Decompositions and representations of monotone operators with linear graphs

by

Liangjin Yao

M.Sc., Yunnan University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The College of Graduate Studies

(Interdisciplinary)

The University Of British Columbia Okanagan

December, 2007

Abstract

We consider the decomposition of a maximal monotone operator into the sum of an antisymmetric operator and the subdifferential of a proper lower semicontinuous convex function. This is a variant of the well-known decomposition of a matrix into its symmetric and antisymmetric part. We analyze in detail the case when the graph of the operator is a linear subspace. Equivalent conditions of monotonicity are also provided.

We obtain several new results on auto-conjugate representations including an explicit formula that is built upon the proximal average of the associated Fitzpatrick function and its Fenchel conjugate. These results are new and they both extend and complement recent work by Penot, Simons and Zălinescu. A nonlinear example shows the importance of the linearity assumption. Finally, we consider the problem of computing the Fitzpatrick function of the sum, generalizing a recent result by Bauschke, Borwein and Wang on matrices to linear relations.

Table of Contents

List of Figures

Acknowledgements

I sincerely thank Heinz Bauschke and Shawn Wang for giving me the opportunity to study with them under their supervision at UBC Okanagan. They have both given me a lot of help and encouragement. I would like to thank Philip Loewen and Yves Lucet for their many pertinent comments on my thesis.

I would also like to thank the many people in the department who have made UBCO into a pleasant environment.

Chapter 1

Introduction

This thesis addresses two issues: decompositions of a monotone operator with a linear graph, and auto-conjugate representations of a monotone operator with a linear graph.

It is well known that every matrix A in $\mathbb{R}^{n\times n}$ can be decomposed into the sum of a symmetric matrix and an antisymmetric matrix by

$$A = \frac{A+A^\intercal}{2} + \frac{A-A^\intercal}{2},$$

where $\frac{A+A^\intercal}{2}$ is a gradient of a quadratic function. Our goal is to decompose more general mappings, namely maximal monotone operators. Both positive semidefinite matrices and gradients of convex functions are maximal monotone.

At present there are two famous decompositions: Asplund decomposition and Borwein-Wiersma decomposition. In 1970, Asplund decomposition was introduced by E. Asplund who showed that a maximal monotone and at most single-valued operator A with $\operatorname{int}\operatorname{dom} A \neq \varnothing$ is Asplund decomposable. In 2006, J. Borwein and H. Wiersma introduced the Borwein-Wiersma decomposition in [12], which is more restrictive. Borwein-Wiersma verified that a maximal monotone operator that is Borwein-Wiersma decomposable is also Asplund decomposable in finite dimensional spaces. One goal of our thesis is to show that maximal monotone operators with linear graphs are Borwein-Wiersma decomposable.

The idea of representing a set by nice functions is classical. For example, for a closed set C, one can define the distance function

$$\mathrm{d}_C(x) := \inf_{c \in C} \Big\{ \|x - c\| \Big\}.$$

Then d_C is 1-Lipschitz and

$$C = \Big\{ x \mid \mathrm{d}_C(x) = 0 \Big\}.$$

One can also define the lower semicontinuous function ι_C where $\iota_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise, then

$$C = \Big\{ x \mid \iota_C(x) = 0 \Big\}.$$

In the later part of this thesis, we want to find auto-conjugate representations for a maximal monotone operator with a linear graph. An auto-conjugate function is a convex function. One very important result provided by Penot, Simons and Zălinescu shows that an auto-conjugate f represents an induced maximal monotone operator $A = G(f)$.

In order to create auto-conjugate representations, we introduce the Fitzpatrick function and the proximal average. In 1988, Simon Fitzpatrick defined a new convex function F_A in [18] which is called the Fitzpatrick function associated with the monotone operator A. Recently, Fitzpatrick functions have turned out to be a very useful tool in the study of maximal monotone operators, see [2, 4, 8, 9, 18]. The proximal average was first introduced in [6] in the context of fixed point theory. In its simplest form, the proximal average is denoted here by $P(f_0, f_1)$, where f_0 and f_1 are proper lower semicontinuous and convex functions. The recent works in [5, 7, 9] give numerous properties that are very attractive to Convex Analysts.

Now we come back to our question. In [24], Penot and Zălinescu showed that a maximal monotone operator A can be represented by an auto-conjugate function h_{F_A}, using a partial epigraphical sum. In [9], Bauschke and Wang showed that $P(F_A, F_A^{*\intercal})$ is an auto-conjugate representation for a maximal monotone operator A. Until now there has been no clear formula for $P(F_A, F_A^{*\intercal})$, even if A is a linear, continuous and monotone operator. In this thesis, we give an explicit formula for $P(F_A, F_A^{*\intercal})$ associated with a maximal monotone operator A with a linear graph. We find that $P(F_A, F_A^{*\intercal}) = h_{F_A}$. This is a new result.

The thesis is organized as follows.

Chapter 2 contains some auxiliary and basic results on monotone operators, subdifferentials and Moore-Penrose inverses.

In Chapter 3, it is shown that the inverse of a linear and monotone operator is Borwein-Wiersma decomposable.

Chapter 4 contains our first main result: *A maximal monotone operator with a linear graph is Borwein-Wiersma decomposable.* In addition, the remainder of this chapter gives some equivalent conditions of monotonicity of operators with linear graphs.

Chapter 5 discusses auto-conjugate representations. We give *an explicit* *formula for* $P(F_A, F_A^{*\intercal})$ *associated with a linear and monotone operator* A, which is our second main result. Furthermore, we show that $P(F_A, F_A^{*\intercal}) = h_{F_A}$.

In Chapter 6, we give a specific example of a nonlinear monotone operator: $\partial(-\ln)$ such that $P(F_{\partial(-\ln)}, F_{\partial(-\ln)}^{*\intercal}) \neq h_{F_{\partial(-\ln)}}$. This illustrates the necessity of the linearity assumption.

Finally, in Chapter 7 we extend auto-conjugate representation results from linear and monotone operators to monotone operators with linear graphs. Here we also discuss one open question: Expressing F_{A+B} in terms of F_A and F_B. We show that $F_{A+B} = F_A \Box_2 F_B$ (Here $\Box_2$ means the inf convolution for the second variable). This generalizes one of the results provided by Bauschke, Borwein and Wang in [4].

Throughout this thesis, X denotes a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, norm $\| \cdot \|$, and Id is the identity mapping in X. The unit ball is

$$B = \Big\{ x \in X \mid \|x\| \leq 1 \Big\}.$$

We further set

$$\mathbb{R}_+ = \Big\{ x \in \mathbb{R} \mid x \geq 0 \Big\}, \quad \mathbb{R}_- = \Big\{ x \in \mathbb{R} \mid x \leq 0 \Big\},$$
$$\mathbb{R}_{++} = \Big\{ x \in \mathbb{R} \mid x > 0 \Big\}, \quad \mathbb{R}_{--} = \Big\{ x \in \mathbb{R} \mid x < 0 \Big\}.$$

For a subset $C \subset X$, the closure of C is denoted by $\overline{C}$. The arrow "$\rightarrow$" is used for a single-valued mapping, whereas "$\rightrightarrows$" denotes a set-valued mapping.

Chapter 2

Auxiliary results

2.1 Definitions

We first introduce some fundamental definitions.

Definition 2.1.1 *Let* $T\colon X \rightrightarrows X$*. We say* T *is* monotone *if*

$$\langle x^* - y^*, x - y\rangle \geq 0,$$

whenever $x^* \in T(x),\ y^* \in T(y)$.

Definition 2.1.2 *Let* $A\colon X \to X$*. We say* A *is* positive semidefinite *if* $\langle x, Ax\rangle \geq 0,\ \forall x \in X$.

Example 2.1.3 *Let* $A\colon X \to X$ *be linear. Then* A *is monotone, if and only if,* A *is positive semidefinite.*

Proof. *"⇒" For every* $x \in X$*, by monotonicity and linearity of* S *we have*

$$\langle Ax, x\rangle = \langle Ax - A0, x - 0\rangle \geq 0. \tag{2.1}$$

(2.1) *holds by* $A0 = 0$ *(since* A *is linear).*

"⇐" For every $x, y \in X$*, since* A *is positive semidefinite, we have*

$$\big\langle Ax - Ay, x - y \big\rangle = \big\langle A(x-y), x-y \big\rangle \geq 0. \tag{2.2}$$

■

Definition 2.1.4 *Let $A\colon X \to X$ be linear. We define q_A by*

$$q_A(x) = \tfrac{1}{2}\langle x, Ax \rangle, \quad \forall x \in X. \tag{2.3}$$

Definition 2.1.5 *Let $A\colon X \to X$ be linear and continuous. Then A^* is the unique linear and continuous operator satisfying*

$$\langle Ax,\ y \rangle = \langle x,\ A^* y \rangle, \quad \forall x, y \in X. \tag{2.4}$$

Remark 2.1.6 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear. Then $A^* = A^\intercal$.*

Definition 2.1.7 *Let $A\colon X \to X$ be linear and continuous. We say that A is* symmetric *if $A^* = A$.*

Remark 2.1.8 *Let $A\colon X \to X$ be linear and continuous. Then A is monotone $\Leftrightarrow$ A^* is monotone.*

Definition 2.1.9 *Let $A\colon X \to X$ be linear and continuous. We say that A is* antisymmetric *if $A^* = -A$.*

Remark 2.1.10 *Let $A\colon X \to X$ be linear and continuous. Then $\frac{A+A^*}{2}$ is symmetric and $\frac{A-A^*}{2}$ is antisymmetric.*

Definition 2.1.11 (Symmetric and antisymmetric part) *Let $A\colon X \to X$ be linear and continuous. Then $A_+ = \frac{1}{2}A + \frac{1}{2}A^*$ is the* symmetric part *of A, and $A_\circ = A - A_+ = \frac{1}{2}A - \frac{1}{2}A^*$ is the* antisymmetric part *of A.*

Remark 2.1.12 *Let $A : X \to X$ be linear and continuous. Then $q_A = q_{A_+}$.*

Proof. Let $x \in X$.

$$\begin{aligned}
2q_{A_+}(x) &= \langle A_+ x, x\rangle = \langle \tfrac{A^*+A}{2}x, x\rangle \\
&= \langle \tfrac{A^*x}{2}, x\rangle + \langle \tfrac{Ax}{2}, x\rangle = \langle \tfrac{x}{2}, Ax\rangle + \langle \tfrac{Ax}{2}, x\rangle \\
&= \langle Ax, x\rangle = 2q_A(x).
\end{aligned}$$

■

Here is a basic property.

Fact 2.1.13 *Let $A : X \to X$ be linear and continuous. Then A is monotone, if and only if, A_+ is monotone.*

Proof. By Example 2.1.3 and Remark 2.1.12. ■

Definition 2.1.14 *Let $T\colon X \rightrightarrows X$ be monotone. We call T* maximal monotone *if for every $(y, y^*) \notin \operatorname{gra} T$ there exists $(x, x^*) \in \operatorname{gra} T$ with $\langle x - y,\ x^* - y^*\rangle < 0$.*

Fact 2.1.15 *Let $A\colon X \rightrightarrows X$ be maximal monotone and $(x_0, x_0^*) \in X \times X$. Let $\widetilde{A}\colon X \rightrightarrows X$ such that $\operatorname{gra} \widetilde{A} = \operatorname{gra} A - (x_0, x_0^*)$ (i.e., a rigid translation of $\operatorname{gra} A$). Then $\widetilde{A}$ is maximal monotone.*

Proof. Follows directly from Definition 2.1.14. ■

Example 2.1.16 *Every continuous monotone operator $A\colon X \to X$ is maximal monotone.*

Proof. [26, Example 12.7]. ■

Let us introduce an essential result that will be used often.

Fact 2.1.17 *Let $A\colon X \to X$ be linear, continuous and monotone. Then* $\ker A = \ker A^*$ *and* $\overline{\operatorname{ran} A} = \overline{\operatorname{ran} A^*}$.

Proof. See [4, Proposition 3.1]. ■

Fact 2.1.18 *Let $A : X \to X$ be linear and continuous. Then*

$$q_A \text{ is convex} \Leftrightarrow A \text{ is monotone} \Leftrightarrow q_A(x) \geq 0, \quad x \in X,$$

and

$$\nabla q_A = A_+.$$

Proof. See [3, Theorem 3.6(i)]. ■

Definition 2.1.19 *Let $f\colon X \to \left]-\infty, +\infty\right]$. We say f is* proper lower semicontinuous and convex *if*

$$f(x_0) < +\infty, \quad \exists\, x_0 \in X,$$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \ \forall \lambda \in \left]0,1\right[, \ \forall x, y \in X,$$

$$\liminf_{y \to x} f(y) = \lim_{\delta \to 0^+} \inf f(x + \delta B) \geq f(x), \quad \forall x \in X.$$

Definition 2.1.20 *Let $f\colon X \to \left]-\infty, +\infty\right]$ be proper lower semicontinuous* *and convex. The* subdifferential mapping $\partial f\colon X \rightrightarrows X$ *is defined by*

$$x \mapsto \partial f(x) := \Big\{x^* \mid \langle x^*, y - x\rangle + f(x) \le f(y), \quad \forall y\Big\}.$$

One of motivations for studying monotone operators comes from the following Fact.

Fact 2.1.21 (Rockafellar) *Let $f\colon X \to\,]-\infty, +\infty]$ be proper lower semicontinuous and convex. Then ∂f is maximal monotone.*

Proof. See [28, page 113] or [31, Theorem 3.28]. ■

Fact 2.1.22 *Let $A\colon X \to X$ be linear and monotone. Then A is maximal monotone and continuous.*

Proof. See [23, Corollary 2.6 and Proposition 3.2.h]. ■

Definition 2.1.23 *For a set $S \subset X$, $\iota_S\colon X \to\,]-\infty, +\infty]$ stands for the* indicator function *defined by*

$$\iota_S(x) := \begin{cases} 0, & \text{if } x \in S; \\ +\infty, & \text{otherwise.} \end{cases}$$

Fact 2.1.24 *Suppose that S is a nonempty convex subset of X. Then ι_S is proper lower semicontinuous and convex, if and only if, S is closed.*

Proof. See [22, Example.(a)]. ■

Definition 2.1.25 *The space ℓ^2 consists of all sequences of real numbers*

$(\xi_1, \xi_2, \ldots)$ *for which*

$$\|(\xi_1, \xi_2, \ldots)\|_2 < \infty,$$

where

$$\|(\xi_1, \xi_2, \ldots)\|_2 := (\sum_{i=1}^{\infty} |\xi_i|^2)^{\frac{1}{2}},$$

and where

$$\langle \xi, \gamma \rangle = \sum_{i=1}^{\infty} \langle \xi_i, \gamma_i \rangle, \quad \forall \xi = (\xi_i)_{i=1}^{\infty},\ \gamma = (\gamma_i)_{i=1}^{\infty} \in \ell^2.$$

Fact 2.1.26 $(\ell^2,\ \|\cdot\|_2)$ *is a Hilbert space.*

Proof. See [27, Example 3.24]. ■

Example 2.1.27 *Let* X *be* $(\ell^2,\ \|\cdot\|_2)$ *space and* $A\colon X \to X\colon (x_n)_{n=1}^{\infty} \mapsto (\frac{x_n}{n})_{n=1}^{\infty}$. *Then* A *is maximal monotone and continuous.*

Proof. Clearly, A is linear. Now we show A is monotone. Let $x = (x_n)_{n=1}^{\infty} \in X$. Then

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

By Example 2.1.3, A is monotone. By Fact 2.1.22, A is maximal monotone and continuous. ■

Lemma 2.1.28 *Let* S *be a linear subspace of* X. *Suppose* $x \in X$ *and* $\alpha \in \mathbb{R}$ *satisfy* $\langle x, s \rangle \leq \alpha, \forall s \in S$. *Then* $x \perp S$.

Proof. Let $s \in S$. By assumption, we have

$$\begin{aligned}\langle x,\ ks\rangle \le \alpha, \quad \forall k \in \mathbb{R} \ \Rightarrow & \langle x,\ s\rangle \le 0, \quad \text{if} \quad k > 0 \\ & \langle x,\ s\rangle \ge 0, \quad \text{if} \quad k < 0 \\ \Rightarrow & \ \langle x, s\rangle = 0, \quad \forall s \in S \\ \Rightarrow & \ x \perp S.\end{aligned}$$

■

Fact 2.1.29 *Suppose that S is a closed linear subspace of X. Then $\partial \iota_S(x) = S^{\perp}, \quad \forall x \in S$.*

Proof. Let $x \in S$. We have

$$\begin{aligned}x^* \in \partial \iota_S(x) &\Leftrightarrow \langle x^*, s - x\rangle \le \iota_S(s) - \iota_S(x), \quad \forall s \in X \\ &\Leftrightarrow \langle x^*, s - x\rangle \le 0, \quad \forall s \in S \\ &\Leftrightarrow x \perp S \quad \text{(by Lemma 2.1.28)}.\end{aligned}$$

■

Fact 2.1.30 *Let $f, g \colon X \to \left]-\infty, +\infty\right]$ be proper lower semicontinuous and convex. Suppose that f is differentiable everywhere. Then*

$$\partial(f + g)(x) = \nabla f(x) + \partial g(x), \quad \forall x \in \operatorname{dom} g.$$

Proof. See [22, Theorem 3.23]. ■

Example 2.1.31 *Suppose that $j(x) = \frac{1}{2}\|x\|^2$, $\forall x \in X$ and $S \subset X$ is a* *closed subspace. Then* $\partial(j + \iota_S)$ *is maximal monotone. In particular,* $\partial(j + \iota_S)(x) = x + S^{\perp}, \quad \forall x \in S$.

Proof. By Fact 2.1.24, ι_S is proper lower semicontinuous and convex. Hence $j + \iota_S$ is proper lower semicontinuous and convex. By Fact 2.1.21, $\partial(j + \iota_S)$ is maximal monotone.

Let $x \in S$. By Fact 2.1.30, Fact 2.1.18 and Fact 2.1.29,

$\partial(j + \iota_S)(x) = \nabla j(x) + \partial\iota_S(x) = \nabla q_{\mathrm{Id}}(x) + \partial\iota_S(x) = x + S^{\perp}$. ■

Fact 2.1.32 *Let* $A\colon X \to X$ *be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then* $\partial\iota_{\operatorname{ran} A}(x) = (\operatorname{ran} A)^{\perp} = \ker A^*, \ \forall x \in \operatorname{ran} A$.

Proof. Let $x \in \operatorname{ran} A$. By Fact 2.1.29, $\partial\iota_{\operatorname{ran} A}(x) = (\operatorname{ran} A)^{\perp}$. Now we show that $(\operatorname{ran} A)^{\perp} = \ker A^*$. We have

$$\begin{aligned} x^* \in (\operatorname{ran} A)^{\perp} &\Leftrightarrow \langle x^*, Ax\rangle = 0, \quad \forall x \in X \\ &\Leftrightarrow \langle A^*x^*, x\rangle = 0, \quad \forall x \in X \\ &\Leftrightarrow A^*x^* = 0 \Leftrightarrow x^* \in \ker A^*. \end{aligned}$$

■

Definition 2.1.33 *Let* $A\colon X \to X$. *The* set-valued inverse mapping, $A^{-1}\colon X \rightrightarrows X$, *is defined by*

$$x \in A^{-1}y \quad \Leftrightarrow \quad Ax = y.$$

The following is the definition of the Moore-Penrose inverse, which will play an important role in our Theorems.

Definition 2.1.34 *Let* $A\colon X \to X$ *be linear and continuous such that* $\operatorname{ran} A$ *is closed. The* Moore-Penrose inverse *of* A*, denoted by* $A^\dagger$*, is defined by*

$$A^\dagger b = \operatorname{argmin}_{A^*Au=A^*b}\|u\|, \quad \forall b \in X.$$

In the following we always let $A^\dagger$ stand for the Moore-Penrose inverse of a linear and continuous operator A.

Remark 2.1.35 *Let* $A\colon X \to X$ *be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then by [19, Theorem 2.1.1],*

$$A^\dagger y \in A^{-1}y, \quad \forall y \in \operatorname{ran} A.$$

In particular, if A *is bijective, then*

$$A^\dagger = A^{-1}.$$

2.2 Properties of $A^\dagger$

By the Remark above, we know that $A^\dagger\,|_{\operatorname{ran} A}$ is a selection for A^{-1}. This raises some questions: What is the relationship between them? If one of them is monotone, can we deduce that the other one is also monotone?

Fact 2.2.1 *Let* $A\colon X \to X$ *be linear and continuous. Then* $\operatorname{ran} A$ *is closed, if and only if,* $\operatorname{ran} A^*$ *is closed.*

Proof. See [19, Theorem 1.2.4]. ■

Fact 2.2.2 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then $A^\dagger$ is linear and continuous.*

Proof. See [19, Corollary 2.1.3]. ■

Fact 2.2.3 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then* $\operatorname{ran} A^\dagger = \operatorname{ran} A^*$.

Proof. See [19, Theorem 2.1.2]. ■

Fact 2.2.4 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then* $(A^\dagger)^* = (A^*)^\dagger$.

Proof. See [19, Exercise 11 on page 111] and [21, Exercise 5.12.16 on page 428]. ■

Proposition 2.2.5 *Let $A\colon X \to X$ be linear, continuous and monotone such that* $\operatorname{ran} A$ *is closed. Then*

$$\operatorname{ran} A = \operatorname{ran} A^* = \operatorname{ran} A^\dagger = \operatorname{ran}(A^\dagger)^*, \quad \ker A = \ker A^* = \ker(A^\dagger) = \ker(A^\dagger)^*.$$

Proof. By Fact 2.1.17 and Fact 2.2.1,

$$\operatorname{ran} A = \operatorname{ran} A^*. \tag{2.5}$$

By Fact 2.2.3 and (2.5), we have

$$\operatorname{ran} A = \operatorname{ran} A^* = \operatorname{ran} A^\dagger. \tag{2.6}$$

By Fact 2.2.1, $\operatorname{ran} A^*$ is closed. By Remark 2.1.8, A^* is monotone. Apply (2.6) with replacing A by A^*, we have

$$\operatorname{ran} A^* = \operatorname{ran} A^{**} = \operatorname{ran}(A^*)^\dagger. \tag{2.7}$$

By Fact 2.2.4 and (2.6), we have

$$\operatorname{ran} A^* = \operatorname{ran} A = \operatorname{ran}(A^\dagger)^* = \operatorname{ran} A^\dagger.$$

Then $(\operatorname{ran} A^*)^\perp = (\operatorname{ran} A)^\perp = \big(\operatorname{ran}(A^\dagger)^*\big)^\perp = (\operatorname{ran} A^\dagger)^\perp$, thus by Fact 2.1.32, $\ker A = \ker A^* = \ker(A^\dagger) = \ker(A^\dagger)^*$. ■

Proposition 2.2.6 *Let $A\colon X \to X$ be linear. Suppose $y \in \operatorname{ran} A$. Then $A^{-1}y = y^* + \ker A,\ \forall y^* \in A^{-1}y$.*

Proof. Let $y^* \in A^{-1}y$ and $z^* \in \ker A$. Then $Ay^* = y$ and

$$A(y^* + z^*) = Ay^* + Az^* = y + 0 = y.$$

Thus $y^* + z^* \in A^{-1}y$. Hence $y^* + \ker A \subset A^{-1}y$.

On the other hand, let $y_1^* \in A^{-1}y$. Then $Ay_1^* = y$ and for each $y^* \in A^{-1}y$,

$$A(y_1^* - y^*) = Ay_1^* - Ay^* = y - y = 0.$$

Thus $y_1^* - y^* \in \ker A$, i.e., $y_1^* \in y^* + \ker A$. Then $A^{-1}y \subset y^* + \ker A$. ■

Corollary 2.2.7 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then $A^{-1}y = A^\dagger y + \ker A, \quad \forall y \in \operatorname{ran} A$.*

Proof. By Remark 2.1.35 and Proposition 2.2.6. ∎

In order to further illustrate the relationship between $A^\dagger$ and A, we introduce the concept of *Projector*.

Fact 2.2.8 *Let M be a closed subspace of X. For every vector $x \in X$, there is a unique vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$. Furthermore, a necessary and sufficient condition that $m_0 \in M$ be the unique minimizing vector is that $x - m_0$ be orthogonal to M.*

Proof. See [20, Theorem 2 on page 51]. ∎

The Fact above ensures that the following mapping is well defined.

Definition 2.2.9 (Projector) *Let M be a closed subspace of X. The* Projector, *$P_M \colon X \to M$, is defined by*

$$P_M x = \operatorname{argmin}_{m \in M} \|x - m\| \,, \quad x \in X. \tag{2.8}$$

Here is a result that will be very helpful for our problems.

Proposition 2.2.10 *Let $A \colon X \to X$ be linear and monotone such that* $\operatorname{ran} A$ *is closed. Then $q_{A^\dagger} = q_{A^\dagger} P_{\operatorname{ran} A}$.*

Proof. Let $x \in X$. Then we have

$$2q_{A^\dagger}(x) = \langle x,\ A^\dagger x \rangle \tag{2.9}$$

$$= \langle P_{\operatorname{ran} A}x + P_{\ker A}x,\ A^\dagger(P_{\operatorname{ran} A}x + P_{\ker A}x) \rangle \tag{2.10}$$

$$= \langle P_{\operatorname{ran} A}x + P_{\ker A}x,\ A^\dagger(P_{\operatorname{ran} A}x) \rangle \tag{2.11}$$

$$= \langle P_{\operatorname{ran} A}x,\ A^\dagger(P_{\operatorname{ran} A}x) \rangle + \langle P_{\ker A}x,\ A^\dagger(P_{\operatorname{ran} A}x) \rangle \tag{2.12}$$

$$= \langle P_{\operatorname{ran} A}x,\ A^\dagger(P_{\operatorname{ran} A}x) \rangle + \langle P_{\operatorname{ran} A}x,\ (A^\dagger)^*(P_{\ker A}x) \rangle \tag{2.13}$$

$$= \langle P_{\operatorname{ran} A}x,\ A^\dagger(P_{\operatorname{ran} A}x) \rangle \tag{2.14}$$

$$= 2q_{A^\dagger}(P_{\operatorname{ran} A}x). \tag{2.15}$$

Note that (2.10) holds since $X = \operatorname{ran} A \oplus \ker A$ by Fact 2.1.32 and Fact 2.1.17.
(2.11) holds since $P_{\ker A}x \in \ker A = \ker A^\dagger$ by Proposition 2.2.5.
(2.14) holds by $(A^\dagger)^*(P_{\ker A}x) = 0$, since $\ker(A^\dagger)^* = \ker A$ by Proposition 2.2.5. ■

Fact 2.2.11 *Let $A \colon X \to X$ be linear and continuous such that $\operatorname{ran} A$ is closed. Then $AA^\dagger = P_{\operatorname{ran} A}$.*

Proof. See [19, Theorem 2.2.2]. ■

Corollary 2.2.12 *Let $A \colon X \to X$ be linear, continuous and monotone such that $\operatorname{ran} A$ is closed. Then $A^\dagger$ is monotone.*

Proof. Since $A^\dagger$ is linear and continuous by Fact 2.2.2, by Fact 2.1.18 it suffices to show that $q_{A^\dagger}(x) \geq 0, \forall x \in X$.

Let $x \in X$ and $y = A^\dagger(P_{\operatorname{ran} A}x)$. Then $Ay = AA^\dagger(P_{\operatorname{ran} A}x) = P_{\operatorname{ran} A}x$ by Fact 2.2.11. By Proposition 2.2.10, we have

$$\begin{aligned} 2q_{A^\dagger}(x) &= 2q_{A^\dagger}(P_{\operatorname{ran} A}x) && (2.16)\\ &= \langle A^\dagger(P_{\operatorname{ran} A}x),\ P_{\operatorname{ran} A}x\rangle && (2.17)\\ &= \langle y,\ Ay\rangle && (2.18)\\ &\geq 0, && (2.19)\end{aligned}$$

in which (2.19) holds since A is monotone. ■

Fact 2.2.13 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then $A^{\dagger\dagger} = A$.*

Proof. See [19, Exercise 7 on page 110]. ■

Theorem 2.2.14 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then A is monotone, if and only if, $A^\dagger$ is monotone.*

Proof. "$\Rightarrow$" By Corollary 2.2.12.
"$\Leftarrow$" Since $\operatorname{ran} A^\dagger = \operatorname{ran} A$ is closed by Proposition 2.2.5, we apply Fact 2.2.13 and Corollary 2.2.12 to $A^\dagger$ to conclude that $A^{\dagger\dagger} = A$ is monotone. ■

Here is a useful result that will be used very often.

Proposition 2.2.15 *Let $A\colon X \to X$ be linear, symmetric and continuous such that* $\operatorname{ran} A$ *is closed. Then*

$$q_{A^\dagger}(x + Ay) = q_{A^\dagger}(x) + q_A(y) + \langle P_{\operatorname{ran} A}x, y\rangle, \quad \forall x, y \in X.$$

Proof. Let $x \in X,\ y \in X$. Then

$$q_{A^\dagger}(x + Ay) \tag{2.20}$$

$$= \tfrac{1}{2}\langle A^\dagger x + A^\dagger Ay,\ x + Ay\rangle \tag{2.21}$$

$$= q_{A^\dagger}(x) + \tfrac{1}{2}\langle A^\dagger Ay, Ay\rangle + \tfrac{1}{2}\langle A^\dagger x, Ay\rangle + \tfrac{1}{2}\langle A^\dagger Ay, x\rangle \tag{2.22}$$

$$= q_{A^\dagger}(x) + \tfrac{1}{2}\langle AA^\dagger Ay, y\rangle + \tfrac{1}{2}\langle AA^\dagger x, y\rangle + \tfrac{1}{2}\big\langle y, (A^\dagger A)^* x\big\rangle \tag{2.23}$$

$$= q_{A^\dagger}(x) + \tfrac{1}{2}\langle P_{\operatorname{ran} A}(Ay), y\rangle + \tfrac{1}{2}\langle P_{\operatorname{ran} A} x, y\rangle + \tfrac{1}{2}\langle y, AA^\dagger x\rangle \tag{2.24}$$

$$= q_{A^\dagger}(x) + q_A(y) + \tfrac{1}{2}\langle P_{\operatorname{ran} A} x, y\rangle + \tfrac{1}{2}\langle y, P_{\operatorname{ran} A} x\rangle \tag{2.25}$$

$$= q_{A^\dagger}(x) + q_A(y) + \langle P_{\operatorname{ran} A} x, y\rangle, \tag{2.26}$$

in which, (2.24) by Fact 2.2.11 and Fact 2.2.4, (2.25) by Fact 2.2.11. ■

Corollary 2.2.16 *Let $A\colon X \to X$ be linear, symmetric and continuous such that* $\operatorname{ran} A$ *is closed. Then*

$$q_{A^\dagger}(Ax) = q_A(x), \quad \forall x \in X.$$

Proof. Apply Proposition 2.2.15 to A with x replaced by 0 and y replaced by x. ■

Fact 2.2.17 *Let $A\colon X \to X$ be linear and continuous such that* $\operatorname{ran} A$ *is closed. Then $A^\dagger A = P_{\operatorname{ran} A^\dagger}$.*

Proof. See [19, Theorem 2.2.2]. ■

Corollary 2.2.18 *Let* $A\colon X \to X$ *be linear, continuous and monotone such that* $\operatorname{ran} A$ *is closed. Then*

$$AA^{\dagger} = A^{\dagger}A = P_{\operatorname{ran} A}.$$

Proof. By Proposition 2.2.5, $\operatorname{ran} A = \operatorname{ran} A^{\dagger}$. Then follows directly from Fact 2.2.11 and Fact 2.2.17. ■

Chapter 3

Inverse of linear monotone operators

It is well known that a linear, continuous and monotone operator A can be decomposed into the sum of a symmetric operator A_+ and an antisymmetric operator $A_\circ$: $A = A_+ + A_\circ$.

By Fact 2.1.18, A is also decomposed into the sum of the subdifferential of a proper lower semicontinuous and convex function ∇q_A and an antisymmetric operator $A_\circ$: $A = \nabla q_A + A_\circ$. Such a decomposition is called a Borwein-Wiersma decomposition.

3.1 Borwein-Wiersma decomposition

Definition 3.1.1 (Borwein-Wiersma decomposition) *We say $A\colon X \rightrightarrows X$ is* Borwein-Wiersma decomposable *or simply* decomposable *if*

$$A = \partial f + S,$$

where f is proper lower semicontinuous and convex, and S is antisymmetric.

What kind of operators are Borwein-Wiersma decomposable?

Definition 3.1.2 *We say* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is* skew *if there exists a linear and antisymmetric operator* B *such that* $B\,|_{\operatorname{dom} A} = A\,|_{\operatorname{dom} A}$.

Fact 3.1.3 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be maximal monotone and at most single-valued. Suppose that* $0 \in \operatorname{dom} A, \operatorname{dom} A$ *is open and* A *is Frechet differentiable on* $\operatorname{dom} A$. *Then* A *is Borwein-Wiersma decomposable, if and only if,* $A - \nabla f_A$ *is skew, where*

$$f_A\colon\ \operatorname{dom} A \to \mathbb{R} : x \mapsto \int_0^1 \langle A(tx), x\rangle \,\mathsf{dt}.$$

Proof. See [12, Theorem 3]. ■

3.2 Asplund decomposition

Here we also introduce another famous decomposition: Asplund decomposition, see [1].

Definition 3.2.1 *We say* $A\colon X \rightrightarrows X$ *is* acyclic *with respect to a subset* C *if*

$$A = \partial f + S,$$

where f *is proper lower semicontinuous and convex, and* S *is monotone, which necessarily implies that* ∂f *is constant on* C. *If no set* C *is given, then* $C = \operatorname{dom} A$.

Definition 3.2.2 (Asplund decomposition) *We say* $A\colon X \rightrightarrows X$ *is* Asplund decomposable *if*

$$A = \partial f + S,$$

where f is proper lower semicontinuous and convex, and S is acyclic with respect to $\operatorname{dom} A$.

The following tells us which operators are Asplund decomposable.

Fact 3.2.3 (Asplund) *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{int}\operatorname{dom} A \neq \varnothing$ *and A is at most single-valued. Then A is Asplund decomposable.*

Proof. See [12, Theorem 13]. ■

By the following result, we can find out the connection between the decompositions.

Fact 3.2.4 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be antisymmetric. Then A is acyclic.*

Proof. See [12, Proposition 15]. ■

Remark 3.2.5 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone and Borwein-Wiersma decomposable via*

$$\partial f + S,$$

where f is proper lower semicontinuous and convex, and S is antisymmetric. Then such a decomposition is also an Asplund decomposition.

3.3 The Borwein-Wiersma decomposition of the inverse

As mentioned earlier, a linear, continuous and monotone operator is Borwein-Wiersma decomposable. It is natural to ask whether its set-valued inverse mapping is also Borwein-Wiersma decomposable.

Theorem 3.3.1 *Let $A\colon X \to X$ be linear, continuous and monotone such that* $\operatorname{ran} A$ *is closed. Then*

$$A^{-1} = \partial f + (A^{\dagger})_{\circ},$$

where $f := q_{A^{\dagger}} + \iota_{\operatorname{ran} A}$ is proper lower semicontinuous and convex, and $(A^{\dagger})_{\circ}$ is antisymmetric. In particular, A^{-1} is decomposable.

Proof. By Fact 2.2.2 and Corollary 2.2.12, $A^{\dagger}$ is linear, continuous and monotone. Then by Fact 2.1.18 we have $q_{A^{\dagger}}$ is convex function, differentiable and $\nabla q_{A^{\dagger}} = (A^{\dagger})_{+}$. Since $\operatorname{ran} A$ is a closed subspace of X, by Fact 2.1.24 $\iota_{\operatorname{ran} A}$ is proper lower semicontinuous and convex. Hence f is proper lower semicontinuous and convex.

We show that the convex function f satisfies

$$\partial f(x) + (A^{\dagger})_{\circ} x = \begin{cases} A^{\dagger} x + \ker A, & \text{if } x \in \operatorname{ran} A; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{3.1}$$

Indeed, since f is convex, $\forall x \in \operatorname{ran} A$ we have

$$\begin{aligned} \partial f(x) &= \partial(q_{A^{\dagger}} + \iota_{\operatorname{ran} A})(x) \\ &= \nabla q_{A^{\dagger}}(x) + \partial \iota_{\operatorname{ran} A}(x) \qquad (\text{ by Fact 2.1.30}) \\ &= (A^{\dagger})_{+} x + \ker A^{*} & (3.2) \\ &= (A^{\dagger})_{+} x + \ker A, & (3.3) \end{aligned}$$

where (3.2) holds by Fact 2.1.32, and (3.3) by Proposition 2.2.5.

Thus

$$\partial f(x) + (A^\dagger)_\circ x = (A^\dagger)_+ x + \ker A + (A^\dagger)_\circ x = A^\dagger x + \ker A, \ \forall x \in \operatorname{ran} A.$$

If $x \notin \operatorname{ran} A = \operatorname{dom} f$, by definition $\partial f(x) = \varnothing$. Hence (3.1) holds.

By Corollary 2.2.7, we have that

$$A^{-1}x = A^\dagger x + \ker A, \ \forall x \in \operatorname{ran} A. \tag{3.4}$$

Thus,

$$A^{-1}x = \begin{cases} A^\dagger x + \ker A, & \text{if } x \in \operatorname{ran} A; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{3.5}$$

By (3.1) and (3.5), we have $A^{-1} = \partial f + (A^\dagger)_\circ$. ■

Proposition 3.3.2 *Assume $T\colon X \rightrightarrows X$ is monotone, then T^{-1} is monotone. Moreover, if T is maximal monotone, then so too is T^{-1}.*

Proof. Use Definition 2.1.1 and Definition 2.1.14 directly. ■

Due to Phelps and Simons, we obtain the following Proposition.

Proposition 3.3.3 *Let $A\colon X \to X$ be linear, continuous and monotone such that A is one-to-one and symmetric. Then*

$$A^{-1} = \partial f,$$

where $f(x) := \sup_{y \in \operatorname{ran} A} \left\{ \langle A^{-1}y, x\rangle - \frac{1}{2}\langle A^{-1}y, y\rangle \right\}$ $(\forall x \in X)$ is proper lower *semicontinuous and convex. If $X = \mathbb{R}^n$, then $A^{-1} = \nabla q_{A^{-1}}$. In particular, A^{-1} is decomposable.*

Proof. By Example 2.1.16, A is maximal monotone. Then by Proposition 3.3.2, A^{-1} is maximal monotone. Since A is linear and one-to-one , A^{-1} is single-valued and linear on $\operatorname{ran} A$.

In the following we show that

$$\langle x, A^{-1}y\rangle = \langle y, A^{-1}x\rangle, \quad \forall x, y \in \operatorname{ran} A.$$

Let $x, y \in \operatorname{ran} A$. Then there exist unique $x_1, y_1 \in X$ such that $x = Ax_1, y = Ay_1$. We have

$$\langle x, A^{-1}y\rangle = \langle Ax_1, y_1\rangle = \langle x_1, Ay_1\rangle = \langle A^{-1}x, y\rangle.$$

By [23, Theorem 5.1], f is proper lower semicontinuous and convex, and $A^{-1} = \partial f$.

If $x = \mathbb{R}^n$, we have A is invertible. By assumption, A^{-1} is symmetric and monotone. By Fact 2.1.18, $A^{-1} = \nabla q_{A^{-1}}$. ■

Chapter 4

Monotone operators with linear graphs

Theorem 3.3.1 tells us that the set-valued inverse A^{-1} of a linear, continuous and monotone operator A is Borwein-Wiersma decomposable. Naturally, this raises the following question: Are maximal monotone operators with linear graphs also Borwein-Wiersma decomposable? This chapter answers the question above. It also gives some important equivalent conditions of monotonicity and maximal monotonicity of operators with linear graphs.

Let us first introduce some interesting results about these operators.

4.1 Linear graph

Fact 4.1.1 *Let S, M be closed linear subspaces of X. Then*

$$S = M \Leftrightarrow S^{\perp} = M^{\perp}, \quad S \neq M \Leftrightarrow S^{\perp} \neq M^{\perp}.$$

Proof. Follows directly by $S^{\perp\perp} = S, M^{\perp\perp} = M$. ■

Definition 4.1.2 *Let* $A\colon X \rightrightarrows X$. *We define* $\operatorname{dom} A, \operatorname{ran} A$ *by*

$$\operatorname{dom} A := \{x \mid Ax \neq \varnothing\}$$
$$\operatorname{ran} A := \{x^* \mid x^* \in Ax,\ \exists x \in \operatorname{dom} A\}.$$

Proposition 4.1.3 *Let* $A\colon X \rightrightarrows X$ *such that* $\operatorname{gra} A$ *is a linear subspace of* $X \times X$. *For every* $x, y \in \operatorname{dom} A$, *the following hold.*

(i) $A0$ *is a linear subspace of* X.

(ii) $Ax = x^* + A0, \quad \forall x^* \in Ax$.

(iii) $\alpha Ax + \beta Ay = A(\alpha x + \beta y), \quad \forall \alpha, \beta \in \mathbb{R}$ *with* $\alpha \neq 0$ *or* $\beta \neq 0$.

(iv) *If* A *is monotone, then* $\operatorname{dom} A \perp A0$, *hence* $\operatorname{dom} A \subset (A0)^{\perp}, A0 \subset (\operatorname{dom} A)^{\perp}$.

(v) *If* A *is monotone, then*

$$\langle x, x^* \rangle \geq 0, \quad \forall (x, x^*) \in \operatorname{gra} A.$$

Proof. Obviously,

$$\operatorname{dom} A = \Big\{x \in X \,|\, (x, y) \in \operatorname{gra} A, \quad \exists y \in X\Big\} \tag{4.1}$$

and $\operatorname{dom} A$ is a linear subspace of X.

(i): $\forall \alpha, \beta \in \mathbb{R},\ \forall x^*, z^* \in A0$ we have

$$(0, x^*) \in \operatorname{gra} A \quad (0, z^*) \in \operatorname{gra} A.$$

As $\operatorname{gra} A$ is a linear subspace of $X \times X$,

$$\alpha(0,\ x^*) + \beta(0,\ z^*) = (0,\ \alpha x^* + \beta z^*) \in \operatorname{gra} A.$$

This gives $\alpha x^* + \beta z^* \in A0$. Hence $A0$ is a linear subspace.

(ii): We first show that

$$x^* + A0 \subset Ax,\ \forall x^* \in Ax.$$

Take $x^* \in Ax,\ z^* \in A0$. Then

$$(x,\ x^*) \in \operatorname{gra} A \quad \text{and} \quad (0,\ z^*) \in \operatorname{gra} A.$$

Since $\operatorname{gra} A$ is a linear subspace,

$$(x,\ x^* + z^*) \in \operatorname{gra} A.$$

That is, $x^* + z^* \in Ax$. Then $x^* + A0 \subset Ax$.

On the other hand, let $x^*, y^* \in Ax$. We have

$$(x,\ x^*) \in \operatorname{gra} A,\ (x,\ y^*) \in \operatorname{gra} A.$$

Since $\operatorname{gra} A$ is a linear subspace,

$$(x - x,\ y^* - x^*) = (0,\ y^* - x^*) \in \operatorname{gra} A.$$

Then $y^* - x^* \in A0$. That is, $y^* \in x^* + A0$. Thus $Ax \subset x^* + A0$. Hence (ii) holds.

(iii): Let $\alpha, \beta \in \mathbb{R}$. Take $x^* \in Ax, y^* \in Ay$. Then we have

$$(x,\ x^*) \in \operatorname{gra} A,\ (y,\ y^*) \in \operatorname{gra} A.$$

Since $\operatorname{gra} A$ is a linear subspace, we have $(\alpha x + \beta y,\ \alpha x^* + \beta y^*) \in \operatorname{gra} A$. That is, $\alpha x^* + \beta y^* \in A(\alpha x + \beta y)$.

Then by (ii) we have

$$Ax = x^* + A0,\ Ay = y^* + A0,\ A(\alpha x + \beta y) = \alpha x^* + \beta y^* + A0. \tag{4.2}$$

Suppose that $\alpha \neq 0$. By (i)

$$\alpha A0 + \beta A0 = A0. \tag{4.3}$$

Then by (4.2) and (4.3),

$$\begin{aligned} \alpha Ax + \beta Ay &= \alpha(x^* + A0) + \beta(y^* + A0) \\ &= \alpha x^* + \beta y^* + (\alpha A0 + \beta A0) \\ &= \alpha x^* + \beta y^* + A0 \\ &= A(\alpha x + \beta y). \end{aligned}$$

(iv): Pick $x \in \operatorname{dom} A$. By (4.1) there exists $x^* \in X$ such that $(x, x^*) \in \operatorname{gra} A$. Then by monotonicity of A, we have

$$\langle x - 0,\ x^* - z^* \rangle \geq 0, \quad \forall z^* \in A0.$$

That is,

$$\langle x,\ x^*\rangle \geq \langle x, z^*\rangle, \quad \forall z^* \in A0. \tag{4.4}$$

Since $A0$ is a linear subspace by (i), by Lemma 2.1.28 and (4.4),

$$\begin{aligned} & x \perp A0, \quad \forall x \in \operatorname{dom} A \\ \Rightarrow\ & \operatorname{dom} A \perp A0 \\ \Rightarrow\ & \operatorname{dom} A \subset (A0)^{\perp}, \quad A0 \subset (\operatorname{dom} A)^{\perp}. \end{aligned}$$

(v): Since $(0,0) \in \operatorname{gra} A$,

$$\langle x, x^*\rangle = \langle x-0, x^*-0\rangle \geq 0, \quad \forall (x,x^*) \in \operatorname{gra} A.$$

■

Remark 4.1.4 *Proposition 4.1.3(ii) is a useful representation. It means*

$$Ax = \widetilde{A}x + A0, \quad \forall x \in \operatorname{dom} A, \widetilde{A}x \in Ax.$$

Later, we will show the selection map $\widetilde{A}$ can be chosen to be linear!

4.2 Maximal monotonicity identification

The next three results are well known.

Fact 4.2.1 *Let $A\colon X \rightrightarrows X$ be maximal monotone. Then Ax is closed and convex, $\forall x \in X$.*

Proof. Fix $x \in X$. If $x \notin \operatorname{dom} A$, then $Ax = \varnothing$ is closed and convex. So suppose $x \in \operatorname{dom} A$. Let $(x_n^*) \subset Ax$ such that $x_n^* \to x^*$. In the following we show that $x^* \in Ax$. For every $(y, y^*) \in \operatorname{gra} A$, by monotonicity of A, we have

$$\langle y - x,\ y^* - x_n^* \rangle \geq 0. \tag{4.5}$$

Letting $n \to \infty$ in (4.5), we see that

$$\langle y - x,\ y^* - x^* \rangle \geq 0, \quad \forall (y, y^*) \in \operatorname{gra} A. \tag{4.6}$$

By (4.6) and maximal monotonicity of A, we have $(x, x^*) \in \operatorname{gra} A$. That is, $x^* \in Ax$. Hence Ax is closed.

Now we show that Ax is convex. Let $\delta \in [0, 1]$. For every $x_1^*, x_2^* \in Ax$, we have

$$\langle y - x,\ y^* - x_1^* \rangle \geq 0 \tag{4.7}$$

$$\langle y - x,\ y^* - x_2^* \rangle \geq 0, \quad \forall (y, y^*) \in \operatorname{gra} A. \tag{4.8}$$

Adding (4.7)$\times\delta$ and (4.8)$\times(1 - \delta)$ yields

$$\langle y - x,\ y^* - (\delta x_1^* + (1 - \delta) x_2^*) \rangle \geq 0, \quad \forall (y, y^*) \in \operatorname{gra} A. \tag{4.9}$$

Since A is maximal monotone, $(x, \delta x_1^* + (1 - \delta) x_2^*) \in \operatorname{gra} A$, i.e., $\delta x_1^* + (1 - \delta) x_2^* \in Ax$. Thus Ax is convex. ■

Proposition 4.2.2 *Let $A, B\colon X \rightrightarrows X$ be monotone. Then $A + B$ is monotone.*

Proof. Let $(x, x^*), (y, y^*) \in \operatorname{gra}(A + B)$. Then there exist

$$(x, x_1^*), (y, y_1^*) \in \operatorname{gra} A$$
$$(x, x_2^*), (y, y_2^*) \in \operatorname{gra} B$$

such that

$$x^* = x_1^* + x_2^*$$
$$y^* = y_1^* + y_2^*.$$

Then

$$\begin{aligned}\langle x - y,\ x^* - y^*\rangle &= \langle x - y,\ x_1^* + x_2^* - y_1^* - y_2^*\rangle \\ &= \langle x - y,\ x_1^* - y_1^*\rangle + \langle x - y,\ x_2^* - y_2^*\rangle \geq 0.\end{aligned}$$

Hence $A + B$ is monotone. ■

Fact 4.2.3 *Let $A\colon X \rightrightarrows X$ be maximal monotone. Then $\overline{\operatorname{dom} A}$ is convex.*

Proof. See [31, Theorem 3.11.12]. ■

Fact 4.2.4 *Let $A\colon X \rightrightarrows X$ be maximal monotone. Then $A + \partial\iota_{\overline{\operatorname{dom} A}} = A$.*

Proof. By Fact 4.2.3 and Fact 2.1.24, $\iota_{\overline{\operatorname{dom} A}}$ is proper lower semicontinuous and convex. Then by Proposition 2.1.21, $\partial\iota_{\overline{\operatorname{dom} A}}$ is monotone. Then by Fact 4.2.2, $A + \partial\iota_{\overline{\operatorname{dom} A}}$ is monotone.

Suppose $x \in \operatorname{dom} A$. Then since $0 \in \partial\iota_{\overline{\operatorname{dom} A}}(x)$,

$$(A + \partial\iota_{\overline{\operatorname{dom} A}})(x) = Ax + \partial\iota_{\overline{\operatorname{dom} A}}(x) \supset Ax.$$

Suppose $x \notin \operatorname{dom} A$. Then since $Ax = \varnothing$,

$$(A + \partial\iota_{\overline{\operatorname{dom} A}})(x) \supset Ax, \quad \forall x \in X.$$

Since A is maximal monotone, $A + \partial\iota_{\overline{\operatorname{dom} A}} = A$. ■

The following are interesting properties about maximal monotonicity of monotone operators with linear graphs.

Proposition 4.2.5 *Let $A\colon X \rightrightarrows X$ be monotone such that* $\operatorname{gra} A$ *is a linear subspace of $X \times X$. Then*

$$A \textit{ is maximal monotone} \Rightarrow \overline{\operatorname{dom} A} = (A0)^{\perp}.$$

Proof. Suppose to the contrary that $\overline{\operatorname{dom} A} \neq (A0)^{\perp}$. By Proposition 4.1.3(i) and Fact 4.2.1, $A0$ is a closed subspace. By Fact 4.1.1, $(\overline{\operatorname{dom} A})^{\perp} \neq (A0)^{\perp\perp} = A0$. Then by Proposition 4.1.3(iv), we have

$$(\operatorname{dom} A)^{\perp} = (\overline{\operatorname{dom} A})^{\perp} \supsetneqq A0. \tag{4.10}$$

Thus there exists $\omega^* \in (\operatorname{dom} A)^{\perp} \setminus A0$. By $\omega^* \in (\operatorname{dom} A)^{\perp}$, we have

$$\langle \omega^*, x\rangle = 0, \quad \forall x \in \operatorname{dom} A. \tag{4.11}$$

Since $\omega^* \notin A0$, $(0, \omega^*) \notin \operatorname{gra} A$. By maximal monotonicity of A, there exists $(x_0, x_0^*) \in \operatorname{gra} A$ such that

$$\langle x_0^*, x_0 \rangle - \langle \omega^*, x_0 \rangle = \langle x_0^* - \omega^*, x_0 - 0 \rangle < 0. \tag{4.12}$$

By (4.11) and (4.12), $\langle x_0^*, x_0 \rangle < 0$, which is a contradiction to Proposition 4.1.3(v). ■

Proposition 4.2.6 *Let $A\colon X \rightrightarrows X$ be monotone such that* $\operatorname{gra} A$ *is a linear subspace of $X \times X$ and $A0$ is closed. Then*

$$\operatorname{dom} A = (A0)^{\perp} \Rightarrow A \textit{ is maximal monotone.}$$

Proof. Let $(x, x^*) \in X \times X$ satisfy that

$$\langle x - y,\ x^* - y^* \rangle \geq 0, \quad \forall (y, y^*) \in \operatorname{gra} A. \tag{4.13}$$

In the following we will verify that $(x, x^*) \in \operatorname{gra} A$.

By (4.13) we have

$$\langle x,\ x^* \rangle \geq \langle x,\ z^* \rangle, \quad \forall z^* \in A0.$$

Since $A0$ is a linear subspace by Proposition 4.1.3(i), by Lemma 2.1.28 we have $x \perp A0$, i.e., $x \in (A0)^{\perp} = \operatorname{dom} A$.

Take $x_0^* \in Ax$. For every $v^* \in Av$, we have $x_0^* + v^* \in A(x + v)$ by Proposition 4.1.3(iii). By (4.13), we have

$$\langle x-(x+v),\ x^*-(x_0^*+v^*)\rangle \geq 0, \quad \forall (v,v^*)\in \operatorname{gra} A.$$

That is,

$$\langle v,\ v^*\rangle \geq \langle v,\ x^*-x_0^*\rangle, \quad \forall (v,v^*)\in \operatorname{gra} A. \tag{4.14}$$

By Proposition 4.1.3(iii), we have $\frac{1}{n}v^*\in A(\frac{1}{n}v)$. Then by (4.14),

$$\langle \tfrac{1}{n}v,\ \tfrac{1}{n}v^*\rangle \geq \langle \tfrac{1}{n}v,\ x^*-x_0^*\rangle, \quad \forall (v,v^*)\in \operatorname{gra} A. \tag{4.15}$$

Multiply (4.15) both sides by n and then let $n\to\infty$ to see that

$$\langle v,\ x^*-x_0^*\rangle \leq 0, \quad \forall v\in \operatorname{dom} A. \tag{4.16}$$

Since $\operatorname{dom} A$ is a linear subspace, by Lemma 2.1.28, $(x^*-x_0^*)\perp \operatorname{dom} A$. Since $A0$ is closed, we have

$$(x^*-x_0^*)\in (\operatorname{dom} A)^{\perp} = (A0)^{\perp\perp} = A0. \tag{4.17}$$

According to Proposition 4.1.3(ii), $x^*\in x_0^*+A0=Ax$. ■

Here is an important result in this chapter.

Theorem 4.2.7 *Let $A\colon X\rightrightarrows X$ be monotone such that $\operatorname{gra} A$ is a linear subspace of $X\times X$ and $\operatorname{dom} A$ is closed. Then*

$$A \text{ is maximal monotone} \Leftrightarrow \operatorname{dom} A=(A0)^{\perp}, A0 \text{ is closed.}$$

Proof. Since A is maximal monotone, $A0$ is closed by Fact 4.2.1. Combine Proposition 4.2.5 and Proposition 4.2.6. ■

Theorem 4.2.7 gives an equivalent condition in infinite-dimensional spaces. When we consider it in finite-dimensional spaces, can we get further results? Now we discuss this in detail.

Proposition 4.2.8 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$. *Then* $\dim(\operatorname{gra} A) = \dim(\operatorname{dom} A) + \dim A0$.

Proof. We shall construct a basis of $\operatorname{gra} A$.
By Proposition 4.1.3(i), $A0$ is a linear subspace. Let $\left\{x_1^*, \cdots, x_k^*\right\}$ be a basis of $A0$ and $\left\{x_{k+1}, \ldots, x_l\right\}$ be a basis of $\operatorname{dom} A$. We show that $\left\{(0, x_1^*), \ldots, (0, x_k^*), (x_{k+1}, x_{k+1}^*), \cdots, (x_l, x_l^*)\right\}$ is a basis of $\operatorname{gra} A$, where $x_i^* \in Ax_i,\ i \in \{k+1, \cdots, l\}$. We first show that

$$\left\{(0, x_1^*), \ldots, (0, x_k^*), (x_{k+1}, x_{k+1}^*), \cdots, (x_l, x_l^*)\right\}$$

is linearly independent. Let $\alpha_i, i \in \{1, \cdots, l\}$, satisfy that

$$\alpha_1(0, x_1^*) + \cdots + \alpha_k(0, x_k^*) + \alpha_{k+1}(x_{k+1}, x_{k+1}^*) + \cdots + \alpha_l(x_l, x_l^*) = 0.$$

Hence

$$\alpha_{k+1}x_{k+1} + \cdots + \alpha_l x_l = 0 \tag{4.18}$$

$$\alpha_1 x_1^* + \cdots + \alpha_k x_k^* + \alpha_{k+1}x_{k+1}^* + \cdots + \alpha_l x_l^* = 0. \tag{4.19}$$

Since $\left\{x_{k+1}, \ldots, x_l\right\}$ is linearly independent, by (4.18) we have $\alpha_i = 0, i \in \{k+1, \cdots, l\}$. Then since $\left\{x_1^*, \cdots, x_k^*\right\}$ is linearly independent, by (4.19) we have $\alpha_i = 0, i \in \{1, \cdots, k\}$. Thus $\alpha_i = 0, i \in \{1, \cdots, l\}$. Hence $\left\{(0, x_1^*), \ldots, (0, x_k^*), (x_{k+1}, x_{k+1}^*), \cdots, (x_l\right.$ is linearly independent.

Let $(x, x^*) \in \operatorname{gra} A$. Then there exists $\beta_i,\ i \in \{k+1, \cdots, l\}$ satisfying that

$$\beta_{k+1} x_{k+1} + \cdots + \beta_l x_l = x.$$

Thus

$$\beta_{k+1} x_{k+1}^* + \cdots + \beta_l x_l^* \in Ax.$$

By Proposition 4.1.3(ii), there exists $z^* \in A0$ such that

$$\beta_{k+1} x_{k+1}^* + \cdots + \beta_l x_l^* + z^* = x^*.$$

Then there exists $\beta_i,\ i \in \{1, \cdots, k\}$ satisfying that

$$z^* = \beta_1 x_1^* + \cdots + \beta_k x_k^*.$$

Thus

$$(x, x^*) = \beta_1(0, x_1^*) + \cdots + \beta_k(0, x_k^*) + \beta_{k+1}(x_{k+1}, x_{k+1}^*) + \cdots + \beta_l(x_l, x_l^*).$$

Hence $\left\{(0, x_1), \ldots, (0, x_k), (x_{k+1}, x_{k+1}^*), \cdots, (x_l, x_l^*)\right\}$ is a basis of $\operatorname{gra} A$. Then

$$\dim(\operatorname{gra} A) = \dim(\operatorname{dom} A) + \dim(A0).$$

■

From Proposition 4.2.8, we now get a satisfactory characterization.

Proposition 4.2.9 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$*. Then*

$$A \textit{ is maximal monotone} \Leftrightarrow \dim \operatorname{gra} A = n.$$

Proof. Since linear subspaces are closed in finite-dimensional spaces, by Proposition 4.1.3(i) and Theorem 4.2.7 we have

$$A \text{ is maximal monotone} \Leftrightarrow \operatorname{dom} A = (A0)^{\perp}. \tag{4.20}$$

Assume that A is maximal monotone. Then

$$\operatorname{dom} A = (A0)^{\perp}.$$

Then Proposition 4.2.8 implies

$$\begin{aligned}
\dim(\operatorname{gra} A) &= \dim(\operatorname{dom} A) + \dim(A0) \\
&= \dim((A0)^{\perp}) + \dim(A0) \\
&= n,
\end{aligned}$$

as $(A0)^{\perp} + A0 = \mathbb{R}^n$.

Conversely, let $\dim(\operatorname{gra} A) = n$. By Proposition 4.2.8, we have

$$\dim(\operatorname{dom} A) = n - \dim(A0).$$

As $\dim((A0)^{\perp}) = n - \dim(A0)$ and $\operatorname{dom} A \subset (A0)^{\perp}$ by Proposition 4.1.3(iv), we have

$$\operatorname{dom} A = (A0)^{\perp}.$$

By (4.20), A is maximal monotone. ■

4.3 Constructing a good selection

When we proved Theorem 3.3.1, most of the much focused on finding a linear, continuous and monotone operator $\widetilde{A}$ such that $\widetilde{A}\,|_{\operatorname{dom} A^{-1}}$ is a selection of A^{-1}. Now for a maximal monotone operator A with a linear graph, we also want to find such an operator.

Fact 4.3.1 *Let S be a nonempty closed convex subset of X. Then for each $x \in X$ there exists a unique $s_0 \in S$ such that*

$$\|x - s_0\| = \min\Big\{\|x - s\| \mid \ s \in S\Big\}.$$

Proof. See [19, Corollary 1.1.5]. ■

By Fact 4.3.1, we can define the projector onto a nonempty closed convex subset of X.

Definition 4.3.2 *Let S be a nonempty closed convex subset of X. We define the* projector $P_S\colon X \to X$ *by*

$$P_S x = \operatorname{argmin}_{s\in S}\|x - s\|, \quad x \in X.$$

Fact 4.3.3 *Let S be a closed linear subspace of X and $x_0 \in X$. Then P_S is linear, continuous and*

$$P_{S+x_0}x = x_0 + P_S(x - x_0),\ \forall x \in X \tag{4.21}$$

$$P_S x + P_{S^\perp} x = x \tag{4.22}$$

$$P_S^* = P_S. \tag{4.23}$$

Proof. (4.21): Let $x \in S$. By Definition 4.3.2,

$$\|x - x_0 - P_S(x - x_0)\| \le \|x - x_0 - s\| = \|x - (x_0 + s)\|, \quad \forall s \in S.$$

By Fact 4.3.1, $P_S(x - x_0) \in S$. Thus $x_0 + P_S(x - x_0) \in S + x_0$. By Definition 4.3.2, $P_{S+x_0}x = x_0 + P_S(x - x_0)$.
(4.22) holds by $S \oplus S^\perp = X$. For the other parts see [27, Theorem 5.51(a)]. ∎

Definition 4.3.4 *Let $A\colon X \rightrightarrows X$ such that* $\operatorname{gra} A$ *is a linear subspace of $X \times X$. We define Q_A by*

$$Q_A x = \begin{cases} P_{Ax}x, & \textit{if } x \in \operatorname{dom} A;\\ \varnothing, & \textit{otherwise.}\end{cases}$$

Proposition 4.3.5 *Let* $A\colon X \rightrightarrows X$ *be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $X \times X$. *Then* Q_A *is single-valued on* $\operatorname{dom} A$ *and a selection of* A.

Proof. By Fact 4.2.1, Ax is nonempty closed convex, for every $x \in \operatorname{dom} A$. Then by Fact 4.3.1, Q_A is single-valued on $\operatorname{dom} A$ and a selection of A. ■

Proposition 4.3.6 *Let* $A\colon X \rightrightarrows X$ *be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $X \times X$. *Then* Q_A *is monotone, and linear on* $\operatorname{dom} A$. *Moreover,*

$$Q_A x = P_{(A0)^{\perp}}(Ax), \quad \forall x \in \operatorname{dom} A. \tag{4.24}$$

Proof. By Proposition 4.1.3(i) and Fact 4.2.1, $A0$ is a closed subspace. Let $x^* \in Ax$. Then

$$\begin{aligned}
Q_A x = P_{Ax}x &= P_{x^*+A0}x && (4.25)\\
&= x^* + P_{A0}(x - x^*) = x^* + P_{A0}x - P_{A0}x^* && (4.26)\\
&= P_{A0}x + P_{(A0)^{\perp}}x^* && (4.27)\\
&= P_{(A0)^{\perp}}x^* && (4.28)\\
&= P_{(A0)^{\perp}}(Ax), && (4.29)
\end{aligned}$$

in which, (4.25) holds by Proposition 4.1.3(ii), (4.26) and (4.27) by Fact 4.3.3. (4.28) holds since $P_{A0}x = 0$ by Proposition 4.1.3(iv).
Thus (4.24) holds. Since $Q_A x$ is single-valued by Remark 4.3.5, then $P_{(A0)^{\perp}}(Ax)$ is single-valued.

Now we show that Q_A is linear on $\operatorname{dom} A$. Take $x, y \in \operatorname{dom} A$ and $\alpha, \beta \in \mathbb{R}$. If $\alpha = \beta = 0$, by Proposition 4.1.3(i), we have

$$Q_A(\alpha x + \beta y) = Q_A 0 = P_{A0} 0 = 0 = \alpha Q_A x + \beta Q_A y. \tag{4.30}$$

Assume that $\alpha \neq 0$ or $\beta \neq 0$. By (4.24), we have

$$\begin{aligned} Q_A(\alpha x + \beta y) &= P_{(A0)^\perp} A(\alpha x + \beta y) && (4.31)\\ &= \alpha P_{(A0)^\perp}(Ax) + \beta P_{(A0)^\perp}(Ay) && (4.32)\\ &= \alpha Q_A x + \beta Q_A y, && (4.33) \end{aligned}$$

where (4.32) holds by Proposition 4.1.3(iii) and Fact 4.3.3, (4.33) by (4.24). By Proposition 4.3.5, Q_A is a selection of A. Since A is monotone, Q_A is monotone. ■

Proposition 4.3.7 *Let Y be a closed linear subspace of X. Let $A\colon X \rightrightarrows X$ be monotone such that A is linear on Y and at most single-valued. Then $P_Y A P_Y$ is linear, continuous and maximal monotone.*

Proof. Clearly, $P_Y A P_Y$ is linear since P_Y is linear by Fact 4.3.3 and A is linear on Y. In the following we show that $P_Y A P_Y$ is monotone.
Let $x \in X$. Then

$$\begin{aligned} \langle x,\ P_Y A P_Y x\rangle = \langle P_Y^* x,\ A(P_Y x) &= \langle P_Y x,\ A(P_Y x)\rangle && (4.34)\\ &\geq 0, && (4.35) \end{aligned}$$

where (4.34) holds by Fact 4.3.3. Inequality (4.35) holds since A is monotone.

By Example 2.1.3, $P_Y A P_Y$ is monotone. Then by Fact 2.1.22, we have $P_Y A P_Y$ is continuous and maximal monotone. ■

Now we show that we found the operator we were looking for.

Corollary 4.3.8 *Let* $A\colon X \rightrightarrows X$ *be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $X \times X$ *and* $\operatorname{dom} A$ *is closed. Then* $P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A}$ *is linear, continuous and maximal monotone. Moreover,* $P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A} = Q_A P_{\operatorname{dom} A}, (P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A})\mid_{\operatorname{dom} A} = Q_A$ *and* $Ax = (P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A})x + A0, \forall x \in \operatorname{dom} A$.

Proof. The former holds by Proposition 4.3.6 and Proposition 4.3.7. By Proposition 4.3.6 and Proposition 4.2.5, we have $(Q_A P_{\operatorname{dom} A})x \in (A0)^{\perp} = \operatorname{dom} A, \forall x \in X$. Then by Proposition 4.3.5, $(P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A})x = Q_A x \in Ax, \forall x \in \operatorname{dom} A$. By Proposition 4.1.3(ii), $Ax = (P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A})x + A0, \forall x \in \operatorname{dom} A$. ■

Remark 4.3.9 *By Corollary 4.3.8, we know that* $Q_A \mid_{\operatorname{dom} A}$ *is continuous on* $\operatorname{dom} A$. *But if we omit the assumption that* $\operatorname{dom} A$ *be closed, then we can't guarantee that* $Q_A \mid_{\operatorname{dom} A}$ *is continuous on* $\operatorname{dom} A$.

Example 4.3.10 *Let* X *be* $(\ell^2, \|\cdot\|_2)$ *space and* $A\colon X \to X\colon (x_n)_{n=1}^{\infty} \mapsto (\frac{x_n}{n})_{n=1}^{\infty}$. *Then* $Q_{A^{-1}} \mid_{\operatorname{dom} A^{-1}}$ *is not continuous on* $\operatorname{dom} A^{-1}$.

Proof. We first show that $Q_{A^{-1}} = A^{-1}$ is maximal monotone with a linear graph, but $\operatorname{dom} A^{-1} = \operatorname{ran} A$ is not closed.

Clearly, A is linear and one-to-one. Thus $Q_{A^{-1}} = A^{-1}$ and $\operatorname{gra} A^{-1}$ is a linear subspace. By Example 2.1.27, A is maximal monotone. By Proposition 3.3.2, A^{-1} is maximal monotone.

By Proposition 4.2.5, $\overline{\operatorname{ran} A} = \overline{\operatorname{dom} A^{-1}} = (A^{-1}0)^{\perp} = (0)^{\perp} = X$. Now we show that $\operatorname{ran} A$ is not closed, i.e, $\operatorname{ran} A \neq X$.

On the contrary, assume $\operatorname{ran} A = X$. Let $x = (1/n)_{n=1}^{\infty} \in X$. Then we have $A^{-1}x = (1)_{n=1}^{\infty} \notin X$. This is a contradiction. Hence $\operatorname{ran} A$ is not closed.

In the following we show that $Q_{A^{-1}} = A^{-1}$ is not continuous on $\operatorname{ran} A = \operatorname{dom} A^{-1}$.

Take $\{\frac{1}{n}e_n\} \subset \operatorname{ran} A$, where $e_n = (0, \cdots, 0, 1, 0, \cdots)$: the nth entry is 1 and the others are 0. Clearly, $\frac{1}{n}e_n \to 0$. But $\|A^{-1}(\frac{1}{n}e_n) - 0\| = \|e_n\| \nrightarrow 0$. Hence $Q_{A^{-1}} = A^{-1}$ is not continuous on $\operatorname{ran} A$. ■

4.4 The first main result

Now we come to our first main result in this thesis.

Theorem 4.4.1 *Let $A\colon X \rightrightarrows X$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace of $X \times X$ and* $\operatorname{dom} A$ *is closed. Then*

$$A = \partial f + \widetilde{A}_{\circ},$$

where $f := q_{\widetilde{A}} + \iota_{\operatorname{dom} A}$ is proper lower semicontinuous and convex, $\widetilde{A} = P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A}$ is linear, continuous and maximal monotone, and $\widetilde{A}_{\circ}$ is antisymmetric. In particular, A is decomposable.

Proof. By Corollary 4.3.8, $\widetilde{A}$ is linear, continuous and maximal monotone. Then by Fact 2.1.18, $q_{\widetilde{A}}$ is convex, differentiable and $\nabla q_{\widetilde{A}} = \widetilde{A}_{+}$. Since

$\operatorname{dom} A$ is a closed subspace, by Fact 2.1.24 $\iota_{\operatorname{dom} A}$ is proper lower semicontinuous and convex. Hence f is proper lower semicontinuous and convex. By Theorem 4.2.7, $(\operatorname{dom} A)^{\perp} = (A0)^{\perp\perp} = A0$. Let $x \in \operatorname{dom} A$. We have

$$\begin{aligned}
\partial f(x) &= \partial(q_{\widetilde{A}} + \iota_{\operatorname{dom} A})(x) \\
&= \nabla q_{\widetilde{A}}(x) + \partial\iota_{\operatorname{dom} A}(x) \qquad \text{(By Fact 2.1.30)} \\
&= \widetilde{A}_{+}x + (\operatorname{dom} A)^{\perp} \quad \text{(by Fact 2.1.18 and Fact 2.1.29)} \\
&= \widetilde{A}_{+}x + A0.
\end{aligned}$$

Thus $\forall x \in \operatorname{dom} A$,

$$\begin{aligned}
\partial f(x) + \widetilde{A}_{\circ}x = \widetilde{A}_{+}x + A0 + \widetilde{A}_{\circ}x &= \widetilde{A}x + A0 && (4.36) \\
&= Ax, && (4.37)
\end{aligned}$$

where (4.37) holds by Corollary 4.3.8. If $x \notin \operatorname{dom} A$, by definition $\partial f(x) = \varnothing = Ax$.

Hence we have $Ax = \partial f(x) + \widetilde{A}_{\circ}x, \quad \forall x \in X.$ ■

In general, a convex cone is not a linear subspace. We wonder if there exists a maximal monotone operator with a convex cone graph such that its graph is not a linear subspace.

The following gives a negative answer.

Fact 4.4.2 *A convex cone K is a linear subspace, if and only if, $-K \subset K$.*

Proof. See [25, Theorem 2.7]. ■

Proposition 4.4.3 *Let $A\colon X \rightrightarrows X$ be maximal monotone such that* $\operatorname{gra} A$ *is a convex cone. Then* $\operatorname{gra} A$ *is a linear subspace of* $X \times X$*.*

Proof. By Fact 4.4.2, it suffices to show that

$$-\operatorname{gra} A \subset \operatorname{gra} A.$$

Assume that $(x, x^*) \in \operatorname{gra} A$. We show that $-(x, x^*) \in \operatorname{gra} A$. Let $(y, y^*) \in \operatorname{gra} A$. As $\operatorname{gra} A$ is a convex cone,

$$(x, x^*) + (y, y^*) = (x + y,\ x^* + y^*) \in \operatorname{gra} A.$$

Thus

$$\langle x + y,\ x^* + y^* \rangle \geq 0. \quad \big(\text{ since } A \text{ is monotone and } (0, 0) \in \operatorname{gra} A\big)$$

This means

$$\begin{aligned} \langle -x - y,\ -x^* - y^* \rangle &\geq 0 \\ \langle (-x) - y,\ (-x^*) - y^* \rangle &\geq 0, \quad \forall (y, y^*) \in \operatorname{gra} A. \end{aligned}$$

Since A is maximal, we conclude that

$$(-x, -x^*) \in \operatorname{gra} A, \quad -(x, x^*) \in \operatorname{gra} A.$$

Hence $\operatorname{gra} A$ is a linear subspace. ■

In [14], Butnariu and Kassay discuss monotone operators with closed convex graphs. Actually, if such operators are maximal monotone, their graphs are affine sets.

Fact 4.4.4 *$C \subset X$ is an affine set $\Leftrightarrow \exists c_0 \in C$, $C - c_0$ is a linear subspace.*

Proof. See [31, page 1]. ■

Proposition 4.4.5 *Let $A\colon X \rightrightarrows X$ be maximal monotone such that* $\operatorname{gra} A$ *is a convex subset. Then* $\operatorname{gra} A$ *is an affine set.*

Proof. Let $(x_0, x_0^*) \in \operatorname{gra} A$ and $\widetilde{A}\colon X \rightrightarrows X$ such that $\operatorname{gra} \widetilde{A} = \operatorname{gra} A - (x_0, x_0^*)$. Thus $\operatorname{gra} \widetilde{A}$ is convex and $(0,0) \in \operatorname{gra} \widetilde{A}$. By Fact 2.1.15, $\widetilde{A}$ is maximal monotone. By Fact 4.4.4, it suffices to verify that $\operatorname{gra} \widetilde{A}$ is a linear subspace. By Proposition 4.4.3, it suffices to show that $\operatorname{gra} \widetilde{A}$ is a cone.
Let $k \geq 0$ and $(x, x^*) \in \operatorname{gra} \widetilde{A}$. We consider two cases.
Case 1: $k \leq 1$.

$$k(x, x^*) = k(x, x^*) + (1-k)(0,0) \in \operatorname{gra} \widetilde{A}. \tag{4.38}$$

Case 2: $k > 1$.
Let $(y, y^*) \in \operatorname{gra} \widetilde{A}$. By (4.38), $\frac{1}{k}(y, y^*) \in \operatorname{gra} \widetilde{A}$. Thus,

$$\langle kx - y,\ kx^* - y^* \rangle = k^2 \langle x - \tfrac{1}{k} y,\ x^* - \tfrac{1}{k} y^* \rangle \geq 0.$$

Since $\widetilde{A}$ is maximal monotone, $k(x, x^*) \in \operatorname{gra} \widetilde{A}$.
Hence $\operatorname{gra} \widetilde{A}$ is a cone. ■

4.5 Monotonicity of operators with linear graphs

In general, it is not easy to identify whether an operator is monotone. But if an operator with a linear graph and a basis is known, then we can use linear algebra to verify monotonicity and strict monotonicity.

Theorem 4.5.1 *Let* $A\colon X \rightrightarrows X$ *and* $\operatorname{gra} A = \operatorname{span}\Big\{(m_1, m_1^*), \cdots, (m_n, m_n^*)\Big\}$. *Then the following are equivalent*

(i) *A is monotone.*

(ii)

$$\textit{The matrix } B := \begin{pmatrix} \langle m_1, m_1^*\rangle & \langle m_1, m_2^*\rangle & \cdots & \langle m_1, m_n^*\rangle \\ \langle m_2, m_1^*\rangle & \langle m_2, m_2^*\rangle & \cdots & \langle m_2, m_n^*\rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle m_n, m_1^*\rangle & \langle m_n, m_2^*\rangle & \cdots & \langle m_n, m_n^*\rangle \end{pmatrix} \textit{ is monotone.}$$

(iii) *B_+ is positive semidefinite.*

Proof. Since $\operatorname{gra} A = \operatorname{span}\Big\{(m_1, m_1^*), \ldots, (m_n, m_n^*)\Big\}$, $\forall (x, x^*) \in \operatorname{gra} A, \exists \alpha_1, \ldots, \alpha_n$ such that

$$(x, x^*) = \sum_{i=1}^{n} \alpha_i (m_i, m_i^*).$$

Then A is monotone

$$\begin{aligned}
&\Leftrightarrow \langle x-y, x^*-y^*\rangle \geq 0, \quad \forall (x,x^*)\in \operatorname{gra} A, \forall (y,y^*)\in \operatorname{gra} A\\
&\Leftrightarrow \langle \sum_{i=1}^n(\alpha_i-\beta_i)m_i,\ \sum_{i=1}^n(\alpha_i-\beta_i)m_i^*\rangle \geq 0\\
&\quad \text{where}\quad (x,x^*)=\sum_{i=1}^n\alpha_i(m_i,m_i^*)=(\sum_{i=1}^n\alpha_i m_i,\ \sum_{i=1}^n\alpha_i m_i^*)\\
&\qquad\qquad\quad (y,y^*)=\sum_{i=1}^n\beta_i(m_i,m_i^*)=(\sum_{i=1}^n\beta_i m_i,\ \sum_{i=1}^n\beta_i m_i^*)\\
&\Leftrightarrow \langle \sum_{i=1}^n\gamma_i m_i,\ \sum_{i=1}^n\gamma_i m_i^*\rangle = \sum_{i=1}^n\sum_{j=1}^n\langle m_i, m_j^*\rangle\gamma_i\gamma_j \geq 0, \quad \forall \gamma_i\in\mathbb{R}\\
&\quad = (\gamma_1,\ldots,\gamma_n)B(\gamma_1,\ldots,\gamma_n)^\intercal \geq 0, \quad \forall\gamma_i\in\mathbb{R}\\
&\Leftrightarrow \nu^\intercal B\nu \geq 0, \quad \forall \nu\in\mathbb{R}^n\\
&\Leftrightarrow B \text{ is monotone} \quad \text{(by Example 2.1.3)}\\
&\Leftrightarrow B_+ \text{ is positive semidefinite} \quad \text{(by Fact 2.1.13 and Example 2.1.3).}
\end{aligned}$$

■

We also have a way to identify whether an operator with a linear graph is strictly monotone. First we give the definition of *strictly monotone*.

Definition 4.5.2 *A* strictly monotone *operator $T\colon X \rightrightarrows X$ is a mapping that satisfies*

$$\langle x^*-y^*,\ x-y\rangle > 0,$$

whenever $x^*\in T(x)$, $y^*\in T(y)$ and $x\neq y$.

Definition 4.5.3 *Let $A\colon X \to X$. We say A is* positive definite *if*

$$\langle x, Ax\rangle > 0, \quad \forall x \neq 0.$$

Theorem 4.5.4 *Let $A\colon X \rightrightarrows X$ and $\operatorname{gra} A = \operatorname{span}\Big\{(m_1, m_1^*), \ldots, (m_n, m_n^*)\Big\}$. Suppose that $\Big\{m_1, \ldots, m_n\Big\}$ is linearly independent. Then A is strictly monotone, if and only if, the matrix*

$$B = \begin{pmatrix} \langle m_1, m_1^*\rangle & \langle m_1, m_2^*\rangle & \cdots & \langle m_1, m_n^*\rangle \\ \langle m_2, m_1^*\rangle & \langle m_2, m_2^*\rangle & \cdots & \langle m_2, m_n^*\rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle m_n, m_1^*\rangle & \langle m_n, m_2^*\rangle & \cdots & \langle m_n, m_n^*\rangle \end{pmatrix}$$

is positive definite.

Proof. Since $\operatorname{gra} A = \operatorname{span}\{(m_i, m_i^*)\}_{i=1}^n$, A is strictly monotone

$$\begin{aligned} &\Leftrightarrow \langle x - y, x^* - y^*\rangle > 0, \quad \forall (x, x^*), (y, y^*) \in \operatorname{gra} A \text{ with } x \neq y \\ &\Leftrightarrow \langle \sum_{i=1}^n (\alpha_i - \beta_i) m_i,\ \sum_{i=1}^n (\alpha_i - \beta_i) m_i^*\rangle > 0 \\ &\quad \text{where} \quad (x, x^*) = \sum_{i=1}^n \alpha_i (m_i, m_i^*) = (\sum_{i=1}^n \alpha_i m_i,\ \sum_{i=1}^n \alpha_i m_i^*) \\ &\qquad\qquad\quad (y, y^*) = \sum_{i=1}^n \beta_i (m_i, m_i^*) = (\sum_{i=1}^n \beta_i m_i,\ \sum_{i=1}^n \beta_i m_i^*) \end{aligned}$$

.

Since $m_1, \ldots, m_n$ are linearly independent,

$$\begin{aligned} x \neq y &\Leftrightarrow (\alpha_1, \ldots, \alpha_n) \neq (\beta_1, \ldots, \beta_n) \\ &\Leftrightarrow \gamma := (\alpha_1 - \beta_1, \ldots, \alpha_n - \beta_n) \neq 0, \end{aligned}$$

A is strictly monotone

$$\Leftrightarrow \langle \sum_{i=1}^{n} \gamma_i m_i,\ \sum_{i=1}^{n} \gamma_i m_i^* \rangle > 0, \quad \text{for } \gamma \neq 0 \tag{4.39}$$

$$\Leftrightarrow \gamma^\intercal B \gamma > 0, \quad \forall \gamma \in \mathbb{R}^n \text{ with } \gamma \neq 0 \tag{4.40}$$

$$\Leftrightarrow B \text{ is positive definite.} \tag{4.41}$$

Just as in the proof of Theorem 4.5.1, we see that (4.40) holds. ■

Chapter 5

Auto-conjugates

5.1 Auto-conjugate representation

Definition 5.1.1 *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty,+\infty\right]$. *We define* $f^{\intercal}$ *by*

$$f^{\intercal}(x,x^*) = f(x^*,x), \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Definition 5.1.2 (Fenchel conjugate) *Let* $f\colon \mathbb{R}^n \to \left]-\infty,+\infty\right]$. *The* Fenchel conjugate *of* f, f^*, *is defined by*

$$f^*(x^*) = \sup_x \Big\{ \langle x^*, x\rangle - f(x) \Big\}, \quad \forall x^* \in \mathbb{R}^n.$$

Fact 5.1.3 *Let* $f\colon \mathbb{R}^n \to \left]-\infty,+\infty\right]$ *be proper lower semicontinuous and convex. Then* $f^{**} = f$.

Proof. See [26, Theorem 11.1]. ■

Proposition 5.1.4 *Let* $f, g\colon \mathbb{R}^n \to \left]-\infty,+\infty\right]$ *satisfy* $f \leq g$. *Then* $f^* \geq g^*$.

Proof. Follows directly by Definition 5.1.2. ■

Definition 5.1.5 (Auto-conjugate) *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty,+\infty\right]$ *be* *proper lower semicontinuous and convex. We say f is* auto-conjugate *if*

$$f^{*\intercal} = f.$$

Here are some examples of auto-conjugate functions.

Example 5.1.6 (Ghoussoub '06/[17]) *Let $\varphi\colon \mathbb{R}^n \to\,]-\infty, +\infty]$ be proper lower semicontinuous and convex, and $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and antisymmetric. Then*

$$f(x, x^*) := \varphi(x) + \varphi^*(x^*)$$
$$f(x, x^*) := \varphi(x) + \varphi^*(-Ax + x^*) \quad (\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n)$$

are auto-conjugate.

Proof. The first function is a special case of the second one when $A = 0$. So, it suffices to show the second case. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then we have

$$\begin{aligned}
& f^*(x^*, x) \\
&= \sup_{(y,y^*)} \Big\{ \langle y, x^*\rangle + \langle y^*, x\rangle - f(y, y^*) \Big\} \\
&= \sup_{(y,y^*)} \Big\{ \langle y, x^*\rangle + \langle y^*, x\rangle - \varphi(y) - \varphi^*(-Ay + y^*) \Big\} \\
&= \sup_{(y,y^*)} \Big\{ \langle y, x^*\rangle + \langle Ay, x\rangle + \langle -Ay + y^*, x\rangle - \varphi(y) - \varphi^*(-Ay + y^*) \Big\} \\
&= \sup_{y} \sup_{y^*} \Big\{ \langle y, x^*\rangle + \langle Ay, x\rangle + \langle -Ay + y^*, x\rangle - \varphi(y) - \varphi^*(-Ay + y^*) \Big\} \\
&= \sup_{y} \Big\{ \langle y, x^*\rangle + \langle y, -Ax\rangle - \varphi(y) + \sup_{y^*} \big\{ \langle -Ay + y^*, x\rangle - \varphi^*(-Ay + y^*) \big\} \Big\} \\
&= \sup_{y} \Big\{ \langle y, -Ax + x^*\rangle - \varphi(y) \Big\} + \varphi^{**}(x) \\
&= \varphi^*(-Ax + x^*) + \varphi(x) \quad \big(\text{by Fact 5.1.3}\big) \\
&= f(x, x^*).
\end{aligned}$$

■

Now we introduce some basic properties of auto-conjugate functions.

Lemma 5.1.7 (Penot-Simons-Zălinescu '05/[24],[29]) *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$ *be auto-conjugate. Then*

$$f(x, x^*) \geq \langle x, x^*\rangle, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$\begin{aligned} f(x,x^*) = f^*(x^*,x) = \sup_{(y,y^*)} \Big\{ \langle y, x^* \rangle + \langle y^*, x \rangle - f(y,y^*) \Big\} \\ \geq \langle x, x^* \rangle + \langle x^*, x \rangle - f(x,x^*). \end{aligned}$$

Thus $2f(x,x^*) \geq 2\langle x, x^* \rangle$. That is , $f(x,x^*) \geq \langle x, x^* \rangle$. ■

Proposition 5.1.8 *Let $f, g\colon \mathbb{R}^n \times \mathbb{R}^n \to \,]-\infty, +\infty]$ be auto-conjugate such that $f \leq g$. Then $f = g$.*

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By assumptions, $f(x,x^*) \leq g(x,x^*)$. On the other hand, by Proposition 5.1.4, $f^*(x^*,x) \geq g^*(x^*,x)$. Since f, g are auto-conjugate, $f(x,x^*) \geq g(x,x^*)$. Hence $f(x,x^*) = g(x,x^*)$. ■

Proposition 5.1.9 *Let $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \,]-\infty, +\infty]$. Then $f^{*\intercal} = f^{\intercal *}$.*

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$\begin{aligned} f^{\intercal *}(x^*,x) &= \sup_{(y,y^*)} \Big\{ \langle (y,y^*), (x^*,x) \rangle - f^{\intercal}(y,y^*) \Big\} \\ &= \sup_{(y,y^*)} \Big\{ \langle (y^*,y), (x,x^*) \rangle - f(y^*,y) \Big\} \\ &= f^*(x,x^*) = f^{*\intercal}(x^*,x). \end{aligned}$$

■

Fact 5.1.10 (Fenchel-Young inequality) *Let $f\colon \mathbb{R}^n \to \,]-\infty, +\infty]$ be proper* *lower semicontinuous and convex, and* $x, x^* \in \mathbb{R}^n$. *Then*

$$f(x) + f^*(x^*) \geq \langle x, x^* \rangle,$$

and equality holds, if and only if, $x^* \in \partial f(x)$.

Proof. See [25, Theorem 23.5] and [25, page 105]. ■.

Definition 5.1.11 *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$. *We define* $G(f)$ *by*

$$x^* \in G(f)x \Leftrightarrow f(x, x^*) = \langle x, x^* \rangle.$$

Here is an important property of auto-conjugates, which provides our main motivation for studying them.

Fact 5.1.12 (Penot-Simons-Zălinescu '05) *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$ *be auto-conjugate. Then* $G(f)$ *is maximal monotone.*

Proof. See [29, Theorem 1.4.(a)]. ■

Definition 5.1.13 (Representation) *Let* $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$ *and* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. *If* $A = G(f)$, *we call* f *a* representation *for* A. *If* f *is auto-conjugate, we call* f *an* auto-conjugate representation *for* A.

Proposition 5.1.14 *Let* $\varphi\colon \mathbb{R}^n \to \left]-\infty, +\infty\right]$ *be proper lower semicontinuous and convex, and* $A\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and antisymmetric. Let* $f(x, x^*) := \varphi(x) + \varphi^*(-Ax + x^*)$ $(\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n)$. *Then* f *is an auto-conjugate representation for* $\partial\varphi + A$.

Proof. By Example 5.1.6, f is auto-conjugate. Then we have

$$\begin{aligned}
&\langle x, x^* \rangle = f(x, x^*) \\
&\Leftrightarrow \langle x, -Ax + x^* \rangle = \varphi(x) + \varphi^*(-Ax + x^*) \\
&\Leftrightarrow x^* - Ax \in \partial\varphi(x) \quad \text{(by Fact 5.1.10)} \\
&\Leftrightarrow (x, x^*) \in \operatorname{gra}(\partial\varphi + A).
\end{aligned}$$

Hence f is an auto-conjugate representation for $\partial\varphi + A$. ■

Definition 5.1.15 *Let $f, g\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$. We define*

$$(f \Box_2 g)(x, x^*) = \inf_{y^*} \Big\{ f(x, x^* - y^*) + g(x, y^*) \Big\}, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Definition 5.1.16 *Let $f, g\colon \mathbb{R}^n \to \left]-\infty, +\infty\right]$. We define*

$$(f \oplus g)(x, x^*) = f(x) + g(x^*), \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Definition 5.1.17 *We define*

$$\pi_1\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n : (x, y) \mapsto x.$$

Fact 5.1.18 *Let $f, g\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$ be proper lower semicontinuous and convex. Set $\varphi = f \Box_2 g$. Assume*

$$\varphi(x, x^*) > -\infty, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n,$$

and

$$\bigcup_{\lambda>0} \lambda\big[\,\pi_1 \operatorname{dom} f - \pi_1 \operatorname{dom} g\big],$$

is a linear subspace of $\mathbb{R}^n$. *Then*

$$\varphi^*(x^*, x) = \min_{y^*} \Big\{ f^*(x^* - y^*, x) + g^*(y^*, x) \Big\}, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. See [29, Theorem 4.2]. ■

Proposition 5.1.19 *Let* $f, g\colon \mathbb{R}^n \times \mathbb{R}^n \to \,]-\infty, +\infty]$ *be auto-conjugate such that* $(\pi_1 \operatorname{dom} f - \pi_1 \operatorname{dom} g)$ *is a linear subspace of* $\mathbb{R}^n$. *Suppose* $M = f \Box_2 g$. *Then*

$$M(x, x^*) = \min_{y^*} \Big\{ f(x, x^* - y^*) + g(x, y^*) \Big\}, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

and M *is an auto-conjugate representation for* $G(f) + G(g)$.

Proof. By Lemma 5.1.7,

$$M(x, x^*) \geq \inf_{y^*} \Big\{ \langle x, y^* \rangle + \langle x, x^* - y^* \rangle \Big\} = \langle x, x^* \rangle, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{5.1}$$

Since $(\pi_1 \operatorname{dom} f - \pi_1 \operatorname{dom} g)$ is a linear subspace, $\bigcup_{\lambda>0} \lambda\big[\,\pi_1 \operatorname{dom} f - \pi_1 \operatorname{dom} g\big] = (\pi_1 \operatorname{dom} f - \pi_1 \operatorname{dom} g)$ is a linear subspace. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By

Fact 5.1.18, we have

$$\begin{aligned} M^*(x^*, x) &= \min_{y^*} \Big\{ f^*(x^* - y^*, x) + g^*(y^*, x) \Big\} \\ &= \min_{y^*} \Big\{ f(x, x^* - y^*) + g(x, y^*) \Big\} \\ &= M(x, x^*). \end{aligned}$$

Hence

$$M(x, x^*) = \min_{y^*} \Big\{ f(x, x^* - y^*) + g(x, y^*) \Big\}, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{5.2}$$

and M is auto-conjugate.

In the following we show that M is a representation for $G(f)+G(g)$. Suppose (x, x^*) satisfies

$$M(x, x^*) = \langle x, x^* \rangle.$$

For every $y^* \in \mathbb{R}^n$, since f, g are auto-conjugate, by Fact 5.1.7 we have

$$\begin{aligned} f(x, x^* - y^*) &\geq \langle x, x^* - y^* \rangle, \\ g(x, y^*) &\geq \langle x, y^* \rangle, \text{ and} \\ M(x, x^*) &\geq \langle x, x^* \rangle. \end{aligned} \tag{5.3}$$

Then by (5.2) and (5.3),

$$
\begin{aligned}
&(x,x^*) \in \operatorname{gra} G(M)\\
&\Leftrightarrow M(x,x^*) = \langle x,x^*\rangle\\
&\Leftrightarrow \exists s^* \text{ such that } \langle x,x^*\rangle = M(x,x^*) = f(x,x^*-s^*) + g(x,s^*)\\
&\Leftrightarrow \exists s^* \text{ such that } 0 = f(x,x^*-s^*) - \langle x,x^*-s^*\rangle + g(x,s^*) - \langle x,s^*\rangle\\
&\Leftrightarrow \exists s^* \text{ such that } \langle x,x^*-s^*\rangle = f(x,x^*-s^*), \langle x,s^*\rangle = g(x,s^*)\\
&\Leftrightarrow \exists s^* \text{ such that } (x,x^*-s^*) \in \operatorname{gra} G(f),\ (x,s^*) \in \operatorname{gra} G(g)\\
&\Leftrightarrow x^* \in \big(G(f)+G(g)\big)x.
\end{aligned}
$$

■

Now this raises the following question: Given a maximal monotone operator A, can we find an auto-conjugate representation for A?

Before answering this question, we introduce some definitions.

5.2 The Fitzpatrick function and the proximal average

Definition 5.2.1 (Fitzpatrick function '88) *Let* $A\colon X \rightrightarrows X$. *The* Fitzpatrick function *of* A *is*

$$F_A\colon (x,x^*) \mapsto \sup_{(y,y^*)\in \operatorname{gra} A} \langle x,y^*\rangle + \langle y,x^*\rangle - \langle y,y^*\rangle.$$

Fact 5.2.2 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is nonempty. Then* F_A *is proper lower semicontinuous and convex.*

Proof. See [8, Fact 1.2]. ■

Fact 5.2.3 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$q_A^*(x) = \iota_{\operatorname{ran} A_+}(x) + q_{(A_+)^\dagger}(x), \quad \forall x \in \mathbb{R}^n.$$

Proof. See [25, page 108] and Corollary 2.2.18. ■

Fact 5.2.4 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$F_A(x,x^*) = \iota_{\operatorname{ran} A_+}(x^* + A^*x) + \tfrac{1}{2}q_{(A_+)^\dagger}(x^* + A^*x), \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By [4, Theorem 2.3],

$$\begin{aligned}
&F_A(x,x^*)\\
&= 2q_{A_+}^*(\tfrac{1}{2}x^* + \tfrac{1}{2}A^*x)\\
&= \iota_{\operatorname{ran} A_+}(\tfrac{1}{2}x^* + \tfrac{1}{2}A^*x) + 2q_{(A_+)^\dagger}(\tfrac{1}{2}x^* + \tfrac{1}{2}A^*x) \quad \text{(by Fact 5.2.3)}\\
&= \iota_{\operatorname{ran} A_+}(x^* + A^*x) + \tfrac{1}{2}q_{(A_+)^\dagger}(x^* + A^*x).
\end{aligned}$$

■

Fact 5.2.5 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$F_A^*(x^*,x) = \iota_{\operatorname{gra} A}(x,x^*) + \langle x, Ax\rangle, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. See [4, Theorem 2.3]. ■

Proposition 5.2.6 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then $A+k\,\mathrm{Id}$ is invertible, for every $k>0$.*

Proof. Let x satisfy that

$$(A+k\,\mathrm{Id})x = 0.$$

Then we have $Ax=-kx$. By the monotonicity of A, we have

$$k\|x\|^2 = \langle x,\ kx\rangle = \langle x,\ -Ax\rangle = -\langle x,\ Ax\rangle \le 0.$$

Then $x=0$. Hence $A+k\,\mathrm{Id}$ is invertible. ■

Definition 5.2.7 (Proximal average) *Let $f_0, f_1\colon \mathbb{R}^n\times\mathbb{R}^n \to \left]-\infty,+\infty\right]$ be proper lower semicontinuous and convex. We define $P(f_0,f_1)$, the* proximal average *of f_0 and f_1, by*

$$\begin{aligned}
&P(f_0,f_1)(x,x^*)\\
&= -\tfrac{1}{2}\|(x,x^*)\|^2 + \inf_{(y_1,y_1^*)+(y_2,y_2^*)=(x,x^*)} \Big\{ \tfrac{1}{2}f_0(2y_1,2y_1^*) + \tfrac{1}{2}f_1\big(2y_2,2y_2^*\big)\\
&\quad + \|(y_1,y_1^*)\|^2 + \|(y_2,y_2^*)\|^2 \Big\}, \quad \forall (x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n.
\end{aligned}$$

Remark 5.2.8 *Let $f_0, f_1\colon \mathbb{R}^n\times\mathbb{R}^n \to \left]-\infty,+\infty\right]$ be proper lower semicontinuous and convex. Then*

$$P(f_0^{\intercal}, f_1^{\intercal}) = \Big(P(f_0,f_1)\Big)^{\intercal}.$$

Fact 5.2.9 *Let f_0 and $f_1\colon \mathbb{R}^n \to \left]-\infty,+\infty\right]$ be proper lower semicontinuous* *and convex. Then*

$$
\begin{aligned}
&P(f_0, f_1)(x, x^*) \\
&= \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} f_0(x+y, x^*+y^*) + \tfrac{1}{2} f_1(x-y, x^*-y^*) + \tfrac{1}{2}\|(y,y^*)\|^2 \Big\}, \\
&\quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}
$$

Proof. Let $(x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$
\begin{aligned}
&P(f_0, f_1)(x, x^*) \\
&= -\tfrac{1}{2}\|(x,x^*)\|^2 + \inf_{(y_1,y_1^*)+(y_2,y_2^*)=(x,x^*)} \Big\{ \tfrac{1}{2} f_0(2y_1, 2y_1^*) + \tfrac{1}{2} f_1\big(2y_2, 2y_2^*\big) \\
&\quad + \|(y_1,y_1^*)\|^2 + \|(y_2,y_2^*)\|^2 \Big\} \\
&= -\tfrac{1}{2}\|(x,x^*)\|^2 + \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} f_0\big(2\tfrac{x+y}{2}, 2\tfrac{x^*+y^*}{2}\big) + \tfrac{1}{2} f_1\big(2\tfrac{x-y}{2}, 2\tfrac{x^*-y^*}{2}\big) \\
&\quad + \big\|(\tfrac{x+y}{2}, \tfrac{x^*+y^*}{2})\big\|^2 + \big\|(\tfrac{x-y}{2}, \tfrac{x^*-y^*}{2})\big\|^2 \Big\} \\
&= \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} f_0(x+y, x^*+y^*) + \tfrac{1}{2} f_1(x-y, x^*-y^*) + \tfrac{1}{2}\|(y,y^*)\|^2 \Big\}.
\end{aligned}
$$

■

Definition 5.2.10 *Let $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \left]-\infty, +\infty\right]$ be proper lower semicontinuous and convex and h_f define by*

$$h_f(x,x^*) = \inf \Big\{ \tfrac{1}{2} f(x, 2x_1^*) + \tfrac{1}{2} f^*(2x_2^*, x) \mid x^* = x_1^* + x_2^* \Big\}, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Now we begin to answer the question above.

Fact 5.2.11 (Bauschke-Wang '07) *Let $A\colon X \rightrightarrows X$ be maximal mono-*

tone. Then $P(F_A, F_A^{*\intercal})$ *is an auto-conjugate representation for* A.

Proof. See [9, Theorem 5.7]. ■

Fact 5.2.12 (Penot-Zălinescu '05) *Let* $A\colon X \rightrightarrows X$ *be maximal monotone such that* $\operatorname{aff}(\operatorname{dom} A)$ *is closed. Then* h_{F_A} *is an auto-conjugate representation for* A.

Proof. See [24, Proposition 4.2].

■

5.3 The second main result

Our main goal is to find a formula for $P(F_A, F_A^{*\intercal})$ associated with a linear and monotone operator A. Until now, there was no explicit formula for that.

Theorem 5.3.1 *Let* $A\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and monotone. Then*

$$P(F_A, F_A^{*\intercal})(x, x^*) = \iota_{\operatorname{ran} A_+}(x^* - Ax) + \langle x, x^* \rangle + q_{(A_+)^\dagger}(x^* - Ax),$$
$$\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Fact 5.2.2 and Fact 5.2.9, we have

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*) \\
&= \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} F_A(x+y, x^*+y^*) + \tfrac{1}{2} F_A^{*\intercal}(x-y, x^*-y^*) && (5.4) \\
&\quad + \tfrac{1}{2}\|(y,y^*)\|^2 \Big\} \\
&= \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} F_A(x+y, x^*+y^*) + \iota_{\operatorname{gra} A}(x-y, x^*-y^*) && (5.5) \\
&\quad + \tfrac{1}{2}\langle x-y, A(x-y)\rangle + \tfrac{1}{2}\|(y,y^*)\|^2 \Big\} \\
&= \inf_{y} \Big\{ \tfrac{1}{2} F_A(x+y, 2x^* - A(x-y)) + \tfrac{1}{2}\langle x-y, A(x-y)\rangle && (5.6) \\
&\quad + \tfrac{1}{2}\|(y, x^* - A(x-y))\|^2 \Big\} \\
&= \inf_{y} \Big\{ \iota_{\operatorname{ran} A_+}\big(2x^* - A(x-y) + A^*x + A^*y\big) && (5.7) \\
&\quad + \tfrac{1}{4} q_{(A_+)^\dagger}\big(2x^* - A(x-y) + A^*x + A^*y\big) \\
&\quad + \tfrac{1}{2}\langle x-y, A(x-y)\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|x^* - A(x-y)\|^2 \Big\},
\end{aligned}
$$

in which, (5.5) holds by Fact 5.2.5, (5.6) by $y^* = x^* - A(x-y)$, and (5.7) by Fact 5.2.4.

Since

$$
\begin{aligned}
&2x^* - A(x-y) + A^*x + A^*y \\
&= 2x^* - 2Ax + Ax + Ay + A^*x + A^*y \\
&= 2x^* - 2Ax + (A + A^*)(x+y) \\
&= 2x^* - 2Ax + 2A_+(x+y), && (5.8)
\end{aligned}
$$

Thus $2x^* - A(x-y) + A^*x + A^*y \in \operatorname{ran} A_+ \Leftrightarrow x^* - Ax \in \operatorname{ran} A_+$. Then

$$\iota_{\operatorname{ran} A_+}(2x^* - A(x-y) + A^*x + A^*y) = \iota_{\operatorname{ran} A_+}(x^* - Ax). \tag{5.9}$$

We consider two cases.

Case 1: $x^* - Ax \notin \operatorname{ran} A_+$. By (5.7) and (5.9), $P(F_A, F_A^{*\intercal})(x, x^*) = \infty$.

Case 2: $x^* - Ax \in \operatorname{ran} A_+$. By Proposition 2.2.15 applied to A_+ with x replaced by $x^* - Ax$ and y replaced by $x + y$, we have

$$\begin{aligned}
&\tfrac{1}{4}q_{(A_+)^\dagger}(2x^* - A(x-y) + A^*x + A^*y)\\
&= \tfrac{1}{4}q_{(A_+)^\dagger}(2x^* - 2Ax + 2A_+(x+y)) \quad \text{(by (5.8))}\\
&= \tfrac{1}{4}\cdot 2^2 q_{(A_+)^\dagger}(x^* - Ax + A_+(x+y))\\
&= q_{(A_+)^\dagger}(x^* - Ax + A_+(x+y))\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle P_{\operatorname{ran} A_+}(x^* - Ax),\ x+y\rangle + q_{A_+}(x+y)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x+y,\ x^* - Ax\rangle + \tfrac{1}{2}\langle x+y,\ A(x+y)\rangle.
\end{aligned} \tag{5.10}$$

By (5.7), (5.9) and (5.10), we have

$$\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \inf_y \Big\{\langle x+y,\ x^* - Ax\rangle + \tfrac{1}{2}\langle x+y,\ A(x+y)\rangle\\
&\quad + \tfrac{1}{2}\langle x-y, A(x-y)\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|x^* - A(x-y)\|^2\Big\}.
\end{aligned}$$

Since

$$\begin{aligned}&\tfrac{1}{2}\langle x+y,\ A(x+y)\rangle+\tfrac{1}{2}\langle x-y, A(x-y)\rangle\\&=\langle x,\ Ax\rangle+\langle y,\ Ay\rangle,\end{aligned}$$

we have

$$\begin{aligned}&\langle x+y,\ x^*-Ax\rangle+\tfrac{1}{2}\langle x+y,\ A(x+y)\rangle+\tfrac{1}{2}\langle x-y, A(x-y)\rangle\\&=\langle x, x^*\rangle-\langle x,\ Ax\rangle+\langle y,\ x^*\rangle-\langle y,\ Ax\rangle+\langle x,\ Ax\rangle+\langle y,\ Ay\rangle\\&=\langle x, x^*\rangle+\langle y,\ Ay\rangle+\langle y,\ x^*\rangle-\langle y,\ Ax\rangle.\end{aligned}$$

Then

$$\begin{aligned}&P(F_A, F_A^{*\intercal})(x,x^*)\\&=q_{(A_+)^\dagger}(x^*-Ax)+\langle x, x^*\rangle+\inf_y\Big\{\langle y,\ Ay\rangle+\langle y,\ x^*\rangle-\langle y,\ Ax\rangle\\&\quad+\tfrac{1}{2}\|y\|^2+\tfrac{1}{2}\|x^*-Ax+Ay\|^2\Big\}\\&=q_{(A_+)^\dagger}(x^*-Ax)+\langle x, x^*\rangle+\inf_y\Big\{\langle y,\ Ay\rangle+\langle y,\ x^*\rangle-\langle y,\ Ax\rangle\\&\quad+\tfrac{1}{2}\|y\|^2+\tfrac{1}{2}\|x^*-Ax\|^2+\tfrac{1}{2}\|Ay\|^2+\langle Ay, x^*-Ax\rangle\Big\}.\end{aligned}$$

Since $\langle y,\ Ay\rangle + \frac{1}{2}\|y\|^2 + \frac{1}{2}\|Ay\|^2 = \frac{1}{2}\|y + Ay\|^2$,

$$\begin{aligned}
&P(F_A, F_A^{*\intercal})(x,x^*)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad + \inf_y \Big\{\tfrac{1}{2}\|y + Ay\|^2 + \langle y, x^* - Ax + A^*(x^* - Ax)\rangle\Big\}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad + \inf_y \Big\{\tfrac{1}{2}\|y + Ay\|^2 + \langle y, (\operatorname{Id} + A^*)(x^* - Ax)\rangle\Big\}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad - \sup_y \Big\{\langle y, (\operatorname{Id} + A^*)(Ax - x^*)\rangle - q_{(\operatorname{Id} + A^*)(\operatorname{Id} + A)}(y)\Big\}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad - q^*_{(\operatorname{Id} + A^*)(\operatorname{Id} + A)}\big((\operatorname{Id} + A^*)(Ax - x^*)\big)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad - q_{\big((\operatorname{Id} + A^*)(\operatorname{Id} + A)\big)^\dagger}\big((\operatorname{Id} + A^*)(Ax - x^*)\big) \quad \text{(by Proposition 5.2.3}\\
&\quad \text{and Proposition 5.2.6)}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2\\
&\quad - q_{(\operatorname{Id} + A)^{-1}(\operatorname{Id} + A^*)^{-1}}\big((\operatorname{Id} + A^*)(Ax - x^*)\big) \quad \text{(by Remark 2.1.35)}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle + \tfrac{1}{2}\|x^* - Ax\|^2 - \tfrac{1}{2}\|x^* - Ax\|^2\\
&= \langle x, x^*\rangle + q_{(A_+)^\dagger}(x^* - Ax).
\end{aligned}$$

Combining the results above, we have

$$P(F_A, F_A^{*\intercal})(x, x^*) = \iota_{\operatorname{ran} A_+}(x^* - Ax) + \langle x, x^* \rangle + q_{(A_+)^\dagger}(x^* - Ax),$$
$$\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

■

Proposition 5.3.2 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$\pi_1 [\operatorname{dom} P(F_A, F_A^{*\intercal})] = \mathbb{R}^n.$$

Proof. By Theorem 5.3.1,

$$P(F_A, F_A^{*\intercal})(x, Ax) = \langle x, Ax \rangle < \infty, \quad \forall x \in \mathbb{R}^n.$$

Thus $(x, Ax) \in \operatorname{dom} P(F_A, F_A^{*\intercal}), \forall x \in \mathbb{R}^n$. Hence

$$\pi_1 [\operatorname{dom} P(F_A, F_A^{*\intercal})] = \mathbb{R}^n.$$

■

Corollary 5.3.3 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Then*

$$P(F_A, F_A^{*\intercal}) = q_A \oplus (\iota_{\operatorname{ran} A} + q_{A^\dagger}).$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Theorem 5.3.1, we have

$$\begin{aligned} P(F_A, F_A^{*\intercal})(x, x^*) &= \iota_{\operatorname{ran} A}(x^* - Ax) + \langle x, x^* \rangle + q_{A^\dagger}(x^* - Ax) \\ &= \iota_{\operatorname{ran} A}(x^*) + \langle x, x^* \rangle + q_{A^\dagger}(x^* - Ax). \end{aligned}$$

Now suppose $x^* \in \operatorname{ran} A$. By Proposition 2.2.15, we have

$$\begin{aligned} q_{A^\dagger}(x^* - Ax) &= q_A(x) + q_{A^\dagger}(x^*) - \langle x, P_{\operatorname{ran} A} x^* \rangle \\ &= q_A(x) + q_{A^\dagger}(x^*) - \langle x, x^* \rangle. \end{aligned}$$

Thus

$$P(F_A, F_A^{*\intercal})(x, x^*) = q_A(x) + q_{A^\dagger}(x^*).$$

Combining the conclusions above,

$$\begin{aligned} P(F_A, F_A^{*\intercal})(x, x^*) &= \iota_{\operatorname{ran} A}(x^*) + q_A(x) + q_{A^\dagger}(x^*) \\ &= \big(q_A \oplus (\iota_{\operatorname{ran} A} + q_{A^\dagger})\big)(x, x^*), \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n. \end{aligned}$$

■

Corollary 5.3.4 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and antisymmetric. Then*

$$P(F_A, F_A^{*\intercal}) = \iota_{\operatorname{gra} A}.$$

Proof. Follows directly by Theorem 5.3.1. ■

Corollary 5.3.5 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$\begin{aligned}\big(\forall (x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n\big)\quad & P(F_A, F_A^{*\intercal})(x,x^*) \geq \langle x, x^*\rangle\\ & P(F_A, F_A^{*\intercal})(x,Ax) = \langle x, Ax\rangle.\end{aligned}$$

Proof. Apply Theorem 5.3.1 and Corollary 2.2.12. ■

For a linear and monotone operator A, Fact 5.2.11 shows $P(F_A, F_A^{*\intercal})$ is auto-conjugate. Now we give a new proof.

Proposition 5.3.6 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then $P(F_A, F_A^{*\intercal})$ is auto-conjugate.*

Proof. Let $(x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n$. By Theorem 5.3.1, we have

$$\begin{aligned}
&P(F_A, F_A^{*\intercal})^*(x^*,x)\\
&= \sup_{(y,\ y^*)}\Big\{\langle (x^*,x),\ (y,y^*)\rangle - \iota_{\operatorname{ran} A_+}(y^*-Ay) - \langle y,y^*\rangle \qquad (5.11)\\
&\quad - q_{(A_+)^\dagger}(y^*-Ay)\Big\}\\
&= \sup_{(y,\ w)}\Big\{\langle (y,\ A_+w+Ay),\ (x^*,x)\rangle - \iota_{\operatorname{ran} A_+}(A_+w) \qquad (5.12)\\
&\quad - \langle y,\ A_+w+Ay\rangle - q_{(A_+)^\dagger}(A_+w)\Big\}\\
&= \sup_{(y,\ w)}\Big\{\langle y,x^*\rangle + \langle A_+w+Ay,\ x\rangle - \langle y,\ A_+w+Ay\rangle - q_{A_+}(w)\Big\} \qquad (5.13)\\
&= \sup_{y}\sup_{w}\Big\{\langle y,x^*\rangle + \langle A_+w+Ay,\ x\rangle - \langle y,\ A_+w+Ay\rangle - q_{A_+}(w)\Big\}\\
&= \sup_{y}\Big\{\langle y,x^*\rangle + \langle Ay,\ x\rangle - \langle y,\ Ay\rangle + \sup_{w}\big\{\langle w, A_+x\rangle - \langle A_+y,\ w\rangle \qquad (5.14)\\
&\quad - q_{A_+}(w)\big\}\Big\}.
\end{aligned}$$

(5.12) holds by $y^* - Ay = A_+ w$ and (5.13) by Corollary 2.2.16.

By (5.14),

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})^*(x^*, x)\\
&= \sup_y \Big\{ \langle y, A^*x + x^* \rangle - \langle y,\ Ay \rangle + q_{A_+}^*(A_+x - A_+y) \Big\}\\
&= \sup_y \Big\{ \langle y, A^*x + x^* \rangle - \langle y,\ Ay \rangle + q_{(A_+)^\dagger}(A_+x - A_+y) \Big\} \qquad (5.15)\\
&= \sup_y \Big\{ \langle y, A^*x + x^* \rangle - \langle y,\ Ay \rangle + q_{A_+}(x - y) \Big\} \qquad (5.16)\\
&= \sup_y \Big\{ \langle y, A^*x + x^* - A_+x \rangle - q_A(y) + q_A(x) \Big\} \qquad (5.17)\\
&= q_A^*(A^*x + x^* - A_+x) + q_A(x)\\
&= \iota_{\operatorname{ran} A_+}(A^*x + x^* - A_+x) + q_{(A_+)^\dagger}(A^*x + x^* - A_+x) + q_A(x) \qquad (5.18)
\end{aligned}
$$

(5.15) and (5.18) holds by Proposition 5.2.3, (5.16) by Corollary 2.2.16 and (5.17) by Remark 2.1.12.

Note

$$A^*x + x^* - A_+x = x^* - Ax + Ax + A^*x - A_+x = x^* - Ax + A_+x. \qquad (5.19)$$

Thus

$$\iota_{\operatorname{ran} A_+}(A^*x + x^* - A_+x) = \iota_{\operatorname{ran} A_+}(x^* - Ax). \qquad (5.20)$$

If $x^* - Ax \notin \operatorname{ran} A_+$. By (5.18) and (5.20), $P(F_A, F_A^{*\intercal})^*(x^*, x) = \infty$.

Now suppose $x^* - Ax \in \operatorname{ran} A_+$. By Proposition 2.2.15 applied to A_+ with

x replaced by $x^* - Ax$ and y replaced by x,

$$\begin{aligned}
&q_{(A_+)^\dagger}(A^*x + x^* - A_+x)\\
&= q_{(A_+)^\dagger}(x^* - Ax + A_+x) \quad \text{(by (5.19))}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \big\langle P_{\operatorname{ran} A_+}(x^* - Ax),\ x\big\rangle + q_{A_+}(x)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x^* - Ax,\ x\rangle + q_A(x) \quad \text{(by Remark 2.1.12)}\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x^*, x\rangle - q_A(x).
\end{aligned}$$

Then by (5.18) and (5.20), $P(F_A, F_A^{*\intercal})^*(x^*, x) = q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle$. Combining the results above, we have

$$\begin{aligned}
P(F_A, F_A^{*\intercal})^*(x^*, x) &= \iota_{\operatorname{ran} A_+}(x^* - Ax) + q_{(A_+)^\dagger}(x^* - Ax) + \langle x, x^*\rangle\\
&= P(F_A, F_A^{*\intercal})(x, x^*) \quad \text{(by Theorem 5.3.1)},\\
&\quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}$$

■

Proposition 5.3.7 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Let $x \in \mathbb{R}^n$. Then*

$$\langle x, Bx\rangle = 0 \Leftrightarrow Bx = 0.$$

Proof. "⇐" Clear.

"⇒" Take $y \in \mathbb{R}^n$ and $\alpha > 0$. We have

$$
\begin{aligned}
0 &\leq \langle \alpha y + x, B(\alpha y + x)\rangle && (5.21)\\
&= \langle x, Bx\rangle + 2\alpha\langle y, Bx\rangle + \alpha^2\langle y, By\rangle \\
&= 2\alpha\langle y, Bx\rangle + \alpha^2\langle y, By\rangle, \quad (\text{by } \langle x, Bx\rangle = 0) \\
\Rightarrow 0 &\leq 2\langle y, Bx\rangle + \alpha\langle y, By\rangle && (5.22)\\
\Rightarrow 0 &\leq \langle y, Bx\rangle, \quad \forall y \in \mathbb{R}^n && (5.23)\\
\Rightarrow Bx &= 0,
\end{aligned}
$$

in which, (5.21) holds by monotonicity of B, (5.22) by multiplying $\frac{1}{\alpha}$ in both sides, and (5.23) by letting $\alpha \to 0^+$. ■

Corollary 5.3.8 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Then*

$$\ker B = \{x \mid q_B(x) = 0\}.$$

Corollary 5.3.9 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Let $x \in \mathbb{R}^n$. Then $(\iota_{\operatorname{ran} B} + q_{B^\dagger})(x) = 0$, if and only if, $x = 0$.*

Proof. "⇐" Clear.

"⇒" By assumption, we have

$$x \in \operatorname{ran} B \tag{5.24}$$

$$0 = \langle x,\ B^\dagger x\rangle. \tag{5.25}$$

By Fact 2.2.2, Fact 2.2.4 and Corollary 2.2.12, $B^\dagger$ is linear, symmetric and monotone. By (5.25) and Proposition 5.3.7 applied to $B^\dagger$, $B^\dagger x = 0$. Then by (5.24) and Fact 2.2.11, $x = P_{\operatorname{ran} B} x = BB^\dagger x = 0$. ■

Corollary 5.3.10 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone.*

$$(\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n) \quad P(F_A, F_A^{*\intercal})(x, x^*) = \langle x, x^* \rangle \Leftrightarrow (x, x^*) \in \operatorname{gra} A.$$

Proof. By Theorem 5.3.1 and Corollary 5.3.9. ■

Corollary 5.3.11 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then $P(F_A, F_A^{*\intercal})$ is an auto-conjugate representation for A.*

Proof. By Proposition 5.3.6 and Corollary 5.3.10. ■

For a linear and monotone operator A, what is h_{F_A}?

Proposition 5.3.12 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$h_{F_A} = P(F_A, F_A^{*\intercal}).$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$
\begin{aligned}
&h_{F_A}(x, x^*) \\
&= \inf \left\{ \tfrac{1}{2} F_A(x, 2x_1^*) + \tfrac{1}{2} F_A^*(2x_2^*, x) \mid x^* = x_1^* + x_2^* \right\} \\
&= \inf_{y^*} \left\{ \tfrac{1}{2} F_A(x, 2(x^* - y^*)) + \tfrac{1}{2} F_A^*(2y^*, x) \right\} \\
&= \inf_{y^*} \left\{ \tfrac{1}{2} F_A(x, 2(x^* - y^*)) + \iota_{\operatorname{gra} A}(x, 2y^*) + q_A(x) \right\} && (5.26) \\
&= \tfrac{1}{2} F_A(x, 2x^* - Ax) + q_{A_+}(x) \quad \text{(by } 2y^* = Ax \text{ and Remark 2.1.12)} \\
&= \iota_{\operatorname{ran} A_+}(2x^* - Ax + A^*x) + \tfrac{1}{4} q_{(A_+)^\dagger}(2x^* - Ax + A^*x) + q_{A_+}(x), && (5.27)
\end{aligned}
$$

where (5.26) holds by Fact 5.2.5, and (5.27) by Fact 5.2.4.

Note that

$$
2x^* - Ax + A^*x = 2x^* - 2Ax + 2A_+x. \tag{5.28}
$$

Then $2x^* - Ax + A^*x \in \operatorname{ran} A_+ \Leftrightarrow x^* - Ax \in \operatorname{ran} A_+$. Thus

$$
\iota_{\operatorname{ran} A_+}(2x^* - Ax + A^*x) = \iota_{\operatorname{ran} A_+}(x^* - Ax). \tag{5.29}
$$

If $x^* - Ax \notin \operatorname{ran} A_+, h_{F_A}(x, x^*) = \infty$ by (5.27) and (5.29).

Now suppose that $x^* - Ax \in \operatorname{ran} A_+$. By Proposition 2.2.15 applied to A_+

with x replaced by $x^* - Ax$ and y replaced by x, we have

$$\begin{aligned}
&\tfrac{1}{4}q_{(A_+)^\dagger}(2x^* - Ax + A^*x)\\
&= \tfrac{1}{4}q_{(A_+)^\dagger}(2x^* - 2Ax + 2A_+x) \quad \text{(by (5.28))}\\
&= q_{(A_+)^\dagger}(x^* - Ax + A_+x)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x,\ P_{\operatorname{ran} A_+}(x^* - Ax)\rangle + q_{A_+}(x)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x,\ x^* - Ax\rangle + q_{A_+}(x)\\
&= q_{(A_+)^\dagger}(x^* - Ax) + \langle x,\ x^*\rangle - q_{A_+}(x) \quad \text{(by Remark 2.1.12).}
\end{aligned}$$

Then by (5.27) and (5.29),

$$h_{F_A}(x, x^*) = \langle x,\ x^*\rangle + q_{(A_+)^\dagger}(x^* - Ax).$$

Combining the results above,

$$\begin{aligned}
h_{F_A}(x, x^*) &= \iota_{\operatorname{ran} A_+}(x^* - Ax) + \langle x,\ x^*\rangle + q_{(A_+)^\dagger}(x^* - Ax)\\
&= P(F_A, F_A^{*\intercal})(x, x^*) \quad \text{(by Theorem 5.3.1)},\\
&\quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}$$

■

Proposition 5.3.13 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is nonempty. Then*

$$h_{F_A} = h_{F_A^{*\intercal}}.$$

Proof. Let $(x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n$. By Definition 5.2.10, we have

$$h_{F_A^{*\intercal}}(x,x^*) = \inf\left\{\tfrac{1}{2}F_A^{*\intercal}(x,2x_1^*)+\tfrac{1}{2}(F_A^{*\intercal})^*(2x_2^*,x)\mid x^*=x_1^*+x_2^*\right\} \tag{5.30}$$

$$= \inf\left\{\tfrac{1}{2}F_A^{*}(2x_1^*,x)+\tfrac{1}{2}F_A(x,2x_2^*)\mid x^*=x_1^*+x_2^*\right\} \tag{5.31}$$

$$= \inf\left\{\tfrac{1}{2}F_A^{*}(2x_2^*,x)+\tfrac{1}{2}(F_A(x,2x_1^*)\mid x^*=x_1^*+x_2^*\right\} \tag{5.32}$$

$$= h_{F_A}(x,x^*), \tag{5.33}$$

where (5.31) holds by Proposition 5.1.9, Fact 5.2.2 and Fact 5.1.3. ■

Corollary 5.3.14 *Let $A\colon\mathbb{R}^n\to\mathbb{R}^n$ be linear and monotone. Then*

$$h_{F_A}=h_{F_A^{*\intercal}}=P(F_A,F_A^{*\intercal}).$$

5.4 An example

In the following we give an example of Theorem 5.3.1.

Example 5.4.1 *Let*

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

be the rotation of angle $\theta\in[0,\frac{\pi}{2}[$. Then $A^*=A^{-1}$ and

$$P(F_A,F_A^{*\intercal})(x,x^*)=\tfrac{1}{2\cos\theta}\|x^*-Ax\|^2+\langle x,x^*\rangle$$
$$=\tfrac{1}{2\cos\theta}\|x^*-\sin\theta Rx\|^2+\tfrac{\cos\theta}{2}\|x\|^2,$$

where

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Proof. By assumptions, we have

$$A_+ = \cos\theta \,\mathrm{Id} \quad AA^* = \mathrm{Id} \quad R^*R = \mathrm{Id}\,. \tag{5.34}$$

By Theorem 5.3.1 and Remark 2.1.35, we have

$$P(F_A, F_A^{*\intercal})(x, x^*) = \tfrac{1}{2\cos\theta}\|x^* - Ax\|^2 + \langle x, x^*\rangle. \tag{5.35}$$

By (5.35), (5.34), and $A = \cos\theta\,\mathrm{Id} + \sin\theta R$, we have

$$\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*)\\
&= \tfrac{1}{2\cos\theta}\Big(\|x^*\|^2 + \langle A^*Ax, x\rangle - 2\langle x^*,\ Ax\rangle\Big) + \langle x, x^*\rangle\\
&= \tfrac{1}{2\cos\theta}\Big(\|x^*\|^2 + \|x\|^2 - 2\langle x^*,\ \cos\theta x + \sin\theta Rx\rangle\Big) + \langle x, x^*\rangle\\
&= \tfrac{1}{2\cos\theta}\Big(\|x^*\|^2 + \|x\|^2 - 2\sin\theta\langle x^*, Rx\rangle\Big)\\
&= \tfrac{1}{2\cos\theta}\Big(\|x^*\|^2 + \sin^2\theta\|x\|^2 - 2\sin\theta\langle x^*, Rx\rangle + \cos^2\theta\|x\|^2\Big)\\
&= \tfrac{1}{2\cos\theta}\|x^* - \sin\theta Rx\|^2 + \tfrac{\cos\theta}{2}\|x\|^2.
\end{aligned}$$

■

Fact 5.4.2 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is nonempty. Then* $F_{A^{-1}}^{\intercal} = F_A$.

Proof. See [8, Fact 1.2]. ■

Corollary 5.4.3 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be monotone such that* $\operatorname{gra} A$ *is nonempty. Then*

$$P(F_{A^{-1}}, F_{A^{-1}}^{*\intercal}) = \Big(P(F_A, F_A^{*\intercal})\Big)^{\intercal}.$$

Proof. By assumptions and Proposition 3.3.2, A^{-1} is monotone and $\operatorname{gra} A^{-1}$ is nonempty. Then by Proposition 5.1.9, Fact 5.4.2 and Remark 5.2.8,

$$P(F_{A^{-1}}, {F^*}^{\intercal}_{A^{-1}}) = P(F_A^{\intercal}, F_{A^{-1}}^{\intercal *}) = P(F_A^{\intercal}, F_A^{*}) = \Big(P(F_A, F_A^{*\intercal})\Big)^{\intercal}.$$

■

Theorem 5.4.4 *Let* $A\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear, monotone and invertible. Let* $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. *Then*

$$0 \in \tfrac{\partial P(F_A, F_A^{*\intercal})(x,x^*)}{\partial x^*} \Leftrightarrow x^* = A_{\circ} x \tag{5.36}$$

$$0 \in \tfrac{\partial P(F_A, F_A^{*\intercal})(x,x^*)}{\partial x} \Leftrightarrow x = (A^{-1})_{\circ} x^*. \tag{5.37}$$

Proof. By Fact 2.2.2, Fact 2.2.4, Fact 2.1.13 and Corollary 2.2.12, $(A_+)^{\dagger}$ is linear, symmetric and monotone.

Now (5.36): Follows from Theorem 5.3.1, Fact 2.1.18 and Fact 2.1.30,

$0 \in \frac{\partial P(F_A, F_A^{*\intercal})(x,x^*)}{\partial x^*}$

$\Leftrightarrow 0 \in \partial \iota_{\operatorname{ran} A_+}(x^* - Ax) + x + (A_+)^{\dagger}(x^* - Ax),\ x^* - Ax \in \operatorname{ran} A_+$

$\Leftrightarrow 0 \in \ker A_+ + x + (A_+)^{\dagger}(x^* - Ax),\ x^* - Ax \in \operatorname{ran} A_+ \quad$ (by Fact 2.1.32)

$\Leftrightarrow 0 \in x + (A_+)^{-1}(x^* - Ax),\ x^* - Ax \in \operatorname{ran} A_+ \quad$ (by Corollary 2.2.7)

$\Leftrightarrow -x \in (A_+)^{-1}(x^* - Ax),\ x^* - Ax \in \operatorname{ran} A_+$

$\Leftrightarrow x^* - Ax \in -A_+ x$

$\Leftrightarrow x^* \in Ax - A_+ x = A_{\circ} x.$

Then (5.37) Follows from Corollary 5.4.3 and (5.36). ■

Proposition 5.4.5 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and $h(x^*) := P(F_A, F_A^{*\intercal})(0, x^*)$ $(\forall x^* \in \mathbb{R}^n)$. Then $\partial h = (A_+)^{-1}$.*

Proof. By Theorem 5.3.1,

$$h(x^*) = \iota_{\operatorname{ran} A_+}(x^*) + q_{(A_+)^{\dagger}}(x^*), \quad \forall x^* \in \mathbb{R}^n.$$

By Fact 2.2.2, Fact 2.2.4, Fact 2.1.13 and Corollary 2.2.12, $(A_+)^{\dagger}$ is linear, symmetric and monotone. Thus by Fact 2.1.30 and Fact 2.1.18,

$$\partial h(x^*) = \begin{cases} \partial \iota_{\operatorname{ran} A_+}(x^*) + (A_+)^{\dagger} x^*, & \text{if } x^* \in \operatorname{ran} A_+; \\ \varnothing, & \text{otherwise.} \end{cases}$$

Now suppose $x^* \in \operatorname{ran} A_+$. By Fact 2.1.32 and Corollary 2.2.7,

$$\partial h(x^*) = \ker A_+ + (A_+)^\dagger x^* = (A_+)^{-1} x^*.$$

Next suppose $x^* \notin \operatorname{ran} A_+$. Clearly, $(A_+)^{-1} x^* = \varnothing = \partial h(x^*)$. In all cases, $\partial h = (A_+)^{-1}$. ■

Remark 5.4.6 *In general, let us consider* $g(x, x^*) := f(x) + f^*(x^*)$ $(\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n)$, *where* $f\colon \mathbb{R}^n \to \mathbb{R}^n$ *is proper lower semicontinuous and convex. Let* $h(x^*) = g(0, x^*) = f(0) + f^*(x^*)$ $(\forall x^* \in \mathbb{R}^n)$. *Thus by [26, Proposition 11.3],*

$$\partial h(x^*) = \partial f^*(x^*) = (\partial f)^{-1}(x^*), \quad \forall x^* \in \mathbb{R}^n.$$

5.5 Relationship among auto-conjugate representations

Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Suppose $f(x, x^*) = q_A(x) + q_A^*(-A_\circ x + x^*)$ $(\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n)$. By Proposition 5.1.14 and Fact 2.1.18, f is an auto-conjugate representation for $A_+ + A_\circ = A$. By Corollary 5.3.11, $P(F_A, F_A^{*\intercal})$ is also an auto-conjugate representation for A. The next Proposition does that.

Proposition 5.5.1 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear, symmetric and monotone, and* $A\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and antisymmetric. Let* $f(x, x^*) = q_B(x) +$

$q_B^*(-Ax+x^*)$ $(\forall (x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n)$. *Then*

$$f = P(F_{(A+B)}, F^{*\intercal}_{(A+B)}).$$

Proof. Let $(x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n$. By Fact 5.2.3,

$$f(x,x^*) = q_B(x) + \iota_{\operatorname{ran} B}(-Ax+x^*) + q_{B^\dagger}(-Ax+x^*). \tag{5.38}$$

By Theorem 5.3.1, we have

$$\begin{aligned}
&P(F_{(A+B)}, F^{*\intercal}_{(A+B)})(x,x^*)\\
&= \iota_{\operatorname{ran} B}\big(x^* - (A+B)x\big) + \langle x, x^*\rangle + q_{B^\dagger}\big(x^* - (A+B)x\big)\\
&= \iota_{\operatorname{ran} B}(x^* - Ax) + \langle x, x^*\rangle + q_{B^\dagger}(x^* - Ax - Bx)
\end{aligned}$$

If $x^* - Ax \notin \operatorname{ran} B$, $P(F_{(A+B)}, F^{*\intercal}_{(A+B)})(x,x^*) = \infty$.

Now suppose $x^* - Ax \in \operatorname{ran} B$. By Proposition 2.2.15 applied to B with x replaced by $x^* - Ax$ and y replaced by $-x$,

$$\begin{aligned}
&q_{B^\dagger}(x^* - Ax - Bx)\\
&= \langle P_{\operatorname{ran} B}(x^* - Ax),\ -x\rangle + q_{B^\dagger}(-Ax+x^*) + q_B(-x)\\
&= \langle x^* - Ax,\ -x\rangle + q_{B^\dagger}(-Ax+x^*) + q_B(-x)\\
&= -\langle x, x^*\rangle + q_{B^\dagger}(-Ax+x^*) + q_B(x) \quad (\text{by } \langle Ax,\ -x\rangle = 0).
\end{aligned}$$

Hence

$$\begin{aligned}P(F_{(A+B)}, F^{*\intercal}_{(A+B)})(x,x^*) &= q_B(x) + \iota_{\operatorname{ran} B}(-Ax + x^*) + q_{B^\dagger}(-Ax + x^*)\\ &= f(x,x^*) \quad \text{(by (5.38))}, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.\end{aligned}$$

■

Proposition 5.5.2 *Let $A, B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Then*

$$P(F_{(A+B)}, F^{*\intercal}_{(A+B)}) = P(F_A, F^{*\intercal}_A)\Box_2 P(F_B, F^{*\intercal}_B).$$

Proof. Let $(x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Theorem 5.3.1,

$$\begin{aligned}&P(F_A, F^{*\intercal}_A)\Box_2 P(F_B, F^{*\intercal}_B)(x,x^*)\\ &= \inf_{y^*} \Big\{ P(F_A, F^{*\intercal}_A)(x, x^* - y^*) + P(F_B, F^{*\intercal}_B)(x, y^*) \Big\}\\ &= \inf_{y^*} \Big\{ \iota_{\operatorname{ran} A_+}(x^* - y^* - Ax) + \langle x, x^* - y^* \rangle + q_{(A_+)^\dagger}(x^* - y^* - Ax)\\ &\quad + \langle x, y^* \rangle + \iota_{\operatorname{ran} B_+}(y^* - Bx) + q_{(B_+)^\dagger}(y^* - Bx) \Big\}\\ &= \langle x, x^* \rangle + \inf_{y^*} \Big\{ \iota_{\operatorname{ran} A_+}(x^* - y^* - Ax) + q_{(A_+)^\dagger}(x^* - y^* - Ax)\\ &\quad + \iota_{\operatorname{ran} B_+}(y^* - Bx) + q_{(B_+)^\dagger}(y^* - Bx) \Big\}\\ &\le \langle x, x^* \rangle + \iota_{\operatorname{ran}(A_+ + B_+)}(x^* - Ax - Bx) \qquad (5.39)\\ &\quad + \inf_{y^*} \Big\{ \iota_{\operatorname{ran} A_+}(x^* - y^* - Ax) + q_{(A_+)^\dagger}(x^* - y^* - Ax)\\ &\quad + \iota_{\operatorname{ran} B_+}(y^* - Bx) + q_{(B_+)^\dagger}(y^* - Bx) \Big\}.\end{aligned}$$

Now suppose $x^*-Ax-Bx \in \operatorname{ran}(A_+ + B_+)$. Let $x^*-Ax-Bx = (A_+ + B_+)p$ and $y_0^* := B_+p + Bx$. Thus

$$\begin{aligned} x^* - y_0^* - Ax &= x^* - B_+p - Bx - Ax = (A_+ + B_+)p - B_+p = A_+p, \\ y_0^* - Bx &= B_+p. \end{aligned}$$

Then by (5.39),

$$\begin{aligned} &P(F_A, F_A^{*\intercal})\Box_2 P(F_B, F_B^{*\intercal})(x,x^*) \\ &\leq \langle x, x^*\rangle + \iota_{\operatorname{ran} A_+}(x^* - y_0^* - Ax) + q_{(A_+)^\dagger}(x^* - y_0^* - Ax) \\ &\quad + \iota_{\operatorname{ran} B_+}(y_0^* - Bx) + q_{(B_+)^\dagger}(y_0^* - Bx) \\ &= \langle x, x^*\rangle + q_{(A_+)^\dagger}(A_+p) + q_{(B_+)^\dagger}(B_+p) \\ &= \langle x, x^*\rangle + q_{A_+}(p) + q_{B_+}(p) \qquad (5.40) \\ &= \langle x, x^*\rangle + q_{(A_+ + B_+)}(p) \\ &= \langle x, x^*\rangle + q_{(A_+ + B_+)^\dagger}(x^* - Ax - Bx), \qquad (5.41) \end{aligned}$$

in which, (5.40) and (5.41) hold by Corollary 2.2.16.

Combining the results above,

$$\begin{aligned} &P(F_A, F_A^{*\intercal})\Box_2 P(F_B, F_B^{*\intercal})(x,x^*) \\ &\leq \iota_{\operatorname{ran}(A_+ + B_+)}(x^* - Ax - Bx) + \langle x, x^*\rangle + q_{(A_+ + B_+)^\dagger}(x^* - Ax - Bx) \\ &= P(F_{(A+B)}, F_{(A+B)}^{*\intercal})(x,x^*) \quad \text{(by Theorem 5.3.1)}, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n. \end{aligned}$$

By Proposition 5.3.2, $\pi_1\left[\operatorname{dom} P(F_A, F_A^{*\intercal})\right] = \pi_1\left[\operatorname{dom} P(F_B, F_B^{*\intercal})\right] = \mathbb{R}^n$. Then by Proposition 5.3.6 and Proposition 5.1.19, $P(F_A, F_A^{*\intercal})\Box_2 P(F_B, F_B^{*\intercal})$

is auto-conjugate. Thus by Proposition 5.3.6 and Proposition 5.1.8,

$P(F_A, F_A^{*\intercal})\Box_2 P(F_B, F_B^{*\intercal}) = P(F_{(A+B)}, F_{(A+B)}^{*\intercal})$. ■

Lemma 5.5.3 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Then*

$$P(F_A, F_A^{*\intercal})(x, Ay) = P(F_A, F_A^{*\intercal})(y, Ax), \quad \forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. By Corollary 5.3.3 and Corollary 2.2.16,

$$P(F_A, F_A^{*\intercal})(x, Ay) = \iota_{\operatorname{ran} A}(Ay) + q_A(x) + q_{A^\dagger}(Ay) = q_A(x) + q_A(y).$$

On the other hand,

$$P(F_A, F_A^{*\intercal})(y, Ax) = \iota_{\operatorname{ran} A}(Ax) + q_A(y) + q_{A^\dagger}(Ax) = q_A(x) + q_A(y).$$

Thus

$$P(F_A, F_A^{*\intercal})(x, Ay) = P(F_A, F_A^{*\intercal})(y, Ax).$$

■

Proposition 5.5.4 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear, symmetric and monotone. Then $f = P(F_A, F_A^{*\intercal})$, if and only if, f is auto-conjugate satisfying $f(x, Ay) = f(y, Ax) \quad (\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n)$ and $f(0,0)$ is finite.*

Proof. "⇒" By Proposition 5.3.6, Lemma 5.5.3 and Corollary 5.3.5.

"⇐" Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. We prove in two steps.

Step 1: We will verify that $f(x, x^*) = \infty$, if $x^* \notin \operatorname{ran} A$.

Since $\mathbb{R}^n = \operatorname{ran} A \oplus \ker A$, $x^* = P_{\operatorname{ran} A}x^* + P_{\ker A}x^*$. Since $x^* \notin \operatorname{ran} A$,

$P_{\ker A}x^* \neq 0$. Thus

$$\langle P_{\ker A}x^*, x^*\rangle = \langle P_{\ker A}x^*, P_{\operatorname{ran} A}x^* + P_{\ker A}x^*\rangle = \|P_{\ker A}x^*\|^2 > 0. \quad (5.42)$$

Thus by assumptions,

$$f(kP_{\ker A}x^*, 0) = f(kP_{\ker A}x^*, A0) = f(0, AkP_{\ker A}x^*) = f(0,0), \quad \forall k \in \mathbb{R}. \quad (5.43)$$

Then by Fact 5.1.10,

$$\begin{aligned} f(x,x^*) + f(0,0) &= f(x,x^*) + f(kP_{\ker A}x^*, 0) = f(x,x^*) + f^*(0, kP_{\ker A}x^*) \\ &\geq \langle x^*, kP_{\ker A}x^*\rangle \to \infty, \text{ as } k \to \infty. \quad \text{(by (5.42))} \quad (5.44) \end{aligned}$$

Since $f(0,0)$ is finite, by (5.44) $f(x,x^*) = \infty$.

Step 2: Suppose that $x^* \in \operatorname{ran} A$. Let $x^* = Ap$. By Fact 5.1.10,

$$\begin{aligned} & f(x,Ap) + f(p,Ax) = f(x,Ap) + f^*(Ax,p) \geq \langle p, Ap\rangle + \langle x, Ax\rangle \\ \Rightarrow \quad & 2f(x,Ap) \geq \langle p, Ap\rangle + \langle x, Ax\rangle \quad (\text{by } f(x,Ap) = f(p,Ax)) \\ \Rightarrow \quad & f(x,x^*) \geq q_A(x) + q_A(p) = q_A(x) + q_{A^\dagger}(x^*), \quad (5.45) \end{aligned}$$

in which, (5.45) by $x^* = Ap$ and Corollary 2.2.16.

By conclusions above and Corollary 5.3.3, we have

$$f(x,x^*) \geq \iota_{\operatorname{ran} A}(x^*) + q_A(x) + q_{A^\dagger}(x^*) = P(F_A, F_A^{*\intercal})(x,x^*), \quad \forall (x,x^*).$$

Then by Corollary 5.3.11 and Proposition 5.1.8 , we have

$$f = P(F_A, F_A^{*\intercal}).$$

■

5.6 Nonuniqueness

We now tackle the following question: Given a linear and monotone operator, are auto-conjugate representations for A unique? The answer is negative. We will give several different auto-conjugate representations for Id. By Corollary 5.3.3, we have

$$P(F_{\mathrm{Id}}, F_{\mathrm{Id}}^{*\intercal}) = \tfrac{1}{2}\|\cdot\|^2 \oplus \tfrac{1}{2}\|\cdot\|^2.$$

Proposition 5.6.1 *Let* $j(x) = \frac{1}{2}x^2,\ \forall x \in \mathbb{R}$. *Assume* g *is such that* $g^*(-x) = g(x) \geq 0,\ \forall x \in \mathbb{R}$. *Then*

$$f(x,y) := j\left(\tfrac{x+y}{\sqrt{2}}\right) + g\left(\tfrac{x-y}{\sqrt{2}}\right) \quad \big(\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n\big)$$

is an auto-conjugate representation for Id .

Proof. We first show that f is auto-conjugate. Let $(x,y)\in\mathbb{R}\times\mathbb{R}$. Then we have

$$\begin{aligned}
&f^*(y,x)\\
&=\sup_{(v,w)}\Big\{\langle v,y\rangle+\langle w,x\rangle-j(\tfrac{v+w}{\sqrt{2}})-g(\tfrac{v-w}{\sqrt{2}})\Big\}\\
&=\sup_{(v,w)}\Big\{\langle\tfrac{v+w}{2},x+y\rangle-\langle\tfrac{v-w}{2},x-y\rangle-j(\tfrac{v+w}{\sqrt{2}})-g(\tfrac{v-w}{\sqrt{2}})\Big\} && (5.46)\\
&=\sup_{(v,w)}\Big\{\langle\tfrac{v+w}{\sqrt{2}},\tfrac{x+y}{\sqrt{2}}\rangle-\langle\tfrac{v-w}{\sqrt{2}},\tfrac{x-y}{\sqrt{2}}\rangle-j(\tfrac{v+w}{\sqrt{2}})-g(\tfrac{v-w}{\sqrt{2}})\Big\}\\
&=\sup_{(s,t)}\Big\{\langle s,\tfrac{x+y}{\sqrt{2}}\rangle-\langle t,\tfrac{x-y}{\sqrt{2}}\rangle-j(s)-g(t)\Big\} && (5.47)\\
&=\sup_{s}\Big\{\langle s,\tfrac{x+y}{\sqrt{2}}\rangle-j(s)\Big\}+\sup_{t}\Big\{-\langle t,\tfrac{x-y}{\sqrt{2}}\rangle-g(t)\Big\}\\
&=j^*(\tfrac{x+y}{\sqrt{2}})+g^*(-\tfrac{x-y}{\sqrt{2}})\\
&=j(\tfrac{x+y}{\sqrt{2}})+g(\tfrac{x-y}{\sqrt{2}})\quad\big(\text{since } j^*=j \text{ by [7, Proposition 3.3(i)]}\big)\\
&=f(x,y).
\end{aligned}$$

Hence f is auto-conjugate.

Note that (5.46) holds since

$$\begin{aligned}
&\langle \tfrac{v+w}{2}, x+y\rangle - \langle \tfrac{v-w}{2}, x-y\rangle \\
&= \tfrac{1}{2}\Big(\langle v+w, x+y\rangle - \langle v-w, x-y\rangle\Big) \\
&= \tfrac{1}{2}\Big(\langle v,x\rangle + \langle v,y\rangle + \langle w,x\rangle + \langle w,y\rangle - \langle v,x\rangle + \langle v,y\rangle \\
&\quad + \langle w,x\rangle - \langle w,y\rangle\Big) \\
&= \tfrac{1}{2}\Big(2\langle v,y\rangle + 2\langle w,x\rangle\Big) \\
&= \langle v,y\rangle + \langle w,x\rangle.
\end{aligned}$$

In the following we show that (5.47) holds. Clearly, for every (v,w) there exists (s,t) such that

$$\begin{aligned}
&\langle \tfrac{v+w}{\sqrt{2}}, \tfrac{x+y}{\sqrt{2}}\rangle - \langle \tfrac{v-w}{\sqrt{2}}, \tfrac{x-y}{\sqrt{2}}\rangle - j(\tfrac{v+w}{\sqrt{2}}) - g(\tfrac{v-w}{\sqrt{2}}) \\
&= \langle s, \tfrac{x+y}{\sqrt{2}}\rangle - \langle t, \tfrac{x-y}{\sqrt{2}}\rangle - j(s) - g(t).
\end{aligned}$$

On the other hand, for every (s,t), there exists $v_0 = \frac{\sqrt{2}}{2}(s+t), w_0 = \frac{\sqrt{2}}{2}(s-t)$ such that

$$\begin{aligned}
&\langle \tfrac{v_0+w_0}{\sqrt{2}}, \tfrac{x+y}{\sqrt{2}}\rangle - \langle \tfrac{v_0-w_0}{\sqrt{2}}, \tfrac{x-y}{\sqrt{2}}\rangle - j(\tfrac{v_0+w_0}{\sqrt{2}}) - g(\tfrac{v_0-w_0}{\sqrt{2}}) \\
&= \langle s, \tfrac{x+y}{\sqrt{2}}\rangle - \langle t, \tfrac{x-y}{\sqrt{2}}\rangle - j(s) - g(t).
\end{aligned}$$

Hence (5.47) holds.

We now show that f is a representation for Id. First we show that $g(0) = 0$.

By assumptions, $g(0) \geq 0$. On the other hand,

$$g(0) = g^*(-0) = g^*(0) = \sup_v\{-g(v)\} \leq 0.$$

Hence $g(0) = 0$.

Then we have

$$\begin{aligned}
&(x, y) \in G(f) \\
&\Leftrightarrow f(x, y) = \langle x, y\rangle \\
&\Leftrightarrow \tfrac{1}{4}\|x + y\|^2 + g(\tfrac{x-y}{\sqrt{2}}) = \langle x, y\rangle \\
&\Leftrightarrow \tfrac{1}{4}\|x - y\|^2 + g(\tfrac{x-y}{\sqrt{2}}) = 0 \\
&\Leftrightarrow \tfrac{1}{4}\|x - y\|^2 = 0,\ g(\tfrac{x-y}{\sqrt{2}}) = 0 \quad \text{(by } g \geq 0\text{)} \\
&\quad \text{by } g(0) = 0 \\
&\Leftrightarrow x = y \Leftrightarrow (x, y) \in \operatorname{gra}\operatorname{Id}.
\end{aligned}$$

Hence f is an auto-conjugate representation for Id. ■

Remark 5.6.2 *If we set $g = j$ in Proposition 5.6.1, $f = P(F_{\mathrm{Id}}, F_{\mathrm{Id}}^{*\intercal})$.*

Now we give three examples based on Proposition 5.6.1.

Example 5.6.3 *The function*

$$g := \iota_{\mathbb{R}_+}$$

satisfies the conditions of Proposition 5.6.1. Figure 5.1 is corresponding to f.

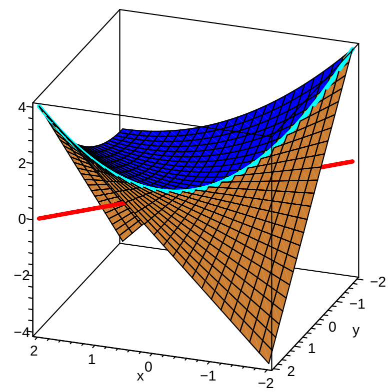

Figure 5.1: The function f (blue) and $z = xy$ (gold), and their intersection line (cyan), $\operatorname{gra}\operatorname{Id}$ (red).

Proof. Let $x \in \mathbb{R}$. We consider two cases.

Case 1: $x \geq 0$. We have

$$g^*(-x) = \sup_v \Big\{ \langle v, -x \rangle - \iota_{\mathbb{R}_+}(v) \Big\} = \sup_{v \geq 0} \Big\{ \langle v, -x \rangle \Big\} = 0 = g(x).$$

Case 2: $x < 0$. Then

$$g^*(-x) = \sup_v \Big\{ \langle v, -x \rangle - \iota_{\mathbb{R}_+}(v) \Big\} = \sup_{v \geq 0} \Big\{ \langle v, -x \rangle \Big\} = \infty = g(x).$$

Hence $g^*(-x) = g(x)$. ■

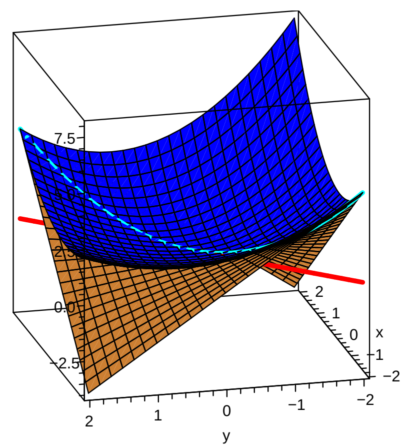

Figure 5.2: The function f (blue) and $z = xy$ (gold), and their intersection line (cyan), $\operatorname{gra}\operatorname{Id}$ (red) .

Example 5.6.4 *Set*

$$g(x) := \begin{cases} x^2, & \text{if } x \geq 0; \\ \frac{1}{4}x^2, & \text{if } x \leq 0 \end{cases}.$$

Then g satisfies the conditions of Proposition 5.6.1. Figure 5.2 is corresponding to f .

Proof. Let $x \in \mathbb{R}$. We consider two cases.

Case 1: $x \geq 0$. We have

$$\begin{aligned} g^*(-x) &= \sup_v \Big\{ \langle v, -x \rangle - g(v) \Big\} \\ &= \sup_{v \leq 0} \Big\{ \langle v, -x \rangle - g(v) \Big\} \quad (\text{since } g \geq 0, g(0) = 0) \\ &= \sup_{v \leq 0} \Big\{ \langle v, -x \rangle - \tfrac{1}{4} v^2 \Big\} \\ &= \sup_{v \leq 0} \Big\{ h_0(v) \Big\}, \end{aligned}$$

where $h_0(v) := \langle v, -x \rangle - \frac{1}{4} v^2$.

Let

$$0 = \nabla h_0(v) = -x - \tfrac{1}{2} v.$$

Then $v_0 = -2x \leq 0$ is a critical point of h_0. Since h_0 is concave on $\mathbb{R}_-$, its critical point is its maximizer. Then

$$g^*(-x) = h_0(v_0) = \langle -2x, -x \rangle - x^2 = x^2 = g(x).$$

Case 2: $x < 0$. We have

$$\begin{aligned} g^*(-x) &= \sup_v \Big\{ \langle v, -x \rangle - g(v) \Big\} \\ &= \sup_{v \geq 0} \Big\{ \langle v, -x \rangle - g(v) \Big\} \quad (\text{since } g \geq 0, g(0) = 0) \\ &= \sup_{v \geq 0} \Big\{ \langle v, -x \rangle - v^2 \Big\} \\ &= \sup_{v \geq 0} \Big\{ h_1(v) \Big\}, \end{aligned}$$

where $h_1(v) := \langle v, -x\rangle - v^2$

Let

$$0 = \nabla h_1(v) = -x - 2v.$$

Then $v_1 = -\frac{1}{2}x \geq 0$ is a critical point of h_1. Since h_1 is concave on $\mathbb{R}_+$, its critical point is its maximizer. Then

$$g^*(-x) = h_1(v_1) = \langle -\tfrac{1}{2}x, -x\rangle - \tfrac{1}{4}x^2 = \tfrac{1}{4}x^2 = g(x).$$

Hence $g^*(-x) = g(x)$. ■

Example 5.6.5 *Set* $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

$$g(x) := \begin{cases} \frac{1}{p}x^p, & \textit{if } x \geq 0; \\ \frac{1}{q}(-x)^q, & \textit{if } x \leq 0. \end{cases}$$

satisfies the conditions of Proposition 5.6.1. Figure 5.3 is corresponding to f.

Proof. Let $x \in \mathbb{R}$. We consider two cases.

Case 1: $x \geq 0$. We have

$$\begin{aligned} g^*(-x) &= \sup_v \Big\{\langle v, -x\rangle - g(v)\Big\} \\ &= \sup_{v\leq 0} \Big\{\langle v, -x\rangle - g(v)\Big\} \quad (\text{since } g \geq 0, g(0) = 0) \\ &= \sup_{v\leq 0} \Big\{\langle v, -x\rangle - \tfrac{1}{q}(-v)^q\Big\} \\ &= \sup_{v\leq 0} \Big\{g_0(v)\Big\}, \end{aligned}$$

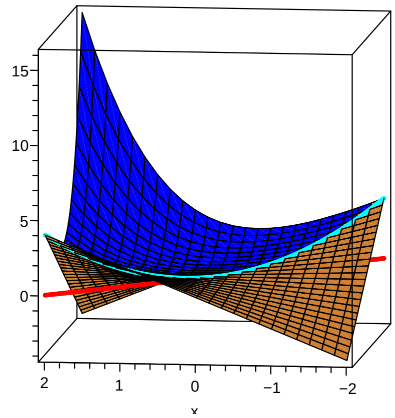

Figure 5.3: The function f (blue) and $z = xy$ (gold) , and their intersection line (cyan), $\operatorname{gra}\operatorname{Id}$ (red), where $p = 4$.

where

$$g_0(v) := \langle v, -x\rangle - \tfrac{1}{q}(-v)^q.$$

Then let

$$0 = \nabla g_0(v) = -x + (-v)^{q-1}.$$

Thus $v_0 := -x^{\frac{1}{q-1}} \leq 0$ is a critical point of g_0.

Since $\nabla^2 g_0(v) = -(q-1)(-v)^{q-2} \leq 0 \quad (\forall v < 0)$, by the continuity of g_0, g_0

is concave on $\mathbb{R}_-$. Then its critical point is its maximizer. Thus

$$\begin{aligned}
g^*(-x) &= g_0(v_0) \\
&= \langle -x^{\frac{1}{q-1}}, -x\rangle - \tfrac{1}{q}x^{\frac{q}{q-1}} \\
&= (1-\tfrac{1}{q})x^{\frac{q}{q-1}} \\
&= \tfrac{1}{p}x^p \quad (\text{by } \tfrac{1}{p}+\tfrac{1}{q}=1) \\
&= g(x).
\end{aligned}$$

Case 2: $x < 0$. We have

$$\begin{aligned}
g^*(-x) &= \sup_v \Big\{\langle v, -x\rangle - g(v)\Big\} \\
&= \sup_{v\geq 0} \Big\{\langle v, -x\rangle - g(v)\Big\} \quad (\text{since } g \geq 0, g(0) = 0) \\
&= \sup_{v\geq 0} \Big\{\langle v, -x\rangle - \tfrac{1}{p}v^p\Big\} \\
&= \sup_{v\geq 0} \Big\{g_1(v)\Big\},
\end{aligned}$$

where

$$g_1(v) := \langle v, -x\rangle - \tfrac{1}{p}v^p.$$

Then let

$$0 = \nabla g_1(v) = -x - v^{p-1}.$$

Thus $v_1 := (-x)^{\frac{1}{p-1}} > 0$ is a critical point of g_1.

Since $\nabla^2 g_1(v) = -(p-1)v^{p-2} \leq 0 \quad (\forall v > 0)$, by the continuity of g_1, g_1 is concave on $\mathbb{R}_+$. Then its critical point is its maximizer. Thus

$$\begin{aligned}
g^*(-x) &= g_1(v_1) \\
&= \langle (-x)^{\frac{1}{p-1}}, -x \rangle - \tfrac{1}{p}(-x)^{\frac{p}{p-1}} \\
&= (1 - \tfrac{1}{p})(-x)^{\frac{p}{p-1}} \\
&= \tfrac{1}{q}(-x)^q \\
&= g(x).
\end{aligned}$$

Hence $g^*(-x) = g(x)$. ■

Remark 5.6.6 *Example 5.6.3, 5.6.4 and 5.6.5 each provide a function f that is an auto-conjugate representation for* Id *with $f \neq P(F_{\mathrm{Id}}, F_{\mathrm{Id}}^{*\intercal})$.*

Chapter 6

Calculation of the auto-conjugates of $\partial(-\ln)$

Throughout this chapter, $-\ln$ is meant in the extended real valued sense, i.e., $-\ln(x) = \infty$ for $x \leq 0$.

In Chapter 5, Proposition 5.3.12 shows that $h_{F_A} = P(F_A, F_A^{*\intercal})$ for a linear and monotone operator A. Now we will show that for the nonlinear monotone operator $\partial(-\ln)$ we have $P(F_{\partial(-\ln)}, F^{*\intercal}_{\partial(-\ln)}) \neq h_{F_{\partial(-\ln)}}$. Throughout the chapter, we denote

$$
\begin{aligned}
C &:= \Big\{(x, x^*) \in \mathbb{R} \times \mathbb{R} \mid x \leq -\tfrac{1}{x^*} < 0\Big\} \\
D &:= \Big\{(x, x^*) \in \mathbb{R} \times \mathbb{R} \mid x^* \leq -\tfrac{1}{2x} < 0\Big\} \\
E &:= \Big\{(x, x^*) \in \mathbb{R} \times \mathbb{R} \mid x^* \leq -\tfrac{1}{4x} < 0\Big\}.
\end{aligned}
$$

6.1 Fitzpatrick function for $\partial(-\ln)$

Fact 6.1.1 *Let $f = -\ln$. Then*

$$F_{\partial f}(x, x^*) = \begin{cases} 1 - 2x^{\frac{1}{2}}(-x^*)^{\frac{1}{2}}, & \text{if } x \geq 0, x^* \leq 0; \\ \infty, & \text{otherwise.} \end{cases}$$

Proof. See [8, Example 3.4]. ■

Fact 6.1.2 *Let $f = -\ln$. Then*

$$F^*_{\partial f}(x^*, x) = -1 + \iota_C(x^*, x).$$

Proof. See [8, Example 3.4]. ■

Fact 6.1.3 (Rockafellar) *Let $f = -\ln$.*

$$f^*(x^*) = \begin{cases} -1 - \ln(-x^*), & \text{if } x^* < 0; \\ \infty, & \text{otherwise,} \end{cases}$$

Proof. See [8, Example 3.4]. ■

Remark 6.1.4 *Let $f = -\ln$. Recall*

$$(f \oplus f^*)(x, x^*) := f(x) + f^*(x^*), \quad \forall (x, x^*) \in \mathbb{R} \times \mathbb{R}.$$

By Fact 6.1.3,

$$\operatorname{dom}(f \oplus f^*) = \mathbb{R}_{++} \times \mathbb{R}_{--}.$$

Proposition 6.1.5

$$D \underset{\neq}{\subsetneq} E \underset{\neq}{\subsetneq} \mathbb{R}_{++} \times \mathbb{R}_{--}.$$

Proof. We first verify that $D \underset{\neq}{\subsetneq} E$.

Let $(x, x^*) \in D$. Thus $x > 0$. Then $-\frac{1}{2x} \le -\frac{1}{4x}$. By $(x, x^*) \in D$,

$$x^* \le -\frac{1}{2x} \le -\frac{1}{4x} < 0.$$

Thus $(x, x^*) \in E$. Then $D \subset E$.

On the other hand, $(1, -\frac{1}{4}) \in E$, but $(1, -\frac{1}{4}) \notin D$. Thus $D \neq E$.

Hence $D \underset{\neq}{\subsetneq} E$.

It is clear we have $E \underset{\neq}{\subsetneq} \mathbb{R}_{++} \times \mathbb{R}_{--}$.

Thus combining the results above, $D \underset{\neq}{\subsetneq} E \underset{\neq}{\subsetneq} \mathbb{R}_{++} \times \mathbb{R}_{--}$. ■

6.2 Proximal average of $\partial(-\ln)$ and $h_{F_{\partial f}}$

Proposition 6.2.1 *Let $f = -\ln$. Then*

$$\operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}) = E.$$

Proof. Let By [7, Theorem 4.6],

$$\operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}) = \tfrac{1}{2} \operatorname{dom} F_{\partial f} + \tfrac{1}{2} \operatorname{dom} F_{\partial f}^{*\intercal}. \tag{6.1}$$

In the following we will show that

$$\operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}) = \tfrac{1}{2} \operatorname{dom} F_{\partial f}^{*\intercal}. \tag{6.2}$$

By Fact 6.1.1, $(0,0) \in \operatorname{dom} F_{\partial f}$, then by (6.1), we have

$$\tfrac{1}{2} \operatorname{dom} F_{\partial f}^{*\intercal} \subset \operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}). \tag{6.3}$$

Next we show that

$$\tfrac{1}{2} \operatorname{dom} F_{\partial f}^{*\intercal} = E. \tag{6.4}$$

Indeed,

$$\begin{aligned}
&(x, x^*) \in \tfrac{1}{2} \operatorname{dom} F_{\partial f}^{*\intercal} \\
&\Leftrightarrow (2x^*, 2x) \in \operatorname{dom} F_{\partial f}^{*} \Leftrightarrow 2x^* \leq -\tfrac{1}{2x} < 0 \quad \text{(by Fact 6.1.2)} \\
&\Leftrightarrow x^* \leq -\tfrac{1}{4x} < 0 \Leftrightarrow (x, x^*) \in E.
\end{aligned}$$

Hence (6.4) holds.

Then by (6.4) and (6.3),

$$E \subset \operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}). \tag{6.5}$$

In the following, we will verify that

$$\tfrac{1}{2} \operatorname{dom} F_{\partial f} + E \subset E. \tag{6.6}$$

Let $(y, y^*) \in \frac{1}{2} \operatorname{dom} F_{\partial f}$ and $(x, x^*) \in E$. By Fact 6.1.1 we have

$$y \geq 0,\ y^* \leq 0,\ x > 0,\ x^* < 0,\ 4xx^* \leq -1.$$

Thus $x + y \geq x > 0, x^* + y^* \leq x^* < 0$. Then we have $4(x+y)(x^*+y^*) \leq 4xx^* \leq -1$, i.e., $(x, x^*) + (y, y^*) \in E$. Hence (6.6) holds. Thus by (6.6), (6.4) and (6.1), $\operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}) \subset E$. Then by (6.5), $\operatorname{dom} P(F_{\partial f}, F_{\partial f}^{*\intercal}) = E$. ■

Lemma 6.2.2 *Let $x, x^*, y^* \in \mathbb{R}$ with $y^* \leq 0$. Then*

$$\iota_C(2x^* - 2y^*, x) = \iota_C(2x^* - 2y^*, x) + \iota_D(x, x^*).$$

Proof. We consider two cases.

Case 1: $(2x^* - 2y^*, x) \notin C$. Clear.

Case 2: $(2x^* - 2y^*, x) \in C$. By assumptions,

$$2x^* \leq 2x^* - 2y^* \leq -\tfrac{1}{x} < 0 \Rightarrow x^* \leq -\tfrac{1}{2x} < 0 \quad (\text{by } y^* \leq 0).$$

Thus $(x, x^*) \in D$. Then $\iota_D(x, x^*) = 0$. ■

Remark 6.2.3 *Let $x, x^* \in \mathbb{R}$. Then*

$$\iota_{\mathbb{R}_+}(x) + \iota_D(x, x^*) = \iota_D(x, x^*).$$

Proof. Follows directly from the definition of D . ■

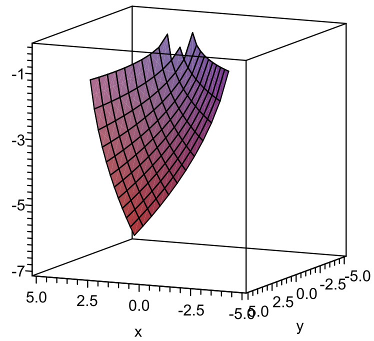

Figure 6.1: The function $h_{F_{\partial f}}$.

Proposition 6.2.4 *Let* $f = -\ln$. *Then*

$$h_{F_{\partial f}}(x, x^*) = -(-1 - 2xx^*)^{\frac{1}{2}} + \iota_D(x, x^*), \quad \forall (x, x^*) \in \mathbb{R} \times \mathbb{R}.$$

Consequently, $\operatorname{dom} h_{F_{\partial f}} = D$. *Figure 6.1 illustrates* $h_{F_{\partial f}}$.

Proof. Let $(x, x^*) \in \mathbb{R} \times \mathbb{R}$. By Fact 6.1.1 and Fact 6.1.2, we have

$$\begin{aligned}
& h_{F_{\partial f}}(x, x^*) \\
&= \inf_{y^*} \left\{ \tfrac{1}{2} F_{\partial f}(x, 2y^*) + \tfrac{1}{2} F^*_{\partial f}(2x^* - 2y^*, x) \right\} \\
&= \inf_{y^* \le 0} \left\{ \tfrac{1}{2} - |x|^{\frac{1}{2}} (-2y^*)^{\frac{1}{2}} + \iota_{\mathbb{R}_+}(x) + \tfrac{1}{2} F^*_{\partial f}(2x^* - 2y^*, x) \right\} \\
&= \inf_{y^* \le 0} \left\{ \tfrac{1}{2} - |x|^{\frac{1}{2}} (-2y^*)^{\frac{1}{2}} + \iota_{\mathbb{R}_+}(x) + \iota_C(2x^* - 2y^*, x) - \tfrac{1}{2} \right\} \\
&= \inf_{y^* \le 0} \left\{ -|x|^{\frac{1}{2}} (-2y^*)^{\frac{1}{2}} + \iota_{\mathbb{R}_+}(x) + \iota_C(2x^* - 2y^*, x) + \iota_D(x, x^*) \right\} \qquad (6.7) \\
&= \inf_{y^* \le 0} \left\{ -|x|^{\frac{1}{2}} (-2y^*)^{\frac{1}{2}} + \iota_C(2x^* - 2y^*, x) + \iota_D(x, x^*) \right\}, \qquad (6.8)
\end{aligned}$$

where (6.7) holds by Lemma 6.2.2, (6.8) by Remark 6.2.3.

Now we consider two cases.

Case 1: $(x, x^*) \notin D$. Thus $h_{F_{\partial f}}(x, x^*) = \infty$.

Case 2: $(x, x^*) \in D$. Thus $x > 0$. Then

$$\begin{aligned}
& h_{F_{\partial f}}(x, x^*) \\
&= \inf_{y^* \leq 0} \Big\{ -x^{\frac{1}{2}}(-2y^*)^{\frac{1}{2}} + \iota_C(2x^* - 2y^*, x) \Big\} \\
&= \inf_{(2x^* - 2y^* \leq -\frac{1}{x} < 0,\ y^* \leq 0)} \Big\{ -x^{\frac{1}{2}}(-2y^*)^{\frac{1}{2}} \Big\} && (6.9) \\
&= -(2x)^{\frac{1}{2}} \sup_{(2x^* - 2y^* \leq -\frac{1}{x} < 0,\ y^* \leq 0)} \Big\{ (-y^*)^{\frac{1}{2}} \Big\} \\
&= -(2x)^{\frac{1}{2}} \sup_{(0 \leq -2y^* \leq -2x^* - \frac{1}{x})} \Big\{ (-y^*)^{\frac{1}{2}} \Big\} \\
&= -(2x)^{\frac{1}{2}} (-\tfrac{1}{2x} - x^*)^{\frac{1}{2}} && (6.10) \\
&= -(-1 - 2xx^*)^{\frac{1}{2}} \quad (\text{by } x > 0),
\end{aligned}$$

where (6.9) holds by letting $2x^* - 2y^* \in C$. (6.10) holds by $0 \leq -\frac{1}{2x} - x^*$ since $(x, x^*) \in D$.

Thus combining the results above,

$$h_{F_{\partial f}}(x, x^*) = -(-1 - 2xx^*)^{\frac{1}{2}} + \iota_D(x, x^*), \quad \forall (x, x^*) \in \mathbb{R} \times \mathbb{R}.$$

■

Corollary 6.2.5 *Let $f = -\ln$. Then $P(F_{\partial f}, F_{\partial f}^{*\intercal})$, $f \oplus f^*$ and $h_{F_{\partial f}}$ are three different functions.*

Proof. By Remark 6.1.4, we have $\operatorname{dom}(f \oplus f^*) = \mathbb{R}_{++} \times \mathbb{R}_{--}$. Then by Proposition 6.2.1 and Proposition 6.2.4, $\operatorname{dom} P(F_{\partial f}, F^{*\intercal}_{\partial f}) = E$ and $\operatorname{dom} h_{F_{\partial f}} = D$. By Proposition 6.1.5,

$$\operatorname{dom} h_{F_{\partial f}} \underset{\neq}{\subsetneq} \operatorname{dom} P(F_{\partial f}, F^{*\intercal}_{\partial f}) \underset{\neq}{\subsetneq} \operatorname{dom}(f \oplus f^*).$$

Hence $P(F_{\partial f}, F^{*\intercal}_{\partial f}), f \oplus f^*$ and $h_{F_{\partial f}}$ are all different. ■

Remark 6.2.6 *We don't have an explicit formula for $P(F_{\partial(-\ln)}, F^{*\intercal}_{\partial(-\ln)})$.*

Chapter 7

Proximal averages of monotone operators with linear graphs

We have given some auto-conjugate representation results for linear and monotone operators. Now we extend them to monotone operators with linear graphs. Background worked on linear relations can be found in the book by Cross [16].

7.1 Adjoint process of operators with linear graphs

Definition 7.1.1 *Let C be a nonempty cone of $\mathbb{R}^n$. The* polar of C, C^-, *is defined by*

$$C^- := \Big\{x^* \mid \langle c, x^* \rangle \leq 0, \forall c \in C\Big\}.$$

Remark 7.1.2 *If C is a linear subspace of $\mathbb{R}^n$, then by Lemma 2.1.28, $C^- = C^\perp$.*

Definition 7.1.3 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be such that* $\operatorname{gra} A$ *is a convex cone. The* adjoint process *of A, A^*, is defined by*

$$\operatorname{gra} A^* := \Big\{(x, x^*) \mid (x^*, -x) \in (\operatorname{gra} A)^- \Big\}.$$

Lemma 7.1.4 *[16, Proposition III.1.3] Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be such that* $\operatorname{gra} A$ *is linear. Suppose $k \in \mathbb{R}$ with $k \neq 0$. Then $(kA)^* = kA^*$.*

Proof. By Remark 7.1.2,

$$\begin{aligned}
(x, x^*) \in \operatorname{gra}(kA)^* &\Leftrightarrow (x^*, -x) \in (\operatorname{gra} kA)^- = (\operatorname{gra} kA)^\perp \\
&\Leftrightarrow \langle (x^*, -x), (v, v^*)\rangle = 0,\ \forall (v, v^*) \in \operatorname{gra}(kA) \\
&\Leftrightarrow \tfrac{1}{k}\langle (x^*, -x), (v, v^*)\rangle = 0,\ \forall (v, v^*) \in \operatorname{gra}(kA) \\
&\Leftrightarrow \langle (\tfrac{1}{k}x^*, -x), (v, \tfrac{1}{k}v^*)\rangle = 0,\ \forall (v, v^*) \in \operatorname{gra}(kA) \\
&\Leftrightarrow \langle (\tfrac{1}{k}x^*, -x), (v, w^*)\rangle = 0,\ \forall (v, w^*) \in \operatorname{gra} A \\
&\Leftrightarrow (x, \tfrac{1}{k}x^*) \in \operatorname{gra} A^* \Leftrightarrow x^* \in kA^*x.
\end{aligned}$$

Hence $(kA)^* = kA^*$. ■

Remark 7.1.5 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be such that* $\operatorname{gra} A$ *is linear. Then* $\operatorname{gra} A^*$ *is a linear graph.*

Remark 7.1.6 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be such that* $\operatorname{gra} A$ *is linear. Then* $A^*0 = (\operatorname{dom} A)^\perp$.

Proof. See [16, Proposition III.1.4 (b)]. ■

Definition 7.1.7 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is linear. We say that* A *is* symmetric *if* $A^* = A$.

Definition 7.1.8 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is linear. We say that* A *is* antisymmetric *if* $A^* = -A$.

Fact 7.1.9 *Let* $A, B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *and* $\operatorname{gra} B$ *are linear. Then* $(A+B)^* = A^* + B^*$.

Proof. See [11, Theorem 7.4]. ■

Fact 7.1.10 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a closed convex cone. Then* $\operatorname{gra} A^{**} = -\operatorname{gra} A$.

Proof. See [13, Exercises 7 page 119]. ■

Corollary 7.1.11 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is linear. Then* $A^{**} = A$.

Proof. Since $\operatorname{gra} A$ is a linear subspace, $-\operatorname{gra} A = \operatorname{gra} A$. Thus by Fact 7.1.10, $\operatorname{gra} A^{**} = \operatorname{gra} A$. Hence $A^{**} = A$. ■

Corollary 7.1.12 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Then* $\operatorname{dom} A^* = (A0)^{\perp}$.

Proof. By Remark 7.1.5 and Remark 7.1.6, we have $(A^*)^*0 = (\operatorname{dom} A^*)^{\perp}$. Then by Corollary 7.1.11, $A0 = (\operatorname{dom} A^*)^{\perp}$. Thus $\operatorname{dom} A^* = (A0)^{\perp}$. ■

Remark 7.1.13 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is linear. By Fact 7.1.9, Remark 7.1.5, Corollary 7.1.11 and Lemma 7.1.4,* $\frac{A+A^*}{2}$ *is symmetric and* $\frac{A-A^*}{2}$ *is antisymmetric.*

Definition 7.1.14 (Symmetric and antisymmetric part) *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is linear. Then* $A_+ = \frac{1}{2}A + \frac{1}{2}A^*$ *is the* symmetric part *of* A, *and* $A_\circ = \frac{1}{2}A - \frac{1}{2}A^*$ *is the* antisymmetric part *of* A.

Remark 7.1.15 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Then by Corollary 7.1.12,* $\operatorname{dom} A_+ = \operatorname{dom} A_\circ = \operatorname{dom} A \cap (A0)^\perp$.

Corollary 7.1.16 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Then* A *can be decomposed into the sum of a symmetric operator with a linear graph and an antisymmetric operator with a linear graph, if and only if,* $\operatorname{dom} A = (A0)^\perp$. *In that case,* A *can be decomposed as :* $A = A_+ + A_\circ$.

Proof. "$\Rightarrow$" Let $B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a symmetric operator with a linear graph and $C\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be an antisymmetric operator with a linear graph such that $A = B + C$. By Fact 7.1.9, $A^* = B^* + C^* = B - C$. Then $\operatorname{dom} A^* = \operatorname{dom} B \cap \operatorname{dom} C = \operatorname{dom} A$. By Corollary 7.1.12, $\operatorname{dom} A = (A0)^\perp$.
"$\Leftarrow$" By Remark 7.1.15, $\operatorname{dom} A_+ = \operatorname{dom} A_\circ = \operatorname{dom} A$. By Corollary 7.1.12, $\operatorname{dom} A^* = (A0)^\perp = \operatorname{dom} A$. Thus, by Remark 7.1.5 and Proposition 4.1.3(iii),

$$
\begin{aligned}
A_+x + A_\circ x &= \tfrac{1}{2}(Ax + A^*x + Ax - A^*x) = Ax + A^*0 \\
&= Ax + (\operatorname{dom} A)^\perp = Ax + A0 \quad \text{(by Remark 7.1.6)} \\
&= Ax \quad \text{(by Proposition 4.1.3(iii))}, \quad \forall x \in \operatorname{dom} A.
\end{aligned}
$$

■

Remark 7.1.17 *Consider an operator* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *with* $\operatorname{gra} A = \{(0,0)\}$. *Then we have* $\operatorname{dom} A = \{0\} \neq \mathbb{R}^n = (A0)^\perp$. *Clearly,* $\operatorname{gra} A^* = \mathbb{R}^n \times \mathbb{R}^n$.

Thus $(A_+ + A_\circ)0 = A_+0 + A_\circ 0 = \mathbb{R}^n + \mathbb{R}^n = \mathbb{R}^n \neq A0$. *Thus* $A \neq A_+ + A_\circ$. *By Proposition 7.1.16, A can not be decomposed into the sum of a symmetric operator with a linear graph and an antisymmetric operator with a linear graph.*

Remark 7.1.18 *Let S be a linear subspace of $\mathbb{R}^n$. Then S is closed.*

Corollary 7.1.19 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then $A = A_+ + A_\circ$.*

Proof. By Remark 7.1.18, Proposition 4.2.5 and Corollary 7.1.16. ■

Definition 7.1.20 *Let C be a nonempty convex subset of $\mathbb{R}^n$ and $x_0 \in \mathbb{R}^n$. The* normal cone of C at x_0, *$N_C(x_0)$, is defined by*

$$N_C(x_0) := \begin{cases} \Big\{x^* \mid \langle x^*, c - x_0\rangle \leq 0, \forall c \in C\Big\}, & \text{if } x_0 \in C; \\ \varnothing, & \text{otherwise.} \end{cases}$$

Fact 7.1.21 *Let C be a nonempty convex subset of $\mathbb{R}^n$ and $x_0 \in C$. Then $N_C(x_0) = \partial\iota_C(x_0)$. If C is a linear subspace of $\mathbb{R}^n$, then $N_C(x_0) = \partial\iota_C(x_0) = C^\perp$ by Fact 2.1.29.*

Remark 7.1.22 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and S be a linear subspace of $\mathbb{R}^n$. Then* $\operatorname{gra}(A + N_S)$ *is a linear subspace of $\mathbb{R}^n \times \mathbb{R}^n$.*

Fact 7.1.23 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and S be a linear subspace of $\mathbb{R}^n$ such that $A = B + N_S$. Then $A^* = B^* + N_S$.*

Proof. Since $\operatorname{gra} A$ is a linear subspace by Remark 7.1.22, then by Remark 7.1.2 and Fact 7.1.21 we have

$$\begin{aligned}
&(x, x^*) \in \operatorname{gra} A^* \\
&\qquad \Leftrightarrow (x^*, -x) \in (\operatorname{gra} A)^- \\
&\qquad \Leftrightarrow (x^*, -x) \in (\operatorname{gra} A)^\perp \\
&\qquad \Leftrightarrow \langle x^*, y\rangle - \langle x, y^*\rangle = 0, \quad \forall y^* \in Ay \\
&\qquad \Leftrightarrow \langle x^*, y\rangle - \langle x, By + S^\perp\rangle = 0, \quad \forall y \in S.
\end{aligned} \tag{7.1}$$

Let $y = 0$ in (7.1). We have $\langle x, S^\perp\rangle = 0$. Thus $x \in S$. Then

$$\begin{aligned}
&(x, x^*) \in \operatorname{gra} A^* \\
&\qquad \Leftrightarrow x \in S,\ \langle x^*, y\rangle - \langle x, By\rangle = 0, \quad \forall y \in S \\
&\qquad \Leftrightarrow x \in S,\ \langle x^* - B^*x, y\rangle = 0, \quad \forall y \in S \\
&\qquad \Leftrightarrow x \in S,\ (x^* - B^*x) \perp S \\
&\qquad \Leftrightarrow x \in S,\ x^* \in B^*x + S^\perp \\
&\qquad \Leftrightarrow x^* \in (B^* + N_S)(x) \quad \text{(by Fact 7.1.21).}
\end{aligned}$$

Hence $A^* = B^* + N_S$. ■

Remark 7.1.24 *Fact 7.1.23 is a special case of [13, Exercises 14 (f) page 120].*

Remark 7.1.25 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and S be a linear subspace of $\mathbb{R}^n$. Suppose $A = B + N_S$. Then by Fact 7.1.23, $A_+ = B_+ + N_S$, $A_\circ =$*

$B_\circ + N_S$ *and* $A = A_+ + A_\circ$.

Now we recall the definition of Q_A.

Definition 7.1.26 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$. *We define* Q_A *by*

$$Q_A x = \begin{cases} P_{Ax}x, & \text{if } x \in \operatorname{dom} A; \\ \varnothing, & \text{otherwise.} \end{cases}$$

Proposition 7.1.27 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Then* Q_A *is single-valued and linear on* $\operatorname{dom} A$, *and* Q_A *is a selection of* A.

Proof. Since $A0$ is a closed subspace by Proposition 4.1.3(i) and Remark 7.1.18, Ax ($\forall x \in \operatorname{dom} A$) is a closed convex by Proposition 4.1.3(ii). By Fact 4.3.1, Q_A is single-valued on $\operatorname{dom} A$ and Q_A is a selection of A. Very similar to the proof of Proposition 4.3.6, we have Q_A is linear on $\operatorname{dom} A$. ■

Corollary 7.1.28 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$Ax = \begin{cases} Q_A P_{\operatorname{dom} A} x + A0, & \text{if } x \in \operatorname{dom} A; \\ \varnothing, & \text{otherwise,} \end{cases}$$

where $Q_A P_{\operatorname{dom} A}$ *is linear.*

Proof. By Proposition 7.1.27 and Proposition 4.1.3(ii). ■

Proposition 7.1.29 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A$ *is a linear subspace. Assume* A *can be decomposed into the sum of a symmetric operator* *with a linear graph and an antisymmetric operator with a linear graph. Then such a decomposition is not unique.*

Proof. By Corollary 7.1.28, Corollary 7.1.16 and Fact 7.1.21,

$$A = Q_A P_{\operatorname{dom} A} + N_{\operatorname{dom} A}.$$

Thus,

$$A = (Q_A P_{\operatorname{dom} A})_+ + \big((Q_A P_{\operatorname{dom} A})_\circ + N_S\big) = ((Q_A P_{\operatorname{dom} A})_+ + N_S) + (Q_A P_{\operatorname{dom} A})_\circ.$$

By Fact 7.1.23, $(Q_A P_{\operatorname{dom} A})_+, (Q_A P_{\operatorname{dom} A})_+ + N_S$ are symmetric and $\big((Q_A P_{\operatorname{dom} A})_\circ + N_S\big), (Q_A P_{\operatorname{dom} A})_\circ$ are antisymmetric. Since $(Q_A P_{\operatorname{dom} A})_+ \neq (Q_A P_{\operatorname{dom} A})_+ + N_S$ and $(Q_A P_{\operatorname{dom} A})_\circ + N_S\big) \neq (Q_A P_{\operatorname{dom} A})_\circ$ as $S \neq \mathbb{R}^n$, the decomposition is not unique. ■

Theorem 7.1.30 *Let* $A, B, C\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be such that* $\operatorname{gra} A, \operatorname{gra} B$ *and* $\operatorname{gra} C$ *are linear subspaces. Assume* B *is symmetric and* C *is antisymmetric such that* $A = B + C$ *and* $\operatorname{dom} B = \operatorname{dom} C$. *Then* $B = A_+, C = A_\circ$.

Proof. By Fact 7.1.9, $A^* = B - C$. Thus by assumptions, $\operatorname{dom} B = \operatorname{dom} C = \operatorname{dom} A = \operatorname{dom} A^*$. Thus $\operatorname{dom} A_+ = \operatorname{dom} A_\circ = \operatorname{dom} B = \operatorname{dom} C = \operatorname{dom} A$. By Corollary 7.1.12, $\operatorname{dom} B = \operatorname{dom} B^* = (B0)^\perp, \operatorname{dom} C = \operatorname{dom} C^* = (C0)^\perp$. Thus $(B0)^\perp = (C0)^\perp$. Since $C0, B0$ are closed linear subspaces by Proposition 4.1.3(i) and Remark 7.1.18, $B0 = C0$. Let $x \in \operatorname{dom} A$. Then by

Proposition 4.1.3(iii) and Proposition 4.1.3(ii),

$$A_+x = \tfrac{1}{2}(Bx + Cx + Bx - Cx) = Bx + C0 = Bx + B0 = Bx,$$

$$A_\circ x = \tfrac{1}{2}(Bx + Cx - Bx + Cx) = B0 + Cx = C0 + Cx = Cx.$$

Hence $B = A_+, C = A_\circ$. ■

Corollary 7.1.31 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$. *Then $A = P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A} + N_{\operatorname{dom} A}$, where $P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A}$ is linear and monotone.*

Proof. Since $\operatorname{dom} A$ is a closed linear subspace by Remark 7.1.18, then by Theorem 4.4.1, $P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A}$ is linear and monotone, and $A = \partial(q_{\widetilde{A}} + \iota_{\operatorname{dom} A}) + \widetilde{A}_\circ$, where $\widetilde{A} = P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A}$.
Then by Fact 2.1.18, Fact 2.1.30 and Fact 7.1.21,

$$A = \widetilde{A}_+ + \partial\iota_{\operatorname{dom} A} + \widetilde{A}_\circ = P_{\operatorname{dom} A} Q_A P_{\operatorname{dom} A} + N_{\operatorname{dom} A}.$$

■

Remark 7.1.32 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n$. *Then* $\operatorname{gra} A^{-1}$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$.

7.2 Fitzpatrick functions of monotone operators with linear graphs

Definition 7.2.1 *Assume* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. *The* set-valued inverse mapping, $A^{-1}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, *is defined by*

$$x \in A^{-1}y \quad \Leftrightarrow \quad y \in Ax.$$

Definition 7.2.2 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *and* S *be a subset of* $\mathbb{R}^n$. *Then* AS *is defined by*

$$AS := \Big\{x^* \mid x^* \in As, \quad \exists s \in S\Big\}.$$

Proposition 7.2.3 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and* S *be a linear subspace of* $\mathbb{R}^n$ *such that* $A = B + N_S$. *Then*

(i) $x \in \operatorname{ran} A \Leftrightarrow x + S^{\perp} \subset \operatorname{ran} A$

(ii) $A^{-1}x = A^{-1}(x + S^{\perp})$.

Proof. (i): By Fact 7.1.21, $\operatorname{ran} A = \operatorname{ran}(B\mid_S) + S^{\perp}$. Thus $S^{\perp} + \operatorname{ran} A = \operatorname{ran} A$. Then

$$x \in \operatorname{ran} A \Leftrightarrow x + S^{\perp} \subset S^{\perp} + \operatorname{ran} A = \operatorname{ran} A.$$

Hence (i) holds.

(ii): Clearly, $A^{-1}x \subset A^{-1}(x + S^{\perp})$. In the following we show $A^{-1}(x + S^{\perp}) \subset A^{-1}x$.

By Fact 7.1.21,

$$\begin{aligned}
y \in A^{-1}(x+S^{\perp}) &\Rightarrow y \in A^{-1}(x+t), \quad \exists t \in S^{\perp} \\
&\Rightarrow x+t \in Ay = By + N_S(y) = By + S^{\perp}, \quad y \in S \\
&\Rightarrow x \in By + S^{\perp} = By + N_S(y) = Ay \\
&\Rightarrow y \in A^{-1}x.
\end{aligned}$$

Thus $A^{-1}(x+S^{\perp}) \subset A^{-1}x$. Hence $A^{-1}x = A^{-1}(x+S^{\perp})$. ∎

Lemma 7.2.4 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and symmetric, and S be a subspace of $\mathbb{R}^n$. Suppose that $x \in \operatorname{ran}(B+N_S)$. Then $\langle x,\ (B+N_S)^{-1}x\rangle$ is single-valued. Moreover, if $y_0 \in (B+N_S)^{-1}x$, then $\langle x,\ (B+N_S)^{-1}x\rangle = \langle y_0, By_0\rangle$.*

Proof. Let $x_1^*, x_2^* \in (B+N_S)^{-1}x$. Then $x_1^*, x_2^* \in S$ and by Fact 7.1.21,

$$x \in (B+N_S)x_1^* = Bx_1^* + S^{\perp}, \quad x \in (B+N_S)x_2^* = Bx_2^* + S^{\perp}. \tag{7.2}$$

Then we have

$$B(x_1^* - x_2^*) \in S^{\perp}. \tag{7.3}$$

By (7.2), there exists $t \in S^{\perp}$ such that $x = Bx_1^* + t$. Then

$$\begin{aligned}
\langle x,\ x_1^* - x_2^* \rangle &= \langle Bx_1^* + t,\ x_1^* - x_2^* \rangle \\
&= \langle Bx_1^*,\ x_1^* - x_2^* \rangle \qquad (7.4) \\
&= \langle x_1^*,\ B(x_1^* - x_2^*) \rangle \\
&= 0, \qquad (7.5)
\end{aligned}$$

in which, (7.4) holds by $t \in S^{\perp}$ and $x_1^* - x_2^* \in S$, and (7.5) holds by (7.3) and $x_1^* \in S$.

Thus $\langle x, x_1^* \rangle = \langle x, x_2^* \rangle$. Hence $\langle x,\ (B + N_S)^{-1}x \rangle$ is single-valued.

Let $y_0 \in (B + N_S)^{-1}x$. Then $y_0 \in S$ and $x \in (B + N_S)y_0 = By_0 + S^{\perp}$ by Fact 7.1.21. Let $t_0 \in S^{\perp}$ such that $x = By_0 + t_0$. Since $\langle x,\ (B + N_S)^{-1}x \rangle$ is single-valued,

$$\langle x,\ (B+N_S)^{-1}x \rangle = \langle x, y_0 \rangle = \langle By_0 + t_0, y_0 \rangle = \langle y_0, By_0 \rangle \text{ (by } y_0 \in S, t_0 \in S^{\perp}).$$

■

Lemma 7.2.5 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and S be a linear subspace of $\mathbb{R}^n$ such that $A = B + N_S$. Suppose $(x, x^*) \in S \times \mathbb{R}^n$. Then*

$$\iota_{\operatorname{ran} A_+}(x^* - Bx) = \iota_{\operatorname{ran} A_+}(x^* - Ax),$$

i.e., $x^* - Bx \in \operatorname{ran} A_+ \Leftrightarrow x^* - Ax \subset \operatorname{ran} A_+$.

Moreover if $x^* - Bx \in \operatorname{ran} A_+$, *then*

$$\big\langle x^* - Bx,\ (A_+)^{-1}(x^* - Bx)\big\rangle = \big\langle x^* - Ax,\ (A_+)^{-1}(x^* - Ax)\big\rangle.$$

Proof. By Fact 7.1.21,

$$Ax = Bx + S^{\perp}. \tag{7.6}$$

By Remark 7.1.25 and Proposition 7.2.3(i) applied to A_+,

$$\begin{aligned}\iota_{\operatorname{ran} A_+}(x^* - Bx) &= \iota_{\operatorname{ran} A_+}\big(x^* - Bx + S^{\perp}\big)\\ &= \iota_{\operatorname{ran} A_+}\big(x^* - Ax\big) \quad \text{(by (7.6))}.\end{aligned}$$

Let $x^* - Bx \in \operatorname{ran} A_+$. By Remark 7.1.25, $(A_+)^{-1}(x^* - Bx) \subset S$, then we have

$$\begin{aligned}&\big\langle x^* - Bx,\ (A_+)^{-1}(x^* - Bx)\big\rangle\\ &= \big\langle x^* - Bx + S^{\perp},\ (A_+)^{-1}(x^* - Bx)\big\rangle\\ &= \big\langle x^* - Bx + S^{\perp},\ (A_+)^{-1}\big(x^* - Bx + S^{\perp}\big)\big\rangle \qquad (7.7)\\ &= \big\langle x^* - Ax,\ (A_+)^{-1}\big(x^* - Ax\big)\big\rangle, \qquad (7.8)\end{aligned}$$

in which, (7.7) holds by Proposition 7.2.3(ii), (7.8) by (7.6). ■

Remark 7.2.6 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and* S *be a linear subspace of* $\mathbb{R}^n$ *such that* $A = B + N_S$. *Suppose* $(x, x^*) \in S \times \mathbb{R}^n$ *such that* $x^* - Bx \in \operatorname{ran} A_+$.

By Remark 7.1.25, Lemma 7.2.4 and Lemma 7.2.5, we see that

$$\big\langle x^* - Ax,\ (A_+)^{-1}(x^* - Ax)\big\rangle$$

is single-valued.

Proposition 7.2.7 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and* S *be a linear subspace of* $\mathbb{R}^n$ *such that* $A = B + N_S$. *Suppose* $(x, x^*) \in S \times \mathbb{R}^n$. *Then* $(x^* - Ax) \subset \operatorname{ran} A_+$ *or* $(x^* - Ax) \cap \operatorname{ran} A_+ = \varnothing$.

Proof. Suppose that $(x^* - Ax) \cap \operatorname{ran} A_+ \neq \varnothing$. By Fact 7.1.21, there exists $t \in S^{\perp}$ such that $x^* - Bx + t \in \operatorname{ran} A_+$. Then by Fact 7.1.21, Remark 7.1.25 and Proposition 7.2.3(i), we obtain $x^* - Ax = x^* - Bx + S^{\perp} = x^* - Bx + t + S^{\perp} \subset \operatorname{ran} A_+$. ■

Definition 7.2.8 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. *We define* $\Phi_A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *by*

$$\Phi_A(x) = \begin{cases} A^{-1}x, & \text{if } x \in \operatorname{ran} A; \\ \{0\}, & \text{otherwise.} \end{cases}$$

Remark 7.2.9 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and* S *be a linear subspace of* $\mathbb{R}^n$ *such that* $A = B + N_S$. *Then by Proposition 7.2.7 and Remark 7.2.6,*

$$\big\langle x^* - Ax,\ \Phi_{A_+}(x^* - Ax)\big\rangle \quad \big((x, x^*) \in S \times \mathbb{R}^n\big)$$

is single-valued. By Lemma 7.2.4 and Remark 7.1.25, $\big\langle x,\ \Phi_{A_+}(x)\big\rangle \quad (x \in \mathbb{R}^n)$ *is single-valued.*

Lemma 7.2.10 *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ such that* $\operatorname{gra} A$ *is a linear subspace of* $\mathbb{R}^n \times \mathbb{R}^n$. *Let $k \in \mathbb{R}$ with $k \neq 0$. Then $\Phi_A(kx) = k\Phi_A(x), \quad \forall x \in \mathbb{R}^n$.*

Proof. Let $x \in \mathbb{R}^n$. We consider two cases.

Case 1: $x \notin \operatorname{ran} A$. Then $kx \notin \operatorname{ran} A$. Thus we have $k\Phi_A(x) = \Phi_A(kx) = 0$.

Case 2: $x \in \operatorname{ran} A$. Then $kx \in \operatorname{ran} A$. Then by Remark 7.1.32 and Proposition 4.1.3(iii), $k\Phi_A(x) = kA^{-1}x = A^{-1}(kx) = \Phi_A(kx)$. ■

Corollary 7.2.11 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and S be a linear subspace of $\mathbb{R}^n$ such that $A = B + N_S$. Let $k \in \mathbb{R}$. Then*

$$\begin{aligned}
&\iota_S(x) + \iota_{\operatorname{ran} A_+}(x^* - Bx) + k\big\langle x^* - Bx,\ \Phi_{A_+}(x^* - Bx)\big\rangle \\
&= \iota_S(x) + \iota_{\operatorname{ran} A_+}(x^* - AP_S x) + k\big\langle x^* - AP_S x,\ \Phi_{A_+}(x^* - AP_S x)\big\rangle, \\
&\quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}$$

Proof. Combine Lemma 7.2.5 and Remark 7.1.25. ■

Fact 7.2.12 *Let $B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Then $B + N_S$ is maximal monotone.*

Proof. See [28, Theorem 41.2]. ■

Fact 7.2.13 *Let $B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be linear, symmetric and monotone, and S be a linear subspace of $\mathbb{R}^n$. Then* $\operatorname{ran}(B + N_S) = \operatorname{ran} B + S^{\perp}$.

Proof. Combine Remark 7.1.25, Fact 7.2.12, [4, Corollary 4.9] and [28, 19.0.3 page 70]. ■

Lemma 7.2.14 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Then*

$$(q_B + \iota_S)^*(x) = \iota_{\operatorname{ran}(B_+ + N_S)}(x) + \tfrac{1}{2}\big\langle x,\ \Phi_{(B_+ + N_S)}(x)\big\rangle, \quad \forall x \in \mathbb{R}^n.$$

Proof. Let $x \in \mathbb{R}^n$. Then

$$(q_B + \iota_S)^*(x) = \sup_{y \in \mathbb{R}^n} \Big\{\langle y, x\rangle - q_B(y) - \iota_S(y)\Big\}.$$

Let

$$g(y) := \langle y, x\rangle - q_B(y) - \iota_S(y).$$

A point y is a maximizer of g, if and only if, it is a critical point. Then by Fact 2.1.30, Fact 2.1.18 and Fact 7.1.21,

$$0 \in \partial g(y) = x - B_+ y - N_S(y) = x - (B_+ + N_S)(y).$$

We consider two cases.

Case 1: $x \in \operatorname{ran}(B_+ + N_S)$. Let y_0 satisfy that $x \in (B_+ + N_S)y_0$. Then $y_0 \in S$ and $x \in B_+ y_0 + S^{\perp}$ by Fact 7.1.21. Let $t \in S^{\perp}$ such that $x = B_+ y_0 + t$.

Since y_0 is a critical point,

$$\begin{aligned}(q_B+\iota_S)^*(x) = g(y_0) &= \langle y_0, x\rangle - \tfrac{1}{2}\langle y_0,\ B_+y_0\rangle \quad \text{(by Remark 2.1.12)}\\ &= \langle y_0, B_+y_0+t\rangle - \tfrac{1}{2}\langle y_0,\ B_+y_0\rangle \quad (\text{by } x = B_+y_0+t)\\ &= \langle y_0, B_+y_0\rangle - \tfrac{1}{2}\langle y_0,\ B_+y_0\rangle \qquad (\text{by } y_0\in S \text{ and } t\in S^{\perp})\\ &= \tfrac{1}{2}\langle y_0,\ B_+y_0\rangle\\ &= \tfrac{1}{2}\langle x,\ (B_++N_S)^{-1}x\rangle \qquad (\text{by Lemma 7.2.4 applied to } B_+)\\ &= \tfrac{1}{2}\big\langle x,\ \Phi_{(B_++N_S)}(x)\big\rangle.\end{aligned}$$

Case 2: $x \notin \operatorname{ran}(B_+ + N_S)$. By Fact 7.2.13, $\operatorname{ran}(B_+ + N_S) = \operatorname{ran} B_+ + S^{\perp}$. Thus by Fact 2.1.32,

$$\big(\operatorname{ran}(B_++N_S)\big)^{\perp} = (\operatorname{ran} B_+ + S^{\perp})^{\perp} = (\operatorname{ran} B_+)^{\perp} \cap (S^{\perp})^{\perp} = \ker B_+ \cap S.$$

Then we have $\mathbb{R}^n = \operatorname{ran}(B_+ + N_S) \oplus (\ker B_+ \cap S)$ and $x = P_{\operatorname{ran}(B_++N_S)}x + P_{\ker B_+\cap S}x$. Since $x \notin \operatorname{ran}(B_+ + N_S)$, $P_{\ker B_+\cap S}x \neq 0$. Thus

$$\begin{aligned}\langle P_{\ker B_+\cap S}x, x\rangle &= \langle P_{\ker B_+\cap S}x, P_{\operatorname{ran}(B_++N_S)}x + P_{\ker B_+\cap S}x\rangle\\ &= \|P_{\ker B_+\cap S}x\|^2 > 0. \end{aligned} \tag{7.9}$$

Then by Fact 5.1.10,

$$\begin{aligned}&(q_B+\iota_S)^*(x) = (q_B+\iota_S)^*(x) + (q_B+\iota_S)(kP_{\ker B_+\cap S}x) \qquad (7.10)\\ &\geq \langle kP_{\ker B_+\cap S}x, x\rangle \to \infty, \text{ as } k\to\infty \quad \big(\text{by } (7.9)\big),\end{aligned}$$

where (7.10) holds since $(q_B + \iota_S)(kP_{\ker B_+ \cap S}x) = 0$ by Remark 2.1.12 and $P_{\ker B_+ \cap S}x \in \ker B_+ \cap S$.

Combining the conclusions above, we have

$$(q_B + \iota_S)^*(x) = \iota_{\operatorname{ran}(B_+ + N_S)}(x) + \tfrac{1}{2}\big\langle x,\ \Phi_{(B_+ + N_S)}(x)\big\rangle, \quad \forall x \in \mathbb{R}^n.$$

■

Proposition 7.2.15 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Then*

$$\begin{aligned}
&F_{(B+N_S)}(x, x^*)\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + x^*) + \tfrac{1}{4}\big\langle B^*x + x^*,\ \Phi_{(B_+ + N_S)}(B^*x + x^*)\big\rangle,\\
&\quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Fact 7.1.21, we have

$$\begin{aligned}
&F_{(B+N_S)}(x, x^*)\\
&= \sup_{(y,y^*)\in \operatorname{gra}(B+N_S)} \Big\{\langle y^*, x\rangle + \langle x^*, y\rangle - \langle y, y^*\rangle\Big\}\\
&= \sup_{y\in S} \Big\{\langle By + S^{\perp}, x\rangle + \langle x^*, y\rangle - \langle y,\ By + S^{\perp}\rangle\Big\}\\
&= \iota_S(x) + \sup_{y\in S} \Big\{\langle By, x\rangle + \langle x^*, y\rangle - \langle y, By\rangle\Big\} && (7.11)\\
&= \iota_S(x) + \sup_{y\in \mathbb{R}^n} \Big\{\langle y,\ B^*x + x^*\rangle - \langle y, By\rangle - \iota_S(y)\Big\}\\
&= \iota_S(x) + 2\sup_{y\in \mathbb{R}^n} \Big\{\langle y, \tfrac{1}{2}(B^*x + x^*)\rangle - q_B(y) - \iota_S(y)\Big\} && (7.12)
\end{aligned}$$

where (7.11) holds by $y \in S$.

By (7.12), we have

$$
\begin{aligned}
F_{(B+N_S)}(x, x^*) &= \iota_S(x) + 2(q_B + \iota_S)^*\big(\tfrac{1}{2}(B^*x + x^*)\big) \\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + x^*) \qquad (7.13)\\
&\quad + \big\langle \tfrac{1}{2}(B^*x + x^*), \Phi_{(B_+ + N_S)}\big(\tfrac{1}{2}(B^*x + x^*)\big)\big\rangle \\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + x^*) \qquad (7.14)\\
&\quad + \tfrac{1}{4}\big\langle (B^*x + x^*), \Phi_{(B_+ + N_S)}(B^*x + x^*)\big\rangle.
\end{aligned}
$$

(7.13) holds by Lemma 7.2.14 and Remark 7.1.22. (7.14) holds by Remark 7.1.22 and Lemma 7.2.10. ■

Remark 7.2.16 *Let S be a linear subspace of $\mathbb{R}^n$. By Fact 7.2.13,* $\operatorname{ran} N_S = S^{\perp}$*. By Proposition 7.2.15,*

$$F_{N_S}(x, x^*) = \iota_S(x) + \iota_{S^{\perp}}(x^*), \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Corollary 7.2.17 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$
\begin{aligned}
&F_A(x, x^*) \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(A^*P_{\operatorname{dom} A}x + x^*) \\
&\quad + \tfrac{1}{4}\big\langle A^*P_{\operatorname{dom} A}x + x^*, \Phi_{A_+}(A^*P_{\operatorname{dom} A}x + x^*)\big\rangle, \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}
$$

Proof. By Corollary 7.1.31, there exists a linear and monotone operator $B : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$A = B + N_{\operatorname{dom} A}.$$

By Proposition 7.2.15 and Remark 7.1.25, we have

$$\begin{aligned}
&F_A(x, x^*) = F_{(B+N_{\operatorname{dom} A})}(x, x^*) \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(B^*x + x^*) + \tfrac{1}{4}\big\langle B^*x + x^*,\ \Phi_{A_+}(B^*x + x^*)\big\rangle \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(-B^*(-x) + x^*) \\
&\quad + \tfrac{1}{4}\big\langle -B^*(-x) + x^*,\ \Phi_{A_+}(-B^*(-x) + x^*)\big\rangle \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}\big(-A^*P_{\operatorname{dom} A}(-x) + x^*\big) \qquad (7.15) \\
&\quad + \tfrac{1}{4}\big\langle -A^*P_{\operatorname{dom} A}(-x) + x^*,\ \Phi_{A_+}\big(-A^*P_{\operatorname{dom} A}(-x) + x^*\big)\big\rangle \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}\big(-A^*(-P_{\operatorname{dom} A}x) + x^*\big) \qquad (7.16) \\
&\quad + \tfrac{1}{4}\big\langle -A^*(-P_{\operatorname{dom} A}x) + x^*,\ \Phi_{A_+}\big(-A^*(-P_{\operatorname{dom} A}x) + x^*\big)\big\rangle \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(A^*P_{\operatorname{dom} A}x + x^*) \qquad (7.17) \\
&\quad + \tfrac{1}{4}\big\langle A^*P_{\operatorname{dom} A}x + x^*,\ \Phi_{A_+}\big(A^*P_{\operatorname{dom} A}x + x^*\big)\big\rangle,
\end{aligned}$$

where (7.15) holds by Fact 7.1.23 and Corollary 7.2.11 applied to B^* and A^*. (7.16) holds by Fact 4.3.3, and (7.17) by Remark 7.1.5 and Proposition 4.1.3(iii). ■

Proposition 7.2.18 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a* *linear subspace of* $\mathbb{R}^n$. *Then*

$$F^*_{(B+N_S)}(x^*,x) = \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \langle x, Bx\rangle,$$
$$\forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Proposition 7.2.15,

$$\begin{aligned}
&F^*_{(B+N_S)}(x^*,x)\\
&= \sup_{(y,y^*)} \Big\{ \langle y, x^*\rangle + \langle y^*, x\rangle - \iota_S(y) - \iota_{\operatorname{ran}(B_+ + N_S)}(B^*y + y^*)\\
&\quad - \tfrac{1}{4}\big\langle B^*y + y^*,\ \Phi_{(B_+ + N_S)}(B^*y + y^*)\big\rangle \Big\}\\
&= \sup_{(y\in S, w\in S)} \Big\{ \langle y, x^*\rangle + \langle B_+w - B^*y + S^\perp,\ x\rangle - \tfrac{1}{4}\langle B_+w, w\rangle \Big\} && (7.18)\\
&= \iota_S(x) + \sup_{(y\in S, w\in S)} \Big\{ \langle y, x^*\rangle + \langle B_+w - B^*y,\ x\rangle - \tfrac{1}{4}\langle B_+w, w\rangle \Big\}\\
&= \iota_S(x) + \sup_{(y\in S,\ w\in S)} \Big\{ \langle y, x^* - Bx\rangle + \langle B_+w, x\rangle - \tfrac{1}{4}\langle w, B_+w\rangle \Big\}\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \sup_{w\in S} \Big\{ \langle B_+w,\ x\rangle - \tfrac{1}{4}\langle w, B_+w\rangle \Big\}\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \tfrac{1}{2}\sup_{w\in S} \Big\{ \langle w,\ 2B_+x\rangle - q_B(w) \Big\} && (7.19)\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx)\\
&\quad + \tfrac{1}{2}\sup_{w\in \mathbb{R}^n} \Big\{ \langle w,\ 2B_+x\rangle - q_B(w) - \iota_S(w) \Big\}\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \tfrac{1}{2}(q_B + \iota_S)^*(2B_+x)\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \iota_{\operatorname{ran}(B_+ + N_S)}(2B_+x) && (7.20)\\
&\quad + \tfrac{1}{4}\langle 2B_+x, \Phi_{(B_+ + N_S)}(2B_+x)\rangle\\
&= \iota_S(x) + \iota_{S^\perp}(x^* - Bx) + \langle x, Bx\rangle, && (7.21)
\end{aligned}$$

in which, (7.18) holds by $B^*y + y^* \in (B_+ + N_S)w = B_+w + S^\perp(w \in S)$ by Fact 7.1.21, and by Lemma 7.2.4. (7.19) holds by Remark 2.1.12. (7.20) holds by Lemma 7.2.14, and (7.21) by Lemma 7.2.4 since $2x \in (B_+ + N_S)^{-1}(2B_+x)$ as $x \in S$ by Fact 7.1.21. ■

Remark 7.2.19 *Let S be a linear subspace of $\mathbb{R}^n$. By Fact 7.2.13,* $\operatorname{ran} N_S = S^\perp$. *Then by Proposition 7.2.18,*

$$F^*_{N_S}(x^*, x) = \iota_S(x) + \iota_{S^\perp}(x^*), \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Corollary 7.2.20 *Let S be a linear subspace of $\mathbb{R}^n$. Then F_{N_S} is autoconjugate.*

Proof. Combine Remark 7.2.16 and Remark 7.2.19. ■

Remark 7.2.21 *Remark 7.2.16 and Remark 7.2.19 are special cases of [8, Example 3.1].*

Remark 7.2.22 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$ such that $A = B + N_S$. Suppose $x \in S$. Then $\langle Ax, x\rangle = \langle x, Bx\rangle$.*

Proof. Apply $Ax = Bx + S^\perp$, which follows from Fact 7.1.21. ■

Corollary 7.2.23 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$F^*_A(x^*, x) = \iota_{\operatorname{dom} A}(x) + \iota_{(\operatorname{dom} A)^\perp}(x^* - AP_{\operatorname{dom} A}x) + \langle x, AP_{\operatorname{dom} A}x\rangle,$$
$$\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Corollary 7.1.31, there exists a linear and monotone operator $B : \mathbb{R}^n \to \mathbb{R}^n$ such that $A = B + N_{\operatorname{dom} A}$. Then by Proposition 7.2.18,

$$F_A^*(x^*, x) = \iota_{\operatorname{dom} A}(x) + \iota_{(\operatorname{dom} A)^\perp}(x^* - Bx) + \langle x, Bx\rangle. \tag{7.22}$$

Suppose $x \in \operatorname{dom} A$. Since $\operatorname{dom} A$ is a subspace of $\mathbb{R}^n$ and

$$x^* - Bx \in (\operatorname{dom} A)^\perp \Leftrightarrow x^* - Bx + (\operatorname{dom} A)^\perp \subset (\operatorname{dom} A)^\perp.$$

By Fact 7.1.21, $x^* - Bx + (\operatorname{dom} A)^\perp = x^* - Ax$. Thus

$$\iota_{(\operatorname{dom} A)^\perp}(x^* - Bx) = \iota_{(\operatorname{dom} A)^\perp}(x^* - Ax) = \iota_{(\operatorname{dom} A)^\perp}(x^* - AP_{\operatorname{dom} A}x). \tag{7.23}$$

By Remark 7.2.22,

$$\langle x, Bx\rangle = \langle Ax, x\rangle = \langle AP_{\operatorname{dom} A}x, x\rangle. \tag{7.24}$$

Thus by (7.22), (7.23) and (7.24),

$$F_A^*(x^*, x) = \iota_{\operatorname{dom} A}(x) + \iota_{(\operatorname{dom} A)^\perp}(x^* - AP_{\operatorname{dom} A}x) + \langle AP_{\operatorname{dom} A}x, x\rangle,$$
$$\forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.$$

■

7.3 The third main result

Lemma 7.3.1 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Suppose that $x \in S, x^* \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^n$. Then*

$$\begin{aligned}
&\iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + y^*) + \iota_{S^\perp}(2x^* - y^* - Bx)\\
&= \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \iota_{S^\perp}(2x^* - y^* - Bx).
\end{aligned} \tag{7.25}$$

Proof. We consider two cases.

Case 1: $2x^* - y^* - Bx \notin S^\perp$. Clear.

Case 2: $2x^* - y^* - Bx \in S^\perp$. Let $t \in S^\perp$ such that $y^* = 2x^* - Bx + t$. Thus

$$\begin{aligned}
&B^*x + y^*\\
&= B^*x + 2x^* - Bx + t = Bx + B^*x + 2x^* - Bx - Bx + t\\
&= 2x^* - 2Bx + 2B_+x + t.
\end{aligned} \tag{7.26}$$

On the other hand, since $t \in S^\perp$, Fact 7.1.21 implies

$$2B_+x + t \in (B_+ + N_S)(2x). \tag{7.27}$$

Then by Remark 7.1.22, (7.26) and (7.27), we have

$$B^*x + y^* \in \operatorname{ran}(B_+ + N_S) \Leftrightarrow x^* - Bx \in \operatorname{ran}(B_+ + N_S). \tag{7.28}$$

Thus $\iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + y^*) = \iota_{\operatorname{ran}(B_+ + N_s)}(x^* - Bx)$.

Hence (7.25) holds. ■

Corollary 7.3.2 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Suppose that $x, x^*, y^* \in \mathbb{R}^n$. Then*

$$\begin{aligned}
&\iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + y^*) + \iota_{S^\perp}(2x^* - y^* - Bx)\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \iota_{S^\perp}(2x^* - y^* - Bx).
\end{aligned}$$

Proof. Apply Lemma 7.3.1. ■

Proposition 7.3.3 *Let $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$\begin{aligned}
&h_{F_A}(x, x^*)\\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(x^* - AP_{\operatorname{dom} A}x) + \langle x, x^*\rangle\\
&\quad + \tfrac{1}{2}\big\langle x^* - AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^* - AP_{\operatorname{dom} A}x)\big\rangle, \quad \forall (x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}$$

Proof. By Corollary 7.1.31, there exists a linear and monotone operator $B : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$A = B + N_S,$$

where $S = \operatorname{dom} A$. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$.

By Proposition 7.2.15 and Proposition 7.2.18,

$$
\begin{aligned}
&h_{F_A}(x, x^*)\\
&= \inf_{y^*} \Big\{ \tfrac{1}{2} F_A(x, 2y^*) + \tfrac{1}{2} F_A^*\big(2(x^* - y^*), x\big) \Big\}\\
&= \inf_{y^*} \Big\{ \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + 2y^*)\\
&\quad + \tfrac{1}{8} \big\langle B^*x + 2y^*,\ \Phi_{(B_+ + N_S)}(B^*x + 2y^*) \big\rangle\\
&\quad + \iota_{S^\perp}(2x^* - 2y^* - Bx) + \tfrac{1}{2}\langle x, Bx\rangle \Big\}\\
&= \iota_S(x) + \tfrac{1}{2}\langle x, Bx\rangle + \inf_{y^*} \Big\{ \iota_{\operatorname{ran}(B_+ + N_S)}(B^*x + 2y^*)\\
&\quad + \tfrac{1}{8} \big\langle B^*x + 2y^*,\ \Phi_{(B_+ + N_S)}(B^*x + 2y^*) \big\rangle\\
&\quad + \iota_{S^\perp}(2x^* - 2y^* - Bx) \Big\}\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \tfrac{1}{2}\langle x, Bx\rangle \qquad (7.29)\\
&\quad + \inf_{y^*} \Big\{ \tfrac{1}{8} \big\langle B^*x + 2y^*,\ \Phi_{(B_+ + N_S)}(B^*x + 2y^*) \big\rangle\\
&\quad + \iota_{S^\perp}(2x^* - 2y^* - Bx) \Big\}\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \tfrac{1}{2}\langle x, Bx\rangle \qquad (7.30)\\
&\quad + \inf_{t \in S^\perp} \Big\{ \tfrac{1}{8} \big\langle B^*x + 2x^* - Bx + t,\ \Phi_{(B_+ + N_S)}(B^*x + 2x^* - Bx + t) \big\rangle \Big\},
\end{aligned}
$$

in which, (7.29) holds by Corollary 7.3.2, (7.30) by $2y^* = 2x^* - Bx + t,\ t \in S^\perp$.

If $x \notin S$ or $x^* - Bx \notin \operatorname{ran}(B_+ + N_S)$, $h_{F_A}(x, x^*) = \infty$ by (7.30).

Now suppose $x \in S = \operatorname{dom} A$ and $x^* - Bx \in \operatorname{ran}(B_+ + N_S)$. Then there exists $y_0 \in S$ such that $x^* - Bx \in (B_+ + N_S)y_0$. Thus by Fact 7.1.21,

$x^* - Bx \in B_+y_0 + S^\perp$. Then

$$\langle B_+x, y_0\rangle = \langle x, B_+y_0\rangle = \langle x, x^* - Bx\rangle = \langle x, x^*\rangle - \langle x, Bx\rangle. \tag{7.31}$$

Note that

$$\begin{aligned} B^*x + 2x^* - Bx + t &= Bx + B^*x + 2x^* - Bx - Bx + t \\ &= 2x^* - 2Bx + 2B_+x + t. \end{aligned} \tag{7.32}$$

By Fact 7.1.21,

$$2B_+x + t \in (B_+ + N_S)(2x). \tag{7.33}$$

Then by Remark 7.1.22, (7.32) and (7.33),

$$B^*x + 2x^* - Bx + t \in (B_+ + N_S)(2y_0 + 2x). \tag{7.34}$$

Then by (7.34), (7.31), (7.30) and Lemma 7.2.4,

$$
\begin{aligned}
& h_{F_A}(x,x^*) \\
&= \tfrac{1}{2}\langle x, Bx\rangle + \tfrac{1}{8}\big\langle B_+(2y_0+2x),\ 2y_0+2x\big\rangle \\
&= \tfrac{1}{2}\langle x, Bx\rangle + \tfrac{1}{2}\big\langle B_+(y_0+x), y_0+x\big\rangle \\
&= \tfrac{1}{2}\langle x, Bx\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle B_+x, y_0\rangle + \tfrac{1}{2}\langle B_+x, x\rangle \\
&= \tfrac{1}{2}\langle x, Bx\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle x, x^*\rangle - \langle x, Bx\rangle + \tfrac{1}{2}\langle B_+x, x\rangle \\
&= \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle x, x^*\rangle \quad \text{(by Remark 2.1.12)} \\
&= \langle x, x^*\rangle + \tfrac{1}{2}\langle x^*-Bx, (B_+ + N_S)^{-1}(x^*-Bx)\rangle \\
&= \langle x, x^*\rangle + \tfrac{1}{2}\langle x^*-Bx, (A_+)^{-1}(x^*-Bx)\rangle \quad \text{(by Remark 7.1.25)} \\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^*-Ax,\ (A_+)^{-1}(x^*-Ax)\big\rangle \quad \text{(by Lemma 7.2.5)} \\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^*-Ax,\ \Phi_{A_+}(x^*-Ax)\big\rangle \quad \text{(by Lemma 7.2.5)} \\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^*-AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^*-AP_{\operatorname{dom} A}x)\big\rangle. \qquad (7.35)
\end{aligned}
$$

Thus combining (7.30) and (7.35),

$$
\begin{aligned}
& h_{F_A}(x,x^*) \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(x^*-Bx) + \langle x, x^*\rangle \quad \text{(by Remark 7.1.25)} \\
&\quad + \tfrac{1}{2}\big\langle x^*-AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^*-AP_{\operatorname{dom} A}x)\big\rangle \\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(x^*-AP_{\operatorname{dom} A}x) + \langle x, x^*\rangle \quad \text{(by Lemma 7.2.5)} \\
&\quad + \tfrac{1}{2}\big\langle x^*-AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^*-AP_{\operatorname{dom} A}x)\big\rangle, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}
$$

■

Remark 7.3.4 *Let S be a linear subspace of $\mathbb{R}^n$ and $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be invertible. Then* $\dim AS = \dim S$.

Proposition 7.3.5 *Let S be a linear subspace of $\mathbb{R}^n$ and $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Suppose that $AS \subset S$. Then* $(\operatorname{Id} + A)S = S$.

Proof. By assumptions, $(\operatorname{Id} + A)S$ is a linear subspace and

$$(\operatorname{Id} + A)S \subset S + AS \subset S. \tag{7.36}$$

Since $(\operatorname{Id} + A)$ is invertible by Proposition 5.2.6, by Remark 7.3.4, $\dim(\operatorname{Id} + A)S = \dim S$. Then by (7.36), $(\operatorname{Id} + A)S = S$. ■

Corollary 7.3.6 *Let S be a linear subspace of $\mathbb{R}^n$ and $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone. Suppose that* $\operatorname{ran} A \subset S$. *Then* $(\operatorname{Id} + A)^{-1} A_+ S \subset S$.

Proof. By Fact 2.1.17, $\operatorname{ran} A^* = \operatorname{ran} A \subset S$. Thus $A_+ S \subset \operatorname{ran} A_+ \subset S$. By Proposition 7.3.5, $A_+ S \subset S = (\operatorname{Id} + A)S$. Then $(\operatorname{Id} + A)^{-1} A_+ S \subset (\operatorname{Id} + A)^{-1}(\operatorname{Id} + A)S = S$. ■

Lemma 7.3.7 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Suppose that $x, y \in S$ and $x^*, y^* \in \mathbb{R}^n$. Then*

$$\begin{aligned}
&\iota_{\operatorname{ran}(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big) + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big) \\
&\quad = \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big).
\end{aligned} \tag{7.37}$$

Proof. We consider two cases.

Case 1: $x^* - y^* - B(x-y) \notin S^\perp$. Clear.

Case 2: $x^* - y^* - B(x-y) \in S^\perp$. Let $t \in S^\perp$ such that $y^* = x^* - B(x-y) + t$. Thus

$$\begin{aligned}
&B^*(x+y) + x^* + y^* \\
&= B^*(x+y) + 2x^* - B(x-y) + t \\
&= B^*(x+y) + B(x+y) - B(x+y) + 2x^* - B(x-y) + t \\
&= 2B_+(x+y) + t + 2x^* - 2Bx. \qquad (7.38)
\end{aligned}$$

On the other hand, since $t \in S^\perp$, Fact 7.1.21 implies

$$2B_+(x+y) + t \in (B_+ + N_S)(2x + 2y). \qquad (7.39)$$

Then by Remark 7.1.22, (7.39) and (7.38), we have

$$B^*(x+y) + x^* + y^* \in \operatorname{ran}(B_+ + N_S) \Leftrightarrow x^* - Bx \in \operatorname{ran}(B_+ + N_S).$$

Thus $\iota_{\operatorname{ran}(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big) = \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx)$. Hence (7.37) holds. ■

Corollary 7.3.8 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Suppose that $x \in \mathbb{R}^n, y \in S$ and $x^*, y^* \in \mathbb{R}^n$. Then*

$$\begin{aligned}
&\iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big) + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big) \\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big).
\end{aligned}$$

Proof. Apply Lemma 7.3.7. ■

Theorem 7.3.9 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$P(F_A, F_A^{*\intercal}) = h_{F_A}.$$

Proof. By Corollary 7.1.31,

$$A = B + N_S,$$

where $B = P_S Q_A P_S, S = \operatorname{dom} A$. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$.

By Fact 5.2.2, Fact 5.2.9, Proposition 7.2.15 and Proposition 7.2.18,

$$\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*) \\
&= \inf_{(y,y^*)} \Big\{ \tfrac{1}{2} F_{(B+N_S)}(x+y, x^*+y^*) + \tfrac{1}{2} F^{*\intercal}_{(B+N_S)}(x-y, x^*-y^*) \\
&\quad + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|y^*\|^2 \Big\} \\
&= \inf_{(y,y^*)} \Big\{ \iota_S(x+y) + \iota_{\operatorname{ran}(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big) + \iota_S(x-y) \\
&\quad + \tfrac{1}{8}\big\langle B^*(x+y) + x^* + y^*,\ \Phi_{(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big)\big\rangle \\
&\quad + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big) + \tfrac{1}{2}\big\langle x-y, B(x-y)\big\rangle + \tfrac{1}{2}\|y\|^2 \\
&\quad + \tfrac{1}{2}\|y^*\|^2 \Big\} \\
&= \iota_S(x) + \inf_{y\in S,\ y^*\in\mathbb{R}^n} \Big\{ \iota_{\operatorname{ran}(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big) \qquad (7.40) \\
&\quad + \tfrac{1}{8}\big\langle B^*(x+y) + x^* + y^*,\ \Phi_{(B_+ + N_S)}\big(B^*(x+y) + x^* + y^*\big)\big\rangle \\
&\quad + \iota_{S^\perp}\big(x^* - y^* - B(x-y)\big) + \tfrac{1}{2}\big\langle x-y, B(x-y)\big\rangle + \tfrac{1}{2}\|y\|^2 \\
&\quad + \tfrac{1}{2}\|y^*\|^2 \Big\},
\end{aligned}$$

where (7.40) holds by $\iota_S(x+y)=0, \iota_S(x-y)=0 \Leftrightarrow x,y\in S$.

By (7.40) and Corollary 7.3.8,

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x,x^*)\\
&= \iota_S(x) + \iota_{\mathrm{ran}(B_+ + N_S)}(x^* - Bx)\\
&\quad + \inf_{y\in S,\ y^*\in\mathbb{R}^n} \Big\{\\
&\quad \tfrac{1}{8}\big\langle B^*(x+y)+x^*+y^*,\ \Phi_{(B_+ + N_S)}\big(B^*(x+y)+x^*+y^*\big)\big\rangle\\
&\quad + \iota_{S^\perp}\big(x^*-y^*-B(x-y)\big) + \tfrac{1}{2}\big\langle x-y, B(x-y)\big\rangle + \tfrac{1}{2}\|y\|^2\\
&\quad + \tfrac{1}{2}\|y^*\|^2\Big\}\\
&= \iota_S(x) + \iota_{\mathrm{ran}(B_+ + N_S)}(x^* - Bx) + \inf_{y\in S,\ t\in S^\perp}\Big\{\tfrac{1}{8}\big\langle C_0,\ \Phi_{(B_+ + N_S)}(C_0)\big\rangle \qquad (7.41)\\
&\quad + \tfrac{1}{2}\big\langle x-y, B(x-y)\big\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|x^*-B(x-y)+t\|^2\Big\},
\end{aligned}
$$

where (7.41) holds by $y^* = x^* - B(x-y) + t,\ t\in S^\perp$, where $C_0 := B^*(x+y) + 2x^* - B(x-y) + t \quad (B^*(x+y)+x^*+y^* = C_0)$.

Suppose that $x\in S$ and $x^*-Bx\in \mathrm{ran}(B_+ + N_S)$. Then exists $y_0\in S$ such that $x^*-Bx\in (B_+ + N_S)y_0$. Thus by Fact 7.1.21, $x^*-Bx\in(B_+ + N_S)y_0 = B_+y_0 + S^\perp$. Thus

$$\langle x, Bx + B_+y_0\rangle = \langle x, x^*\rangle \quad (\text{by } x\in S). \tag{7.42}$$

On one hand,

$$
\begin{aligned}
C_0 &= B^*(x+y) + B(x+y) - B(x+y) + 2x^* - B(x-y) + t \\
&= 2B_+(x+y) + t + 2x^* - 2Bx. \qquad (7.43)
\end{aligned}
$$

On the other hand, since $t \in S^\perp$, Fact 7.1.21 implies

$$
2B_+(x+y) + t \in (B_+ + N_S)(2x+2y). \qquad (7.44)
$$

Then by Remark 7.1.22, (7.44) and (7.43)

$$
C_0 \in (B_+ + N_S)(2x + 2y + 2y_0). \qquad (7.45)
$$

Then by (7.45), (7.41) and Lemma 7.2.4,

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*) \\
&= \inf_{y \in S,\ t \in S^\perp} \Big\{ \tfrac{1}{8} \big\langle 2B_+(x+y+y_0),\ 2x + 2y + 2y_0 \big\rangle \\
&\quad + \tfrac{1}{2} \big\langle x - y, B(x-y) \big\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|x^* - B(x-y) + t\|^2 \Big\} \\
&= \inf_{y \in S,\ t' \in S^\perp} \Big\{ \tfrac{1}{2} \big\langle B_+(x+y+y_0),\ x + y + y_0 \big\rangle \\
&\quad + \tfrac{1}{2} \big\langle x - y, B(x-y) \big\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|B_+ y_0 + By + t'\|^2 \Big\} \\
&\leq \inf_{y \in S} \Big\{ \tfrac{1}{2} \langle B_+(x+y+y_0),\ x + y + y_0 \rangle \\
&\quad + \tfrac{1}{2} \langle x - y, B(x-y) \rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|B_+ y_0 + By\|^2 \Big\}.
\end{aligned}
$$

Note that (by Remark 2.1.12)

$$
\begin{aligned}
&\tfrac{1}{2}\big\langle B_+(x+y+y_0),\ x+y+y_0\big\rangle + \tfrac{1}{2}\big\langle x-y, B(x-y)\big\rangle\\
&= \tfrac{1}{2}\big\langle B_+(x+y_0) + B_+y,\ (x+y_0)+y\big\rangle + \tfrac{1}{2}\big\langle x-y, B_+(x-y)\big\rangle\\
&= \tfrac{1}{2}\big\langle B_+(x+y_0),\ x+y_0\big\rangle + \langle y, B_+(x+y_0)\rangle + \tfrac{1}{2}\langle y, B_+y\rangle + \tfrac{1}{2}\langle x, B_+x\rangle\\
&\quad + \tfrac{1}{2}\langle y, B_+y\rangle - \langle y, B_+x\rangle\\
&= \tfrac{1}{2}\big\langle B_+(x+y_0),\ x+y_0\big\rangle + \langle y, B_+y_0\rangle + \langle y, B_+y\rangle + \tfrac{1}{2}\langle x, B_+x\rangle\\
&= \tfrac{1}{2}\langle B_+x,\ x\rangle + \langle B_+y_0, x\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle y, B_+y_0\rangle + \langle y, B_+y\rangle + \tfrac{1}{2}\langle x, B_+x\rangle\\
&= \langle x, Bx\rangle + \langle B_+y_0, x\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle y, B_+y_0\rangle + \langle y, B_+y\rangle\\
&= \big\langle x, Bx + B_+y_0\big\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle y, B_+y_0\rangle + \langle y, B_+y\rangle\\
&= \langle x, x^*\rangle + \tfrac{1}{2}\langle B_+y_0, y_0\rangle + \langle y, B_+y_0\rangle + \langle y, B_+y\rangle \quad \text{(by (7.42))}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x,x^*)\\
&\le \tfrac{1}{2}\langle y_0, B_+y_0\rangle + \langle x, x^*\rangle \qquad (7.46)\\
&\quad + \inf_{y\in S}\Big\{\langle y, B_+y_0\rangle + \langle y, B_+y\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|B_+y_0 + By\|^2\Big\}
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\langle y, B_+y_0\rangle + \langle y, B_+y\rangle + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|B_+y_0 + By\|^2\\
&= \langle y, B_+y_0\rangle + \langle y, B_+y\rangle + \tfrac{1}{2}\|y\|^2 + \langle y, B^*B_+y_0\rangle + \tfrac{1}{2}\langle y, B^*By\rangle + \tfrac{1}{2}\|B_+y_0\|^2\\
&= \big\langle y,\ (\mathrm{Id} + B^*)B_+y_0\big\rangle + \tfrac{1}{2}\big\langle y,\ \big(B + B^* + \mathrm{Id} + B^*B\big)y\big\rangle + \tfrac{1}{2}\|B_+y_0\|^2\\
&= \big\langle y,\ (\mathrm{Id} + B^*)B_+y_0\big\rangle + \tfrac{1}{2}\big\langle y,\ (\mathrm{Id} + B^*)(\mathrm{Id} + B)y\big\rangle + \tfrac{1}{2}\|B_+y_0\|^2.
\end{aligned}
$$

Then by (7.46), we have

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x, x^*)\\
&\leq \tfrac{1}{2}\langle y_0, B_+y_0\rangle + \tfrac{1}{2}\|B_+y_0\|^2 + \langle x, x^*\rangle \qquad (7.47)\\
&\quad + \inf_{y\in S}\Big\{\big\langle y,\ (\mathrm{Id} + B^*)B_+y_0\big\rangle + \tfrac{1}{2}\big\langle y, (\mathrm{Id} + B^*)(\mathrm{Id} + B)y\big\rangle\Big\}\\
&\leq \tfrac{1}{2}\langle y_0, B_+y_0\rangle + \langle x, x^*\rangle + \tfrac{1}{2}\|B_+y_0\|^2 - \tfrac{1}{2}\|B_+y_0\|^2 \qquad (7.48)\\
&= \tfrac{1}{2}\langle y_0, B_+y_0\rangle + \langle x, x^*\rangle\\
&= \langle x, x^*\rangle + \tfrac{1}{2}\langle x^* - Bx, (B_+ + N_S)^{-1}(x^* - Bx)\rangle \qquad (7.49)\\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^* - Ax,\ (A_+)^{-1}(x^* - Ax)\big\rangle \qquad (7.50)\\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^* - Ax,\ \Phi_{A_+}(x^* - Ax)\big\rangle \quad \text{(by Lemma 7.2.5)}\\
&= \langle x, x^*\rangle + \tfrac{1}{2}\big\langle x^* - AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^* - AP_{\operatorname{dom} A}x)\big\rangle, \qquad (7.51)
\end{aligned}
$$

in which (7.48) holds by letting $y = -(\mathrm{Id} + B)^{-1}B_+y_0 \in S$, where $y \in S$ by Corollary 7.3.6.

(7.49) holds Lemma 7.2.4, (7.50) by Remark 7.1.25 and Lemma 7.2.5.

Combining (7.41) and (7.51),

$$
\begin{aligned}
&P(F_A, F_A^{*\intercal})(x,x^*)\\
&\leq \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \langle x, x^*\rangle\\
&\quad + \tfrac{1}{2}\big\langle x^* - AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^* - AP_{\operatorname{dom} A}x)\big\rangle\\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(x^* - AP_{\operatorname{dom} A}x) + \langle x, x^*\rangle \text{ (by Remark 7.1.25, Lemma 7.2.5)}\\
&\quad + \tfrac{1}{2}\big\langle x^* - AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^* - AP_{\operatorname{dom} A}x)\big\rangle\\
&= h_{F_A}(x,x^*) \quad \text{(by Proposition 7.3.3)}, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}
$$

By Fact 5.2.11, Fact 5.2.12 and Proposition 5.1.8, $P(F_A, F_A^{*\intercal}) = h_{F_A}$. ■

Corollary 7.3.10 *Let* $A\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be maximal monotone such that* $\operatorname{gra} A$ *is a linear subspace. Then*

$$P(F_A, F_A^{*\intercal}) = h_{F_A} = h_{F_A^{*\intercal}}.$$

Proof. Combine Theorem 7.3.9 and Proposition 5.3.13. ■

Theorem 7.3.11 *Let* $B\colon \mathbb{R}^n \to \mathbb{R}^n$ *be linear and monotone, and* S *be a linear subspace of* $\mathbb{R}^n$. *Then*

$$
\begin{aligned}
&P(F_{(B+N_S)}, F_{(B+N_S)}^{*\intercal})(x,x^*) = h_{F_{(B+N_S)}}(x,x^*) = h_{F_{(B+N_S)}^{*\intercal}}(x,x^*)\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \langle x, x^*\rangle\\
&\quad + \tfrac{1}{2}\big\langle x^* - Bx,\ \Phi_{(B_+ + N_S)}(x^* - Bx)\big\rangle, \quad \forall (x,x^*) \in \mathbb{R}^n \times \mathbb{R}^n.
\end{aligned}
$$

Proof. Let $A = B + N_S$. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Fact 7.2.12 and Remark 7.1.22, A is maximal monotone with a linear graph. Thus by Proposition 7.3.3, Corollary 7.3.10, Remark 7.1.25 and Corollary 7.2.11,

$$\begin{aligned}
&P(F_{(B+N_S)}, F^{*\intercal}_{(B+N_S)})(x, x^*) = h_{F_{(B+N_S)}}(x, x^*) = h_{F^{*\intercal}_{(B+N_S)}}(x, x^*)\\
&= \iota_{\operatorname{dom} A}(x) + \iota_{\operatorname{ran} A_+}(x^* - AP_{\operatorname{dom} A}x) + \langle x, x^*\rangle\\
&\quad + \tfrac{1}{2}\big\langle x^* - AP_{\operatorname{dom} A}x,\ \Phi_{A_+}(x^* - AP_{\operatorname{dom} A}x)\big\rangle\\
&= \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* - Bx) + \langle x, x^*\rangle\\
&\quad + \tfrac{1}{2}\big\langle x^* - Bx,\ \Phi_{(B_+ + N_S)}(x^* - Bx)\big\rangle.
\end{aligned}$$

■

7.4 The Fitzpatrick function of the sum

Fact 7.4.1 *Let $A, B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be monotone. Then*

$$F_{A+B} \leq F_A \Box_2 F_B. \tag{7.52}$$

Proof. See [8, Proposition 4.2]. ■

In (7.52), equality doesn't always hold, see [8, Example 4.7]. It would be interesting to characterize the pairs of monotone operators (A, B) that satisfy the identity $F_{A+B} = F_A \Box_2 F_B$.

Lemma 7.4.2 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S be a linear subspace of $\mathbb{R}^n$. Then $F_{(B+N_S)} = F_B \Box_2 F_{N_S}$.*

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. By Fact 5.2.4 and Remark 7.2.16, we have

$$
\begin{aligned}
&(F_B \Box_2 F_{N_S})(x, x^*) \\
&= \inf_{y^*} \Big\{ F_B(x, y^*) + F_{N_S}(x, x^* - y^*) \Big\} \\
&= \inf_{y^*} \Big\{ \iota_{\operatorname{ran} B_+}(y^* + B^*x) \\
&\quad + \tfrac{1}{2} q_{(B_+)^\dagger}(y^* + B^*x) + \iota_S(x) + \iota_{S^\perp}(x^* - y^*) \Big\} \\
&= \iota_S(x) \\
&\quad + \inf_{y^*} \Big\{ \iota_{\operatorname{ran} B_+}(y^* + B^*x) + \tfrac{1}{2} q_{(B_+)^\dagger}(y^* + B^*x) + \iota_{S^\perp}(x^* - y^*) \Big\} \\
&\leq \iota_S(x) + \iota_{\operatorname{ran}(B_+ + N_S)}(x^* + B^*x) \qquad (7.53) \\
&\quad + \inf_{y^*} \Big\{ \iota_{\operatorname{ran} B_+}(y^* + B^*x) + \tfrac{1}{2} q_{(B_+)^\dagger}(y^* + B^*x) + \iota_{S^\perp}(x^* - y^*) \Big\}.
\end{aligned}
$$

Next we will show that $(F_A \Box_2 F_{N_S})(x, x^*) \leq F_{(B+N_S)}(x, x^*)$.

Now suppose $x \in S$ and $x^* + B^*x \in \operatorname{ran}(B_+ + N_S)$. Then there exists $y_0 \in S$ such that

$$x^* + B^*x \in (B_+ + N_S)y_0. \tag{7.54}$$

By Fact 7.1.21, there exists $t \in S^\perp$ such that $x^* + B^*x = B_+ y_0 + t$.

Let $y_0^* = x^* - t$. Then by $x^* + B^*x = B_+ y_0 + t$,

$$y_0^* + B^*x = x^* + B^*x - t = B_+ y_0. \tag{7.55}$$

By (7.53), (7.55) and Lemma 7.2.4,

$$\begin{aligned}
&(F_B\Box_2 F_{N_S})(x,x^*)\\
&\le \iota_{\operatorname{ran} B_+}(y_0^*+B^*x)+\tfrac{1}{2}q_{(B_+)^\dagger}(y_0^*+B^*x)+\iota_{S^\perp}(x^*-y_0^*)\\
&=\tfrac{1}{2}q_{(B_+)^\dagger}(B_+y_0)\\
&=\tfrac{1}{2}q_{B_+}(y_0)\quad\text{(by Corollary 2.2.16)}\\
&=\tfrac{1}{4}\big\langle B^*x+x^*,\ (B_++N_S)^{-1}(B^*x+x^*)\big\rangle\quad\text{(by (7.54))}\\
&=\tfrac{1}{4}\big\langle B^*x+x^*,\ \Phi_{(B_++N_S)}(B^*x+x^*)\big\rangle. \qquad (7.56)
\end{aligned}$$

Thus combining (7.53) and (7.56),

$$\begin{aligned}
&(F_A\Box_2 F_{N_S})(x,x^*)\\
&\le \iota_S(x)+\iota_{\operatorname{ran}(B_++N_S)}(x^*+B^*x)+\tfrac{1}{4}\big\langle B^*x+x^*,\ \Phi_{(B_++N_S)}(B^*x+x^*)\big\rangle\\
&=F_{(B+N_S)}(x,x^*)\quad\text{(by Proposition 7.2.15)},\quad \forall (x,x^*)\in\mathbb{R}^n\times\mathbb{R}^n.
\end{aligned}$$

By Fact 7.4.1, $F_{(B+N_S)}=(F_B\Box_2 F_{N_S})$. ■

Fact 7.4.3 *Let $A,B\colon\mathbb{R}^n\to\mathbb{R}^n$ be linear and monotone. Then*

$$F_{(A+B)}=F_A\Box_2 F_B.$$

Proof. See [4, Corollary 5.7]. ■

Fact 7.4.4 *Let S_1,S_2 be linear subspaces of $\mathbb{R}^n$. Then*

$$F_{(N_{S_1}+N_{S_2})}=F_{N_{S_1}}\Box_2 F_{N_{S_2}}.$$

Proof. See [8, Example 4.4]. ■

Fact 7.4.5 *Let S_1, S_2 be linear subspaces of $\mathbb{R}^n$. Then $N_{S_1} + N_{S_2} = N_{S_1 \cap S_2}$.*

Proof. Clearly, $\operatorname{dom}(N_{S_1} + N_{S_2}) = S_1 \cap S_2 = \operatorname{dom} N_{S_1 \cap S_2}$. Let $x \in S_1 \cap S_2$. By Fact 7.1.21,

$$\begin{aligned}
(N_{S_1} + N_{S_2})(x) &= (S_1)^{\perp} + (S_1)^{\perp} = (S_1 \cap S_2)^{\perp} \quad \text{(by [27, Exercises 3.17])} \\
&= N_{S_1 \cap S_2}(x).
\end{aligned}$$

■

Proposition 7.4.6 *Let $B_1, B_2 \colon \mathbb{R}^n \to \mathbb{R}^n$ be linear and monotone, and S_1, S_2 be linear subspaces of $\mathbb{R}^n$ such that $A_1 = B_1 + N_{S_1}, A_2 = B_2 + N_{S_2}$. Then $F_{(A_1+A_2)} = F_{A_1} \Box_2 F_{A_2}$.*

Proof. Let $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$. Then we have

$$\begin{aligned}
&(F_{N_{A_1}} \Box_2 F_{N_{A_2}})(x, x^*) \\
&= \big(F_{(B_1+N_{S_1})} \Box_2 F_{(B_2+N_{S_2})}\big)(x, x^*) \\
&= \inf_{y^*+z^*=x^*} \Big\{ F_{(B_1+N_{S_1})}(x, y^*) + F_{(B_2+N_{S_2})}(x, z^*) \Big\} \\
&= \inf_{y^*+z^*=x^*} \Big\{ \inf_{y_1^*+y_2^*=y^*} \big\{ F_{B_1}(x, y_1^*) + F_{N_{S_1}}(x, y_2^*) \big\} \quad \text{(by Lemma 7.4.2)} \\
&\quad + \inf_{z_1^*+z_2^*=z^*} \big\{ F_{B_2}(x, z_1^*) + F_{N_{S_2}}(x, z_2^*) \big\} \Big\} \quad \text{(by Lemma 7.4.2)} \\
&= \inf_{y_1^*+y_2^*+z_1^*+z_2^*=x^*} \Big\{ F_{B_1}(x, y_1^*) + F_{N_{S_1}}(x, y_2^*) + F_{B_2}(x, z_1^*) + F_{N_{S_2}}(x, z_2^*) \Big\}.
\end{aligned}$$

Thus

$$
\begin{aligned}
&(F_{N_{A_1}} \Box_2 F_{N_{A_2}})(x, x^*) \\
&= \inf_{u^*+v^*=x^*} \Big\{ \inf_{u_1^*+u_2^*=u^*} \big\{F_{B_1}(x, u_1^*) + F_{B_2}(x, u_2^*)\big\} \\
&\quad + \inf_{v_1^*+v_2^*=v^*} \big\{F_{N_{S_1}}(x, v_1^*) + F_{N_{S_2}}(x, v_2^*)\big\}\Big\} \\
&\qquad \text{(by Fact 7.4.3 and Fact 7.4.4)} \\
&= \inf_{u^*+v^*=x^*} \Big\{F_{(B_1+B_2)}(x, u^*) + F_{(N_{S_1}+N_{S_2})}(x, v^*)\Big\} \\
&= \inf_{u^*+v^*=x^*} \Big\{F_{(B_1+B_2)}(x, u^*) + F_{N_{S_1 \cap S_2}}(x, v^*)\Big\} \quad \text{(by Fact 7.4.5)} \\
&= F_{(B_1+B_2+N_{S_1 \cap S_2})}(x, x^*) \quad \text{(by Lemma 7.4.2)} \\
&= F_{(B_1+B_2+N_{S_1}+N_{S_2})}(x, x^*) \quad \text{(by Fact 7.4.5)} \\
&= F_{(A_1+A_2)}(x, x^*).
\end{aligned}
$$

■

Corollary 7.4.7 *Let $A, B\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximal monotone such that* $\operatorname{gra} A$ *and* $\operatorname{gra} B$ *are linear subspaces. Then $F_{(A+B)} = F_A \Box_2 F_B$.*

Proof. By Corollary 7.1.31, there exist linear and monotone operators $A_1, B_1\colon \mathbb{R}^n \to \mathbb{R}^n$ such that $A = A_1 + N_{\operatorname{dom} A}$ and $B = B_1 + N_{\operatorname{dom} B}$. Since $\operatorname{dom} A$ and $\operatorname{dom} B$ are subspaces of $\mathbb{R}^n$, by Proposition 7.4.6,

$$F_{(A+B)} = F_A \Box_2 F_B.$$

■

Remark 7.4.8 *Corollary 7.4.7 generalizes the result of Bauschke, Borwein and Wang in [4].*

Chapter 8

Future work

Our future work is the following

- Simplify some of earlier technic proofs.
- Extend main results to a Hilbert space and a possibly (reflexive) Banach space.
- Since Asplund's decomposition of monotone operators is based on Zorn's Lemma, it would be very interesting to find a constructive proof.

Bibliography

[1] E. Asplund, “A monotone convergence theorem for sequences of nonlinear mappings,” *Nonlinear Functional Analysis*, Proceedings of Symposia in Pure Mathematics, American Mathematical Society, vol. XVIII Part 1, Chicago, pp. 1–9, 1970.

[2] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, “Fitzpatrick functions, cyclic monotonicity, and Rockafellar’s antiderivative,” *Nonlinear Analysis*, No. 5, pp. 1198-1223, 2007.

[3] H. H. Bauschke and J. M. Borwein, “Maximal monotonicity of dense type, local maximal monotonicity, and monotonicity of the conjugate are all the same for continuous linear operators,” *Pacific Journal of Mathematics,* vol. 189, pp. 1–20, 1999.

[4] H. H. Bauschke, J. M. Borwein, and X. Wang, “Fitzpatrick functions and continuous linear monotone operators,” *SIAM Journal on Optimization,* 18, pp. 789–809, 2007.

[5] H. H. Bauschke, Y. Lucet, and M. Trienis, “How to transform one convex function continuously into another,” *SIAM Review*, in press, 2008.

[6] H. H. Bauschke, E. Matouskova, and S. Reich, "Projection and proximal point methods: Convergence results and counterexamples," *Nonlinear Analysis,* vol. 56, pp. 715-739, 2004.

[7] H. H. Bauschke, R. G. Goebel, Y. Lucet, and X. Wang, "The proximal average: Basic Theory," submitted.

[8] H. H. Bauschke, D. A. McLaren, and H. S. Sendov, "Fitzpatrick functions: inequalities, examples and remarks on a problem by S. Fitzpatrick," *Journal of Convex Analysis* 13(3+4), pp. 499-523,2006.

[9] H. H. Bauschke and X. Wang, "The kernel average for two convex functions and its applications to the extension and representation of monotone operators," to appear in *Transactions of the American Mathematical Society*, 2007.

[10] H. H. Bauschke and X. Wang, "A convex-analytical approach to extension results for n-cyclically monotone operators," *Set-Valued Analysis,* No.3, pp. 297-306, 2007.

[11] J. M. Borwein , "Adjoint process duality," *Mathematics of Operations Research*, Vol. 8, No.3, pp. 403-434, 1983.

[12] J. M. Borwein and H. Wiersma, "Asplund decomposition of monotone operators," *SIAM Journal on Optimization,* Vol. 18, No.3, pp. 946-960, 2007.

[13] J. M. Borwein and A. S. Lewis, "Convex Analysis and Nonlinear Optimization," Springer-Verlag, 2000.

[14] D. Butnariu and G. Kassay, "A proximal-projection method for finding zeros of set-valued operators," submitted.

[15] F. H. Clarke, "Optimization and nonsmooth analysis," Wiley, New York, 1983.

[16] R. Cross, "Multivalued linear operators," Marcel Dekker, Inc, New York, 1998.

[17] N. Ghoussoub, "Maximal monotone operators are selfdual vector fields and vice-versa," 2006, preprint.

[18] S. Fitzpatrick, "Representing monotone operators by convex functions," *Workshop/Miniconference on Functional Analysis and Optimization (Canberra 1988)*, Proceedings of the Centre for Mathematical Analysis, Australian National University vol. 20, Canberra, Australia, pp. 59–65, 1988.

[19] C. W. Groetsch, "Generalized Inverses of Linear Operators," Marcel Dekker, Inc, New York, 1977.

[20] D. G. Luenberger, "Optimization by Vector Space Methods," John Wiley Sons, Inc, 2000.

[21] C. D. Meyer, "Matrix Analysis and Applied Linear Algebra," SIAM, New York, 2000.

[22] R.R. Phelps, "Convex functions, Monotone Operators and Differentiability," Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1993.

[23] R. R. Phelps and S. Simons, "Unbounded linear monotone operators on nonreflexive Banach spaces," *Journal of Convex Analysis,* vol. 5, pp. 303–328, 1998.

[24] J. P. Penot and C. Zălinescu, "Some problems about the representation of monotone operators by convex functions," *The Australian New Zealand Industrial and Applied mathematics Journal,* vol. 47, No.1, pp. 1–20, 2005.

[25] R. T. Rockafellar, "Convex Analysis," Princeton University Press, 1970.

[26] R. T. Rockafellar and R. J. B. Wets, "Variational Analysis," Springer-Verlag, 1998.

[27] B. P. Rynne and M. A. Youngson, "Linear Functional Analysis," London, 2000.

[28] S. Simons, "Minimax and Monotonicity," Springer, New York, 1998.

[29] S. Simons and C. Zălinescu, "Fenchel duality, Fitzpatrick functions and maximal monotonicity," *Journal of Nonlinear and Convex Analysis,* 6, pp. 1–22, 2005.

[30] M. D. Voisei, "The sum theorem for linear maximal monotone operators," *Mathematical Sciences Research Journal,* 10 (4), pp. 83–85, 2006.

[31] C. Zălinescu, "Convex Analysis in General Vector Spaces," World Scientific Publishing, 2002.