U \) and \( V \) are nonempty closed convex subsets of \( X \), that \((P_UT^n x)_{n \in \mathbb{N}}\) is bounded and its weak cluster
points lie in $U \cap (v + V)$. Theorem 14.11 yields the much stronger result that $(P_U T^nx)_{n \in \mathbb{N}}$ actually converges weakly to a point in $U \cap (v + V)$. We point out that the feasibility setting is actually of practical importance and occurs in real-world applications. We also stress the fact that Theorem 14.11 is entirely new even in finite-dimensional spaces.

![Figure 14.2: A GeoGebra [78] snapshot that illustrates Theorem 14.11. Two nonintersecting polyhedral sets in $\mathbb{R}^2$, $U$ and $V$. The first few iterates of the sequences $(T^n x)_{n \in \mathbb{N}}$ (red points) and $(P_U T^n x)_{n \in \mathbb{N}}$ (blue points) are also depicted. Shown is the minimal displacement vector $v$ as well.](image)

In proving Corollary 14.14, which is another variant of Theorem 14.10, we require the following auxiliary result.

**Proposition 14.13.** Suppose that $U$ is a closed affine subspace of $X$, that $A = N_U$, that $B$ is paramonotone and that $v = P_{\text{ran}(\text{Id} - T)} 0 \in \text{ran}(\text{Id} - T)$. Suppose also that $\text{zer}(vA) \cap \text{zer}(Bv) \neq \emptyset$ and let $x \in X$. Then the following hold:

(i) $Z_v = \text{zer}(vA) \cap \text{zer}(Bv) = \text{zer}(-v + N_U) \cap \text{zer}(B(\cdot - v))$.

(ii) $0 \in K_v$. 

167
14.3. Convergence analysis in the inconsistent case

(iii) \( v \in (\text{par } U) \cap \).
(iv) \( Z_\varphi \subseteq U. \)
(v) \( Z_\varphi \subseteq Z_\varphi + K_\varphi = \text{Fix}(T_{-\varphi}) \subseteq \text{Fix}(v + T). \)
(vi) \( (\forall n \in \mathbb{N})(\forall y \in Z_\varphi) J_A T^n y = P_U T^n y = y. \)
(vii) \( P_U T^n x - J_B R_U T^n x = T^n x - T^{n+1} x \to v. \)
(viii) \( (J_A T^n x)_{n \in \mathbb{N}} = (P_U T^n x)_{n \in \mathbb{N}} \) is bounded.
(ix) \( (J_B R_U T^n x)_{n \in \mathbb{N}} \) is bounded.

Proof. (i)&(ii): Apply Theorem 4.32(i) to \((\varphi A, B_\varphi) = (-v + N_U, B(\cdot - v)).\)
(iii): In view of (14.5) we learn that \( Z_\varphi \not= \emptyset. \) Therefore (i) implies that \( \text{zer}(\varphi A) = \text{zer}(-v + N_U) \not= \emptyset. \) Hence \( (\exists x \in X) 0 \in -v + N_U x = -v + (\text{par } U)^+; \) equivalently, \( v \in (\text{par } U)^+. \) (iv): Using (i) we see that \( Z_\varphi \subseteq \text{zer}(-v + N_U). \) We claim that \( \text{zer}(-v + N_U) = U. \) Indeed, let \( y \in X. \) In view of (iii) we have \( 0 \in -v + N_U y \Leftrightarrow [y \in U \text{ and } 0 \in -v + (\text{par } U)^+ = (\text{par } U)^+] \). (v): The left-hand inclusion follows from (ii). Note that, as a subdifferential operator, \( A \) is paramonotone. To prove the identity, apply Corollary 5.16(v) to \((\varphi A, B_\varphi)\) and use Proposition 11.6. The right-hand inclusion follows from Proposition 7.8(iv). (vi): Let \( y \in Z_\varphi \) and note that \( y \in U \cap \text{Fix}(T_{-\varphi}) \) by (iv) and (v). Therefore by Proposition 8.4(i)
\( T^n y = y - n v. \) Now use Lemma 2.8 and (iii). (vii): By definition of \( T \) we have \( P_U T^n x - J_B R_U T^n x = (\text{Id} - T) T^n x, \) which proves the first identity. The limit follows from Fact 8.2. (viii): It follows from Proposition 8.4(ii) that the sequence \( (T^n x + n v)_{n \in \mathbb{N}} \) is Fejér monotone with respect to \( \text{Fix}(T_{-\varphi}), \) hence bounded. In view of Lemma 2.8 and (iii) we learn that \( (\forall n \in \mathbb{N}) J_A T^n x = P_U T^n x = P_U (T^n x + n v) \) and therefore \( (J_A T^n x)_{n \in \mathbb{N}} = (P_U T^n x)_{n \in \mathbb{N}} \) is bounded. (ix): Combine (vii) and (viii). 

Here is another instance of Theorem 14.10.

Corollary 14.14. Let \( x \in X. \) Suppose that \( U \) is a closed affine subspace of \( X, \) that \( A = N_U, \) that \( B \) is paramonotone, that \( v = P_{\text{ran}(\text{Id} - T)} 0 \in \text{ran}(\text{Id} - T), \) that \( \text{zer}(\varphi A) \cap \text{zer}(B_\varphi) \not= \emptyset \) and that all weak cluster points of \( (J_A T^n x)_{n \in \mathbb{N}} = (P_U T^n x)_{n \in \mathbb{N}} \) lie in \( Z_\varphi. \) Then \( (J_A T^n x)_{n \in \mathbb{N}} = (P_U T^n x)_{n \in \mathbb{N}} \) converges weakly to some point in \( Z_\varphi. \)

Proof. This follows from combining Theorem 14.10 and Proposition 14.13(v), (vi), and (viii).
14.3. Convergence analysis in the inconsistent case

Before we proceed, we need to recall that by Theorem 12.10(ii) we have
\[ \text{ran}(\text{Id} - T) = (\text{dom } A - \text{dom } B) \cap (\text{ran } A + \text{ran } B). \] (14.16)

The following provides an example of Corollary 14.14.

**Example 14.15.** Suppose that \( U \) is a closed linear subspace of \( X \), that \( b \in U^\perp \setminus \{0\} \), that \( A = N_U \) and that \( B = \text{Id} + N_{(-b+U)} \). Then \( Z = \emptyset \), \( v = b \in \text{ran}(\text{Id} - T) \), \( Z_v = \{0\} \) and \( K_v = U^\perp \). Moreover, \( (\forall x \in X) \ (\forall n \in \mathbb{N}) \) \( P_UT^n x = \frac{1}{n} P_U x \to 0 \) and \( \|P_UT^n x\| \to +\infty \), and \( 0 \in \text{zer}(vA) \cap \text{zer}(B_v) \).

**Proof.** By the Brezis–Harrau theorem (see Fact 3.35) we have \( X = \text{int } X \subseteq \text{int } \text{ran } B = \text{int } (\text{ran } \text{Id} + \text{ran } N_{(-b+U)}) \subseteq X \), hence \( \text{ran } B = X \). Using (14.16) \( \text{ran}(\text{Id} - T) = (U + b - U) \cap (U^\perp + X) = b + U \). Consequently, using (14.2) and [12, Proposition 3.17] we have
\[ v = P_{\text{ran}(\text{Id} - T)} 0 = P_{b+U} 0 = b + P_U(-b) = b \in U^\perp \setminus \{0\}. \] (14.17)

Note that \( \text{dom } vA = \text{dom } A = U \) and \( \text{dom } B_v = v + \text{dom } B = b - b + U = U \), hence \( \text{dom } (vA + B_v) = U \cap U = U \). Let \( x \in U \). Using (14.17) we have
\[ x \in Z_v \Leftrightarrow 0 \in N_U x - b + x - b + N_{-b+U}(x - b) \quad (14.18a) \]
\[ = N_U x - b + x - b + N_U x \quad (14.18b) \]
\[ \Leftrightarrow 0 \in U^\perp - b + x - b + U^\perp = x \Leftrightarrow x \in U^\perp, \quad (14.18c) \]
hence \( Z_v = \{0\} \), as claimed. As subdifferential operators, both \( A \) and \( B \) are paramonotone, and so are the translated operators \( vA \) and \( B_v \). Since \( Z_v = \{0\} \), in view of Remark 4.29 and (14.17) we learn that
\[ K_v = (N_U 0 - b) \cap (0 - b + N_{-b+U}(0 - b)) \quad (14.19a) \]
\[ = (U^\perp - b) \cap (-b + U^\perp) = U^\perp. \quad (14.19b) \]

Next we claim that
\[ (\forall x \in X) \quad P_UT x = \frac{1}{2} P_U x. \] (14.20)

Indeed, note that \( J_B = (\text{Id} + B)^{-1} = (2 \text{Id} + N_{-b+U})^{-1} = (2 \text{Id} + 2N_{-b+U})^{-1} = (\text{Id} + N_{-b+U})^{-1} \circ (\frac{1}{2} \text{Id}) = P_{-b+U} \circ (\frac{1}{2} \text{Id}) = -b + \frac{1}{2} P_U \), where the last identity follows from [12, Proposition 3.17] and (14.17). Now, using that \( P_U R_U = P_U \) (see Lemma 10.4(ii)) and (14.17) we have \( P_UT = P_U(P_{U^\perp} + J_B R_U) = P_U J_B R_U = P_U (-b + \frac{1}{2} P_U) R_U = P_U (-b + \frac{1}{2} P_U) = \frac{1}{2} P_U \). To show that \( (\forall x \in X) \ (\forall n \in \mathbb{N}) \ P_UT^n x = \frac{1}{n} P_U x \), we use induction. Let \( x \in X \). Clearly, when \( n = 0 \), the base case holds. Now suppose that for some \( n \in \mathbb{N} \)
14.3. Convergence analysis in the inconsistent case

\[ \mathbb{N}, \text{we have, for every } x \in X, P_U T^n x = \frac{1}{n!} P_U x. \] Now applying the inductive hypothesis with \( x \) replaced by \( T x \), and using (14.20), we have \( P_U T^{n+1} x = P_U T^n (T x) = \frac{1}{n!} P_U T x = \frac{1}{n!} \left( \frac{1}{n!} P_U x \right) = \frac{1}{n!} \frac{1}{n!} P_U x \), as claimed. Finally, using (14.17) and [116, Corollary 6(a)] we have \( \| T^n x \| \to +\infty \), hence

\[ \| P_{U^\perp} T^n x \|^2 = \| T^n x \|^2 - \| P_U T^n x \|^2 = \| T^n x \|^2 - \frac{1}{n!} \| P_U x \|^2 \to +\infty. \] (14.21)

\[ \square \]

In fact, as we shall now see, the shadow sequence may be unbounded in the general case, even when one of the operators is a normal cone operator.

**Remark 14.16. (shadows in the presence of normal solutions)**

(i) Example 14.15 illustrates that even when normal solutions exist, the shadows need not converge. Indeed, we have \( K_v = U^\perp \neq \emptyset \) but the dual shadows satisfy \( \| P_{U^\perp} T^n x \| \to +\infty \).

(ii) Suppose that \( A \) and \( B \) are as defined in Example 14.15. Set \( \tilde{A} = A^{-1}, \tilde{B} = B^{-\perp} \) and \( \tilde{Z} = Z_{(\tilde{A}, \tilde{B})} \). By Proposition 4.4(v) \( \tilde{Z} \neq \emptyset \iff K_{(\tilde{A}, \tilde{B})} = Z_{(A, B)} \neq \emptyset \), hence \( \tilde{Z} = \emptyset \). Moreover, Remark 11.20 and Remark 11.12 imply that \( v = b \in \text{ran} (\text{Id} - T) \) and \( \tilde{Z}_v = U^\perp + b = U^\perp \neq \emptyset \). However, in the light of (i) the primal shadows satisfy \( \| J_A T^n x \| = \| J_{A^{-1}} T^n x \| = \| P_{U^\perp} T^n x \| \to +\infty \).

(iii) Concerning Theorem 14.10, it would be interesting to find other conditions sufficient for weak convergence of the shadow sequence or to even characterize this behaviour.

The assumption that \( \text{zer}(\cdot A) \cap \text{zer}(B_v) \neq \emptyset \) is critical in the conclusions of Proposition 14.13 and Corollary 14.14 as we illustrate in Example 14.17. It also provides another example where the shadow sequence diverges to infinity.

**Example 14.17.** Suppose that \( U \) is a closed linear subspace of \( X \), that \( w \in X \setminus U^\perp \), that \( A = N_U \) and that \( \forall x \in X \) \( B x = w \). Let \( x \in X \). Then \( Z = \emptyset \), \( v = P_U w \notin U^\perp \), \( \tilde{Z}_v = U, T = P_U - w \) and \( (\forall n \in \mathbb{N}) P_U T^n x = P_U x - n P_U w \), hence \( \| P_U T^n x \| \to +\infty \). Note that \( \text{zer}(\cdot A) = \text{zer}(B_v) = \text{zer}(\cdot A) \cap \text{zer}(B_v) = \emptyset \).

**Proof.** One can readily verify that \( \forall y \in X \) \( N_U y + B y \subseteq U^\perp + w = U^\perp + P_U w \neq 0 \) and therefore \( Z = \emptyset \). Using (14.16) we have \( \text{ran}(\text{Id} - T) = (U -
14.3. Convergence analysis in the inconsistent case

\[ X \cap (U^\perp + w) = w + U^\perp. \] Therefore using (14.2) and Fact 2.6 we have \( v = P_{w+U}0 = w - P_{U^\perp}w = P_Uw \notin U^\perp. \) Now let \( y \in X. \) Then \( y \in Z_v \iff \) \( y \in U \) and \( 0 \in -v + N_Uy + B(y - v) = U^\perp - P_Uw + w = U^\perp + P_{U^\perp}w = U^\perp). \) Hence \( Z_v = U. \) Clearly \( J_Bx = (\text{Id} + B)^{-1}x = x - w, \) therefore \( T = \text{Id} - P_U + J_B(2P_U - \text{Id}) = \text{Id} - P_U + 2P_U - \text{Id} - w = P_U - w. \) Using simple induction we get \((\forall n \geq 1) T^n x = P_U - (n - 1)P_Uw - w,\) now apply \( P_U \) and notice that \( P_U \) is linear.

We point out that the assumption that \( Z_v \subseteq \text{Fix}(v + T) \) in Theorem 14.10 does not hold true in general, even in the consistent case when \( v = 0 \) as we illustrate in the next two examples.

**Example 14.18** \((v \neq 0 \text{ and } Z_v \not\subseteq \text{Fix}(v + T)).\) Suppose that \( A \) and \( B \) are as defined in Example 14.17 and that \( w \notin U. \) Then

\[ U = Z_v \not\subseteq \text{Fix}(v + T) = -w + U. \] (14.22)

Proof. Let \( x \in X. \) Then \( x \in \text{Fix}(v + T) = \text{Fix}(P_Uw + P_U - w) = \text{Fix}(P_U - P_Uw) \iff x = P_Ux - P_Uw \iff P_{U^\perp}x = P_{U^\perp}(x + w) = 0 \iff x + w \in U \iff x \in -w + U. \) Using Example 14.17 we have \( U = Z_v \subseteq \text{Fix}(v + T) = -w + U \iff w \in U \) which is not true. \( \square \)

**Example 14.19** \((v = 0 \text{ and } Z \not\subseteq \text{Fix } T).\) Suppose that \((a, b) \in X \times (X \setminus \{0\})\) and that \((\forall x \in X) Ax = x - a \text{ and } Bx = x - b.\) Then

\[
\{ \frac{1}{2}(a + b) \} = Z \not\subseteq \text{Fix } T = \{ b \}
\] (14.23)

whenever \( a \neq b. \)

Proof. Clearly \( Z = \frac{1}{2}(a + b). \) One can readily verify that \((\forall x \in x) J_Ax = \frac{1}{2}(x + a), J_Bx = \frac{1}{2}(x + b), \) hence \( R_Ax = a, R_Bx = b \) and therefore \( R_BR_Ax = b. \) Using (5.2) we learn that \( \text{Fix } T = \text{Fix}(R_BR_A) = \{ b \} \) and the conclusion follows. \( \square \)

We conclude this chapter with the following application of Theorem 8.31 to the Douglas–Rachford algorithm when \( A \) and \( B \) are affine relations (possibly not normal cone operators) whose sum does not necessarily have a zero.

**Proposition 14.20.** Suppose that \( A \) and \( B \) are affine such that \( v \in \text{ran}(\text{Id} - T). \) Let \( x \in X \) and set

\[
(\forall n \in \mathbb{N}) \quad x_n = T^n x + n(T^n x - T^{n+1} x).
\] (14.24)
Then the following hold:

(i) \( x_n \to P_{\text{Fix}(v+T)}x = P_{\text{Fix}(T,v)}x = P_{\text{Fix}(T(v,A,B_v))}x. \)

(ii) \( J_Ax_n \to \text{some point in } Z_v. \)

Proof. Note that \( J_A \) and \( J_B \) are affine and so is \( T. \) (i): The limit follows from Theorem 8.31 whereas the identities follow from Theorem 8.8(v) applied with \( n = 1. \) (ii): In view of (i) we learn that \( J_Ax_n \to J_A(P_{\text{Fix}(T(v,A,B_v))}x) \subseteq Z_v, \) where the inclusion follows from applying Lemma 11.14(i). \( \blacksquare \)
Chapter 15

The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces

15.1 Overview

In this chapter, we carefully study the case when $U$ and $V$ are closed affine subspaces that do not necessarily intersect. This case has applications e.g., in image restoration of a spatially limited image from partial knowledge of its Fourier transform (see, e.g., [58]). The problem reduces to a convex feasibility problem for two affine subspaces, and one approach to solve the problem is the Gerchberg–Papoulis algorithm (see [79] and [114]), which reduces to the method of alternating projections in this case, to find a solution. Furthermore, the Douglas–Rachford method for two closed affine subspaces has recently shown to be useful for solving the nonconvex sparse affine feasibility problem (see [87] and [88]) and basis pursuit problem (see [68]). Other possible applications arise via the parallel splitting method (see [12, Proposition 25.7]), which adapts the Douglas–Rachford method to handle a finite sum of monotone operators. When specializing the operators further to normal cone operators of affine subspaces, the problem reduces to a (possibly inconsistent) convex feasibility problem involving an affine subspace and the diagonal subspace in the product space.

The main results of this chapter appear in Theorem 15.4 and are summarized as follows:

- Let $x \in X$. We prove the strong convergence of the shadow sequence $(P_U T^n x)_{n \in \mathbb{N}}$ when $U$ and $V$ are affine subspaces that do not have to intersect (see Theorem 15.4).

- We identify the limit to be the closest best approximation solution to the
15.2. The Douglas–Rachford operator for two affine subspaces

starting point; moreover, the rate of convergence is linear when $U + V$ is closed.

Our proofs critically rely on the new results developed in Theorem 8.8 and Proposition 15.3, the well-developed results in the consistent case in [30] and those of the normal problem studied in Chapter 11.

All the new results in this chapter are published in [15].

15.2 The Douglas–Rachford operator for two affine subspaces

In the following we assume that

\[ v = P_{\text{ran}(\text{Id} - T)} 0 \in \text{ran (Id} - T), \]

that

\[ U \text{ and } V \text{ are nonempty closed convex subsets of } X \]

and that

\[ A = N_U \text{ and } B = N_V. \]

Using Example 3.14(ii), (5.2) becomes

\[ T_{U,V} = T_{(N_U,N_V)} = \text{Id} - P_U + P_V R_U, \]

where $R_U = 2P_U - \text{Id}$. In this case (see Proposition 11.22)

\[ v = P_{U+V} 0, \]

or equivalently

\[ -v \in N_{U+V} v. \]

The normal problem now is to find $x \in X$ such that

\[ 0 \in N_U x - v + N_V (x - v). \]

We start with the following result:

**Proposition 15.1.** Suppose that $U$ is a closed affine subspace of $X$. Then the following hold:
15.2. The Douglas–Rachford operator for two affine subspaces

(i) $K_v = N_{U - v} v.$

(ii) $T_{(-v + U, N_v (-v))} = T_{(N_U, N_v (-v))} = T_{U, v + v}.$

Proof. (i): Let $z \in U \cap (v + V) = Z_v$ and note that, as subdifferential operators, $N_U$ and $N_V$ are paramonotone (see Example 3.24(i)) and so are the translated operators $-v + U$ and $N_v (\cdot - v)$. Therefore, in view of Remark 4.29, Fact 14.3(ii) and Fact 14.2(i) we have

\[
K_v = (-v + N_U z) \cap (-N_V (z - v)) \tag{15.8a}
\]

\[
= (-v + (\text{par } U) \perp) \cap (-N_V (z - v)) \tag{15.8b}
\]

\[
= (\text{par } U) \perp \cap (-N_V (z - v)) \tag{15.8c}
\]

\[
= (N_U z) \cap (-N_V (z - v)) = N_{U - v} v. \tag{15.8d}
\]

(ii): It follows from Lemma 3.21(i), Fact 14.3(ii) and Lemma 2.8 that $J_{-v + U} = P_U (\cdot + v) = P_U$. Moreover, Lemma 3.21(ii) implies that $N_V (\cdot - v) = N_{v + V}$. Altogether, we have

\[
T_{(-v + N_U, N_V (-v))} = T_{(-v + N_U, N_{v + V})} \tag{15.9a}
\]

\[
= \text{Id} - J_{-v + N_U} + J_{N_V (-v)} (2J_{-v + N_U} - \text{Id}) \tag{15.9b}
\]

\[
= \text{Id} - P_U + J_{N_V (-v)} (2P_U - \text{Id}) = T_{(N_U, N_V (-v))} \tag{15.9c}
\]

\[
= T_{(N_U, N_{v + V})} = T_{U, v + v}. \tag{15.9d}
\]

The conclusions of Proposition 15.1 may fail if we drop the assumption that $U$ is an affine subspace as we illustrate in the next example that was motivated by [20, Example 3.7].

Example 15.2. Let $X = \mathbb{R}$, $U = [1, +\infty[$, and $V = \{0\}$. Then the following hold:

(i) $v = 1$.

(ii) $Z_v = \{1\}$.

(iii) $]-\infty, -1] = K_v \neq N_{U - v} = ]-\infty, 0]$.

(iv) $(\forall x \in ]0, +\infty[) 0 = T_{(-v + U, N_V (-v))} x \neq T_{U, v + v} x = \min\{1, x\}.$
15.2. The Douglas–Rachford operator for two affine subspaces

Proof. (i): Clear in view of (15.5). (ii): Combine (i) and Proposition 11.22. (iii): It follows from (15.8a) and (ii) that \( K_v = (-1 + N_U(1)) \cap (N_V(0)) = (-1 + [-\infty, 0]) \cap \mathbb{R} = [-\infty, -1] \). On the other hand \( N_{U^\perp}v = N_U(1) = [-\infty, 0] \). (iv): By Lemma 3.21(ii) we learn that \( N_V(-v) = N_{v^+}. \) Let \( x \in X \) and note that \( v + v = \{1\} \), hence \( \forall x \in X \) \( P_{v^+}x = 1 \). In view of Lemma 2.8 and Lemma 3.21(ii) we have \( T_{(-v+N_U,N_v(-v))}x = T_{(U+v)}x \Leftrightarrow T_{(-v+N_U,N_v(-v))}x = T_{(U,v+v)}x \Leftrightarrow x - P_U(x+v) + P_{v^+}(2P_U(x+v) - x) = x - P_Ux + P_{v^+}(2P_Ux - x) \Leftrightarrow P_U(x+v) = P_Ux \Leftrightarrow x \leq 0. \) Now suppose that \( x > 0 \). Using the same tools as above we obtain \( T_{(-v+N_U,N_v(-v))}x = x - P_U(x+v) + P_{v^+}(2P_U(x+v) - x) = x - P_U(x+1) + 1 = x - (x+1) + 1 = 0. \) On the other hand, \( T_{(U,v+v)}x = x - P_Ux + P_{v^+}(2P_Ux - x) = x - P_Ux + 1 = \min\{1,x\} > 0, \) which completes the proof.

Proposition 15.3. Suppose that \( U \) and \( V \) are closed affine subspaces of \( X \) and that \( T = T_{U,V}. \) Then the following hold:

(i) \( T \) is affine and \( T = \text{Id} - P_U - P_V + 2P_V P_U. \)

(ii) \( v \in (\text{par } U)^\perp \cap (\text{par } V)^\perp. \)

(iii) \( \forall x \in X \) \( \forall \alpha \in \mathbb{R} \) \( P_Ux = P_U(x + \alpha v). \)

(iv) \( \forall x \in X \) \( \forall \alpha \in \mathbb{R} \) \( P_Vx = P_V(x + \alpha v). \)

(v) \( T_{-v} = v + T = T_{N_U,N_v(-v)} = T_{U,v+v}. \)

(vi) \( Z_v = U \cap (v + V). \)

(vii) \( K_v = N_{U^\perp}v = (\text{par } U)^\perp \cap (\text{par } V)^\perp. \)

(viii) \( \text{Fix}(T_{-v}) = \text{Fix}(v + T) = Z_v + K_v = (U \cap (v + V)) + ((\text{par } U)^\perp \cap (\text{par } V)^\perp). \)

Proof. (i): Note that \( P_U \) and \( P_V \) are affine (see Fact 2.7(i)). Using (15.4) we have \( T = \text{Id} - P_U + P_V(2P_U - \text{Id}) = \text{Id} - P_U + 2P_V P_U - P_V. \) Since the class of affine operators is closed under addition, subtraction and composition we deduce that \( T \) is affine.

(ii): This follows from Fact 14.3(ii). (iii) and (iv): Combine (ii) with Lemma 2.8.

(v) Combine Theorem 8.8(v) (applied with \( n = 1 \)) and Proposition 15.1(ii). (vi): See Proposition 11.22.
15.2. The Douglas–Rachford operator for two affine subspaces

(vii): The first identity follows from Proposition 15.1(i). To prove the second identity note that \( N_{U-V} = (\text{par}(U-V))^\perp = (\text{par } U + \text{par } V)^\perp = (\text{par } U)^\perp \cap (\text{par } V)^\perp \).

(viii): Since \(-v + N_U\) and \(N_V(\cdot - v)\) are paramonotone, it follows from (v), (vii) and Corollary 5.16(v) applied to the normal pair \((vA, B_v)\) that \(\text{Fix}(T_{-v}) = \text{Fix}(v + T) = Z_v + K_v\). Now combine with (vi) and (vii) to get the desired result.

We are now ready for our main result. The proof of Theorem 15.4 relies on the work leading up to this point as well as the convergence analysis of the consistent case in [30].

**Theorem 15.4 (Douglas–Rachford algorithm for two affine subspaces).**

Suppose that \(U\) and \(V\) are closed affine subspaces of \(X\) and let \(x \in X\). Then \((\forall n \in \mathbb{N})\) we have

\[
P_U T^n x = P_U((T_{-v})^n x) = P_U T_{U,V + v}^n x = J_{-v + N_U}((T_{-v})^n x),
\]

and

\[
P_U T^n x \rightharpoonup P_{Z_v} x = P_{U \cap (v + V)} x.
\]

Moreover, if \(\text{par } U + \text{par } V\) is closed (as is always the case when \(X\) is finite-dimensional) then the convergence is linear with rate being the cosine of the Friedrichs angle

\[
c_F(\text{par } U, \text{par } V) = \sup_{\substack{u \in \text{par } U \cap W^\perp \cap \text{ball}(0;1) \\ v \in \text{par } V \cap W^\perp \cap \text{ball}(0;1)}} |\langle u, v \rangle| < 1,
\]

where \(W = \text{par } U \cap \text{par } V\) and \(\text{ball}(0;1)\) is the closed unit ball.

**Proof.** Let \(n \in \mathbb{N}\). Using Proposition 15.3(iii) with \((x, \alpha)\) replaced by \((T^n x, n)\) we learn that \(P_U T^n x = P_U(T^n x + nv)\). Now combine with Theorem 8.8(iv) to get the second identity. The third identity follows from applying Proposition 15.3(v). Finally note that using the first identity, Proposition 15.3(iii) with \((x, \alpha)\) replaced by \(((T_{-v})^n x, 1)\) and Lemma 3.21(ii) we learn that \(P_U T^n x = P_U((T_{-v})^n x + v) = J_{-v + N_U}((T_{-v})^n x)\). Now we prove (15.11). It follows from (15.1), Lemma 11.14(iii) and Proposition 15.3(vi) that \(Z_v = U \cap (v + V) \neq \emptyset\). Now apply [30, Corollary 4.5].
15.2. The Douglas–Rachford operator for two affine subspaces

Figure 15.1: Two nonintersecting affine subspaces $U$ (blue line) and $V$ (purple line) in $\mathbb{R}^3$. Shown are also the first few iterates of $(T^n x_0)_{n \in \mathbb{N}}$ (red points) and $(P_U T^n x_0)_{n \in \mathbb{N}}$ (blue points).

Figure 15.1 shows a GeoGebra snapshot [78] of the Douglas–Rachford iterates and its shadows for two nonintersecting nonparallel lines $U$ and $V$ in $\mathbb{R}^3$.

Let us now comment on the comparison of the Douglas–Rachford algorithm to the method of alternating projections.

Remark 15.5. Let $x \in X$ and $n \in \{1, 2, \ldots \}$. A straightforward induction yields

$$ (P_V P_U)^n x = -v + (P_{v+V} P_U)^n x $$

while Theorem 15.4 results in

$$ P_U T^n x = P_U T^n_{U \cup V} x. $$

We conclude that executing the Douglas–Rachford algorithm or the Method of Alternating Projection (MAP), to solve the best approximation problem for $U$ and $V$, is essentially the same (up to a shift with $v$ in the case of MAP) as executing the two algorithms on the (consistent) normal problem. Consequently, we can use the numerical results established in [30, Sections 7 and 8] to illustrate the efficiency of our algorithm. In passing, we recall that the numerical experiments presented in [30, Section 7] show that Douglas–Rachford algorithm reveal a faster rate of convergence than MAP when the angle between the subspaces is small.
15.2. The Douglas–Rachford operator for two affine subspaces

In the light of refined conclusions of Theorem 15.4, under the additional assumptions that the sets are affine subspaces, in comparison to those of Theorem 14.11, it is natural to ask whether or not we can obtain refined/better conclusions in the general inconsistent case when the operators are affine relations.

The following simple example shows that for two affine (but not normal cone) operators, the shadow sequence may fail to converge.

Example 15.6 (both primal and dual shadows are unbounded). Suppose that $X = \mathbb{R}^2$ and let $S : \mathbb{R}^2 \to \mathbb{R}^2 : (x_1, x_2) \mapsto (-x_2, x_1)$, be the counterclockwise rotator by $\pi/2$. Let $b \in \mathbb{R}^2 \setminus \{(0, 0)\}$. Suppose that $A = S$ and set $B = -S + b$. Then $\text{zer } A \neq \emptyset$, $\text{zer } B \neq \emptyset$ yet $\text{zer}(A + B) = \emptyset$. Moreover, $v = (\text{Id} + S)(b)$, the set of normal solutions $Z_v = \mathbb{R}^2$ and for every $x \in \mathbb{R}^2$ we have $\|J_A T^u x\| \to +\infty$ and $\|J_{A^{-1}} T^u x\| \to +\infty$.

**Proof.** Let $x \in \mathbb{R}^2$ and note that $S$ and $-S = S^{-1}$ are both linear, continuous, single-valued, monotone and $S^2 = (-S)^2 = \text{Id}$. It follows from Proposition 3.16 that $J_A x = J_S x = \frac{1}{2}(\text{Id} - S)x = \frac{1}{2}(x - Sx)$ and $J_{A^{-1}} x = J_{-S} x = \frac{1}{2}(\text{Id} + S)x = \frac{1}{2}(x + Sx)$. Similarly using Fact 3.20(i) we can see that $J_B x = \frac{1}{2}(x - b + Sx - Sb)$. Therefore we have $R_A x = -Sx$ and $R_B x = -b + Sx - Sb$. Hence $R_B R_A x = S(-Sx) - Sb - b = -S^2 x - Sb - b = x - Sb - b = x - (\text{Id} + S)b$. Consequently we have

\[
(\forall x \in \mathbb{R}^2) \quad T x = \frac{1}{2} (\text{Id} + R_B R_A) x = x - \frac{1}{2}(\text{Id} + S)b. \tag{15.15}
\]

It follows from (15.15) that $\text{ran}(\text{Id} - T) = \{\frac{1}{2}(\text{Id} + S)b\}$, hence $v = \frac{1}{2}(\text{Id} + S)b$ and $Tx = x - v$. Therefore, using Theorem 8.8(v) $\text{Fix}(v + T) = \text{Fix}(T_{-v}) = \mathbb{R}^2$. Moreover, using Lemma 11.14(i) we learn that $Z_v = J_{A^{-1}}(\text{Fix}(T_{-v})) = \mathbb{R}^2$. In view of Proposition 7.8(i) we have $T^u x = x - nv$. Hence, using that $J_A$ is linear, we get $J_A T^u x = J_A (x - nv) = J_A x - nJ_A v$. Now $J_A v = \frac{1}{2}(\text{Id} - S)(\frac{1}{2}(\text{Id} + S)b) = \frac{1}{2}b \neq (0, 0)$. Similarly $J_{A^{-1}} T^u x = J_{A^{-1}}(x - nv) = J_{A^{-1}} x - nJ_{A^{-1}} v$ and $J_{A^{-1}} v = \frac{1}{2}(\text{Id} + S)(\frac{1}{2}(\text{Id} + S)b) = \frac{1}{2}Sb \neq (0, 0)$.

With some more work, we can construct an example of the same type where the involved operators are even subdifferential operators:

Example 15.7 (The shadows may fail to converge to a normal solution). Suppose that $U$ and $V$ are affine subspaces of $X$ such that $U \cap V = \emptyset$, and set $A = N_U^{-1}$ and $B = N_V^{-\ominus}$. Then $Z = K = \emptyset$, $Z_v = v + \left( (\text{par } U)^\perp \cap (\text{par } V)^\perp \right)$, $K_v = -v + (U \cap (v + V))$, \tag{15.16}
15.2. The Douglas–Rachford operator for two affine subspaces

\[ \| J_{\tilde{A}^{-1}} T^n x \| \to +\infty, \] and \( J_{\tilde{A}^{-1}} T^n x = P_U T^n x \to P_{U \cap (v + V)} x. \) Consequently, even though \( Z_v \neq \emptyset, \) the sequence of primal shadows has no convergent subsequence.

**Proof.** Note that \( \tilde{A}^{-1} = N_U, \tilde{B}^{-\emptyset} = N_V, \) therefore using Fact 5.2(ii) we have \( T_{(\tilde{A}, \tilde{B})} = T_{(\tilde{A}^{-1}, \tilde{B}^{-\emptyset})} = T_{U, V}. \) Clearly, \( K = (N_U + N_V)^{-1}(0) = U \cap V = \emptyset, \) hence by (10.3) and Fact 5.2(ii), \( Z = \emptyset \) and \( \text{Fix} T_{(\tilde{A}, \tilde{B})} = \emptyset. \) Using Remark 11.12 and Proposition 15.3(vii)&(vi), we conclude that \( Z_v = v + ((\text{par } U)^\perp \cap (\text{par } V)^\perp) \) and \( K_v = -v + (U \cap (v + V)). \) Let \( x \in X. \) It follows from Theorem 15.4 with \( A \) and \( B \) replaced with \( \tilde{A}^{-1} \) and \( \tilde{B}^{-\emptyset} \) that \( J_{\tilde{A}^{-1}} T^n x = J_{N_U} T^n x = P_U T^n x \to P_{U \cap (v + V)} x. \) By Fact 8.1(i) \( \| T^n x \| \to +\infty. \) Moreover the inverse resolvent identity (see (3.15)) implies \( J_{\tilde{A}^{-1}} T^n x = (\text{Id} - P_U) T^n x = T^n x - P_U T^n x. \) Finally, \( \| J_{\tilde{A}^{-1}} T^n x \| \geq \| T^n x \| - \| P_U T^n x \| \to +\infty. \)

Example 14.15 provides another instance where the shadow sequence is unbounded with both operators being affine subdifferential operators.
Chapter 16

The Douglas–Rachford algorithm in the affine-convex case

16.1 Overview

In this chapter we provide convergence results when one constraint set is an affine subspace. As a consequence, we extend a result by Spingarn from halfspaces to general closed convex sets admitting least-squares solutions. Let $x \in X$ and suppose that $(A, B) = (N_U, N_V)$ where $U$ and $V$ are nonempty closed convex subsets of $X$. Our main results in this chapter can be summarized as follows:

- When $U$ is a closed affine subspace we provide a simpler proof that the shadow sequence $(P_U T^n x)_{n \in \mathbb{N}}$ will always converge to a normal solution (see Theorem 16.1). We tighten our conclusion by showing that the other shadow sequence $(P_V T^n x)_{n \in \mathbb{N}}$ may or may not be bounded (see Example 16.2).

- In contrast, when $V$ is a closed affine subspace, we obtain the convergence of both shadow sequences $(P_U T^n x)_{n \in \mathbb{N}}$ and $(P_V T^n x)_{n \in \mathbb{N}}$ (see Theorem 16.5).

- As a consequence we obtain a far-reaching refinement of Spingarn’s splitting method introduced in [131] (see Corollary 16.6).

All the new results in this chapter are published in [36].

16.2 Convergence results

In this chapter we work under the assumptions that

\[ U \text{ and } V \text{ are nonempty closed convex subsets of } X, \quad (16.1) \]
16.2. Convergence results

that

\[ A = N_U, \quad B = N_V \quad \text{and} \quad v = P_{\overline{U-V}}0, \quad (16.2) \]

and that

\[ v = v(U,V) = P_{\overline{U-V}}0 \in U - V. \quad (16.3) \]

In view of (16.3) we have

\[ U \cap (v + V) \neq \emptyset \quad \text{and} \quad (-v + U) \cap V \neq \emptyset. \quad (16.4) \]

For sufficient conditions on when \( v \in U - V \) (or equivalently the sets \( U \cap (v + V) \) and \( (-v + U) \cap V \) are nonempty) we refer the reader to [10, Facts 5.1].

Here is our first main result in this chapter.

**Theorem 16.1 (convergence of Douglas–Rachford algorithm when \( U \) is a closed affine subspace).** Suppose that \( U \) is a closed affine subspace of \( X \), and let \( x \in X \). Then

\[ (\forall n \in \mathbb{N}) \quad P_U T^n x = P_U (T^n x + n v). \quad (16.5) \]

Moreover the following hold:

(i) The shadow sequence \( (P_U T^n x)_{n \in \mathbb{N}} \) converges weakly to some point in \( U \cap (v + V) \).

(ii) No general conclusion can be drawn about the boundedness of the sequence \( (P_V T^n x)_{n \in \mathbb{N}} \).

**Proof.** After translating the sets \( U \) and \( V \) by a vector if necessary, we can and do assume that \( U \) is a closed linear subspace of \( X \). Using Fact 14.3(ii) and Lemma 2.8 we learn that \( (\forall n \in \mathbb{N}) \quad P_U T^n x = P_U (T^n x + n v) \). (i): Note that \( U \cap (v + V) \subseteq U \). Now combine Proposition 8.4(ii), Fact 14.6(iii)(b) and Lemma 6.4 with \( C = U \cap (v + V) \), and \( (x_n)_{n \in \mathbb{N}} \) replaced by \( (T^n x + n v)_{n \in \mathbb{N}} \). (ii): In fact, \( (P_V T^n x)_{n \in \mathbb{N}} \) can be unbounded (see Example 16.2 below) or bounded (e.g., when \( U = V = X \)).

**Example 16.2.** Suppose that \( X = \mathbb{R}^2 \), that \( U = \mathbb{R} \times \{0\} \) and that \( V = \text{epi}(\cdot | + 1) \). Then \( U \cap V = \emptyset \) and for the starting point \( x \in [-1, 1] \times \{0\} \) we have \( (\forall n \in \{1, 2, \ldots\}) \quad T^n x = (0, n) \in V \) and therefore \( \|P_V T^n x\| = \|T^n x\| = n \to +\infty \).
Proof. Let \( x = (\alpha, 0) \) with \( \alpha \in [-1, 1] \). We proceed by induction. When \( n = 1 \) we have \( T(x) = P_{U^\perp}(\alpha, 0) + P_V R_U(\alpha, 0) = P_V(\alpha, 0) = (0, 1) \). Now suppose that for some \( n \in \{1, 2, \ldots\} \) \( T^n x = (0, n) \). Then \( T^{n+1} x = T(0, n) = P_{U^\perp}(0, n) + P_V R_U(0, n) = (0, n) + P_V(0, -n) = (0, n + 1) \in V \) and the conclusion follows.

\[
\begin{align*}
&\text{Figure 16.1: A GeoGebra [78] snapshot that illustrates Example 16.2.}
\end{align*}
\]

Remark 16.3. We point out that even though the conclusion of Theorem 16.1(i) follows from Theorem 14.11, in Theorem 16.1(i) we provide an alternative simpler proof that requires less tools than what we need to prove the general case presented in Theorem 14.11. The simpler proof as well as (16.5) are both products of the additional assumption that one set is an affine subspace.

It is tempting to conjecture that (16.5) holds true when \( U \) is just convex and not necessarily a subspace. We now show that, even though we still get the convergence of the shadow sequence \( (P_U T^n x)_{n \in \mathbb{N}} \) in the light of Theorem 14.11, this conjecture fails, in general.

Example 16.4. Suppose that \( X = \mathbb{R} \), that \( U = [1, 2] \) and that \( V = \{0\} \). Then \( v = 1 \) and \( U \cap (v + V) = \{1\} \). Let \( x = 4 \). We have \( (T^n x)_{n \in \mathbb{N}} = (4, 2, 0, -1, -2, -3, \ldots) \), \( P_U T^n x \to 1 \in U \cap (v + V) \) and \( (\forall n \in \{2, 3, 4, \ldots\}) \) \( T^n x + n v = -(n - 2) + n(1) = 2 \in U \) and \( P_U (T^n x + n v) = 2 \in U \setminus (U \cap (v + V)) \). In the proof of Theorem 16.1(i), we had \( (P_U T^n x)_{n \in \mathbb{N}} = (P_U (T^n x + n v))_{n \in \mathbb{N}} \) which is strikingly false here.

183
16.3. Spingarn’s method

When \( V \) is an affine subspace, the convergence theory is even more satisfying:

**Theorem 16.5 (convergence of Douglas–Rachford algorithm when \( V \) is a closed affine subspace).** Suppose that \( V \) is a closed affine subspace of \( X \), and let \( x \in X \). Then the following hold:

(i) The shadow sequence \( (P^U T^n x)_{n \in \mathbb{N}} \) converges weakly to some point in \( U \cap (v + V) \).

(ii) The sequence \( (P^n V T x)_{n \in \mathbb{N}} \) converges weakly to some point in \( (U - v) \cap V \).

**Proof.** (i): Apply Theorem 16.1(i) with \((U, V)\) replaced by \((V, U)\) and use Corollary 10.10. (ii): Combine (i) and Fact 14.7. □

16.3 Spingarn’s method

In this section we discuss the problem to find a least-squares solution of \( \bigcap_{i=1}^M C_i \), i.e., to

\[
\text{find a minimizer of } \sum_{i=1}^M d_{C_i}^2 \tag{16.6}
\]

where \( C_1, \ldots, C_M \) are nonempty closed convex (possibly nonintersecting) subsets of \( X \) with corresponding distance functions \( d_{C_1}, \ldots, d_{C_M} \). Following Pierra [120], we now consider the product Hilbert space \( X = X^M \), with the inner product \( ((x_1, \ldots, x_M), (y_1, \ldots, y_M)) \mapsto \sum_{i=1}^M \langle x_i, y_i \rangle \). We set

\[
U = \{ (x, \ldots, x) \in X \mid x \in X \} \quad \text{and} \quad V = C_1 \times \cdots \times C_M. \tag{16.7}
\]

Then the projections of \( x = (x_1, \ldots, x_M) \in X \) onto \( U \) and \( V \) are given by, respectively, \( P_U x = \left( \frac{1}{M} \sum_{i=1}^M x_i, \ldots, \frac{1}{M} \sum_{i=1}^M x_i \right) \) and \( P_V x = (P_{C_1} x_1, \ldots, P_{C_M} x_M) \). Now assume that

\[
v = (v_1, \ldots, v_M) = P_{U - V} 0 \in U - V. \tag{16.8}
\]

Then we have \( U \cap (v + V) \neq \emptyset \) and

\[
(x, \ldots, x) \in U \cap (v + V) \iff x \in \bigcap_{i=1}^M (v_i + C_i). \tag{16.9}
\]
16.3. Spingarn’s method

Using [10, Section 6], we see that the M-set problem (16.6) is equivalent to the two-set problem

\[
\text{find a least-squares solution of } U \cap V. \quad (16.10)
\]

It follows from (16.8) and (16.9) that v is the unique vector in \( U - V \) that satisfies

\[
\{ (w_1, w_2, \ldots, w_M) \not= (v_1, v_2, \ldots, v_M), \quad \text{and} \quad \bigcap_{i=1}^M (C_i + w_i) \not= \emptyset \} \Rightarrow \sum_{i=1}^M \|w_i\|^2 > \sum_{i=1}^M \|v_i\|^2.
\]

We have the following result for the problem of finding a least-squares solution for the intersection of a finite family of sets.

**Corollary 16.6.** Suppose that \( C_1, \ldots, C_M \) are closed convex subsets of \( X \). Let \( T = \text{Id} - P_U + P_V R_U \), let \( x \in X \) and recall assumption (16.8). Then the shadow sequence \( (P_U T^n x)_{n \in \mathbb{N}} \) converges to \( \bar{x} = (\bar{x}, \ldots, \bar{x}) \in U \cap (v + V) \), where \( \bar{x} \in \bigcap_{i=1}^M (C_i + v_i) \) and \( \bar{x} \) is a least-squares solution of (16.6).

**Proof.** Combine Theorem 16.1 with (16.11) and (16.9). \qed

In Figure 16.2 below we visualize Corollary 16.6 in the case when \( X = \mathbb{R}^2 \), and \( M = 3 \).

**Remark 16.7.** When we particularize Corollary 16.6 from convex sets to half-spaces and \( X \) is finite-dimensional, we recover Spingarn’s [131, Theorem 1]. Note that in this case, in view of [10, Facts 5.1(ii)] we have \( v \in U - V \). Recall that Spingarn used the following version of his method of partial inverses from [129]:

\[
(u_0, v_0) \in U \times U^\perp \quad (16.12)
\]

and

\[
(\forall n \in \mathbb{N}) \quad \begin{cases} u'_n = P_V(u_n + v_n), & v'_n = u_n + v_n - u'_n, \\ u_{n+1} = P_U u'_n, & v_{n+1} = v'_n - P_U v'_n. \end{cases} \quad (16.13)
\]

This method is the Douglas–Rachford algorithm in \( X \), applied to \( U \) and \( V \) with starting point \( (u_0 - v_0) \) (see [37, Lemma 2.17]).
16.3. Spingarn’s method

Figure 16.2: A GeoGebra [78] snapshot that illustrates Corollary 16.6. Three nonintersecting closed convex sets, $C_1$ (the blue triangle), $C_2$ (the red polygon) and $C_3$ (the green circle), are shown along with their translations forming the generalized intersection. The first few terms of the sequence $(e(P_U T^n x))_{n \in \mathbb{N}}$ (yellow points) are also depicted. Here $e : U \rightarrow \mathbb{R}^2 : (x, x, x) \mapsto x$. 
Chapter 17

Conclusion

In this thesis we presented a novel analysis of the behaviour of the Douglas–Rachford algorithm when the underlying sum problem is possibly inconsistent, i.e., the problem had no solution.

Our results represent considerable progress which helps to better understand the inconsistent case. A significant consequence of our new analysis was that we were able to provide a proof of weak convergence of the shadow sequence to a best approximation solution when the algorithm is specialized to solving convex feasibility problems.

The work in this thesis is based on the author’s joint work that appears in the publications [13], [14], [15], [26], [31], [33], [36], [38] and [107] and the submitted manuscripts [39] and [110]. We start by summarizing the main contributions of this thesis.

17.1 Main results

We introduced the new concept of the normal problem and associated normal solutions for the sum of two maximally monotone operators. A nice feature of the theory is that it recovers, as a particular case, the notion of normal solution (obtained by the least squares method) of linear systems. Another interesting and very useful feature is that the theory fits perfectly with Attouch–Théra duality, in the sense that the Douglas–Rachford splitting operators for the primal and dual problems coincide.

To complete the theory, we were able to provide an elegant description of the range of the displacement mapping associated with the Douglas–Rachford operator. This result is of particular importance because it gives an explicit tool to calculate the minimal displacement vector that defines the normal problem.

As a natural follow up, we looked at the algorithmic consequences of our new structure. We provided sufficient conditions for the convergence of the shadow sequence produced by the Douglas–Rachford operator in the possibly inconsistent case. When specialized to the convex feasibility set-
17.2 Side results

We now turn to our new results that arose either as auxiliary tools or as byproducts of the analysis of the inconsistent case.

We systematically studied Attouch–Théra duality for the sum problem problem. We provide new results related to Passty’s parallel sum, to Eckstein and Svaiter’s extended solution set, and to Combettes’ fixed point description of the set of primal solutions. Furthermore, we prove paramonotonicity is a key property because it allows for the recovery of all primal solutions given just one arbitrary dual solution.

We gave a systematic study of nearly convex sets, provided further characterizations, and extended known results on calculus, relative interiors, and applications. Although nearly convex sets need not be convex, we showed that many results on convex sets do extend.

As noticed, even though the sum problem does not depend on the order of the operators, the corresponding Douglas–Rachford operator does. Another question that we tackled was, what can we say about the two Douglas–Rachford operators that can solve the sum problem? We proved that the reflectors of the underlying operators act as bijections between the sets of fixed points of the two operators. Moreover, in some special cases, we were able to relate the two sequences obtained by iterating the two operators.

To make further progress in our convergence analysis, we presented some new Fejér monotonicity principles, which serve as major building blocks in the convergence results.

We developed a comprehensive study of iterates of nonexpansive operators, their corresponding inner/outer shifts and the associated sets of fixed points. When the operators were affine, we got strikingly elegant as
well as useful results. As a consequence, we were able to extend and refine some of the well-known results of linear and strong convergence from linear operators to affine operators.

Finally we presented a new proof to the convergence of the shadow sequence produced by the Douglas–Rachford algorithm in the consistent case.

17.3 Future work

In the following we discuss some possible directions of future research:

• Suppose that \( T \) is nonexpansive. We present a list of open problems that may be easier than the general question (8.46), namely: under which conditions on \( T \) must \( (T^n x - T^n y)_{n \in \mathbb{N}} \) converge weakly? Let \( x \) and \( y \) be in \( X \).

P1: Suppose that \( X = \mathbb{R}, v = 0 \) but \( \text{Fix} T = \emptyset \). Is \( (T^n x - T^n y)_{n \in \mathbb{N}} \) convergent?

P2: Suppose that \( X = \mathbb{R}, v \neq 0 \) but \( \text{Fix}(v + T) = \emptyset \). Is \( (T^n x - T^n y)_{n \in \mathbb{N}} \) convergent?

P3: Does Corollary 8.38 remain true if \( \dim(X) \geq 3 \)? Example 8.39 suggests that the answer to this question might be positive.

P4: What can be said about (8.46) if we replace “weakly” by “strongly”?

• Note that the minimal displacement vector can be found by using Fact 8.2. Conceptionally, we can thus first find \( v_{(A,B)} \) via either iterations in (8.2), and proceed then by iterating the operator \( x \mapsto T(x + v_{(A,B)}) \) to find a normal solution. It would be desirable to devise an algorithm that approximates \( v_{(A,B)} \) and a corresponding normal solution (should it exist) simultaneously. Proposition 11.29, which leads us to solving a quadratic optimization problem, suggests that this may indeed be possible in general.

• In view of Example 14.15, Remark 14.16 and Example 15.6 it is natural to ask the question: what conditions imposed on the operators \( A \) and \( B \) guarantee that the shadow sequence of the Douglas–Rachford algorithm converges to a normal solution, provided there is one?

• One natural extension of the work of this thesis is to examine whether or not this can be extended to the nonconvex feasibility setting, where
17.3. Future work

the operators $A$ and $B$ are normal cones of two sets, each of which is a finite union of convex sets. Convergence analysis of this setting in the consistent case are indeed developed in [16].

• Regarding the inconsistent case, it is tempting to explore the behaviour of the alternating direction method of multipliers (ADMM) in this case.
Bibliography


[9] H. H. Bauschke. New demiclosedness principles for (firmly) non-expansive operators. In Computational and analytical mathematics,


\rightarrow\) pages 67

\rightarrow\) pages 67

\rightarrow\) pages 48, 62, 96, 113

\rightarrow\) pages 103, 109, 114

\rightarrow\) pages 29

\rightarrow\) pages 1

\rightarrow\) pages 1

\rightarrow\) pages 30, 37, 112, 144

\rightarrow\) pages 1

\rightarrow\) pages 173


[99] Maple software. www.maplesoft.com. \(\rightarrow\) pages 90


Bibliography


201


Index

3*- monotone, 17
R-linear convergence, 90

affine hull, 5
affine relation, 11
asymptotically regular operator, 89
asymptotically regular sequence, 72
Attouch–Théra dual problem, 30
averaged operator, 95

best approximation pair, 162
Brezis–Haraux Theorem, 19

cone, 5
conic hull, 5
conjugate function, 9
convex feasibility, 162
convex function, 7
convex hull, 5
convex set, 5
discrete Laplacian, 63
domain of an operator, 11
Douglas–Rachford splitting operator, 47
dual cone, 5
dual normal solutions, 133
dual pair, 30
dual solutions, 30
eigenvalues, 24
epigraph of a function, 7
Fejér monotone, 67
Fejér monotone sequence, 67
Fenchel conjugate, 9
Fermat’s rule, 9
firmly nonexpansive operator, 10
fixed points, 13
governing sequence, 101
graph of an operator, 11
Halpern-type algorithm, 111
Haugazeau-type algorithm, 111
Hilbert space, 5
inconsistent feasibility problem, 164
indicator function, 8
inner perturbation, 128
inner shift, 74
inverse of an operator, 11
inverse resolvent identity, 14
Krasnosel’ski–Mann iteration, 102
Kronecker product of matrices, 24
Laplace operator, 63
least squares solutions, 127
linear convergence, 93
linear relation, 11
lower semicontinuous function, 7
maximally monotone operator, 12
### Index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimal displacement vector</td>
<td>76, 133, 156</td>
</tr>
<tr>
<td>minimizer of a function</td>
<td>8</td>
</tr>
<tr>
<td>Minty parametrization</td>
<td>15</td>
</tr>
<tr>
<td>monotone operator</td>
<td>12</td>
</tr>
<tr>
<td>nearly convex</td>
<td>18</td>
</tr>
<tr>
<td>nearly equal</td>
<td>18</td>
</tr>
<tr>
<td>nonexpansive operator</td>
<td>10</td>
</tr>
<tr>
<td>normal cone</td>
<td>5</td>
</tr>
<tr>
<td>normal problem</td>
<td>126, 133</td>
</tr>
<tr>
<td>normal solutions</td>
<td>133</td>
</tr>
<tr>
<td>operator norm</td>
<td>89</td>
</tr>
<tr>
<td>Ostrowski’s Theorem</td>
<td>67</td>
</tr>
<tr>
<td>outer perturbation</td>
<td>128</td>
</tr>
<tr>
<td>outer shift</td>
<td>74</td>
</tr>
<tr>
<td>parallel space</td>
<td>6</td>
</tr>
<tr>
<td>parallel splitting</td>
<td>110</td>
</tr>
<tr>
<td>parallel sum</td>
<td>34</td>
</tr>
<tr>
<td>paramonotone operator</td>
<td>16</td>
</tr>
<tr>
<td>pointwise convergence of linear operators</td>
<td>90</td>
</tr>
<tr>
<td>polar cone</td>
<td>5</td>
</tr>
<tr>
<td>positive semidefinite</td>
<td>20</td>
</tr>
<tr>
<td>primal normal solutions</td>
<td>133</td>
</tr>
<tr>
<td>primal problem</td>
<td>30</td>
</tr>
<tr>
<td>primal solutions</td>
<td>30</td>
</tr>
<tr>
<td>projection operator</td>
<td>6</td>
</tr>
<tr>
<td>proper function</td>
<td>7</td>
</tr>
<tr>
<td>proximal mapping</td>
<td>8</td>
</tr>
<tr>
<td>range of an operator</td>
<td>11</td>
</tr>
<tr>
<td>recession cone</td>
<td>5</td>
</tr>
<tr>
<td>rectangular</td>
<td>17</td>
</tr>
<tr>
<td>reflected resolvent</td>
<td>13</td>
</tr>
<tr>
<td>relative interior</td>
<td>5</td>
</tr>
<tr>
<td>resolvent</td>
<td>13</td>
</tr>
<tr>
<td>root linear convergence</td>
<td>90</td>
</tr>
<tr>
<td>shadow sequence</td>
<td>101</td>
</tr>
<tr>
<td>shift operator</td>
<td>128</td>
</tr>
<tr>
<td>solution mappings</td>
<td>34</td>
</tr>
<tr>
<td>Spingarn’s method</td>
<td>184</td>
</tr>
<tr>
<td>strictly monotone operator</td>
<td>12</td>
</tr>
<tr>
<td>strong relative interior</td>
<td>5</td>
</tr>
<tr>
<td>subdifferential operator</td>
<td>8</td>
</tr>
<tr>
<td>symmetric matrix</td>
<td>24</td>
</tr>
<tr>
<td>Toeplitz matrix</td>
<td>21</td>
</tr>
<tr>
<td>total duality</td>
<td>45</td>
</tr>
<tr>
<td>tridiagonal matrix</td>
<td>21</td>
</tr>
<tr>
<td>uniform convergence of linear operators</td>
<td>90</td>
</tr>
<tr>
<td>uniformly monotone operator</td>
<td>12</td>
</tr>
<tr>
<td>zeros of a set-valued operator</td>
<td>11</td>
</tr>
</tbody>
</table>