The Resolvent Average: An Expansive Analysis of Firmly Nonexpansive Mappings and Maximally Monotone Operators

by

Sarah Michelle Moffat

B.Sc. Mathematics Honours, The University of British Columbia, 2007
M.Sc., The University of British Columbia, 2009

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 2014

Abstract

Monotone operators and firmly nonexpansive mappings are essential to modern optimization and fixed point theory. Minty first discovered the link between these two classes of operators; every resolvent of a monotone operator is firmly nonexpansive and every firmly nonexpansive mapping is a resolvent of a monotone operator.

This thesis provides an in-depth study of the relationship between firmly nonexpansive mappings and maximally monotone operators. First, corresponding properties between maximally monotone operators and their resolvents are collected. Then a new method of averaging monotone operators is presented, called the *resolvent average*, which is based on the convex combination of the resolvents of monotone operators. Several new results are given concerning the asymptotic regularity of compositions and convex combinations of firmly nonexpansive mappings. Finally, the resolvent average is studied with respect to which properties the average inherits from the averaged operators.

Preface

This thesis is based on the papers [15, 18–21]. The papers [18–21], which form the basis of Chapters 4–6 and Chapter 8 are based on joint work with my supervisors, Dr. Heinz H. Bauschke and Dr. Xianfu Wang. The paper [15] which is the basis of Chapter 7 is based on joint work with Dr. Heinz H. Bauschke, Dr. Victoria Martin-Marquez, and Dr. Xianfu Wang.

For all co-authored papers, each author contributed equally.

Table of Contents

List of Tables

List of Figures

Glossary of Notation

$\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$	The set of continuous linear mappings from $\mathcal{H}_1$ to $\mathcal{H}_2$. 5
$\mathcal{H}$	A real Hilbert space. 3
$\mathbb{N}$	Strictly positive integers, $1, 2, 3, \cdots$. 3
$\mathbb{Q}$	Rational numbers. 3
$\mathbb{R}$	Real numbers. 3
$\mathbb{R}_+$	Nonnegative real numbers. 3
$\mathbb{R}_{++}$	Strictly positive real numbers. 3
$\mathbb{R}^n$	n-dimensional Euclidean space. 3
$\mathbb{S}^N$	Space of $N \times N$ real symmetric matrices. 4
$\mathbb{S}^N_+$	Set of $N \times N$ real symmetric positive semidefinite matrices. 4
$\mathbb{S}^N_{++}$	Set of $N \times N$ real symmetric positive definite matrices. 4
$x_n \to x$	Strong convergence. 4
$x_n \rightharpoonup x$	Weak convergence. 4
$\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})$	Arithmetic average. 17
$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})$	Geometric average. 18
$\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})$	Harmonic average. 17
$\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$	Proximal average. 18
$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$	Resolvent average. 64
$\Gamma_0(\mathcal{H})$	The class of proper lower semi-continuous convex functions from $\mathcal{H} \to \left]-\infty, +\infty\right]$. 12
$\nabla f(x)$	The Gâteaux gradient of f at x, unless otherwise specified to be the Fréchet gradient. 7
∂f	Subdifferential of f. 12
$\operatorname{epi} f$	Epigraph of f. 12
f^*	Fenchel conjugate of f. 12
$\alpha \star f$	Epi-multiplication. 12

$f \square g$	Infimal convolution of f and g. 12
$\operatorname{prox}_f x$	The proximal mapping of f at x. 15
$\langle \cdot, \cdot \rangle$	Inner product. 3
$\|x\|$	The norm of x. 3
$A : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$	Set valued operator A from $\mathcal{H}_1$ to $\mathcal{H}_2$. 7
A^*	The adjoint of A. 5
A_+	The symmetric part of an $N \times N$ matrix A. 4
${}^{\lambda}A$	Yosida λ-regularization of A. 31
$A^\dagger$	Moore-Penrose inverse. 11
$\operatorname{dom} A$	The domain of a set valued operator. 7
$\operatorname{Fix} T$	Fixed points of operator T. 26
$\operatorname{gra} A$	Graph of the operator A. 7
Id	Identity mapping. 5
J_A	The resolvent of the monotone operator A. 31
$\ker A$	The nullspace of the operator A. 5
$\mathfrak{q}$	The quadratic form $\mathfrak{q} = \mathfrak{q}_{\mathrm{Id}} = \langle x, x \rangle$. 6
$\mathfrak{q}_A$	The quadratic form $\mathfrak{q}_A x = \langle x, Ax \rangle$. 6
$\operatorname{ran} T$	The range of a single valued operator. 4
$\operatorname{ran} A$	The range of a set valued operator. 7
$\operatorname{aff} C$	The affine hull of the set C. 8
$\overline{C}$	The closure of C. 8
$\operatorname{cone} C$	The conical hull of the set C. 8
$\operatorname{conv} C$	The convex hull of the set C. 8
$C^\perp$	The orthogonal complement of C. 5
ι_C	The indicator function of the set C. 10
$N_C x$	The normal cone to the set C at x. 11
$P_C x$	The projection of x onto the set C. 10
$\operatorname{ri} C$	The relative interior of the set C. 8
$\operatorname{span} C$	The span of the set C. 8
$A \approx B$	The set A is nearly equal to the set B. 79
$B \preceq A$	$A - B \in \mathbb{S}^N_+$. 4
$B \prec A$	$A - B \in \mathbb{S}^N_{++}$. 4

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr. Heinz Bauschke and Dr. Shawn Wang, for all of their help and guidance throughout my graduate studies. They have been excellent mentors and without them this thesis would not have been possible. I would also like to thank Dr. Yves Lucet and Dr. Stephen Simons for serving on my committee and providing me with valuable insights, and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding this research.

Finally, I would like to recognize the support of all of the staff of Unit 5, who have made it a pleasure to work at the university; in particular, my lab mates Walaa Moursi and Liangjin Yao who supported me along this journey.

Dedication

For Jim, Zoe, and Hailey.

Chapter 1

Introduction

The study of firmly nonexpansive mappings and their connection to monotone operators is motivated by the large number of problems to which these types of operators have been applied. Signal processing, image restoration, and phase retrieval problems are examples that can be solved using projection methods, where projections are a type of firmly nonexpansive operator. The general problem is to find a point x in the intersection of n convex subsets of a Hilbert space. That is, for convex subsets $C_1, \ldots, C_n$ and $C = \bigcap_{i=1}^{n} C_i \neq \varnothing$,

$$\text{Find } x \in C. \tag{1.1}$$

Equation (1.1) is referred to as the *convex feasibility problem.* Numerous algorithms have been created to solve these types of problems and those algorithms make use of the operators studied in this thesis. The simplest example of such an algorithm is the method of alternating projections, where C_1 and C_2 are convex sets with $C_1 \cap C_2 \neq \varnothing$ and the update formula is

$$x_{k+1} = P_{C_1} \circ P_{C_2} x_k,$$

where P_C denotes the projection operator, discussed in more detail in Chapter 2. The method of alternating projections, and variations thereof, was the driving force behind the study of compositions and convex combinations of firmly nonexpansive mappings. The majority of the background theory used in this thesis can be found in Rockafeller's *Convex Analysis*, [58]; Rockafeller and Wet's *Variational Analysis*, [59]; and Bauschke and Combette's *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, [11].

The rest of this thesis is organized as follows:

Chapter 2 gives notations and background information on operators, convex analysis, and methods of averaging operators. Chapter 3 covers known results on nonexpansive mappings and monotone operators.

My contribution begins in Chapter 4, with an in-depth look at which properties of resolvents correspond to properties of their associated monotone operator. Dual and self-dual properties are also identified. The material in this chapter is based on [19].

Chapter 5 then introduces the *resolvent average*, a new method of averaging monotone operators based on the convex combination of the resolvents of the operators. Basic properties of the resolvent average of monotone operators are gathered and properties specific to positive semidefinite matrices are also derived. Results in this chapter can be found in [18] and [21]

Chapter 6, based on the paper [20], uses the notions of near equality and near convexity to study convex combinations of monotone operators and firmly nonexpansive mappings.

In Chapter 7, it is shown that compositions and convex combinations of asymptotically regular mappings maintain asymptotic regularity. This chapter is based on [15].

Chapter 8, based on [21], looks at how properties of monotone operators and their resolvents extend to the resolvent average. Properties are classified according to whether they are

(i) *dominant,* i.e. only one averaged operator needs the property to ensure the average maintains the property, or

(ii) *recessive,* i.e. all average operators need to have the property to ensure the average has the property.

Finally, the key results of this thesis are summarized in Chapter 9.

Chapter 2

Preliminary Details

2.1 Normed vector spaces

We work in a variety of spaces throughout this thesis, most commonly Hilbert and Euclidean spaces, which are both subclasses of the class of Banach spaces.

Definition 2.1. A *Banach space*, $\mathcal{X}$, is a complete normed vector space.

Definition 2.2. A *Hilbert space*, $\mathcal{H}$, is a complete inner product space.

Let $\mathcal{H}$ denote a real Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. The n-dimensional Euclidean space, $\mathbb{R}^n$, is a classic example of a Hilbert space. The real numbers, nonnegative real numbers, and strictly positive real numbers are indicated by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ respectively. We also denote the strictly positive integers, $1, 2, 3, \ldots$ by $\mathbb{N}$ and the rational numbers by $\mathbb{Q}$. Let I be an index set with $I = \{1, 2, \ldots, m\}$ for some integer m and let

$$\mathcal{H}^m = \{\mathbf{x} = (x_i)_{i \in I} \mid (\forall i \in I) \quad x_i \in \mathcal{H}\},$$

denote the *Hilbert product space* with inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i \in I} x_i y_i$.

Clearly, every Euclidean space is a Hilbert space and every Hilbert space is a Banach space.

Example 2.3. [37, Example 1.19(2)] The space of square summable sequences,

$$\ell^2(\mathbb{N}) = \{(x_n)_{n \in \mathbb{N}} \mid \sum_{n=1}^{\infty} |x_n|^2 < \infty\},$$

with inner product $\langle x, y \rangle = \sum\limits_{n=1}^{\infty} x_n y_n$ is a Hilbert space.

Example 2.4. [32, pg. 2] Let $n \geq 2$. Then $\mathbb{R}^n$ with the infinity norm,

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\},$$

is a Banach space that is not a Hilbert space since $\|x\|_\infty$ is not induced by an inner product.

A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ *converges strongly* to a point x if

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

This is written $x_n \to x$. The sequence *converges weakly* to x, or $x_n \rightharpoonup x$, if for every $u \in \mathcal{H}$,

$$\lim_{n\to\infty} \langle x_n, u\rangle = \langle x, u\rangle .$$

In the space $\mathbb{S}^N$ of $N \times N$ real symmetric matrices, $\mathbb{S}^N_+$ denotes the set of $N \times N$ positive semidefinite matrices, and $\mathbb{S}^N_{++}$ the set of positive definite matrices. For $A, B \in \mathbb{S}^N$, we write $B \preceq A$ if $A - B \in \mathbb{S}^N_+$ and $B \prec A$ if $A - B \in \mathbb{S}^N_{++}$.

Example 2.5. [11, Example 2.4] $\mathbb{S}^N$ with inner product $\langle A, B\rangle = \operatorname{tr}(AB)$ is a Hilbert space, where tr is the trace function defined by $\operatorname{tr} A = \sum_{i=1}^{N} a_{ii}$.

A 2×2 matrix A is called a *rotation matrix* if A is of the form

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

for some angle θ. A matrix A is *orthogonal* if $A^{-1} = A^T$, where A^T denotes the *transpose* of A.

Fact 2.6. *[48, 3.7.16] For an invertible matrix A,*

$$\left(A^{-1}\right)^T = \left(A^T\right)^{-1}.$$

The *symmetric part* of an $N \times N$ matrix A is

$$A_+ = \frac{1}{2}(A + A^T).$$

2.2 Operators

2.2.1 Single-valued operators

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces with $D \subseteq \mathcal{H}_1$. Let $T : D \to \mathcal{H}_2$ denote an operator (or mapping) T that maps every point $x \in D$ to a point $Tx \in \mathcal{H}_2$. The range of T is

$$\operatorname{ran} T = \{y \in \mathcal{H}_2 \mid \exists x \in \mathcal{H}_1 \text{ with } Tx = y\}.$$

We set

$$\mathcal{B}(\mathcal{H}_1,\mathcal{H}_2) = \{T : \mathcal{H}_1 \to \mathcal{H}_2 \mid T \text{ is linear and continuous}\}.$$

Fact 2.7. [37, 8.25] *For* $T \in \mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)$, *the* adjoint *of* T *is the unique operator* T^* *that satisfies*

$$(\forall x \in \mathcal{H}_1)(\forall y \in \mathcal{H}_2) \quad \langle Tx, y\rangle = \langle x, T^* y\rangle\,.$$

Example 2.8. Let R be the cyclic right-shift operator,

$$R : \mathcal{H}^m \to \mathcal{H}^m : (x_1, x_2, \dots, x_m) \mapsto (x_m, x_1, \dots, x_{m-1}).$$

Let $x = (x_1, \dots, x_m)$ and $y = (y_1, \dots, y_m) \in \mathcal{H}^m$, then R^* satisfies

$$\begin{aligned}\langle Rx, y\rangle = \langle x, R^* y\rangle &\Leftrightarrow \langle (x_m, x_1, \dots, x_{m-1}), (y_1, y_2, \dots y_m)\rangle = \langle x, R^* y\rangle\\ &\Leftrightarrow \langle x, R^* y\rangle = y_1x_m + y_2x_1 + \dots + y_mx_{m-1}\\ &\Leftrightarrow \langle x, R^* y\rangle = x_1y_2 + x_2y_3 + \dots + x_{m-1}y_m + x_my_1.\end{aligned}$$

Thus R^* is the *cyclic left-shift operator*

$$R^* : \mathcal{H}^m \to \mathcal{H}^m : (x_1, x_2, \dots, x_m) \mapsto (x_2, x_3, \dots, x_m, x_1).$$

The kernel of T is $\ker T = \{x \in \mathcal{H} \mid Tx = 0\}$.

Fact 2.9. [37, Lemma 8.33(2)] *Let* $T \in \mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)$. *Then*

$$(\ker T)^\perp = \overline{\operatorname{ran} T^*},$$

where $(\ker T)^\perp$ *denotes the* orthogonal complement *of* $\ker T$, *i.e.,*

$$(\ker T)^\perp = \{u \in \mathcal{H}_1 \mid (\forall x \in \ker T)\ \langle x, u\rangle = 0\}.$$

If $\mathcal{H}_2 \subseteq \mathcal{H}_1$, then $T^n x$ denotes the n-fold composition of T. The identity mapping is the operator $\mathrm{Id} : \mathcal{H} \to \mathcal{H} : x \mapsto x$.

Definition 2.10. Let $T : \mathcal{H} \to \mathcal{H}$. T is *Lipschitz continuous* with constant β if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \le \beta\|x - y\|.$$

If $\beta \in\,]0, 1[$, then T is called a *Banach contraction.*

Definition 2.11. T is *sequentially weakly continuous* if for every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $x_n \rightharpoonup x$, then $Tx_n \rightharpoonup Tx$.

Definition 2.12. A mapping $T : \mathcal{H} \to \mathcal{H}$ is an *isometry* if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| = \|x - y\|. \tag{2.1}$$

When $T : \mathcal{H} \to \mathcal{H}$ is linear, the quadratic form $\mathfrak{q}_T : \mathcal{H} \to \mathbb{R}$ is defined by

$$\mathfrak{q}_T(x) = \tfrac{1}{2}\langle Tx, x\rangle \quad \forall x \in \mathcal{H},$$

and $\mathfrak{q}_{\mathrm{Id}} = \mathfrak{q}$ is used interchangeably.

Fact 2.13. [11, Corollary 15.34] *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces and $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. Then* $\operatorname{ran} T$ *is closed if and only if* $\operatorname{ran} T^*$ *is closed.*

Fact 2.14 (Closed Range Theorem)**.** [37, Theorem 8.18] *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces and $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2) \setminus \{0\}$. Then the following are equivalent:*

(i) *T has closed range;*

(ii) *There exists $\rho > 0$ such that $\|Tx\| \geq \rho\|x\|$ for all $x \in (\ker T)^\perp$;*

(iii) *$\rho := \inf\{\|Tx\| \mid x \in (\ker T)^\perp, \|x\| = 1\} > 0$.*

Definition 2.15 (Gâteaux differentiability)**.** Let $T : C \to \mathcal{X}$, with $C \subseteq \mathcal{H}$ and $\mathcal{X}$ a real Banach space. Let $x \in C$ be such that $(\forall y \in \mathcal{H})(\exists \alpha \in \mathbb{R}_{++})$ $[x, x + \alpha y] \subseteq C$. Then T is *Gâteaux differentiable* at x if there exists an operator $DT(x) \in \mathcal{B}(\mathcal{H}, \mathcal{X})$, called the *Gâteaux derivative* of T at x, such that

$$(\forall y \in \mathcal{H}) \quad DT(x)y = \lim_{\alpha \to 0^+} \frac{T(x + \alpha y) - T(x)}{\alpha}.$$

Remark 2.16. Unless otherwise noted, when differentiability is mentioned then Gâteaux differentiability is assumed.

Definition 2.17 (Fréchet differentiability)**.** Let $x \in \mathcal{H}$ and let $T : U \to \mathcal{X}$, where U is an open subset of $\mathcal{H}$ and $\mathcal{X}$ is a real Banach space. Then T is *Fréchet differentiable* at x if there exists an operator $DT(x) \in \mathcal{B}(\mathcal{H}, \mathcal{X})$, called the *Fréchet derivative* of T at x such that

$$\lim_{0 \neq \|y\| \to 0} \frac{\|T(x + y) - Tx - DT(x)y\|}{\|y\|} = 0.$$

Fact 2.18 (Fréchet-Riesz Representation Theorem)**.** [37, Theorem 6.10] *Let* $f \in \mathcal{B}(\mathcal{H}, \mathbb{R})$*. Then there exists a unique vector* $u \in \mathcal{H}$ *such that*

$$(\forall x \in \mathcal{H}) \quad f(x) = \langle x, u \rangle .$$

Moreover, $\|f\| = \|u\|$*.*

Let $C \subseteq \mathcal{H}$, $f : C \to \mathbb{R}$ and suppose that f is Fréchet differentiable at $x \in C$. Then by Fact 2.18, there exists a unique vector $\nabla f(x) \in \mathcal{H}$ such that

$$(\forall y \in \mathcal{H}) \quad Df(x)y = \langle y, \nabla f(x) \rangle .$$

We call $\nabla f(x)$ the *Fréchet gradient* of f at x. If f is Fréchet differentiable on C the *gradient operator* is

$$\nabla f : C \to \mathcal{H} : x \mapsto \nabla f(x).$$

The *Gâteaux gradient* is defined similarly.

2.2.2 Set-valued operators

An operator $A : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$ is *set-valued* if $(\forall x \in \mathcal{H}_1)$ $Ax \subseteq \mathcal{H}_2$. For set-valued operators,

$$\operatorname{dom} A = \{x \mid \ Ax \neq \varnothing\},$$

and

$$\operatorname{ran} A = \bigcup_{x \in \operatorname{dom} A} Ax.$$

A set-valued operator A is characterized by its graph

$$\operatorname{gra} A = \{(x, u) \in \mathcal{H}_1 \times \mathcal{H}_2 \mid u \in Ax\}.$$

The set-valued inverse A^{-1} of A is defined by

$$(y, x) \in \operatorname{gra} A^{-1} \;\Leftrightarrow\; (x, y) \in \operatorname{gra} A.$$

The operator A is *at most single-valued* if Ax is a singleton or $Ax = \varnothing$. The sum of operators,

$$A + B : x \mapsto Ax + Bx,$$

and therefore $\operatorname{gra}(A+B) = \{(x, u+v) \in \mathcal{H}_1 \times \mathcal{H}_2 \mid (x, u) \in \operatorname{gra} A \text{ and } (x, v) \in \operatorname{gra} B\}$ and $\operatorname{dom}(A + B) = \operatorname{dom} A \cap \operatorname{dom} B$.

Definition 2.19. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$. Then A is a *linear relation* if $\operatorname{gra} A$ is a linear subspace of $\mathcal{H} \times \mathcal{H}$. Similarly, A is an *affine relation* if $\operatorname{gra} A$ is an affine subspace of $\mathcal{H} \times \mathcal{H}$, i.e. if

$$\operatorname{gra} A \neq \varnothing \text{ and } (\forall \lambda \in \mathbb{R}) \quad \operatorname{gra} A = \lambda \operatorname{gra} A + (1-\lambda) \operatorname{gra} A.$$

See [36] for more on linear relations.

Fact 2.20. [36, I.2.3 and I.4] *Let A, B be linear relations on $\mathcal{H}$. Then A^{-1} and $A + B$ are linear relations.*

Definition 2.21. An operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is *disjointly injective* if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \ \Rightarrow \ Ax \cap Ay = \varnothing. \tag{2.2}$$

2.3 Convex analysis

2.3.1 Convex sets

A subset C of $\mathcal{H}$ is *convex* if for all $x, y \in C$ and $\lambda \in \left]0,1\right[$,

$$\lambda x + (1-\lambda) y \in C.$$

The *closure* of C is denoted by $\overline{C}$. A set C is *sequentially weakly closed* if every weakly convergent sequence $(x_n)_{n\in\mathbb{N}}$ in C has its weak limit x also in C.

The intersection of all the convex sets containing C is called the *convex hull* of C, and is denoted by $\operatorname{conv} C$. The intersection of all affine subspaces containing C is likewise called the *affine hull* of C and is denoted by $\operatorname{aff} C$.

A subset C of $\mathcal{H}$ is a cone if $C = \mathbb{R}_{++}C$. That is, $x \in C$ and $\lambda > 0$ implies $\lambda x \in C$. A *convex cone* is a set that is both convex and a cone. The conical hull of C, $\operatorname{cone} C$, is the intesection of all the cones in $\mathcal{H}$ containing C. The smallest linear subspace of $\mathcal{H}$ containing C is $\operatorname{span} C$. The *interior* of C is the largest open set contained in C,

$$\operatorname{int} C = \{x \mid \exists \epsilon > 0, B(x,\epsilon) \subseteq C\}.$$

The *relative interior* of C is

$$\operatorname{ri} C = \{x \in \operatorname{aff} C \mid \exists \epsilon > 0, B(x,\epsilon) \cap \operatorname{aff} C \subseteq C\},$$

where $B(x, \epsilon)$ is a ball centered at x with radius ϵ.

Lemma 2.22. *Let A and B be subsets of $\mathbb{R}^n$ such that $A \subseteq B$ and $\operatorname{aff} A = \operatorname{aff} B$. Then $\operatorname{ri} A \subseteq \operatorname{ri} B$.*

Proof. This follows directly from the definition. $\square$

Fact 2.23. [58, pg. 44] *Let A be a subset of $\mathbb{R}^n$. Then $\overline{A} \subseteq \operatorname{aff} A$.*

Lemma 2.24. *Let A and B be subsets of $\mathbb{R}^n$ such that $\overline{A} = \overline{B}$. Then $\operatorname{aff} A = \operatorname{aff} B$.*

Proof. Let $x \in \operatorname{aff} A$. Then $x = \lambda_1 a_1 + \cdots + \lambda_m a_m$ for $a_i \in A$, $\lambda_i \in \mathbb{R}$, $i = 1, \dots, m$, and $\sum_{i=1}^{m} \lambda_i = 1$. Clearly, each $a_i \in \overline{A} = \overline{B}$ and by Fact 2.23, $\overline{B} \subseteq \operatorname{aff} B$, so x is an affine combination of elements in $\operatorname{aff} B$. Hence $x \in \operatorname{aff} B$. Altogether, $\operatorname{aff} A \subseteq \operatorname{aff} B$. Similarly, you can show $\operatorname{aff} B \subseteq \operatorname{aff} A$, and thus $\operatorname{aff} A = \operatorname{aff} B$. $\square$

Fact 2.25 (Rockafellar)**.** *Let C and D be convex subsets of $\mathbb{R}^n$, and let $\lambda \in \mathbb{R}$. Then the following hold.*

(i) $\operatorname{ri} C$ *and* $\overline{C}$ *are convex.*

(ii) $C \neq \varnothing \Rightarrow \operatorname{ri} C \neq \varnothing$.

(iii) $\overline{\operatorname{ri} C} = \overline{C}$.

(iv) $\operatorname{ri} C = \operatorname{ri} \overline{C}$.

(v) $\operatorname{aff} \operatorname{ri} C = \operatorname{aff} C = \operatorname{aff} \overline{C}$.

(vi) $\operatorname{ri} C = \operatorname{ri} D \Leftrightarrow \overline{C} = \overline{D} \Leftrightarrow \operatorname{ri} C \subseteq D \subseteq \overline{C}$.

(vii) $\operatorname{ri} \lambda C = \lambda \operatorname{ri} C$.

(viii) $\operatorname{ri}(C + D) = \operatorname{ri} C + \operatorname{ri} D$.

Proof. (i)&(ii): See [58, Theorem 6.2]. (iii)&(iv): See [58, Theorem 6.3]. (v): See [58, Theorem 6.2]. (vi): See [58, Corollary 6.3.1]. (vii): See [58, Corollary 6.6.1]. (viii): See [58, Corollary 6.6.2]. $\square$

Fact 2.26. [58, Theorem 6.5] *Let C_i be a convex set in $\mathbb{R}^n$ for $i = 1, \dots, m$ such that $\bigcap_{i=1}^{m} \operatorname{ri} C_i \neq \varnothing$. Then*

$$\overline{\bigcap_{i=1}^{m} C_i} = \bigcap_{i=1}^{m} \overline{C_i},$$

and

$$\operatorname{ri}\bigcap_{i=1}^{m} C_i = \bigcap_{i=1}^{m} \operatorname{ri} C_i.$$

Fact 2.27. [59, Proposition 2.40] *Let $C \neq \varnothing$ be a convex subset of $\mathbb{R}^n$. Then* $\operatorname{ri} C$ *is nonempty and convex with*

$$\overline{\operatorname{ri} C} = \overline{C}.$$

Definition 2.28. Let C be a convex subset of $\mathcal{H}$, the *indicator function* of C at x is

$$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{if } x \notin C. \end{cases} \tag{2.3}$$

Fact 2.29 (projection)**.** [11, Definition 3.7]*Let C be a nonempty, closed, convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then there exists a unique vector $p \in C$ such that*

$$\|x - p\| = \inf_{y \in C} \|x - y\|,$$

and p is called the projection *of x onto the set C, denoted by $P_C x$.*

Fact 2.30 (projection characterization)**.** [11, Theorem 3.14] *Let C be a closed convex subset of $\mathcal{H}$. For every x and p in $\mathcal{H}$, $p = P_C x$ if and only if*

$$p \in C \text{ and } (\forall y \in C) \quad \langle y - p, x - p \rangle \leq 0. \tag{2.4}$$

Example 2.31. Let $x \in \mathbb{R}^2$ and $C = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = x_2\}$. Then $P_C(x) = \frac{1}{2}(x_1 + x_2, x_1 + x_2)$.

Proof. Let $x = (x_1, x_2)$ and $y = (y_1, y_1) \in C$. Clearly,

$$p = \frac{1}{2}(x_1 + x_2, x_1 + x_2) \in C,$$

and

$$\begin{aligned}
&\langle y - p, x - p \rangle \\
&= \left\langle (y_1, y_1) - \frac{1}{2}(x_1 + x_2, x_1 + x_2), (x_1, x_2) - \frac{1}{2}(x_1 + x_2, x_1 + x_2) \right\rangle \\
&= \frac{1}{2} y_1 (x_1 - x_2) - \frac{1}{4}(x_1 + x_2)(x_1 - x_2) - \frac{1}{2} y_1 (x_1 - x_2) \\
&\qquad - \frac{1}{4}(x_1 + x_2)(x_2 - x_1) \\
&= 0.
\end{aligned}$$

Thus by Fact 2.30, $p = \frac{1}{2}(x_1 + x_2, x_1 + x_2) = P_C(x)$. $\square$

Example 2.32. Let $\mathbf{\Delta} \subseteq \mathcal{H}^m$ with $\mathbf{\Delta} = \big\{\mathbf{x} = (x)_{i\in I} \;\big|\; x \in \mathcal{H}\big\}$ and $\mathbf{x} \in \mathcal{H}^m$, then $P_{\mathbf{\Delta}}\mathbf{x}$ satisfies

$$\begin{aligned} \|\mathbf{x} - P_{\mathbf{\Delta}}\mathbf{x}\|^2 &= \inf_{\mathbf{y}\in\mathbf{\Delta}} \|\mathbf{x} - \mathbf{y}\|^2 \\ &= \inf_{\mathbf{y}\in\mathbf{\Delta}} \left((x_1 - y)^2 + (x_2 - y)^2 + \cdots + (x_m - y)^2\right). \end{aligned} \tag{2.5}$$

Since $\|\cdot\|^2$ is convex, differentiating (2.5) with respect to y, setting it equal to zero and solving for y yields

$$\begin{gathered} -2(x_1 - y) - 2(x_2 - y) - \cdots - 2(x_m - y) = 0 \\ \Leftrightarrow y = \frac{1}{m}\sum_{i=1}^{m} x_i \end{gathered}$$

Thus, $P_{\mathbf{\Delta}}\mathbf{x} = \left(\frac{1}{m}\sum\limits_{i=1}^{m} x_i, \frac{1}{m}\sum\limits_{i=1}^{m} x_i, \ldots, \frac{1}{m}\sum\limits_{i=1}^{m} x_i\right)$.

Definition 2.33. Let C be a nonempty convex subset of $\mathcal{H}$ and $x \in \mathcal{H}$. Then the *normal cone operator* to C at x is

$$N_C x = \begin{cases} \{u \in \mathcal{H} \mid \sup \langle C - x, u\rangle \le 0\} & \text{if } x \in C; \\ \varnothing & \text{otherwise.} \end{cases}$$

Definition 2.34. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces, let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, let $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$. Then x is a *least-squares solution* to the equation $Tz = y$ if

$$\|Tx - y\| = \min_{z\in\mathcal{H}_1} \|Tz - y\|.$$

Fact 2.35. [11, Proposition 3.25] *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces, let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ be such that* $\operatorname{ran} T$ *is closed, and let $y \in \mathcal{H}_2$. Then the equation $Tz = y$ has at least one least-squares solution. Moreover, for every $x \in \mathcal{H}_1$, the following are equivalent:*

(i) *x is a least-squares solution,*

(ii) *$Tx = P_{\operatorname{ran} T}\, y$,*

(iii) *$T^*Tx = T^*y$ (normal equation).*

Definition 2.36 (Moore-Penrose inverse)**.** Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces, let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ be such that $\operatorname{ran} T$ is closed and for every $y \in \mathcal{H}_2$, set $C_y = \{x \in \mathcal{H}_1 \mid T^*Tx = T^*y\}$. The *Moore-Penrose inverse* of T is

$$T^\dagger : \mathcal{H}_2 \to \mathcal{H}_1 : y \mapsto P_{C_y} 0.$$

See [41] for more on the Moore-Penrose inverse.

Fact 2.37. [11, Proposition 3.28(v)] *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces, let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ be such that* $\operatorname{ran} T$ *is closed. Then*

$$\operatorname{ran} T^\dagger = \operatorname{ran} T^*.$$

2.3.2 Convex functions

A function $f : \mathcal{H} \to \,]-\infty, +\infty] = \mathbb{R} \cup \{+\infty\}$ is said to be *convex* if its (essential) domain, $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$, is a convex set and $\forall\, x, y \in \mathcal{H}$, $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \tag{2.6}$$

with f being *strictly convex* if (2.6) becomes a strict inequality whenever $x \neq y$. A function f is *proper* if

$$(\forall x \in \mathcal{H})\ f(x) > -\infty \text{ and } (\exists x_0 \in \mathcal{H}) \text{ such that } f(x_0) < +\infty.$$

A function f is *lower semi-continuous* if for every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$,

$$x_n \to x \Rightarrow f(x) \leq \liminf_{n\to\infty} f(x_n).$$

The *epigraph* of f is

$$\operatorname{epi} f = \{(x, r) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq r\}.$$

For $\alpha > 0$, *epi-multiplication* is

$$\alpha \star f = \alpha f(\cdot/\alpha).$$

The *lower semi-continuous hull* of f is the function whose epigraph is the closure in $\mathcal{H} \times \mathbb{R}$ of the epigraph of f.

The class of proper lower semi-continuous convex functions from $\mathcal{H} \to \,]-\infty, +\infty]$ will be denoted by $\Gamma_0(\mathcal{H})$. For $f \in \Gamma_0(\mathcal{H})$, ∂f denotes its *convex subdifferential*,

$$\partial f(x) = \{x^* \in \mathcal{H} :\ f(y) \geq f(x) + \langle x^*, y - x\rangle\ \ \forall y \in \mathcal{H}\}.$$

If f is continuous and differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$, see [63, Theorem 2.4.4(i)]. The function f^* denotes its *Fenchel conjugate* given by

$$(\forall x^* \in \mathcal{H})\ f^*(x^*) = \sup_x \{\langle x^*, x\rangle - f(x)\}.$$

If $f, g \in \Gamma_0(\mathcal{H})$, $f \Box g$ stands for the *infimal convolution* of f with g given by

$$(\forall x \in \mathcal{H}) \quad (f \Box g)(x) = \inf\{f(x_1) + g(x_2) :\ x_1 + x_2 = x\}.$$

Fact 2.38. [11, Example 16.12] *Let C be a convex subset of $\mathcal{H}$. Then*

$$\partial\iota_C = N_C.$$

Example 2.39. Set $C = \{0\}$ and let $x \in \mathcal{H}$. Then by Fact 2.38, we have

$$\begin{aligned}\partial\iota_{\{0\}}(x) &= N_{\{0\}}x \\ &= \begin{cases}\{u \in \mathcal{H} \mid \sup\langle 0, u\rangle \le 0\} & \text{if } x \in \{0\}; \\ \varnothing & \text{otherwise.}\end{cases} \\ &= \begin{cases}\mathcal{H} & \text{if } x = 0; \\ \varnothing & \text{otherwise.}\end{cases}\end{aligned}$$

Definition 2.40. Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and let $\beta \in \mathbb{R}_{++}$. Then f is *strongly convex* with constant β if $(\forall x \in \operatorname{dom} f)$, $(\forall y \in \operatorname{dom} f)$ and $(\forall \lambda \in \left]0, 1\right[)$,

$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\frac{\beta}{2}\|x - y\|^2 \le \lambda f(x) + (1-\lambda)f(y).$$

Fact 2.41. [11, Proposition 10.6] *Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and let $\beta \in \mathbb{R}_{++}$. Then f is strongly convex with constant β if and only if $f - \beta\mathfrak{q}$ is convex.*

Fact 2.42. [11, Lemma 2.13] *Let $(x_i)_{i\in I}$ and $(u_i)_{i\in I}$ be finite families in $\mathcal{H}$, and let $(\alpha_i)_{i\in I}$ be a family in $\mathbb{R}$ such that $\sum_{i\in I}\alpha_i = 1$. Then the following holds*

$$\left\langle \sum_{i\in I}\alpha_i x_i, \sum_{j\in I}\alpha_j u_j \right\rangle + \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\left\langle x_i - x_j, u_i - u_j\right\rangle/2 = \sum_{i\in I}\alpha_i\left\langle x_i, u_i\right\rangle.$$

In particular, $\|\cdot\|^2$ is strongly convex and

$$\|\sum_{i\in I}\alpha_i x_i\|^2 = \sum_{i\in I}\alpha_i\|x_i\|^2 - \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\|x_i - x_j\|^2/2. \tag{2.7}$$

Fact 2.43. [10, Theorem 2.1] *Let $f \in \Gamma_0(\mathcal{H})$ and let $\beta \in \mathbb{R}_{++}$. Then the following are equivalent*

(i) *f is Fréchet differentiable on $\mathcal{H}$ and ∇f is β-Lipschitz continuous.*

(ii) *f^* is $\frac{1}{\beta}$-strongly convex.*

Definition 2.44. [8, 58] A proper convex function f on $\mathbb{R}^N$ is *essentially strictly convex* if f is strictly convex on every convex subset of $\operatorname{dom} \partial f$.

Definition 2.45. [8, 58] A proper convex function f on $\mathbb{R}^N$ is *essentially smooth* if it satisfies the following conditions for $C := \operatorname{int}(\operatorname{dom} f)$:

(i) C is not empty;

(ii) f is differentiable throughout C;

(iii) $\lim_{n\to\infty} |\nabla f(x_n)| = +\infty$ whenever $(x_n)_{n\in\mathbb{N}}$ is a sequence in C converging to a point $x \in \operatorname{bdry} C := \overline{C} \setminus \operatorname{int} C$.

Definition 2.46. Let $f \in \Gamma_0(\mathbb{R}^N)$. Then f is *Legendre* if f is essentially smooth and essentially strictly convex.

Fact 2.47. [58, Theorem 26.1] *Let f be a closed proper convex function. Then ∂f is a single-valued mapping if and only if f is essentially smooth.*

Fact 2.48. [58, Theorem 26.3] *A closed proper convex function f is essentially strictly convex if and only if f^* is essentially smooth.*

Definition 2.49. Let $f : \mathcal{H} \to [-\infty, +\infty]$. Then f is *coercive* if

$$\lim_{\|x\|\to+\infty} f(x) = +\infty,$$

and f is *supercoercive* if

$$\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Fact 2.50. [11, Proposition 12.15] *Let $f \in \Gamma_0(\mathcal{H})$. Then the infimal convolution,*

$$f \Box \mathfrak{q} : \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \inf_{y\in\mathcal{H}} \left(f(y) + \mathfrak{q}(x-y)\right),$$

is convex, real-valued, continuous, and the infimum is uniquely attained.

Remark 2.51. In Fact 2.50 the existence of a minimizer follows from the supercoercivity of $\mathfrak{q}$ while the uniqueness follows from the strict convexity of $\mathfrak{q}$. This motivates the next definition.

Definition 2.52 (proximal mapping)**.** [11, Definition 12.23] Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then $\operatorname{prox}_f x$ is the unique point in $\mathcal{H}$ which satisfies

$$\min_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2}\|x - y\|^2 \right) = f(\operatorname{prox}_f x) + \frac{1}{2}\|x - \operatorname{prox}_f x\|^2.$$

The operator $\operatorname{prox}_f x : \mathcal{H} \to \mathcal{H}$ is the *proximal mapping* or *proximity operator* of f.

Fact 2.53. [11, Proposition 16.34] *Let $f \in \Gamma_0(\mathcal{H})$ and let $x, p \in \mathcal{H}$. Then*

$$p = \operatorname{prox}_f x \Leftrightarrow x - p \in \partial f(p).$$

In other words,

$$\operatorname{prox}_f x = (\operatorname{Id} + \partial f)^{-1}.$$

Fact 2.54. [59, Exercise 11.27] or [11, Remark 14.4] *Let $f \in \Gamma_0(\mathcal{H})$. Then*

$$\operatorname{prox}_f = \nabla(f^* \,\square\, \mathfrak{q}),$$

where $\nabla(f^* \,\square\, \mathfrak{q})$ is the Fréchet gradient of $f^* \,\square\, \mathfrak{q}$.

Example 2.55. Set $f = \|\cdot\|$, then

$$\operatorname{prox}_f x = \begin{cases} \left(1 - \frac{1}{\|x\|}\right) x & \text{if } \|x\| > 1; \\ 0 & \text{if } \|x\| \leq 1. \end{cases}$$

Proof. Set $g(y) = \|y\| + \frac{1}{2}\|x - y\|^2$. If $x = 0$, then clearly $y = 0$ is the minimizer of $g(y)$. We consider two cases: Case 1: $\|x\| \leq 1$. If $y = 0$, $g(0) = \frac{1}{2}\|x\|^2$. If $\|y\| = \|x\|$, we have

$$\begin{aligned} g(y) &= \|y\| + \frac{1}{2}\|x\|^2 - \langle x, y\rangle + \frac{1}{2}\|y\|^2 \\ &\geq \|y\| + \frac{1}{2}\|x\|^2 - \|x\|\|y\| + \frac{1}{2}\|y\|^2 \\ &= \|x\| + \frac{1}{2}\|x\|^2 - \|x\|\|x\| + \frac{1}{2}\|x\|^2 = \|x\| \\ &\geq \frac{1}{2}\|x\|^2, \end{aligned}$$

so any y such that $\|y\| = \|x\|$ is not the minimizer. Clearly if $\|y\| > \|x\|$ then $g(y) > g(0)$. Finally, if $\|y\| = \lambda\|x\|$ for some $\lambda \in \,]0, 1[$ we have

$$\begin{aligned} g(y) &\geq \|y\| + \frac{1}{2}\|x\|^2 - \|x\|\|y\| + \frac{1}{2}\|y\|^2 \\ &= \lambda\|x\| + \frac{1}{2}\|x\|^2 - \lambda\|x\|^2 + \frac{\lambda^2}{2}\|x\|^2 \\ &= \lambda(\|x\| - \|x\|^2) + \frac{1}{2}(1 + \lambda^2)\|x\|^2. \end{aligned}$$

The first term is greater than or equal to zero, while the second term is greater than $\frac{1}{2}\|x\|^2$. Altogether $y = 0$ is the minimizer if $\|x\| < 1$.

Case 2: $\|x\| > 1$. When $\|x\| > 1$, $g(x) < g(0)$. Thus $y = 0$ is not the minimizer and since $g(y)$ is convex, differentiating with respect to y and settling equal to zero will yield the minimizer. Doing this, we have

$$\begin{aligned} &\frac{y}{\|y\|} - (x-y) = 0 \\ \Leftrightarrow\ & x = y\left(\frac{1}{\|y\|} + 1\right). \end{aligned} \tag{2.8}$$

Taking norms of each side of (2.8) gives,

$$\begin{aligned} &\|x\| = \|y\|\left(\frac{1}{\|y\|} + 1\right) \\ \Leftrightarrow\ &\|x\| = 1 + \|y\| \Leftrightarrow \|y\| = \|x\| - 1. \end{aligned}$$

So if $\|x\| > 1$, (2.8) becomes $y = x\left(1 - \frac{1}{\|x\|}\right)$ and if $\|x\| \leq 1$, $y = 0$ is the minimizer. $\square$

Fact 2.56. [58, Theorem 16.4] *Let $f_1, \cdots, f_n$ be proper convex functions on $\mathbb{R}^N$. Then*

$$(f_1 \square \cdots \square f_n)^* = f_1^* + \cdots + f_n^*. \tag{2.9}$$

If the sets $\mathrm{ri}(\mathrm{dom}\, f_i)$, $i = 1, \cdots, n$ *have a point in common, then*

$$(f_1 + \cdots + f_n)^*(x^*) = \inf_{x_1^* + \cdots + x_n^* = x^*} \left(f_1^*(x_1^*) + \cdots + f_n^*(x_n^*)\right), \tag{2.10}$$

where for each x^* the infimum is attained.

Fact 2.57. [58, page 108] *Let $A \in \mathbb{S}_{++}^N$. Then*

$$(\mathfrak{q}_A)^* = \mathfrak{q}_{A^{-1}}.$$

Fact 2.58. [58, Theorem 12.3] *Let $A \in \mathbb{S}_{++}^N$ be an injective linear operator, a and $b \in \mathbb{R}^N$ and $r \in \mathbb{R}$. Set*

$$f(x) = \mathfrak{q}_A(x-a) + \langle x, b\rangle + r,$$

Then

$$f^*(x^*) = \mathfrak{q}_{A^{-1}}(x^* - b) + \langle x^*, a\rangle - \langle a, b\rangle - r.$$

Fact 2.59. [58, Theorem 23.5] *Let* $f \in \Gamma_0(\mathbb{R}^n)$. *Then*

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

Fact 2.60. [58, Theorem 25.7] *Let* C *be a nonempty open convex subset of* $\mathbb{R}^N$*, and let* f *be a convex function which is finite and differentiable on* C. *Let* $f_1, f_2, \ldots,$ *be a sequence of convex functions finite and differentiable on* C *such that* $\lim_{i\to\infty} f_i(x) = f(x)$ *for every* $x \in C$. *Then*

$$\lim_{i\to\infty} \nabla f_i(x) = \nabla f(x), \quad \forall x \in C.$$

In fact, the sequence of gradients ∇f_i *converges to* ∇f *uniformly on every compact subset of* C.

2.4 Averages

There are many methods of averaging; this section gathers the definitions of some methods that will be of interest.

2.4.1 Arithmetic and harmonic averages

The most commonly used averages are the arithmetic, harmonic, and geometric averages. Let A_i, $i = 1, \ldots, n$ be $N \times N$ positive semidefinite matrices, λ_i be strictly positive real coefficients with $\sum_{i=1}^{n} \lambda_i = 1$, $\mathbf{A} = (A_1, \ldots, A_n)$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$.

Definition 2.61. (Arithmetic average) The $\boldsymbol{\lambda}$-weighted arithmetic average of $\mathbf{A}$ is

$$\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}) = \lambda_1 A_1 + \cdots + \lambda_n A_n. \tag{2.11}$$

Definition 2.62. (Harmonic average) The $\boldsymbol{\lambda}$-weighted harmonic average of $\mathbf{A}$ is

$$\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) = (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1}. \tag{2.12}$$

2.4.2 Geometric mean

For matrices $A, B \in \mathbb{S}^N_{++}$, the geometric mean is defined by

$$A \sharp B = A^{\frac{1}{2}} \big(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}\big)^{\frac{1}{2}} A^{\frac{1}{2}}.$$

There have been several suggestions for how to define the geometric mean of $A_1, \ldots, A_n \in \mathbb{S}^N_+$ for $n \geq 3$, [1, 44, 51, 54].

Definition 2.63. (Geometric mean) Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ with $x_i > 0$ for all $i = 1, \dots, n$. The $\boldsymbol{\lambda}$-weighted geometric average of $\mathbf{x}$ is

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}.$$

The weighted geometric mean always has the following properties:

Fact 2.64. *Let* $\mathbf{x} = (x_1, \dots, x_n)$ *and* $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ *such that* $(\forall\ i)\ x_i > 0, y_i > 0$*, and* $\mathbf{x}^{-1} = (x_1^{-1}, \dots, x_n^{-1})$*. Let* $\lambda_i \in \mathbb{R}_{++}$ *such that* $\sum_{i=1}^{n} \lambda_i = 1$*. Then we have*

(i) (**harmonic-geometric-arithmetic mean inequality**)*:*

$$\left(\lambda_1 x_1^{-1} + \cdots + \lambda_n x_n^{-1}\right)^{-1} \leq \mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq \lambda_1 x_1 + \cdots + \lambda_n x_n.$$

Moreover, $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ *if and only* $x_1 = \cdots = x_n$*.*

(ii) (**self-duality**)*:* $[\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})]^{-1} = \mathcal{G}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda})$*.*

(iii) *If* $\boldsymbol{x} = (x_1, \dots, x_1)$*, then* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1$*.*

(iv) *If* $\mathbf{z} = (x_1, x_1^{-1}, x_2, x_2^{-1}, \dots, x_n, x_n^{-1})$ *and* $\mu = (\frac{1}{2n}, \dots, \frac{1}{2n})$*, then* $\mathcal{G}(\mathbf{z}, \mu) = 1$*.*

(v) *The function* $\boldsymbol{x} \mapsto \mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})$ *is concave on* $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$*.*

(vi) *If* $\boldsymbol{x} \succeq \boldsymbol{y}$*, then* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \geq \mathcal{G}(\boldsymbol{y}, \boldsymbol{\lambda})$*.*

Proof. (i): See [58, page 29]. (ii)-(iv) and (vi) are simple. (v): See [59, Example 2.53]. ☐

2.4.3 Proximal average

One key tool used later is the *proximal average of convex functions*, which finds its roots in [16, 50, 52], and which has been further systematically studied in [12–14, 22].

Definition 2.65 (proximal average)**.** Let $(\forall i)$ $f_i \in \Gamma_0(\mathcal{H})$ and λ_i be strictly positive real numbers with $\sum_{i=1}^{n} \lambda_i = 1$. The $\boldsymbol{\lambda}$-weighted proximal average of $\boldsymbol{f} = (f_1, \dots, f_n)$ with parameter $\mu > 0$ is defined by

$$\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \left(\lambda_1 (f_1 + \tfrac{1}{\mu}\mathfrak{q})^* + \lambda_2 (f_2 + \tfrac{1}{\mu}\mathfrak{q})^* + \cdots + \lambda_n (f_n + \tfrac{1}{\mu}\mathfrak{q})^*\right)^* - \tfrac{1}{\mu}\mathfrak{q}. \tag{2.13}$$

The function $\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is a proper lower semi-continuous convex function on $\mathcal{H}$, and it inherits many desirable properties from each underlying function f_i; see [12, 13]. The next fact is a fundamental property of the proximal average.

Fact 2.66. [12, Theorem 5.1]

$$\big(\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \mathcal{P}_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}).$$

Lemma 2.67. [11, Lemma 2.13(ii)] *Let $x_1, \dots x_n \in \mathbb{R}^N$ and $\lambda_i \in \mathbb{R}_{++}$ such that $\sum\limits_{i=1}^n \lambda_i = 1$. Then the following identity holds:*

$$\sum_{i=1}^n \lambda_i \mathfrak{q}(x_i) - \mathfrak{q}(\sum_{i=1}^n \lambda_i x_i) = \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \|x_i - x_j\|^2.$$

Proof. From Fact 2.42, we have

$$\|\sum_{i=1}^n \lambda_i x_i\|^2 - \sum_{i=1}^n \lambda_i \|x_i\|^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \|x_i - x_j\|^2. \tag{2.14}$$

Multiplying (2.14) by $\frac{1}{2}$ on each side gives the desired identity. □

The following reformulation of the proximal average will be useful.

Proposition 2.68. *Let $f_1, \dots, f_n \in \Gamma_0(\mathbb{R}^N)$ and $\lambda_1, \dots, \lambda_n > 0$ with $\sum_{i=1}^n \lambda_i = 1$. Then for every $x \in \mathbb{R}^N$,*

$$\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$$

$$= \min_{x_1 + \cdots + x_n = x} \left\{ \lambda_1 (f_1 + \frac{1}{\mu}\mathfrak{q})(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n (f_n + \frac{1}{\mu}\mathfrak{q})(\frac{x_n}{\lambda_n}) \right\} - \frac{1}{\mu}\mathfrak{q}(x) \tag{2.15}$$

$$= \min_{x_1 + \cdots + x_n = x} \left\{ \lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{4\mu} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \|\frac{x_i}{\lambda_i} - \frac{x_j}{\lambda_j}\|^2 \right\} \tag{2.16}$$

$$= \min_{\lambda_1 y_1 + \cdots + \lambda_n y_n = x} \Big\{ \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{\mu} [\lambda_1 \mathfrak{q}(y_1) + \cdots + \lambda_n \mathfrak{q}(y_n) - \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n)] \Big\} \tag{2.17}$$

$$= \min_{\lambda_1 y_1+\cdots+\lambda_n y_n = x} \left\{ \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{4\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \|y_i - y_j\|^2 \right\} \tag{2.18}$$

$$= \min_{x_1+\cdots+x_n = x} \Big\{ \lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{\mu} \Big[\lambda_1 \mathfrak{q}(x - \frac{x_1}{\lambda_1}) + \cdots$$
$$+ \lambda_n \mathfrak{q}(x - \frac{x_n}{\lambda_n}) \Big] \Big\}. \tag{2.19}$$

Proof. Indeed, as

$$\left(f_i + \frac{1}{\mu} \mathfrak{q} \right)^* = f_i^* \Box (\mu \mathfrak{q}),$$

it is finite-valued everywhere, we write

$$f = \lambda_1 \star (f_1 + \frac{1}{\mu}\mathfrak{q}) \Box \cdots \Box \lambda_n \star (f_n + \frac{1}{\mu}\mathfrak{q}) - \frac{1}{\mu}\mathfrak{q},$$

by Fact 2.56. That is, for every x,

$$f(x) = \inf_{x_1+\cdots+x_n = x} \left\{ \lambda_1 (f_1 + \frac{1}{\mu}\mathfrak{q})(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n (f_n + \frac{1}{\mu}\mathfrak{q})(\frac{x_n}{\lambda_n}) \right\} - \frac{1}{\mu}\mathfrak{q}(x),$$

and the infimum is attained, again by Fact 2.56. Hence, replacing inf with min we get (2.15).

Now rewrite (2.15) as

$$\min_{x_1+\cdots+x_n = x} \Big\{ \lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{\mu} \Big[\lambda_1 \mathfrak{q}(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n \mathfrak{q}(\frac{x_n}{\lambda_n})$$
$$- \mathfrak{q}(x_1 + \cdots + x_n) \Big] \Big\}, \tag{2.20}$$
$$= \min_{\lambda_1 y_1+\cdots+\lambda_n y_n = x} \Big\{ \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{\mu} [\lambda_1 \mathfrak{q}(y_1) + \cdots + \lambda_n \mathfrak{q}(y_n)$$
$$- \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n)] \Big\}.$$

Thus, (2.16)–(2.18) follow by using Lemma 2.67. Next, recall that

$$x = x_1 + \cdots + x_n,$$

and observe that by expanding and simplifying we get

$$\begin{aligned}
&\lambda_1\mathfrak{q}(x_1+\cdots+x_n-\frac{x_1}{\lambda_1})+\cdots+\lambda_n\mathfrak{q}(x_1+\cdots+x_n-\frac{x_n}{\lambda_n})\\
&=\frac{\lambda_1}{2}\left\|x-\frac{x_1}{\lambda_1}\right\|^2+\cdots+\frac{\lambda_n}{2}\left\|x-\frac{x_n}{\lambda_n}\right\|^2\\
&=\sum_{i=1}^n\frac{\lambda_i}{2}\left\langle x-\frac{x_i}{\lambda_i},x-\frac{x_i}{\lambda_i}\right\rangle\\
&=\sum_{i=1}^n\frac{\lambda_i}{2}\left(\|x\|^2-2\left\langle x,\frac{x_i}{\lambda_i}\right\rangle+\|\frac{x_i}{\lambda_i}\|^2\right)\\
&=\frac{1}{2}\|x\|^2-\|x\|^2+\sum_{i=1}^n\lambda_i\mathfrak{q}(\frac{x_i}{\lambda_i})\\
&=\lambda_1\mathfrak{q}(\frac{x_1}{\lambda_1})+\cdots+\lambda_n\mathfrak{q}(\frac{x_n}{\lambda_n})-\mathfrak{q}(x_1+\cdots+x_n),
\end{aligned}$$

thus we have (2.19) by (2.20). □

Fact 2.69 (inequalities)**.** [12, Theorem 5.4]

$$(\lambda_1 f_1^*+\cdots+\lambda_n f_n^*)^*\le\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})\le\lambda_1 f_1+\cdots+\lambda_n f_n.$$

Fact 2.70. [12, Example 4.5] *Let $\alpha_1,\ldots,\alpha_n$ be strictly positive real numbers and suppose that $(\forall i)$ $f_i=\alpha_i\mathfrak{q}$. Then*

$$\begin{aligned}
\mathcal{P}_{\mu^{-1}}(\boldsymbol{f},\boldsymbol{\lambda})&=\left(\sum_{i=1}^n\lambda_i(\alpha_i\mathfrak{q}+\mu\mathfrak{q})^*\right)^*-\mu\mathfrak{q}=\left(\sum_{i=1}^n\frac{\lambda_i}{\alpha_i+\mu}\mathfrak{q}\right)^*-\mu\mathfrak{q}\\
&=\left(\sum_{i=1}^n\frac{\lambda_i}{\alpha_i+\mu}\right)^{-1}\mathfrak{q}-\mu\mathfrak{q}.
\end{aligned}$$

And thus,

$$\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})=\left(\left(\sum_{i=1}^n\frac{\lambda_i}{\alpha_i+\mu^{-1}}\right)^{-1}-\mu^{-1}\right)\mathfrak{q}.$$

Fact 2.71. [12, Corollary 7.7] *Suppose that at least one function f_i is essentially smooth and that $\lambda_i>0$. Then $\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ is essentially smooth.*

Fact 2.72. [12, Theorem 8.5] *Let $x\in\mathbb{R}^N$. Then the function*

$$]0,+\infty[\ \to\ ]-\infty,+\infty]:\mu\mapsto\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)\quad\textit{is decreasing.}\tag{2.21}$$

Consequently, $\lim_{\mu\to 0^+} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *and* $\lim_{\mu\to+\infty} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *exist. In fact,*

$$\lim_{\mu\to 0^+} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \sup_{\mu>0} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 f_1 + \cdots + \lambda_n f_n\big)(x) \tag{2.22}$$

and

$$\lim_{\mu\to+\infty} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \inf_{\mu>0} \mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 \star f_1 \Box \cdots \Box \lambda_n \star f_n\big)(x). \tag{2.23}$$

We have now covered the building blocks needed for the main focus of this thesis, nonexpansive mappings and monotone operators. In the next chapter, we introduce several different notions of "nonexpansiveness" and monotonicity and cover many of the known results about these kinds of operators.

Chapter 3

Nonexpansive Mappings and Monotone Operators

This chapter contains a collection of known results involving nonexpansive mappings and monotone operators. We begin with the concept of a nonexpansive mapping.

3.1 Nonexpansive mappings

Definition 3.1. Let D be a nonempty subset of $\mathcal{H}$. A mapping $T : D \to \mathcal{H}$ is

(i) *nonexpansive*, or *Lipschitz continuous* with constant 1, if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq \|x - y\|; \tag{3.1}$$

(ii) *strictly nonexpansive* if

$$(\forall x \in D)(\forall y \in D) \quad x \neq y \Rightarrow \|Tx - Ty\| < \|x - y\|; \tag{3.2}$$

(iii) *firmly nonexpansive* if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx-Ty\|^2 + \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2 \leq \|x-y\|^2; \tag{3.3}$$

(iv) a *Banach contraction*, or *Lipschitz continuous* with constant β, if there exists $\beta \in [0, 1[$ such that

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leq \beta\|x - y\|; \tag{3.4}$$

(v) *strongly nonexpansive* if T is nonexpansive and whenever $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are sequences in D such that $(x_n - y_n)_{n\in\mathbb{N}}$ is bounded and $\|x_n - y_n\| - \|Tx_n - Ty_n\| \to 0$, it follows that $(x_n - y_n) - (Tx_n - Ty_n) \to 0$.

Remark 3.2. Clearly, both firmly nonexpansive and strongly nonexpansive imply nonexpansive. And in Hilbert spaces, Bruck and Reich showed that firmly nonexpansive implies strongly nonexpansive, see Fact 3.23. The opposite implication does not hold, see Example 3.6. Thus we have

$$\text{firmly nonexpansive} \Rightarrow \text{strongly nonexpansive} \Rightarrow \text{nonexpansive}.$$

Fact 3.3. [11, Proposition 4.2] *Let D be a nonempty subset of $\mathcal{H}$ and $T\colon D \to \mathcal{H}$. Then the following are equivalent:*

(i) *T is firmly nonexpansive.*

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

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

(iv) *$(\forall x \in D)(\forall y \in D)\ \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle$.*

(v) *$(\forall x \in D)(\forall y \in D)\ 0 \leq \langle Tx - Ty, (\operatorname{Id} - T)x - (\operatorname{Id} - T)y\rangle$.*

Example 3.4. The identity mapping is both strongly nonexpansive and firmly nonexpansive. However, when $T = -\operatorname{Id}$, T is nonexpansive but it fails to be strongly nonexpansive, and consequently fails to be firmly nonexpansive.

To see that, let $x, y \in \mathcal{H}$. To see that $T = -\operatorname{Id}$ is not strongly nonexpansive, set $x_n = Tx_{n-1} = T^n x_0$ and $y_n = Ty_{n-1} = T^n y_0$. Then $(x_n - y_n)$ is bounded and for all $n \in \mathbb{N}$,

$$\|x_n - y_n\| - \|Tx_n - Ty_n\| = \|x_n - y_n\| - \| - x_n + y_n\| = 0.$$

But,

$$\begin{aligned}(x_n - y_n) - (Tx_n - Ty_n) &= (x_n - y_n) - (-x_n + y_n) = 2(x_n - y_n)\\ &= 2(T^n x_0 - T^n y_0),\end{aligned}$$

which only goes to zero if $x_0 = y_0$, so T is not strongly nonexpansive, and consequently not firmly nonexpansive.

Example 3.5. [11, Proposition 4.8] Let C be a nonempty closed convex subset of $\mathcal{H}$. Then the projection operator P_C is firmly nonexpansive.

Example 3.6. Let $x, y \in \mathbb{R}^2$, let $C = \mathbb{R} \times \{0\}$ and $D = \{x \in \mathbb{R}^2 \mid x_1 = x_2\}$. Clearly, $P_C(x) = (x_1, 0)$ and by Example 2.31 $P_D(x) = \frac{1}{2}(x_1 + x_2, x_1 + x_2)$.

By Example 3.5, P_C and P_D are firmly nonexpansive. Then consider $T(x) = P_C P_D(x) = \frac{1}{2}(x_1 + x_2, 0)$ with the points $x = (1, 2)$ and $y = (1, 3)$,

$$\|Tx - Ty\|^2 = \left\| \frac{1}{2}(3,0) - \frac{1}{2}(4,0) \right\|^2 = \frac{1}{4}$$

and

$$\begin{aligned} \langle x - y, Tx - Ty \rangle &= \left\langle (1,2) - (1,3), \frac{1}{2}(3,0) - \frac{1}{2}(4,0) \right\rangle \\ &= \left\langle (0,-1), (-\frac{1}{2}, 0) \right\rangle = 0. \end{aligned}$$

Thus $\|Tx - Ty\|^2 > \langle x - y, Tx - Ty \rangle$, so by Fact 3.3(iv) T is not firmly nonexpansive.

T is strongly nonexpansive though. Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}^2$ such that $(x_n - y_n)$ is bounded and $\|x_n - y_n\| - \|Tx_n - Ty_n\| \to 0$. Set $x_n = (x_1^n, x_2^n)$, $y_n = (y_1^n, y_2^n)$, $d_n = x_1^n - y_1^n$ and $e_n = x_2^n - y_2^n$. Now,

$$\begin{aligned} &\|x_n - y_n\| - \|Tx_n - Ty_n\| \to 0 \\ \Leftrightarrow\ &\|x_n - y_n\|^2 - \|Tx_n - Ty_n\|^2 \to 0 \\ \Leftrightarrow\ &(x_1^n - y_1^n)^2 + (x_2^n - y_2^n)^2 - \left(\frac{1}{2}(x_1^n - y_1^n + x_2^n - y_2^n)\right)^2 \to 0 \\ \Leftrightarrow\ &d_n^2 + e_n^2 - \frac{1}{4}(d_n + e_n)^2 \to 0 \\ \Leftrightarrow\ &\frac{3}{4}d_n^2 + \frac{3}{4}e_n^2 - \frac{1}{2}d_n e_n \to 0 \\ \Leftrightarrow\ &\frac{1}{4}\left(2d_n^2 + 2e_n^2 + (d_n - e_n)^2\right) \to 0. \end{aligned}$$

Thus we have

$$e_n^2 \to 0 \text{ and } (d_n - e_n)^2 \to 0. \tag{3.5}$$

And we see that

$$\begin{aligned} (x_n - y_n) - (Tx_n - Ty_n) &= (d_n, e_n) - (\frac{1}{2}d_n + \frac{1}{2}e_n, 0) \\ &= (\frac{1}{2}d_n - \frac{1}{2}e_n, e_n). \end{aligned}$$

Taking the norm,

$$\|(\frac{1}{2}d_n - \frac{1}{2}e_n, e_n)\|^2 = \frac{1}{4}(d_n - e_n)^2 + e_n^2,$$

which goes to zero by (3.5). Thus T is strongly nonexpansive.

Remark 3.7. Example 3.6 shows both that strongly nonexpansive does not imply firmly nonexpansive and that the composition of two firmly nonexpansive operators may fail to be firmly nonexpansive.

Definition 3.8. Let $D \subseteq \mathcal{H}$ with $D \neq \varnothing$ and $T : D \to \mathcal{H}$ be nonexpansive. Let $\alpha \in]0, 1[$. Then T is *averaged* with constant α if there exists a nonexpansive operator $N : D \to \mathcal{H}$ such that $T = (1 - \alpha)\operatorname{Id} + \alpha N$.

Fact 3.9. *T is firmly nonexpansive if and only if T is $1/2$-averaged.*

Proof. This follows directly from Fact 3.3(iii). ☐

Definition 3.10. Let $D \subseteq \mathcal{H}$ with $D \neq \varnothing$ and $T : D \to \mathcal{H}$ and let $\beta \in \mathbb{R}_{++}$. Then T is *β-cocoercive* if βT is firmly nonexpansive. That is,

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y, Tx - Ty\rangle \geq \beta\|Tx - Ty\|^2.$$

Remark 3.11. T being β-cocoercive is the same as T^{-1} being β strongly monotone, see Definition 3.31(iv). Thus β-cocoercive is also referred to as being β-inverse strongly monotone.

Fact 3.12 (Baillon-Haddad Theorem)**.** [4, Corollaire 10] or [11, Corollary 18.16] *Let $f : \mathcal{H} \to \mathbb{R}$ be a Fréchet differentiable convex function and let $\beta \in \mathbb{R}_{++}$. Then ∇f is β-Lipschitz continuous if and only if ∇f is $(1/\beta)$-cocoercive. In particular, ∇f is nonexpansive if and only if ∇f is firmly nonexpansive.*

Remark 3.13. For more on the Baillon-Haddad theorem, see [4] and [10].

Definition 3.14. T is *cyclically firmly nonexpansive* if for every set of points $\{x_1, \ldots, x_n\} \subseteq \mathcal{H}$, where $n \in \{2, 3, \ldots\}$ and $x_{n+1} = x_1$, we have

$$\sum_{i=1}^{n} \langle x_i - Tx_i, Tx_i - Tx_{i+1}\rangle \geq 0. \tag{3.6}$$

3.2 Fixed points and asymptotic regularity

Several problems in science and engineering can be formulated as fixed point problems, where the set of desired solutions is the set of fixed points of T,

$$\operatorname{Fix} T := \big\{x \in \mathcal{H} \;\big|\; x = Tx\big\}. \tag{3.7}$$

If T is firmly nonexpansive and $\operatorname{Fix} T \neq \varnothing$, then the sequence of iterates

$$(T^n x)_{n\in\mathbb{N}} \tag{3.8}$$

converges weakly to a fixed point [29]. The iterates $x_{n+1} = Tx_n$, for all $n \in \mathbb{N}$, are referred to as *Banach-Picard* iterates. However, if the mapping is simply nonexpansive then this result does not hold. For example, $T = -\operatorname{Id}$ is nonexpansive with $\operatorname{Fix} T = \{0\}$, but $(T^n x)_{n\in\mathbb{N}}$ converges only if you begin at the fixed point $x = 0$.

Fact 3.15. [11, Corollary 4.15] *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $T : C \to \mathcal{H}$ be nonexpansive. Then* $\operatorname{Fix} T$ *is closed and convex.*

Fact 3.16. [62, Lemma 1.8, Corollary 2] *Let C be a closed convex subset of $\mathcal{H}$ and let $T : \mathcal{H} \to \mathcal{H}$ be a firmly nonexpansive mapping such that*

$$\operatorname{ran} T \subseteq \operatorname{Fix} T = C.$$

Then $T = P_C$.

Proof. Let $x \in \mathcal{H}$ and $y \in C$. Then,

$$Tx \in \operatorname{ran} T \subseteq C \text{ and } y = Ty \in C = \operatorname{Fix} T.$$

Since T is firmly nonexpansive, by Fact 3.3(v)

$$\begin{aligned} 0 &\leq \langle Tx - Ty, (x - Tx) - (y - Ty)\rangle \\ \Leftrightarrow 0 &\leq \langle Tx - y, x - Tx\rangle \\ \Leftrightarrow \langle y - Tx, x - Tx\rangle &\leq 0. \end{aligned}$$

Thus by Fact 2.30, $Tx = P_C x$. $\square$

Definition 3.17. A mapping $T : \mathcal{H} \to \mathcal{H}$ is *asymptotically regular* if

$$(\forall x \in \mathcal{H}) \quad T^n x - T^{n+1} x \to 0.$$

T is *weakly asymptotically regular* if the convergence is weak.

Fact 3.18. [3, Theorem 1.2] *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping. Then $(T^n x)_{n\in\mathbb{N}}$ converges weakly to a fixed point of T if and only if* $\operatorname{Fix} T \neq \varnothing$ *and T is weakly asymptotically regular.*

Fact 3.19. [3, Corollary 2.2] *Let C be a closed convex subset of $\mathcal{H}$. Let $U : C \to \mathcal{H}$ be an averaged nonexpansive mapping. Then* $\operatorname{Fix} U = \varnothing$ *if and only if $\lim\limits_{n\to\infty} \|U^n x\| = \infty$ for all x in C.*

Fact 3.20. [3, Corollary 2.3] *Let C be a closed convex subset of $\mathcal{H}$ and let $U : C \to X$ be an averaged nonexpansive mapping. Then for each $x \in C$*

$$\lim_{n\to\infty}(U^n x - U^{n+1}x) \to v,$$

where v is the element of least norm in $\overline{\operatorname{ran}(\operatorname{Id} - U)}$.

Remark 3.21. Facts 3.18 - 3.20 were originally formulated in a Banach space with additional structure. Details for how those results apply in Hilbert spaces are provided in Appendix A.

Remark 3.22. Suppose $T\colon \mathcal{H} \to \mathcal{H}$ is asymptotically regular. Then, for every $x \in \mathcal{H}$,

$$\begin{aligned} T^n x - T^{n+1}x &\to 0 \\ \Leftrightarrow (\operatorname{Id} - T)T^n x &\to 0 \end{aligned}$$

and hence $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T)}$. The opposite implication fails in general (consider $T = -\operatorname{Id}$), but it is true for strongly nonexpansive mappings, see Fact 3.24.

The next result illustrates that strongly nonexpansive mappings generalize the notion of firmly nonexpansive mappings. In addition, the class of strongly nonexpansive mappings is closed under compositions.

Fact 3.23 (Bruck and Reich)**.** [30, Proposition 2.1 and Proposition 1.1] *In a Hilbert space $\mathcal{H}$, the following hold.*

(i) *Every firmly nonexpansive mapping is strongly nonexpansive.*

(ii) *The composition of finitely many strongly nonexpansive mappings is also strongly nonexpansive.*

In contrast, the composition of two (necessarily firmly nonexpansive) projectors may fail to be firmly nonexpansive, see Example 3.6

The sequences of iterates and of differences of iterates have striking convergence properties as we shall see now.

Fact 3.24 (Bruck and Reich)**.** [30, Corollary 1.5, Corollary 1.4, and Corollary 1.3] *Let $S\colon \mathcal{H} \to \mathcal{H}$ be strongly nonexpansive and let $x \in \mathcal{H}$. Then the following hold.*

(i) *The sequence $(S^n x - S^{n+1}x)_{n\in\mathbb{N}}$ converges strongly to the unique element of least norm in $\overline{\operatorname{ran}(\operatorname{Id} - S)}$.*

(ii) *If* $\operatorname{Fix} S = \varnothing$*, then* $\|S^n x\| \to +\infty$*.*

(iii) *If* $\operatorname{Fix} S \neq \varnothing$*, then* $(S^n x)_{n\in\mathbb{N}}$ *converges weakly to a fixed point of* S*.*

Fact 3.25. [57, Corollary 2] *Let* D *be a subset of* $\mathcal{H}$ *and let* $T : D \to D$ *be firmly nonexpansive. Set* $d = \inf\limits_{y\in D} \|y - Ty\|$*, then for each* $x \in D$*,*

$$\lim_{n\to\infty} \|T^{n+1}x - T^n x\| = d.$$

3.3 Monotone operators

We now look at known results for monotone operators.

Definition 3.26. A set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is *monotone* if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y, u - v\rangle \geq 0. \tag{3.9}$$

A monotone operator A is *maximally monotone* if there exists no monotone operator B such that $\operatorname{gra} A \subset \operatorname{gra} B$. That is, for every $(x,u) \in \mathcal{H} \times \mathcal{H}$,

$$(x,u) \in \operatorname{gra} A \Leftrightarrow (\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y, u - v\rangle \geq 0. \tag{3.10}$$

Lemma 3.27. *Let* $A : \mathcal{H} \to \mathcal{H}$ *be linear. Then* A *is monotone if and only if*

$$(\forall z \in \mathcal{H}) \quad \langle z, Az\rangle \geq 0.$$

Proof. Since A is linear it is single-valued, thus (3.9) becomes

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y, Ax - Ay\rangle \geq 0.$$

Set $z = x - y$ and by linearity we get

$$\langle z, Az\rangle \geq 0.$$

$\square$

Lemma 3.28. *Let* $A : \mathcal{H} \rightrightarrows \mathcal{H}$ *and* $\lambda \in \mathbb{R}_{++}$*. Then*

$$(x,u) \in \operatorname{gra} A \Leftrightarrow (x, \lambda u) \in \operatorname{gra} \lambda A.$$

Proof. Take $(x,u) \in \operatorname{gra} A$. Then $u \in Ax \Rightarrow \lambda u \in \lambda Ax$, i.e. $(x, \lambda u) \in \operatorname{gra} \lambda A$. On the other hand, let $(x, \lambda u) \in \operatorname{gra} \lambda A$, then $\lambda u \in \lambda Ax \Rightarrow u \in Ax$. Altogether, $(x,u) \in \operatorname{gra} A \Leftrightarrow (x, \lambda u) \in \operatorname{gra} \lambda A$. $\square$

Proposition 3.29. *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone and $\lambda \in \mathbb{R}_{++}$. Then λA is maximally monotone.*

Proof. Let $(x, u) \in \operatorname{gra} \lambda A$, then by Lemma 3.28 $(x, \lambda^{-1}u) \in \operatorname{gra} A$. Since A is maximally monotone, (3.10) gives

$$(x, \lambda^{-1}u) \in \operatorname{gra} A \Leftrightarrow (\forall (y, \lambda^{-1}v) \in \operatorname{gra} A) \quad \left\langle x - y, \lambda^{-1}u - \lambda^{-1}v \right\rangle \geq 0.$$

Then for all $(y, v) \in \operatorname{gra} \lambda A$,

$$\langle x - y, u - v \rangle = \lambda \left\langle x - y, \lambda^{-1}u - \lambda^{-1}v \right\rangle \geq 0.$$

Conversely, let $(x, u) \in \mathcal{H} \times \mathcal{H}$ such that $(\forall (y, v) \in \operatorname{gra} \lambda A)$ $\langle x - y, u - v \rangle \geq 0$. Then

$$\langle x - y, u - v \rangle = \lambda \left\langle x - y, \lambda^{-1}u - \lambda^{-1}v \right\rangle \geq 0 \Rightarrow \left\langle x - y, \lambda^{-1}u - \lambda^{-1}v \right\rangle \geq 0.$$

That is, for every $(x, \lambda^{-1}u) \in \mathcal{H} \times \mathcal{H}$ and for every $(y, \lambda^{-1}v) \in \operatorname{gra} A$, $\left\langle x - y, \lambda^{-1}u - \lambda^{-1}v \right\rangle \geq 0$. Thus by (3.10), $(x, \lambda^{-1}u) \in \operatorname{gra} A$ and therefore $(x, u) \in \operatorname{gra} \lambda A$. □

Fact 3.30 (monotonicity versus convexity)**.** [59, Theorem 12.17] *Let $\mathcal{H}$ be finite dimensional and let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximal monotone, and f is essentially strictly convex if and only if ∂f is strictly monotone.*

Definition 3.31. An operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is

(i) *paramonotone* if it is monotone and

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \qquad \langle x - y, u - v \rangle = 0 \Rightarrow (x, v) \in \operatorname{gra} A.$$

(ii) *strictly monotone* if

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad x \neq y \Rightarrow \langle x - y, u - v \rangle > 0.$$

(iii) *uniformly monotone* with modulus $\phi : \mathbb{R}_+ \to [0, +\infty]$ if ϕ is increasing, vanishes only at zero, and

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad \langle x - y, u - v \rangle \geq \phi\left(\|x - y\|\right).$$

(iv) *strongly monotone* with constant $\beta \in \mathbb{R}_{++}$ if $A - \beta \operatorname{Id}$ is monotone. That is,

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad \langle x - y, u - v \rangle \geq \beta \|x - y\|^2.$$

Example 3.32. Let $z \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_{++}$. Set $\phi(z) = \beta z^2$, then it is clear that every operator that is strongly monotone with constant β is also uniformly monotone with modulus ϕ.

Example 3.33. [11, Example 22.3(iv)] Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and strongly convex with constant $\beta \in \mathbb{R}_{++}$. Then ∂f is strongly monotone with constant β.

Definition 3.34. Let $A : \mathcal{H} \to \mathcal{H}$ and $\alpha \in \mathbb{R}$. A is *hemicontinuous* if for every $(x, y, z) \in \mathcal{H}^3$,

$$\lim_{\alpha \to 0^+} \langle z, A(x + \alpha y) \rangle = \langle z, Ax \rangle .$$

Fact 3.35. [11, Example 22.9(iii)] *Let $A : \mathcal{H} \to \mathcal{H}$ be strongly monotone and hemicontinuous, and let $r \in \mathcal{H}$. Then the equation $Ax = r$ has exactly one solution.*

Now let A be a monotone operator from $\mathcal{H} \rightrightarrows \mathcal{H}$ and denote the associated *resolvent* by

$$J_A = (\mathrm{Id} + A)^{-1}. \tag{3.11}$$

For $\lambda > 0$, the *Yosida λ-regularization* of A is,

$${}^{\lambda}A = \lambda^{-1}(\mathrm{Id} - J_{\lambda A}). \tag{3.12}$$

The resolvent satisfies the useful *resolvent identity,*

$$J_A = \mathrm{Id} - J_{A^{-1}}, \tag{3.13}$$

which allows for the *Minty parametrization*

$$\operatorname{gra} A = \big\{ (J_A x, x - J_A x) \;\big|\; x \in \operatorname{dom} J_A \big\} \tag{3.14}$$

of the graph of A, which provides the bijection $x \mapsto (J_A x, x - J_A x)$ from $\operatorname{dom} J_A$ onto $\operatorname{gra} A$, with inverse $(x, u) \mapsto x + u$. The Yosida regularization is related to the resolvent through the following identity,

$${}^{\lambda}A = (\lambda \,\mathrm{Id} + A^{-1})^{-1} = \lambda^{-1}[\mathrm{Id} - (\mathrm{Id} + \lambda A)^{-1}]. \tag{3.15}$$

When $A = \partial f$ for some $f \in \Gamma_0(\mathcal{H})$ then Fact 2.53 yields that

$$J_{\partial f} = \operatorname{prox}_f . \tag{3.16}$$

Minty observed that J_A is in fact a firmly nonexpansive operator from $\mathcal{H}$ to $\mathcal{H}$ and that, conversely, every firmly nonexpansive operator arises this way:

Fact 3.36. (See [38] and [50].) *Let $T\colon \mathcal{H} \to \mathcal{H}$ and let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$. Then the following hold.*

(i) *If T is firmly nonexpansive, then $B := T^{-1} - \mathrm{Id}$ is maximally monotone and $J_B = T$.*

(ii) *If A is maximally monotone, then J_A has full domain, and is single-valued and firmly nonexpansive, and $A = J_A^{-1} - \mathrm{Id}$.*

Definition 3.37. [11, Definition 21.9] Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $x \in \mathcal{H}$. Then A is *locally bounded* at x if there exists $\delta \in \mathbb{R}_{++}$ such that $A(B(x;\delta))$ is bounded, where $B(x;\delta)$ is the closed ball centered at x with radius δ.

Fact 3.38. [11, Corollary 21.19] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone. Then A is surjective if and only if A^{-1} is locally bounded everywhere on $\mathcal{H}$.*

Fact 3.39. [11, Corollary 21.21] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone with bounded domain. Then A is surjective.*

Fact 3.40. [11, Proposition 20.22] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let u and z be in $\mathcal{H}$ and $\gamma \in \mathbb{R}_{++}$. Then A^{-1} and $x \mapsto u + \gamma A(x+z)$ are maximally monotone.*

Fact 3.41. [11, Example 20.41] *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then N_C is maximally monotone.*

Fact 3.42. [11, Example 23.4] *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then*

$$J_{N_C} = (\mathrm{Id} + N_C)^{-1} = \mathrm{prox}_{\iota_C} = P_C.$$

Fact 3.43 (Minty's Theorem)**.** [11, Theorem 21.1] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. Then A is maximally monotone if and only if* $\mathrm{ran}(\mathrm{Id} + A) = \mathcal{H}$.

Remark 3.44. Minty's Theorem provides a characterization for maximal monotonicity which allows for determining maximality without having to show graph inclusions.

Fact 3.45. [11, Proposition 23.11] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone and let $\beta \in \mathbb{R}_{++}$. Then A is strongly monotone with constant β if and only if J_A is $(\beta+1)$-cocoercive, in which case J_A is Lipschitz continuous with constant* $1/(\beta+1) \in \left]0,1\right[$.

Fact 3.46. [11, Example 20.26] *Let $T : \mathcal{H} \to \mathcal{H}$ be nonexpansive and let $\alpha \in [-1,1]$. Then $\mathrm{Id} + \alpha T$ is maximally monotone.*

While the sum of two monotone operators is still monotone, the sum of two maximally monotone operators can fail to be maximally monotone.

Example 3.47. Let $\mathcal{H} = \mathbb{R}^2$ and set C to be the closed unit ball centered at $(-1, 0)$ and D be the closed unit ball centered at $(1, 0)$. By Fact 3.41, both N_C and N_D are maximally monotone and we have

$$\operatorname{dom} N_C \cap \operatorname{dom} N_D = \{(0,0)\} \neq \varnothing.$$

But given Fact 3.39 and the fact that

$$\operatorname{ran}(N_C + N_D) = \mathbb{R} \times \{0\},$$

$N_C + N_D$ is not maximally monotone.

The next fact gives some constraint qualifications under which the sum is maximally monotone.

Fact 3.48 (Rockafellar)**.** [59, Theorem 12.44] and [11, Corollary 24.4] *Let A and B be maximally monotone on $\mathcal{H}$. Suppose one of the following holds:*

(i) $\operatorname{dom} A \cap \operatorname{int} \operatorname{dom} B \neq \varnothing$.

(ii) *If* $\mathcal{H} = \mathbb{R}^N$, $\operatorname{ri} \operatorname{dom} A \cap \operatorname{ri} \operatorname{dom} B \neq \varnothing$.

Then $A + B$ is maximally monotone.

Definition 3.49. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and let $n \in \mathbb{N}$ be such that $n \geq 2$. Then A is *n-cyclically monotone* if, for every $(x_1, \ldots, x_{n+1}) \in \mathcal{H}^{n+1}$ and $(u_1, \ldots, u_n) \in \mathcal{H}^n$,

$$(x_1, u_1) \in \operatorname{gra} A, \ldots, (x_n, u_n) \in \operatorname{gra} A, x_{n+1} = x_1 \Rightarrow \sum_{i=1}^{n} \langle x_{i+1} - x_i, u_i \rangle \leq 0.$$

If A is n-cyclically monotone for every integer $n \geq 2$, then A is *cyclically monotone*. If A is cyclically monotone and there exists no cyclically monotone operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, then A is *maximally cyclically monotone*.

Fact 3.50 (Rockafellar)**.** [11, Theorem 22.14] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$. Then A is maximally cyclically monotone if and only if there exists $f \in \Gamma_0(\mathcal{H})$ such that $A = \partial f$.*

Fact 3.51. [5, Theorem 6.6] *Suppose $\mathcal{H}$ is a real Hilbert space and let $T : \mathcal{H} \to \mathcal{H}$. Then T is the resolvent of the maximally cyclically monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ if and only if T has full domain, T is firmly nonexpansive, and T is cyclically firmly nonexpansive. That is, for every set of points $\{x_1, \ldots, x_n\}$ where $n \in \mathbb{N}$, $n \geq 2$ and $x_{n+1} = x_1$, one has*

$$\sum_{i=1}^{n} \langle x_i - Tx_i, Tx_i - Tx_{i+1} \rangle \geq 0. \tag{3.17}$$

3.4 Rectangular monotone operators

The notion of rectangularity for monotone operators requires the use of the Fitzpatrick function.

Definition 3.52 (Fitzpatrick function)**.** (See [40], [31] or [47].) Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$. Then the *Fitzpatrick function* associated with A is

$$\begin{aligned} F_A\colon \mathcal{H} \times \mathcal{H} \to [-\infty, +\infty] :& \\ (x, x^*) \mapsto & \sup_{(a,a^*) \in \operatorname{gra} A} \big(\langle x, a^* \rangle + \langle a, x^* \rangle - \langle a, a^* \rangle \big) && (3.18) \\ &= \langle x, x^* \rangle - \inf_{(a,a^*) \in \operatorname{gra} A} \left(\langle x - a, x^* - a^* \rangle \right). && (3.19) \end{aligned}$$

Example 3.53 (energy)**.** [17, Example 3.10] The Fitzpatrick function of the identity operator is

$$F_{\mathrm{Id}}\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}\colon (x, x^*) \mapsto \tfrac{1}{4}\|x + x^*\|^2.$$

Definition 3.54 (Brézis-Haraux)**.** (See [28].) Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone. Then A is *rectangular* (which is also known as star-monotone or 3^* monotone), if

$$\operatorname{dom} A \times \operatorname{ran} A \subseteq \operatorname{dom} F_A. \tag{3.20}$$

Remark 3.55. If $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is maximally monotone and rectangular, then one obtains the "rectangle" $\overline{\operatorname{dom} F_A} = \overline{\operatorname{dom} A} \times \overline{\operatorname{ran} A}$, which prompted Simons [61] to call such an operator rectangular. Such operators are also referred to as *star-monotone* in [53] or *(BH)-operators* in [33].

Proposition 3.56. *A monotone operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is rectangular if*

$$(\forall (x, y^*) \in \operatorname{dom} A \times \operatorname{ran} A) \qquad \sup_{(z,z^*) \in \operatorname{gra} A} \langle x - z, z^* - y^* \rangle < +\infty.$$

Proof. This follows from (3.20) and (3.19). □

Fact 3.57 (Rank-Nullity Theorem)**.** *[48, (4.4.15)] Let $A : \mathbb{R}^N \to \mathbb{R}^N$ be an $N \times N$ matrix. Then*

$$\dim \operatorname{ran} A + \dim \ker A = N.$$

Fact 3.58. *Let $A : \mathbb{R}^N \to \mathbb{R}^N$ be a linear maximally monotone operator. Then the following hold:*

(i) *A is paramonotone if and only if A is rectangular;*

(ii) *A is paramonotone if and only if $\operatorname{rank} A = \operatorname{rank} A_+$ if and only if $\operatorname{ran} A = \operatorname{ran} A_+$.*

Proof. (i) See [9, Remark 4.11] or [24, Corollary 4.11]. (ii) Since A is monotone, we have $\operatorname{ran} A_+ \subseteq \operatorname{ran} A$. Thus, the result follows from [24, Corollary 4.11] and Fact 3.57. □

Fact 3.59. [11, Proposition 24.15] *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. Then A is rectangular $\Leftrightarrow$ A^{-1} is rectangular.*

Fact 3.60. [11, Proposition 24.18] *Let A and B be monotone operators from $\mathcal{H} \rightrightarrows \mathcal{H}$ such that $(\operatorname{dom} A \cap \operatorname{dom} B) \times \mathcal{H} \subseteq \operatorname{dom} F_B$. Then $A + B$ is rectangular.*

Example 3.61. (See [28, Example 3] or [2, Example 6.5.2(iii)].) Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone. Then $A + \operatorname{Id}$ and $(A + \operatorname{Id})^{-1}$ are maximally monotone and rectangular.

Proof. Combining Fact 3.60 and Example 3.53, we see that $A + \operatorname{Id}$ is rectangular. Furthermore, $A + \operatorname{Id}$ is maximally monotone by Fact 3.48. Using Fact 3.59, we see that $(\operatorname{Id} + A)^{-1}$ is maximally monotone and rectangular. □

Proposition 3.62. [17, Proposition 4.2] *Let A and B be monotone on $\mathcal{H}$, and let $(x, x^*) \in \mathcal{H} \times \mathcal{H}$. Then $F_{A+B}(x, x^*) \leq \big(F_A(x, \cdot) \,\Box\, F_B(x, \cdot)\big)(x^*)$.*

Lemma 3.63. [20, Lemma 3.11] *Let A and B be rectangular on $\mathcal{H}$. Then $A + B$ is rectangular.*

Proof. Clearly, $\operatorname{dom}(A+B) = (\operatorname{dom} A)\cap(\operatorname{dom} B)$, and $\operatorname{ran}(A+B) \subseteq \operatorname{ran} A + \operatorname{ran} B$. Take $x \in \operatorname{dom}(A+B)$ and $y^* \in \operatorname{ran}(A+B)$. Then there exist $a^* \in \operatorname{ran} A$ and $b^* \in \operatorname{ran} B$ such that $a^* + b^* = y^*$. Furthermore, $(x, a^*) \in (\operatorname{dom} A) \times (\operatorname{ran} A) \subseteq \operatorname{dom} F_A$ and $(x, b^*) \in (\operatorname{dom} B) \times (\operatorname{ran} B) \subseteq \operatorname{dom} F_B$. Using Proposition 3.62 and the assumption that A and B are rectangular, we obtain

$$F_{A+B}(x, y^*) \leq F_A(x, a^*) + F_B(x, b^*) < +\infty. \tag{3.21}$$

Therefore, $\operatorname{dom}(A+B)\times\operatorname{ran}(A+B) \subseteq \operatorname{dom} F_{A+B}$ and $A+B$ is rectangular. $\square$

Fact 3.64. [24, Theorem 6.1] *Let $T : \mathcal{H} \to \mathcal{H}$ be nonexpansive and define the corresponding* displacement mapping *by*

$$A = \operatorname{Id} - T.$$

Then the following hold:

(i) *A is maximally monotone.*

(ii) *A is $\frac{1}{2}$-cocoercive, i.e. $\frac{1}{2}A$ is firmly nonexpansive.*

(iii) *A is rectangular.*

(iv) *A^{-1} is strongly monotone with constant $\frac{1}{2}$.*

(v) *A^{-1} is strictly monotone.*

(vi) *A is paramonotone.*

Example 3.65. [24, Example 6.2] Let N be a strictly positive integer and let

$$R : \mathcal{H}^N \to \mathcal{H}^N : (x_1, \ldots, x_N) \mapsto (x_N, x_1, \ldots, x_{N-1}),$$

be the cyclic right-shift operator in $\mathcal{H}^N$. Since $\|Rx\| = \|x\|$ for all $x \in \mathcal{H}$, R is nonexpansive and therefore by Fact 3.64, $\operatorname{Id} - R$ is maximally monotone, rectangular, and paramonotone.

Fact 3.66 (Brézis-Haraux)**.** [2, Theorem 6.5.1(b) and Theorem 6.5.2] *Let A and B be monotone on a Hilbert space $\mathcal{H}$ such that $A + B$ is maximally monotone. Suppose that one of the following holds.*

(i) *A and B are rectangular.*

(ii) $\operatorname{dom} A \subseteq \operatorname{dom} B$ *and B is rectangular.*

Then $\overline{\operatorname{ran}(A+B)} = \overline{\operatorname{ran} A + \operatorname{ran} B}$*,* $\operatorname{int}(\operatorname{ran}(A+B)) = \operatorname{int}(\operatorname{ran} A + \operatorname{ran} B)$*, and if* $\mathcal{H}$ *is finite dimensional* $\operatorname{ri}\operatorname{conv}(\operatorname{ran} A + \operatorname{ran} B) \subseteq \operatorname{ran}(A+B)$*.*

In this chapter we have seen many properties of firmly nonexpansive mappings and monotone operators. We have also seen how the two concepts are linked through the resolvent of a maximally monotone operator. This will be fundamental to the results in chapters 4–8.

Chapter 4

Correspondence of Properties

This chapter contains new results concerning the closely-knit nature of firmly nonexpansive mappings and maximally monotone operators and is based on [19].

4.1 Maximally monotone operators and firmly nonexpansive mappings

The first result in this section provides a comprehensive list of corresponding properties of firmly nonexpansive mappings and maximally monotone operators, building on Minty's Fact 3.36.

Theorem 4.1. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \mathrm{Id}$. Then the following hold:*

(i) $\operatorname{ran} T = \operatorname{dom} A$.

(ii) *T is surjective if and only if $\operatorname{dom} A = \mathcal{H}$.*

(iii) *$\mathrm{Id} - T$ is surjective if and only if A is surjective.*

(iv) *T is injective if and only if A is at most single-valued.*

(v) *T is an isometry if and only if there exists $z \in \mathcal{H}$ such that $A\colon x \mapsto z$, in which case $T\colon x \mapsto x - z$.*

(vi) *T satisfies*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad Tx \neq Ty \;\Rightarrow\; \|Tx - Ty\|^2 < \langle x - y, Tx - Ty\rangle \tag{4.1}$$

if and only if A is strictly monotone, i.e.,

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \qquad x \neq y \;\Rightarrow\; \langle x-y, u-v\rangle > 0. \tag{4.2}$$

(vii) *T is strictly monotone if and only if A is at most single-valued.*

(viii) *T is strictly firmly nonexpansive, i.e.,*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \;\Rightarrow\; \|Tx - Ty\|^2 < \langle x-y, Tx-Ty\rangle \tag{4.3}$$

if and only if A is at most single-valued and strictly monotone.

(ix) *T is strictly nonexpansive, i.e.,*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \;\Rightarrow\; \|Tx - Ty\| < \|x-y\| \tag{4.4}$$

if and only if A is disjointly injective, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \;\Rightarrow\; Ax \cap Ay = \varnothing. \tag{4.5}$$

(x) *T is injective and strictly nonexpansive, i.e.,*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \;\Rightarrow\; 0 < \|Tx - Ty\| < \|x-y\| \tag{4.6}$$

if and only if A is at most single-valued and disjointly injective.

(xi) *Suppose that $\varepsilon \in \,]0,+\infty[$. Then $(1+\varepsilon)T$ is firmly nonexpansive if and only if A is strongly monotone with constant ε, i.e., $A - \varepsilon \operatorname{Id}$ is monotone, in which case T is a Banach contraction with constant $(1+\varepsilon)^{-1}$.*

(xii) *Suppose that $\gamma \in \,]0,+\infty[$. Then $(1+\gamma)(\operatorname{Id} - T)$ is firmly nonexpansive if and only if A is γ-cocoercive , i.e.,*

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x-y, u-v\rangle \geq \gamma \|u-v\|^2. \tag{4.7}$$

(xiii) *Suppose that $\beta \in \,]0,1[$. Then T is a Banach contraction with constant β if and only if A satisfies*

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \qquad \frac{1-\beta^2}{\beta^2}\|x-y\|^2 \leq 2\langle x-y, u-v\rangle + \|u-v\|^2. \tag{4.8}$$

(xiv) *Suppose that* $\phi\colon [0,+\infty[\to [0,+\infty]$ *is increasing and vanishes only at* 0. *Then* T *satisfies*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle Tx - Ty, (x - Tx) - (y - Ty)\rangle \geq \phi\big(\|Tx - Ty\|\big) \tag{4.9}$$

if and only if A *is uniformly monotone with modulus* ϕ, *i.e.*,

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y, u - v\rangle \geq \phi\big(\|x - y\|\big). \tag{4.10}$$

(xv) T *satisfies*

$$\begin{aligned}&(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\\ &\qquad \|Tx - Ty\|^2 = \langle x - y, Tx - Ty\rangle \;\Rightarrow\; \begin{cases} Tx = T\big(Tx + y - Ty\big) \\ Ty = T\big(Ty + x - Tx\big)\end{cases}\end{aligned} \tag{4.11}$$

if and only if A *is paramonotone, i.e.*

$$\begin{aligned}&(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A)\\ &\qquad\qquad \langle x - y, u - v\rangle = 0 \;\Rightarrow\; \{(x,v),(y,u)\} \subseteq \operatorname{gra} A.\end{aligned} \tag{4.12}$$

(xvi) (Bartz et al., [5]) T *is cyclically firmly nonexpansive, i.e.*,

$$\sum_{i=1}^{n} \langle x_i - Tx_i, Tx_i - Tx_{i+1}\rangle \geq 0, \tag{4.13}$$

for every set of points $\{x_1, \ldots, x_n\} \subseteq \mathcal{H}$, *where* $n \in \{2, 3, \ldots\}$ *and* $x_{n+1} = x_1$, *if and only if* A *is a subdifferential operator, i.e., there exists* $f \in \Gamma_0(\mathcal{H})$ *such that* $A = \partial f$.

(xvii) T *satisfies*

$$(\forall x \in \mathcal{H})(y \in \mathcal{H}) \quad \inf_{z \in \mathcal{H}} \langle Tx - Tz, (y - Ty) - (z - Tz)\rangle > -\infty \tag{4.14}$$

if and only if A *is rectangular, i.e.*,

$$(\forall x \in \operatorname{dom} A)(\forall v \in \operatorname{ran} A) \quad \inf_{(z,w) \in \operatorname{gra} A} \langle x - z, v - w\rangle > -\infty. \tag{4.15}$$

(xviii) T *is linear if and only if* A *is a linear relation, i.e.*, $\operatorname{gra} A$ *is a linear subspace of* $\mathcal{H} \times \mathcal{H}$.

(xix) *T is affine if and only if A is an affine relation, i.e.,* $\operatorname{gra} A$ *is an affine subspace of $\mathcal{H} \times \mathcal{H}$.*

(xx) (Zarantonello) $\operatorname{ran} T = \operatorname{Fix} T := C$ *if and only if A is a normal cone operator, i.e., $A = \partial \iota_C$; equivalently, T is a projection (nearest point) mapping P_C.*

(xxi) *T is sequentially weakly continuous if and only if* $\operatorname{gra} A$ *is sequentially weakly closed.*

Proof. Let x, y, u, v be in $\mathcal{H}$.

(i): Clear.

(ii): This follows from (i).

(iii): Clear from the Minty parametrization (3.14).

(iv): Assume first that T is injective and that $\{u, v\} \subseteq Ax$. Then

$$\{x + u, x + v\} \subseteq (\operatorname{Id} + A)x,$$

and hence

$$x = T(x + u) = T(x + v).$$

Since T is injective, it follows that $x + u = x + v$ and hence that $u = v$. Thus, A is at most single-valued.

Conversely, let us assume that A is at most single-valued and that $Tu = Tv = x$. Then

$$\{u, v\} \subseteq (\operatorname{Id} + A)x = x + Ax,$$

and hence

$$\{u - x, v - x\} \subseteq Ax$$

Since A is at most single-valued, we have $u - x = v - x$ and so $u = v$. Thus, T is injective.

(v): Assume first that T is an isometry. Then by (2.1) and (3.3),

$$\|Tx - Ty\|^2 = \|x - y\|^2 \geq \|Tx - Ty\|^2 + \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2.$$

Thus,

$$0 \geq \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2.$$

It follows that there exists $z \in \mathcal{H}$ such that $T\colon w \mapsto w - z$. Thus, $T^{-1}\colon w \mapsto w + z$. On the other hand, $T^{-1} = \operatorname{Id} + A\colon w \mapsto w + Aw$. Hence $A\colon w \mapsto z$, as claimed. Conversely, let us assume that there exits $z \in \mathcal{H}$ such that $A\colon w \mapsto z$. Then $\mathrm{Id} + A\colon w \mapsto w + z$ and hence

$$T = J_A = (\mathrm{Id} + A)^{-1}\colon w \mapsto w - z.$$

Thus, T is an isometry.

(vi): Assume first that T satisfies (4.1), that $\{(x, u), (y, v)\} \subseteq \operatorname{gra} A$, and that $x \neq y$. Set $p = x + u$ and $q = y + v$. Then

$$(x, u) = (Tp, p - Tp),$$

and

$$(y, v) = (Tq, q - Tq).$$

Since $x \neq y$, it follows that $Tp \neq Tq$ and therefore that

$$\|Tp - Tq\|^2 < \langle p - q, Tp - Tq\rangle\,,$$

because T satisfies (4.1). Hence

$$0 < \langle (p - Tp) - (q - Tq), Tp - Tq\rangle = \langle u - v, x - y\rangle\,.$$

Thus, A is strictly monotone. Conversely, let us assume that A is strictly monotone and that $x = Tu \neq Tv = y$. Then $\{(x, u - x), (y, v - y)\} \subseteq \operatorname{gra} A$. Since $x \neq y$ and A is strictly monotone, we have

$$\begin{aligned}\langle x - y, (u - x) - (v - y)\rangle > 0 &\Leftrightarrow \|x - y\|^2 < \langle x - y, u - v\rangle\\ &\Leftrightarrow \|Tu - Tv\|^2 < \langle Tv - Tu, u - v\rangle\,.\end{aligned}$$

Thus, T satisfies (4.1).

(vii): In view of (vi) it suffices to show that T is injective if and only if T is strictly monotone. Assume first that T is injective and that $x \neq y$. Then $Tx \neq Ty$ and hence

$$0 < \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle\,.$$

Thus, T is strictly monotone. Conversely, assume that T is strictly monotone and that $x \neq y$. Then $\langle x - y, Tx - Ty\rangle > 0$ and hence $Tx \neq Ty$. Thus, T is injective.

(viii): Observe that T is strictly firmly nonexpansive if and only if T is injective and T satisfies (4.1). Thus, the result follows from combining (iv) and (vi).

(ix): Assume first that T is strictly nonexpansive, that $x \neq y$, and that $u \in Ax \cap Ay$. Then

$$x + u \in (\mathrm{Id} + A)x \text{ and } y + u \in (\mathrm{Id} + A)y;$$

equivalently,

$$T(x+u) = x \neq y = T(y+u).$$

Since T is strictly nonexpansive, we have

$$\|x - y\| = \|T(x+u) - T(y+u)\| < \|(x+u) - (y+u)\| = \|x - y\|,$$

which gives a contradiction. Thus, A is disjointly injective. Conversely, assume that A is disjointly injective, that $u \neq v$, and that $\|Tu - Tv\| = \|u - v\|$. Since T is firmly nonexpansive, we deduce that

$$u - Tu = v - Tv.$$

Assume that $x = u - Tu = v - Tv$. Then,

$$Tu = u - x \text{ and } Tv = v - x;$$

equivalently,

$$u \in (\mathrm{Id} + A)(u - x) \text{ and } v \in (\mathrm{Id} + A)(v - x).$$

Thus, $x \in A(Tu) \cap A(Tv)$, which contradicts the assumption on disjoint injectivity of A.

(x): Combine (iv) and (ix).

(xi): Assume first that $(1+\varepsilon)T$ is firmly nonexpansive and that

$$\{(x,u),(y,v)\} \subseteq \operatorname{gra} A.$$

Then $x = T(x+u)$ and $y = T(y+v)$. Hence by Fact 3.3(iv),

$$\begin{aligned} &\langle (x+u) - (y+v), x - y\rangle \geq (1+\varepsilon)\|x-y\|^2 \\ \Leftrightarrow\ &\langle x - y, u - v\rangle \geq \varepsilon\|x - y\|^2. \end{aligned}$$

Thus, $A - \varepsilon\,\mathrm{Id}$ is monotone. Conversely, assume that $A - \varepsilon\,\mathrm{Id}$ is monotone and that

$$\{(x,u),(y,v)\} \subseteq \operatorname{gra} T.$$

Then $\{(u, x-u),(v, y-v)\} \subseteq \operatorname{gra} A$ and hence

$$\begin{aligned} &\langle u - v, (x-u) - (y-v)\rangle \geq \varepsilon\|u - v\|^2 \\ \Leftrightarrow\ &\langle x - y, u - v\rangle \geq (1+\varepsilon)\|u-v\|^2. \end{aligned}$$

Thus, $(1+\varepsilon)T$ is firmly nonexpansive. Alternatively, this result follows from Fact 3.45.

(xii): Applying (xi) to $\operatorname{Id} - T$ and A^{-1}, we see that $(1+\gamma)(\operatorname{Id} - T)$ is firmly nonexpansive if and only if $A^{-1} - \gamma \operatorname{Id}$ is monotone, which is equivalent to A being γ-cocoercive.

(xiii): Assume first that T is a Banach contraction with constant β and that $\{(x,u),(y,v)\} \subseteq \operatorname{gra} A$. Set $p = x + u$ and $y = y + v$. Then

$$\begin{aligned}(x,u) &= (Tp, p - Tp),\\ (y,v) &= (Tq, q - Tq), \text{ and}\\ \|Tp - Tq\| &\leq \beta\|p - q\|,\end{aligned}$$

i.e.,

$$\begin{aligned}\|x-y\|^2 &\leq \beta^2\|(x+u)-(y+v)\|^2 = \beta^2\|(x-y)+(u-v)\|^2 \\ &= \beta^2\big(\|x-y\|^2 + 2\langle x-y, u-v\rangle + \|u-v\|^2\big).\end{aligned} \tag{4.16}$$

Thus, (4.8) holds. The converse is proved similarly.

(xiv): The equivalence is immediate from the Minty parametrization (3.14).

(xv): Assume first that T satisfies (4.11) and that $\{(x,u),(y,v)\} \subseteq \operatorname{gra} A$ with $\langle x-y, u-v\rangle = 0$. Set $p = x + u$ and $q = y + v$. Then

$$(x,u) = (Tp, p - Tp) \text{ and } (y,v) = (Tq, q - Tq),$$

and we have

$$\begin{aligned}&\langle Tp - Tq, (p - Tp) - (q - Tq)\rangle = 0\\ \Leftrightarrow\ &\|Tp - Tq\|^2 = \langle p - q, Tp - Tq\rangle .\end{aligned}$$

By (4.11),

$$\begin{aligned}Tp = T(Tp + q - Tq) &\Leftrightarrow x = T(x+v)\\ &\Leftrightarrow x + v \in x + Ax \Leftrightarrow v \in Ax.\end{aligned}$$

And similarly,

$$\begin{aligned}Tq = T(Tq + p - Tp) &\Leftrightarrow y = T(y+u)\\ &\Leftrightarrow y + u \in y + Ay \Leftrightarrow u \in Ay.\end{aligned}$$

Thus, A is paramonotone. Conversely, assume that A is paramonotone, that $\|Tu - Tv\|^2 = \langle u - v, Tu - Tv\rangle$, that $x = Tu$, and that $y = Tv$. Then,

$$\{(x, u - x), (y, v - y)\} \subseteq \operatorname{gra} A,$$

and

$$\langle x - y, (u - x) - (v - y)\rangle = 0.$$

Since A is paramonotone, we deduce that

$$\begin{aligned} v - y \in Ax &\Leftrightarrow x - y + v \in (\operatorname{Id} + A)x \\ &\Leftrightarrow x = T(x - y + v) \Leftrightarrow Tu = T(Tu + v - Tv). \end{aligned}$$

And similarly,

$$\begin{aligned} u - x \in Ay &\Leftrightarrow y - x + u \in (\operatorname{Id} + A)y \\ &\Leftrightarrow y = T(y - x + u) \Leftrightarrow Tv = T(Tv + u - Tu). \end{aligned}$$

Thus, T satisfies (4.11).

(xvi): This follows from Fact 3.51.

(xvii): The equivalence is immediate from the Minty parametrization (3.14).

(xviii): Indeed,

$$\begin{aligned} T = J_A \text{ is linear} &\Leftrightarrow (A + \operatorname{Id})^{-1} \text{ is a linear relation,} \\ &\Leftrightarrow A + \operatorname{Id} \text{ is a linear relation,} \\ &\Leftrightarrow A \text{ is a linear relation.} \end{aligned}$$

(xix): This follows from (xviii).

(xx): Assume $C := \operatorname{Fix} T = \operatorname{ran} T$. Fact 3.15 yields C is a closed convex set and by Fact 3.16, $T = P_C$. On the other hand, if $A = N_C$, then by Fact 3.42, $T = J_{N_C} = P_C$ which gives $\operatorname{Fix} T = C$ and $\operatorname{ran} T = C$. The fact that $N_C = \partial \iota_C$ follows from Fact 2.38.

(xxi): Assume that T is sequentially weakly continuous. Let $(x_n, u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$ that converges weakly to $(x, u) \in \mathcal{H} \times \mathcal{H}$. Then $(x_n + u_n)_{n\in\mathbb{N}}$ converges weakly to $x + u$. On the other hand, $\operatorname{Id} - T$ is sequentially weakly continuous because T is. Altogether,

$$\begin{aligned} (x_n, u_n)_{n\in\mathbb{N}} &= (T(x_n + u_n), (\operatorname{Id} - T)(x_n + u_n))_{n\in\mathbb{N}} \\ &\rightharpoonup (T(x + u), (\operatorname{Id} - T)(x + u)). \end{aligned}$$

But $(x_n, u_n) \rightharpoonup (x, u)$ and thus $(x, u) = (T(x+u), (\mathrm{Id} - T)(x+u)) \in \operatorname{gra} A$. Therefore $\operatorname{gra} A$ is sequentially weakly closed. Conversely, let us assume that $\operatorname{gra} A$ is sequentially weakly closed. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that is weakly convergent to x. Our goal is to show that $Tx_n \rightharpoonup Tx$. Since T is nonexpansive, the sequence $(Tx_n)_{n\in\mathbb{N}}$ is bounded. After passing to a subsequence and relabeling if necessary, we can and do assume that $(Tx_n)_{n\in\mathbb{N}}$ converges weakly to some point $y \in \mathcal{H}$. Now $(Tx_n, x_n - Tx_n)_{n\in\mathbb{N}}$ lies in $\operatorname{gra} A$, and this sequence converges weakly to $(y, x-y)$. Since $\operatorname{gra} A$ is sequentially weakly closed, it follows that $(y, x-y) \in \operatorname{gra} A$. Therefore,

$$x - y \in Ay \Leftrightarrow x \in (\mathrm{Id} + A)y \Leftrightarrow y = Tx,$$

which implies the result. □

Example 4.2. Concerning items (xi) and (xiii) in Theorem 4.1, it was previously known that if A is strongly monotone, then T is a Banach contraction, see Fact 3.45. The converse, however, is false. Consider the case $\mathcal{H} = \mathbb{R}^2$ and set

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{4.17}$$

Then $(\forall z \in \mathcal{H})\ \langle z, Az \rangle = 0$ so A cannot be strongly monotone. On the other hand,

$$T = J_A = (\mathrm{Id} + A)^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{4.18}$$

is linear and $\|Tz\|^2 = \frac{1}{2}\|z\|^2$, which implies that T is a Banach contraction with constant $1/\sqrt{2}$.

Corollary 4.3. *Let $A\colon \mathcal{H} \to \mathcal{H}$ be continuous, linear, and maximally monotone. Then the following hold.*

(i) *If J_A is a Banach contraction, then A is (disjointly) injective.*

(ii) *If $\operatorname{ran} A$ is closed and A is (disjointly) injective, then J_A is a Banach contraction.*

Proof. The result is trivial if $\mathcal{H} = \{0\}$ so we assume that $\mathcal{H} \neq \{0\}$. Let x and y be in $\mathcal{H}$.

(i): Assume that J_A is a Banach contraction, with constant $\beta \in [0, 1[$. If $\beta = 0$, then $J_A \equiv 0 \Leftrightarrow A = N_{\{0\}}$, which contradicts the single-valuedness of A. Thus, $0 < \beta < 1$. By Theorem 4.1(xiii),

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \frac{1-\beta^2}{\beta^2}\|x-y\|^2 \leq 2\langle x-y, Ax-Ay\rangle + \|Ax-Ay\|^2. \tag{4.19}$$

If $x \neq y$, then the left side of (4.19) is strictly positive, which implies that $Ax \neq Ay$. Thus, A is (disjointly) injective.

(ii): Let us assume that $\operatorname{ran} A$ is closed and that A is (disjointly) injective. Then $\ker A = \{0\}$ and hence, by Fact 2.14, there exists $\rho \in \,]0, +\infty[$ such that $(\forall z \in \mathcal{H})\ \|Az\| \geq \rho\|z\|$. Thus,

$$(\forall z \in \mathcal{H}) \quad \|Az\|^2 - \rho^2\|z\|^2 \geq 0. \tag{4.20}$$

Set $\beta = 1/\sqrt{1+\rho^2}$ and $z = x - y$. Then $\rho^2 = (1-\beta^2)/\beta^2$ and hence by (4.20) and Lemma 3.27,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$$
$$\frac{1-\beta^2}{\beta^2}\|x-y\|^2 \leq \|Ax - Ay\|^2 \leq 2\langle x-y, Ax-Ay\rangle + \|Ax-Ay\|^2. \tag{4.21}$$

Again by Theorem 4.1(xiii), J_A is a Banach contraction with constant $\beta \in \,]0,1[$. ☐

Example 4.4. Suppose that $\mathcal{H} = \ell_2(\mathbb{N})$, the space of square-summable sequences, i.e., $x = (x_n) \in \mathcal{H}$ if and only if $\sum_{n=1}^{\infty} |x_n|^2 < +\infty$, and set

$$A\colon \mathcal{H} \to \mathcal{H}\colon (x_n) \mapsto \big(\tfrac{1}{n}x_n\big). \tag{4.22}$$

Then A is continuous, linear, maximally monotone, and $\operatorname{ran} A$ is a dense, proper subspace of $\mathcal{H}$ that is not closed. The resolvent $T = J_A$ is

$$T\colon \mathcal{H} \to \mathcal{H}\colon (x_n) \mapsto \big(\tfrac{n}{n+1}x_n\big). \tag{4.23}$$

Now denote the n^{th} unit vector in $\mathcal{H}$ by $\mathbf{e}_n$ (i.e. $\mathbf{e}_n$ has a one at position n and zeros otherwise). Then $\|T\mathbf{e}_n - T0\| = \frac{n}{n+1}\|\mathbf{e}_n - 0\|$. Since $\frac{n}{n+1} \to 1$, it follows that T is not a Banach contraction.

Remark 4.5. When A is a subdifferential operator, then it is impossible to get the behavior witnessed in Example 4.2, as we see next in Proposition 4.6.

Proposition 4.6. *Let $f \in \Gamma_0(\mathcal{H})$ and let $\varepsilon \in \,]0,+\infty[$. Then $(1+\varepsilon)\operatorname{prox}_f$ is firmly nonexpansive if and only if prox_f is a Banach contraction with constant $(1+\varepsilon)^{-1}$.*

Proof. Set $\beta = (1+\varepsilon)^{-1}$. It is clear that if $(1+\varepsilon)\operatorname{prox}_f$ is firmly nonexpansive, then $(1+\varepsilon)\operatorname{prox}_f$ is nonexpansive and hence, for x and y in $\mathcal{H}$

$$\begin{aligned}&\|(1+\varepsilon)\operatorname{prox}_f x - (1+\varepsilon)\operatorname{prox}_f y\| \le \|x-y\|\\ \Leftrightarrow\ &\|\operatorname{prox}_f x - \operatorname{prox}_f y\| \le (1+\varepsilon)^{-1}\|x-y\|,\end{aligned}$$

thus prox_f is a Banach contraction with constant β. Conversely, assume that prox_f is a Banach contraction with constant β. Since prox_f is the Fréchet gradient mapping of the continuous convex function $f^*\Box\frac{1}{2}\|\cdot\|^2\colon \mathcal{H}\to\mathbb{R}$ (see Fact 2.54), the Baillon-Haddad theorem (Fact 3.12) guarantees that $\beta^{-1}\operatorname{prox}_f$ is firmly nonexpansive. $\square$

Remark 4.7. If $n=2$, then (4.13) reduces to

$$\begin{aligned}&\langle x_1 - Tx_1, Tx_1 - Tx_2\rangle + \langle x_2 - Tx_2, Tx_2 - Tx_1\rangle \ge 0\\ \Leftrightarrow\ &\langle(\operatorname{Id}-T)x_1 - (\operatorname{Id}-T)x_2, Tx_1 - Tx_2\rangle \ge 0\end{aligned}$$

i.e., to firm nonexpansiveness of T (see Fact 3.3(v)).

4.2 Duality

There is a natural duality for firmly nonexpansive mappings and maximally monotone operators; namely,

$$T \mapsto \operatorname{Id}-T \text{ and } A \mapsto A^{-1},$$

respectively. Note that the dual of the dual is the original property, e.g. $\operatorname{Id}-(\operatorname{Id}-T) = T$. Every property considered in Theorem 4.1 has a dual property. We have considered all dual properties and we shall explicitly single those out that we found to have simple and pleasing descriptions. Among these properties, those that are "self-dual", that is the property is identical to its dual property, stand out even more. First, we more explicitly define the notion of dual properties.

Definition 4.8 (dual and self-dual properties)**.** Let (p) and (p^*) be properties for firmly nonexpansive mappings defined on $\mathcal{H}$. If, for every firmly nonexpansive mapping $T\colon \mathcal{H}\to\mathcal{H}$,

$$T \text{ satisfies } (p) \text{ if and only if } \operatorname{Id}-T \text{ satisfies } (p^*), \tag{4.24}$$

then (p^*) is dual to (p), and hence (p) is dual to (p^*). If $(p) = (p^*)$, we say that (p) is self-dual. Analogously, let (q) and (q^*) be properties of maximally monotone operators defined on $\mathcal{H}$. If

$$A \text{ satisfies } (q) \text{ if and only if } A^{-1} \text{ satisfies } (q^*) \tag{4.25}$$

for every maximally monotone operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, then (q^*) is dual to (q), and hence (q) is dual to (q^*). If $(q) = (q^*)$, we say that (q) is self-dual.

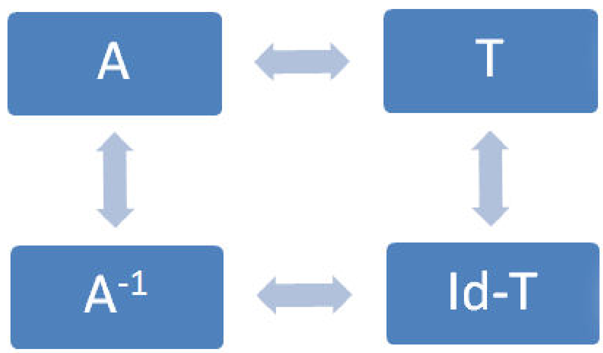

Figure 4.1: Duality of a monotone operator, A, and its associated resolvent, $T = J_A$.

Theorem 4.9. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \mathrm{Id}$. Then the following are equivalent:*

(i) *T is surjective.*

(ii) *A has full domain.*

(iii) *A^{-1} is surjective.*

Thus for maximally monotone operators, surjectivity and full domain are properties that are dual to each other. These properties are not self-dual; for example, $A = 0$ has full domain while $A^{-1} = \partial\iota_{\{0\}}$ does not.

Proof. (i)⇔(ii): Theorem 4.1(ii). (ii)⇔(iii): Obvious. □

Theorem 4.10. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \mathrm{Id}$. Then the following are equivalent:*

(i) *T is strictly nonexpansive.*

(ii) *A is disjointly injective.*

(iii) *$\mathrm{Id} - T$ is injective.*

(iv) *A^{-1} is at most single-valued.*

Thus for firmly nonexpansive mappings, strict nonexpansiveness and injectivity are dual to each other; and correspondingly for maximally monotone operators disjoint injectivity and at most single-valuedness are dual to each other. These properties are not self-dual, $T \equiv 0$ is strictly nonexpansive, but $\mathrm{Id} - T = \mathrm{Id}$ is not. Correspondingly, $A = \partial\iota_{\{0\}}$ is disjointly injective but $A^{-1} = 0$ is not.

Proof. We know that (i)⇔(ii) by Theorem 4.1(ix). We also know that (iii)⇔(iv) by Theorem 4.1(iv) (applied to A^{-1} and $\mathrm{Id} - T$). It thus suffices to show that (ii)⇔(iv). Assume first that A is disjointly injective and that $\{x, y\} \subseteq A^{-1}u$. Then $u \in Ax \cap Ay$. Since A is disjointly injective, we have $x = y$. Thus, A^{-1} is at most single-valued. Conversely, assume that A^{-1} is at most single-valued and that $u \in Ax \cap Ay$. Then $\{x, y\} \subseteq A^{-1}u$ and so $x = y$. It follows that A is disjointly injective. □

Theorem 4.11. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \mathrm{Id}$. Then the following are equivalent:*

(i) *T satisfies* (4.1) *i.e.,*

$$Tx \neq Ty \;\Rightarrow\; \|Tx - Ty\|^2 < \langle x - y, Tx - Ty\rangle .$$

(ii) *A is strictly monotone.*

(iii) *$\mathrm{Id} - T$ satisfies*

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad & \big(\mathrm{Id} - (\mathrm{Id} - T)\big)x \neq \big(\mathrm{Id} - (\mathrm{Id} - T)\big)y \\ & \Rightarrow\; \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 < \langle x - y, (\mathrm{Id} - T)x - (\mathrm{Id} - T)y\rangle .\end{aligned} \tag{4.26}$$

(iv) A^{-1} *satisfies*

$$(\forall (x,u) \in \operatorname{gra} A^{-1})(\forall (y,v) \in \operatorname{gra} A^{-1}) \quad u \neq v \;\Rightarrow\; \langle x-y, u-v\rangle > 0. \tag{4.27}$$

Thus for firmly nonexpansive mappings, properties (4.1) *and* (4.26) *are dual to each other; and correspondingly for maximally monotone operators strict monotonicity and* (4.27) *are dual to each other. These properties are not self-dual;* $T = 0$ *trivially satisfies* (4.1)*, but* $\operatorname{Id} - 0 = \operatorname{Id}$ *does not.*

Proof. (i)⇔(ii): Theorem 4.1(vi). (i)⇔(iii): Indeed, (4.26) and (4.1) are equivalent as is easily seen by expansion and rearranging. (ii)⇔(iv): Clear. ☐

Theorem 4.12 (self-duality of strict firm nonexpansiveness)**.** *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \operatorname{Id}$*. Then the following are equivalent:*

(i) T *is strictly firmly nonexpansive.*

(ii) A *is at most single-valued and strictly monotone.*

(iii) $\operatorname{Id} - T$ *is strictly firmly nonexpansive.*

(iv) A^{-1} *is at most single-valued and strictly monotone.*

Consequently, strict firm nonexpansive is a self-dual property for firmly nonexpansive mappings; correspondingly, being both strictly monotone and at most single-valued is self-dual for maximally monotone operators.

Proof. Note that T is strictly firmly nonexpansive if and only if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad x \neq y \;\Rightarrow\; 0 < \langle Tx - Ty, (\operatorname{Id} - T)x - (\operatorname{Id} - T)y\rangle\,, \tag{4.28}$$

which is obviously self-dual. In view of Theorem 4.1(viii), the corresponding property for A is being both at most single-valued and strictly monotone. ☐

Theorem 4.12 illustrates the technique of obtaining self-dual properties by fusing any property and its dual. Here is another example of this type.

Theorem 4.13 (self-duality of strict nonexpansiveness and injectivity)**.** *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \operatorname{Id}$*. Then the following are equivalent:*

(i) *T is strictly nonexpansive and injective.*

(ii) *A is at most single-valued and disjointly injective.*

(iii) $\operatorname{Id} - T$ *is strictly nonexpansive and injective.*

(iv) *A^{-1} is at most single-valued and disjointly injective.*

Consequently, being both strictly nonexpansive and injective is a self-dual property for firmly nonexpansive mappings; correspondingly, being both disjointly injective and at most single-valued is self-dual for maximally monotone operators.

Proof. Clear from Theorem 4.1(x). □

Remark 4.14. In Theorem 4.12 and Theorem 4.13, arguing directly (or by using the characterization with monotone operators via Theorem 4.1), it is easy to verify the implication

$$\begin{aligned} &T \text{ is strictly firmly nonexpansive} \\ &\qquad\qquad \Rightarrow T \text{ is injective and strictly nonexpansive.} \end{aligned} \tag{4.29}$$

The converse of implication (4.29) is false in general, see Example 4.15. In contrast, we see in Corollary 4.17 that when $\mathcal{H}$ is finite-dimensional and $T = J_A$ is a proximal mapping (i.e., A is a subdifferential operator), then the converse implication of (4.29) is true.

Example 4.15. Consider $\mathcal{H} = \mathbb{R}^2$, and let A denote the counter-clockwise rotation by $\pi/2$, which we utilized already in (4.17). Clearly, A is a linear single-valued maximally monotone operator that is (disjointly) injective, but A is not strictly monotone. Accordingly, $T = J_A$ is linear, injective and strictly nonexpansive, but not strictly firmly nonexpansive.

Lemma 4.16. *Suppose that $\mathcal{H}$ is finite-dimensional and let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) *∂f is disjointly injective.*

(ii) *$(\partial f)^{-1} = \partial f^*$ is at most single-valued.*

(iii) *f^* is essentially smooth.*

(iv) *f is essentially strictly convex.*

(v) *∂f is strictly monotone.*

(vi) prox_f *is strictly nonexpansive.*

(vii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \operatorname{prox}_f x \neq \operatorname{prox}_f y$

$$\Rightarrow \|\operatorname{prox}_f x - \operatorname{prox}_f y\|^2 < \big\langle x - y, \operatorname{prox}_f x - \operatorname{prox}_f y \big\rangle.$$

Proof. "(i)⇔(ii)": Theorem 4.10. "(ii)⇔(iii)": Fact 2.47. "(iii)⇔(iv)": Fact 2.48. "(iv)⇔(v)": Fact 3.30. "(i)⇔(vi)": Theorem 4.1(ix) and Fact 2.53. "(v)⇔(vii)": Theorem 4.1(vi). □

Lemma 4.16 admits a dual counterpart that contains various characterizations of essential smoothness. The following consequence of these characterizations is also related to Remark 4.14. Recall that for a finite-dimensional $\mathcal{H}$, a function $f \in \Gamma_0(\mathcal{H})$ is Legendre if it is both essentially smooth and essentially strictly convex.

Corollary 4.17 (Legendre self-duality)**.** *Suppose that $\mathcal{H}$ is finite-dimensional and let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) ∂f *is disjointly injective and at most single-valued.*

(ii) ∂f *is strictly monotone and at most single-valued.*

(iii) f *is Legendre.*

(iv) prox_f *is strictly firmly nonexpansive.*

(v) prox_f *is strictly nonexpansive and injective.*

(vi) ∂f^* *is disjointly injective and at most single-valued.*

(vii) ∂f^* *is strictly monotone and at most single-valued.*

(viii) f^* *is Legendre.*

(ix) $\operatorname{prox}_{f^*}$ *is strictly firmly nonexpansive.*

(x) $\operatorname{prox}_{f^*}$ *is strictly nonexpansive and injective.*

Proof. Combine Theorem 4.13 and Lemma 4.16 using Fact 2.53. □

Theorem 4.18 (self-duality of paramonotonicity)**.** *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \mathrm{Id}$. Then A is paramonotone if and only if A^{-1} is paramonotone; consequently, T satisfies* (4.11) *if and* *only if* $\operatorname{Id} - T$ *satisfies* (4.11) *(with* T *replaced by* $\operatorname{Id} - T$*). Consequently, being paramonotone is a self-dual property for maximally monotone operators; correspondingly, satisfying* (4.11) *is a self-dual property for firmly nonexpansive mappings.*

Proof. Self-duality is immediate from the definition of paramonotonicity, and the corresponding result for firmly nonexpansive mappings follows from Theorem 4.1(xv). ☐

Theorem 4.19 (self-duality of cyclical firm nonexpansiveness and cyclical monotonicity)**.**
Let $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, let* $f \in \Gamma_0$*, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \operatorname{Id}$*. Then the following are equivalent:*

(i) T *is cyclically firmly nonexpansive.*

(ii) A *is cyclically monotone.*

(iii) $A = \partial f$*.*

(iv) $\operatorname{Id} - T$ *is cyclically firmly nonexpansive.*

(v) A^{-1} *is cyclically monotone.*

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

Consequently, cyclic firm nonexpansiveness is a self-dual property for firmly nonexpansive mappings; correspondingly, cyclic monotonicity is a self-dual property for maximally monotone operators.

Proof. The fact that cyclically maximal monotone operators are subdifferential operators is due to Rockafellar and well known, see Fact 3.50, as is the identity $(\partial f)^{-1} = \partial f^*$, see Fact 2.59. The result thus follows from Theorem 4.1(xvi). ☐

Theorem 4.20 (self-duality of rectangularity)**.** *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \operatorname{Id}$*. Then the following are equivalent:*

(i) T *satisfies* (4.14)*.*

(ii) A *is rectangular.*

(iii) $\operatorname{Id} - T$ *satisfies* (4.14)*.*

(iv) A^{-1} *is rectangular.*

Consequently, rectangularity is a self-dual property for maximally monotone operators; correspondingly, (4.14) *is a self-dual property for firmly nonexpansive mappings.*

Proof. It is obvious from the definition that either property is self-dual; the equivalences thus follow from Theorem 4.1(xvii). $\square$

Theorem 4.21 (self-duality of linearity)**.** *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \mathrm{Id}$*. Then the following are equivalent:*

(i) T *is linear.*

(ii) A *is a linear relation.*

(iii) $\mathrm{Id} - T$ *is linear.*

(iv) A^{-1} *is a linear relation.*

Consequently, linearity is a self-dual property for firmly nonexpansive mappings; correspondingly, being a linear relation is a self-dual property for maximally monotone operators.

Proof. It is clear that T is linear if and only if $\mathrm{Id} - T$ is; thus, the result follows from Theorem 4.1(xviii). $\square$

Theorem 4.22 (self-duality of affineness)**.** *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be firmly nonexpansive, let* $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ *be maximally monotone, and suppose that* $T = J_A$ *or equivalently that* $A = T^{-1} - \mathrm{Id}$*. Then the following are equivalent:*

(i) T *is affine.*

(ii) A *is an affine relation.*

(iii) $\mathrm{Id} - T$ *is affine.*

(iv) A^{-1} *is an affine relation.*

Consequently, affineness is a self-dual property for firmly nonexpansive mappings; correspondingly, being an affine relation is is a self-dual property for maximally monotone operators.

Proof. It is clear that T is affine if and only if $\mathrm{Id} - T$ is; therefore, the result follows from Theorem 4.1(xix). $\square$

Remark 4.23 (projection). Concerning Theorem 4.1(xx), note that being a projection is not a self-dual: indeed, suppose that $\mathcal{H} \neq \{0\}$ and let T be the projection onto the closed unit ball. Then $\operatorname{Id} - T$ is not a projection since $\operatorname{Fix}(\operatorname{Id} - T) = \{0\} \subsetneqq \mathcal{H} = \operatorname{ran}(\operatorname{Id} - T)$.

Theorem 4.24 (self-duality of sequential weak continuity)**.** *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and suppose that $T = J_A$ or equivalently that $A = T^{-1} - \operatorname{Id}$. Then the following are equivalent:*

(i) *T is sequentially weakly continuous.*

(ii) *$\operatorname{gra} A$ is sequentially weakly closed.*

(iii) *$\operatorname{Id} - T$ is sequentially weakly continuous.*

(iv) *$\operatorname{gra} A^{-1}$ is sequentially weakly closed.*

Consequently, sequential weak continuity is a self-dual property for firmly nonexpansive mappings; correspondingly, having a sequentially weakly closed graph is a self-dual property for maximally monotone operators.

Proof. Since Id is weakly continuous, it is clear that T is sequentially weakly continuous if and only if $\operatorname{Id} - T$ is; thus, the result follows from Theorem 4.1(xxi). $\square$

The self dual properties of this section are summarized in Table 4.1.

4.3 Reflected resolvents

In the previous two sections, the correspondence between firmly nonexpansive mappings and maximally monotone operators was extensively utilized. However, Fact 3.3 provides another correspondence with nonexpansive mappings:

$$T \text{ is firmly nonexpansive if and only if } N = 2T - \operatorname{Id} \text{ is nonexpansive.} \tag{4.30}$$

Note that N is also referred to as a *reflected resolvent*. The corresponding dual of N within the set of nonexpansive mappings is simply

$$-N. \tag{4.31}$$

Table 4.1: Summary of self dual properties between monotone operators and their resolvents.

Monotone Operator A and A^{-1}	Resolvent T and $(\mathrm{Id} - T)$
At most single-valued and strictly monotone	Strictly firmly nonexpansive
At most single-valued and disjointly injective	Strictly nonexpansive and injective
Paramonotone	Satisfies (4.11)
Cyclically monotone	Cyclically firmly nonexpansive
Linear relation	Linear
Affine relation	Affine
Sequentially weakly closed	Sequentially weakly continuous

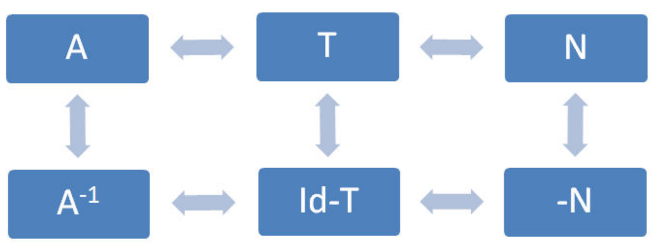

Figure 4.2: Duality of a monotone operator A, its associated resolvent, T, and its reflected resolvent, $N = 2T - \mathrm{Id}$.

Thus, all results have counterparts formulated for nonexpansive mappings. These counterparts are most easily derived from the firmly nonexpansive formulation, by simply replacing T by $\frac{1}{2}\mathrm{Id} + \frac{1}{2}N$.

Theorem 4.25 (strict firm nonexpansiveness). *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly* *nonexpansive, let $N\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive, and suppose that $N = 2T - \mathrm{Id}$. Then T is strictly firmly nonexpansive if and only if N is strictly nonexpansive.*

Proof. Let x and y be in $\mathcal{H}$. T is strictly firmly nonexpansive if $x \neq y$ implies

$$\begin{aligned} &\|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 < \|x - y\|^2 \\ \Leftrightarrow\ &\frac{1}{4}\|(\mathrm{Id} + N)x - (\mathrm{Id} + N)y\|^2 + \frac{1}{4}\|(\mathrm{Id} - N)x - (\mathrm{Id} - N)y\|^2 < \|x - y\|^2. \end{aligned}$$

Now expand and simplify to yield the result. □

Remark 4.26.

(i) We know from Theorem 4.12 that strict firm nonexpansiveness is a self-dual property with respect to monotone operators and firmly nonexpansive mappings. This can also be seen within the realm of nonexpansive mappings since N is strictly nonexpansive if and only if $-N$ is.

(ii) Furthermore, combining Theorem 4.12 with Theorem 4.25 yields the following: a maximally monotone operator A is at most single-valued and strictly monotone if and only if its reflected resolvent $2J_A - \mathrm{Id}$ is strictly nonexpansive. This characterization was observed by Rockafellar and Wets; see [59, Proposition 12.11].

(iii) In passing, we note that when $\mathcal{H}$ is finite-dimensional, the iterates of a strictly nonexpansive mapping converge to the unique fixed point (assuming it exists). For this and more, see, e.g., [39].

Theorem 4.27 (strong monotonicity). *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $N\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive, suppose that $N = 2J_A - \mathrm{Id}$ and that $\varepsilon \in\ ]0, +\infty[$. Then A is strongly monotone with constant ε if and only if $\varepsilon \,\mathrm{Id} + (1 + \varepsilon)N$ is nonexpansive.*

Proof. We know from Theorem 4.1(xi) that A is strongly monotone with constant ε if and only if $(1 + \varepsilon)T$ is firmly nonexpansive. This is equivalent to

$$2(1 + \varepsilon)T - \mathrm{Id} = (1 + \varepsilon)(2T - \mathrm{Id}) + \varepsilon \,\mathrm{Id},$$

is nonexpansive. □

Theorem 4.28 (reflected resolvent as Banach contraction)**.** *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, and let $N\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive. Suppose that $T = J_A$, that $N = 2T - \mathrm{Id}$, and that $\beta \in [0, 1]$. Then the following are equivalent:*

(i) $(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A)$

$$(1 - \beta^2)(\|x - y\|^2 + \|u - v\|^2) \leq 2(1 + \beta^2) \langle x - y, u - v\rangle$$

(ii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$

$$(1 - \beta^2)\|x - y\|^2 \leq 4 \langle Tx - Ty, (\mathrm{Id} - T)x - (\mathrm{Id} - T)y\rangle$$

(iii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$

$$\|Nx - Ny\| \leq \beta\|x - y\|.$$

Proof. In view of the Minty parametrization, (3.14), item (i) is equivalent to

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad & (1 - \beta^2)\big(\|Tx - Ty\|^2 + \|(x - Tx) - (y - Ty)\|^2\big) \\ & \leq 2\big(1 + \beta^2\big) \langle Tx - Ty, (x - Tx) - (y - Ty)\rangle . \end{aligned} \tag{4.32}$$

Simple algebraic manipulations show that (4.32) is equivalent to (ii), which in turn is equivalent to (iii). ☐

It is clear that the properties (i)–(iii) in Theorem 4.28 are self-dual (for fixed β). The following result is a simple consequence.

Corollary 4.29 (self-duality of reflected resolvents that are Banach contractions)**.** *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $N\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive, and suppose that $T = J_A$ and $N = 2T - \mathrm{Id}$. Then the following are equivalent:*

(i) $\inf \left\{ \dfrac{\langle x - y, u - v\rangle}{\|x - y\|^2 + \|u - v\|^2} \;\middle|\; \{(x, u), (y, v)\} \subseteq \operatorname{gra} A, (x, u) \neq (y, v) \right\} > 0.$

(ii) $\inf \left\{ \dfrac{\langle Tx - Ty, (\mathrm{Id} - T)x - (\mathrm{Id} - T)y\rangle}{\|x - y\|^2} \;\middle|\; \{x, y\} \subseteq \mathcal{H}, x \neq y \right\} > 0.$

(iii) *N is a Banach contraction.*

Furthermore, these properties are self-dual for their respective classes of operators.

Remark 4.30. Precisely when $A\colon x \mapsto x - z$ for some fixed vector $z \in \mathcal{H}$, we compute $T\colon x \mapsto (x+z)/2$ and therefore we reach the extreme case of Corollary 4.29 where $N\colon x \mapsto z$ is a Banach contraction with constant 0.

Corollary 4.31. *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $N\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive, and suppose that $T = J_A$ and $N = 2T - \mathrm{Id}$. Then the following are equivalent:*

(i) *Both A and A^{-1} are strongly monotone.*

(ii) *There exists $\gamma \in \left]1, +\infty\right[$ such that both γT and $\gamma(\mathrm{Id} - T)$ are firmly nonexpansive.*

(iii) *N is a Banach contraction.*

Proof. Let us assume that A and A^{-1} are both strongly monotone; equivalently, there exists $\varepsilon \in \left]0, +\infty\right[$ such that $A - \varepsilon \mathrm{Id}$ and $A^{-1} - \varepsilon \mathrm{Id}$ are monotone. Let $\{(x,u),(y,v)\} \subseteq \operatorname{gra} A$. Then $\{(u,x),(v,y)\} \subseteq \operatorname{gra} A^{-1}$ and

$$\langle x - y, u - v\rangle \geq \varepsilon\|x - y\|^2 \text{ and } \langle u - v, x - y\rangle \geq \varepsilon\|u - v\|^2. \tag{4.33}$$

Adding these inequalities yields $2\langle x - y, u - v\rangle \geq \varepsilon(\|x-y\|^2 + \|u-v\|^2)$. Thus, item (i) of Corollary 4.29 holds. Conversely, if item (i) of Corollary 4.29 holds, then both A and A^{-1} are strongly monotone. Therefore, by Corollary 4.29, (i) and (iii) are equivalent. Finally, in view of Theorem 4.1(xi), we see that (i) and (ii) are also equivalent. ☐

Additional characterizations are available for subdifferential operators:

Proposition 4.32. *Let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) *f and f^* are strongly convex.*

(ii) *f and f^* are everywhere differentiable, and both ∇f and ∇f^* are Lipschitz continuous.*

(iii) prox_f *and* $\mathrm{Id} - \mathrm{prox}_f$ *are Banach contractions.*

(iv) $2\,\mathrm{prox}_f - \mathrm{Id}$ *is a Banach contraction.*

Proof. It is well known that for functions, strong convexity is equivalent to strong monotonicity of the subdifferential operators, see Example 3.33. In view of Proposition 4.6 and Corollary 4.31, we obtain the equivalence of items (i), (iii), and (iv). Finally, the equivalence of (i) and (ii) follows from Fact 2.43. ☐

We now turn to linear relations.

Proposition 4.33. *Let $A\colon \mathcal{H} \to \mathcal{H}$ be a maximally monotone linear relation. Then the following are equivalent:*

(i) *Both A and A^{-1} are strongly monotone.*

(ii) *A is a continuous surjective linear operator on $\mathcal{H}$ and*

$$\inf_{z\in\mathcal{H}\smallsetminus\{0\}} \frac{\langle z, Az\rangle}{\|z\|^2 + \|Az\|^2} > 0.$$

(iii) $2J_A - \mathrm{Id}$ *is a Banach contraction.*

If $\mathcal{H}$ is finite-dimensional, then (i)–(iii) *are also equivalent to*

(iv) *$A\colon \mathcal{H} \to \mathcal{H}$ satisfies* $(\forall z \in \mathcal{H} \smallsetminus \{0\})\ \langle z, Az\rangle > 0$.

Proof. "(i)⇔(iii)": Clear from Corollary 4.31.

"(i)⇒(ii)": By Fact 3.35 A and A^{-1} are single-valued surjective operators with full domain. Since A and A^{-1} are linear, Fact 3.38 implies that A and A^{-1} are continuous. Thus, (ii) holds.

"(i)⇐(ii)": (ii) implies that item (i) of Corollary 4.29 holds. Thus, (i) follows from Corollary 4.29 and Corollary 4.31.

"(ii)⇒(iv)": Clear.

"(ii)⇐(iv)": Since A is injective and $\mathcal{H}$ is finite-dimensional, A is bijective and continuous. To see that the infimum in item (ii) is strictly positive, note that we may take the infimum over the unit sphere, which is a compact subset of $\mathcal{H}$. □

Example 4.34. In Proposition 4.33(iv), if $\mathcal{H}$ is infinite dimensional then the equivalence does not hold. Consider again the case in Example 4.4 where $\mathcal{H} = \ell^2(\mathbb{N})$ and $A : \mathcal{H} \to \mathcal{H} : (x_n) \mapsto (\frac{1}{n}x_n)$. Then for $x \in \ell^2(\mathbb{N}) \setminus \{0\}$,

$$\langle x, Ax\rangle = \sum_{n=1}^{\infty} \frac{1}{n} x_n^2 > 0,$$

so (iv) holds. But take the unit vectors $\mathbf{e}_n$ and $\mathbf{e}_{n+1}$ and we see that

$$\langle \mathbf{e}_n - \mathbf{e}_{n+1}, A\mathbf{e}_n - A\mathbf{e}_{n+1}\rangle = \frac{1}{n} + \frac{1}{n+1} = \frac{2n+1}{n^2+n} \to 0.$$

So $\nexists \beta \in \mathbb{R}_{++}$ such that $\langle x - y, Ax - Ay\rangle \geq \beta\|x - y\|^2\ \forall x, y \in \ell^2(\mathbb{N})$, thus A is not strongly monotone and thus (i) does not hold. Similarly,

$$\inf \frac{\langle \mathbf{e}_n, A\mathbf{e}_n\rangle}{\|\mathbf{e}_n\|^2 + \|A\mathbf{e}_n\|^2} = \frac{\frac{1}{n}}{1 + \frac{1}{n^2}} \to 0,$$

so (ii) does not hold. Finally, we have $T\mathbf{e}_n = J_A\mathbf{e}_n = (\frac{n}{n+1}\mathbf{e}_n)$ and thus

$$\|(2T - \mathrm{Id})\mathbf{e}_n\| = \left\|\frac{2n}{n+1}\mathbf{e}_n - \mathbf{e}_n\right\| = \left\|\frac{n-1}{n+1}\mathbf{e}_n\right\| \to 1,$$

and so $2T - \mathrm{Id}$ is not a Banach contraction and (iii) does not hold.

We shall conclude this chapter with some comments regarding applications of the above results to splitting methods. See also [11] for further information and various variants. Here is a technical lemma, which is well known and whose simple proof is omitted.

Lemma 4.35. *Let $T_1, \ldots, T_n$ be finitely many nonexpansive mappings from $\mathcal{H}$ to $\mathcal{H}$, and let $\lambda_1, \ldots, \lambda_n$ be in $]0, 1]$ such that $\lambda_1 + \cdots + \lambda_n = 1$. Then the following hold:*

(i) *The composition $T_1T_2\cdots T_n$ is nonexpansive.*

(ii) *The convex combination $\lambda_1T_1 + \cdots + \lambda_nT_n$ is nonexpansive.*

(iii) *If some T_i is strictly nonexpansive, then $T_1T_2\cdots T_n$ is strictly nonexpansive.*

(iv) *If some T_i is strictly nonexpansive, then $\lambda_1T_1 + \cdots + \lambda_nT_n$ is strictly nonexpansive.*

(v) *If some T_i is a Banach contraction, then $T_1T_2\cdots T_n$ is a Banach contraction.*

(vi) *If some T_i is a Banach contraction, then $\lambda_1T_1 + \cdots + \lambda_nT_n$ is a Banach contraction.*

Corollary 4.36 (backward-backward iteration)**.** *Let A_1 and A_2 be two maximally monotone operators from $\mathcal{H}$ to $\mathcal{H}$, and assume that one of these is disjointly injective. Then the (backward-backward) composition T_1T_2 is strictly nonexpansive.*

Proof. Combine Theorem 4.1(ix) and Lemma 4.35. □

Corollary 4.37 (Douglas-Rachford iteration)**.** *Let A_1 and A_2 be two maximally monotone operators from $\mathcal{H}$ to $\mathcal{H}$, and assume that one of these is both at most single-valued and strictly monotone (as is, e.g., the subdifferential operator of a convex Legendre function when $\mathcal{H}$ is finite-dimensional; see Corollary 4.17). Denote the resolvents of A_1 and A_2 by T_1 and T_2, respectively. Then the operator governing the Douglas-Rachford iteration, i.e.,*

$$T := \tfrac{1}{2}(2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id}) + \tfrac{1}{2}\,\mathrm{Id}, \tag{4.34}$$

is not just firmly nonexpansive but also strictly nonexpansive; consequently, $\operatorname{Fix} T$ *is either empty or a singleton.*

Proof. In view of Theorem 4.12 and Theorem 4.25, we see that $2T_1 - \mathrm{Id}$ and $2T_2 - \mathrm{Id}$ are both nonexpansive, and one of these two is strictly nonexpansive. By Lemma 4.35(iii), $(2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id})$ is strictly nonexpansive. Hence, by Lemma 4.35(iv), T is strictly nonexpansive. $\square$

Remark 4.38. Consider Corollary 4.37, and assume that A_i, where $i \in \{1, 2\}$, satisfies condition (i) in Corollary 4.29. Then $2T_i - \mathrm{Id}$ is a Banach contraction by Corollary 4.29. Furthermore, Lemma 4.35 now shows that the Douglas-Rachford operator T defined in (4.34) is a Banach contraction. Thus, $\operatorname{Fix} T$ is a singleton and the unique fixed point may be found as the strong limit of any sequence of Banach-Picard iterates for T.

This chapter gave a comprehensive list of how properties of firmly nonexpansive mappings translate to the corresponding maximally monotone operators. The duality of these properties was also examined, and those properties that are self-dual were identified. Finally, some applications to operators occurring in splitting methods, including reflected resolvents, were given.

Chapter 5

The Resolvent Average of Monotone Operators

This chapter is based on the papers [18] and [21]. We begin this chapter with a new method of averaging monotone operators.

Definition 5.1 (Resolvent average)**.** Let $A_i, i = 1, \dots, n$ be monotone operators, $\lambda_i > 0$ with $\sum_{i=1}^n \lambda_i = 1$, and $\mu > 0$. For $\boldsymbol{A} = (A_1, \dots, A_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ the *resolvent average* of $\boldsymbol{A}$ is,

$$\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) := \big[\lambda_1(A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\operatorname{Id})^{-1}\big]^{-1} - \mu^{-1}\operatorname{Id}. \tag{5.1}$$

The name "resolvent average" is motivated from the fact that when $\mu = 1$

$$\big(\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) + \operatorname{Id}\big)^{-1} = \lambda_1\big(A_1 + \operatorname{Id}\big)^{-1} + \cdots + \lambda_n\big(A_n + \operatorname{Id}\big)^{-1}, \tag{5.2}$$

which says that the resolvent of $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is the arithmetic average of resolvents of the A_i, with weight $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$. The resolvent average provides a novel averaging technique, and having the parameter μ in $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ will allow us to take limits which compare the resolvent average with the arithmetic and harmonic averages.

5.1 Basic properties

In this section, we give some basic properties of $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$.

Proposition 5.2. *We have*

$$J_{\mu\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n}, \tag{5.3}$$

$$^{\mu}\big(\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})\big) = \lambda_1\, {}^{\mu}A_1 + \cdots + \lambda_n\, {}^{\mu}A_n. \tag{5.4}$$

Proof. It follows from (5.1) that

$$\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) + \mathrm{Id} = \left[\lambda_1(\mu A_1 + \mathrm{Id})^{-1} + \cdots + \lambda_n(\mu A_n + \mathrm{Id})^{-1}\right]^{-1}.$$

Then (5.3) follows by taking inverses on both sides and using the definition of the resolvent, (3.11).

By (5.3), we obtain that

$$(\mathrm{Id} - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}) = \lambda_1(\mathrm{Id} - J_{\mu A_1}) + \cdots + \lambda_n(\mathrm{Id} - J_{\mu A_n}).$$

Dividing both sides by μ,

$$\mu^{-1}(\mathrm{Id} - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}) = \lambda_1\mu^{-1}(\mathrm{Id} - J_{\mu A_1}) + \cdots + \lambda_n\mu^{-1}(\mathrm{Id} - J_{\mu A_n}).$$

Then apply the definition of the Yosida regularization, (3.12). $\square$

Theorem 5.3. *For all $i \in I$, let A_i be a monotone operator from $\mathcal{H} \rightrightarrows \mathcal{H}$. Then $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ is monotone. Moreover,*

$$\operatorname{dom} J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \operatorname{dom} J_{\mu A_1} \cap \cdots \cap \operatorname{dom} J_{\mu A_n}, \ \textit{i.e.,} \tag{5.5}$$

$$\operatorname{ran}(\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) + \mathrm{Id}) = \operatorname{ran}(\mu A_1 + \mathrm{Id}) \cap \cdots \cap \operatorname{ran}(\mu A_n + \mathrm{Id}).$$

Consequently, $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ is maximal monotone if and only if $(\forall i)$ A_i is maximal monotone.

Proof. Since A_i is monotone, $(\mu A_i + \mathrm{Id})^{-1}$ is firmly nonexpansive, so there exists a nonexpansive mapping N_i such that $J_{\mu A_i} = \frac{N_i + \mathrm{Id}}{2}$. Then

$$\lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n} = \frac{(\lambda_1 N_1 + \cdots + \lambda_n N_n) + \mathrm{Id}}{2},$$

is firmly nonexpansive, since $\lambda_1 N_1 + \cdots + \lambda_n N_n$ is nonexpansive. This means that there exists a monotone operator B such that $(\mu B + \mathrm{Id})^{-1} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n}$. Then

$$\mu B = (\lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n})^{-1} - \mathrm{Id} = \mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}),$$

therefore $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) = B$ is monotone. Since $J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n}$, this gives

$$\operatorname{dom} J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \operatorname{dom} J_{\mu A_1} \cap \cdots \cap \operatorname{dom} J_{\mu A_n},$$

which is (5.5). If each A_i is maximal monotone, then μA_i is maximal monotone, and thus by Fact 3.43 $\operatorname{dom} J_{\mu A_i} = \mathcal{H}$. By (5.5), $\operatorname{dom} J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \mathcal{H}$ and since $\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ is maximal monotone, so is $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$. On the other hand, if $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ is maximal monotone, then $\operatorname{dom} J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \mathcal{H}$. It follows from (5.5) that $(\forall i \in I) \operatorname{dom} J_{\mu A_i} = \mathcal{H}$, thus μA_i must be maximal monotone and therefore A_i is maximal monotone. $\square$

Proposition 5.4. *For all $i \in I$, let A_i be a maximally monotone operator from $\mathcal{H} \rightrightarrows \mathcal{H}$. Let $\boldsymbol{A} = (A_1, A_1^{-1}, \ldots, A_m, A_m^{-1})$, $\boldsymbol{\lambda} = (\frac{1}{2m}, \frac{1}{2m}, \ldots, \frac{1}{2m})$, and $\mu = 1$. Then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathrm{Id}$.*

Proof. This follows directly from the definition of $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$, (5.2), and the resolvent identity, (3.13). $\square$

Proposition 5.5. *Let $\boldsymbol{A} = (A_1, \ldots, A_1)$. Then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = A_1$.*

Proof. We have

$$\begin{aligned}\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) &= \big((\lambda_1 + \cdots + \lambda_n)(A_1 + \mu^{-1}\,\mathrm{Id})^{-1}\big)^{-1} - \mu^{-1}\,\mathrm{Id} \\ &= \big((A_1 + \mu^{-1}\,\mathrm{Id})^{-1}\big)^{-1} - \mu^{-1}\,\mathrm{Id} = A_1 + \mu^{-1}\,\mathrm{Id} - \mu^{-1}\,\mathrm{Id} = A_1,\end{aligned}$$

which proves the result. $\square$

For clarification, in the following result we write $\mathcal{R}_\mu(A_1, \lambda_1, \cdots, A_n, \lambda_n)$ for $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$.

Proposition 5.6 (recursion)**.** *We have*

$$\mathcal{R}_\mu(A_1, \lambda_1, \ldots, A_n, \lambda_n) = \mathcal{R}_\mu\left(\mathcal{R}_\mu\left(A_1, \tfrac{\lambda_1}{1-\lambda_n}, \ldots, A_{n-1}, \tfrac{\lambda_{n-1}}{1-\lambda_n}\right), 1-\lambda_n, A_n, \lambda_n\right).$$

In particular, for $\lambda_1 = \cdots = \lambda_n = \frac{1}{n}$ one has

$$\mathcal{R}_\mu\left(A_1, \tfrac{1}{n}, \cdots, A_n, \tfrac{1}{n}\right) = \mathcal{R}_\mu\left(\mathcal{R}_\mu\left(A_1, \tfrac{1}{n-1}, \ldots, A_{n-1}, \tfrac{1}{n-1}\right), 1-\tfrac{1}{n}, A_n, \tfrac{1}{n}\right).$$

Proof. This follows from the definition of $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$. Indeed,

$$\begin{aligned}\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) &= \Big[\lambda_1(A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1}\Big]^{-1} - \mu^{-1}\,\mathrm{Id} \\ &= \Big[(1-\lambda_n)\Big(\frac{\lambda_1}{1-\lambda_n}(A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \frac{\lambda_{n-1}}{1-\lambda_n}(A_{n-1} + \mu^{-1}\,\mathrm{Id})^{-1}\Big) \\ &\qquad + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1}\Big]^{-1} - \mu^{-1}\,\mathrm{Id} \\ &= \Big[(1-\lambda_n)\Big(\mathcal{R}_\mu(A_1, \lambda_1/(1-\lambda_n), \cdots, A_{n-1}, \lambda_{n-1}/(1-\lambda_n)) + \mu^{-1}\,\mathrm{Id}\Big)^{-1} \\ &\qquad + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1}\Big]^{-1} - \mu^{-1}\,\mathrm{Id} \\ &= \mathcal{R}_\mu\left(\mathcal{R}_\mu\left(A_1, \tfrac{\lambda_1}{1-\lambda_n}, \ldots, A_{n-1}, \tfrac{\lambda_{n-1}}{1-\lambda_n}\right), 1-\lambda_n, A_n, \lambda_n\right).\end{aligned}$$

$\square$

Proposition 5.7 (Minty parametrization of $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$). *For all $i \in I$, let A_i be a maximally monotone operator from $\mathcal{H} \rightrightarrows \mathcal{H}$. Then for every $x \in \mathcal{H}$, we have*

$$\begin{aligned}&(J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x), x - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x)) = \\ &\qquad \lambda_1(J_{\mu A_1}(x), x - J_{\mu A_1}(x)) + \cdots + \lambda_n(J_{\mu A_n}(x), x - J_{\mu A_n}(x)).\end{aligned} \tag{5.6}$$

Consequently,

$$\operatorname{gra}\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \subset \lambda_1 \operatorname{gra}\mu A_1 + \cdots + \lambda_n \operatorname{gra}\mu A_n.$$

In particular,

$$\operatorname{gra}\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \subset \lambda_1 \operatorname{gra} A_1 + \cdots + \lambda_n \operatorname{gra} A_n.$$

Proof. As Minty's parametrization of $\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ is

$$\operatorname{gra}\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) = \big\{(J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x), x - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x)) \bigm| x \in \mathcal{H}\big\},$$

then applying (5.3) and

$$\operatorname{Id} - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \lambda_1(\operatorname{Id} - J_{\mu A_1}) + \cdots + \lambda_n(\operatorname{Id} - J_{\mu A_n}),$$

we have

$$\begin{aligned}&\operatorname{gra}\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \\ &= \{\left(\sum_{i=1}^n \lambda_i J_{\mu A_i} x, \sum_{i=1}^n \lambda_i(\operatorname{Id} - J_{\mu A_i})x\right) \mid x \in \mathcal{H}\} \\ &= \big\{(\lambda_1 J_{\mu A_1} x, \lambda_1(\operatorname{Id} - J_{\mu A_1})x) + \cdots + (\lambda_n J_{\mu A_n} x, \lambda_n(\operatorname{Id} - J_{\mu A_n})x) \bigm| x \in \mathcal{H}\big\} \\ &= \big\{\lambda_1 (J_{\mu A_1} x, (\operatorname{Id} - J_{\mu A_1})x) + \cdots + \lambda_n (J_{\mu A_n} x, (\operatorname{Id} - J_{\mu A_n})x) \bigm| x \in \mathcal{H}\big\} \\ &\subset \lambda_1 \operatorname{gra}\mu A_1 + \cdots + \lambda_n \operatorname{gra}\mu A_n.\end{aligned}$$

$\square$

Theorem 5.8 (self-duality). *For all $i \in I$, let A_i be a monotone operator on $\mathcal{H}$ and $\mu > 0$. Assume that $\sum_{i=1}^n \lambda_i = 1$ with $\lambda_i > 0$. Then*

$$(\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}))^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}), \text{ i.e.,} \tag{5.7}$$

$$\begin{aligned}&\left[\left(\lambda_1(A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\operatorname{Id})^{-1}\right)^{-1} - \mu^{-1}\operatorname{Id}\right]^{-1} = \\ &\left(\lambda_1(A_1^{-1} + \mu\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n^{-1} + \mu\operatorname{Id})^{-1}\right)^{-1} - \mu\operatorname{Id}.\end{aligned}$$

Proof. By (3.15) we have,

$$(A_i + \mu^{-1}\,\mathrm{Id})^{-1} = \mu\big(\,\mathrm{Id} - (\mathrm{Id} + \mu^{-1}A_i^{-1})^{-1}\big).$$

This and the fact that $\sum_{i=1}^{n}\lambda_i = 1$ gives

$$\begin{aligned}
&\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})\\
&= \Big[\lambda_1\mu\big(\,\mathrm{Id} - (\mathrm{Id} + \mu^{-1}A_1^{-1})^{-1}\big) + \cdots + \lambda_n\mu\big(\,\mathrm{Id} - (\mathrm{Id} + \mu^{-1}A_n^{-1})^{-1}\big)\Big]^{-1}\\
&\quad - \mu^{-1}\,\mathrm{Id}\\
&= \Big[\mu\big(\sum_{i=1}^{n}\lambda_i\,\mathrm{Id} - \sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\big)\Big]^{-1} - \mu^{-1}\,\mathrm{Id}\\
&= \Big(\,\mathrm{Id} + \big(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\big)\Big)^{-1} \circ (\mu^{-1}\,\mathrm{Id}) - \mu^{-1}\,\mathrm{Id}\,.
\end{aligned}$$

By (3.13) we have,

$$\left(\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)\right)^{-1} = \mathrm{Id} - \left(\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right)^{-1}.$$

Then,

$$\begin{aligned}
\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) &= \left[\mathrm{Id} - \left(\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right)^{-1}\right] \circ (\mu^{-1}\,\mathrm{Id}) - \mu^{-1}\,\mathrm{Id}\\
&= \mu^{-1}\,\mathrm{Id} - \left(\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right)^{-1} \circ (\mu^{-1}\,\mathrm{Id}) - \mu^{-1}\,\mathrm{Id}\\
&= -\left[\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right]^{-1} \circ (\mu^{-1}\,\mathrm{Id}).
\end{aligned}$$

$$
\begin{aligned}
&= -\left[\mu\left(\mathrm{Id} + \left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right)\right]^{-1}\\
&= -\left[\mu\,\mathrm{Id} + \mu\left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)^{-1}\right]^{-1}\\
&= -\left[\mu\,\mathrm{Id} + \left(\left(-\sum_{i=1}^{n}\lambda_i J_{\mu^{-1}A_i^{-1}}\right)\circ(\mu^{-1}\,\mathrm{Id})\right)^{-1}\right]^{-1}\\
&= -\big[\mu\,\mathrm{Id} + \big((-\lambda_1(\mathrm{Id} + \mu^{-1}A_1^{-1})^{-1}\circ(\mu^{-1}\,\mathrm{Id}) \;- \cdots\\
&\qquad -\; \lambda_n(\mathrm{Id} + \mu^{-1}A_n^{-1})^{-1}\circ(\mu^{-1}\,\mathrm{Id})\big)\big)^{-1}\big]^{-1}\\
&= -\Big[\mu\,\mathrm{Id} + \Big(-\lambda_1\left(\mu(\mathrm{Id} + \mu^{-1}A_1^{-1})\right)^{-1} - \cdots\\
&\qquad -\lambda_n\left(\mu(\mathrm{Id} + \mu^{-1}A_n^{-1})\right)^{-1}\Big)^{-1}\Big]^{-1}\\
&= -\left[\mu\,\mathrm{Id} + \left(-\lambda_1\left(\mu\,\mathrm{Id} + A_1^{-1}\right)^{-1} - \cdots - \lambda_n\left(\mu\,\mathrm{Id} + A_n^{-1}\right)^{-1}\right)^{-1}\right]^{-1}.
\end{aligned}
$$

To continue, we write

$$
\begin{aligned}
&\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})\\
&= -\left[\mu\,\mathrm{Id} + \left(-\left(\lambda_1\left(\mu\,\mathrm{Id} + A_1^{-1}\right)^{-1} + \cdots + \lambda_n\left(\mu\,\mathrm{Id} + A_n^{-1}\right)^{-1}\right)\right)^{-1}\right]^{-1}\\
&= -\left[\mu\,\mathrm{Id} + \left(\lambda_1\left(\mu\,\mathrm{Id} + A_1^{-1}\right)^{-1} + \cdots + \lambda_n\left(\mu\,\mathrm{Id} + A_n^{-1}\right)^{-1}\right)^{-1}\circ(-\,\mathrm{Id})\right]^{-1}\\
&= \left[\left(\mu\,\mathrm{Id} + \left(\lambda_1\left(\mu\,\mathrm{Id} + A_1^{-1}\right)^{-1} + \cdots + \lambda_n\left(\mu\,\mathrm{Id} + A_n^{-1}\right)^{-1}\right)^{-1}\circ(-\,\mathrm{Id})\right)\circ(-\,\mathrm{Id})\right]^{-1}\\
&= \left[-\mu\,\mathrm{Id} + \left(\lambda_1\left(\mu\,\mathrm{Id} + A_1^{-1}\right)^{-1} + \cdots + \lambda_n\left(\mu\,\mathrm{Id} + A_n^{-1}\right)^{-1}\right)^{-1}\right]^{-1}\\
&= \left(\mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda})\right)^{-1},
\end{aligned}
$$

which gives $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) = \left(\mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda})\right)^{-1}$. Taking inverses on both sides we obtain (5.7). $\square$

Corollary 5.9. *Let $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone operators for all $i = 1,\dots,n$, λ_i be strictly positive real numbers such that $\sum_{i=1}^{n}\lambda_i = 1$, and $\mu > 0$. Set $\boldsymbol{A} = (A_1,\dots,A_n)$ and $\boldsymbol{A}^{-1} = (A_1^{-1},\dots,A_n^{-1})$. Then*

$$
J_{(\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}))^{-1}} = J_{\mu^{-1}\mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda})} = \lambda_1 J_{\mu^{-1}A_1^{-1}} + \cdots + \lambda_n J_{\mu^{-1}A_n^{-1}}. \tag{5.8}
$$

In particular,

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}} = \lambda_1 J_{A_1^{-1}} + \cdots + \lambda_n J_{A_n^{-1}}. \tag{5.9}$$

Proof. Combine (5.3) with Theorem 5.8. □

5.2 The resolvent average of positive semidefinite matrices

This section covers results specific to positive definite and positive semidefinite $N \times N$ matrices. Recall the following fact for these types of matrices:

Fact 5.10. [42, Corollary 7.7.4.(a)] and [46, Section 16.E] or [26, page 55]. *Let $A, B \in \mathbb{S}^N_{++}$, we have*

$$A \succeq B \quad \Leftrightarrow \quad A^{-1} \preceq B^{-1} \tag{5.10}$$

and

$$A \succ B \quad \Leftrightarrow \quad A^{-1} \prec B^{-1}; \tag{5.11}$$

Proposition 5.11. *Assume that* $(\forall\ i)$ $A_i, B_i \in \mathbb{S}^N_+$ *and* $A_i \succeq B_i$. *Then*

$$\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \succeq \mathcal{R}_\mu(\boldsymbol{B},\boldsymbol{\lambda}). \tag{5.12}$$

Furthermore, if additionally some $A_j \succ B_j$, then $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \succ \mathcal{R}_\mu(\boldsymbol{B},\boldsymbol{\lambda})$.

Proof. Note that $\forall\ \mu > 0$,

$$A_i + \mu^{-1}\,\mathrm{Id} \succeq B_i + \mu^{-1}\,\mathrm{Id} \succ 0,$$

so that

$$0 \prec (A_i + \mu^{-1}\,\mathrm{Id})^{-1} \preceq (B_i + \mu^{-1}\,\mathrm{Id})^{-1},$$

by (5.10). As $\mathbb{S}^N_+$ and $\mathbb{S}^N_{++}$ are convex cones, we obtain that

$$0 \prec \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} \preceq \sum_{i=1}^n \lambda_i (B_i + \mu^{-1}\,\mathrm{Id})^{-1}. \tag{5.13}$$

Using (5.10) on (5.13), followed by subtracting $\mu^{-1}\,\mathrm{Id}$, gives

$$\Big[\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1}\Big]^{-1} - \mu^{-1}\,\mathrm{Id} \succeq \Big[\sum_{i=1}^n \lambda_i (B_i + \mu^{-1}\,\mathrm{Id})^{-1}\Big]^{-1} - \mu^{-1}\,\mathrm{Id},$$

which establishes (5.12). The "furthermore" part follows analogously using (5.11). □

Theorem 5.12. *Assume that* $(\forall\ i)\ A_i \in \mathbb{S}^N_+$. *Then* $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \in \mathbb{S}^N_+$. *Furthermore, if additionally some* $A_j \in \mathbb{S}^N_{++}$, *then* $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$.

Proof. This follows from Proposition 5.11 (with each $B_i = 0$) and Proposition 5.5. ☐

5.2.1 Inequalities among means

In this section, we derive an inequality comparing the resolvent average to the arithmetic and harmonic averages when $(\forall\ i)\ A_i \in \mathbb{S}^N_{++}$. We start by computing the proximal average of general linear-quadratic functions thereby extending Fact 2.70.

Lemma 5.13. *Let* $A_i \in \mathbb{S}^N_+$, $b_i \in \mathbb{R}^N$, $r_i \in \mathbb{R}$. *If each* $f_i = \mathfrak{q}_{A_i} + \langle b_i, \cdot\rangle + r_i$, *i.e., linear-quadratic, then* $\forall x^* \in \mathbb{R}^N$,

$$\begin{aligned}
&\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x^*)\\
&= \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x^*) + \left\langle x^*, \Big(\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\operatorname{Id})^{-1}\Big)^{-1} \sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\operatorname{Id})^{-1} b_i \right\rangle\\
&\qquad + \mathfrak{q}_{\left(\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\operatorname{Id})^{-1}\right)^{-1}} \left(\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\operatorname{Id})^{-1} b_i\right)\\
&\qquad - \sum_{i=1}^n \lambda_i \left(\mathfrak{q}_{(A_i+\mu^{-1}\operatorname{Id})^{-1}}(b_i) - r_i\right).
\end{aligned} \tag{5.14}$$

In particular, if $(\forall i)$ f_i *is quadratic, i.e.,* $b_i = 0, r_i = 0$, *then* $\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ *is quadratic with*

$$\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})};$$

If $(\forall i)$ f_i *is affine, i.e.,* $A_i = 0$, *then* $\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ *is affine.*

Proof. We have $f_i + \mu^{-1}\mathfrak{q} = \mathfrak{q}_{(A_i+\mu^{-1}\operatorname{Id})} + \langle b_i, \cdot\rangle + r_i$ and applying Fact 2.58 and then expanding we get

$$\begin{aligned}
(f_i+\mu^{-1}\mathfrak{q})^*(x^*) &= \mathfrak{q}_{(A_i+\mu^{-1}\operatorname{Id})^{-1}}(x^*-b_i) - r_i\\
&= \mathfrak{q}_{(A_i+\mu^{-1}\operatorname{Id})^{-1}}(x^*) - \left\langle x^*, (A_i+\mu^{-1}\operatorname{Id})^{-1} b_i\right\rangle\\
&\qquad + \mathfrak{q}_{(A_i+\mu^{-1}\operatorname{Id})^{-1}}(b_i) - r_i.
\end{aligned}$$

Then $\big(\lambda_1(f_1+\mu^{-1}\mathfrak{q})^*+\cdots+\lambda_n(f_n+\mu^{-1}\mathfrak{q})^*\big)(x^*) =$

$$\sum_{i=1}^n \lambda_i\bigg(\mathfrak{q}_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^*) - \big\langle x^*, (A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i\big\rangle + \mathfrak{q}_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\bigg)$$
$$= \mathfrak{q}_{\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^*) - \left\langle x^*, \sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i\right\rangle$$
$$+ \sum_{i=1}^n \lambda_i\big(\mathfrak{q}_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big).$$

It follows again from Fact 2.58 that

$$\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x^*) = \mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^* + \sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i)$$
$$- \sum_{i=1}^n \lambda_i\big(\mathfrak{q}_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big) - \mathfrak{q}_{\mu^{-1}\,\mathrm{Id}}(x^*).$$

Since

$$\mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^* + \sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i)$$
$$= \mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^*)$$
$$+ \left\langle x^*, [\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i\right\rangle$$
$$+ \mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i),$$

we obtain that

$$\mathcal{P}_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x^*) = \mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}-\mu^{-1}\,\mathrm{Id}}(x^*)$$
$$+ \left\langle x^*, [\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i\right\rangle$$
$$+ \mathfrak{q}_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}b_i)$$
$$- \sum_{i=1}^n \lambda_i\big(\mathfrak{q}_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big),$$

which is (5.14). The remaining claims are immediate from (5.14) and that $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = 0$ when $(\forall\ i)\ A_i = 0$ by Proposition 5.5. $\square$

We are ready for the main result of this section:

Theorem 5.14 (harmonic-resolvent-arithmetic average inequality and limits)**.**
Let $A_1, \ldots, A_n \in \mathbb{S}^N_{++}$. We have

(i)
$$\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) \preceq \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \preceq \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}); \tag{5.15}$$
In particular, $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$.

(ii) $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})$ *when* $\mu \to 0^+$.

(iii) $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})$ *when* $\mu \to +\infty$.

Proof. (i): According to Fact 2.69,

$$(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* \le \mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \le \lambda_1 f_1 + \cdots + \lambda_n f_n. \tag{5.16}$$

Let $f_i = \mathfrak{q}_{A_i}$. Using $(\mathfrak{q}_{A_i})^* = \mathfrak{q}_{A_i^{-1}}$ (by Fact 2.57) and Lemma 5.13 we have

$$\begin{aligned}(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* &= (\lambda_1 \mathfrak{q}_{A_1^{-1}} + \cdots + \lambda_n \mathfrak{q}_{A_n^{-1}})^* = (\mathfrak{q}_{\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1}})^* \\ &= \mathfrak{q}_{(\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1}} = \mathfrak{q}_{\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})}. &(5.17)\\ \lambda_1 f_1 + \cdots + \lambda_n f_n &= \mathfrak{q}_{\lambda_1 A_1 + \cdots + \lambda_n A_n} = \mathfrak{q}_{\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})}, &(5.18)\\ \mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) &= \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}. &(5.19)\end{aligned}$$

Then (5.16) becomes

$$\mathfrak{q}_{\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})} \le \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})} \le \mathfrak{q}_{\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})}.$$

As $\mathfrak{q}_X \le \mathfrak{q}_Y \Leftrightarrow X \preceq Y$, (5.15) is established. Since $A_i \in \mathbb{S}^N_{++}, A_i^{-1} \in \mathbb{S}^N_{++}, \lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1} \in \mathbb{S}^N_{++}$, we have $\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) = (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1} \in \mathbb{S}^N_{++}$, thus $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$ by (5.15). (Alternatively, apply Theorem 5.12.)

(ii) and (iii): Observe that $(\forall i)$ $\big(\lambda_i \star f_i\big)^* = \lambda_i f_i^* = \lambda_i \mathfrak{q}_{A_i^{-1}}$ has full domain. By Fact 2.56,

$$(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = (\lambda_1 \star f_1) \,\square \cdots \square\, (\lambda_n \star f_n)\,.$$

By Fact 2.72, $\forall x \in \mathbb{R}^N$ one has

$$\lim_{\mu\to 0^+} \mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 f_1 + \cdots + \lambda_n f_n)(x),$$

$$\lim_{\mu\to +\infty} \mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^*(x).$$

Since $(\forall i)$ f_i, f_i^* are differentiable on $\mathbb{R}^N$, so is $\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ by Fact 2.71. According to Fact 2.60, for all x

$$\lim_{\mu\to 0^+} \nabla\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \lambda_1 \nabla f_1(x) + \cdots + \lambda_n \nabla f_n(x), \tag{5.20}$$

$$\lim_{\mu\to +\infty} \nabla\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \nabla(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^*(x). \tag{5.21}$$

Moreover, the convergences in (5.20)-(5.21) are uniform on every closed bounded subset of $\mathbb{R}^N$. Now it follows from (5.17)-(5.19) that

$$\nabla\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}),$$
$$\nabla(\lambda_1 f_1 + \cdots + \lambda_n f_n) = \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}),$$
$$\nabla(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}).$$

(5.20)–(5.21) becomes

$$\lim_{\mu\to 0^+} \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})x = \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})x, \tag{5.22}$$

$$\lim_{\mu\to +\infty} \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})x = \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})x, \tag{5.23}$$

where the convergences are uniform on every closed bounded subset of $\mathbb{R}^N$. Hence (ii) and (iii) follow from (5.22) and (5.23). $\square$

Note that in Theorem 5.14(ii) and (ii), there is no ambiguity since all norms in finite dimensional spaces are equivalent.

Definition 5.15. A function $g\colon \mathbb{D} \to \mathbb{S}^N$, where $\mathbb{D}$ is a convex subset of $\mathbb{S}^N$, is *matrix convex* if $\forall A_1, A_2 \in \mathbb{D}$, $\forall \lambda \in [0, 1]$,

$$g(\lambda A_1 + (1-\lambda)A_2) \preceq \lambda g(A_1) + (1-\lambda) g(A_2).$$

Matrix concave functions are defined similarly.

It is easy to see that a symmetric matrix valued function g is matrix concave if and only if $\forall x \in \mathbb{R}^N$ the function $A \mapsto \mathfrak{q}_{g(A)}(x)$ is concave. Similarly, g is matrix convex if and only if $A \mapsto \mathfrak{q}_{g(A)}(x)$ is convex. Some immediate consequences of Theorem 5.14 on matrix-valued functions are:

Corollary 5.16. *Assume that* $(\forall i)\ A_i \in \mathbb{S}^N_{++}$ *and* $\sum_{i=1}^n \lambda_i = 1$ *with* $\lambda_i > 0$. *Then*

$$(\lambda_1 A_1 + \cdots + \lambda_n A_n)^{-1} \preceq \lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1}.$$

Consequently, the matrix function $X \mapsto X^{-1}$ *is matrix convex on* $\mathbb{S}^N_{++}$.

Proof. Apply (5.15) with $\boldsymbol{A} = (A_1^{-1}, \cdots, A_n^{-1})$. □

Corollary 5.17. *For every* $\mu > 0$, *the resolvent average matrix function* $\boldsymbol{A} \mapsto \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ *given by*

$$(A_1, \cdots, A_n) \mapsto [\lambda_1(A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\operatorname{Id})^{-1}]^{-1} - \mu^{-1}\operatorname{Id}$$

is matrix concave on $\mathbb{S}^N_{++} \times \cdots \times \mathbb{S}^N_{++}$. (5.24)

For each $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$ *with* $\sum_{i=1}^n \lambda_i = 1$ *and* $\lambda_i > 0\ \forall i$, *the harmonic average matrix function*

$$(A_1, \cdots, A_n) \mapsto (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1} \text{ is matrix concave} \tag{5.25}$$

on $\mathbb{S}^N_{++} \times \cdots \times \mathbb{S}^N_{++}$. *Consequently, the harmonic average function*

$$(x_1, \cdots, x_n) \mapsto \frac{1}{x_1^{-1} + \cdots + x_n^{-1}} \text{ is concave} \tag{5.26}$$

on $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$.

Proof. Set $f_i = \mathfrak{q}_{A_i}$. Then $\forall x \in \mathbb{R}^N$, we have from (2.17)

$$\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \min_{\lambda_1 x_1 + \cdots + \lambda_n x_n = x} \Big(\big(\lambda_1 \mathfrak{q}_{A_1}(x_1) + \cdots + \lambda_n \mathfrak{q}_{A_n}(x_n)\big) + \big(\mu^{-1}\lambda_1 \mathfrak{q}(x_1) + \cdots + \mu^{-1}\lambda_n \mathfrak{q}(x_n)\big)\Big) - \mu^{-1}\mathfrak{q}(x).$$

Since for each fixed $(x_1, \ldots, x_n)$,

$$(A_1, \cdots, A_n) \mapsto \big(\lambda_1 \mathfrak{q}_{A_1}(x_1) + \cdots + \lambda_n \mathfrak{q}_{A_n}(x_n)\big),$$

is affine, being the minimum of affine functions we have that $\forall x$ the function

$$(A_1, \cdots, A_n) \mapsto \mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x),$$

is concave. As $\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}(x)$ by Lemma 5.13, this shows that $\forall x \in \mathbb{R}^N$ the function

$$\boldsymbol{A} = (A_1, \cdots, A_n) \mapsto \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}(x) \text{ is concave,}$$

so $\boldsymbol{A} \mapsto \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is matrix concave.

Now by Theorem 5.14(iii), $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})$ when $\mu \to +\infty$. This and (5.24) implies that

$$\boldsymbol{A} \mapsto \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}),$$

is also matrix concave. Equation (5.26) follows from (5.25) by setting $N = 1$ and $\lambda_1 = \cdots = \lambda_n = 1/n$. ☐

Remark 5.18. Corollary 5.16 is well-known, cf. [59, Proposition 2.56]. Equation (5.26) is also well-known, cf. [27, Exercise 3.17].

The next theorem provides a simplified proof of Theorem 5.8 when the operators are positive definite matrices.

Theorem 5.19 (self-duality)**.** *Let* $(\forall\ i)\ A_i \in \mathbb{S}^N_{++}$ *and* $\mu > 0$. *Assume that* $\sum_{i=1}^n \lambda_i = 1$ *with* $\lambda_i > 0$. *Then*

$$\big[\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})\big]^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}), \tag{5.27}$$

i.e.,

$$\left[\Big(\lambda_1(A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\operatorname{Id})^{-1}\Big)^{-1} - \mu^{-1}\operatorname{Id}\right]^{-1} =$$

$$\Big(\lambda_1(A_1^{-1} + \mu\operatorname{Id})^{-1} + \cdots + \lambda_n(A_n^{-1} + \mu\operatorname{Id})^{-1}\Big)^{-1} - \mu\operatorname{Id}.$$

In particular, for $\mu = 1$, $[\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})]^{-1} = \mathcal{R}_1(\boldsymbol{A}^{-1}, \boldsymbol{\lambda})$.

Proof. Let $f_i = q_{A_i}$. By Fact 2.66, $\big(\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \mathcal{P}_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})$, taking subgradients on both sides, followed by using Fact 2.59, we obtain that

$$\partial\big(\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \big(\partial\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^{-1} = \partial\big(\mathcal{P}_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})\big).$$

By Lemma 5.13, $\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mathfrak{q}_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}, \mathcal{P}_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})) = \mathfrak{q}_{\mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda})}$, we have

$$\partial\mathcal{P}_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}),$$
$$\partial\mathcal{P}_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}) = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}).$$

Hence

$$\big[\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})\big]^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}),$$

as claimed. ☐

Remark 5.20. Although the harmonic and arithmetic average lack self-duality, they are dual to each other:

$$[\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda})]^{-1} = \lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1} = \mathcal{A}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}),$$
$$[\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda})]^{-1} = \big[\lambda_1(A_1^{-1})^{-1} + \cdots + \lambda_n(A_n^{-1})^{-1}\big]^{-1} = \mathcal{H}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}).$$

5.2.2 A comparison to weighted geometric means

To compare the resolvent average with the well-known geometric mean, we restrict our attention to non-negative real numbers (1×1 matrices). When $\boldsymbol{A} = \boldsymbol{x} = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}_+$ and $\mu = 1$, we write

$$\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \big(\lambda_1(x_1+1)^{-1} + \cdots + \lambda_n(x_n+1)^{-1}\big)^{-1} - 1,$$

and $\boldsymbol{x}^{-1} = (1/x_1, \ldots, 1/x_n)$ when $(\forall i)\ x_i \in \mathbb{R}_{++}$.

Proposition 5.21. *Let $(\forall\ i)\ x_i > 0, y_i > 0$. We have*

(i) (**harmonic-resolvent-arithmetic mean inequality**)*:*

$$\big(\lambda_1 x_1^{-1} + \cdots + \lambda_n x_n^{-1}\big)^{-1} \leq \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq \lambda_1 x_1 + \cdots + \lambda_n x_n. \quad (5.28)$$

Moreover, $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ if and only if $x_1 = \cdots = x_n$.

(ii) (**self-duality**)*: $[\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})]^{-1} = \mathcal{R}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda})$.*

(iii) *If $\boldsymbol{x} = (x_1, \ldots, x_1)$, then $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1$.*

(iv) *If $\boldsymbol{x} = (x_1, x_1^{-1}, x_2, x_2^{-1}, \ldots, x_n, x_n^{-1})$ and $\boldsymbol{\lambda} = (\frac{1}{2n}, \ldots, \frac{1}{2n})$, then $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = 1$.*

(v) *The function $\boldsymbol{x} \mapsto \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$ is concave on $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$.*

(vi) *If $\boldsymbol{x} \succeq \boldsymbol{y}$, then $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) \geq \mathcal{R}(\boldsymbol{y}, \boldsymbol{\lambda})$.*

Proof. (i): For (5.28), apply Theorem 5.14(i) with $\mu = 1$. Now

$$\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$$

is equivalent to

$$\big(\lambda_1(x_1+1)^{-1} + \cdots + \lambda_n(x_n+1)^{-1}\big)^{-1} = \lambda_1 x_1 + \cdots + \lambda_n x_n + 1, \quad (5.29)$$

As $\sum_{i=1}^n \lambda_i = 1$, (5.29) is the same as

$$\lambda_1 \frac{1}{(x_1+1)} + \cdots + \lambda_n \frac{1}{(x_n+1)} = \frac{1}{\lambda_1(x_1+1) + \cdots + \lambda_n(x_n+1)}.$$

Since the function $x \mapsto 1/x$ is strictly convex on $\mathbb{R}_{++}$, we must have

$$x_1 = \cdots = x_n.$$

(ii): Theorem 5.19. (iii): Proposition 5.5. (iv): Proposition 5.4. (v): Corollary 5.17. (vi): Proposition 5.11. □

Proposition 5.21 and Fact 2.64 demonstrate that $\mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$ and $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})$ have strikingly similar properties. Are they the same?

Example 5.22.

(i) Let $\lambda = (\frac{1}{2}, \frac{1}{2})$. When $x = (0,1)$, $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda}) = 0$ but $\mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{3}$, so $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda}) \neq \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$.

(ii) *Is it right that $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda}) \leq \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$ for all $x \in \mathbb{R}^2_{++}$?* The answer is also no. Assume that $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda}) \leq \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$, $\forall\ \boldsymbol{x} \in \mathbb{R}_{++} \times \mathbb{R}_{++}$. Taking the inverse of both sides, followed by applying the self-duality of $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})$ and $\mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$, gives

$$\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})^{-1} \geq \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})^{-1} = \mathcal{R}(\boldsymbol{x}^{-1},\boldsymbol{\lambda}) \geq \mathcal{G}(\boldsymbol{x}^{-1},\boldsymbol{\lambda}) = \mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})^{-1},$$

and this gives that $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})^{-1} = \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})^{-1}$ so that $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda}) = \mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$. This is a contradiction to (i), thus $\mathcal{R}(\boldsymbol{x},\boldsymbol{\lambda})$ is distinct from $\mathcal{G}(\boldsymbol{x},\boldsymbol{\lambda})$.

Chapter 6

Near Equality, Near Convexity, Sums of Maximally Monotone Operators, and Averages of Firmly Nonexpansive Mappings

In this chapter, based on [20], we introduce near equality for sets and show that this notion is useful in the study of nearly convex sets. These results are the key to study ranges of sums of maximal monotone operators in the next section. Recall that I denotes an index set

$$I = \{1, 2, \ldots, m\},$$

for a strictly positive integer m.

6.1 Near equality and near convexity

Definition 6.1 (near equality)**.** Let A and B be subsets of $\mathbb{R}^n$. We say that A and B are *nearly equal*, if

$$\overline{A} = \overline{B} \text{ and } \operatorname{ri} A = \operatorname{ri} B. \tag{6.1}$$

and denote this by $A \approx B$.

Remark 6.2. The following holds:

$$A \approx B \;\Rightarrow\; \operatorname{int} A = \operatorname{int} B. \tag{6.2}$$

Observe that if $\operatorname{int} A \neq \varnothing$ then there exists $x \in A$ such that $B(x, \epsilon) \subseteq A$ for some $\epsilon > 0$. This is an n-dimensional convex set in $\mathbb{R}^n$ and thus $\operatorname{aff} A = \mathbb{R}^n$

and so $\operatorname{int} A = \operatorname{ri} A = \operatorname{ri} B$. Thus, $B(x,\epsilon) \subseteq B$ and so $\operatorname{int} A = \operatorname{int} B$. Similarly, $\operatorname{int} B \neq \varnothing \Rightarrow \operatorname{int} A \neq \varnothing$, thus $\operatorname{int} A = \varnothing \Leftrightarrow \operatorname{int} B = \varnothing$. Altogether, $A \approx B \Rightarrow \operatorname{int} A = \operatorname{int} B$.

Proposition 6.3 (equivalence relation)**.** *The following hold for any subsets A, B, C of $\mathbb{R}^n$.*

(i) $A \approx A$.

(ii) $A \approx B \Rightarrow B \approx A$.

(iii) $A \approx B$ *and* $B \approx C \Rightarrow A \approx C$.

Proposition 6.4 (squeeze theorem)**.** *Let A, B, C be subsets of $\mathbb{R}^n$ such that $A \approx C$ and $A \subseteq B \subseteq C$. Then $A \approx B \approx C$.*

Proof. By assumption, $\overline{A} = \overline{C}$ and $\operatorname{ri} A = \operatorname{ri} C$. Thus $\overline{A} = \overline{B} = \overline{C}$ and by Lemma 2.24, $\operatorname{aff}(A) = \operatorname{aff}(\overline{A}) = \operatorname{aff}(\overline{C}) = \operatorname{aff}(C)$. Hence $\operatorname{aff} A = \operatorname{aff} B = \operatorname{aff} C$ and so, by Lemma 2.22, $\operatorname{ri} A \subseteq \operatorname{ri} B \subseteq \operatorname{ri} C$. Since $\operatorname{ri} A = \operatorname{ri} C$, we deduce that $\operatorname{ri} A = \operatorname{ri} B = \operatorname{ri} C$. Therefore, $A \approx B \approx C$. ☐

The equivalence relation "$\approx$" is best suited for studying nearly convex sets (defined next), as we do have that, e.g., $\mathbb{Q} \approx \mathbb{R} \smallsetminus \mathbb{Q}$!

Definition 6.5 (near convexity)**.** [59, Theorem 12.41] Let A be a subset of $\mathbb{R}^n$. Then A is *nearly convex* if there exists a convex subset C of $\mathbb{R}^n$ such that $C \subseteq A \subseteq \overline{C}$.

Lemma 6.6. *Let A be a nearly convex subset of $\mathbb{R}^n$, say $C \subseteq A \subseteq \overline{C}$, where C is a convex subset of $\mathbb{R}^n$. Then*

$$A \approx \overline{A} \approx \operatorname{ri} A \approx \operatorname{conv} A \approx \operatorname{ri} \operatorname{conv} A \approx C. \tag{6.3}$$

In particular, the following hold.

(i) $\overline{A}$ *and* $\operatorname{ri} A$ *are convex.*

(ii) *If* $A \neq \varnothing$, *then* $\operatorname{ri} A \neq \varnothing$.

Proof. We have

$$C \subseteq A \subseteq \operatorname{conv} A \subseteq \overline{C} \quad \text{and} \quad C \subseteq A \subseteq \overline{A} \subseteq \overline{C}. \tag{6.4}$$

Since $C \approx \overline{C}$ by Fact 2.25(iv), it follows from Proposition 6.4 that

$$A \approx \overline{A} \approx \operatorname{conv} A \approx C. \tag{6.5}$$

This implies

$$\operatorname{ri}(\operatorname{ri} A) = \operatorname{ri}(\operatorname{ri} C) = \operatorname{ri} C = \operatorname{ri} A \tag{6.6}$$

and

$$\overline{\operatorname{ri} A} = \overline{\operatorname{ri} C} = \overline{C} = \overline{A} \tag{6.7}$$

by Fact 2.25(iii). Therefore, $\operatorname{ri} A \approx A$. Applying this to $\operatorname{conv} A$, which is nearly convex, it also follows that $\operatorname{ri}\operatorname{conv} A \approx \operatorname{conv} A$. Finally, (i) holds because $A \approx C$ while (ii) follows from $\operatorname{ri} A = \operatorname{ri} C$ and Fact 2.25(ii). □

Remark 6.7. The assumption of near convexity in Lemma 6.6 is necessary: consider $\mathbb{R}$ with $A = \mathbb{Q}$. Then $\operatorname{ri}\mathbb{Q} = \varnothing$ but $\mathbb{Q}$ is obviously not. Thus (6.3) fails, although (i) still holds.

Lemma 6.8 (characterization of near convexity)**.** *Let $A \subseteq \mathbb{R}^n$. Then the following are equivalent.*

(i) *A is nearly convex.*

(ii) *$A \approx \operatorname{conv} A$.*

(iii) *A is nearly equal to a convex set.*

(iv) *A is nearly equal to a nearly convex set.*

(v) *$\operatorname{ri}\operatorname{conv} A \subseteq A$.*

Proof. "(i)⇒(ii)": Apply Lemma 6.6. "(ii)⇒(v)": Indeed, $\operatorname{ri}\operatorname{conv} A = \operatorname{ri} A \subseteq A$. "(v)⇒(i)": Set $C = \operatorname{ri}\operatorname{conv} A$. By Fact 2.25(iii), $C \subseteq A \subseteq \overline{\operatorname{conv}} A = \overline{\operatorname{ri}\operatorname{conv} A} = \overline{C}$. "(ii)⇒(iii)": Clear. "(iii)⇒(i)": Suppose that $A \approx C$, where C is convex. Then, using Fact 2.25(iii), $\operatorname{ri} C = \operatorname{ri} A \subseteq A \subseteq \overline{A} = \overline{C} = \overline{\operatorname{ri} C}$. Hence A is nearly convex. "(iii)⇒(iv)": Clear. "(iv)⇒(iii)": (The following simple proof was suggested by a referee of [20]) Suppose $A \approx B$, where B is nearly convex. Then, applying the already verified implications "(i)⇒(ii)" and "(ii)⇒(iii)" to the set B, we see that $B \approx C$ for some convex set C. Using Proposition 6.3(iii), we conclude that $A \approx C$. □

Remark 6.9. The condition appearing in Lemma 6.8(v) was also used by Minty [49] and named "almost-convex".

Remark 6.10. Brézis and Haraux [28] define, for two subsets A and B of $\mathbb{R}^n$,

$$A \simeq B \quad \Leftrightarrow \quad \overline{A} = \overline{B} \quad \text{and} \quad \operatorname{int} A = \operatorname{int} B. \tag{6.8}$$

(i) In view of (6.2), it is clear that $A \approx B \Rightarrow A \simeq B$.

(ii) On the other hand, $A \simeq B \not\Rightarrow A \approx B$: indeed, consider $\mathbb{R}^2$ with $A = \mathbb{Q} \times \{0\}$, and $B = \mathbb{R} \times \{0\}$. Then $\operatorname{int} A = \operatorname{int} B = \varnothing$, but $\operatorname{ri} A \neq \operatorname{ri} B$.

(iii) The implications (iii)⇒(i) and (ii)⇒(i) in Lemma 6.8 fail for $\simeq$: indeed, consider $\mathbb{R}^2$ with $A = (\mathbb{R} \smallsetminus \{0\}) \times \{0\}$ and $C = \operatorname{conv} A = \mathbb{R} \times \{0\}$. Then C is convex and $A \simeq C$. However, A is not nearly convex because $\operatorname{ri} A \neq \operatorname{ri} \overline{A}$.

Proposition 6.11. *Let A and B be nearly convex subsets of $\mathbb{R}^n$. Then the following are equivalent.*

(i) $A \approx B$.

(ii) $\overline{A} = \overline{B}$.

(iii) $\operatorname{ri} A = \operatorname{ri} B$.

(iv) $\overline{\operatorname{conv} A} = \overline{\operatorname{conv} B}$.

(v) $\operatorname{ri} \operatorname{conv} A = \operatorname{ri} \operatorname{conv} B$.

Proof. "(i)⇒(ii)": This is clear from the definition of $\approx$. "(ii)⇒(iii)": $\operatorname{ri} \overline{A} = \operatorname{ri} A$ and $\operatorname{ri} \overline{B} = \operatorname{ri} B$ by Lemma 6.6. "(iii)⇒(iv)": $\overline{\operatorname{ri} A} = \overline{\operatorname{conv} A}$ and $\overline{\operatorname{ri} B} = \overline{\operatorname{conv} B}$ by Lemma 6.6. "(iv)⇒(v)": $\operatorname{ri} \overline{\operatorname{conv} A} = \operatorname{ri} \operatorname{conv} A$ and $\operatorname{ri} \overline{\operatorname{conv} B} = \operatorname{ri} \operatorname{conv} B$. "(v)⇒(i)": Lemma 6.6 gives that $\operatorname{ri} \operatorname{conv} A = \operatorname{ri} A$ and $\operatorname{ri} \operatorname{conv} B = \operatorname{ri} B$ so that $\operatorname{ri} A = \operatorname{ri} B$, $\overline{\operatorname{ri} \operatorname{conv} A} = \overline{\operatorname{conv} A} = \overline{A}$ and $\overline{\operatorname{ri} \operatorname{conv} B} = \overline{\operatorname{conv} B} = \overline{B}$ so that $\overline{A} = \overline{B}$. Hence (i) holds. ☐

The next results generalize Rockafellar's Fact 2.26 to nearly convex sets.

Lemma 6.12. *Let A_1 and A_2 be nearly convex sets in $\mathbb{R}^n$ such that $\operatorname{ri} A_1 \cap \operatorname{ri} A_2 \neq \varnothing$. Then $A_1 \cap A_2$ is nearly convex and*

$$\operatorname{ri}(A_1 \cap A_2) = \operatorname{ri} A_1 \cap \operatorname{ri} A_2.$$

Proof. From Lemma 6.8(v) and the definition of the convex hull, we have

$$\begin{aligned} \operatorname{ri}(\operatorname{conv} A_1) &\subseteq A_1 \subseteq \operatorname{conv} A_1 \\ \operatorname{ri}(\operatorname{conv} A_2) &\subseteq A_2 \subseteq \operatorname{conv} A_2. \end{aligned}$$

This implies that

$$\operatorname{ri}(\operatorname{conv} A_1) \cap \operatorname{ri}(\operatorname{conv} A_2) \subseteq A_1 \cap A_2 \subseteq \operatorname{conv} A_1 \cap \operatorname{conv} A_2. \tag{6.9}$$

From Lemma 6.6 we have $\operatorname{ri} A_1 = \operatorname{ri}\operatorname{conv} A_1$ and $\operatorname{ri} A_2 = \operatorname{ri}\operatorname{conv} A_2$ and thus $\operatorname{ri}\operatorname{conv} A_1 \cap \operatorname{ri}\operatorname{conv} A_2 \neq \varnothing$. Then we can apply Fact 2.26 and Lemma 6.6(i) to get,

$$\operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2) = \operatorname{ri}(\operatorname{conv} A_1) \cap \operatorname{ri}(\operatorname{conv} A_2),$$

so by (6.9),

$$\operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2) \subseteq A_1 \cap A_2 \subseteq \operatorname{conv} A_1 \cap \operatorname{conv} A_2.$$

Thus by Fact 2.27, $A_1 \cap A_2$ is nearly convex and by Lemma 6.6,

$$A_1 \cap A_2 \approx \operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2).$$

Using (6.6), this means

$$\operatorname{ri}(A_1 \cap A_2) = \operatorname{ri}\left(\operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2)\right) = \operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2). \quad (6.10)$$

Now we also have, $A_1 \approx \operatorname{conv} A_1$ and $A_2 \approx \operatorname{conv} A_2$ by Lemma 6.8(ii). That and Fact 2.26 yield,

$$\operatorname{ri} A_1 \cap \operatorname{ri} A_2 = \operatorname{ri}\operatorname{conv} A_1 \cap \operatorname{ri}\operatorname{conv} A_2 = \operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2). \quad (6.11)$$

Combining (6.10) and (6.11) we get

$$\operatorname{ri}(A_1 \cap A_2) = \operatorname{ri}(\operatorname{conv} A_1 \cap \operatorname{conv} A_2) = \operatorname{ri} A_1 \cap \operatorname{ri} A_2,$$

which proves the result. □

Theorem 6.13. *Let A_i be a nearly convex set in $\mathbb{R}^n$ for $i = 1, \ldots, m$ such that $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$. Then $\bigcap_{i=1}^{m} A_i$ is nearly convex and*

$$\bigcap_{i=1}^{m} \operatorname{ri} A_i = \operatorname{ri} \bigcap_{i=1}^{m} A_i.$$

Proof. Clearly, when $m = 1$ this holds. When $m = 2$, by Lemma 6.12 we have $A_1 \cap A_2$ is nearly convex and

$$\operatorname{ri}(A_1 \cap A_2) = \operatorname{ri} A_1 \cap \operatorname{ri} A_2.$$

Thus we proceed via induction and assume that when $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$,

$$\bigcap_{i=1}^{m} \operatorname{ri} A_i = \operatorname{ri} \bigcap_{i=1}^{m} A_i \text{ and } \bigcap_{i=1}^{m} A_i \text{ is nearly convex.}$$

Then consider $\bigcap_{i=1}^{m+1} \operatorname{ri} A_i$ such that $\bigcap_{i=1}^{m+1} \operatorname{ri} A_i \neq \varnothing$. We have

$$\bigcap_{i=1}^{m+1} \operatorname{ri} A_i = \bigcap_{i=1}^{m} \operatorname{ri} A_i \cap \operatorname{ri} A_{m+1} \neq \varnothing. \tag{6.12}$$

Since $\bigcap_{i=1}^{m+1} \operatorname{ri} A_i \neq \varnothing$ we must have $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$. Thus, by the inductive hypothesis (6.12) becomes,

$$\bigcap_{i=1}^{m+1} \operatorname{ri} A_i = \operatorname{ri} \bigcap_{i=1}^{m} A_i \cap \operatorname{ri} A_{m+1} \neq \varnothing, \tag{6.13}$$

and since $\bigcap_{i=1}^{m} A_i$ is nearly convex we can apply Lemma 6.12 to the sets $\bigcap_{i=1}^{m} A_i$ and A_{m+1} to get $\bigcap_{i=1}^{m} A_i \cap A_{m+1} = \bigcap_{i=1}^{m+1} A_i$ is nearly convex and (6.13) becomes

$$\bigcap_{i=1}^{m+1} \operatorname{ri} A_i = \operatorname{ri} \left(\bigcap_{i=1}^{m} A_i \cap A_{m+1} \right) = \operatorname{ri} \bigcap_{i=1}^{m+1} A_i.$$

□

Lemma 6.14. *Let A_1 and A_2 be nearly convex sets such that* $\operatorname{ri} A_1 \cap \operatorname{ri} A_2 \neq \varnothing$. *Then*

$$\overline{A_1 \cap A_2} = \overline{A_1} \cap \overline{A_2}.$$

Proof. Clearly we always have

$$\overline{A_1 \cap A_2} \subseteq \overline{A_1} \cap \overline{A_2}.$$

On the other hand, by Fact 2.25(iii) and Fact 2.26,

$$\overline{A_1} \cap \overline{A_2} = \overline{\operatorname{ri} A_1} \cap \overline{\operatorname{ri} A_2} = \overline{\operatorname{ri} A_1 \cap \operatorname{ri} A_2} \subseteq \overline{A_1 \cap A_2}.$$

Thus $\overline{A_1 \cap A_2} = \overline{A_1} \cap \overline{A_2}$. □

Theorem 6.15. *Let A_i be a nearly convex set in $\mathbb{R}^n$ for $i = 1, \ldots, m$ such that $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$. Then*

$$\overline{\bigcap_{i=1}^{m} A_i} = \bigcap_{i=1}^{m} \overline{A_i}. \tag{6.14}$$

Proof. This clearly holds when $m = 1$. When $m = 2$, (6.14) holds by Lemma 6.14. Continuing via induction, we assume that when $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$, we have

$$\overline{\bigcap_{i=1}^{m} A_i} = \bigcap_{i=1}^{m} \overline{A_i}.$$

Then we consider $\bigcap_{i=1}^{m+1} \overline{A_i}$ such that $\bigcap_{i=1}^{m+1} \operatorname{ri} A_i \neq \varnothing$. We have

$$\bigcap_{i=1}^{m+1} \overline{A_i} = \bigcap_{i=1}^{m} \overline{A_i} \cap \overline{A_{m+1}}. \tag{6.15}$$

Since $\bigcap_{i=1}^{m+1} \operatorname{ri} A_i \neq \varnothing$, then clearly $\bigcap_{i=1}^{m} \operatorname{ri} A_i \neq \varnothing$. Thus, by the inductive hypothesis, (6.15) becomes

$$\bigcap_{i=1}^{m+1} \overline{A_i} = \overline{\bigcap_{i=1}^{m} A_i} \cap \overline{A_{m+1}}. \tag{6.16}$$

Now, by Theorem 6.13, $\bigcap_{i=1}^{m} A_i$ is nearly convex and $\operatorname{ri} \bigcap_{i=1}^{m} A_i = \bigcap_{i=1}^{m} \operatorname{ri} A_i$ so,

$$\operatorname{ri}\left(\bigcap_{i=1}^{m} A_i\right) \cap \operatorname{ri} A_{m+1} = \bigcap_{i=1}^{m} \operatorname{ri} A_i \cap \operatorname{ri} A_{m+1} = \bigcap_{i=1}^{m+1} \operatorname{ri} A_i \neq \varnothing.$$

Thus, apply Lemma 6.14 to the sets $\bigcap_{i=1}^{m} A_i$ and A_{m+1}, and (6.16) becomes,

$$\bigcap_{i=1}^{m+1} \overline{A_i} = \overline{\bigcap_{i=1}^{m} A_i \cap A_{m+1}} = \overline{\bigcap_{i=1}^{m+1} A_i},$$

which proves the desired result. $\square$

In order to study addition of nearly convex sets, we require the following result.

Lemma 6.16. *Let $(A_i)_{i\in I}$ be a family of nearly convex subsets of $\mathbb{R}^n$, and let $(\lambda_i)_{i\in I}$ be a family of real numbers. Then $\sum_{i\in I} \lambda_i A_i$ is nearly convex, and* $\operatorname{ri}(\sum_{i\in I} \lambda_i A_i) = \sum_{i\in I} \lambda_i \operatorname{ri} A_i$.

Proof. For every $i \in I$, there exists a convex subset C_i of $\mathbb{R}^n$ such that $C_i \subseteq A_i \subseteq \overline{C_i}$. We have

$$\sum_{i\in I} \lambda_i C_i \subseteq \sum_{i\in I} \lambda_i A_i \subseteq \sum_{i\in I} \lambda_i \overline{C_i} \subseteq \overline{\sum_{i\in I} \lambda_i C_i}, \tag{6.17}$$

which yields the near convexity of $\sum_{i\in I} \lambda_i A_i$ and $\sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i C_i$ by Lemma 6.6. Moreover, by Fact 2.25(vii)&(viii) and Lemma 6.6,

$$\operatorname{ri}\Big(\sum_{i\in I} \lambda_i A_i\Big) = \operatorname{ri}\Big(\sum_{i\in I} \lambda_i C_i\Big) = \sum_{i\in I} \operatorname{ri}\big(\lambda_i C_i\big) = \sum_{i\in I} \lambda_i \operatorname{ri} C_i = \sum_{i\in I} \lambda_i \operatorname{ri} A_i. \tag{6.18}$$

This completes the proof. □

Theorem 6.17. *Let $(A_i)_{i\in I}$ be a family of nearly convex subsets of $\mathbb{R}^n$, and let $(B_i)_{i\in I}$ be a family of subsets of $\mathbb{R}^n$ such that $A_i \approx B_i$, for every $i \in I$. Then $\sum_{i\in I} A_i$ is nearly convex and $\sum_{i\in I} A_i \approx \sum_{i\in I} B_i$.*

Proof. Lemma 6.8 implies that B_i is nearly convex, for every $i \in I$. By Lemma 6.16, we have that $\sum_{i\in I} A_i$ is nearly convex and

$$\operatorname{ri}\sum_{i\in I} A_i = \sum_{i\in I} \operatorname{ri} A_i = \sum_{i\in I} \operatorname{ri} B_i = \operatorname{ri}\sum_{i\in I} B_i. \tag{6.19}$$

Furthermore,

$$\overline{\sum_{i\in I} A_i} = \overline{\sum_{i\in I} \overline{A_i}} = \overline{\sum_{i\in I} \overline{B_i}} = \overline{\sum_{i\in I} B_i} \tag{6.20}$$

and the result follows. □

Remark 6.18. Theorem 6.17 fails without the near convexity assumption: indeed, consider $\mathbb{R}$ and $m = 2$, with $A_1 = A_2 = \mathbb{Q}$ and $B_1 = B_2 = \mathbb{R} \smallsetminus \mathbb{Q}$. Then $A_i \approx B_i$, for every $i \in I$, yet $A_1 + A_2 = \mathbb{Q} \not\approx \mathbb{R} = B_1 + B_2$.

Theorem 6.19. *Let $(A_i)_{i\in I}$ be a family of nearly convex subsets of $\mathbb{R}^n$, and let $(\lambda_i)_{i\in I}$ be a family of real numbers. For every $i \in I$, take $B_i \in \big\{A_i, \overline{A_i}, \operatorname{conv} A_i, \operatorname{ri} A_i, \operatorname{ri}\operatorname{conv} A_i\big\}$. Then*

$$\sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i B_i. \tag{6.21}$$

Proof. By Lemma 6.6, $A_i \approx B_i$ for every $i \in I$. Now apply Theorem 6.17. □

Corollary 6.20. *Let $(A_i)_{i\in I}$ be a family of nearly convex subsets of $\mathbb{R}^n$, and let $(\lambda_i)_{i\in I}$ be a family of real numbers. Suppose that $j \in I$ is such that $\lambda_j \neq 0$. Then*

$$\big(\operatorname{int}\lambda_j A_j\big) + \sum_{i\in I\smallsetminus\{j\}} \lambda_i \overline{A_i} \subseteq \operatorname{int}\sum_{i\in I}\lambda_i A_i; \tag{6.22}$$

consequently, if $0 \in (\operatorname{int} A_j) \cap \bigcap_{i\in I\smallsetminus\{j\}} \overline{A_i}$, *then* $0 \in \operatorname{int}\sum_{i\in I}\lambda_i A_i$.

Proof. By Theorem 6.19, $\operatorname{ri}(\lambda_j A_j + \sum_{i\in I\smallsetminus\{j\}} \lambda_i\overline{A_i}) = \operatorname{ri}\sum_{i\in I}\lambda_i A_i$. Since

$$\big(\operatorname{int}\lambda_j A_j\big) + \sum_{i\in I\smallsetminus\{j\}} \lambda_i \overline{A_i} \subseteq \operatorname{ri}\Big(\lambda_j A_j + \sum_{i\in I\smallsetminus\{j\}} \lambda_i\overline{A_i}\Big), \tag{6.23}$$

and $\big(\operatorname{int}\lambda_j A_j\big) + \sum_{i\in I\smallsetminus\{j\}} \lambda_i\overline{A_i}$ is an open set, (6.22) follows. In turn, the "consequently" follows from (6.22). ☐

We develop a complementary cancelation result whose proof relies on Rådström's cancelation:

Fact 6.21. (See [55].) *Let A be a nonempty subset of $\mathbb{R}^n$, let E be a nonempty bounded subset of $\mathbb{R}^n$, and let B be a nonempty closed convex subset of $\mathbb{R}^n$ such that $A + E \subseteq B + E$. Then $A \subseteq B$.*

Theorem 6.22. *Let A and B be nonempty nearly convex subsets of $\mathbb{R}^n$, and let E be a nonempty compact subset of $\mathbb{R}^n$ such that $A + E \approx B + E$. Then $A \approx B$.*

Proof. We have $A + E \subseteq \overline{A+E} = \overline{B+E} = \overline{B} + E$. Fact 6.21 implies $A \subseteq \overline{B}$; hence, $\overline{A} \subseteq \overline{B}$. Analogously, $\overline{B} \subseteq \overline{A}$ and thus $\overline{A} = \overline{B}$. Now apply Proposition 6.11. ☐

Finally, we give a result concerning the interior of nearly convex sets.

Proposition 6.23. *Let A be a nearly convex subset of $\mathbb{R}^n$. Then* $\operatorname{int} A = \operatorname{int}\operatorname{conv} A = \operatorname{int}\overline{A}$.

Proof. By Lemma 6.6, $A \approx B$, where $B \in \big\{\overline{A}, \operatorname{conv} A\big\}$. Now recall (6.2). ☐

6.2 Maximally monotone operators

Fact 6.24 (Minty). [59, Theorem 12.41] *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone. Then* $\operatorname{dom} A$ *and* $\operatorname{ran} A$ *are nearly convex.*

Theorem 6.25. *Let A and B be monotone on $\mathbb{R}^n$ such that $A+B$ is maximally monotone. Suppose that one of the following holds.*

(i) *A and B are rectangular.*

(ii) $\operatorname{dom} A \subseteq \operatorname{dom} B$ *and B is rectangular.*

Then $\operatorname{ran}(A+B)$ *is nearly convex, and* $\operatorname{ran}(A+B) \approx \operatorname{ran} A + \operatorname{ran} B$.

Proof. The near convexity of $\operatorname{ran}(A+B)$ follows from Fact 6.24. Using Fact 3.66 and Fact 2.25(iii),

$$\begin{aligned} \operatorname{ri}\operatorname{conv}(\operatorname{ran} A + \operatorname{ran} B) &\subseteq \operatorname{ran}(A+B) \\ &\subseteq \operatorname{ran} A + \operatorname{ran} B \\ &\subseteq \overline{\operatorname{conv}(\operatorname{ran} A + \operatorname{ran} B)} \\ &= \overline{\operatorname{ri}\operatorname{conv}(\operatorname{ran} A + \operatorname{ran} B)}. \end{aligned}$$

Proposition 6.4 and Lemma 6.6 imply $\operatorname{ran}(A+B) \approx \operatorname{ran} A + \operatorname{ran} B \approx \operatorname{ri}\operatorname{conv}(\operatorname{ran} A + \operatorname{ran} B)$. ☐

Remark 6.26. Considering $A+0$, where A is the rotator by $\pi/2$ on $\mathbb{R}^2$ which is not rectangular, we see that $A+B$ need not be rectangular under assumption (ii) in Theorem 6.25.

If we let $S_i = \operatorname{ran} A_i$ and $\lambda_i = 1$ for every $i \in I$ in Theorem 6.27, then we obtain a result that is related to Pennanen's [53, Corollary 6].

Theorem 6.27. *Let $(A_i)_{i\in I}$ be a family of maximally monotone rectangular operators on $\mathbb{R}^n$ with $\bigcap_{i\in I} \operatorname{ri}\operatorname{dom} A_i \neq \varnothing$, let $(S_i)_{i\in I}$ be a family of subsets of $\mathbb{R}^n$ such that*

$$(\forall i \in I) \quad S_i \in \big\{ \operatorname{ran} A_i, \overline{\operatorname{ran} A_i}, \operatorname{ri}(\operatorname{ran} A_i)\big\}, \tag{6.24}$$

and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers. Then $\sum_{i\in I} \lambda_i A_i$ is maximally monotone, rectangular, and $\operatorname{ran} \sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i S_i$ is nearly convex.

Proof. We have $I = \{1, \ldots, m\}$. To see that $\sum_{i\in I} \lambda_i A_i$ is maximally monotone we proceed using induction on m. When $m = 1$, since $\lambda_i \in \mathbb{R}_{++}$ and A_i is maximally monotone, then $\lambda_i A_i$ is maximally monotone by Proposition 3.29. When $m = 2$, by assumption we have

$$\operatorname{ri}\operatorname{dom} A_1 \cap \operatorname{ri}\operatorname{dom} A_2 \neq \varnothing \Rightarrow \operatorname{ri}\operatorname{dom} \lambda_1 A_1 \cap \operatorname{ri}\operatorname{dom} \lambda_2 A_2 \neq \varnothing.$$

so by Fact 3.48(ii) $\lambda_1 A_1 + \lambda_2 A_2$ is maximally monotone. Now assume this holds for $\lambda_1 A_1 + \ldots + \lambda_m A_m$ with $\bigcap\limits_{i=1}^{m} \operatorname{ri}\operatorname{dom} A_i \neq \varnothing$. Then consider

$$\lambda_1 A_1 + \cdots + \lambda_{m+1} A_{m+1} = (\lambda_1 A_1 + \cdots + \lambda_m A_m) + \lambda_{m+1} A_{m+1}.$$

By the inductive hypothesis, $\lambda_1 A_1 + \ldots + \lambda_m A_m$ is maximally monotone. We have

$$\begin{aligned}
&\operatorname{ri}\operatorname{dom}(\lambda_1 A_1 + \cdots + \lambda_m A_m) \cap \operatorname{ri}\operatorname{dom} \lambda_{m+1} A_{m+1} \\
&\qquad = \operatorname{ri}\left(\bigcap_{i=1}^{m} \operatorname{dom} \lambda_i A_i\right) \cap \operatorname{ri}\operatorname{dom} \lambda_{m+1} A_{m+1}. \qquad (6.25)
\end{aligned}$$

By Fact 6.24, $\operatorname{dom} \lambda_i A_i$ is nearly convex for all $i \in I$, so apply Lemma 6.13 and (6.25) becomes

$$\begin{aligned}
&\operatorname{ri}\operatorname{dom}(\lambda_1 A_1 + \cdots + \lambda_m A_m) \cap \operatorname{ri}\operatorname{dom} \lambda_{m+1} A_{m+1} \\
&\qquad = \left(\bigcap_{i=1}^{m} \operatorname{ri}\operatorname{dom} \lambda_i A_i\right) \cap \operatorname{ri}\operatorname{dom} \lambda_{m+1} A_{m+1} = \bigcap_{i=1}^{m+1} \operatorname{ri}\operatorname{dom} \lambda_i A_i \neq \varnothing.
\end{aligned}$$

Thus by Fact 3.48(ii), $\lambda_1 A_1 + \cdots + \lambda_{m+1} A_{m+1}$ is maximally monotone. Using Lemma 3.63 and induction we have $\lambda_1 A_1 + \cdots + \lambda_{m+1} A_{m+1}$ is rectangular.

With Theorem 6.17, Fact 3.48 and Lemma 3.63 in mind, Theorem 6.25(i) and induction yields $\operatorname{ran} \sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i \operatorname{ran} A_i$ and the near convexity. Finally, as $\operatorname{ran} A_i$ is nearly convex for every $i \in I$ by Fact 6.24, $\operatorname{ran} \sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i S_i$ follows from Theorem 6.19. ☐

The main result of this section is the following.

Theorem 6.28. *Let $(A_i)_{i\in I}$ be a family of maximally monotone rectangular operators on $\mathbb{R}^n$ such that $\bigcap_{i\in I} \operatorname{ri}\operatorname{dom} A_i \neq \varnothing$, let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers, and let $j \in I$. Set*

$$A = \sum_{i\in I} \lambda_i A_i. \qquad (6.26)$$

Then the following hold.

(i) *If* $\sum_{i\in I}\lambda_i \operatorname{ran} A_i = \mathbb{R}^n$, *then* $\operatorname{ran} A = \mathbb{R}^n$.

(ii) *If* A_j *is surjective, then* A *is surjective.*

(iii) *If* $0 \in \bigcap_{i\in I}\overline{\operatorname{ran} A_i}$, *then* $0 \in \overline{\operatorname{ran} A}$.

(iv) *If* $0 \in (\operatorname{int}\operatorname{ran} A_j) \cap \bigcap_{i\in I\smallsetminus\{j\}}\overline{\operatorname{ran} A_i}$, *then* $0 \in \operatorname{int}\operatorname{ran} A$.

Proof. Theorem 6.27 implies that $\operatorname{ran}\sum_{i\in I}\lambda_i A_i \approx \sum_{i\in I}\lambda_i \operatorname{ran} A_i$ is nearly convex. Hence

$$\operatorname{ri}\operatorname{ran} A = \operatorname{ri}\operatorname{ran}\sum_{i\in I}\lambda_i A_i = \operatorname{ri}\Big(\sum_{i\in I}\lambda_i \operatorname{ran} A_i\Big) = \sum_{i\in I}\lambda_i \operatorname{ri}\operatorname{ran} A_i \tag{6.27}$$

and

$$\overline{\operatorname{ran} A} = \overline{\operatorname{ran}\sum_{i\in I}\lambda_i A_i} = \overline{\sum_{i\in I}\lambda_i \operatorname{ran} A_i}\,. \tag{6.28}$$

(i): Indeed, using (6.27),

$$\mathbb{R}^n = \operatorname{ri}\mathbb{R}^n = \operatorname{ri}\sum_{i\in I}\lambda_i \operatorname{ran} A_i = \operatorname{ri}\operatorname{ran} A \subseteq \operatorname{ran} A \subseteq \mathbb{R}^n.$$

(ii): Clear from (i). (iii): It follows from (6.28) that

$$0 \in \sum_{i\in I}\lambda_i\overline{\operatorname{ran} A_i} \subseteq \overline{\sum_{i\in I}\lambda_i \operatorname{ran} A_i} = \overline{\operatorname{ran} A}.$$

(iv): By Fact 6.24, $\operatorname{ran} A_i$ is nearly convex for every $i \in I$. Thus, $0 \in \operatorname{int}\sum_{i\in I}\lambda_i \operatorname{ran} A_i$ by Corollary 6.20. On the other hand, (6.27) implies that

$$\operatorname{int}\sum_{i\in I}\lambda_i \operatorname{ran} A_i \subseteq \operatorname{ri}\sum_{i\in I}\lambda_i \operatorname{ran} A_i = \operatorname{ri}\operatorname{ran} A.$$

Altogether, $0 \in \operatorname{ri}\operatorname{ran} A = \operatorname{int}\operatorname{ran} A$ because $\operatorname{int}\operatorname{ran} A \neq \varnothing$. ☐

6.3 Firmly nonexpansive mappings

In this section, we apply the result of Section 6.2 to firmly nonexpansive mappings.

Corollary 6.29. *Let* $T\colon \mathbb{R}^n \to \mathbb{R}^n$ *be firmly nonexpansive. Then* T *is maximally monotone and rectangular, and* $\operatorname{ran} T$ *is nearly convex.*

Proof. Combine Example 3.61, Fact 3.36(i), and Fact 6.24. ☐

It is also known that the class of firmly nonexpansive mappings is closed under taking convex combinations. For completeness, we include a short proof of this result.

Lemma 6.30. *Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on $\mathbb{R}^n$, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Then $\sum_{i\in I}\lambda_i T_i$ is also firmly nonexpansive.*

Proof. Set $T = \sum_{i\in I}\lambda_i T_i$. By Fact 3.3, $2T_i - \mathrm{Id}$ is nonexpansive for every $i \in I$, so $2T - \mathrm{Id} = \sum_{i\in I}\lambda_i(2T_i - \mathrm{Id})$ is also nonexpansive. Applying Fact 3.3 once more, we deduce that T is firmly nonexpansive. ☐

We are now ready for the first main result of this section.

Theorem 6.31 (averages of firmly nonexpansive mappings)**.** *Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on $\mathbb{R}^n$, let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$, and let $j \in I$. Set $T = \sum_{i\in I}\lambda_i T_i$. Then the following hold.*

(i) *T is firmly nonexpansive and $\operatorname{ran} T \approx \sum_{i\in I}\lambda_i \operatorname{ran} T_i$ is nearly convex.*

(ii) *If T_j is surjective, then T is surjective.*

(iii) *If $0 \in \bigcap_{i\in I}\overline{\operatorname{ran} T_i}$, then $0 \in \overline{\operatorname{ran} T}$.*

(iv) *If $0 \in (\operatorname{int}\operatorname{ran} T_j) \cap \bigcap_{i\in I\smallsetminus\{j\}}\overline{\operatorname{ran} T_i}$, then $0 \in \operatorname{int}\operatorname{ran} T$.*

Proof. By Corollary 6.29, each T_i is maximally monotone, rectangular and $\operatorname{ran} T_i$ is nearly convex. (i): Lemma 6.6, Lemma 6.30, and Theorem 6.27. (ii): Theorem 6.28(ii). (iii): Theorem 6.28(iii). (iv): Theorem 6.28(iv). ☐

The following averaged-projection operator plays a role in methods for solving (potentially inconsistent) convex feasibility problems because its fixed point set consists of least-squares solutions; see, e.g., [7, Section 6], [23] and [35] for further information.

Example 6.32. Let $(C_i)_{i\in I}$ be a family of nonempty closed convex subsets of $\mathbb{R}^n$ with associated projection operators P_i, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Then

$$\operatorname{ran}\sum_{i\in I}\lambda_i P_i \approx \sum_{i\in I}\lambda_i C_i. \tag{6.29}$$

Proof. This follows from Theorem 6.31(i) since $(\forall i \in I)\ \operatorname{ran} P_i = C_i$. ☐

Remark 6.33. Let C_1 and C_2 be nonempty closed convex subsets of $\mathbb{R}^n$ with associated projection operators P_1 and P_2 respectively, and—instead of averaging as in Example 6.32—consider the composition $T = P_2 \circ P_1$, which is still *nonexpansive*. It is obvious that $\operatorname{ran} T \subseteq \operatorname{ran} P_2 = C_2$, but $\operatorname{ran} T$ need not be even nearly convex: indeed, in $\mathbb{R}^2$, let C_2 be the unit ball centered at 0 of radius 1, and let $C_1 = \mathbb{R} \times \{2\}$. Then $\operatorname{ran} T$ is the intersection of the open upper halfplane and the boundary of C_2, which is very far from being nearly convex. Thus the near convexity part of Corollary 6.29 has no counterpart for nonexpansive mappings.

Remark 6.34. Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be firmly nonexpansive. Recall the set of *fixed points* is denoted by

$$\operatorname{Fix} T = \big\{x \in \mathbb{R}^n \bigm| x = Tx\big\}, \tag{6.30}$$

and that T is *asymptotically regular* if there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^n$ such that $x_n - Tx_n \to 0$.

If the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a point, say $\bar{x}$, then continuity of T implies that $\bar{x} \in \operatorname{Fix} T$.

The next result is a consequence of fundamental work by Baillon, Bruck and Reich [3] .

Theorem 6.35. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be firmly nonexpansive. Then T is asymptotically regular if and only if for every $x_0 \in \mathbb{R}^n$, the sequence defined by*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n \tag{6.31}$$

satisfies $x_n - x_{n+1} \to 0$. Moreover, if $\operatorname{Fix} T \neq \varnothing$, then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a fixed point; otherwise, $\|x_n\| \to +\infty$.

Proof. T is firmly nonexpansive $\Leftrightarrow$ T is $\frac{1}{2}$-averaged, so by Fact 3.20, $T^n x - T^{n+1}x \to v$ where v is the element of minimum norm in $\overline{\operatorname{ran}(\operatorname{Id} - T)}$. Since T is asymptotically regular, $v = 0$ and thus $x_n - x_{n+1} \to 0$. By Fact 3.18, if $\operatorname{Fix} T \neq \varnothing$, then $(x_n)_{n\in\mathbb{N}} \rightharpoonup x \in \operatorname{Fix} T$. And by Fact 3.19 if $\operatorname{Fix} T = \varnothing$, then $\|x_n\| \to +\infty$. ☐

Here is the second main result of this chapter.

Theorem 6.36 (asymptotic regularity of the average)**.** *Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on $\mathbb{R}^n$, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I} \lambda_i = 1$. Suppose that T_i is asymptotically regular, for every $i \in I$. Then $\sum_{i\in I} \lambda_i T_i$ is also asymptotically regular.*

Proof. Set $T = \sum_{i\in I} \lambda_i T_i$. Then

$$\operatorname{Id} - T = \sum_{i\in I} \lambda_i (\operatorname{Id} - T_i).$$

Since each $\operatorname{Id} - T_i$ is firmly nonexpansive and $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$ by the asymptotic regularity of T_i, the conclusion follows from Theorem 6.31(iii). $\square$

Remark 6.37. Consider Theorem 6.36. Even when $\operatorname{Fix} T_i \neq \varnothing$, for every $i \in I$, it is impossible to improve the conclusion to $\operatorname{Fix} \sum_{i\in I} \lambda_i T_i \neq \varnothing$. Indeed, in $\mathbb{R}^2$, set $C_1 = \mathbb{R} \times \{0\}$ and $C_2 = \operatorname{epi} \exp$. Set $T = \frac{1}{2} P_{C_1} + \frac{1}{2} P_{C_2}$. Then $\operatorname{Fix} T_1 = C_1$ and $\operatorname{Fix} T_2 = C_2$, yet $\operatorname{Fix} T = \varnothing$.

The proof of the following useful result is straightforward and hence omitted.

Lemma 6.38. *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone. Then J_A is asymptotically regular if and only if $0 \in \overline{\operatorname{ran} A}$.*

We conclude this chapter with an application to the resolvent average of monotone operators.

Corollary 6.39 (resolvent average)**.** *Let $(A_i)_{i\in I}$ be a family of maximally monotone operators on $\mathbb{R}^n$, let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I} \lambda_i = 1$, let $j \in I$, and set*

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \left(\sum_{i\in I} \lambda_i (\operatorname{Id} + A_i)^{-1} \right)^{-1} - \operatorname{Id}. \tag{6.32}$$

Then the following hold.

(i) $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ *is maximally monotone.*

(ii) $\operatorname{dom} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) \approx \sum_{i\in I} \lambda_i \operatorname{dom} A_i$.

(iii) $\operatorname{ran} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) \approx \sum_{i\in I} \lambda_i \operatorname{ran} A_i$.

(iv) *If* $0 \in \bigcap_{i\in I} \overline{\operatorname{ran} A_i}$, *then* $0 \in \overline{\operatorname{ran} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})}$.

(v) *If* $0 \in \operatorname{int} \operatorname{ran} A_j \cap \bigcap_{i\in I\setminus\{j\}} \overline{\operatorname{ran} A_i}$, *then* $0 \in \operatorname{int} \operatorname{ran} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$.

(vi) *If* $\operatorname{dom} A_j = \mathbb{R}^n$, *then* $\operatorname{dom} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathbb{R}^n$.

(vii) *If* $\operatorname{ran} A_j = \mathbb{R}^n$, *then* $\operatorname{ran} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathbb{R}^n$.

Proof. Observe that

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum_{i\in I} \lambda_i J_{A_i} \tag{6.33}$$

and

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}} = \sum_{i\in I} \lambda_i J_{A_i^{-1}} \tag{6.34}$$

by using (3.13). Furthermore, using (3.14), we note that

$$\operatorname{ran} J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \operatorname{dom} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \quad \text{and} \quad \operatorname{ran} J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}} = \operatorname{ran} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}). \tag{6.35}$$

(i): This follows from (6.33) and Fact 3.36. (ii): Apply Theorem 6.31(i) to $(J_{A_i})_{i\in I}$, and use (6.33) and (6.35). (iii): Apply Theorem 6.31(i) to $(\operatorname{Id} - J_{A_i})_{i\in I}$, and use (3.13) and (6.35). (iv): Combine Theorem 6.36 and Lemma 6.38, and use (6.33). (v): Apply Theorem 6.31(iv) to (6.34), and use (6.35). (vi) and (vii): These follow from (ii) and (iii), respectively. □

Remark 6.40 (proximal average). In Corollary 6.39, one may also start from a family $(f_i)_{i\in I}$ of functions on $\mathbb{R}^n$ that are convex, lower semi-continuous, and proper, and with corresponding subdifferential operators $(A_i)_{i\in I} = (\partial f_i)_{i\in I}$. This relates to the *proximal average*, $\mathcal{P}$, of the family $(f_i)_{i\in I}$, where $\partial\mathcal{P}$ is the resolvent average of the family $(\partial f_i)_{i\in I}$. See [12] for further information and references. Corollary 6.39(vii) essentially states that $\mathcal{P}$ is *supercoercive* provided that some f_j is. Analogously, Corollary 6.39(v) shows that that *coercivity* of $\mathcal{P}$ follows from the coercivity of some function f_j.

Chapter 7

Compositions and Convex Combinations of Asymptotically Regular Firmly Nonexpansive Mappings

In this chapter, based on [15], we extend some of the results in Chapter 6 into a Hilbert space setting. Even though the main results are formulated in the given Hilbert space $\mathcal{H}$, it will turn out that the key space to work in is the product space,

$$\mathcal{H}^m = \big\{\mathbf{x} = (x_i)_{i\in I} \;\big|\; (\forall i \in I)\; x_i \in \mathcal{H}\big\}, \tag{7.1}$$

where $m \in \{2, 3, 4, \ldots\}$ and $I = \{1, 2, \ldots, m\}$. This product space contains an embedding of the original space $\mathcal{H}$ via the *diagonal* subspace

$$\mathbf{\Delta} = \big\{\mathbf{x} = (x)_{i\in I} \;\big|\; x \in \mathcal{H}\big\}. \tag{7.2}$$

We also assume that we are given m firmly nonexpansive operators $T_1, \ldots, T_m$; equivalently, m resolvents of maximally monotone operators $A_1, \ldots, A_m$. We now define various pertinent operators acting on $\mathcal{H}^m$. We start with the Cartesian product operators

$$\mathbf{T}\colon \mathcal{H}^m \to \mathcal{H}^m\colon (x_i)_{i\in I} \mapsto (T_i x_i)_{i\in I} \tag{7.3}$$

and

$$\mathbf{A}\colon \mathcal{H}^m \rightrightarrows \mathcal{H}^m\colon (x_i)_{i\in I} \mapsto (A_i x_i)_{i\in I}. \tag{7.4}$$

Denoting the identity on $\mathcal{H}^m$ by $\mathbf{Id}$, we observe that

$$J_{\mathbf{A}} = (\mathbf{Id} + \mathbf{A})^{-1} = T_1 \times \cdots \times T_m = \mathbf{T}. \tag{7.5}$$

Of central importance will be the *cyclic right-shift operator*

$$\mathbf{R}\colon \mathcal{H}^m \to \mathcal{H}^m\colon (x_1, x_2, \dots, x_m) \mapsto (x_m, x_1, \dots, x_{m-1}) \tag{7.6}$$

and for convenience we set

$$\mathbf{M} = \mathbf{Id} - \mathbf{R}. \tag{7.7}$$

We also fix strictly positive *convex coefficients* (or weights) $(\lambda_i)_{i\in I}$, i.e.,

$$(\forall i \in I) \quad \lambda_i \in \left]0,1\right[\quad \text{and} \quad \sum_{i\in I} \lambda_i = 1. \tag{7.8}$$

Let us make $\mathcal{H}^m$ into the Hilbert product space

$$\boldsymbol{\mathcal{H}} = \mathcal{H}^m, \quad \text{with} \quad \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i\in I} \langle x_i, y_i \rangle\,. \tag{7.9}$$

Fact 7.1. [11, Proposition 25.4(i)] *Set* $\boldsymbol{\Delta} = \big\{\mathbf{x} = (x)_{i\in I} \;\big|\; x \in \mathcal{H}\big\}$*. The orthogonal complement of* $\boldsymbol{\Delta}$ *with respect to this standard inner product is*

$$\boldsymbol{\Delta}^\perp = \Big\{\mathbf{x} = (x_i)_{i\in I} \;\Big|\; \sum_{i\in I} x_i = 0\Big\}. \tag{7.10}$$

7.1 Properties of the operator M

In this section, we collect several useful properties of the operator $\mathbf{M}$, including its Moore-Penrose inverse. To that end, the following result will be useful.

Proposition 7.2. *Let* Y *be a real Hilbert space and let* B *be a continuous linear operator from* $\mathcal{H}$ *to* Y *with adjoint* B^* *and such that* $\operatorname{ran} B$ *is closed. Then the Moore-Penrose inverse of* B *satisfies*

$$B^\dagger = P_{\operatorname{ran} B^*} \circ B^{-1} \circ P_{\operatorname{ran} B}. \tag{7.11}$$

Proof. Take $y \in Y$. Define the corresponding set of least squares solutions (see Fact 2.35) by $C = B^{-1}(P_{\operatorname{ran} B} y)$. By Fact 2.13, since $\operatorname{ran} B$ is closed, so is $\operatorname{ran} B^*$; hence, by Fact 2.9 and setting $U = (\ker B)^\perp$ we have

$$U = (\ker B)^\perp = \overline{\operatorname{ran} B^*} = \operatorname{ran} B^*.$$

Thus,

$$C = B^\dagger y + \ker B = B^\dagger y + U^\perp.$$

Therefore, since by Fact 2.37 $\operatorname{ran} B^\dagger = \operatorname{ran} B^*$,

$$P_U(C) = P_U B^\dagger y = B^\dagger y,$$

as claimed. □

Theorem 7.3. *Define*

$$\mathbf{L}\colon \boldsymbol{\Delta}^\perp \to \boldsymbol{\mathcal{H}}\colon \mathbf{y} \mapsto \sum_{i=1}^{m-1} \frac{m-i}{m}\mathbf{R}^{i-1}\mathbf{y}. \tag{7.12}$$

Then the following hold.

(i) $\mathbf{M}$ *is continuous, linear, and maximally monotone with* $\operatorname{dom}\mathbf{M} = \boldsymbol{\mathcal{H}}$*.*

(ii) $\mathbf{M}$ *is rectangular.*

(iii) $\ker\mathbf{M} = \ker\mathbf{M}^* = \boldsymbol{\Delta}$*.*

(iv) $\operatorname{ran}\mathbf{M} = \operatorname{ran}\mathbf{M}^* = \boldsymbol{\Delta}^\perp$ *is closed.*

(v) $\operatorname{ran}\mathbf{L} = \boldsymbol{\Delta}^\perp$*.*

(vi) $\mathbf{M}\circ\mathbf{L} = \mathbf{Id}|_{\boldsymbol{\Delta}^\perp}$*.*

(vii) $\mathbf{M}^{-1}\colon \boldsymbol{\mathcal{H}} \rightrightarrows \boldsymbol{\mathcal{H}}\colon \mathbf{y} \mapsto \begin{cases} \mathbf{L}\mathbf{y} + \boldsymbol{\Delta}, & \text{if } \mathbf{y} \in \boldsymbol{\Delta}^\perp; \\ \varnothing, & \text{otherwise.} \end{cases}$

(viii) $\mathbf{M}^\dagger = P_{\boldsymbol{\Delta}^\perp}\circ\mathbf{L}\circ P_{\boldsymbol{\Delta}^\perp} = \mathbf{L}\circ P_{\boldsymbol{\Delta}^\perp}$*.*

(ix) $\mathbf{M}^\dagger = \displaystyle\sum_{k=1}^{m} \frac{m-(2k-1)}{2m}\mathbf{R}^{k-1}$*.*

Proof. (i): Clearly, $\operatorname{dom}\mathbf{M} = \boldsymbol{\mathcal{H}}$ and $(\forall \mathbf{x} \in \boldsymbol{\mathcal{H}})$ $\|\mathbf{R}\mathbf{x}\| = \|\mathbf{x}\|$. Thus, $\mathbf{R}$ is nonexpansive and therefore by Fact 3.46, $\mathbf{M} = \mathbf{Id} - \mathbf{R}$ is maximally monotone.

(ii): This is Fact 3.65.

(iii): The definitions of $\mathbf{M}$ and $\mathbf{R}$ and the fact that $\mathbf{R}^*$ is the cyclic left shift operator (see Example 2.8) readily imply that

$$\begin{aligned}\ker\mathbf{M} &= \{\mathbf{x} \in \boldsymbol{\mathcal{H}} \mid \mathbf{M}\mathbf{x} = 0\} \\ &= \{\mathbf{x} \in \boldsymbol{\mathcal{H}} \mid \mathbf{Id} - \mathbf{R} = 0\},\end{aligned}$$

which yields $x_1 = x_m$ and $x_i = x_{i+1}$ for all $i \in I$. That is, $\ker \mathbf{M} = \boldsymbol{\Delta}$. Similarly, $\ker \mathbf{M}^* = \boldsymbol{\Delta}$ and thus,

$$\ker \mathbf{M} = \ker \mathbf{M}^* = \boldsymbol{\Delta}.$$

(iv), (vi), and (vii): Let $\mathbf{y} = (y_1, \ldots, y_m) \in \boldsymbol{\mathcal{H}}$. Assume first that $\mathbf{y} \in \operatorname{ran} \mathbf{M}$. Then there exists $\mathbf{x} = (x_1, \ldots, x_m)$ such that $y_1 = x_1 - x_m$, $y_2 = x_2 - x_1$, $\ldots$, and $y_m = x_m - x_{m-1}$. It follows that $\sum_{i \in I} y_i = 0$, i.e., $\mathbf{y} \in \Delta^{\perp}$ by Fact 7.1. Thus,

$$\operatorname{ran} \mathbf{M} \subseteq \boldsymbol{\Delta}^{\perp}. \tag{7.13}$$

Conversely, assume now that $\mathbf{y} \in \boldsymbol{\Delta}^{\perp}$. Now set

$$\mathbf{x} := \mathbf{L}\mathbf{y} = \sum_{i=1}^{m-1} \frac{m-i}{m} \mathbf{R}^{i-1} \mathbf{y}. \tag{7.14}$$

Then

$$\begin{aligned}
\mathbf{x} &= \frac{m-1}{m}\mathbf{R}^0\mathbf{y} + \frac{m-2}{m}\mathbf{R}y + \frac{m-3}{m}\mathbf{R}^2 y + \cdots + \frac{1}{m}\mathbf{R}^{m-2} y \\
&= \frac{m-1}{m}(y_1, \ldots, y_m) + \frac{m-2}{m}(y_m, y_1, \ldots, y_{m-1}) + \cdots + \frac{1}{m}(y_3, \ldots, y_1, y_2).
\end{aligned}$$

It will be notationally convenient to wrap indices around, i.e., $y_{m+1} = y_1$, $y_0 = y_m$ and likewise. We then get

$$(\forall i \in I) \quad x_i = \frac{m-1}{m} y_i + \frac{m-2}{m} y_{i-1} + \cdots + \frac{1}{m} y_{i+2}. \tag{7.15}$$

Therefore,

$$\begin{aligned}
\sum_{i \in I} x_i &= \frac{m-1}{m} \sum_{i \in I} y_i + \frac{m-2}{m} \sum_{i \in I} y_i + \cdots + \frac{1}{m} \sum_{i \in I} y_i \\
&= \frac{m(m-1) - \sum\limits_{i}^{m-1} i}{m} \sum_{i \in I} y_i \\
&= \frac{m(m-1) - \frac{m(m-1)}{2}}{m} \sum_{i \in I} y_i \\
&= \frac{m-1}{2} \sum_{i \in I} y_i = 0.
\end{aligned}$$

Thus $\mathbf{x} \in \boldsymbol{\Delta}^{\perp}$ and

$$\operatorname{ran} \mathbf{L} \subseteq \Delta^{\perp}. \tag{7.16}$$

Furthermore, $(\forall i \in I)$

$$\begin{aligned}
x_i - x_{i-1} &= \left(\frac{m-1}{m} y_i + \frac{m-2}{m} y_{i-1} + \cdots + \frac{1}{m} y_{i+2}\right) \\
&\quad - \left(\frac{m-1}{m} y_{i-1} + \cdots + \frac{1}{m} y_{(i-1)+2}\right) \\
&= \frac{m-1}{m} y_i + \left(\frac{(m-2)-(m-1)}{m}\right) y_{i-1} + \cdots \\
&\quad + \left(\frac{(m-(m-2))-1}{m}\right) y_{i+2} - \frac{1}{m} y_{i+1} \\
&= \frac{m-1}{m} y_i - \frac{1}{m} y_{i-1} - \frac{1}{m} y_{i-2} - \cdots - \frac{1}{m} y_{i+1} \\
&= y_i - \frac{1}{m} \sum_{j \in I} y_j = y_i.
\end{aligned}$$

Hence $\mathbf{M}\mathbf{x} = \mathbf{x} - \mathbf{R}\mathbf{x} = \mathbf{y}$ and thus $\mathbf{y} \in \operatorname{ran} \mathbf{M}$. Moreover, in view of (iii),

$$\mathbf{M}^{-1}\mathbf{y} = \mathbf{x} + \ker \mathbf{M} = \mathbf{x} + \boldsymbol{\Delta}. \tag{7.17}$$

We thus have shown

$$\boldsymbol{\Delta}^{\perp} \subseteq \operatorname{ran} \mathbf{M}. \tag{7.18}$$

Combining (7.13) and (7.18), we obtain $\operatorname{ran} \mathbf{M} = \boldsymbol{\Delta}^{\perp}$. We thus have verified (vi), and (vii). Since $\operatorname{ran} \mathbf{M}$ is closed, so is $\operatorname{ran} \mathbf{M}^*$, by Fact 2.13. Thus (iv) holds.

(viii)&(v): We have seen in Proposition 7.2 that

$$\mathbf{M}^{\dagger} = P_{\operatorname{ran} \mathbf{M}^*} \circ \mathbf{M}^{-1} \circ P_{\operatorname{ran} \mathbf{M}}. \tag{7.19}$$

Now let $\mathbf{z} \in \boldsymbol{\mathcal{H}}$. Then, by (iv),

$$\mathbf{y} := P_{\operatorname{ran} \mathbf{M}} \mathbf{z} = P_{\boldsymbol{\Delta}^{\perp}} \mathbf{z} \in \boldsymbol{\Delta}^{\perp}.$$

By (vii), $\mathbf{M}^{-1}\mathbf{y} = \mathbf{L}\mathbf{y} + \boldsymbol{\Delta}$. So,

$$\begin{aligned}
\mathbf{M}^{\dagger}\mathbf{z} &= P_{\operatorname{ran} \mathbf{M}^*} \mathbf{M}^{-1} P_{\operatorname{ran} \mathbf{M}} \mathbf{z} = P_{\operatorname{ran} \mathbf{M}^*} \mathbf{M}^{-1} \mathbf{y} \\
&= P_{\boldsymbol{\Delta}^{\perp}}(\mathbf{L}\mathbf{y} + \boldsymbol{\Delta}) = P_{\boldsymbol{\Delta}^{\perp}} \mathbf{L}\mathbf{y} = \mathbf{L}\mathbf{y} \\
&= (\mathbf{L} \circ P_{\boldsymbol{\Delta}^{\perp}})\mathbf{z},
\end{aligned}$$

because $\operatorname{ran}\mathbf{L} \subseteq \mathbf{\Delta}^{\perp}$ by (7.16). Hence (viii) holds. Furthermore, by (iv) and Fact 2.37, $\operatorname{ran}\mathbf{L} = \operatorname{ran}\mathbf{L} \circ P_{\mathbf{\Delta}^{\perp}} = \operatorname{ran}\mathbf{M}^{\dagger} = \operatorname{ran}\mathbf{M}^{*} = \mathbf{\Delta}^{\perp}$ and so (v) holds.

(ix): Note that $P_{\mathbf{\Delta}^{\perp}} = \mathbf{Id} - P_{\mathbf{\Delta}}$ and from Example 2.32,

$$P_{\mathbf{\Delta}} = m^{-1} \sum_{j \in I} \mathbf{R}^{j}.$$

Hence,

$$P_{\mathbf{\Delta}^{\perp}} = \mathbf{Id} - \frac{1}{m} \sum_{j \in I} \mathbf{R}^{j}. \tag{7.20}$$

Thus, by (viii) and (7.12),

$$\begin{aligned}
\mathbf{M}^{\dagger} = \mathbf{L} \circ P_{\mathbf{\Delta}^{\perp}} &= \frac{1}{m} \sum_{i=1}^{m-1} (m-i)\mathbf{R}^{i-1} \circ \Big(\mathbf{Id} - \frac{1}{m} \sum_{j \in I} \mathbf{R}^{j} \Big) \\
&= \frac{1}{m} \sum_{i=1}^{m-1} (m-i)\mathbf{R}^{i-1} - \frac{1}{m^2} \sum_{i=1}^{m-1} (m-i) \sum_{j \in I} \mathbf{R}^{i+j-1} \\
&= \frac{1}{m} \sum_{i=1}^{m-1} (m-i)\mathbf{R}^{i-1} - \frac{1}{m^2} \Bigg((m-1) \sum_{j=1}^{m} \mathbf{R}^{j} + \\
&\qquad + (m-2) \sum_{j=1}^{m} \mathbf{R}^{j+1} + \cdots + (1) \sum_{j=1}^{m} \mathbf{R}^{m-2+j} \Bigg).
\end{aligned}$$

Using the fact that $\mathbf{R}^{m} = \mathbf{R}^{0}$, $\mathbf{R}^{m+1} = \mathbf{R}^{1}$, etc., and noting that

$$\sum_{i=1}^{m-1} (m-i)\mathbf{R}^{i-1} = \sum_{i=1}^{m} (m-i)\mathbf{R}^{i-1},$$

we get

$$\begin{aligned}
\mathbf{M}^{\dagger} &= \frac{1}{m} \big((m-1)\mathbf{R}^{0} + (m-2)\mathbf{R}^{1} + \cdots + \mathbf{R}^{m-2} \big) \\
&\quad - \frac{1}{m^2} \big((m-1)(\mathbf{R}^{1} + \cdots + \mathbf{R}^{m}) + (m-2)(\mathbf{R}^{2} + \cdots + \mathbf{R}^{m+1}) + \cdots \\
&\quad + (\mathbf{R}^{m-1} + \cdots + \mathbf{R}^{2(m-1)}) \big) \\
&= \sum_{k=1}^{m} \left(\frac{m-k}{m} - \frac{1}{m^2} \sum_{i=1}^{m-1} (m-i) \right) \mathbf{R}^{k-1}
\end{aligned}$$

$$= \sum_{k=1}^{m} \left(\frac{m-k}{m} - \frac{m(m-1)}{m^2} + \frac{1}{m^2} \sum_{i=1}^{m-1} i \right) \mathbf{R}^{k-1}$$
$$= \sum_{k=1}^{m} \left(\frac{2m(m-k)}{2m^2} - \frac{2m(m-1)}{2m^2} + \frac{m(m-1)}{2m^2} \right) \mathbf{R}^{k-1}$$
$$= \sum_{k=1}^{m} \frac{m-(2k-1)}{2m} \mathbf{R}^{k-1}$$

Thus,

$$\mathbf{M}^{\dagger} = (\mathbf{Id} - \mathbf{R})^{\dagger} = \sum_{k=1}^{m} \frac{m-(2k-1)}{2m} \mathbf{R}^{k-1}. \tag{7.21}$$

□

Remark 7.4. Suppose that $\widetilde{\mathbf{L}} \colon \mathbf{\Delta}^{\perp} \to \boldsymbol{\mathcal{H}}$ satisfies $\mathbf{M} \circ \widetilde{\mathbf{L}} = \mathbf{Id}|_{\mathbf{\Delta}^{\perp}}$. Then

$$\mathbf{M}^{-1} \colon \boldsymbol{\mathcal{H}} \rightrightarrows \boldsymbol{\mathcal{H}} \colon \mathbf{y} \mapsto \begin{cases} \widetilde{\mathbf{L}}\mathbf{y} + \mathbf{\Delta}, & \text{if } \mathbf{y} \in \mathbf{\Delta}^{\perp}; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{7.22}$$

One may show that $\mathbf{M}^{\dagger} = P_{\mathbf{\Delta}^{\perp}} \circ \widetilde{\mathbf{L}} \circ P_{\mathbf{\Delta}^{\perp}}$ and that $P_{\mathbf{\Delta}^{\perp}} \circ \widetilde{\mathbf{L}} = \mathbf{L}$ (see (7.12)). Concrete choices for $\widetilde{\mathbf{L}}$ and $\mathbf{L}$ are

$$\mathbf{\Delta}^{\perp} \to \boldsymbol{\mathcal{H}} \colon (y_1, y_2, \ldots, y_m) \mapsto (y_1, y_1 + y_2, \ldots, y_1 + y_2 + y_3 + \cdots + y_m); \tag{7.23}$$

however, the range of the latter operator is not equal to $\mathbf{\Delta}^{\perp}$ whenever $\mathcal{H} \neq \{0\}$.

Corollary 7.5. *The operator $\boldsymbol{A} + \mathbf{M}$ is maximally monotone and*

$$\overline{\operatorname{ran}(\boldsymbol{A} + \mathbf{M})} = \overline{\mathbf{\Delta}^{\perp} + \operatorname{ran} \boldsymbol{A}}.$$

Proof. Since each A_i is maximally monotone and recalling Theorem 7.3(i), we see that $\boldsymbol{A}$ and $\mathbf{M}$ are maximally monotone. On the other hand, $\operatorname{dom} \mathbf{M} = \boldsymbol{\mathcal{H}}$. Thus, by Fact 3.48, $\boldsymbol{A} + \mathbf{M}$ is maximally monotone. Furthermore, by Theorem 7.3(ii) and (iv), $\mathbf{M}$ is rectangular and $\operatorname{ran} \mathbf{M} = \mathbf{\Delta}^{\perp}$. The result therefore follows from Fact 3.66(ii). □

7.2 Composition

We now use Corollary 7.5 to study the composition.

Theorem 7.6. *Suppose that* $(\forall i \in I)$ $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$. *Then the following hold.*

(i) $\mathbf{0} \in \overline{\operatorname{ran}(\mathbf{A} + \mathbf{M})}$.

(ii) $(\forall \varepsilon > 0)$ $(\exists (\mathbf{b}, \mathbf{x}) \in \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{H}})$ $\|\mathbf{b}\| \leq \varepsilon$ *and* $\mathbf{x} = \mathbf{T}(\mathbf{b} + \mathbf{R}\mathbf{x})$.

(iii) $(\forall \varepsilon > 0)$ $(\exists (\mathbf{c}, \mathbf{x}) \in \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{H}})$ $\|\mathbf{c}\| \leq \varepsilon$ *and* $\mathbf{x} = \mathbf{c} + \mathbf{T}(\mathbf{R}\mathbf{x})$.

(iv) $(\forall \varepsilon > 0)$ $(\exists \mathbf{x} \in \boldsymbol{\mathcal{H}})$ $(\forall i \in I)$

$$\|T_{i-1} \cdots T_1 x_m - T_i T_{i-1} \cdots T_1 x_m - x_{i-1} + x_i\| \leq (2i - 1)\varepsilon,$$

where $x_0 = x_m$.

(v) $(\forall \varepsilon > 0)$ $(\exists x \in \mathcal{H})$ $\|x - T_m T_{m-1} \cdots T_1 x\| \leq m^2 \varepsilon$.

Proof. (i): The assumptions and the Minty parametrization (3.14) imply that $(\forall i \in I)$,

$$\begin{aligned} 0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)} &\Leftrightarrow \exists x_i \in \mathcal{H} \text{ such that } (J_{A_i} x_i, 0) \in \overline{\operatorname{gra} A_i} \\ &\Leftrightarrow 0 \in \overline{\operatorname{ran} A_i}. \end{aligned}$$

Hence, $\mathbf{0} \in \overline{\operatorname{ran} \mathbf{A}}$. Obviously, $\mathbf{0} \in \boldsymbol{\Delta}^\perp$. It follows that $\mathbf{0} \in \overline{\boldsymbol{\Delta}^\perp + \operatorname{ran} \mathbf{A}}$. Thus, by Corollary 7.5, $\mathbf{0} \in \overline{\operatorname{ran}(\mathbf{A} + \mathbf{M})}$.

(ii): Fix $\varepsilon > 0$. In view of (i), there exists $\mathbf{x} \in \boldsymbol{\mathcal{H}}$ and $\mathbf{b} \in \boldsymbol{\mathcal{H}}$ such that $\|\mathbf{b}\| \leq \varepsilon$ and $\mathbf{b} \in \mathbf{A}\mathbf{x} + \mathbf{M}\mathbf{x}$. Hence $\mathbf{b} + \mathbf{R}\mathbf{x} \in (\mathbf{Id} + \mathbf{A})\mathbf{x}$ and thus $\mathbf{x} = J_{\mathbf{A}}(\mathbf{b} + \mathbf{R}\mathbf{x}) = \mathbf{T}(\mathbf{b} + \mathbf{R}\mathbf{x})$.

(iii): Let $\varepsilon > 0$. By (ii), there exists $(\mathbf{b}, \mathbf{x}) \in \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{H}}$ such that $\|\mathbf{b}\| \leq \varepsilon$ and $\mathbf{x} = \mathbf{T}(\mathbf{b} + \mathbf{R}\mathbf{x})$. Set $\mathbf{c} = \mathbf{x} - \mathbf{T}(\mathbf{R}\mathbf{x}) = \mathbf{T}(\mathbf{b} + \mathbf{R}\mathbf{x}) - \mathbf{T}(\mathbf{R}\mathbf{x})$ Then, since $\mathbf{T}$ is nonexpansive, $\|\mathbf{c}\| = \|\mathbf{T}(\mathbf{b} + \mathbf{R}\mathbf{x}) - \mathbf{T}(\mathbf{R}\mathbf{x})\| \leq \|\mathbf{b}\| \leq \varepsilon$.

(iv): Take $\varepsilon > 0$. Then, by (iii), there exists $\mathbf{x} \in \boldsymbol{\mathcal{H}}$ and $\mathbf{c} \in \boldsymbol{\mathcal{H}}$ such that $\|\mathbf{c}\| \leq \varepsilon$ and $\mathbf{x} = \mathbf{c} + \mathbf{T}(\mathbf{R}\mathbf{x})$. Let $i \in I$. Then $x_i = c_i + T_i x_{i-1}$. Since $\|c_i\| \leq \|\mathbf{c}\| \leq \varepsilon$ and T_i is nonexpansive, we have

$$\begin{aligned} \|T_i T_{i-1} \cdots T_1 x_0 - x_i\| &\leq \|T_i T_{i-1} \cdots T_1 x_0 - T_i x_{i-1}\| + \|T_i x_{i-1} - x_i\| \\ &\leq \|T_i T_{i-1} \cdots T_1 x_0 - T_i x_{i-1}\| + \varepsilon \\ &\leq \|T_{i-1} \cdots T_1 x_0 - x_{i-1}\| + \varepsilon \\ &\leq \|T_{i-1} \cdots T_1 x_0 - T_{i-1} x_{i-2}\| + \|T_{i-1} x_{i-2} - x_{i-1}\| + \varepsilon \\ &\leq \|T_{i-2} \cdots T_1 x_0 - x_{i-2}\| + 2\varepsilon. \end{aligned}$$

Continuing similarly, we thus obtain

$$\|T_i T_{i-1} \cdots T_1 x_0 - x_i\| \leq i\varepsilon. \tag{7.24}$$

Hence,

$$\|T_{i-1}\cdots T_1x_0 - x_{i-1}\| \le (i-1)\varepsilon. \tag{7.25}$$

Adding (7.24) and (7.25), and recalling the triangle inequality and the fact that $x_m = x_0$,

$$\begin{aligned}
&\|T_{i-1}\cdots T_1x_m - T_iT_{i-1}\cdots T_1x_m - x_{i-1} + x_i\| \\
&\le \|T_{i-1}\cdots T_1x_0 - x_{i-1}\| + \|T_iT_{i-1}\cdots T_1x_0 - x_i\| \\
&\le (i-1)\varepsilon + i\varepsilon = (2i-1)\varepsilon,
\end{aligned}$$

as stated.

(v): Let $\varepsilon > 0$. In view of (iv), there exists $\mathbf{x} \in \boldsymbol{\mathcal{H}}$ such that

$$(\forall i \in I) \quad \|T_{i-1}\cdots T_1x_m - T_iT_{i-1}\cdots T_1x_m - x_{i-1} + x_i\| \le (2i-1)\varepsilon \tag{7.26}$$

where $x_0 = x_m$. Now set,

$$(\forall i \in I) \quad e_i = T_{i-1}\cdots T_1x_m - T_iT_{i-1}\cdots T_1x_m - x_{i-1} + x_i.$$

Then $(\forall i \in I)$ $\|e_i\| \le (2i-1)\varepsilon$. Set $x = x_m$. Then

$$\begin{aligned}
\sum_{i=1}^m e_i &= \sum_{i=1}^m T_{i-1}\cdots T_1x_m - T_iT_{i-1}\cdots T_1x_m - x_{i-1} + x_i && (7.27)\\
&= x - T_mT_{m-1}\cdots T_1x. && (7.28)
\end{aligned}$$

This, (7.26), and the triangle inequality imply that

$$\|x - T_mT_{m-1}\cdots T_1x\| \le \sum_{i=1}^m \|e_i\| \le \sum_{i=1}^m (2i-1)\varepsilon = m^2\varepsilon. \tag{7.29}$$

This completes the proof. $\square$

Remark 7.7. When $m = 2$, then Theorem 7.6(v) also follows from [56, p. 124].

Corollary 7.8. *Suppose that* $(\forall i \in I)$ $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$. *Then*

$$0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_mT_{m-1}\cdots T_1)}.$$

Proof. This follows from Theorem 7.6(v). $\square$

Remark 7.9. The converse implication in Corollary 7.8 fails in general: indeed, consider the case when $\mathcal{H} \neq \{0\}$, $m = 2$, and $v \in \mathcal{H} \setminus \{0\}$. Now set $T_1 \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto x + v$ and set $T_2 \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto x - v$. Then $0 \notin \overline{\operatorname{ran}(\operatorname{Id} - T_1)} = \{-v\}$ and $0 \notin \overline{\operatorname{ran}(\operatorname{Id} - T_2)} = \{v\}$; however, $T_2 T_1 = \operatorname{Id}$ and $\overline{\operatorname{ran}(\operatorname{Id} - T_2 T_1)} = \{0\}$.

Remark 7.10. Corollary 7.8 is optimal in the sense that even if $(\forall i \in I)$ we have $0 \in \operatorname{ran}(\operatorname{Id} - T_i)$, we cannot deduce that $0 \in \operatorname{ran}(\operatorname{Id} - T_m T_{m-1} \cdots T_1)$: indeed, suppose that $\mathcal{H} = \mathbb{R}^2$ and $m = 2$. Set $C_1 := \operatorname{epi} \exp$ and $C_2 := \mathbb{R} \times \{0\}$. Suppose further that $T_1 = P_{C_1}$ and $T_2 = P_{C_2}$. Then $(\forall i \in I)$ $0 \in \operatorname{ran}(\operatorname{Id} - T_i)$; however, $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_2 T_1)} \setminus \operatorname{ran}(\operatorname{Id} - T_2 T_1)$.

7.3 Asymptotic regularity

In this section we show that the composition of asymptotically regular mappings is still asymptotically regular. The following results are corollaries to Bruck and Reich's Fact 3.24.

Corollary 7.11. *Let $S \colon \mathcal{H} \to \mathcal{H}$ be strongly nonexpansive. Then S is asymptotically regular if and only if $0 \in \overline{\operatorname{ran}(\operatorname{Id} - S)}$.*

Proof. "⇒": Recall that S is asymptotically regular if

$$\begin{aligned}(\forall x \in \mathcal{H}) \quad S^n x - S^{n+1} x \to 0 &\Leftrightarrow S^n x - S(S^n x) \to 0 \\ &\Rightarrow 0 \in \overline{\operatorname{ran}(\operatorname{Id} - S)}\end{aligned}$$

"⇐": Fact 3.24(i). ☐

Remark 7.12. Under the assumption that T is firmly nonexpansive, the previous result also follows from Fact 3.25.

Corollary 7.13. *Set $S = T_m T_{m-1} \cdots T_1$. Then S is asymptotically regular if and only if $0 \in \overline{\operatorname{ran}(\operatorname{Id} - S)}$.*

Proof. Since each T_i is firmly nonexpansive, it is also strongly nonexpansive by Fact 3.23(i). By Fact 3.23(ii), S is strongly nonexpansive. Now apply Corollary 7.11. Alternatively, $0 \in \overline{\operatorname{ran}(\operatorname{Id} - S)}$ by Corollary 7.8 and again Corollary 7.11 applies. ☐

We are now ready for our first main result.

Theorem 7.14. *Suppose that each T_i is asymptotically regular. Then the composition $T_m T_{m-1} \cdots T_1$ is asymptotically regular as well.*

Proof. Theorem 7.6(v) implies that $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_m T_{m-1} \cdots T_1)}$. The conclusion thus follows from Corollary 7.13. $\square$

Remark 7.15. (i) When $m = 2$, then the conclusion of Theorem 7.14 also follows from [56, p. 124].

(ii) As an application of Theorem 7.14, we obtain the main result of [6], Example 7.16.

Example 7.16. Let $C_1, \dots, C_m$ be nonempty closed convex subsets of $\mathcal{H}$. Then the composition of the corresponding projectors, $P_{C_m} P_{C_{m-1}} \cdots P_{C_1}$ is asymptotically regular.

Proof. For every $i \in I$, the projector P_{C_i} is firmly nonexpansive, hence strongly nonexpansive, and $\operatorname{Fix} P_{C_i} = C_i \neq \varnothing$. Suppose that $(\forall i \in I)$ $T_i = P_{C_i}$, which is thus asymptotically regular by Corollary 7.11. Now apply Theorem 7.14. $\square$

7.4 Convex combination

In this section, we use our fixed weights $(\lambda_i)_{i \in I}$ to turn $\mathcal{H}^m$ into a Hilbert product space *different from* $\boldsymbol{\mathcal{H}}$ considered in the previous sections. Specifically, we set

$$\mathbf{Y} = \mathcal{H}^m \quad \text{with} \quad \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i \in I} \lambda_i \langle x_i, y_i \rangle \tag{7.30}$$

so that $\|\mathbf{x}\|^2 = \sum_{i \in I} \lambda_i \|x_i\|^2$. We also set

$$\mathbf{Q} \colon \mathcal{H}^m \to \mathcal{H}^m \colon \mathbf{x} \mapsto (\bar{x})_{i \in I}, \text{ where } \bar{x} = \sum_{i \in I} \lambda_i x_i. \tag{7.31}$$

Fact 7.17. [11, Proposition 28.13] *In the Hilbert product space* $\mathbf{Y}$*, we have* $P_{\boldsymbol{\Delta}} = \mathbf{Q}$.

Corollary 7.18. *In the Hilbert product space* $\mathbf{Y}$*, the operator* $\mathbf{Q}$ *is firmly nonexpansive and strongly nonexpansive. Furthermore,* $\operatorname{Fix} \mathbf{Q} = \boldsymbol{\Delta} \neq \varnothing$*,* $\mathbf{0} \in \operatorname{ran}(\mathbf{Id} - \mathbf{Q})$*, and* $\mathbf{Q}$ *is asymptotically regular.*

Proof. By Fact 7.17, the operator $\mathbf{Q}$ is equal to the projector $P_{\boldsymbol{\Delta}}$ and hence firmly nonexpansive. Now apply Fact 3.23(i) to deduce that $\mathbf{Q}$ is strongly nonexpansive. It is clear that $\operatorname{Fix} \mathbf{Q} = \boldsymbol{\Delta}$ and that $\mathbf{0} \in \operatorname{ran}(\mathbf{Id} - \mathbf{Q})$. Finally, recall Corollary 7.11 to see that $\mathbf{Q}$ is asymptotically regular. $\square$

Proposition 7.19. *In the Hilbert product space* $\mathbf{Y}$, *the operator* $\mathbf{T}$ *is firmly nonexpansive.*

Proof. Since each T_i is firmly nonexpansive, $(\forall \mathbf{x} = (x_i)_{i\in I} \in \mathbf{Y})$ $(\forall \mathbf{y} = (y_i)_{i\in I} \in \mathbf{Y})$ we have

$$\|T_i x_i - T_i y_i\|^2 \leq \langle x_i - y_i, T_i x_i - T_i y_i \rangle \,,$$

which implies,

$$\begin{aligned}
\|\mathbf{Tx} - \mathbf{Ty}\|^2 &= \sum_{i\in I} \lambda_i \|T_i x_i - T_i y_i\|^2 \\
&\leq \sum_{i\in I} \lambda_i \langle x_i - y_i, T_i x_i - T_i y_i \rangle \\
&= \langle \mathbf{x} - \mathbf{y}, \mathbf{Tx} - \mathbf{Ty} \rangle \,.
\end{aligned}$$

Thus $\mathbf{T}$ is firmly nonexpansive. ☐

Theorem 7.20. *Suppose that* $(\forall i \in I)$ $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$. *Then the following hold in the Hilbert product space* $\mathbf{Y}$.

(i) $\mathbf{0} \in \overline{\operatorname{ran}(\mathbf{Id} - \mathbf{T})}$.

(ii) $\mathbf{T}$ *is asymptotically regular.*

(iii) $\mathbf{Q} \circ \mathbf{T}$ *is asymptotically regular.*

Proof. (i): This follows because $(\forall \mathbf{x} = (x_i)_{i\in I})$

$$\|\mathbf{x} - \mathbf{Tx}\|^2 = \sum_{i\in I} \lambda_i \|x_i - T_i x_i\|^2.$$

(ii): Combine Fact 3.23(i) with Corollary 7.11.

(iii): On the one hand, $\mathbf{Q}$ is firmly nonexpansive and asymptotically regular by Corollary 7.18. On the other hand, $\mathbf{T}$ is firmly nonexpansive and asymptotically regular by Proposition 7.19 and Theorem 7.20(ii). Altogether, the result follows from Theorem 7.14. ☐

We are now ready for the second main result of this chapter, which concerns convex combinations of asymptotically regular mappings.

Theorem 7.21. *Suppose that each* T_i *is asymptotically regular. Then*

$$\sum_{i\in I} \lambda_i T_i,$$

is asymptotically regular as well.

Proof. Set $S = \sum_{i\in I} \lambda_i T_i$. Fix $x_0 \in \mathcal{H}$ and set $(\forall n \in \mathbb{N})$ $x_{n+1} = Sx_n$. Set $\mathbf{x}_0 = (x_0)_{i\in I} \in \mathcal{H}^m$ and $(\forall n \in \mathbb{N})$ $\mathbf{x}_{n+1} = (\mathbf{Q} \circ \mathbf{T})\mathbf{x}_n$. Then $(\forall n \in \mathbb{N})$ $\mathbf{x}_n = (x_n)_{i\in I}$. Now $\mathbf{Q} \circ \mathbf{T}$ is asymptotically regular by Theorem 7.20(iii); hence, $\mathbf{x}_n - \mathbf{x}_{n+1} = (x_n - x_{n+1})_{i\in I} \to \mathbf{0}$. Thus, $x_n - x_{n+1} \to 0$ and therefore S is asymptotically regular. □

Remark 7.22. Theorem 7.21 extends Theorem 6.36 from Euclidean to Hilbert space.

Remark 7.23. Similarly to Remark 7.10, one cannot deduce that if each T_i has fixed points, then $\sum_{i\in I} \lambda_i T_i$ has fixed points as well: indeed, consider the setting described in Remark 7.10 for an example.

We conclude this chapter by showing that it was necessary to work in $\mathbf{Y}$ and not in $\boldsymbol{\mathcal{H}}$; indeed, viewed in $\boldsymbol{\mathcal{H}}$, the operator $\mathbf{Q}$ is generally not even nonexpansive. The following fact is needed:

Fact 7.24. [11, Proposition 25.4(iii)] *In the Hilbert product space $\boldsymbol{\mathcal{H}}$, set $\boldsymbol{\Delta} = \{\mathbf{x} = (x)_{i\in I} \mid x \in \mathcal{H}\}$ and $\mathbf{j} : \mathcal{H} \to \boldsymbol{\Delta} : x \mapsto (x, \dots, x)$. Then*

$$P_{\boldsymbol{\Delta}}\mathbf{x} = \mathbf{j}\left(\frac{1}{m}\sum_{i\in I} x_i\right).$$

Theorem 7.25. *Suppose that $\mathcal{H} \neq \{0\}$. Then the following are equivalent in the Hilbert product space $\boldsymbol{\mathcal{H}}$.*

(i) $(\forall i \in I)$ $\lambda_i = 1/m$.

(ii) $\mathbf{Q}$ *coincides with the projector* $P_{\boldsymbol{\Delta}}$.

(iii) $\mathbf{Q}$ *is firmly nonexpansive.*

(iv) $\mathbf{Q}$ *is nonexpansive.*

Proof. "(i)⇒(ii)": Fact 7.24. "(ii)⇒(iii)": Clear. "(iii)⇒(iv)": Clear. "(iv)⇒(i)": Take $e \in \mathcal{H}$ such that $\|e\| = 1$. Set $\mathbf{x} := (\lambda_i e)_{i\in I}$ and $y := \sum_{i\in I} \lambda_i^2 e$. Then $\mathbf{Q}\mathbf{x} = (y)_{i\in I}$. We compute $\|\mathbf{Q}\mathbf{x}\|^2 = m\|y\|^2 = m\big(\sum_{i\in I} \lambda_i^2\big)^2$ and $\|\mathbf{x}\|^2 = \sum_{i\in I} \lambda_i^2$. Since $\mathbf{Q}$ is nonexpansive, we must have that $\|\mathbf{Q}\mathbf{x}\|^2 \leq \|\mathbf{x}\|^2$, which is equivalent to

$$m\left(\sum_{i\in I} \lambda_i^2\right)^2 \leq \sum_{i\in I} \lambda_i^2 \tag{7.32}$$

and to

$$m \sum_{i \in I} \lambda_i^2 \leq 1. \tag{7.33}$$

On the other hand, applying the Cauchy-Schwarz inequality to the vectors $(\lambda_i)_{i \in I}$ and $(1)_{i \in I}$ in $\mathbb{R}^m$ yields

$$1 = 1^2 = \left(\sum_{i \in I} \lambda_i \cdot 1 \right)^2 \leq \left\| (\lambda_i)_{i \in I} \right\|^2 \left\| (1)_{i \in I} \right\|^2 = m \sum_{i \in I} \lambda_i^2. \tag{7.34}$$

In view of (7.33) and the Cauchy-Schwarz inequality, (7.34) is actually an equality which implies that $(\lambda_i)_{i \in I}$ is a multiple of $(1)_{i \in I}$. We deduce that $(\forall i \in I)\ \lambda_i = 1/m$. $\square$

In this chapter, we have show that the composition $T_m T_{m-1} \cdots T_1$ and the convex combination $\sum_{i \in I} \lambda_i T_i$ of asymptotically regular firmly nonexpansive mappings in a Hilbert space are asymptotically regular (Theorem 7.6 and Theorem 7.21). Theorem 7.21 also extended a previous result, Theorem 6.36, into the more general Hilbert space setting. In the next chapter we continue with the notion of averages "inheriting" properties from the averaged operators with a look at the resolvent average.

Chapter 8

Inheritance of Properties of the Resolvent Average of Monotone Operators

The resolvent average was previously defined in Chapter 5. In this chapter, based on [21], we determine which properties the average inherits from the averaged operators and provide new results in monotone operator theory. Specifically, we cover the properties provided in Theorem 4.1, as well as k-cyclic monotonicity, orthogonality, and difference maps.

Definition 8.1 (inheritance of properties)**.** For all $i \in I$, let $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators, and $\mathbf{A} = (A_1, \ldots, A_n)$. A property (P) is:

(i) *Dominant* if for some $j \in I$, A_j has property (P) implies that $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ has property (P);

(ii) *Recessive* if for all $i \in I$, A_i has property (P) implies that $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ has property (P), but property (P) is not dominant;

(iii) *Indeterminant* if the property is neither dominant nor recessive.

All of the theorems in this chapter build upon Theorem 5.3, which showed $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ maintains monotonicity when all of the averaged operators are monotone.

8.1 Dominant properties

8.1.1 Single-valuedness of $\mathcal{R}_\mu(\mathbf{A}, \lambda)$

Lemma 8.2. *For all $i \in I$, let T_i be a firmly nonexpansive mapping and let $T = \sum_{i \in I} \lambda_i T_i$. Then for every x and y in $\mathcal{H}$, we have*

$$\begin{aligned}
&\|Tx - Ty\|^2 \\
&= \sum_i \lambda_i \|T_i x - T_i y\|^2 - \tfrac{1}{2} \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 && (8.1)\\
&\leq \|x - y\|^2 - \|x - Tx - y + Ty\|^2 - \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
& && (8.2)\\
&\leq \|x - y\|^2 - \|x - Tx - y + Ty\|^2. && (8.3)
\end{aligned}$$

Consequently, T is firmly nonexpansive.

Proof. Let x and y be in $\mathcal{H}$. By (2.7) and since each T_i is firmly nonexpansive, we have

$$\begin{aligned}
\|Tx - Ty\|^2 &= \|\textstyle\sum_i \lambda_i (T_i x - T_i y)\|^2 \\
&= \textstyle\sum_i \lambda_i \|T_i x - T_i y\|^2 - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
&\leq \textstyle\sum_i \lambda_i \big(\|x - y\|^2 - \|(x - T_i x) - (y - T_i y)\|^2\big) \\
&\qquad\qquad\qquad - \textstyle\frac{1}{2} \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
&= \|x - y\|^2 - \big(\textstyle\sum_i \lambda_i \|(x - T_i x) - (y - T_i y)\|^2 \\
&\qquad\qquad\qquad + \textstyle\frac{1}{2} \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2\big) \\
&= \|x - y\|^2 - \big(\|\textstyle\sum_i \lambda_i (x - T_i x - y + T_i y)\|^2 \\
&\qquad\qquad\qquad + \textstyle\sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2\big) \\
&= \|x - y\|^2 - \|x - Tx - y + Ty\|^2 \\
&\qquad\qquad\qquad - \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
&\leq \|x - y\|^2 - \|x - Tx - y + Ty\|^2,
\end{aligned}$$

and the result follows. ☐

Corollary 8.3. *For all $i \in I$, let T_i be firmly nonexpansive on $\mathcal{H}$, λ_i be strictly positive real numbers such that $\sum_{i \in I} \lambda_i = 1$, and set $T = \sum_{i=1}^n \lambda_i T_i$. Let x and y be in $\mathcal{H}$ such that $Tx = Ty$. Then $(\forall i \in I)$ $T_i x = T_i y$. Consequently, if some T_i is injective, so is T.*

Proof. By Lemma 8.2, we have

$$
\begin{aligned}
0 = \|Tx - Ty\|^2 &= \sum_i \lambda_i \|T_i x - T_i y\|^2 - \tfrac{1}{2} \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
&\le \|x - y\|^2 - \|x - y\|^2 - \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \\
&= - \sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 \le 0.
\end{aligned}
\tag{8.4}
$$

Thus $\sum_{i,j} \lambda_i \lambda_j \|T_i x - T_i y - T_j x + T_j y\|^2 = 0$ so we must have $(\forall i \in I)(\forall j \in I)$ $T_i x - T_i y = T_j x - T_j y$ and therefore

$$
0 = Tx - Ty = \sum_{i=1}^n \lambda_i T_i x - \sum_{i=1}^n \lambda_i T_i y = T_i x - T_i y.
$$

Thus $T_i x = T_i y$ and the result follows. □

Corollary 8.4. *For all $i \in I$, let T_i be firmly nonexpansive on $\mathcal{H}$, λ_i be strictly positive real numbers such that $\sum_{i \in I} \lambda_i = 1$, and set $T = \sum_{i=1}^n \lambda_i T_i$. Let z, u, v in $\mathcal{H}$ be given such that $u = T(u + z)$ and $(\forall i \in I)$ $v = T_i(v + z)$. Then $v = T(v + z)$ and $(\forall i \in I)$ $T_i(u + z) = u$.*

Proof. It is clear that $v = T(v + z)$. Now set $x = u + z$ and $y = v + z$ in Lemma 8.2 to deduce

$$
\|u - v\|^2 \le \|u - v\|^2 - \sum_{i,j} \lambda_i \lambda_j \|T_i(u + z) - T_j(u + z)\|^2. \tag{8.5}
$$

Hence $(\forall i \in I)(\forall j \in I)$ $T_i(u + z) = T_j(u + z)$. Since $T(u + z) = u$, we must have

$$
T(u + z) = \sum_{i=1}^n \lambda_i T_i(u + z) = T_i(u + z) = u.
$$

Thus $(\forall i \in I)$ $T_i(u + z) = u$. □

We are now ready for the main result of this section.

Theorem 8.5 (single-valuedness is dominant)**.** *For all $i \in I$, let $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone and assume that some A_j is at most single-valued. Then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is also at most single-valued.*

Proof. By Theorem 4.1(iv), a maximally monotone operator is at most single-valued if and only if its resolvent is injective. Hence $J_{\mu A_j}$ is injective. By Corollary 8.3 and (5.3), $J_{\mu \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}$ is injective. Hence $\mu \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ and therefore $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is at most single-valued. □

8.1.2 Domain and range

Proposition 8.6 (resolvents are rectangular). *Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive. Then T is rectangular.*

Proof. Since T is firmly nonexpansive, it is the resolvent of a monotone operator, A, such that $T = (\mathrm{Id} + A)^{-1}$. From the definition, it follows that T is rectangular if and only if $T^{-1} = \mathrm{Id} + A$ is rectangular. Now by Example 3.53, $F_{\mathrm{Id}}(x, x^*) = \frac{1}{4}\|x + x^*\|^2$. We know that $F_A(x, x^*) = \langle x, x^*\rangle$ if $(x, x^*) \in \operatorname{gra} A$. To show that $\mathrm{Id} + A$ is rectangular, we must show that $\operatorname{dom}(A + \mathrm{Id}) \times \operatorname{ran}(A + \mathrm{Id}) \subseteq \operatorname{dom} F_{A+\mathrm{Id}}$, which by Fact 3.43 means that

$$\operatorname{dom} A \times \mathcal{H} \subseteq \operatorname{dom} F_{A+\mathrm{Id}}.$$

To this end, let $(x, u) \in \operatorname{dom} A \times \mathcal{H}$ and take $x^* \in Ax$. Then $F_{A+\mathrm{Id}}(x, u) \leq (F_A(x, \cdot)\Box F_{\mathrm{Id}}(x, \cdot))(u) \leq F_A(x, x^*) + F_{\mathrm{Id}}(x, u - x^*) < +\infty$. Hence $(x, u) \in \operatorname{dom} F_{A+\mathrm{Id}}$ and thus $A + \mathrm{Id}$ is rectangular. $\square$

Proposition 8.7. *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, and let $\gamma > 0$. Then A is rectangular if and only if γA is rectangular.*

Proof. Use the definitions and $F_{\gamma A}(x, x^*) = \gamma F_A(x, x^*/\gamma)$. $\square$

Proposition 8.8 (surjective). *For all $i \in I$, let T_i be a firmly nonexpansive. If some T_j is surjective, then so is $T = \sum_{i=1}^n \lambda_i T_i$.*

Proof. First, consider the case where $T = \lambda_1 T_1 + \lambda_2 T_2$. Without loss of generality, assume that T_1 is surjective. By Proposition 8.6, since T is firmly nonexpansive, it is a resolvent and hence rectangular. Each T_i is rectangular as well, as is each $\lambda_i T_i$ by Proposition 8.7. Using Fact 3.66(i),

$$\operatorname{int}\operatorname{ran}(\lambda_1 T_1 + \lambda_2 T_2) = \operatorname{int}(\operatorname{ran}\lambda_1 T_1 + \operatorname{ran}\lambda_2 T_2),$$

we see that T is surjective. The case $n > 2$ now follows inductively. $\square$

Theorem 8.9 (full domain is dominant). *For all $i \in I$, let A_i be maximally monotone and suppose that for some $j \in I$, $\operatorname{dom} A_j = \mathcal{H}$. Then $\operatorname{dom}\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathcal{H}$.*

Proof. Since $\operatorname{dom} A_j = \mathcal{H}$, then $\operatorname{dom}\mu A_j = \mathcal{H}$. By Theorem 4.1(ii), $\operatorname{dom}\mu A_j = \mathcal{H}$ if and only if $J_{\mu A_j}$ is surjective. Then by (5.3) and Proposition 8.8, $J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \sum\limits_{i\in I} \lambda_i J_{\mu A_i}$ is surjective and thus $\operatorname{dom}\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathcal{H}$. $\square$

Theorem 8.10 (surjectivity is dominant)**.** *For all $i \in I$, let A_i be maximally monotone and suppose that for some $j \in I$, A_j is surjective. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is surjective.*

Proof. By Theorem 4.1(iii) and (3.13),

$$\begin{aligned} A_j \text{ is surjective } &\Leftrightarrow \mathrm{Id} - J_{A_j} \text{ is surjective.} \\ &\Leftrightarrow J_{A_j^{-1}} \text{ is surjective.} \end{aligned}$$

Thus, by Corollary 5.9, and Proposition 8.8, $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}}$ is surjective. Applying Theorem 4.1(iii) and (3.13) again, we have

$$\begin{aligned} J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}} \text{ is surjective } &\Leftrightarrow \mathrm{Id} - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} \text{ is surjective.} \\ &\Leftrightarrow \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) \text{ is surjective.} \end{aligned}$$

□

8.1.3 Strict monotonicity

Lemma 8.11. *For all $i \in I$, let T_i be firmly nonexpansive and λ_i be strictly positive real numbers such that $\sum_{i \in I} \lambda_i = 1$. If there exists $j \in I$ such that T_j is strictly firmly nonexpansive then $T = \sum_{i \in I} \lambda_i T_i$ is strictly firmly nonexpansive.*

Proof. We know from Lemma 8.2 that T is firmly nonexpansive. To show T is strictly firmly nonexpansive we need to show that for $u, v \in \operatorname{dom} T$ if $Tu \neq Tv$ then $\|Tu - Tv\|^2 < \langle Tu - Tv, u - v\rangle$. Suppose to the contrary that

$$Tu \neq Tv \text{ and } \|Tu - Tv\|^2 = \langle u - v, Tu - Tv\rangle . \tag{8.6}$$

We know from (8.1) and the firm nonexpansiveness of the T_i that

$$\begin{aligned} \|Tu - Tv\|^2 = \|\sum_{i \in I} \lambda_i (T_i u - T_i v)\|^2 &\leq \sum_{i \in I} \lambda_i \|T_i u - T_i v\|^2 \\ &\leq \sum_{i \in I} \lambda_i \langle u - v, T_i u - T_i v\rangle = \langle u - v, Tu - Tv\rangle . \end{aligned} \tag{8.7}$$

Since $\|Tu - Tv\|^2 = \langle u - v, Tu - Tv\rangle$, (8.7) yields

$$\begin{aligned} \|Tu - Tv\|^2 = \|\sum_{i \in I} \lambda_i (T_i u - T_i v)\|^2 &= \sum_{i \in I} \lambda_i \|T_i u - T_i v\|^2 \\ &= \sum_{i \in I} \lambda_i \langle T_i u - T_i v, u - v\rangle \end{aligned} \tag{8.8}$$

Since $\|\cdot\|^2$ is strictly convex, we have

$$T_i u - T_i v = T_j u - T_j v \quad (\forall i \in I). \tag{8.9}$$

Since each T_i is firmly nonexpansive, $\|T_i u - T_i v\|^2 \leq \langle T_i u - T_i v, u - v\rangle$ and the third equality of (8.8) gives

$$\|T_i u - T_i v\|^2 = \langle T_i u - T_i v, u - v\rangle \quad (\forall i \in I). \tag{8.10}$$

By definition of T, (8.9) and the fact that $Tu \neq Tv$ we also have

$$Tu - Tv = \sum_{i\in I} \lambda_i (T_i u - T_i v) = T_j u - T_j v \neq 0.$$

Then $\|T_j u - T_j v\|^2 < \langle T_j u - T_j v, u - v\rangle$, since T_j is strictly firmly nonexpansive. But this contradicts (8.10), and thus (8.6) is false. Therefore,

$$\|Tu - Tv\|^2 < \langle Tu - Tv, u - v\rangle \quad \text{whenever } Tu \neq Tv,$$

and hence T is strictly firmly nonexpansive. □

Theorem 8.12 (strict monotonicity is dominant)**.** *For all $i \in I$, let A_i be monotone and additionally assume that some A_j is strictly monotone. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is strictly monotone.*

Proof. By Theorem 4.1(vi), since A_j is strictly monotone then J_{A_j} is strictly firmly nonexpansive, thus by Lemma 8.11 we have $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum\limits_{i\in I} \lambda_i J_{A_i}$ is strictly firmly nonexpansive. Apply Theorem 4.1(vi) again to see that $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is strictly monotone. □

8.1.4 Banach contraction

Proposition 8.13 (Banach contraction)**.** *For all $i \in I$, let T_i be a firmly nonexpansive mapping and $\lambda_i \in \mathbb{R}_{++}$ such that $\sum_{i=1}^n \lambda_i = 1$. If some T_j is a Banach contraction with constant β, then $T = \sum_{i\in I} \lambda_i T_i$ is a Banach contraction with constant $(1 - \lambda_j(1-\beta))$.*

Proof. Suppose that T_j is β-Lipschitz, with $0 \leq \beta < 1$. Let x and y be in

$\mathcal{H}$. Then

$$\begin{aligned}\|Tx - Ty\| &\leq \sum_{i\in I} \lambda_i \|T_i x - T_i y\| \\ &\leq \sum_{i\neq j} \lambda_i \|x-y\| + \lambda_j \beta \|x-y\| \\ &= \sum_{i\in I} \lambda_i \|x-y\| - \lambda_j(1-\beta)\|x-y\| \\ &= (1-\lambda_j(1-\beta))\|x-y\|.\end{aligned}$$

□

Theorem 8.14. *For all $i \in I$, let A_i be maximally monotone operators from $\mathcal{H} \rightrightarrows \mathcal{H}$ and assume that for some $j \in I$ and J_{A_j} is a Banach contraction with constant β. Then $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}$ is a Banach contraction with constant $\gamma = (1-\lambda_j(1-\beta))$ and $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ satisfies*

$$(\forall (x,u) \in \operatorname{gra} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))\,(\forall (y,v) \in \operatorname{gra} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))$$
$$\frac{1-\gamma^2}{\gamma^2}\|x-y\|^2 \leq 2\langle x-y, u-v\rangle + \|u-v\|^2,$$

Proof. Since, J_{A_j} is a Banach contraction with constant β, applying Proposition 8.13 and (5.3), $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum\limits_{i\in I} \lambda_i J_{A_i}$ is a Banach contraction with constant $\gamma = (1-\lambda_j(1-\beta))$. Therefore, Theorem 4.1(xiii) yields that $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ satisfies

$$(\forall (x,u) \in \operatorname{gra} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))\,(\forall (y,v) \in \operatorname{gra} \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))$$
$$\frac{1-\gamma^2}{\gamma^2}\|x-y\|^2 \leq 2\langle x-y, u-v\rangle + \|u-v\|^2,$$

where $\gamma = (1-\lambda_j(1-\beta))$. □

8.1.5 Rectangularity and paramonotonicity

While rectangularity and paramonotonicity are not typically dominant properties, in the special case where the operators are linear on $\mathbb{R}^N$, they are dominant.

Theorem 8.15 (linear rectangularity and paramonotonicity are dominant on $\mathbb{R}^N$)**.** *Assume that $(\forall i \in I)$ $A_i : \mathbb{R}^N \to \mathbb{R}^N$ is linear and at least one A_i is paramonotone (equivalently rectangular). Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is linear and paramonotone (equivalently rectangular).*

Proof. This is Theorem 8.24(iii) combined with Fact 3.58(i). □

8.2 Dominant or recessive properties

In this section we gather the properties that are at least recessive, but for which there is not yet a proof or counterexample for dominance.

8.2.1 Strong monotonicity

Theorem 8.16. *For all $i \in I$, let T_i be $(1+\epsilon_i)$ firmly nonexpansive with $\epsilon_i \geq 0$. Then $T = \sum\limits_{i \in I} \lambda_i T_i$ is $(1+\epsilon)$ firmly nonexpansive, where $\epsilon = \min_{i \in I} \epsilon_i$.*

Proof. T_i is $(1+\epsilon)$ firmly nonexpansive with $\epsilon \geq 0$ gives

$$\|T_i x - T_i y\|^2 \leq (1+\epsilon)^{-1} \langle T_i x - T_i y, x - y \rangle . \tag{8.11}$$

By the convexity of $\|\cdot\|^2$, and (8.11), we have

$$\begin{aligned} \|Tx - Ty\|^2 &\leq \sum_{i \in I} \lambda_i \|T_i x - T_i y\|^2 \\ &\leq \sum_{i \in I} \lambda_i (1+\epsilon_i)^{-1} \langle T_i x - T_i y, x - y \rangle \\ &\leq \sum_{i=1}^{n} \lambda_i (1+\epsilon)^{-1} \langle T_i x - T_i y, x - y \rangle \\ &= (1+\epsilon)^{-1} \langle Tx - Ty, x - y \rangle . \end{aligned}$$

Thus T is $(1+\epsilon)$ firmly nonexpansive, where $\epsilon = \min_{i \in I} \epsilon_i$. □

Theorem 8.17 (strong monotonicity)**.** *For all $i \in I$, let A_i be strongly monotone with constant ϵ_i. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is strongly monotone with constant $\epsilon = \min_{i \in I} \epsilon_i$.*

Proof. By Theorem 4.1(xi), J_{A_i} is $(1+\epsilon_i)$ firmly nonexpansive for all $i \in I$. Then by Theorem 8.16, $\sum\limits_{i \in I} \lambda_i J_{A_i}$ is $(1+\epsilon)$ firmly nonexpansive where $\epsilon = \min_{i \in I} \epsilon_i$. Thus by (5.3) and Theorem 4.1(xi), $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is ϵ strongly monotone. □

8.2.2 γ-cocoercive

Theorem 8.18 (γ-cocoercive)**.** *For all $i \in I$, let A_i be maximally monotone and γ_i-cocoercive for $\gamma_i > 0$ and $\lambda_i \in \mathbb{R}_{++}$ such that $\sum_{i=1}^{n} \lambda_i = 1$. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is γ- cocoercive, where $\gamma = \min_{i \in I} \gamma_i$.*

Proof. By Theorem 4.1(xii), (3.13) and Theorem 4.1(xi),

$$
\begin{aligned}
& A_i \text{ is } \gamma_i\text{-cocoercive} \\
& \Leftrightarrow (1+\gamma_i)(\operatorname{Id} - J_{A_i}) \text{ is firmly nonexpansive;} \\
& \Leftrightarrow (1+\gamma_i) J_{A_i^{-1}} \text{ is firmly nonexpansive.} \\
& \Leftrightarrow (\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \| J_{A_i^{-1}} x - J_{A_i^{-1}} y \|^2 \\
& \qquad\qquad \le (1+\gamma_i)^{-1} \left\langle x - y, J_{A_i^{-1}} x - J_{A_i^{-1}} y \right\rangle .
\end{aligned}
$$

Then by (8.1), $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$

$$
\begin{aligned}
& \| \sum_{i \in I} \lambda_i J_{A_i^{-1}} x - \sum_{i \in I} \lambda_i J_{A_i^{-1}} y \|^2 && (8.12) \\
& \le \sum_{i \in I} \lambda_i \| J_{A_i^{-1}} x - J_{A_i^{-1}} y \|^2 \\
& \le \sum_{i \in I} \lambda_i (1+\gamma_i)^{-1} \left\langle x - y, J_{A_i^{-1}} x - J_{A_i^{-1}} y \right\rangle \\
& \le \sum_{i \in I} \lambda_i (1+\gamma)^{-1} \left\langle x - y, J_{A_i^{-1}} x - J_{A_i^{-1}} y \right\rangle \\
& = (1+\gamma)^{-1} \left\langle x - y, \sum_{i \in I} \lambda_i J_{A_i^{-1}} x - \sum_{i \in I} \lambda_i J_{A_i^{-1}} y \right\rangle , && (8.13)
\end{aligned}
$$

where $\gamma = \min_{i \in I} \gamma_i$. Applying Theorem 4.1(xi), Theorem 5.19, and Theorem 4.1(xii) to (8.13) we have, $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$

$$
\begin{aligned}
& \| \sum_{i \in I} \lambda_i J_{A_i^{-1}} x - \sum_{i \in I} \lambda_i J_{A_i^{-1}} y \|^2 \\
& \qquad\qquad \le (1+\gamma)^{-1} \left\langle x - y, \sum_{i \in I} \lambda_i J_{A_i^{-1}} x - \sum_{i \in I} \lambda_i J_{A_i^{-1}} y \right\rangle \\
& \Leftrightarrow (1+\gamma) J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})^{-1}} \text{ is firmly nonexpansive;} \\
& \Leftrightarrow (1+\gamma)(\operatorname{Id} - J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})}) \text{ is firmly nonexpansive;} \\
& \Leftrightarrow \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) \text{ is } \gamma\text{-cocoercive.}
\end{aligned}
$$

$\square$

8.3 Recessive properties

8.3.1 Maximality and linearity

Theorem 8.19 (maximal monotonicity is recessive)**.** *$\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is maximally monotone if and only if for all $i \in I$, A_i is maximally monotone.*

Proof. This is a consequence of Theorem 5.3. $\square$

Theorem 8.20 (linear relations are recessive)**.** *For all $i \in I$, let A_i be a maximal monotone linear relation. Then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is a maximal monotone linear relation.*

Proof. Since each A_i is a linear relation, Fact 2.20 shows that $A_i + \mathrm{Id}$ is a linear relation and therefore $(A_i + \mathrm{Id})^{-1}$ is a linear relation. The maximality of A_i and Fact 3.43 imply that $(A_i + \mu^{-1}\,\mathrm{Id})^{-1}$ is single-valued and full-domain. Thus, $(A_i + \mu^{-1}\,\mathrm{Id})^{-1}$ is a linear mapping. Using

$$(\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) + \mu^{-1}\,\mathrm{Id})^{-1} = \lambda_1(A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1},$$

we see that $(\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) + \mathrm{Id})^{-1}$ is linear, therefore $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) + \mathrm{Id}$ is a linear relation. Then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = (\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) + \mathrm{Id}) - \mathrm{Id}$ is a linear relation. $\square$

The next example shows that linearity is not a dominant property.

Example 8.21. Set $f = \|\cdot\|$ and $A_1 = \partial f$ and $A_2 = \mathbf{0}$. Then by (3.16) and Example 2.55

$$J_{A_1}x = \mathrm{prox}_f\, x = \begin{cases} \left(1 - \frac{1}{\|x\|}\right) x, & \text{if } \|x\| > 1; \\ 0, & \text{if } \|x\| \le 1, \end{cases}$$

and we have $J_{A_2} = \mathrm{Id}$. Then J_{A_1} is not linear and J_{A_2} is linear. However,

$$J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})}x = \lambda J_{A_1}x + (1-\lambda)J_{A_2}x = \begin{cases} \left(1 - \lambda\frac{1}{\|x\|}\right) x, & \text{if } \|x\| > 1; \\ (1-\lambda)x, & \text{if } \|x\| \le 1, \end{cases}$$

which is not linear. Thus $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is not a linear relation.

8.3.2 Rectangularity and paramonotonicity

Theorem 8.22 (rectangularity is recessive)**.** *Assume that for each $i \in I$, $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$ is a rectangular maximally monotone operator. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is rectangular.*

Proof. By Theorem 4.1(xvii), we need to show $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$,

$$\inf_{z \in \mathcal{H}} \left\langle J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}z, (y - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y) - (z - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}z)\right\rangle > -\infty. \tag{8.14}$$

Using $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum_{i=1}^n \lambda_i J_{A_i}$, (8.14) becomes $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$,

$$\inf_{z \in \mathcal{H}} \left\langle \sum_{i=1}^n \lambda_i (J_{A_i}x - J_{A_i}z), \sum_{j=1}^n \lambda_j [(y-z) - (J_{A_j}y - J_{A_j}z)] \right\rangle > -\infty. \tag{8.15}$$

Since A_i is rectangular, by Theorem 4.1(xvii) we have for each $i \in I$,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \inf_{z \in \mathcal{H}} \langle J_{A_i}x - J_{A_i}z, (y - J_{A_i}y) - (z - J_{A_i}z)\rangle > -\infty. \tag{8.16}$$

Setting $x_i = J_{A_i}x - J_{A_i}z$, $u_i = (y-z) - (J_{A_i}y - J_{A_i}z)$, and

$$c_{ij} = -\left\langle J_{A_i}x - J_{A_j}x, J_{A_i}y - J_{A_j}y\right\rangle + \frac{1}{4}\|(J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y)\|^2,$$

we have $\langle x_i - x_j, u_i - u_j\rangle =$

$$\begin{aligned}
&\left\langle (J_{A_i}x - J_{A_j}x) - (J_{A_i}z - J_{A_j}z), -(J_{A_i}y - J_{A_j}y) + (J_{A_i}z - J_{A_j}z)\right\rangle \\
&= -\left\langle J_{A_i}x - J_{A_j}x, J_{A_i}y - J_{A_j}y\right\rangle \\
&\quad + \left\langle (J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y), J_{A_i}z - J_{A_j}z\right\rangle - \|J_{A_i}z - J_{A_j}z\|^2 \\
&= -\left\langle J_{A_i}x - J_{A_j}x, J_{A_i}y - J_{A_j}y\right\rangle \qquad (8.17)\\
&\quad - \left\| \frac{(J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y)}{2} - (J_{A_i}z - J_{A_j}z)\right\|^2 \\
&\quad + \frac{\|(J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y)\|^2}{4} \\
&= c_{ij} - \left\| \frac{(J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y)}{2} - (J_{A_i}z - J_{A_j}z)\right\|^2. \qquad (8.18)
\end{aligned}$$

Then for given $x, y \in \mathcal{H}$, using Fact 2.42, (8.18) and (8.16),

$$\begin{aligned}
&\inf_{z \in \mathcal{H}} \left\langle \sum_{i=1}^n \lambda_i (J_{A_i}x - J_{A_i}z), \sum_{j=1}^n \lambda_j [(y-z) - (J_{A_j}y - J_{A_j}z)] \right\rangle \\
&= \inf_{z \in \mathcal{H}} \Bigg[\sum_{i=1}^n \lambda_i \langle J_{A_i}x - J_{A_i}z, (y-z) - (J_{A_i}y - J_{A_i}z)\rangle - \\
&\qquad \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \frac{1}{2}\left(c_{ij} - \left\| \frac{(J_{A_i}x - J_{A_j}x) + (J_{A_i}y - J_{A_j}y)}{2} - (J_{A_i}z - J_{A_j}z)\right\|^2 \right)\Bigg]
\end{aligned}$$

$$\geq \inf_{z\in\mathcal{H}} \Big(\sum_{i=1}^{n} \lambda_i \left\langle J_{A_i}x - J_{A_i}z, (y-z) - (J_{A_i}y - J_{A_i}z) \right\rangle \Big) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \frac{c_{ij}}{2}$$
$$\geq \sum_{i=1}^{n} \lambda_i \inf_{z\in\mathcal{H}} \left\langle J_{A_i}x - J_{A_i}z, (y-z) - (J_{A_i}y - J_{A_i}z) \right\rangle - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \frac{c_{ij}}{2}$$
$$= \sum_{i=1}^{n} \lambda_i \inf_{z\in\mathcal{H}} \left\langle J_{A_i}x - J_{A_i}z, (y - J_{A_i}y) - (z - J_{A_i}z) \right\rangle - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \frac{c_{ij}}{2} > -\infty.$$

Hence (8.15) holds and $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is rectangular. ☐

Remark 8.23. Theorem 8.22 is not true if only one A_i is rectangular. See Example 8.27 for a counterexample.

Theorem 8.24. *Let A_i be maximally monotone operators from $\mathcal{H} \rightrightarrows \mathcal{H}$ for all $i \in I$. The following hold:*

(i) *Assume that $(\forall i \in I)$ A_i is paramonotone. Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is paramonotone.*

(ii) *Assume that at least one A_i is paramonotone and at most single-valued. Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is paramonotone and at most single-valued.*

(iii) *Assume that $(\forall i \in I)$ A_i is linear and at least one A_i is paramonotone. Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is linear and paramonotone.*

Proof. By Theorem 4.1(xv), we need to show that $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$,

$$\|J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y\|^2 = \big\langle x - y, J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y \big\rangle \tag{8.19}$$
$$\Rightarrow \begin{cases} J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}\big(J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x + y - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y\big) \\ J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}\big(J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y + x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x\big). \end{cases} \tag{8.20}$$

Using $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum_{i=1}^{n} \lambda_i J_{A_i}$, (8.19) becomes

$$\|\sum_{i=1}^{n} \lambda_i (J_{A_i}x - J_{A_i}y)\|^2 = \sum_{i=1}^{n} \lambda_i \left\langle x - y, J_{A_i}x - J_{A_i}y \right\rangle. \tag{8.21}$$

By the strong convexity of $\|\cdot\|^2$, (2.7) with $x_i = J_{A_i}x - J_{A_i}y$, gives

$$\|\sum_{i=1}^{n} \lambda_i (J_{A_i}x - J_{A_i}y)\|^2 = \sum_{i=1}^{n} \lambda_i \|J_{A_i}x - J_{A_i}y\|^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \frac{\|(J_{A_i}x - J_{A_i}y) - (J_{A_j}x - J_{A_j}y)\|^2}{2}. \tag{8.22}$$

Then it follows from (8.21) and (8.22) that (8.19) is equivalent to

$$\sum_{i=1}^{n}\lambda_i\|J_{A_i}x-J_{A_i}y\|^2-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\frac{\|(J_{A_i}x-J_{A_i}y)-(J_{A_j}x-J_{A_j}y)\|^2}{2}$$
$$=\sum_{i=1}^{n}\lambda_i\left\langle x-y,J_{A_i}x-J_{A_i}y\right\rangle, \quad (8.23)$$

That is,

$$\sum_{i=1}^{n}\lambda_i(\|J_{A_i}x-J_{A_i}y\|^2-\langle x-y,J_{A_i}x-J_{A_i}y\rangle)$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\frac{\|(J_{A_i}x-J_{A_i}y)-(J_{A_j}x-J_{A_j}y)\|^2}{2}\geq 0. \quad (8.24)$$

Since J_{A_i} is firmly nonexpansive, Fact 3.3(iv) gives

$$(\forall i\in I)\quad \|J_{A_i}x-J_{A_i}y\|^2-\langle x-y,J_{A_i}x-J_{A_i}y\rangle\leq 0 \quad (8.25)$$

Then because $\lambda_i>0$, (8.24) and (8.25) indicates that

$$(\forall i\in I)\quad \|J_{A_i}x-J_{A_i}y\|^2=\left\langle x-y,J_{A_i}x-J_{A_i}y\right\rangle, \quad (8.26)$$

and

$$(\forall\ i\in I)(\forall j\in I)\quad J_{A_i}x-J_{A_i}y=J_{A_j}x-J_{A_j}y=d \quad (8.27)$$

where $d\in\mathcal{H}$. In particular, multiplying (8.27) by λ_i, followed by summation, gives

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x-J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y=d. \quad (8.28)$$

With these, we are ready to show:

(i) If each A_i is paramonotone, then (8.26) and (4.11) gives

$$(\forall i\in I)\quad J_{A_i}x=J_{A_i}(J_{A_i}x+y-J_{A_i}y),$$
$$(\forall\ i\in I)\quad J_{A_i}y=J_{A_i}(J_{A_i}y+x-J_{A_i}x).$$

In view of (8.27) this is,

$$(\forall i\in I)\quad J_{A_i}x=J_{A_i}(d+y), \quad (8.29)$$
$$(\forall i\in I)\quad J_{A_i}y=J_{A_i}(-d+x). \quad (8.30)$$

Then multiplying (8.29) and (8.30) by λ_i, taking summations and using (8.28) leads to

$$\sum_{i=1}^{n} \lambda_i J_{A_i} x = \sum_{i=1}^{n} \lambda_i J_{A_i}(d+y) = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}(J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x + y - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y),$$

$$\sum_{i=1}^{n} \lambda_i J_{A_i} y = \sum_{i=1}^{n} \lambda_i J_{A_i}(-d+x) = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}(J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y + x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x),$$

which is (8.20). Thus $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is paramonotone.

(ii) If at least one A_i is paramonotone, say A_i, then for this A_i, (8.26) and (4.11) gives

$$J_{A_i}x = J_{A_i}(J_{A_i}x + y - J_{A_i}y),$$
$$J_{A_i}y = J_{A_i}(J_{A_i}y + x - J_{A_i}x).$$

By Theorem 4.1(iv), since A_i is at most single-valued then J_{A_i} is injective, therefore

$$x = J_{A_i}x + y - J_{A_i}y,$$
$$y = J_{A_i}y + x - J_{A_i}x. \tag{8.31}$$

Both of which imply,

$$x - y = J_{A_i}x - J_{A_i}y. \tag{8.32}$$

Note that (8.27) and (8.32) together signify

$$(\forall j \in I) \quad x - y = J_{A_j}x - J_{A_j}y$$

which gives us

$$(\forall\ j \in I) \quad x = J_{A_j}x + y - J_{A_j}y,$$
$$(\forall\ j \in I) \quad y = J_{A_j}y + x - J_{A_j}x.$$

Multiplying both sides by λ_j, followed by summation, gives

$$x = \sum_{j=1}^{n} \lambda_j J_{A_j}x + y - \sum_{j=1}^{n} \lambda_j J_{A_j}y = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x + y - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y,$$

$$y = \sum_{j=1}^{n} \lambda_j J_{A_j}y + x - \sum_{j=1}^{n} \lambda_j J_{A_j}x = J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}y + x - J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x.$$

Hence (8.20) follows immediately. Therefore, $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is paramonotone. By Theorem 8.5 $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is at most single-valued.

(iii) follows from (ii). □

Remark 8.25. Theorem 8.15 gives an improved version of Theorem 8.22 when each A_i is linear and monotone.

Remark 8.26. Example 8.27 demonstrates that Theorems 8.22 and 8.24 are almost optimal, and they cannot be significantly improved.

Example 8.27. Define $A_1 : \mathbb{R}^2 \mapsto \mathbb{R}^2$ to be the normal cone operator of the set $\mathbb{R} \times \{0\}$. That is,

$$A_1 := N_{\mathbb{R}\times\{0\}}.$$

Let $A^2 : \mathbb{R}^2 \to \mathbb{R}^2$ be the skew operator such that

$$A_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } J_{A_2} = (\mathrm{Id} + A_2)^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Then by Fact 3.42, J_{A_1} is the projector on $\mathbb{R} \times \{0\}$,

$$J_{A_1} = P_{\mathbb{R}\times\{0\}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Then for $\lambda_1 = \lambda_2 = \frac{1}{2}$ we have,

$$\begin{aligned}
\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) &= (\frac{1}{2}J_{A_1} + \frac{1}{2}J_{A_2})^{-1} - \mathrm{Id} \\
&= \left(\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\right)^{-1} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}^{-1} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix},
\end{aligned}$$

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})_+ = \frac{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) + \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})^\intercal}{2} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Clearly, $\operatorname{rank} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = 2$ while $\operatorname{rank} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})_+ = 1$. As $\operatorname{rank} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) \neq \operatorname{rank} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})_+$, Fact 3.58(ii) implies that $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is not paramonotone and equivalently $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is not rectangular.

Remark 8.28. Note that in Example 8.27 we have demonstrated the following:

(i) A_1 is rectangular and A_2 is not rectangular, and we have $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is not rectangular. Therefore the requirement for all A_i to be rectangular in Theorem 8.22 is optimal.

(ii) A_1 is paramonotone and A_2 is not paramonotone, and we have $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is not paramonotone. This implies that Theorem 8.24(i) is optimal.

(iii) A_1 is paramonotone but A_1 is not single valued, and $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is not paramonotone. Thus Theorem 8.24(ii) is optimal.

Theorem 8.29. *For every $i \in I$ suppose there exists $z \in \mathcal{H}$ such that $A_i \colon x \mapsto z_i$. Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) : x \mapsto \sum_{i\in I} \lambda_i z_i$.*

Proof. By Theorem 4.1(v) there exists $z_i \in \mathcal{H}$ such that $A_i \colon x \mapsto z_i$ if and only if J_{A_i} is an isometry, in which case $J_{A_i} \colon x \mapsto x - z_i$. Then,

$$\begin{aligned} J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}x &= \sum_{i\in I} \lambda_i J_{A_i} x \\ &= \sum_{i\in I} \lambda_i (x - z_i) = x - \sum_{i\in I} \lambda_i z_i. \end{aligned}$$

Thus $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})}$ is an isometry, so $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) : x \mapsto \sum_{i\in I} \lambda_i z_i$. ☐

8.3.3 k-cyclical monotonicity

Recall that for an operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$, A is *k-cyclically monotone* if for all $(x_1, u_1), \ldots, (x_k, u_k) \in \operatorname{gra} A$ and $x_{k+1} = x_1$ one has

$$\sum_{i=1}^{k} \langle u_i, x_{i+1} - x_i \rangle \leq 0. \tag{8.33}$$

The operator A is *cyclically monotone* if $\forall k \in \{2, 3, \ldots\}$, A is k-cyclically monotone.

Example 8.30. [5, Example 4.6] Let $\mathcal{H} = \mathbb{R}^2$ and let $n \in \{2, 3, \ldots\}$. Denote the matrix corresponding to the counter-clockwise rotation by π/n by R_n. That is,

$$R_n = \begin{pmatrix} \cos\frac{\pi}{n} & -\sin\frac{\pi}{n} \\ \sin\frac{\pi}{n} & \cos\frac{\pi}{n} \end{pmatrix}.$$

Then R_n is maximally monotone and n-cyclically monotone, but R_n is not $(n+1)$-cyclically monotone.

Lemma 8.31. *Let A and B be k-cyclically monotone operators from $\mathcal{H} \rightrightarrows \mathcal{H}$. Then*

(i) *αA is k-cyclically monotone for $\alpha > 0$.*

(ii) *$A + B$ is k-cyclically monotone;*

(iii) *A^{-1} is k-cyclically monotone.*

Proof. (i): Let $(x_i, u_i) \in \operatorname{gra} \alpha A$ for $i = 1, \dots, k+1$ with $x_{k+1} = x_1$. Then $(x_i, \alpha^{-1} u_i) \in \operatorname{gra} A$ for $i = 1, \dots, k+1$, and we have

$$\sum_{i=1}^{k} \langle x_{i+1} - x_i, u_i \rangle = \alpha \sum_{i=1}^{k} \left\langle x_{i+1} - x_i, \alpha^{-1} u_i \right\rangle \leq 0,$$

by the k-cyclical monotonicity of A. Thus αA is k-cyclically monotone.

(ii): Let $(x_i, u_i + v_i) \in \operatorname{gra}(A + B)$ for $i = 1, \dots, k+1$ with $x_{k+1} = x_1$, $(x_i, u_i) \in \operatorname{gra} A, (x_i, v_i) \in \operatorname{gra} B$. Since A and B are k-cyclic, by definition we have,

$$\sum_{i=1}^{k} \langle x_{i+1} - x_i, u_i \rangle \leq 0, \text{ and } \sum_{i=1}^{k} \langle x_{i+1} - x_i, v_i \rangle \leq 0.$$

Adding these two inequalities yields, $\sum_{i=1}^{k} \langle x_{i+1} - x_i, u_i + v_i \rangle \leq 0$. Thus $A + B$ is k-cyclic.

(iii): Let $(u_i, x_i) \in \operatorname{gra} A^{-1}$ for $i = 1, \dots, k$. Then $(\forall i)$ $(x_i, u_i) \in \operatorname{gra} A$. By the k-cyclical monotonicity of A, one can do the k-cyclical summation for points arranged in

$$(x_{k+1}, u_{k+1}) = (x_1, u_1), (x_k, u_k), (x_{k-1}, u_{k-1}), \cdots, (x_2, u_2),$$

to obtain that

$$\sum_{i=k+1}^{2} \langle x_{i-1} - x_i, u_i \rangle \leq 0 \quad \Leftrightarrow \quad \sum_{i=1}^{k} \langle x_i - x_{i+1}, u_{i+1} \rangle \leq 0. \tag{8.34}$$

Now

$$\begin{aligned} \sum_{i=1}^{k} \langle x_i - x_{i+1}, u_{i+1} \rangle &= \sum_{i=1}^{k} \langle x_i, u_{i+1} \rangle - \sum_{i=1}^{k} \langle x_{i+1}, u_{i+1} \rangle \\ &= \sum_{i=1}^{k} \langle x_i, u_{i+1} \rangle - \sum_{i=1}^{k} \langle x_i, u_i \rangle = \sum_{i=1}^{k} \langle x_i, u_{i+1} - u_i \rangle, \end{aligned}$$

so (8.34) transpires to $\sum_{i=1}^{k} \langle x_i, u_{i+1} - u_i \rangle \leq 0$. Hence A^{-1} is k-cyclically monotone. $\square$

Fact 8.32. [5, Theorem 6.6] *Let $T : \mathcal{H} \to \mathcal{H}$. Then T is the resolvent of the maximal monotone and k-cyclic operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ if and only if T has full domain, T is firmly nonexpansive, and the mapping $Tx \mapsto x - Tx$ is k-cyclic, i.e., for every set of points $\{x_1, \ldots, x_k\}$, where $x_{k+1} = x_1$, one has*

$$\sum_{i=1}^{k} \langle x_i - Tx_i, Tx_i - Tx_{i+1} \rangle \geq 0.$$

Proposition 8.33. *Suppose that A_1 and A_2 are two maximal monotone and k-cyclical mappings from $\mathcal{H} \rightrightarrows \mathcal{H}$. Further, let $\alpha \in]0,1[$. Then there exists a k-cyclical monotone operator B such that*

$$(\mathrm{Id} + B)^{-1} = \alpha(\mathrm{Id} + A_1)^{-1} + (1-\alpha)(\mathrm{Id} + A_2)^{-1}. \tag{8.35}$$

Hence the set of resolvents

$$\{J_A : \ A \textit{ is } k\textit{-cyclically monotone}\},$$

is a convex set.

Proof. Set $T_1 = J_{A_1}$ and $T_2 = J_{A_2}$. Then T_1 and T_2 are firmly nonexpansive with full domain as they are the resolvents of maximally monotone operators. Let $\alpha \in]0,1[$, then $T := \alpha T_1 + (1-\alpha)T_2$ is firmly nonexpansive with full domain, and is thus the resolvent of a maximally monotone operator, B. To show that B is k-cyclically monotone, by Fact 8.32 we need to show that

$$\begin{aligned}\sum_{i=1}^{k} \langle x_i - (\alpha T_1 x_i + (1-\alpha)T_2 x_i), (\alpha T_1 x_i + (1-\alpha)T_2 x_i) \\ - (\alpha T_1 x_{i+1} + (1-\alpha)T_2 x_{i+1}) \rangle \geq 0.\end{aligned} \tag{8.36}$$

For each i we have,

$$\begin{aligned}
&\langle x_i - (\alpha T_1 x_i + (1-\alpha)T_2 x_i), (\alpha T_1 x_i + (1-\alpha)T_2 x_i) \\
&\qquad - (\alpha T_1 x_{i+1} + (1-\alpha)T_2 x_{i+1})\rangle \\
&= \alpha^2 \langle x_i - T_1 x_i, T_1 x_i - T_1 x_{i+1}\rangle + (1-\alpha)^2 \langle x_i - T_2 x_i, T_2 x_i - T_2 x_{i+1}\rangle \\
&\qquad + \alpha(1-\alpha) \langle x_i - T_1 x_i, T_2 x_i - T_2 x_{i+1}\rangle \\
&\qquad + \alpha(1-\alpha) \langle x_i - T_2 x_i, T_1 x_i - T_1 x_{i+1}\rangle \\
&= \big(\alpha^2 + \alpha(1-\alpha)\big) \langle x_i - T_1 x_i, T_1 x_i - T_1 x_{i+1}\rangle \\
&\qquad + \big((1-\alpha)^2 + \alpha(1-\alpha)\big) \langle x_i - T_2 x_i, T_2 x_i - T_2 x_{i+1}\rangle \\
&\qquad + \alpha(1-\alpha) \langle T_1 x_i - T_2 x_i, (T_1 x_i - T_1 x_{i+1}) - (T_2 x_i - T_2 x_{i+1})\rangle \\
&= \alpha \langle x_i - T_1 x_i, T_1 x_i - T_1 x_{i+1}\rangle + (1-\alpha) \langle x_i - T_2 x_i, T_2 x_i - T_2 x_{i+1}\rangle \\
&\qquad + \alpha(1-\alpha) \langle T_1 x_i - T_2 x_i, (T_1 x_i - T_1 x_{i+1}) - (T_2 x_i - T_2 x_{i+1})\rangle . \qquad (8.37)
\end{aligned}$$

By Fact 8.32,

$$\sum_{i=1}^{k} \langle x_i - T_1 x_i, T_1 x_i - T_1 x_{i+1}\rangle \geq 0 \text{ and } \sum_{i=1}^{k} \langle x_i - T_2 x_i, T_2 x_i - T_2 x_{i+1}\rangle \geq 0. \qquad (8.38)$$

Since Id is cyclically monotone, then any points $x_1, \ldots, x_k$ satisfy

$$\sum_{i=1}^{k} \langle x_i, x_i - x_{i+1}\rangle \geq 0,$$

where $x_{k+1} = x_1$. Thus,

$$\sum_{i=1}^{k} \langle T_1 x_i - T_2 x_i, (T_1 x_i - T_1 x_{i+1}) - (T_2 x_i - T_2 x_{i+1})\rangle \geq 0. \qquad (8.39)$$

Altogether, (8.37), (8.38), and (8.39) yield,

$$\begin{aligned}
\sum_{i=1}^{k} &\langle x_i - (\alpha T_1 x_i + (1-\alpha)T_2 x_i), (\alpha T_1 x_i + (1-\alpha)T_2 x_i) \\
&\qquad\qquad - (\alpha T_1 x_{i+1} + (1-\alpha)T_2 x_{i+1})\rangle \geq 0,
\end{aligned}$$

which is (8.36). The convexity of $C := \{J_A : \ A \text{ is } k\text{-cyclically monotone}\}$ then follows from induction. Clearly, if $n = 1$ then $J_{A_1} \in C$. Assume that

$\lambda_1 J_{A_1} + \cdots + \lambda_{n-1} J_{A_{n-1}} \in C$, then

$$\begin{aligned}
&\lambda_1 J_{A_1} + \cdots + \lambda_n J_{A_n}\\
&= \lambda_1 J_{A_1} + \cdots + \lambda_{n-1} J_{A_{n-1}} + \lambda_n J_{A_n}\\
&= (1-\lambda_n)\left(\frac{\lambda_1}{\lambda_1 + \cdots + \lambda_{n-1}} J_{A_1} + \cdots + \frac{\lambda_{n-1}}{\lambda_1 + \cdots + \lambda_{n-1}} J_{A_{n-1}}\right) + \lambda_n J_{A_n}\\
&= (1-\lambda_n) J_{\tilde{A}} + \lambda_n J_{A_n}
\end{aligned}$$

We know that $\tilde{A}$ is k-cylic by the induction assumption, thus apply (8.35) to get $\lambda_1 J_{A_1} + \cdots + \lambda_n J_{A_n} \in C$. □

Theorem 8.34 (k-cyclic monotonicity is recessive)**.** *For all $i \in I$, let A_i be maximal monotone and k-cyclic, then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is k-cyclic. In particular, $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is cyclic if each A_i is cyclic.*

Proof. By Theorem 8.33, $\mu\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is k-cyclic since,

$$J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n},$$

Apply Lemma 8.31(i) with $\alpha = \mu^{-1}$ to get $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is k-cyclic. □

To see that k-cyclic monotonicity is not dominant, we look at the following example.

Example 8.35. Let $\mathcal{H} = \mathbb{R}^2$ and set $\boldsymbol{\lambda} = (1/2, 1/2)$. Define

$$A_1 = \begin{pmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{pmatrix}.$$

By Example 8.30, A_1 is 2-cyclically monotone, but not 3-cyclically monotone and A_2 is 3-cyclically monotone. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is not 3-cyclically monotone, as can be verified using (8.33) and the points

$$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, x_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ and } x_4 = x_1.$$

The code that can be used to verify this example can be found in Appendix B.1.

8.3.4 Displacement mappings

Theorem 8.36. *Let T be a mapping from $\mathcal{H}$ to $\mathcal{H}$. Then $T \circ (2\,\mathrm{Id})$ is a displacement mapping, i.e.*

$$T \circ (2\,\mathrm{Id}) = \mathrm{Id} - N,$$

for some nonexpansive mapping $N : \mathcal{H} \to \mathcal{H}$ if and only if T is firmly nonexpansive.

Proof. Assume first that $T \circ (2\,\mathrm{Id}) = \mathrm{Id} - N$. Then

$$Tx = (\mathrm{Id} - N)(\frac{1}{2}x) = \frac{x}{2} - N(\frac{1}{2}x) = \frac{x - 2N(\frac{1}{2}x)}{2}.$$

As N is nonexpansive, we have $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})$ $\|Nx - Ny\| \leq \|x - y\|$. So,

$$\begin{aligned} &\left\| N\left(\frac{1}{2}x\right) - N\left(\frac{1}{2}y\right) \right\| \leq \left\| \frac{1}{2}x - \frac{1}{2}y \right\| \\ \Leftrightarrow &\left\| 2N\left(\frac{1}{2}x\right) - 2N\left(\frac{1}{2}y\right) \right\| \leq \|x - y\|. \end{aligned}$$

So $2N \circ \left(\frac{1}{2}\,\mathrm{Id}\right)$ is nonexpansive, hence by Fact 3.3(iii) T is firmly nonexpansive.

Conversely, assume T is firmly nonexpansive. Consider

$$N = (\mathrm{Id} - T \circ (2\,\mathrm{Id})),$$

we will show N is nonexpansive.

$$\begin{aligned} \|Nx - Ny\|^2 &= \|(x - T(2x)) - (y - T(2y))\|^2 \\ &= \|(x - y) - (T(2x) - T(2y))\|^2 \\ &= \|x - y\|^2 - 2\langle x - y, T(2x) - T(2y)\rangle + \|T(2x) - T(2y)\|^2 \\ &= \|x - y\|^2 - \big(\langle 2x - 2y, T(2x) - T(2y)\rangle - \|T(2x) - T(2y)\|^2\big) \\ &\leq \|x - y\|^2, \end{aligned}$$

since T is firmly nonexpansive. Thus N is nonexpansive and

$$T \circ (2\,\mathrm{Id}) = \mathrm{Id} - N.$$

$\square$

Theorem 8.37. *Let $T : \mathcal{H} \to \mathcal{H}$ be a displacement mapping, i.e. $T = \operatorname{Id} - N$ for some nonexpansive mapping $N : \mathcal{H} \to \mathcal{H}$. Then $T = 2J_A$ for some monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$.*

Proof. By Fact 3.3(iii) and Fact 3.36, $N = 2J_A - \operatorname{Id}$ for some monotone operator, A. Then using the resolvent identity we have,

$$T = \operatorname{Id} - (2J_A - \operatorname{Id}) = 2(\operatorname{Id} - J_A) = 2J_{A^{-1}}.$$

□

Theorem 8.38. *$A : \mathcal{H} \rightrightarrows \mathcal{H}$ is $\frac{1}{2}$-strongly monotone if and only if $A^{-1} = \operatorname{Id} - N$ for some nonexpansive mapping, i.e. A^{-1} is a displacement mapping.*

Proof. Assume A is $\frac{1}{2}$-strongly monotone. Then $A = \frac{1}{2}\operatorname{Id} + B$ for some monotone operator B, and we have

$$\begin{aligned} A^{-1} &= (\frac{1}{2}\operatorname{Id} + B)^{-1} = \left(\frac{1}{2}(\operatorname{Id} + 2B)\right)^{-1} \\ &= (\operatorname{Id} + 2B)^{-1} \circ (2\operatorname{Id}) = J_{2B} \circ (2\operatorname{Id}). \end{aligned}$$

Thus by Theorem 8.36, A^{-1} is a displacement mapping.

On the other hand, assume A^{-1} is a displacement mapping. Then by Theorem 8.37, $A^{-1} = \operatorname{Id} - N = 2J_B$ for some monotone operator B and

$$\begin{aligned} A^{-1} = 2(\operatorname{Id} + B)^{-1} &\Leftrightarrow A = (\operatorname{Id} + B) \circ (\frac{1}{2}\operatorname{Id}) = \frac{1}{2}\operatorname{Id} + B\left(\frac{1}{2}\operatorname{Id}\right) \\ &\Leftrightarrow B\left(\frac{1}{2}\operatorname{Id}\right) = A - \frac{1}{2}\operatorname{Id}. \end{aligned}$$

Since B is monotone, A is $\frac{1}{2}$-strongly monotone. □

Theorem 8.39. *Assume that for all $i \in I$, A_i is a displacement mapping, i.e. $A_i = \operatorname{Id} - N_i$ for some nonexpansive N_i. Then*

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = (\lambda_1 J_{A_1} + \cdots + \lambda_n J_{A_n})^{-1} - \operatorname{Id},$$

is a displacement mapping, i.e. $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \operatorname{Id} - N$ for some nonexpansive mapping N.

Proof. Using Corollary 5.9,

$$J_{(\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))^{-1}} = \lambda_1 J_{A_1^{-1}} + \cdots + \lambda_n J_{A_n^{-1}}.$$

By Theorem 8.38, A_i^{-1} is $\frac{1}{2}$-strongly monotone for all $i \in I$, and therefore by Theorem 4.1(xi) $J_{A_i^{-1}}$ is $(1+\frac{1}{2})$-firmly nonexpansive. By Theorem 8.16, $J_{(\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))^{-1}} = \sum_{i=1}^{n} \lambda_i J_{A_i^{-1}}$ is $(1+\frac{1}{2})$-firmly nonexpansive. Then Theorem 4.1(xi) gives $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})^{-1}$ is $\frac{1}{2}$-strongly monotone and thus Theorem 8.38 yields that $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is a displacement mapping. □

8.3.5 Nonexpansive monotone operators

Theorem 8.40 (nonexpansiveness is recessive)**.** *For all $i \in I$, let A_i be a nonexpansive monotone mapping, i.e. $A_i = 2T_i - \mathrm{Id}$ and $T_i = J_{B_i}$ for some monotone operator B_i. Then $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is nonexpansive and*

$$\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) = 2T - \mathrm{Id},$$

where $T = J_B$, $B = \sum_{i=1}^{n} \lambda_i B_i$, and B_i is nonexpansive for all $i \in I$.

Proof. We have

$$J_{A_i} = (\mathrm{Id} + 2J_{B_i} - \mathrm{Id})^{-1} = (2J_{B_i})^{-1} = (\mathrm{Id} + B_i) \circ (\frac{1}{2}\,\mathrm{Id}). \tag{8.40}$$

Thus $J_{A_i} \circ (2\,\mathrm{Id}) = \mathrm{Id} + B_i$, and using Theorem 8.36

$$B_i = -N_i, \tag{8.41}$$

for some nonexpansive mapping N_i. So we have

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} = \sum_{i=1}^{n} \lambda_i J_{A_i} = \sum_{i=1}^{n} \lambda_i (\mathrm{Id} + B_i) \circ (\frac{1}{2}\,\mathrm{Id}) = \sum_{i=1}^{n} \lambda_i (\mathrm{Id} - N_i) \circ (\frac{1}{2}\,\mathrm{Id}) \tag{8.42}$$

Or, $J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} \circ (2\,\mathrm{Id}) = \sum_{i=1}^{n} \lambda_i (\mathrm{Id} - N_i)$. On the other hand, by Theorem 8.36

$$J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} \circ (2\,\mathrm{Id}) = \mathrm{Id} - N. \tag{8.43}$$

for some nonexpansive N. Then we have $N = \sum_{i=1}^{n} \lambda_i N_i$.

We also have, from (8.43)

$$(\mathrm{Id} + \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}))^{-1} = (\mathrm{Id} - N) \circ (\frac{1}{2}\,\mathrm{Id})$$

$$\Leftrightarrow \mathrm{Id} + \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) = \left((\mathrm{Id} - N) \circ (\frac{1}{2}\,\mathrm{Id}) \right)^{-1} = 2(\mathrm{Id} - N)^{-1} \tag{8.44}$$

$$\Leftrightarrow \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) = 2(\mathrm{Id} - N)^{-1} - \mathrm{Id}\,. \tag{8.45}$$

But $N = \sum_{i=1}^{n} \lambda_i N_i = -\sum_{i=1}^{n} \lambda_i B_i$, thus we have

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = 2(\mathrm{Id} + \sum_{i=1}^{n} \lambda_i B_i)^{-1} - \mathrm{Id} = 2J_B - \mathrm{Id},$$

where $B = \sum_{i=1}^{n} \lambda_i B_i$. □

Lemma 8.41. *Let A be a real 2×2 matrix of the form $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Then A is a rotation matrix if and only if $A^T = A^{-1}$, i.e. A is an orthogonal matrix.*

Proof. First, assume $A^T = A^{-1}$ and $A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Then $A^T A = \mathrm{Id} \Rightarrow a^2 + b^2 = 1$, i.e. a and b lie on the unit circle. Therefore, converting to polar coordinates using $a = \cos\theta$ and $b = \sin\theta$ gives that A is a rotation matrix.

On the other hand, assume A is a rotation matrix, then

$$A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix},$$

and

$$A^T A = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, $A^T = A^{-1}$. □

Example 8.42 (2×2 rotation matrices)**.** Let A_α and A_θ be the 2×2 rotation matrices,

$$A_\alpha = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \quad A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

with $\alpha, \theta \in [-\pi/2, \pi/2]$. Then using mathematical software, it is easy to verify that $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is a matrix of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, with $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})^T = \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})^{-1}$ (see Appendix B.2), thus by Lemma 8.41, $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is a rotation matrix.

Remark 8.43. Example 8.42 shows that the resolvent average of two rotators is also a rotator. The same is not true of the arithmetic average. Consider,

$$A = \lambda A_\alpha + (1-\lambda)A_\theta = \begin{bmatrix} \lambda\cos\alpha + (1-\lambda)\cos\theta & -\lambda\sin\alpha - (1-\lambda)\sin\theta \\ \lambda\sin\alpha + (1-\lambda)\sin\theta & \lambda\cos\alpha + (1-\lambda)\cos\theta \end{bmatrix}.$$

Then we have

$$A^TA = \begin{bmatrix} 2(\lambda-\lambda^2)\cos(\alpha-\theta) + 2\lambda^2 - 2\lambda + 1 & 0 \\ 0 & 2(\lambda-\lambda^2)\cos(\alpha-\theta) + 2\lambda^2 - 2\lambda + 1 \end{bmatrix}.$$

By Lemma 8.41, A is a rotator if and only if $A^TA = \mathrm{Id}$. This implies that $\cos(\alpha-\theta) = 1$. That is, $\alpha = \theta + 2k\pi$ for $k = 0, 1, \ldots$. So the arithmetic average of two rotation matrices only produces another rotation matrix under very specific circumstances.

Remark 8.44. Although we can see that $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is a rotation matrix, even in very simple cases it is difficult to see the relationship between the original rotation matrices and the resulting rotation matrix. See Figure 8.1 to see how the angle of rotation varies with certain values of θ and α.

Theorem 8.45 (orthogonality is recessive)**.** *Let A_i be monotone orthogonal matrices for all $i \in I$. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is an orthogonal matrix, i.e. $A^{-1} = A^T$.*

Proof. By Theorem 5.19, the orthogonality of each A_i and Fact 2.6, we have

$$\begin{aligned}
(\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}))^{-1} &= \mathcal{R}_1(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}) \\
&= \mathcal{R}_1(\mathbf{A}^T, \lambda) \\
&= \left(\lambda_1(\mathrm{Id} + A_1^T)^{-1} + \cdots + \lambda_m(\mathrm{Id} + A_m^T)^{-1}\right)^{-1} - \mathrm{Id} \\
&= \left(\lambda_1(\mathrm{Id}^T + A_1^T)^{-1} + \cdots + \lambda_m(\mathrm{Id}^T + A_m^T)^{1}\right)^{-1} - \mathrm{Id} \\
&= \left(\lambda_1((\mathrm{Id} + A_1)^T)^{-1} + \cdots + \lambda_m((\mathrm{Id} + A_m)^T)^{-1}\right)^{-1} - \mathrm{Id} \\
&= \left(\lambda_1((\mathrm{Id} + A_1)^{-1})^T + \cdots + \lambda_m((\mathrm{Id} + A_m)^{-1})^T\right)^{-1} - \mathrm{Id} \\
&= \left(\left(\lambda_1(\mathrm{Id} + A_1)^{-1} + \cdots + \lambda_m(\mathrm{Id} + A_m)^{-1}\right)^T\right)^{-1} - \mathrm{Id} \\
&= \left(\left(\lambda_1(\mathrm{Id} + A_1)^{-1} + \cdots + \lambda_m(\mathrm{Id} + A_m)^{-1}\right)^{-1}\right)^T - \mathrm{Id} \\
&= \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})^T.
\end{aligned}$$

Thus $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is orthogonal. $\square$

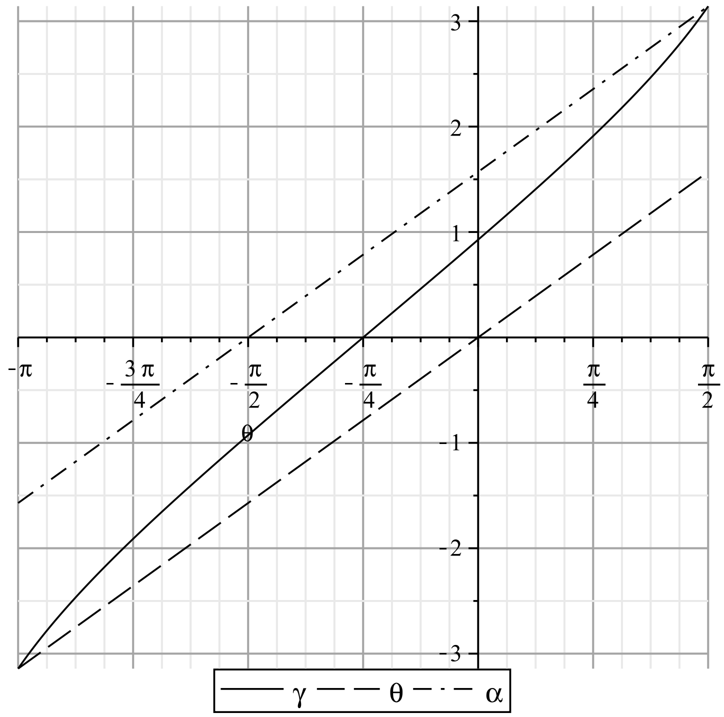

Figure 8.1: Resulting angle γ of the rotation of $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ of Example 8.42 with $A_\alpha = A_{\theta+\frac{\pi}{2}}$.

Remark 8.46. As you would expect, Theorem 8.45 only holds when $\mu = 1$. For example, take $A_1 = \mathrm{Id}$, A_2 be the 2×2 rotation by $\pi/2$, $\lambda_1 = \lambda_2 = \frac{1}{2}$ and $\mu = 2$, then $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is not orthogonal.

Remark 8.47 (Pythagorean triples). As an interesting aside, note that when we set $\alpha = \pi/2$ and $\theta = 0$ in Example 8.42 then any rational value for λ produces a pythagorean triple, i.e. three numbers x, y, and z such that $x^2 + y^2 = z^2$. We begin with the matrix,

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \begin{pmatrix} \frac{1-\lambda^2}{1+\lambda^2} & \frac{-2\lambda}{1+\lambda^2} \\ \frac{2\lambda}{1+\lambda^2} & \frac{1-\lambda^2}{1+\lambda^2} \end{pmatrix}.$$

Substituting $\lambda = \frac{a}{b}$, where a and b are integer values, we have

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = \begin{pmatrix} \frac{b^2-a^2}{a^2+b^2} & \frac{-2ab}{a^2+b^2} \\ \frac{2ab}{a^2+b^2} & \frac{b^2-a^2}{a^2+b^2} \end{pmatrix}.$$

Theorem 8.45 shows that $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ maintains orthogonality and by Example 8.42 $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is a rotation matrix. So the entries correspond to

$$\cos\beta = \frac{b^2-a^2}{a^2+b^2} \text{ and } \sin\beta = \frac{2ab}{a^2+b^2},$$

for some $\beta \in [0, \pi/2]$. Thus the angles formed in $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ are the angles in right triangles with all integer sides such that

$$\left(b^2-a^2\right)^2 + (2ab)^2 = \left(a^2+b^2\right)^2. \tag{8.46}$$

In fact, all possible triples can be generated this way. By [60, Theorem 11.1], all pythagorean triples can be generated by relatively prime integers m and n such that

$$x = m^2 - n^2, \quad y = 2mn, \quad z = m^2 + n^2,$$

then $x^2 + y^2 = z^2$, which is exactly (8.46). The code used to generate this example can be found in Appendix B.3.

Lemma 8.48. *For all $i \in I$, let A_i be monotone, and let some A_j be strongly monotone with constant $\beta \in \mathbb{R}_{++}$. Then $A = \sum_{i=1}^{n} \lambda_i A_i$ is strongly monotone with constant $\lambda_j \beta$.*

Proof. Since A_j is strongly monotone with constant β then $A_j - \beta\,\mathrm{Id}$ is monotone. Then $\sum_{i=1}^{n} \lambda_i A_i - \lambda_j \beta\,\mathrm{Id}$ is monotone, as it is the sum of monotone operators. Thus, A is strongly monotone with constant $\lambda_j\beta$. □

Lemma 8.49. *Let $\beta \in \mathbb{R}_{++}$. An operator A is strongly monotone with constant $\beta \Leftrightarrow A \circ \left(\frac{1}{2}\,\mathrm{Id}\right)$ is strongly monotone with constant $\frac{\beta}{2}$.*

Proof. Let $u \in A\left(\frac{1}{2}x\right)$ and $v \in A\left(\frac{1}{2}y\right)$, then

$$(x,u) \in \operatorname{gra} A \circ \left(\frac{1}{2}\,\mathrm{Id}\right) \text{ and } (y,v) \in \operatorname{gra} A \circ \left(\frac{1}{2}\,\mathrm{Id}\right).$$

As well,

$$(\frac{1}{2}x, u) \in \operatorname{gra} A \text{ and } (\frac{1}{2}y, v) \in \operatorname{gra} A.$$

Since A is strongly monotone with constant β, we have

$$\begin{aligned} \left\langle \frac{1}{2}x - \frac{1}{2}y, u - v \right\rangle &\geq \beta \| \frac{1}{2}x - \frac{1}{2}y \|^2 \\ \Leftrightarrow \frac{1}{2} \langle x - y, u - v \rangle &\geq \frac{\beta}{4} \|x - y\|^2 \\ \Leftrightarrow \langle x - y, u - v \rangle &\geq \frac{\beta}{2} \|x - y\|^2. \end{aligned}$$

Thus $A \circ \left(\frac{1}{2}\,\mathrm{Id}\right)$ is strongly monotone with constant $\frac{\beta}{2}$. $\square$

Lemma 8.50. *Let $\beta > 0$ and $\alpha \geq 1$. Then A is strongly monotone with constant $\beta \Leftrightarrow \alpha A$ is strongly monotone with constant $\alpha\beta$.*

Proof. Let (x, u) and $(y, v) \in \operatorname{gra} A$. Then $(x, \alpha u)$ and $(y, \alpha v) \in \operatorname{gra} \alpha A$. Since A is strongly monotone with constant β,

$$\langle x - y, \alpha u - \alpha v \rangle = \alpha \langle x - y, u - v \rangle \geq \alpha\beta \|x - y\|^2.$$

Thus αA is strongly monotone with constant $\alpha\beta$. $\square$

Lemma 8.51. *Assume A is both nonexpansive and strongly monotone with constant β. Then A^{-1} is strongly monotone with constant β.*

Proof. Let $(x, u) \in \operatorname{gra} A$ and $(y, v) \in \operatorname{gra} A$. Then $(u, x) \in \operatorname{gra} A^{-1}$ and $(v, y) \in \operatorname{gra} A^{-1}$. By the strong monotonicity and then nonexpasiveness of A, we have

$$\langle x - y, u - v \rangle \geq \beta \|x - y\|^2 \geq \beta \|u - v\|^2,$$

i.e. A^{-1} is strongly monotone with constant β. $\square$

Theorem 8.52. *Let $A_i = 2T_i - \mathrm{Id}$ be monotone for all $i \in I$ and $T_i = J_{B_i}$ for a monotone operator B_i. Additionally, assume some $A_j = 2T_j - \mathrm{Id}$ is a Banach contraction. Then $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is a Banach contraction.*

Proof. By Corollary 4.31, A_j is a Banach contraction $\Leftrightarrow B_j$ and B_j^{-1} are strongly monotone. Using Theorem 8.40,

$$\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) = 2J_B - \mathrm{Id},$$

where $B = \sum\limits_{i=1}^{n} \lambda_i B_i$. Setting $N = 2J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} - \mathrm{Id}$, we have

$$\begin{aligned} N &= 2(\mathrm{Id} + 2J_B - \mathrm{Id})^{-1} - \mathrm{Id} \\ &= 2(\mathrm{Id} + B) \circ \left(\frac{1}{2}\,\mathrm{Id}\right) - \mathrm{Id} \\ &= \mathrm{Id} + 2B \circ \left(\frac{1}{2}\,\mathrm{Id}\right) - \mathrm{Id} \\ &= 2\sum_{i=1}^{n} \lambda_i B_i \circ \left(\frac{1}{2}\,\mathrm{Id}\right). \end{aligned}$$

Combining Lemma 8.48, Lemma 8.49, and Lemma 8.50 we have N is strongly monotone. By Theorem 8.40, B_i is nonexpansive for all $i = 1, \dots, n$ therefore $\sum\limits_{i=1}^{n} \lambda_i B_i$ is nonexpansive and we get

$$\begin{aligned} &\|\sum_{i=1}^{n} \lambda_i B_i \circ \left(\frac{1}{2}x\right) - \sum_{i=1}^{n} \lambda_i B_i \circ \left(\frac{1}{2}y\right)\| \le \|\frac{x}{2} - \frac{y}{2}\| \\ \Leftrightarrow\ &\|2\sum_{i=1}^{n} \lambda_i B_i \circ \left(\frac{1}{2}x\right) - 2\sum_{i=1}^{n} \lambda_i B_i \circ \left(\frac{1}{2}y\right)\| \le \|x - y\|. \end{aligned}$$

Thus, N is both nonexpansive and strongly monotone, so by Lemma 8.51, N^{-1} is strongly monotone. Therefore by Corollary 4.31, $2J_N - \mathrm{Id}$ is a Banach contraction. Now,

$$\begin{aligned} 2J_N - \mathrm{Id} &= 2(\mathrm{Id} + 2J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})} - \mathrm{Id})^{-1} - \mathrm{Id} \\ &= 2(2J_{\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})})^{-1} - \mathrm{Id} \\ &= 2(\mathrm{Id} + \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})) \circ \left(\frac{1}{2}\,\mathrm{Id}\right) - \mathrm{Id} \\ &= 2\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}\,\mathrm{Id}\right). \end{aligned}$$

So $2\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}\,\mathrm{Id}\right)$ is a Banach contraction, i.e. there exists $\beta \in [0, 1[$ such that for all $x, y \in \mathcal{H}$,

$$\begin{aligned} &\|2\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}x\right) - 2\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}y\right)\| \le \beta\|x - y\| \\ \Leftrightarrow\ &\|\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}x\right) - \mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) \circ \left(\frac{1}{2}y\right)\| \le \beta\|\frac{x}{2} - \frac{y}{2}\|. \end{aligned}$$

Thus $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is a Banach contraction. $\square$

8.4 Indeterminate properties

We conclude this chapter with a couple of examples of properties that do not satisfy the definition of dominant or recessive.

Example 8.53 (projections are indeterminant)**.** Let A_1 and A_2 be the projections in $\mathbb{R}^2$ onto $\mathbb{R} \times \{0\}$ and $\{0\} \times \mathbb{R}$, respectively. That is,

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Then A_1 and A_2 are both projections, but

$$\begin{aligned} \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}) &= (\lambda J_{A_1} + (1-\lambda) J_{A_2})^{-1} - \mathrm{Id} \\ &= \begin{pmatrix} \frac{\lambda}{2-\lambda} & 0 \\ 0 & \frac{1-\lambda}{\lambda+1} \end{pmatrix}, \end{aligned}$$

is not a projection, since $(\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda}))^2 \neq \mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ unless $\lambda = 0$ or $\lambda = 1$.

Example 8.54 (normal cones are indeterminant)**.** Let $A_1 = N_{C_1}$ and $A_2 = N_{C_2}$ be the normal cones operators of $C_1 = \mathbb{R} \times \{0\}$ and $C_2 = \{0\} \times \mathbb{R}$. Then,

$$J_{A_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } J_{A_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

A_1 and A_2 are both normal cones, but $\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})$ is not a normal cone by Theorem 4.1(xx), since

$$J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})} = \begin{pmatrix} \lambda & 0 \\ 0 & (1-\lambda) \end{pmatrix},$$

and thus $\operatorname{ran} J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})} \neq \operatorname{Fix} J_{\mathcal{R}_1(\boldsymbol{A}, \boldsymbol{\lambda})}$.

Remark 8.55. In Examples 8.53 and 8.54 we have shown that the resolvent average is not necessarily a projection (or normal cone) even if all of the averaged operators are projections (normal cones).

The classifications of each of the properties considered in this chapter are summarized in Table 8.1 and Table 8.2.

Table 8.1: Summary of completely classified properties of the resolvent average.

Dominant Properties	**Recessive Properties**
Single-valuedness	Maximal monotonicity
Full domain	Linearity
Surjectivity	k-cyclic monotonicity
Strict monotonicity	Displacement mappings
Banach contraction	Orthogonality
Linear rectangularity	Nonlinear rectangularity
Linear paramonotonicty	Nonlinear paramonotonicity
	Nonexpansiveness

Table 8.2: Summary of incompletely classified and indeterminant properties of the resolvent average.

Dominant or Recessive Properties	**Indeterminant Properties**
Strong monotonicity	Projection operators
Cocoercivity	Normal cone operators

Chapter 9

Conclusion

9.1 Key results

This thesis has provided a comprehensive study of the relationship between maximally monotone operators and firmly nonexpansive mappings as well as defining a new method for averaging monotone operators. The key results presented are outlined below.

Theorem 4.1 lists the corresponding properties between maximally monotone operators and their associated resolvents. This theorem covers twenty-one properties of interest in monotone operator theory.

Definition 5.1 describes the resolvent average of monotone operators, a new average that maintains several desirable properties that the arithmetic average does not.

Theorem 5.3 demonstrates that the resolvent average is maximally monotone if and only if all of the averaged operators are maximally monotone. This is a stronger result than for the arithmetic average, which requires the additional constraint qualifications found in Fact 3.48.

The resolvent average satisfies the beautiful duality presented in Theorem 5.8,

$$(\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1}, \boldsymbol{\lambda}).$$

Theorem 5.14 develops an inequality between the arithmetic, resolvent, and harmonic averages for positive semidefinite matrices,

$$\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) \preceq \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \preceq \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}),$$

and shows the limits

$$\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}) \text{ when } \mu \to 0^+,$$

and

$$\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) \text{ when } \mu \to +\infty.$$

Theorem 6.28 provides results on the range of convex combinations of rectangular maximally monotone operators on $\mathbb{R}^n$.

Theorem 6.31 and Theorem 6.36 give results on convex combinations of firmly nonexpansive mappings in $\mathbb{R}^n$.

Theorem 7.14 shows that the composition of asymptotically regular mappings is again asymptotically regular in a Hilbert space.

Theorem 7.21 derives the asymptotic regularity of a convex combination of asymptotically regular mappings, extending Theorem 6.36 to a Hilbert space setting.

Chapter 8 classified properties of monotone operators and/or their resolvents as dominant, recessive, or indeterminant. Dominant properties include:

(i) single valuedness,

(ii) full domain,

(iii) surjectivity,

(iv) strict monotonicity,

(v) Banach contraction,

(vi) linear paramonotonicity (equivalently rectangularity).

Dominant or recessive properties are:

(i) γ-cocoercive, and

(ii) strong monotonicity.

Recessive properties are:

(i) maximal monotonicity,

(ii) linear relations,

(iii) rectangularity (except as noted above),

(iv) paramonotonicity (except as noted above),

(v) k-cyclic monotonicity,

(vi) displacement mappings, and

(vii) orthogonality.

Altogether, these results have expanded on the known theory regarding maximally monotone operators and firmly nonexpansive mappings.

9.2 Future work

Areas to consider for future research include specializing the inheritance properties of Chapter 8 to positive semidefinite matrices, similar to Section 5.2. Also, of the properties listed in Theorem 4.1, we have shown that strong monotonicity and being γ-cocoercive are at least recessive properties, but there is no proof or counterexample provided for dominance. Uniform monotonicity was also not classified as dominant, recessive, or indeterminant and will likely require strong constraint qualifications on the function ϕ in order to do so.

Example 8.42 also leaves room for future research. What is the relationship between the angle of rotation of the resolvent average of rotation matrices and the averaged matrices?

Finally, the broadest area of possible future research involves applications of the resolvent average. The work in the realm of positive semidefinite matrices has already been thoroughly cited in [43] and [45]. Are there other applications for the resolvent average in science and engineering?

Bibliography

[1] T. Ando, C.-K. Li, and R. Mathias, *Geometric means*, Linear Algebra Appl., 385 (2004), pp. 305–334. → pages 17

[2] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer-Verlag, 2003. → pages 35, 36

[3] J.-B. Baillon, R. E. Bruck, and S. Reich, *Asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston J. Math., 4 (1978), pp. 1–9. → pages 27, 28, 92, 154

[4] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones*, Israel J. Math., 26 (1977), pp. 137–150. → pages 26

[5] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, *Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative*, Nonlin. Anal., 66 (2007), pp. 1198–1223. → pages 34, 40, 124, 126

[6] H. H. Bauschke, *The composition of finitely many projections onto closed convex sets in Hilbert space is asymptotically regular.*, Proc. Am. Math. Soc., 56 (2003), pp. 715–738. → pages 105

[7] H. H. Bauschke and J. M. Borwein, *Dykstra's alternating projection algorithm for two sets*, J. Approx. Theory, 79 (1994), pp. 418–443. → pages 91

[8] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Comm. Contemp. Math, 3 (2000), pp. 615–647. → pages 14

[9] H. H. Bauschke, J. M. Borwein, and X. Wang, *Fitzpatrick functions and continuous linear monotone operators*, SIAM J. Optim., 18 (2007), pp. 789–809. → pages 35

[10] H. H. Bauschke and P.L. Combettes, *The Baillon-Haddad theorem revisited*, J. Convex Anal., 17 (2010), pp. 781–787. → pages 13, 26

[11] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011. → pages 1, 4, 6, 10, 11, 12, 13, 14, 15, 19, 24, 26, 27, 31, 32, 33, 35, 62, 96, 105, 107

[12] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, *The proximal average: Basic theory*, SIAM J. Optim., 19 (2008), pp. 768–785. → pages 18, 19, 21, 94

[13] H. H. Bauschke, Y. Lucet, and M. Trienis, *How to transform one convex function continuously into another*, SIAM Rev., 50 (2008), pp. 115–132. → pages 19

[14] H. H. Bauschke, Y. Lucet, and X. Wang, *Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators*, SIAM J. Control Optim., 46 (2007), pp. 2031–2051. → pages 18

[15] H. H. Bauschke, V. Martin-Marquez, S. M. Moffat, and X. Wang, *Compositions and convex combinations of asymptotically regular firmly nonexpansive mappings are also asymptotically regular*, Fixed Point Theory Appl., 53 (2012). → pages iii, 2, 95

[16] H. H. Bauschke, E. Matoušková, and S. Reich, *Projection and proximal point methods: convergence results and counterexamples*, Nonlin. Anal., 56 (2004), pp. 715–738. → pages 18

[17] H. H. Bauschke, D. A. McLaren, and H. S. Sendov, *Fitzpatrick functions: inequalities, examples, and remarks on a problem by S. Fitzpatrick*, J. Convex Anal., 13 (2006), pp. 499–523. → pages 34, 35

[18] H. H. Bauschke, S. M. Moffat, and X. Wang, *The resolvent average for positive semidefinite matrices*, Linear Algebra Appl., 432 (2010), pp. 1757–1771. → pages iii, 2, 64

[19] ——, *Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality*, Set-Valued Var. Anal., 20 (2012), pp. 131–153. → pages 1, 38

[20] ——, *Near equality, near convexity, sums of maximally monotone operators, and averages of firmly nonexpansive mappings*, Math. Program., 139 (2013), pp. 55–70. → pages 2, 35, 79, 81

[21] ——, *Inheritance of properties of the resolvent average of monotone operators*, preprint, (2014). → pages iii, 2, 64, 109

[22] H. H. Bauschke and X. Wang, *The kernel average for two convex functions and its applications to the extension and representation of monotone operators*, Trans. Am. Math. Soc., 319 (2009), pp. 5947–5965. → pages 18

[23] H. H. Bauschke, X. Wang, and C. J. S. Wylie, *Fixed points of averages of resolvents: geometry and algorithms*, SIAM J. Opt., 22 (2012), pp. 24–40. → pages 91

[24] H. H. Bauschke, X. Wang, and L. Yao, *Rectangularity and paramonotonicity of maximally monotone operators*, Optimization, (2012). → pages 35, 36

[25] T. D. Benavides, G. L. Acedo, and H.-K. Xu, *Constructions of sunny nonexpansive retractions in Banach spaces*, Bull. Austral. Math. Soc., 66 (2002), pp. 9–16. → pages 154, 155

[26] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, 2000. → pages 70

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. → pages 76

[28] H. Brézis and A. Haraux, *Image d'une somme d'operateurs monotones et applications*, Israel J. Math., 23 (1976), pp. 165–168. → pages 34, 35, 81

[29] F. E. Browder, *Convergence theorems for sequences of nonlinear operators in Banach spaces*, Math. Z., 100 (1967), pp. 201–225. → pages 27

[30] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. Math., 3 (1977), pp. 459–470. → pages 28

[31] R. S. Burachik and B. F. Svaiter, *Maximal monotone operators, convex functions, and a special family of enlargements*, Set-Valued Anal., 10 (2002), pp. 297–316. → pages 34

[32] N. L. Carothers, *A Short Course on Banach Space Theory*, Cambridge University Press, 2005. → pages 3, 153

[33] L.-J. Chu, *On the sum of monotone operators*, Mich. Math. J., 43 (1996), pp. 273–289. → pages 34

[34] F. H. Clark, R. J Stern, and G. Sabidussi, *Nonlinear analysis, differential equations, and control*, Springer, 1999. → pages 154

[35] P. L. Combettes, *Inconsistent signal feasibility problems: least-squares solutions in a product space*, IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966. → pages 91

[36] R. Cross, *Multivalued Linear Operators*, Marcel Dekker, 1998. → pages 8

[37] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer-Verlag, 2001. → pages 3, 5, 6, 7, 153

[38] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Prog. Ser. A, 55 (1992), pp. 293–318. → pages 32

[39] L. Elsner, I. Koltracht, and M. Neumann, *Convergence of sequential and asynchronous nonlinear paracontractions*, Numer. Math., 62 (1992), pp. 305–319. → pages 58

[40] S. Fitzpatrick, *Representing monotone operators by convex functions*, Proc. Centre Math. Anal. Austral. Nat. Univ., 20 (1988), pp. 59–65. → pages 34

[41] C. W. Groetsch, *Generalized Inverses of Linear Operators: Representation and Approximation*, Marcel Dekker, 1977. → pages 12

[42] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985. → pages 70

[43] S. Kim, J. Lawson, and Y. Lim, *The matrix geometric mean of parameterized, weighted arithmetic and harmonic means*, Linear Algebra Appl., 435 (2011), pp. 2114–2131. → pages 142

[44] F. Kubo and T. Ando, *Means of positive linear operators*, Math. Ann., 246 (1980), pp. 205–224. → pages 17

[45] S. Kum, *Resolvent average on second-order cone*, Taiwanese J. Math., 15 (2011), pp. 2733–2750. → pages 142

[46] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979. → pages 70

[47] J.-E. Martínez-Legaz and M. Théra, *A convex representation of maximal monotone operators*, J. Nonlinear Convex Anal., 2 (2001), pp. 243–247. → pages 34

[48] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000. → pages 4, 35

[49] G. J. Minty, *On the maximal domain of a 'monotone' function*, Mich. Math. J., 8 (1961), pp. 135–137. → pages 81

[50] ——, *Monotone (nonlinear) operators in Hilbert spaces*, Duke Math. J., 29 (1962), pp. 341–346. → pages 18, 32

[51] M. Moakher, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 735–747. → pages 17

[52] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299. → pages 18

[53] T. Pennanen, *On the range of monotone composite mappings*, J. Nonlinear Convex Anal., 2 (2001), pp. 193–202. → pages 34, 88

[54] D. Petz, *Means of positive matrices: geometry and a conjecture*, Ann. Math. Inform., 32 (2005), pp. 129–139. → pages 17

[55] H. Rådström, *An embedding theorem for spaces of convex sets*, Proc. Am. Math. Soc., 3 (1952), pp. 165–169. → pages 87

[56] S. Reich, *On the asymptotic behavior of nonlinear semigroups and the range of accretive operators*, J. Math. Anal. Appl., 79 (1981), pp. 113–126. → pages 103, 105

[57] S. Reich and I. Shafrir, *The asymptotic behavior of firmly nonexpansive mappings*, Proc. Am. Math. Soc., 101 (1987), pp. 246–250. → pages 29

[58] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970. → pages 1, 9, 14, 16, 17, 18

[59] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, 1998. → pages 1, 10, 15, 18, 30, 33, 58, 76, 80, 88

[60] K. H. Rosen, *Elementary Number Theory and its Applications*, Addison-Wesley, 1986. → pages 135

[61] S. Simons, *LC-functions and maximal monotonicity*, J. Nonlinear Convex Anal., 7 (2006), pp. 123–137. → pages 34

[62] E. H. Zarantonello, ed., *Contributions to Nonlinear Functional Analysis*, Academic Press, 1971, ch. Projections on convex sets in Hilbert space and spectral theory I. Projections on convex sets, pp. 237–341. → pages 27

[63] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co. Inc., 2002. → pages 12

Index

Appendices

Appendix A

Uniformly Convex Banach Spaces

A *Banach space* is a complete normed linear space, whereas a *Hilbert space* is a complete inner product space. Because each Hilbert space has a norm induced by its inner product, every Hilbert space is a Banach space. The *dual space* of an inner product space X is the set X^* of all bounded linear functionals on X. The dual space of a Hilbert space is isomorphic to the original space [37, Theorem 6.10]. That is, $\mathcal{H}^* = \mathcal{H}$.

Definition A.1. [32, Equation 11.1] A normed linear space X is *uniformly convex* if, for each $\epsilon > 0$, $\exists \delta = \delta(\epsilon) > 0$ such that

$$\|x\| < 1, \quad \|y\| < 1, \quad \|x - y\| > \epsilon \Rightarrow \left\| \frac{x+y}{2} \right\| \leq 1 - \delta.$$

Lemma A.2. *[parallelogram identity] Let $x, y \in \mathcal{H}$. Then*

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Proof.

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

□

Lemma A.3. *Every Hilbert space is uniformly convex.*

Proof. Let $x, y \in \mathcal{H}$ such that $\|x\| < 1$, $\|y\| < 1$ and $\|x - y\| \geq \epsilon$. By Lemma A.2 we have

$$\begin{aligned} \|x + y\|^2 &= 2\|x\|^2 + 2\|y\|^2 - \|x - y\|^2 \\ &\leq 4 - \epsilon^2. \end{aligned}$$

Set $\delta = 1 - \frac{1}{2}\sqrt{4-\epsilon^2}$, then

$$\|x+y\|^2 \leq 4-\epsilon^2 \Leftrightarrow \left\|\frac{x+y}{2}\right\| \leq 1-\delta.$$

Since the parallelogram identity holds for every Hilbert space, every Hilbert space is uniformly convex. □

Definition A.4. [34, pg. 126]A normed linear space X has a *uniformly Gâteaux differentiable* norm if for each $y \in X$ and each $\epsilon > 0$, there exists $\delta(\epsilon, y) > 0$ such that for every $x \in X$, $\|x\| = 1$, there is a continuous linear functional f_x on X and

$$\left|\frac{\|x+ty\| - \|x\|}{t} - f_x(y)\right| < \epsilon \text{ for all } 0 < t < \delta(\epsilon, y).$$

Fact A.5. [34, pg. 127] *Every Hilbert space has a uniformly Gâteaux differentiable norm.*

Fact A.6. [3, Theorem 1.2] *Let T be a nonexpansive mapping and let X be a uniformly convex Banach space with a weakly sequentially continuous duality map, then $(T^n x)_{n\in\mathbb{N}}$ converges weakly to a fixed point of T if and only if* $\operatorname{Fix} T \neq \varnothing$ *and T is weakly asymptotically regular.*

Fact A.7. [3, Corollary 2.2] *Let X be a Banach space and C be a closed convex subset of X. Let $U : C \to X$ be an averaged nonexpansive mapping. If X is uniformly convex, then* $\operatorname{Fix} U = \varnothing$ *if and only if* $\lim_{n\to\infty} |U^n x| = \infty$ *for all x in C.*

Definition A.8. [25] Let C be a nonempty closed convex subset of a Banach space X and let D be a nonempty subset of C. A *retraction* from C to D is a mapping $T : C \to D$ such that $Tx = x$ for all $x \in D$.

Definition A.9. [25] A retraction $T : C \to D$ is *sunny* if it satisfies the property

$T(Tx+\lambda(x-Tx)) = Tx$ for $x \in C$ and $\lambda > 0$ whenever $Tx+\lambda(x-Tx) \in C$.

A retraction $T : C \to D$ is sunny nonexpansive if it it both sunny and nonexpansive.

Fact A.10. [3, Corollary 2.3] *Let X be a Banach space and C be a closed convex subset of X. Let $U : C \to X$ be an averaged nonexpansive mapping. Suppose that the norm of X is uniformly Gâteaux differentiable while the* *norm of X^* is Fréchet differentiable. If C is a sunny nonexpansive retract of X, then for each $x \in C$*

$$\lim_{n\to\infty}(U^n x - U^{n+1}x) \to v,$$

where v is the element of least norm in $\overline{\operatorname{ran}(\operatorname{Id} - U)}$.

Remark A.11. Sunny nonexpansive retracts are unique, if they exist. If C is a nonempty closed convex subset of a Hilbert space $\mathcal{H}$ then the projection operator P_C is the sunny nonexpansive retraction [25].

Appendix B

Maple Code

The following sections provide the code that was used to verify examples. All code was run using Maplesoft's Maple 15 software.

B.1 Code to verify Example 8.35

> restart: with(LinearAlgebra):

> $A1 := \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$

$$\begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

> $A2 := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

> $Id := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

> A1 := subs(alpha = (1/2)*Pi, A1);

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

> A2 := subs(theta = (1/3)*Pi, A2);

$$\begin{bmatrix} 1/2 & -1/2\sqrt{3} \\ 1/2\sqrt{3} & 1/2 \end{bmatrix}$$

> R := simplify(MatrixInverse((1/2)*MatrixInverse(Id+A1)
+(1/2)*MatrixInverse(Id+A2))-Id);

$$\begin{bmatrix} -\frac{-4+\sqrt{3}}{8+\sqrt{3}} & -\frac{6+2\sqrt{3}}{8+\sqrt{3}} \\ \frac{6+2\sqrt{3}}{8+\sqrt{3}} & -\frac{-4+\sqrt{3}}{8+\sqrt{3}} \end{bmatrix}$$

> $x_1 := \begin{bmatrix} 0 \\ 0 \end{bmatrix};\ x_2 := \begin{bmatrix} 1 \\ 0 \end{bmatrix};\ x_3 := \begin{bmatrix} 0 \\ 1 \end{bmatrix};\ x_4 := x_1;$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

> $u_1 := Multiply(R, x_1); u_2 := Multiply(R, x_2); u_3 := Multiply(R, x_3);$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -\frac{-4+\sqrt{3}}{8+\sqrt{3}} \\ \frac{6+2\sqrt{3}}{8+\sqrt{3}} \end{bmatrix}$$

$$\begin{bmatrix} -\frac{6+2\sqrt{3}}{8+\sqrt{3}} \\ -\frac{-4+\sqrt{3}}{8+\sqrt{3}} \end{bmatrix}$$

> $simplify(sum((x_{i+1}[1] - x_i[1]) * u_i[1] + (x_{i+1}[2] - x_i[2]) * u_i[2], i = 1..3));\ evalf(\%);$

$$2\,\frac{-1+2\sqrt{3}}{8+\sqrt{3}}$$

$$0.5063889748$$

B.2 Code to verify Example 8.42

> restart: with(LinearAlgebra):

$> \ A1 := \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$

$$\begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

$> \ A2 := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$> \ Id := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$> \ JA1 := MatrixInverse(Id + A1)$

$$\begin{bmatrix} \frac{1+\cos(\alpha)}{1+2\cos(\alpha)+(\cos(\alpha))^2+(\sin(\alpha))^2} & \frac{\sin(\alpha)}{1+2\cos(\alpha)+(\cos(\alpha))^2+(\sin(\alpha))^2} \\ -\frac{\sin(\alpha)}{1+2\cos(\alpha)+(\cos(\alpha))^2+(\sin(\alpha))^2} & \frac{1+\cos(\alpha)}{1+2\cos(\alpha)+(\cos(\alpha))^2+(\sin(\alpha))^2} \end{bmatrix}$$

$> \ JA2 := MatrixInverse(Id + A2)$

$$\begin{bmatrix} \frac{1+\cos(\theta)}{1+2\cos(\theta)+(\cos(\theta))^2+(\sin(\theta))^2} & \frac{\sin(\theta)}{1+2\cos(\theta)+(\cos(\theta))^2+(\sin(\theta))^2} \\ -\frac{\sin(\theta)}{1+2\cos(\theta)+(\cos(\theta))^2+(\sin(\theta))^2} & \frac{1+\cos(\theta)}{1+2\cos(\theta)+(\cos(\theta))^2+(\sin(\theta))^2} \end{bmatrix}$$

$> \ R := simplify(MatrixInverse(lambda * JA1 + (1 - lambda) * JA2) - Id)$

$$[[-(\lambda+\cos(\theta)+\cos(\alpha)\cos(\theta)-\lambda\cos(\theta)+\lambda\cos(\alpha)-\lambda\cos(\alpha)\cos(\theta)$$
$$+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)+\lambda^2\sin(\alpha)\sin(\theta)-\lambda^2\big)$$
$$/\left(-\lambda\cos(\alpha)\cos(\theta)+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\cos(\theta)+\lambda\cos(\alpha)-1+\lambda\right.$$
$$\left.+\lambda^2\sin(\alpha)\sin(\theta)-\lambda^2-\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)\right),$$
$$-((-\lambda\sin(\alpha)-\lambda\sin(\alpha)\cos(\theta)-\sin(\theta)-\sin(\theta)\cos(\alpha)+\sin(\theta)\lambda$$
$$+\sin(\theta)\lambda\cos(\alpha))/\left(-\lambda\cos(\alpha)\cos(\theta)+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\cos(\theta)\right.$$
$$\left.+\lambda\cos(\alpha)-1+\lambda+\lambda^2\sin(\alpha)\sin(\theta)-\lambda^2-\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)\right))\big],$$
$$[(-\lambda\sin(\alpha)-\lambda\sin(\alpha)\cos(\theta)-\sin(\theta)-\sin(\theta)\cos(\alpha)+\sin(\theta)\lambda$$
$$+\sin(\theta)\lambda\cos(\alpha))/\left(-\lambda\cos(\alpha)\cos(\theta)+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\cos(\theta)\right.$$
$$\left.+\lambda\cos(\alpha)-1+\lambda+\lambda^2\sin(\alpha)\sin(\theta)-\lambda^2-\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)\right),$$
$$-(\lambda+\cos(\theta)+\cos(\alpha)\cos(\theta)-\lambda\cos(\theta)+\lambda\cos(\alpha)-\lambda\cos(\alpha)\cos(\theta)$$
$$+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)+\lambda^2\sin(\alpha)\sin(\theta)-\lambda^2/(-\lambda\cos(\alpha)\cos(\theta)$$
$$+\lambda^2\cos(\theta)\cos(\alpha)-\lambda\cos(\theta)+\lambda\cos(\alpha)-1+\lambda+\lambda^2\sin(\alpha)\sin(\theta)$$
$$-\lambda^2-\cos(\alpha)-\lambda\sin(\alpha)\sin(\theta)\big)\big]\big]$$

```
> R[1, 1]-R[2, 2];
```

$$0$$

```
> R[1, 2]+R[2, 1];
```

$$0$$

```
> simplify(Multiply(Transpose(R), R))
```

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

B.3 Code to verify Remark 8.47

Define $A1$, $A2$, Id and R as in Section B.2.

> $R2 := factor(simplify(subs([theta = 0, alpha = (1/2) * Pi], R)));$

$$\begin{bmatrix} -\frac{-1+\lambda^2}{1+\lambda^2} & -2\,\frac{\lambda}{1+\lambda^2} \\ 2\,\frac{\lambda}{1+\lambda^2} & -\frac{-1+\lambda^2}{1+\lambda^2} \end{bmatrix}$$

> $factor(simplify(subs(lambda = a/b, R2)));$

$$\begin{bmatrix} -\frac{-b^2+a^2}{b^2+a^2} & -2\,\frac{ab}{b^2+a^2} \\ 2\,\frac{ab}{b^2+a^2} & -\frac{-b^2+a^2}{b^2+a^2} \end{bmatrix}$$