Subgradient Projectors: Theory, Extensions and Algorithms

by

Jia Xu

M.Sc., Xiamen University, 2012

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

DOCTOR OF PHILOSOPHY

in

THE COLLEGE OF GRADUATE STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

April 2016

The undersigned certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis entitled:

Subgradient Projectors: Theory, Extensions and Algorithms

Submitted by Jia Xu in partial fulfillment of the requirements of

The degree of PhD .

Dr. Heinz Bauschke, Irving K. Barber School of Arts & Sciences, UBCO
Supervisor, Professor (please print name and faculty/school above the line)

Dr. Yves Lucet, Irving K. Barber School of Arts & Sciences, UBCO
Supervisory Committee Member, Professor (please print name and faculty/school in the line above)

Dr. Richard Klukas, School of Engineering, UBCO
Supervisory Committee Member, Professor (please print name and faculty/school in the line above)

Dr. Mohamed Tawhid, Mathematics, Thompson Rivers University
University Examiner, Professor (please print name and faculty/school in the line above)

External Examiner, Professor (please print name and university in the line above)

April 22, 2016
(Date submitted to Grad Studies)

Abstract

The subgradient projector plays an important role in convex optimization, since the subgradient projection algorithm is a classical method for solving convex feasibility problems. This motivates us to explore fundamental properties of the subgradient projector. One can define the subgradient projector of a convex function via a selection of the subdifferential operator. This opens the door to define the subgradient projector of nonconvex functions by using the Mordukhovich limiting subdifferential operator.

This thesis offers a systematic study of the subgradient projector. First, motivated by Polyak and Crombez, we present and analyze a more general algorithm for finding a fixed point of a cutter which is assumed to have a fixed point set with nonempty interior. Our results complement and extend conclusions by Crombez for cutters and by Polyak for subgradient projectors.

We also present a comprehensive list of properties of the subgradient projectors which complements work by Pauwels. Special attention is given to continuity, nonexpansiveness and monotonicity. Finally, for subgradient projectors associated with nonconvex functions, we obtain various characterizations.

Preface

This thesis is based on two published papers [12, 13] and two preprint manuscripts [10, 11]. The papers [12, 13] form the basis of Chapters 3 and 4. Chapters 5 and 6 are based on manuscripts [10, 11], respectively. These are joint works with my supervisors, Dr. Heinz H. Bauschke and Dr. Shawn X. Wang, and also with Dr. Caifang Wang, a visiting scholar from Shanghai Maritime University.

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

Table of Contents

List of Tables

List of Figures

Glossary of Notation

$A : \mathcal{H}_1 \rightrightarrows \mathcal{H}_2$	Set valued operator A from $\mathcal{H}_1$ to $\mathcal{H}_2$. 12
A^*	The adjoint of operator A. 8
$C^\perp$	Orthogonal complement of set C. 19
$G_f x$	Subgradient projector of f at x. 44
$K^\ominus$	The polar cone of set K. 18
$K^\oplus$	The dual cone of set K. 18
$N_\epsilon(x)$	A neighborhood of x. 16
$P_C x$	The metric projection of x onto set C. 17
R_C	Reflection of P_C, $R_C = 2P_C - \mathrm{Id}$. 32
$U^\intercal$	The transpose of matrix U. 121
$U_{r,\eta}$	Relaxation of U_r. 30
V	A real vector space. 4
$\operatorname{Fix} T$	Fixed points of T, $\operatorname{Fix} T = \{x \in \mathcal{H} : x = Tx\}$. 9
$\mathcal{G}_f x$	The set of subgradient projectors associated with selection s. 48
$\operatorname{ball}(x, \epsilon)$	A closed ball centre at x with radius ϵ. 16
$\operatorname{cone} C$	Conical hull of a set C. 18
$\operatorname{conv} C$	The convex hull of a set C. 16
$\operatorname{dom} A$	The domain of A. 20
$\operatorname{gra} A$	The graph of A. 12
ι_C	Indicator function of a set C. 21
$\langle \cdot, \cdot \rangle$	Inner product. 4
$\mathbb{N}$	Strictly positive integers, 1,2,3,$\cdots$. 7
$\mathbb{R}^n$	n-dimentional Euclidean space. 7
$\mathbb{R}_+$	Nonnegative real numbers. 19
$\mathbb{R}_{++}$	Strictly positive real numbers. 17, 23
$\mathbb{R}$	Real numbers. 4
$\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$	The set of linear bounded mappings from $\mathcal{H}_1$ to $\mathcal{H}_2$. 23
$\mathcal{H}$	A real Hilbert space. 5

$\nabla f(x)$	The Gâteaux gradient of f at x. 23, 26
∂f	Subdifferential of f. 25
Prox_f	The proximal mapping of a function f. 23
$\mathrm{rec}\, C$	Recession cone of C. 64
$\mathrm{sgn}(x)$	signum function. 30
${}^{\gamma}f$	Moreau envelope of index γ of a function f. 22
d_C	Distance function to a set C. 21
$f \Box g$	Infimal convolution of the functions f and g. 21
f^+	Positive part of a function f. 46
$x_n \rightharpoonup x$	Weak convergence. 6
$x_n \to x$	Strong convergence. 7
$\|\| \cdot \|\|^*$	The Fenchel conjugate of norm. 51
$\|\|x\|\|$	The norm of x. 5

Acknowledgements

First, I would like to express my heartfelt thanks to my supervisors Dr. Heinz Bauschke and Dr. Shawn Wang, for their careful guidance throughout my thesis. Their enthusiasm and passion have deeply affected me in Mathematics. Without their help, I could not have completed my thesis. It is my great pleasure to work with them.

Also, thanks to my committee member Dr. Yves Lucet for providing many valuable suggestions. I also thank the other members of the examination committee, Dr. Mohamed Tawhid and Dr. Richard Klukas, for their comments, and Dr. Susan Murch for chairing the defense.

Finally, I want to express my gratitude to my colleagues Chayne Planiden, Walaa Moursi, Yipin Guo, Sarah Moffat and Dongying Wang who supported me during my study. They helped me a lot.

Dedication

For William Lee.

Chapter 1

Introduction

The study of subgradient projectors and their properties is motivated by a large number of problems, for example, *convex feasibility problems*. The convex feasibility problem is to find a point x in the intersection of finitely many convex subsets C_i of a Hilbert space $\mathcal{H}$. A lot of algorithms have been created to solve this problem by making use of subgradient projectors. The most famous instance of such algorithms is Polyak's subgradient projection algorithm (see Theorem 3.1). Therefore, subgradient projectors are beneficial to the study of the convex feasibility problem and we are encouraged to obtain a more complete understanding of the subgradient projector.

On the other hand, we can formulate the convex feasibility problem as a *common fixed point problem*, where each convex subset $C_i = \operatorname{Fix} U_i$ for some operator U_i. If we are dealing with only one operator U, then our results generalize works by Polyak [50] and by Crombez [31].

The majority of the theoretical background can be found in Bauschke and Combettes' *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* [7]; Cegielski's *Iterative Methods for Fixed Point Problems in Hilbert Spaces* [20]; Mordukhovich's *Variational Analysis and Generalized Differentiation I* [40]; and Rockafeller and Wets' *Variational Analysis* [53].

The rest of this thesis is organized as follows.

Chapter 2 gives basic background information on operators, convex analysis, and some related known results. Chapters 3-6 contain new results.

My contribution begins in Chapter 3, with a new method for finding a fixed point of a cutter. Related properties are also identified. The material in this chapter is based on [13], which appeared in the *Journal of Optimization Theory and Applications*.

Chapter 4 provides fundamental properties of the subgradient projector, such as continuity, nonexpansiveness and monotonicity. The Yamagishi-Yamada operator is also discussed. Results in this chapter can be found in [12], which appeared in the *SIAM Journal on Optimization*.

Chapter 5 is based on the manuscript [10]. In the previous chapter, we consider the subgradient projectors associated with convex functions. Complementarily for nonconvex functions, the associated subgradient projectors

also exist. In this part, we provide the basic theory of subgradient projectors for possibly nonconvex functions.

Chapter 6 mainly focuses on charaterizations of linear subgradient projectors. We explicitly offer the closed expression of the subgradient projector, which itself is linear and symmetric. This chapter is based on the manuscript [11].

Finally, the key results of my thesis and directions for future research are presented in Chapter 7.

Chapter 2

Auxiliary Results

2.1 Overview

Throughout this thesis, we work with Hilbert spaces and Euclidean spaces. This chapter contains some basic mathematical definitions and also provides various examples and results, mostly from Convex Analysis and Monotone Operator Theory [7].

2.2 Vector Spaces

Definition 2.1. A real *vector space* (also called *linear space*) is a set V of vectors together with two operations: vector addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{R} \times V \to V$ with the following properties:

(1) vector addition satisfies:

(i) (commutative law) $\forall u, v \in V$, then $u + v = v + u$;

(ii) (associative law) $\forall u, v, w \in V$, then $(u + v) + w = u + (v + w)$;

(iii) (additive identity) there exists a vector $0 \in V$ such that $\forall u \in V$, $0 + u = u + 0 = u$;

(iv) (additive inverse) for $\forall u \in V$, there is a vector $-u \in V$ such that $u + (-u) = 0$;

(2) scalar multiplication satisfies:

(i) (distributive law for vectors) $\forall c \in \mathbb{R}$, $\forall u, v \in V$, then $c \cdot (u + v) = c \cdot u + c \cdot v$;

(ii) (distributive law for real numbers) $\forall a, b \in \mathbb{R}$, $\forall u \in V$, then $(a+b) \cdot u = a \cdot u + b \cdot u$;

(iii) (associative law for real numbers) $\forall a, b \in \mathbb{R}$, $\forall u \in V$, then $a \cdot (b \cdot u) = (a \cdot b) \cdot u$;

(iv) (unitary law) there is a vector $1 \in V$ such that $\forall u \in V$, $1 \cdot u = u$.

Definition 2.2. [54, Definition 13.1] Let S be a set and suppose $d : S \times S \to [0, \infty)$ is a function on S satisfying

(i) $d(x, x) = 0$ for all $x \in S$ and $d(x, y) > 0$ for distinct $x, y \in S$;

(ii) $d(x, y) = d(y, x)$ for all $x, y \in S$;

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in S$.

Such a function d is called a *metric* on S. A *metric space*, denoted as (S, d), is a set S together with a metric d on it.

Definition 2.3. [36] An *inner product* on a real vector space V is a function on $V \times V \to \mathbb{R}$, $(x, y) \to \langle x, y \rangle$ with the following properties:

(i) $(\forall x, y, z \in V)$ $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$;

(ii) $(\forall a \in \mathbb{R}, \forall x, z \in V)$ $\langle ax, z \rangle = a\langle x, z \rangle$;

(iii) $(\forall x, y \in V)$ $\langle x, y \rangle = \langle y, x \rangle$, $\forall x, y \in V$;

(iv) $(\forall x \in V)$ $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

An *inner product space*, denoted as $(V, \langle \cdot, \cdot \rangle)$, is a vector space with an inner product defined on it.

Definition 2.4. [36, Definition 2.2-1] A *norm* on a real vector space V is a real-valued function on V whose value at $x \in V$ is denoted by $\|x\|$ and which has the following properties

(i) $(\forall x \in V)$ $\|x\| \geq 0$. $\|x\| = 0$ if and only if $x = 0$;

(ii) $(\forall a \in \mathbb{R}, \forall x \in V)$ $\|ax\| = |a|\|x\|$;

(iii) $(\forall x, y \in V)$ $\|x + y\| \leq \|x\| + \|y\|$.

The *normed space*, denoted as $(V, \|\cdot\|)$, is a vector space with a norm defined on it.

An inner product on V defines a norm on V which is given by

$$(\forall x \in V) \quad \|x\| = \sqrt{\langle x, x \rangle} \tag{2.1}$$

A norm on V defines a metric d on V which is given by

$$(\forall x, y \in V) \quad d(x, y) = \|x - y\|. \tag{2.2}$$

Remark 2.5. [36] A norm on an inner product space given by (2.1) satisfies the important *parallelogram equality*[1]. If a norm does not satisfy the paral-

[1] Parallelogram equality: $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$

lelogram equality, then it is not derived from an inner product by the use of (2.1).

Definition 2.6. [7] A sequence $(x_n)_{n=1}^{\infty}$ in an inner product space V is a *Cauchy sequence* if $\forall \epsilon > 0$, there exists $N \in \mathbb{N}$, when $n, m > N$, we have

$$d(x_m, x_n) < \epsilon. \tag{2.3}$$

An inner product space V is *complete* if every Cauchy sequence in V converges to a point in V.

Definition 2.7. [32, Definition 7.7.2] A real *Hilbert space*, $\mathcal{H}$, is a complete inner product space.

Example 2.8. [36] The space of square summable sequences,

$$\ell^2 = \{(x_n)_{n=1}^{\infty} : \sum_{n=1}^{\infty} |x_n|^2 < \infty\}, \tag{2.4}$$

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

In Hilbert space $\mathcal{H}$, the following is an important inequality we will use frequently.

Fact 2.9. *(Cauchy-Schwarz inequality) Let $x, y \in \mathcal{H}$. Then*

$$|\langle x, y\rangle| \leq \|x\|\|y\|. \tag{2.5}$$

In a real Hilbert space $\mathcal{H}$, a *ball* with centre $x \in \mathcal{H}$ and radius $\epsilon > 0$ is given by

$$\text{ball}(x, \epsilon) = \{y \in \mathcal{H} : \ \|y - x\| \leq \epsilon\}. \tag{2.6}$$

Definition 2.10. [36] We say a sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H}$ is *bounded* if $(x_n)_{n=1}^{\infty} \subset \text{ball}(x_0, r)$ for some $x_0 \in \mathcal{H}$ and $r > 0$.

Definition 2.11. [36, Definition 1.4-1] A sequence $(x_n)_{n=1}^{\infty}$ in a real Hilbert space $\mathcal{H}$ is said to be *convergent* if there is an $x \in \mathcal{H}$, for $\forall \epsilon > 0$, there is $N \in \mathbb{N}$, when $n > N$, then

$$\|x_n - x\| < \epsilon. \tag{2.7}$$

x is called the *limit* of $(x_n)_{n=1}^{\infty}$ and we write $x_n \to x$.

Definition 2.12. (Weak convergence)[20] Let $x \in \mathcal{H}$ and $(x_n)_{n=1}^{\infty}$ be a sequence in $\mathcal{H}$. We say that $(x_n)_{n=1}^{\infty}$ *converges weakly* to x, denoted as $x_n \rightharpoonup x$, if

$$(\forall z \in \mathcal{H}) \quad \langle x_n, z\rangle \to \langle x, z\rangle. \tag{2.8}$$

The following result is well known (see e.g. [36, Theorem 4.8-4(a)]); we give its proof for completeness.

Fact 2.13. *Every convergent sequence in $\mathcal{H}$ is weakly convergent.*

Proof. Suppose $(x_n)_{n=1}^{\infty}$ is a convergent sequence in $\mathcal{H}$ and x is its limit. For every $z \in \mathcal{H}$, by the Cauchy Schwarz inequality, we have

$$0 \leq |\langle x_n, z\rangle - \langle x, z\rangle| = |\langle x_n - x, z\rangle| \leq \|x_n - x\|\|z\| \to 0. \tag{2.9}$$

So $(x_n)_{n=1}^{\infty}$ is weakly convergent to x. □

The converse of Fact 2.13 is not true.

Example 2.14. Consider $(e_n)_{n=1}^{\infty}$: $e_1 = (1, 0, 0, 0, \cdots)$, $e_2 = (0, 1, 0, 0, \cdots)$, $e_3 = (0, 0, 1, 0, \cdots), \cdots$, the sequence of standard unit vectors in ℓ^2. $\forall x \in \ell^2$, by *Bessel inequality*[2], we have

$$\sum_{n=1}^{\infty} |\langle e_n, x\rangle|^2 \leq \|x\|^2. \tag{2.10}$$

Thus the series $\sum_{n=1}^{\infty} |\langle e_n, x\rangle|^2$ is convergent. Then its corresponding sequence $(|\langle e_n, x\rangle|^2)_{n=1}^{\infty}$ converges to 0, which implies

$$\langle e_n, x\rangle \to \langle 0, x\rangle = 0 \ (\forall x \in \ell^2). \tag{2.11}$$

Therefore, $(e_n)_{n=1}^{\infty}$ converges weakly to 0. Since $\|e_n - 0\| = \|e_n\| = 1$, we deduce that e_n does not converge to 0.

The following result is well known (see e.g., [36, Theorem 4.8-4(c)]); we give its proof for completeness.

Fact 2.15. *A weakly convergent sequence of a finite dimensional space is convergent.*

[2]Bessel inequality: Let $(e_n)_{n\in\mathbb{N}}$ be the standard orthonormal basis of $\mathcal{H}$, $(\forall x \in \mathcal{H})$, we have $\sum \langle x, e_n\rangle^2 \leq \|x\|^2$.

Proof. Let $(x_n)_{n=1}^{\infty} \in \mathbb{R}^m$, where $x_n = (x_n(1), x_n(2), \cdots, x_n(m))$, converges weakly to $x \in \mathbb{R}^m$, by definition of weak convergence, choosing $z = e_i$, $i \in \{1, 2, \cdots, m\}$, a standard unit vector in $\mathbb{R}^m$, then we have

$$\langle x_n, e_i \rangle \longrightarrow \langle x, e_i \rangle, \tag{2.12}$$

which is equivalent to $x_n(i) \to x(i)$. Meanwhile,

$$\begin{aligned}
x_n(i) \to x(i) &\Rightarrow |x_n(i) - x(i)| \to 0, && (2.13)\\
&\Rightarrow |x_n(i) - x(i)|^2 \to 0, && (2.14)\\
&\Rightarrow \sum_{i=1}^{m} |x_n(i) - x(i)|^2 \to 0, && (2.15)\\
&\Rightarrow \|x_n - x\|^2 \to 0, && (2.16)\\
&\Rightarrow \|x_n - x\| \to 0, && (2.17)\\
&\Rightarrow x_n \to x. && (2.18)
\end{aligned}$$

□

By the Uniform Boundedness Principle, we will see that a weakly convergent sequence in a real Hilbert space is bounded. See Fact 2.24 below.

The following powerful notion of the Fejér monotone sequence is central in the study of various iterative methods.

Definition 2.16. A sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H}$ is called *Fejér monotone* with respect to a nonempty subset S of $\mathcal{H}$ if and only if for $\forall s \in S,\ \forall n \in \mathbb{N}$,

$$\|x_{n+1} - s\| \leq \|x_n - s\|. \tag{2.19}$$

Example 2.17. [7, Example 5.2] Let's assume that $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ is bounded and monotonically increasing, i.e.,

$$(\forall n \in \mathbb{N})\ x_n \leq x_{n+1} \leq \cdots \leq \beta \tag{2.20}$$

for some constant $\beta > 0$. Then sequence $(x_n)_{n=1}^{\infty}$ is Fejér monotone with respect to $C \subseteq [\beta, +\infty)$.

Clearly, every Fejér monotone sequence is bounded. Actually, if $(x_n)_{n=1}^{\infty}$ is a Fejér monotone sequence with respect to a nonempty set C, then $\forall c \in C$,

$$\|x_{n+1} - c\| \leq \|x_n - c\| \leq \cdots \leq \|x_1 - c\|. \tag{2.21}$$

That is to say, $(x_n)_{n=1}^{\infty}$ lies in $\mathrm{ball}(c, \|x_1 - c\|)$.

2.3 Operators

2.3.1 Linear Bounded Operators

Definition 2.18. [36, Definition 2.6-1] A *linear operator* T is an operator such that

(i) the domain $\operatorname{dom} T$ of T is a real vector space and the range $\operatorname{ran} T$ of T lies in a real vector space;

(ii) for all $x, y \in \operatorname{dom} T$ and $a, b \in \mathbb{R}$,

$$T(ax + by) = aTx + bTy. \tag{2.22}$$

Definition 2.19. [36, Definition 2.7-1] Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces and $T : \operatorname{dom} T \to \mathcal{H}_2$ a linear operator, where $\operatorname{dom} T \subset \mathcal{H}_1$. The operator T is said to be *bounded* if there is a real number c such that for all $x \in \operatorname{dom} T$,

$$\|Tx\| \leq c\|x\|. \tag{2.23}$$

We denote $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ the set of all linear bounded operators from $\mathcal{H}_1$ to $\mathcal{H}_2$.

Definition 2.20. [36, page 92] The norm of operator T is defined as:

$$\|T\| = \sup_{\|x\| \leq 1} \|Tx\|. \tag{2.24}$$

Definition 2.21. [7] Let $T \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, the *adjoint* of T is the unique operator $T^* \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ that satisfies

$$(\forall x \in \mathcal{H}_1)(\forall y \in \mathcal{H}_2) \;\; \langle Tx, y\rangle = \langle x, T^* y\rangle. \tag{2.25}$$

Definition 2.22. [20, Definition 1.1.19] We say that a bounded linear operator $A : \mathcal{H} \to \mathcal{H}$ is *self-adjoint* if $A^* = A$.

Theorem 2.23. *(Uniform Boundedness Principle) [36, Theorem 4.7-3] Let (T_n) be a sequence of bounded linear operators $T_n : X \to Y$ from a Banach space X into a normed space Y such that $(\|T_n x\|)$ is bounded for every $x \in X$, say $\|T_n x\| \leq c_x$, where c_x is a real number depends on x. Then the sequence of the operator norms $\|T_n\|$ is bounded, that is $\exists c > 0$ such that $\|T_n\| \leq c$.*

Now let us prove that a weakly convergent sequence in a real Hilbert space $\mathcal{H}$ is bounded.

Fact 2.24. *Every weakly convergent sequence in a real Hilbert space $\mathcal{H}$ is bounded.*

Proof. Let $x_n \rightharpoonup x$. Then for every $z \in \mathcal{H}$, we have $\langle x_n, z\rangle \to \langle x, z\rangle$. So we have $|\langle x_n, z\rangle| \to |\langle x, z\rangle|$. This implies $(|\langle x_n, z\rangle|)_{n\in\mathbb{N}}$ is bounded. Define $T_n : \mathcal{H} \to \mathbb{R} : y \mapsto \langle x_n, y\rangle$. Then $(T_n)_{n=1}^{\infty}$ is pointwise bounded. By the Uniform Boundness Principle, we deduce that $(\|T_n\|)_{n=1}^{\infty}$ is bounded. Finally, we show that $\|x_n\| = \|T_n\|$. Indeed,

$$\|T_n\| = \sup_{\|z\|\leq 1} |\langle x_n, z\rangle| \leq \sup_{\|z\|\leq 1} \|x_n\|\|z\| = \|x_n\|. \tag{2.26}$$

If $x_n = 0$, then $\|x_n\| = 0$ and so $\|T_n\| = 0$. If $x_n \neq 0$, we have

$$\|T_n\| \geq \left|\langle x_n, \frac{x_n}{\|x_n\|}\rangle\right| = \frac{\|x_n\|^2}{\|x_n\|} = \|x_n\|. \tag{2.27}$$

Altogether, we have $\|T_n\| = \|x_n\|$. □

2.3.2 Cutters

Definition 2.25. [20, Definition 2.1.30] We call $T : \mathcal{H} \to \mathcal{H}$ a *cutter* if and only if $\operatorname{Fix} T \neq \varnothing$ and for $\forall x \in \mathcal{H},\ \forall y \in \operatorname{Fix} T$,

$$\langle y - Tx, x - Tx\rangle \leq 0. \tag{2.28}$$

Example 2.26. (Subgradient projector) Let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and continuous such that $\{x \in \mathcal{H} :\ f(x) \leq 0\} \neq \varnothing$, and let $s\colon \mathcal{H} \to \mathcal{H}$ be a selection of ∂f (for definition of subgradient, see Definition 2.66), i.e., $(\forall x \in \mathcal{H})\ s(x) \in \partial f(x)$. Then the associated *subgradient projector*, defined by

$$(\forall x \in \mathcal{H}) \quad G_f x = \begin{cases} x - \dfrac{f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > 0; \\ x, & \text{otherwise,} \end{cases} \tag{2.29}$$

is a cutter. For details, see Fact 4.1 (iv).

The following fact says that a cutter is strongly Fejér monotone[3] with respect to the set of its fixed points, .

[3](see [20, page 108]) We say an operator $T : \mathcal{H} \to \mathcal{H}$ is *strongly Fejér monotone* with respect to the set of its fixed points if there exists a constant $\alpha > 0$ such that

$$\|Tx - z\|^2 \leq \|x - z\|^2 - \alpha\|Tx - x\|^2,$$

for all $z \in \operatorname{Fix} T$.

Fact 2.27. *Suppose that $T : \mathcal{H} \to \mathcal{H}$ has a fixed point.*

(i) A mapping $T : \mathcal{H} \to \mathcal{H}$ is a cutter if and only if

$$(\forall x \in \mathcal{H})(y \in \operatorname{Fix} T)\ \|x - Tx\|^2 \leq \|x - y\|^2 - \|Tx - y\|^2.$$

(ii) Let $T : \mathcal{H} \to \mathcal{H}$ be a cutter. Then T is always continuous on $\operatorname{Fix} T$.

(iii) Let $T : \mathcal{H} \to \mathcal{H}$ be a cutter. Then $\operatorname{Fix} T$ is closed and convex.

Proof. (i) Let $x \in \mathcal{H}, \forall y \in \operatorname{Fix} T$.

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

Thus we can conclude that

$$\|x - Tx\|^2 + \|y - Tx\|^2 \leq \|x - y\|^2 \Longleftrightarrow \langle y - Tx, x - Tx\rangle \leq 0. \quad (2.30)$$

(ii) Suppose that $T : \mathcal{H} \to \mathcal{H}$ is a cutter. Let $y \in \operatorname{Fix} T$. From (i) we have

$$\|Tx - Ty\|^2 = \|Tx - y\|^2 \leq \|x - y\|^2 - \|x - Tx\|^2 \leq \|x - y\|^2. \quad (2.31)$$

Now $\forall \varepsilon > 0$, let $\delta = \varepsilon$. Then

$$\|Tx - Ty\| \leq \varepsilon. \quad (2.32)$$

when $\|x - y\| \leq \delta$. Therefore T is continuous on $\operatorname{Fix} T$.

(iii) Suppose that $T : \mathcal{H} \to \mathcal{H}$ is a cutter. By [20, Lemma 2.1.36], we have

$$\operatorname{Fix} T = \bigcap_{x \in \mathcal{H}} \{y \in \mathcal{H} :\ \langle y - Tx, x - Tx\rangle \leq 0\}. \quad (2.33)$$

Denote the right hand side of (2.33) as

$$S := \bigcap_{x \in \mathcal{H}} \{y \in \mathcal{H} :\ \langle x - Tx, y - Tx\rangle \leq 0\}. \quad (2.34)$$

First we show $S \subseteq \operatorname{Fix} T$. Indeed, let $y \in S$, i.e., $\langle x - Tx, y - Tx\rangle \leq 0$ for all $x \in \mathcal{H}$. If we take $x = y$, we get $\|y - Ty\|^2 \leq 0$ and hence, $y = Ty$, i.e., $y \in \operatorname{Fix} T$. Conversely, let $y \in \operatorname{Fix} T$. Since T is a cutter, we have $\langle x - Tx, y - Tx\rangle \leq 0$. Then $y \in S$. Hence we have $\operatorname{Fix} T = S$. Finally, we observe that S is the intersection of a family of sets that are closed and convex, so is $\operatorname{Fix} T$.

□

2.3.3 Single-valued and Set-valued Monotone Operators

Definition 2.28. [20, Definition 1.1.28] We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is *monotone* if

$$(\forall x, y \in \mathcal{H})\ \langle Tx - Ty, x - y\rangle \geq 0. \tag{2.35}$$

The above definition is just for a single-valued operator. We now extend this to multivalued or set-valued operators. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ i.e., $(\forall x \in \mathcal{H})$, $Ax \subseteq \mathcal{H}$ be a set-valued operator. The *graph* of A is defined by

$$\operatorname{gra} A = \{(x, x^*) \in \mathcal{H} \times \mathcal{H} :\ x^* \in Ax\}. \tag{2.36}$$

Definition 2.29. [7] $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is *monotone* if $\forall (x, x^*)\ (y, y^*) \in \operatorname{gra} A$,

$$\langle x - y, x^* - y^*\rangle \geq 0. \tag{2.37}$$

Let's introduce the definition of maximally monotone operator.

Definition 2.30. [7] Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. Then A is *maximally monotone* if $B : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone and $\operatorname{gra} A \subseteq \operatorname{gra} B \Rightarrow A = B$.

2.3.4 Nonexpansive Mappings

Definition 2.31. Let $T : \mathcal{H} \to \mathcal{H}$. Then T is

(i) *firmly nonexpansive* if $\forall x, y \in \mathcal{H}$

$$\|Tx - Ty\|^2 + \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2 \leq \|x - y\|^2. \tag{2.38}$$

(ii) *nonexpansive* if $\forall x, y \in \mathcal{H}$,

$$\|Tx - Ty\| \leq \|x - y\|. \tag{2.39}$$

(iii) *quasi-nonexpansive* if $\forall x \in \mathcal{H}$, $\forall y \in \operatorname{Fix} T$,

$$\|Tx - Ty\| \leq \|x - y\|. \tag{2.40}$$

We record some properties of nonexpansive maps that are most useful for us. It is clear that firm nonexpansiveness implies nonexpansiveness, which itself implies quasinonexpansiveness. However, the converse is not true. See following results and counterexamples.

Theorem 2.32. *[20, Theorem 2.2.4] A firmly nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ is monotone and nonexpansive.*

Proof. Let T be firmly nonexpansive. Then

$$\|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leq \|x - y\|^2. \tag{2.41}$$

$$\|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 = \|(x - y) - (Tx - Ty)\|^2 \tag{2.42}$$

$$= \|x - y\|^2 + \|Tx - Ty\|^2 - 2\langle x - y, Tx - Ty\rangle. \tag{2.43}$$

Then we have

$$\text{T is firmly nonexpansive} \Leftrightarrow \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle \tag{2.44}$$

By the Cauchy-Schwarz inequality, we have

$$\|Tx - Ty\| \cdot \|x - y\| \geq \langle Tx - Ty, x - y\rangle \geq \|Tx - Ty\|^2 \geq 0, \tag{2.45}$$

for all $x, y \in \mathcal{H}$, which yields the monotonicity and the nonexpansivity of T. □

Actually, firmly nonexpansiveness is equivalent to monotonicity and nonexpansiveness on the real line. Indeed, necessity is obvious from the above theorem. For sufficiency, let $T : \mathbb{R} \to \mathbb{R}$ be a monotone and nonexpansive operator. Then we have

$$(\forall x, y \in \mathbb{R})\ (x - y)(Tx - Ty) \geq 0 \text{ and } |Tx - Ty| \leq |x - y|. \tag{2.46}$$

Then

$$|Tx - Ty|^2 = |Tx - Ty||Tx - Ty| \tag{2.47}$$

$$\leq |x - y||Tx - Ty| \tag{2.48}$$

$$= |(x - y)(Tx - Ty)| \tag{2.49}$$

$$= (x - y)(Tx - Ty). \tag{2.50}$$

Thus, a monotone nonexpansive operator on the real line is firmly nonexpansive. However, if not on the real line, nonexpansivenss and monotonicity will not imply firmly nonexpansiveness. See Example 2.35.

Lemma 2.33. *[20, Lemma 2.1.20] A nonexpansive operator $U : \mathcal{H} \to \mathcal{H}$ with a fixed point is quasi-nonexpansive.*

The converse of Lemma 2.33 is not true.

Example 2.34. [20] Let $T : \mathbb{R} \to \mathbb{R}$,

$$Tx = \begin{cases} x^2, & \text{if } |x| \leq \frac{3}{4}, \\ |x| - \frac{3}{16}, & \text{if } |x| > \frac{3}{4}. \end{cases} \tag{2.51}$$

Then T is a continuous quasi-nonexpansive operator, but not nonexpansive.

Proof. (i) T is continuous at $\frac{3}{4}$ as $\lim_{x\to(\frac{3}{4})^+} Tx = \lim_{x\to(\frac{3}{4})^+} x - \frac{3}{16} = \frac{9}{16}$, $\lim_{x\to(\frac{3}{4})^-} Tx = \lim_{x\to(\frac{3}{4})^-} x^2 = \frac{9}{16}$ and $T(\frac{3}{4}) = \frac{9}{16}$. Similarly, we have T is continuous at $-\frac{3}{4}$. Then T is continuous.

(ii) T is quasi-nonexpansive. It is easy to calculate that $\operatorname{Fix} T = \{0\}$. If $|x| \leq \frac{3}{4} < 1$, then

$$|0 - Tx| = |Tx| = |x^2| \leq |x| = |0 - x|. \tag{2.52}$$

If $|x| > \frac{3}{4} > \frac{3}{16}$, then

$$|0 - Tx| = \left||x| - \frac{3}{16}\right| = |x| - \frac{3}{16} < |x| = |0 - x|. \tag{2.53}$$

So T is quasi-nonexpansive.

(iii) T is not nonexpansive when $x = \frac{1}{2}, y = 1$. Indeed,

$$|Tx - Ty| = \left|\frac{1}{4} - (1 - \frac{3}{16})\right| = \frac{9}{16} > \left|\frac{1}{2} - 1\right| = |x - y|. \tag{2.54}$$

Therefore T is a continuous quasi-nonexpansive operator but not nonexpansive. □

The converse of Theorem 2.32 does not hold.

Example 2.35. [20] Let $T : \mathbb{R}^2 \to \mathbb{R}^2$,

$$Tx = (x_1 \cos\theta - x_2 \sin\theta, x_1 \sin\theta + x_2 \cos\theta) \tag{2.55}$$

is nonexpansive and monotone for $\theta \in (0, \frac{\pi}{2}]$, but T is not firmly nonexpansive.

Proof. Let $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R}^2$, then

(i) T is monotone. Since

$$Tx - Ty \tag{2.56}$$
$$= ((x_1 - y_1)\cos\theta - (x_2 - y_2)\sin\theta, (x_1 - y_1)\sin\theta + (x_2 - y_2)\cos\theta). \tag{2.57}$$

Therefore we have

$$\langle x - y, Tx - Ty\rangle = (x_1 - y_1)^2\cos\theta - (x_1 - y_1)(x_2 - y_2)\sin\theta \tag{2.58}$$
$$+ (x_1 - y_1)(x_2 - y_2)\sin\theta + (x_2 - y_2)^2\cos\theta \tag{2.59}$$
$$= \left((x_1 - y_1)^2 + (x_2 - y_2)^2\right)\cos\theta. \tag{2.60}$$

Since $\theta \in (0, \frac{\pi}{2}]$, then $\cos\theta > 0$, so $\langle x - y, Tx - Ty\rangle \geq 0$.

(ii) T is nonexpansive. Indeed,

$$\|Tx - Ty\|^2 \tag{2.61}$$
$$= ((x_1 - y_1)\cos\theta - (x_2 - y_2)\sin\theta)^2 \tag{2.62}$$
$$+ ((x_1 - y_1)\sin\theta + (x_2 - y_2)\cos\theta)^2 \tag{2.63}$$
$$= (x_1 - y_1)^2(\cos\theta)^2 + (x_2 - y_2)^2(\sin\theta)^2 \tag{2.64}$$
$$-2(x_1 - y_1)(x_2 - y_2)\sin\theta\cos\theta \tag{2.65}$$
$$+(x_1 - y_1)^2(\sin\theta)^2 + (x_2 - y_2)^2(\cos\theta)^2 \tag{2.66}$$
$$+2(x_1 - y_1)(x_2 - y_2)\sin\theta\cos\theta \tag{2.67}$$
$$= (x_1 - y_1)^2 + (x_2 - y_2)^2 \tag{2.68}$$
$$= \|x - y\|^2. \tag{2.69}$$

(iii) T is not firmly nonexpansive.
From (i), we have

$$\langle x - y, Tx - Ty\rangle = \left((x_1 - y_1)^2 + (x_2 - y_2)^2\right)\cos\theta = \cos\theta\|x - y\|^2. \tag{2.70}$$

From (ii), we have $\|Tx - Ty\|^2 = \|x - y\|^2$.
Then $\|Tx - Ty\|^2 > \langle x - y, Tx - Ty\rangle$ for $\forall\theta \in (0, \frac{\pi}{2}]$ and distinct x, y.

□

2.4 Convex Analysis

2.4.1 Convex Sets

Definition 2.36. A subset C of $\mathcal{H}$ is *convex* if for all $x, y \in C$ and $\lambda \in (0,1)$,

$$\lambda x + (1-\lambda)y \in C. \tag{2.71}$$

(a) convex set, (b) nonconvex set.

Figure 2.1: Examples of convex and nonconvex sets

For example, a ball is convex. See Fig. 2.1 (a)

Definition 2.37. [7, Definition 3.3] Let $C \subset \mathcal{H}$. The *convex hull* of C is the intersection of all the convex subsets of $\mathcal{H}$ containing C, i.e., the smallest convex subset of $\mathcal{H}$ containing C. It is denoted by $\operatorname{conv} C$.

Proposition 2.38. *[7, Proposition 3.4] Let $C \subset \mathcal{H}$. Then*

$$\operatorname{conv} C = \left\{ \sum_{i\in I} \alpha_i x_i : \ I \text{ finite}, \ (x_i)_{i\in I} \subset C, \ (\alpha_i)_{i\in I} \subset \left]0,1\right], \sum_{i\in I} \alpha_i = 1 \right\}. \tag{2.72}$$

A *neighborhood* of $x \in \mathcal{H}$, is an open ball of radius ϵ, which is defined as

$$N_\epsilon(x) = \{y \in \mathcal{H} : \|y - x\| < \epsilon\}. \tag{2.73}$$

The *interior of* C, is the set of all interior points of C, denoted as $\operatorname{int} C$,

$$\operatorname{int} C = \{x \in \mathcal{H} : \ \exists \epsilon > 0, N_\epsilon(x) \subseteq C\}, \tag{2.74}$$

and is also the largest open set contained in C.

Definition 2.39. [20] Let $C \subseteq \mathcal{H}$ be a nonempty subset and $x \in \mathcal{H}$. If there exists a point $y \in C$ such that

$$\|y - x\| \leq \|z - x\| \tag{2.75}$$

for any $z \in C$, then y is called a *metric projection* of x onto C and is denoted by $P_C x$ (see Fig. 2.2).

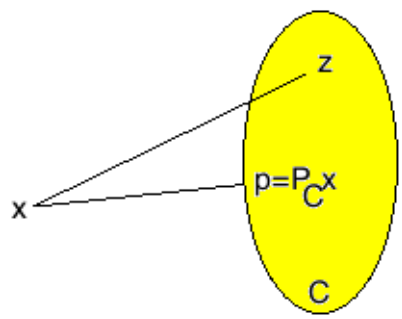


Figure 2.2: Metric projection.

Fact 2.40. *(Projection characterization) [7, Theorem 3.14] Let C be a nonempty closed convex subset of $\mathcal{H}$. Then for every x and p in $\mathcal{H}$,*

$$p = P_C x \text{ if and only if } (\forall z \in C)\ \langle z - p, x - p \rangle \leq 0 \text{ and } p \in C.$$

Definition 2.41. [7] Let $C \subseteq \mathcal{H}$. We say C is a *cone* if

$$C = \mathbb{R}_{++} C. \tag{2.76}$$

$\mathbb{R}_{++} := \{\xi \in \mathbb{R} : \xi > 0\}$.

Definition 2.42. [7, Definition 6.1] Let C be a subset of $\mathcal{H}$. The *conical hull* of C is the intersection of all the cones in $\mathcal{H}$ containing C, denoted by $\operatorname{cone} C$.

Actually, from [7, Proposition 6.2], we have $\operatorname{cone} C = \mathbb{R}_{++}C$.

Definition 2.43. [7, Definition 6.21] Let K be a subset of $\mathcal{H}$. The *polar cone* of K is

$$K^{\ominus} = \{u \in \mathcal{H} : \sup\langle K, u\rangle \leq 0\}, \tag{2.77}$$

and the *dual cone* of K is $K^{\oplus} = -K^{\ominus}$. If K is a nonempty convex cone, then K is *self-dual* if $K = K^{\oplus}$. If K is a nonempty closed convex cone in $\mathcal{H}$, K is *acute* if $K \subset K^{\oplus}$, and K is *obtuse* if $K^{\oplus} \subset K$.

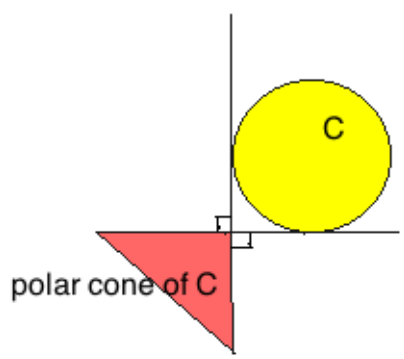


Figure 2.3: Illustration of polar cone.

Example 2.44. Let $\mathcal{H} = \mathbb{R}$ and $K = \mathbb{R}_+$,

$$\begin{aligned} K^{\ominus} &= \{u \in \mathcal{H} : \sup\langle K, u\rangle \leq 0\} && (2.78)\\ &= \{u \in \mathbb{R} : y \cdot u \leq 0, \forall y \in \mathbb{R}_+\} && (2.79)\\ &= \mathbb{R}_- && (2.80)\\ &= -K. && (2.81) \end{aligned}$$

By definitions above, we have

$$K^{\oplus} = -K^{\ominus} = K. \tag{2.82}$$

Since $K = \mathbb{R}_+$ is a nonempty closed convex cone, then $K = \mathbb{R}_+$ is a acute cone and an obtuse cone as well.

Definition 2.45. [7] The *orthogonal complement* of a subset C of $\mathcal{H}$ is denoted by $C^{\perp}$, i.e.,

$$C^{\perp} = \{u \in \mathcal{H} : \ \langle x, u \rangle = 0, \forall x \in \mathcal{H}\}. \tag{2.83}$$

Proposition 2.46. *[7, Propostion 6.22] Let C be a linear subspace of $\mathcal{H}$. Then $C^{\ominus} = C^{\perp}$.*

Definition 2.47. [7] Let D be a nonempty convex subset of X. The *recession cone* of D is

$$\operatorname{rec} D = \{x \in X : \ x + D \subset D\}. \tag{2.84}$$

Proposition 2.48. *[7, Corollary 6.49] Let C be a nonempty closed convex cone in $\mathcal{H}$. Then* $\operatorname{rec} C = C$.

Proposition 2.49. *Let $T : \mathcal{H} \to \mathcal{H}$ be a cutter. Then*

$$\operatorname{ran}(\operatorname{Id} - T) \subseteq (\operatorname{rec}(\operatorname{Fix} T))^{\ominus}. \tag{2.85}$$

Consequently, when $\operatorname{Fix} T$ *is a linear subspace,*

$$\operatorname{ran}(\operatorname{Id} - T) \subseteq (\operatorname{Fix} T)^{\perp}. \tag{2.86}$$

In other words,

$$\operatorname{ran}(\operatorname{Id} - T) \subseteq (\ker(\operatorname{Id} - T))^{\perp}. \tag{2.87}$$

Proof. Let $x - Tx \in \operatorname{ran}(\operatorname{Id} - T)$ and $v \in \operatorname{rec}(\operatorname{Fix} T)$. Then for every $k > 0$, $u \in \operatorname{Fix} T$, $u + kv \in \operatorname{Fix} T$. We have

$$\langle x - Tx, u + kv - Tx \rangle \leq 0 \quad \Rightarrow \quad \langle x - Tx, u/k + v - Tx/k \rangle \leq 0. \tag{2.88}$$

When $k \to \infty$ this gives

$$\langle x - Tx, v \rangle \leq 0. \tag{2.89}$$

Since $v \in \operatorname{rec}(\operatorname{Fix} T)$ was arbitrary, $x - Tx \in (\operatorname{rec}(\operatorname{Fix} T))^{\ominus}$.

When $\operatorname{Fix} T$ is a linear subspace, $\operatorname{Fix} T = \operatorname{rec}(\operatorname{Fix} T)$ and $(\operatorname{rec}(\operatorname{Fix} T))^{\ominus} = (\operatorname{rec}(\operatorname{Fix} T))^{\perp}$. $\square$

2.4.2 Convex Functions

Let $f : \mathcal{H} \to [-\infty, +\infty]$. The function is *proper* if $-\infty \notin f(\mathcal{H})$ and $\operatorname{dom} f \neq \varnothing$.

Definition 2.50. A function $f : \mathcal{H} \to]-\infty, +\infty]$ is said to be *convex* if its domain, $\operatorname{dom} f = \{x \in \mathcal{H} : \ f(x) < +\infty\}$ is a convex set and $\forall x, y \in \mathcal{H}, \lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{2.90}$$

Definition 2.51. Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function. Then f is *strictly convex* if $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \lambda \in]0, 1[)$,

$$x \neq y \Rightarrow f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y). \tag{2.91}$$

Definition 2.52. A function f is *lower semi-continuous* if for every sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H}$,

$$x_n \to x \Rightarrow f(x) \leq \liminf_{n\to\infty} f(x_n). \tag{2.92}$$

In [54], the *limit inferior* of sequence (s_n) in $\mathbb{R}$, is defined by

$$\liminf s_n = \lim_{N\to\infty} \inf\{s_n : n > N\}. \tag{2.93}$$

The *limit superior* is defined by

$$\limsup s_n = \lim_{N\to\infty} \sup\{s_n : n > N\}. \tag{2.94}$$

Example 2.53. The ceiling function $f(x) = \lceil x \rceil = \min\{n \in \mathbb{Z} : n \geq x\}$ (see Fig. 2.4) is lower semicontinuous.

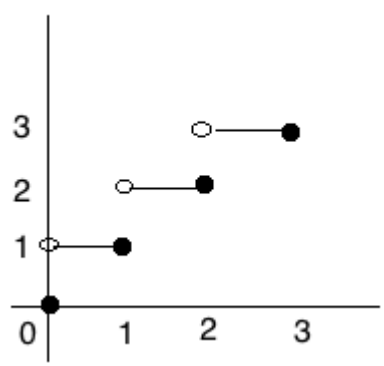


Figure 2.4: The graph of the ceiling function, an example of a lower semi-continuous function.

In the following, we define the infimal convolution operator. See [7] for more details.

Definition 2.54. Let f and g be functions from $\mathcal{H}$ to $]-\infty,+\infty]$. The *infimal convolution* of f and g is

$$f\Box g : \mathcal{H} \to]-\infty,+\infty] : x \mapsto \inf_{y\in\mathcal{H}} (f(y) + g(x-y)). \tag{2.95}$$

This operator gives us an alternative way to express certain functions, for example the distance function.

The *indicator function* of a set $C \subset \mathcal{H}$, is the function defined by

$$\iota_C : \mathcal{H} \to [-\infty,+\infty] : \ x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.96}$$

Fact 2.55. *The distance from $x \in \mathcal{H}$ to a set $C \subseteq \mathcal{H}$ is defined by*

$$d_C(x) = \inf_{y\in C} \|x-y\|. \tag{2.97}$$

Then we have $d_C(x) = (\iota_C \Box \|\cdot\|)(x)$.

Proof. For $x \in \mathcal{H}$, we have

$$\begin{aligned}
(\iota_C \Box \|\cdot\|)(x) &= \inf_{y\in\mathcal{H}} (\iota_C(y) + \|x-y\|) && (2.98)\\
&= \inf_{y\in C} (\iota_C(y) + \|x-y\|) \quad (\iota_C(y) = +\infty \text{ for } y \notin C) && (2.99)\\
&= \inf_{y\in C} \|x-y\| \qquad\qquad (\iota_C(y) = 0 \text{ for } y \in C) && (2.100)\\
&= d_C(x). && (2.101)
\end{aligned}$$

□

Now let's concentrate on a special case of infimal convolution, the Moreau envelope, as it is from there that the proximal mapping is defined.

Definition 2.56. [7, Definition 12.20] Let $f : \mathcal{H} \to]-\infty,+\infty]$ and let $\gamma \in \mathbb{R}_{++}$. The *Moreau envelope* of f of parameter γ is

$${}^{\gamma}f = f\Box(\frac{1}{2\gamma}\|\cdot\|^2). \tag{2.102}$$

The following is a simple example using the distance and the indicator functions.

Example 2.57. [7, Example 12.21] Let $C \subset \mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Then ${}^{\gamma}\iota_C = \frac{1}{2\gamma} d_C^2$.

Proof. For $x \in \mathcal{H}$, we have

$$
\begin{aligned}
{}^{\gamma}\iota_C(x) &= \inf_{y \in \mathcal{H}} \left(\iota_C(y) + \frac{1}{2\gamma}\|x-y\|^2 \right) && (2.103)\\
&= \inf_{y \in C} \left(\iota_C(y) + \frac{1}{2\gamma}\|x-y\|^2 \right) \quad (\iota_C(y) = +\infty \text{ for } y \notin C) && (2.104)\\
&= \inf_{y \in C} \frac{1}{2\gamma}\|x-y\|^2 \quad (\iota_C(y) = 0 \text{ for } y \in C) && (2.105)\\
&= \frac{1}{2\gamma} d_C^2(x). && (2.106)
\end{aligned}
$$

□

Actually, the infimal convolution of a convex function with a power of the norm has some particular porperties.

Remark 2.58. [7, Proposition 12.15] Let f be a proper lower semicontinuous convex function on $\mathcal{H}$, let $\gamma \in \mathbb{R}_{++}$ and let $p \in]1, +\infty[$. Then the infimal convolution

$$
f \square \frac{1}{\gamma p}\|\cdot\|^p : \mathcal{H} \to]-\infty, +\infty] : x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{\gamma p}\|x-y\|^p \tag{2.107}
$$

is convex, real-valued, continuous and exact[4]. Moreover, for every $x \in \mathcal{H}$, the infimum in (2.107) is uniquely attained. Indeed, since $\frac{1}{\gamma p}\|\cdot\|^p$ is stricly convex and supercoercive[5], then by [7, Corollary 11.15 (i)], we have shown that the infimum is uniquely attained.

In the case $p = 2$, Remark 2.58 motivated the following definition.

Definition 2.59. [7, Definition 12.23] Let f be a proper lower semicontinuous convex function on $\mathcal{H}$ and let $x \in \mathcal{H}$. Then $\operatorname{Prox}_f x$ is the unique point

[4] The infimal convolution is exact at a point $x \in \mathcal{H}$ if $(f \square g)(x) = \min_{y \in \mathcal{H}} f(y) + g(x-y)$, i.e. there exists a point $y \in \mathcal{H}$ such that $(f \square g)(x) = f(y) + g(x-y)$.

[5] Let $f : \mathcal{H} \to [-\infty, +\infty]$. Then f is *supercoercive* if $\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty$.

in $\mathcal{H}$ that satisfies

$$^1 f(x) = \min_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2}\|x - y\|^2 \right) = f(\operatorname{Prox}_f x) + \frac{1}{2}\|x - \operatorname{Prox}_f x\|^2. \quad (2.108)$$

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

Definition 2.60. [7] (Gâteaux differentiability) Let $f : C \to \mathbb{R}$, with $C \subseteq \mathcal{H}$, and let $x \in C$ be such that $(\forall y \in \mathcal{H})(\exists \alpha \in \mathbb{R}_{++})\ [x, x + \alpha y] \subseteq C$. We say that f is *Gâteaux differentiable* at x if there exists an operator $Df(x) \in \mathcal{B}(\mathcal{H}, \mathcal{H})$, called the *Gâteaux derivative* of f at x, such that

$$(\forall y \in \mathcal{H}) \quad \lim_{\alpha \downarrow 0} \frac{f(x + \alpha y) - f(x)}{\alpha} = Df(x)y. \quad (2.109)$$

The Gâteaux derivative $Df(x)$ in Definition 2.60 is unique whenever it exists. Moreover, since $Df(x)$ is linear, for every $y \in \mathcal{H}$, we have $Df(x)y = -Df(x)(-y)$, and we can therefore replace (2.109) by

$$(\forall y \in \mathcal{H}) \quad \lim_{0 \neq \alpha \to 0} \frac{f(x + \alpha y) - f(x)}{\alpha} = Df(x)y. \quad (2.110)$$

Remark 2.61. Let C be a subset of $\mathcal{H}$, let $f : C \to \mathbb{R}$, and suppose that f is Gâteaux differentiable at $x \in C$. Then by Riesz-Fréchet representation[6], there exists a unique vector $\nabla f(x) \in \mathcal{H}$ such that

$$(\forall y \in \mathcal{H}) \quad Df(x)y = \langle y, \nabla f(x) \rangle. \quad (2.111)$$

We call $\nabla f(x)$ the *Gâteaux gradient* of f at x.

Definition 2.62. [7] (Fréchet differentiability) Let $x \in \mathcal{H}$ and let $f : U \to]-\infty, +\infty]$ where U is a neighborhood of x. Then f is *Fréchet differentiable* at x if there exists an operator $Df(x) \in \mathcal{B}(\mathcal{H}, \mathcal{H})$, called the *Fréchet derivative* of f at x such that

$$\lim_{0 \neq \|y\| \to 0} \frac{\|f(x + y) - f(x) - Df(x)y\|}{\|y\|} = 0. \quad (2.112)$$

The *Fréchet gradient* of a function $f : U \to \mathbb{R}$ at $x \in U$ is defined as in Remark 2.61.

[6] [7, Fact 2.17] Riesz-Fréchet representation: Let $g \in \mathcal{B}(\mathcal{H}, \mathbb{R})$. Then there exists a unique vector $u \in \mathcal{H}$ such that $(\forall x \in \mathcal{H})\ g(x) = \langle x, u \rangle$. Moreover, $\|g\| = \|u\|$.

Fact 2.63. *[20, Theorem 1.1.13, Theorem 1.1.14]*

(i) If f is Fréchet differentiable at $x \in \mathcal{H}$, then it is Gâteaux differentiable at x.

(ii) If f is Gâteaux differentiable in a neighborhood of x and its Gâteaux derivative is continuous (strong-to-strong) at x, then f is Fréchet differentiable at x.

Remark 2.64. [16] For locally Lipschitz functions on $\mathbb{R}^n$, Gâteaux and Fréchet differentiability coincide.

Example 2.65. [57, Example 1.1.2] Consider the function

$$f : \mathbb{R}^2 \to \mathbb{R} : \ (x_1, x_2) \mapsto \begin{cases} \dfrac{x_1 x_2^3}{x_1^2 + x_2^4}, & \text{if } (x_1, x_2) \neq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.113}$$

Then f is continuous and Gâteaux differentiable at $(0,0)$, but not Fréchet differentiable at $(0,0)$.

Proof. Let $x = (x_1, x_2)$. Observe that $f(x) = 0$ if $x_1 x_2 = 0$. Now assume that $x_1 \neq 0$ and that $x_2 \neq 0$. Since

$$0 \leq (|x_1| - x_2^2)^2 \Leftrightarrow 2|x_1|x_2^2 \leq x_1^2 + x_2^4, \tag{2.114}$$

we estimate

$$|f(x)| = \frac{|x_1||x_2|^3}{x_1^2 + x_2^4} \leq \frac{|x_1||x_2|^3}{2|x_1|x_2^2} = \frac{|x_2|}{2} \to 0 \text{ as } x \to 0. \tag{2.115}$$

Thus, f is continuous at $(0,0)$.
Let $t \in \mathbb{R}_{++}$. If $x_1 = 0$, then $\frac{f(tx)}{t} = 0$; otherwise

$$\frac{f(tx)}{t} = \frac{f(tx_1, tx_2)}{t} = \frac{t^4 x_1 x_2^3}{t^2 x_1^2 + t^4 x_2^4} = \frac{t^2 x_1 x_2^3}{x_1^2 + t^2 x_2^4} \to 0 \text{ as } t \downarrow 0. \tag{2.116}$$

Thus f is Gâteaxu differentiable at $(0,0)$, with $\nabla f(0,0) = (0,0)$.
It remains to show that f is not Fréchet differentiable at $(0,0)$. Since $f(0,0) = 0$ and $\nabla f(0,0) = 0$, the quotient of interest is simply $\frac{f(y)}{\|y\|}$, where $y = (y_1, y_2) \neq (0,0)$. We will show that this quotient does not tend to 0 as $y \to 0$. Observe that

$$\left(\frac{f(y)}{\|y\|}\right)^2 = \frac{y_1^2 y_2^6}{(y_1^2 + y_2^2)(y_2^2 + y_2^4)^2}, \tag{2.117}$$

which when $y_1 = y_2^2$ leads to

$$\frac{y_1^2 y_2^6}{(y_1^2+y_2^2)(y_2^2+y_2^4)^2} = \frac{y_2^1 0}{y_2^2(1+y_2^2)4y_2^8} = \frac{1}{4(1+y_2^2)} \to \frac{1}{4} \text{ as } y_2 \to 0. \quad (2.118)$$

Hence $\overline{\lim}_{0\neq\|y\|\to 0} \frac{f(y)}{\|y\|} \geq \frac{1}{2}$ and therefore f is not Fréchet differentiable at $(0,0)$. □

Definition 2.66. [7, Definition 16.1] Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper convex function. The *subdifferential* of f is the set-valued operator

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \{u \in \mathcal{H} : \langle u, y-x\rangle + f(x) \leq f(y),\ \forall y \in \mathcal{H}\}. \quad (2.119)$$

The elements of $\partial f(x)$ are the *subgradients* of f at x.

Fact 2.67. *[20, Theorem 1.1.57] A continuous convex function $f : \mathcal{H} \to \mathbb{R}$ is Gâteaux differentiable at $x \in \mathcal{H}$ if and only if its subdifferential $\partial f(x)$ consists of one point. In this case, $\partial f(x) = \{Df(x)\} = \{\nabla f(x)\}$.*

Example 2.68. (1) The absolute value $f(x) = |x|$ is subdifferentiable at every $x \in \mathbb{R}$. Precisely,

$$(\forall x \in \mathbb{R}) \quad \partial f(x) = \begin{cases} 1, & \text{if } x > 0, \\ [-1,1], & \text{if } x = 0, \\ -1. & \text{if } x < 0. \end{cases} \quad (2.120)$$

(2) For a differentiable function $f(x) = x^2$ on $\mathbb{R}$, then its subdifferential is a singleton, which is the gradient of f, that is $\partial f(x) = \{\nabla f(x)\} = \{2x\}$ for $\forall x \in \mathbb{R}$.
(3) The subdifferential of the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^n$ has the form

$$(\forall x \in \mathbb{R}^n) \quad \partial(\|x\|) = \begin{cases} \left\{\dfrac{x}{\|x\|}\right\}, & \text{if } x \neq 0, \\ \text{ball}(0,1), & \text{if } x = 0. \end{cases} \quad (2.121)$$

We have now covered the elementary results needed for the following chapters of this thesis, including subgradient projectors and cutters. In the next chapter, we present a new method for finding a fixed point of a cutter.

Chapter 3

Finite Convergence Result of a Projected Cutter Method

3.1 Overview

This chapter is based on the paper [13]. We obtain finite convergence results for a general class of algorithms provided a constraint qualification is satisfied. Our results complement and extend conclusions by Crombez for cutters and by Polyak for subgradient projectors. Comparisons of our results with existing algorithms are also included. We begin with the motivation for this research.

Let a finite family of closed convex subsets $C_i, i \in I$, of a Hilbert space $\mathcal{H}$ be given. Denote $C = \cap_{i \in I} C_i$. If $C \neq \varnothing$, then the *convex feasibility problem* is to find a point $x \in C$. A particular way to solve the convex feasibility problem is to express the solution set as the fixed point set of a proper algorithmic operator. The most famous instance is Polyak's subgradient projector of a convex function.

Theorem 3.1. *[50] (Polyak's subgradient projection algorithm) Let C be a nonempty closed convex subset of $\mathcal{H}$. Let $f : \mathcal{H} \to \mathbb{R}$ be a continuous convex function. Assume* $\operatorname{argmin} f = \{x \in \mathcal{H} : \ f(x) = \inf f(C)\} \neq \varnothing$. *Without loss of generality, we assume* $\min f(C) = 0$. *Assume ∂f is bounded on bounded sets. Let $(\alpha_n) \subseteq]0, 2[$ such that $\exists\, \epsilon > 0$, $\epsilon \leq \alpha_n \leq 2 - \epsilon$. Let $s \in \partial f$ be a selection and define*

$$x_{n+1} = \begin{cases} P_C\left(x_n - \alpha_n \dfrac{f(x_n)}{\|s(x_n)\|^2} s(x_n)\right), & \text{if } f(x_n) > 0, \\ x_n, & \text{if } f(x_n) \leq 0. \end{cases} \tag{3.1}$$

Then (x_n) converges weakly to a minimizer of f over C.

For the proof of this theorem, see [50].

Theorem 3.2. *[20, Theorem 3.5.1] (Opial, 1967) Let $C \subseteq \mathcal{H}$ be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$ and $U : C \to C$ be a nonexpansive*

and asymptotically regular operator with a fixed point. Then, for any $x \in C$, the sequence $(U^k x)_{k=1}^{\infty}$ converges weakly to a point $z \in \operatorname{Fix} U$.

If the suitable algorithmic operator is nonexpansive and asymptotically regular[7], then Opial's result implies that the iterates of the operator converge weakly to a solution.

However, whether this solution can be reached within a finite number of steps is not fully answered. Censor et al. [21] studied the modified cyclic subgradient projection (MCSP) and provided finite convergence conditions while Crombez [31] focussed on the iterations of a cutter whose fixed points set contains a ball of a known radius. Unfortunately, this radius is not necessarily known in practice.

3.2 Basic Properties

Given $r \geq 0$, we follow Crombez [31] and define the operator $U_r \colon \mathcal{H} \to \mathcal{H}$ at $x \in \mathcal{H}$ by

$$U_r x := \begin{cases} Tx + \dfrac{r}{\|Tx - x\|}(Tx - x), & \text{if } x \neq Tx, \\ x, & \text{otherwise.} \end{cases} \tag{3.2}$$

Obviously, we can rewrite $U_r x$ as

$$U_r x = x + \frac{r + \|Tx - x\|}{\|Tx - x\|}(Tx - x) \tag{3.3}$$

when $x \notin \operatorname{Fix} T$. When T is a subgradient projector, then U_r was also studied by Polyak [50]. Note that $\operatorname{Fix} U_r = \operatorname{Fix} T$.

Lemma 3.3. *[13, Lemma 2.1] Let $y \in \operatorname{Fix} T$, let $r \in \mathbb{R}_{++}$, and suppose that $\operatorname{ball}(y; r) \subseteq \operatorname{Fix} T$ and that $x \in \mathcal{H} \smallsetminus \operatorname{Fix} T$. Set*

$$\tau_x := \langle x - y, \frac{x - Tx}{\|x - Tx\|} \rangle - \big(r + \|x - Tx\|\big). \tag{3.4}$$

Then the following hold:

(i) $\tau_x \geq 0$.

[7] An operator $U : \mathcal{H} \to \mathcal{H}$ is called *asymptotically regular* if for all $x \in \mathcal{H}$, then we have

$$\lim_{k \to \infty} \|U^{k+1} x - U^k x\| = 0.$$

(ii) $\|U_r x - y\|^2 = \|Tx - y\|^2 - r^2 - 2r\tau_x \leq \|Tx - y\|^2 - r^2$.

(iii) $\|U_r x - y\|^2 = \|x - y\|^2 - (r + \|x - Tx\|)^2 - 2\tau_x(r + \|x - Tx\|)$
$\leq \|x - y\|^2 - (r + \|x - Tx\|)^2 \leq \|x - y\|^2 - r^2 - \|x - Tx\|^2$.

Proof. (i): Set $z = y + r(x - Tx)/\|x - Tx\|$. Then $z \in \operatorname{ball}(y; r) \subseteq \operatorname{Fix} T$. Since T is a cutter, we obtain

$$0 \geq \langle z - Tx, x - Tx\rangle \tag{3.5a}$$
$$= \langle y + r(x - Tx)/\|x - Tx\| - Tx\, x - Tx\rangle \tag{3.5b}$$
$$= \langle y - Tx, x - Tx\rangle + r\|x - Tx\| \tag{3.5c}$$
$$= \langle y - x, x - Tx\rangle + \|x - Tx\|^2 + r\|x - Tx\|. \tag{3.5d}$$

Rearranging and dividing by $\|x - Tx\|$ yields

$$\langle x - y, (x - Tx)/\|x - Tx\|\rangle \geq r + \|x - Tx\| \tag{3.6}$$

and hence $\tau_x \geq 0$.

(ii): Using (3.2), we derive the identity from

$$\|U_r x - y\|^2 \tag{3.7a}$$
$$= \big\|x + (\|x - Tx\| + r)/\|x - Tx\|(Tx - x) - y\big\|^2 \tag{3.7b}$$
$$= \big\|(Tx - y) + r(Tx - x)/\|Tx - x\|\big\|^2 \tag{3.7c}$$
$$= \|Tx - y\|^2 + r^2 + 2r\langle (Tx - x) + (x - y), (Tx - x)/\|Tx - x\|\rangle \tag{3.7d}$$
$$= \|Tx - y\|^2 + r^2 + 2r\|x - Tx\| - 2r\langle x - y, (x - Tx)/\|x - Tx\|\rangle \tag{3.7e}$$
$$= \|Tx - y\|^2 - r^2 - 2r\tau_x. \tag{3.7f}$$

The inequality follows immediately from (i).

(iii): Using (ii), we obtain

$$= \|U_r x - y\|^2 \tag{3.8a}$$
$$= \|(x - y) + (Tx - x)\|^2 - r^2 - 2r\tau_x \tag{3.8b}$$
$$= \|x - y\|^2 + \|x - Tx\|^2 + 2\langle x - y, Tx - x\rangle - r^2 - 2r\tau_x \tag{3.8c}$$
$$= \|x - y\|^2 - \|x - Tx\|^2 - 2(\tau_x + r)\|x - Tx\| - r^2 - 2r\tau_x \tag{3.8d}$$
$$= \|x - y\|^2 - (r + \|x - Tx\|)^2 - 2\tau_x(r + \|x - Tx\|). \tag{3.8e}$$

The inequalities now follow from (i). □

Example 3.4. (U_r need not be a cutter) [13, Example 2.2] Suppose that $\mathcal{H} = \mathbb{R}$ and that T is the subgradient projector associated with the function $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2 - 1$. Then $\operatorname{Fix} T = [-1, 1]$. Let

$$r \in \mathbb{R}_+ := \{\xi \in \mathbb{R} : \xi \geq 0\}. \tag{3.9}$$

Then for $\forall x \in \mathbb{R} \smallsetminus \operatorname{Fix} T$, we have

$$U_r x = \frac{x}{2} + \frac{1}{2x} - r \operatorname{sgn}(x), \tag{3.10}$$

where $\operatorname{sgn}(\cdot)$ is the signum function defined as

$$\operatorname{sgn}(x) := \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0. \end{cases} \tag{3.11}$$

Choosing $y := 1 \in \operatorname{Fix} T$ and $x := y + \varepsilon \notin \operatorname{Fix} T$, where $\varepsilon \in \mathbb{R}_{++}$ and $r = 1$, then

$$y - U_r x = 1 - (\frac{1+\varepsilon}{2} + \frac{1}{2(1+\varepsilon)} - 1) = 2 - \frac{1+\varepsilon}{2} - \frac{1}{2(1+\varepsilon)} > 0, \tag{3.12}$$

when ε is sufficient small. And

$$x - U_r x = 1 + \varepsilon - (\frac{1+\varepsilon}{2} + \frac{1}{2(1+\varepsilon)} - 1) = 2 + \varepsilon - \frac{1+\varepsilon}{2} - \frac{1}{2(1+\varepsilon)} > 0. \tag{3.13}$$

Then we have $\langle y - U_r x, x - U_r x \rangle > 0$. By definition of a cutter, we can easily check that U_r is not a cutter when ε is sufficiently small.

The following result concerns a relaxed version of U_r.

Corollary 3.5. *[13, Corollary 2.1] Let $y \in \operatorname{Fix} T$, let $r \in \mathbb{R}_{++}$, let $\eta \in \mathbb{R}_+$, and suppose that* $\operatorname{ball}(y; r) \subseteq \operatorname{Fix} T$ *and that $x \in \mathcal{H} \smallsetminus \operatorname{Fix} T$. Set*

$$U_{r,\eta} x := x + \eta \frac{r + \|x - Tx\|}{\|Tx - x\|}(Tx - x). \tag{3.14}$$

Then the following hold

(i) $U_{r,\eta} x = (1-\eta)x + \eta U_r x$.

(ii) $\|U_{r,\eta} x - y\|^2 = \eta \|U_r x - y\|^2 + (1-\eta)\|x - y\|^2 - \eta(1-\eta)\|x - U_r x\|^2$.

(iii) $\|U_r x - x\| = r + \|x - Tx\|$.

(iv) $\|U_r x - y\|^2 \leq \|x-y\|^2 - (r + \|x - Tx\|)^2 = \|x-y\|^2 - \|x - U_r x\|^2$.

(v) $\|U_{r,\eta} x - y\|^2 \leq \|x-y\|^2 - \eta(2-\eta)(r + \|x - Tx\|)^2$
$= \|x-y\|^2 - \eta^{-1}(2-\eta)\|x - U_{r,\eta} x\|^2$.

Proof. (i):

$$U_{r,\eta} x = x + \eta \frac{r + \|x - Tx\|}{\|Tx - x\|}(Tx - x) \tag{3.15}$$

$$= x + \eta(x + \frac{r + \|x - Tx\|}{\|x - Tx\|}(Tx - x) - x) \tag{3.16}$$

$$= x + \eta(U_r x - x) \tag{3.17}$$

$$= (1-\eta)x + \eta U_r x. \tag{3.18}$$

(ii): Using (i), we obtain $\|U_{r,\eta} x - y\|^2 = \|(1-\eta)(x-y) + \eta(U_r x - y)\|^2$. Now use [7, Corollary 2.14] to obtain the identity.

(iii): Since $U_r x = x + \frac{r + \|x - Tx\|}{\|x - Tx\|}(Tx - x)$, then

$$U_r x - x = \frac{r + \|x - Tx\|}{\|x - Tx\|}(Tx - x). \tag{3.19}$$

So we have

$$\|U_r x - x\| = r + \|x - Tx\|. \tag{3.20}$$

(iv): Combine (iii) with Lemma 3.3 (iii).

(v): Combine (i)-(iv). Indeed, more precisely,

$$\|U_{r,\eta} x - y\|^2 \tag{3.21}$$

$$= \eta\|U_r x - y\|^2 + (1-\eta)\|x-y\|^2 - \eta(1-\eta)\|x - U_r x\|^2 \tag{3.22}$$

$$\leq \eta(\|x-y\|^2 - \|x - U_r x\|^2) + (1-\eta)\|x-y\|^2 - \eta(1-\eta)\|x - U_r x\|^2 \tag{3.23}$$

$$= \|x-y\|^2 - \eta(2-\eta)\|x - U_r x\|^2 \tag{3.24}$$

$$= \|x-y\|^2 - \eta(2-\eta)(r + \|x - Tx\|)^2. \tag{3.25}$$

And by definition of $U_{r,\eta}$, we have $\|u_{r,\eta} x - x\| = \eta(r + \|x - Tx\|)$, therefore

$$\|U_{r,\eta} x - y\|^2 \leq \|x-y\|^2 - \eta(2-\eta)(r + \|x - Tx\|)^2 \tag{3.26}$$

$$= \|x-y\|^2 - \eta(2-\eta)(\frac{1}{\eta}\|u_{r,\eta} x - x\|)^2 \tag{3.27}$$

$$= \|x-y\|^2 - \eta^{-1}(2-\eta)\|x - U_{r,\eta} x\|^2. \tag{3.28}$$

□

Definition 3.6. We call $Q\colon \mathcal{H} \to \mathcal{H}$ a *quasi projector* of C if and only if $\operatorname{ran} Q = \operatorname{Fix} Q = C$ and for $\forall x \in \mathcal{H},\ \forall c \in C$

$$\|Qx - c\| \leq \|x - c\|. \tag{3.29}$$

Example 3.7. (Projectors are quasi projectors) [13, Example 2.3] P_C is a quasi projector of C. Since $\|P_C x - x\| = \inf_{c\in C} \|x - c\| \leq \|x - c\|$ and $\operatorname{ran} P_C = \operatorname{Fix} P_C = C$. More generally if $R\colon \mathcal{H} \to \mathcal{H}$ is quasi nonexpansive, i.e., $\forall x \in \mathcal{H}, \forall y \in \operatorname{Fix} R$, then $\|Rx - y\| \leq \|x - y\|$ and $C \subseteq \operatorname{Fix} R$, then $P_C \circ R$ is a quasi projector of C. Indeed, for $\forall c \in C \subseteq \operatorname{Fix} R$, then

$$\|P_C \circ Rx - c\| = \|P_C \circ Rx - P_C c\| \leq \|Rx - c\| \leq \|x - c\|. \tag{3.30}$$

And also we have $\operatorname{ran} P_C \circ R = \operatorname{Fix} P_C \circ R = C$.

It can be shown (see [1, Proposition 3.4.4]) that when C is an *affine subspace* (i.e., if $C \neq \varnothing$, and $\forall \lambda \in \mathbb{R}$, $C = \lambda C + (1 - \lambda)C$), then the only quasi projector of C is the projector. However, we will now see that for certain cones there are quasi projectors different from projectors.

Proposition 3.8. *(Reflector of an obtuse cone) [13, Proposition 2.1] Suppose that C is an obtuse convex cone, i.e., $\mathbb{R}_+ C = C$ and $C^\ominus := \{x \in \mathcal{H} : \sup\langle C, x\rangle = 0\} \subseteq -C$. Then the reflector $R_C := 2P_C - \operatorname{Id}$ is nonexpansive and $\operatorname{ran} R_C = \operatorname{Fix} R_C = C$.*

Proof. (i) $\operatorname{ran} R_C = \operatorname{Fix} R_C = C$. For $\forall x \in \mathcal{H}$,

$$\begin{aligned} R_C x &= 2P_C x - x = 2P_C x - (P_C x + P_{C^\ominus} x) && (3.31)\\ &= P_C x - P_{C^\ominus} x && (3.32)\\ &\in C - C^\ominus && (3.33)\\ &\subseteq C + C = C. && (3.34) \end{aligned}$$

So $\operatorname{ran} R_C = C$.
Let $x = R_C x$, then $x = 2P_C x - x$, which is equivalent to $x = P_C x$, then $x \in C$. So we have $\operatorname{Fix} R_C = C$.

(ii) Since P_C is firmly nonexpansive, then $2P_C - \operatorname{Id}$ is nonexpansive. Then inequality (3.29) is satisfied.

Therefore, the reflector of an obtuse cone is a quasi projector of C. □

Corollary 3.9. *[13, Corollary 2.2] Suppose that C is an obtuse cone and let $\lambda\colon \mathcal{H} \to [1, 2]$. Then*

$$Q\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \big(1 - \lambda(x)\big)x + \lambda(x) P_C x \tag{3.35}$$

is a quasi projector of C.

Proof. Since, for every $x \in \mathcal{H}$, when $\lambda(x) = 1$, then $Q(x) = P_C x$, when $\lambda(x) = 2$, $Q(x) = R_C x$. So we have $Q(x) \in [P_C x, R_C x]$. Both P_C and R_C are quasi projectors of C from Proposition 3.8. Therefore Q is a quasi projector of C. □

Example 3.10. [13, Example 2.4] Suppose $\mathcal{H} = \mathbb{R}^d$ and $C = \mathbb{R}^d_+$. Then R_C is a quasi projector.

Proof. Referring to Example 2.44, we have $C^{\ominus} = -C$. Then C is an obtuse cone. The conclusion follows from Corollary 3.9 with $\lambda(x) \equiv 2$. □

Remark 3.11. [13, Remark 2.1] A quasi projector need not be continuous because we may choose λ in Corollary 3.9 discontinuously. For instance, let $\mathcal{H} = \mathbb{R}$, $C = \mathbb{R}_+$, from Example 2.44, we have C is an obtuse cone. Define

$$\lambda(x) = \begin{cases} 1, & \text{if } x \leq -1, \\ 2, & \text{if } x > -1. \end{cases} \tag{3.36}$$

When $x \leq -1$, then $\lambda(x) = 1$, $P_C x = 0$. Then $Q(x) = 0$. When $-1 < x < 0$, then $\lambda(x) = 2$, $P_C x = 0$. Then $Q(x) = -x \to 1$ $(x \to -1^+)$. Thus Q is discontinuous at $x = -1$.

Fact 3.12. *(Raik)[51] Let $(x_n)_{n=1}^{\infty}$ be a sequence in $\mathcal{H}$ that is* Fejér *monotone with respect to a subset S of $\mathcal{H}$. If $\operatorname{int} S \neq \varnothing$, then $\sum_{n\in\mathbb{N}} \|x_n - x_{n+1}\| < +\infty$ and $(x_n)_{n=1}^{\infty}$ converges strongly to some point in $\mathcal{H}$.*

Lemma 3.13. *[13, Lemma 2.2] Suppose that $\mathcal{H}$ is finite-dimensional, let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and Fréchet differentiable such that $\inf f(\mathcal{H}) < 0$. Then for every $\rho \in \mathbb{R}_{++}$, we have*

$$\inf\{\|\nabla f(x)\| : x \in \operatorname{ball}(0;\rho) \cap f^{-1}(\mathbb{R}_{++})\} > 0. \tag{3.37}$$

Proof. Let $\rho \in \mathbb{R}_{++}$ and assume to the contrary that the conclusion fails. Since $\mathcal{H}$ is finite dimensional space, by Bolzano-Weierstrass Theorem[8], then there exists a sequence $(x_n)_{n=1}^{\infty}$ in $\operatorname{ball}(0;\rho) \cap f^{-1}(\mathbb{R}_{++})$ and a point $x \in \operatorname{ball}(0;\rho)$ such that $x_n \to x$ and $\nabla f(x_n) \to 0$. Since f is Fréchet differentiable, then ∇f is continuous. Then $\nabla f(x_n) \to \nabla f(x) = 0$ and $f(x_n) \to f(x) \geq 0$. On the other hand, $\inf f(x) < 0$ and $\nabla f(x) = 0$, then we have $f(x) < 0$, which is absurd. □

[8] Bolzano-Weierstrass Theorem: In $\mathbb{R}^k$, every bounded sequence has a convergent subsequence.

3.3 Finitely Convergent Cutter Method

Assume that $(r_n)_{n=1}^{\infty}$ is a sequence in $\mathbb{R}_{++}$ such that $r_n \to 0$ and $(\eta_n)_{n=1}^{\infty}$ is a sequence in $]0,2]$. Let Q_C be a quasi projector of C. We further assume that $x_0 \in C$ and that $(x_n)_{n=1}^{\infty}$ is generated by

$$x_{n+1} := \begin{cases} Q_C\big(x_n + \eta_n(U_{r_n}x_n - x_n)\big), & \text{if } x_n \notin \operatorname{Fix} T, \\ x_n, & \text{otherwise.} \end{cases} \tag{3.38}$$

Note that $(x_n)_{n=1}^{\infty}$ lies in C. Also observe that if x_n lies in $\operatorname{Fix} T$, then so does x_{n+1}.

Theorem 3.14. *[13, Theorem 3.1] Assume that $(r_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{R}_{++}$ such that $r_n \to 0$ and $(\eta_n)_{n\in\mathbb{N}}$ is a sequence in $]0,2]$. Suppose that $\operatorname{int}(C\cap\operatorname{Fix} T) \neq \varnothing$ and that $\sum_{n\in\mathbb{N}} \eta_n r_n = +\infty$. Then $(x_n)_{n=1}^{\infty}$ lies eventually in $C \cap \operatorname{Fix} T$.*

Proof. We argue by contradiction. If the conclusion is false, then *no* term of the sequence in $(x_n)_{n=1}^{\infty}$ lies in $\operatorname{Fix} T$, i.e., $(x_n)_{n=1}^{\infty}$ lies in $\mathcal{H} \smallsetminus \operatorname{Fix} T$. By assumption, there exist $z \in C \cap \operatorname{Fix} T$ and $r \in \mathbb{R}_{++}$ and such that $\operatorname{ball}(z;2r) \subseteq C \cap \operatorname{Fix} T$. Hence

$$\big(\forall y \in \operatorname{ball}(z;r)\big) \quad \operatorname{ball}(y;r) \subseteq C \cap \operatorname{Fix} T. \tag{3.39}$$

Since $r_n \to 0$, there exists $m \in$ such that $n \geq m$ implies $r_n \leq r$. Now let $n \geq m$ and $y \in \operatorname{ball}(z;r)$. Using the assumption that Q_C is a quasi projector of C, that $y \in C$, (3.39) and Corollary 3.5 (v), we obtain

$$\begin{aligned} \|x_{n+1} - y\| &= \big\|Q_C\big(x_n + \eta_n(U_{r_n}x_n - x_n)\big) - y\big\| && \text{(3.40a)} \\ &\leq \|x_n + \eta_n(U_{r_n}x_n - x_n) - y\| && \text{(3.40b)} \\ &\leq \|x_n - y\|. && \text{(3.40c)} \end{aligned}$$

Hence the sequence

$$(x_m, x_m + \eta_m(U_{r_m}x_m - x_m), x_{m+1}, x_{m+1} + \eta_{m+1}(U_{r_{m+1}}x_{m+1} - x_{m+1}), \ldots) \tag{3.41}$$

is Fejér monotone with respect to $\operatorname{ball}(z;r)$. It follows from Fact 3.12 and Corollary 3.5 (iii) that

$$+\infty > \sum_{n\geq m} \eta_n\|x_n - U_{r_n}x_n\| = \sum_{n\geq m} \eta_n\big(r_n + \|x_n - Tx_n\|\big) \geq \sum_{n\geq m} \eta_n r_n, \tag{3.42}$$

which is absurd because $\sum_{n\in\mathbb{N}} \eta_n r_n = +\infty$. □

Compared to Theorem 3.14, the following theorem has a less restrictive assumption on $(\operatorname{Fix} T, C)$ but a more restrictive one on the parameters (r_n, η_n).

Theorem 3.15. *[13, Theorem 3.2] Assume that $(r_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{R}_{++}$ such that $r_n \to 0$ and $(\eta_n)_{n\in\mathbb{N}}$ is a sequence in $]0,2]$. Suppose that $C \cap \operatorname{int}\operatorname{Fix} T \neq \varnothing$ and that $\sum_{n\in\mathbb{N}} \eta_n(2-\eta_n)r_n^2 = +\infty$. Then $(x_n)_{n=1}^{\infty}$ lies eventually in $C \cap \operatorname{Fix} T$.*

Proof. Similarly to the proof of Theorem 3.14, we argue by contradiction and assume the conclusion is false. Then $(x_n)_{n=1}^{\infty}$ must lie in $\mathcal{H} \smallsetminus \operatorname{Fix} T$. By assumption, there exist $y \in \operatorname{Fix} T$ and $r \in \mathbb{R}_{++}$ such that $\operatorname{ball}(y; r) \subseteq \operatorname{Fix} T$. Because $r_n \to 0$, there exists $m \in$ such that $n \geq m$ implies $r_n \leq r$. Let $n \geq m$. Using also the assumption that Q_C is a quasi projector of C and Corollary 3.5 (v), we deduce that

$$\begin{aligned}
\|x_{n+1} - y\|^2 &= \big\|Q_C\big(x_n + \eta_n(U_{r_n}x_n - x_n)\big) - y\big\|^2 && \text{(3.43a)}\\
&\leq \|x_n + \eta_n(U_{r_n}x_n - x_n) - y\|^2 && \text{(3.43b)}\\
&\leq \|x_n - y\|^2 - \eta_n(2-\eta_n)\big(r_n + \|x_n - Tx_n\|\big)^2 && \text{(3.43c)}\\
&\leq \|x_n - y\|^2 - \eta_n(2-\eta_n)r_n^2. && \text{(3.43d)}
\end{aligned}$$

This implies

$$\|x_m - y\|^2 \geq \sum_{n\geq m}\big(\|x_n - y\|^2 - \|x_{n+1} - y\|^2\big) \geq \sum_{n\geq m} \eta_n(2-\eta_n)r_n^2 = +\infty, \tag{3.44}$$

which contradicts our assumption on the parameters. □

Theorem 3.14 and Theorem 3.15 have various applications. Since every resolvent of a maximally monotone operator (see [7, Definition 20.20]) is firmly nonexpansive (see [7, Definition 4.1 (i)]) and hence a cutter, we obtain the following result.

Corollary 3.16. *[13, Corollary 3.1] Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, suppose that $Q_C = P_C$, that $T = (\operatorname{Id} + A)^{-1}$, and that one of the following holds:*

(i) $\operatorname{int}(C \cap A^{-1}0) \neq \varnothing$ *and* $\sum_{n\in\mathbb{N}} \eta_n r_n = +\infty$.

(ii) $C \cap \operatorname{int} A^{-1}0 \neq \varnothing$ *and* $\sum_{n\in\mathbb{N}} \eta_n(2-\eta_n)r_n^2 = +\infty$.

Then $(x_n)_{n=1}^{\infty}$ lies eventually in $C \cap A^{-1}0$.

Proof. Let A be maximally monotone, $T = (\mathrm{Id} + A)^{-1}$, By Theorem 2.2.5 in [20], T is firmly nonexpansive, then T is a cutter. $\forall x \in \operatorname{Fix} T$, then $Tx = (\mathrm{Id} + A)^{-1}x = x$, so $(\mathrm{Id} + A)x = x$, we have $Ax = 0$, $x \in A^{-1}0$. Therefore $\operatorname{Fix} T \subseteq A^{-1}0$. Similary proof for inverse. Then the conlusion follows by Theorem 3.14 and Theorem 3.15. □

Corollary 3.16 applies in particular to finding a constrained critical point of a convex function. When specializing further to a normal cone operator, we obtain the following result.

Example 3.17. (Convex feasibility) [13, Example 3.1] Let D be a nonempty, closed and convex subset of $\mathcal{H}$, and suppose that $Q_C = P_C$, that $T = P_D$, and that one of the following holds:

(i) $\operatorname{int}(C \cap D) \neq \varnothing$ and $\sum_{n\in\mathbb{N}} r_n = +\infty$.

(ii) $C \cap \operatorname{int} D \neq \varnothing$ and $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$.

Then the sequence $(x_n)_{n=1}^{\infty}$ generated by

$$x_{n+1} := P_C\left(P_D x_n + r_n \frac{P_D x_n - x_n}{\|P_D x_n - x_n\|}\right) \tag{3.45}$$

if $x_n \notin D$ and $x_{n+1} := x_n$ if $x_n \in D$, lies eventually in $C \cap D$.

Remark 3.18. (Relationship to Polyak's work) [13, Remark 3.1] In [50], B.T. Polyak considers random algorithms for solving constrained systems of convex inequalities. Suppose that only one consistent constrained convex inequality is considered. Hence the cutters used are all subgradient projectors. Then his algorithm coincides with the one considered in this section and thus is comparable. We note that our Theorem 3.14 is more flexible because Polyak requires $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$ (see [50, Theorem 1 and Section 4.2]) provided that $0 < \inf_{n\in\mathbb{N}} \eta_n \leq \sup_{n\in\mathbb{N}} \eta_n < 2$ while we require only $\sum_{n\in\mathbb{N}} r_n = +\infty$ in this case. Regarding our Theorem 3.15, we note that our proof essentially follows his proof which actually works for cutters — not just subgradient projectors — and under a less restrictive constraint qualification.

Remark 3.19. (Relationship to Crombez's work) [13, Remark 3.2] In [31], G. Crombez considers asynchronous parallel algorithms for finding a point in the intersection of the fixed point sets of finitely many cutters — without the constraint set C. Again, we consider the case when we are dealing with only one cutter. Then Crombez's convergence result (see [31, Theorem 2.7]) is

similar to Theorem 3.15; however, he requires that the radius r of some ball contained in $\operatorname{Fix} T$ be known which may not always be realistic in practical applications.

Actually, it is not too difficult to extend Theorem 3.14 and Theorem 3.15 to deal with finitely many cutter. We have opted here for simplicity rather than maximal generality.

3.4 Limiting Examples

In this section, we collect several examples that illustrate the boundaries of the theory. We start by showing that the conclusion of Theorem 3.14 and Theorem 3.15 both may fail to hold if the divergent-series condition is not satisfied.

Example 3.20. (Divergent-series condition is important) Suppose that $\mathcal{H} = C = \mathbb{R}$, that $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2 - 1$, and that $T = G_f$ is the subgradient projector associated with f. Suppose that $x_0 > 1$, set $r_{-1} := x_0 - 1 > 0$ and $(\forall n \in \mathbb{N})\ r_n := r_{n-1}^2/(4(1 + r_{n-1}))$. Then $(r_n)_{n\in\mathbb{N}}$ lies in $\mathbb{R}_{++}$, $r_n \to 0$, and $\sum_{n\in\mathbb{N}} r_n < +\infty$ and hence $\sum_{n\in\mathbb{N}} r_n^2 < +\infty$. However, the sequence $(x_n)_{n=1}^{\infty}$ generated by (3.38) lies in $]1, +\infty[$ and hence does not converge finitely to a point in $\operatorname{Fix} T = [-1, 1]$. Furthermore, the classical subgradient projector iteration $(\forall n \in \mathbb{N})\ y_{n+1} = Ty_n$ converges to some point in $\operatorname{Fix} T$, but not finitely when $y_0 \notin \operatorname{Fix} T$.

Proof. It is clear that $\operatorname{Fix} T = [-1, 1]$. Observe that $(\forall n \in \mathbb{N})\ 0 < r_n \leq (1/4)r_{n-1} \leq (1/4)^{n+1}r_{-1}$. It follows that $r_n \to 0$ and that $\sum_{n\in\mathbb{N}} r_n$ and $\sum_{n\in\mathbb{N}} r_n^2$ are both convergent series. Now suppose that $r_{n-1} = x_n - 1 > 0$ for some $n \in \mathbb{N}$. It then follows from Example 3.4 that

$$x_{n+1} = \frac{x_n}{2} + \frac{1}{2x_n} - r_n = \frac{(x_n - 1)^2}{2x_n} + 1 - r_n = \frac{r_{n-1}^2}{2(1 + r_{n-1})} + 1 - r_n = r_n + 1. \tag{3.46}$$

Hence, by induction, $(\forall n \in \mathbb{N})\ x_n = 1 + r_{n-1}$ and therefore $x_n \to 1^+$.

As for the sequence $(y_n)_{n\in\mathbb{N}}$, it is follows from Polyak's seminal work (see [48]) that $(y_n)_{n\in\mathbb{N}}$ converges to some point in $\operatorname{Fix} T$. However, by e.g. [12, Proposition 9.9], $(y_n)_{n\in\mathbb{N}}$ lies outside $\operatorname{Fix} T$ whenever y_0 does. $\square$

The next example illustrates that in the context of Theorem 3.14 and Theorem 3.15, we cannot expect finite convergence if the interior of $\operatorname{Fix} T$ is empty.

Example 3.21. (Nonempty-interior condition is important) Suppose that $\mathcal{H} = C = \mathbb{R}$, that $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$, and that $T = G_f$ is the subgradient projector associated with f. Then $\operatorname{Fix} T = \{0\}$ and hence $\operatorname{int} \operatorname{Fix} T = \varnothing$. Set $x_0 := 1/2$, and set $(\forall n \in \mathbb{N})\ w_n := (n+1)^{-1/2}$ and $r_n = w_n$ if $U_{w_n} x_n \neq 0$ and $r_n = 2w_n$ if $U_{w_n} x_n = 0$. Then $r_n \to 0$ and $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$. The sequence $(x_n)_{n=1}^{\infty}$ generated by (3.38) converges to 0 but not finitely.

Proof. The statements concerning $(r_n)_{n\in\mathbb{N}}$ are clear. It follows readily from the definition that $(\forall x \in \mathbb{R})(\forall r \in \mathbb{R}_+)\ Tx = x/2$ and $U_r x = x/2 - r \operatorname{sgn}(x)$. Since $x_0 = 1/2$, $w_0 = 1$, $U_1 x_0 = -3/4 \neq 0$, and $r_0 = w_0 = 1$, it follows that $0 < |x_0/2| < r_0$. We now show that for every $n \in \mathbb{N}$,

$$0 < |x_n/2| < r_n. \tag{3.47}$$

This is clear for $n = 0$. Now assume (3.47) holds for some $n \in \mathbb{N}$.

Case 1: $|x_n| = 2w_n$.
Then $U_{w_n} x_n = x_n/2 - \operatorname{sgn}(x_n) w_n = 0$. Hence $r_n = 2w_n$ and thus $x_{n+1} = U_{r_n} x_n = x_n/2 - 2w_n \operatorname{sgn}(x_n) = \operatorname{sgn}(x_n) w_n - 2w_n \operatorname{sgn}(x_n) = -\operatorname{sgn}(x_n) w_n$. Thus $0 < |x_{n+1}/2| = w_n/2 = 1/(2\sqrt{n+1}) < 1/\sqrt{n+2} = w_{n+1} \leq r_{n+1}$, which yields (3.47) with n replaced by $n+1$.

Case 2: $|x_n| \neq 2w_n$.
Then $U_{w_n} x_n = x_n/2 - \operatorname{sgn}(x_n) w_n \neq 0$. Hence $r_n = w_n$ and thus $x_{n+1} = U_{r_n} x_n = x_n/2 - r_n \operatorname{sgn}(x_n)$. It follows that $|x_{n+1}| = r_n - |x_n/2| > 0$. Hence $0 < |x_{n+1}/2|$ and also $|x_{n+1}| < r_n = w_n < 2w_{n+1} \leq 2r_{n+1}$. Again, this is (3.47) with n replaced by $n+1$.

It follow now by induction that (3.47) holds for every $n \in \mathbb{N}$. □

We now illustrate that when $\operatorname{Fix} T = \varnothing$, then $(x_n)_{n=1}^{\infty}$ may fail to converge.

Example 3.22. Suppose that $\mathcal{H} = C = \mathbb{R}$, that $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2 + 1$, and that $T = G_f$ is the subgradient projector associated with f. Let $y_0 \in \mathbb{R}$ and suppose that $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$. Then $(x_n)_{n\in\mathbb{N}}$ is either not well defined or it diverges. Suppose that $x_0 > 1/\sqrt{3}$, set $k_0 = x_0 - 1/\sqrt{3} > 0$ and $(\forall n \in \mathbb{N})\ k_{n+1} = \sqrt{(n+1)/(n+2)}\, k_n$. Suppose that

$$(\forall n \in \mathbb{N}) \quad r_n = \frac{1}{2}\left(\sqrt{3} + 2k_{n+1} + k_n - \frac{1}{k_n + 1/\sqrt{3}}\right). \tag{3.48}$$

Then $r_n \to 0^+$ and $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$. Moreover, the sequence $(x_n)_{n=1}^{\infty}$ generated by (3.38) diverges.

Proof. Clearly, $\operatorname{Fix} T = \varnothing$ and one checks that

$$(\forall r \in \mathbb{R}_+)(\forall x \in \mathbb{R} \smallsetminus \{0\}) \quad U_r x = \frac{x}{2} - \frac{1}{2x} - r \operatorname{sgn}(x). \tag{3.49}$$

If some $x_n = 0$, then the sequence $(x_n)_{n\in\mathbb{N}}$ is not well defined.

Case 1: $(\exists\, n \in \mathbb{N})\ x_n = 1/\sqrt{3}$.
Then $x_{n+1} = Tx_n = U_0 x_n = x_n/2 - 1/(2x_n) = -1/\sqrt{3} = -x_n$ and similarly $x_{n+2} = -x_{n+1} = x_n$. Hence the sequence eventually oscillates between $1/\sqrt{3}$ and $-1/\sqrt{3}$.

Case 2: $(\exists\, n \in \mathbb{N})\ |x_n| = 1$.
Then $x_{n+1} = 0$ and the sequence is not well defined.

Case 3: $(\forall n \in \mathbb{N})\ |x_n| \notin \{1, 1/\sqrt{3}\}$.
Using the Arithmetic Mean–Geometric Mean inequality, we obtain

$$|x_{n+1} - x_n| = \left|\frac{x_n}{2} - \frac{1}{2x_n} - x_n\right| = \frac{1}{2}\left|x_n + \frac{1}{x_n}\right| = \frac{1}{2}\left(|x_n| + \frac{1}{|x_n|}\right) \geq 1 \tag{3.50}$$

for every $n \in \mathbb{N}$. Therefore, $(x_n)_{n\in\mathbb{N}}$ is divergent or not well defined.

We now turn to the sequence $(x_n)_{n=1}^{\infty}$. Observe that

$$0 < k_n = \sqrt{n/(n+1)}\,k_{n-1} = \cdots = k_0/\sqrt{n+1} \to 0^+ \tag{3.51}$$

and hence $(k_n)_{n\in\mathbb{N}}$ is strictly decreasing. It follows that $r_n \to 0^+$ and that $r_n > (2k_{n+1} + k_n)/2 > 3k_{n+1}/2 = 3k_0/(2\sqrt{n+2})$. Thus, $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$. Next, (3.49) yields

$$x_1 = \frac{x_0}{2} - \frac{1}{2x_0} - r_0 \tag{3.52a}$$

$$= \frac{k_0 + 1/\sqrt{3}}{2} - \frac{1}{2\big(k_0 + 1/\sqrt{3}\big)} - \frac{1}{2}\left(\sqrt{3} + 2k_1 + k_0 - \frac{1}{k_0 + 1/\sqrt{3}}\right) \tag{3.52b}$$

$$= -\frac{1}{\sqrt{3}} - k_1. \tag{3.52c}$$

Hence $x_1 < 0$ and we then see analogously that $x_2 = 1/\sqrt{3} + k_2 > 0$. We inductively obtain

$$(\forall n \in \mathbb{N}) \quad 0 < x_{2n} = \frac{1}{\sqrt{3}} + k_{2n} \ \text{ and } \ 0 > x_{2n+1} = -\frac{1}{\sqrt{3}} - k_{2n+1}. \tag{3.53}$$

It follows that $(-1)^n x_n \to 1/\sqrt{3}$; therefore, $(x_n)_{n=1}^{\infty}$ is divergent. $\square$

3.5 Comparison

We assume that $f\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a level set $\{x \in \mathcal{H} : f(x) \leq 0\} \neq \varnothing$ and that

$$T = G_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \begin{cases} x - \dfrac{f(x)}{\|\nabla f(x)\|^2}\nabla f(x), & \text{if } f(x) > 0, \\ x, & \text{otherwise} \end{cases} \tag{3.54}$$

is the associated subgradient projector. Then (3.2) turns into

$$U_r x = \begin{cases} x - \dfrac{f(x) + r\|\nabla f(x)\|}{\|\nabla f(x)\|^2}\nabla f(x), & \text{if } f(x) > 0, \\ x, & \text{otherwise} \end{cases} \tag{3.55}$$

and (3.38) into

$$x_{n+1} = \begin{cases} Q_C\Bigg(x_n - \eta_n \dfrac{f(x_n) + r_n\|\nabla f(x_n)\|}{\|\nabla f(x_n)\|^2}\nabla f(x_n)\Bigg), & \text{if } f(x_n) > 0, \\ x_n, & \text{otherwise.} \end{cases} \tag{3.56}$$

In the algorithmic setting, Polyak uses $\eta_n \equiv \eta \in \,]0, 2[\,$, (e.g. $\eta = 1.8$; see [50, Section 4.3]). In the present setting, his framework requires $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$. When $C = \mathcal{H}$, one also has the following similar yet different update formula

$$y_{n+1} = \begin{cases} y_n - \eta_n \dfrac{f(y_n) + \varepsilon_n}{\|\nabla f(y_n)\|^2}\nabla f(y_n), & \text{if } f(y_n) > 0, \\ y_n, & \text{otherwise,} \end{cases} \tag{3.57}$$

where $0 < \inf_{n\in\mathbb{N}} \eta_n \leq \sup_{n\in\mathbb{N}} \eta_n < 2$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ is a strictly decreasing sequence in $\mathbb{R}_{++}$ with $\sum_{n\in\mathbb{N}} \varepsilon_n = +\infty$. In this setting, this is also known as the *Modified Cyclic Subgradient Projection Algorithm (MCSPA)*, which finds its historical roots in works by Fukushima [34], by De Pierro and Iusem [47], and by Censor and Lent [22]. Note that MCSPA requires the existence of a *Slater point*, i.e., $\inf f(\mathcal{H}) < 0$, which is more restrictive than our assumptions (consider, e.g., the squared distance to the unit ball). Let us now link the assumption on the parameters of the MCSPA (3.57) to (3.56).

Proposition 3.23. *[13, Proposition 5.1] Suppose that $\mathcal{H} = C$ is finite-dimensional, $\inf f(\mathcal{H}) < 0$, $\eta_n \equiv 1$, $\sum_{n\in\mathbb{N}} r_n = +\infty$ and $(\forall n \in \mathbb{N})$ $\varepsilon_n = r_n\|\nabla f(x_n)\| > 0$. Then $\varepsilon_n \to 0$ and $\sum_{n\in\mathbb{N}} \varepsilon_n = +\infty$.*

Proof. Corollary 3.5 (iv) implies that $(x_n)_{n=1}^{\infty}$ is bounded. Because ∇f is continuous, we obtain that $\sigma := \sup_{n\in\mathbb{N}} \|\nabla f(x_n)\| < +\infty$. By Lemma 3.13, there exists $\alpha \in \mathbb{R}_{++}$ such that if $f(x_n) > 0$, then $\|\nabla f(x_n)\| \geq \alpha$. Hence

$$(\forall n \in \mathbb{N}) \quad f(x_n) > 0 \quad \Rightarrow \quad 0 < \alpha r_n \leq \|\nabla f(x_n)\| r_n = \varepsilon_n \leq \sigma r_n, \tag{3.58}$$

and therefore $\sum_{n\in\mathbb{N}} \varepsilon_n = +\infty$. □

The following example shows that our assumptions are independent of those of the MCSPA.

Example 3.24. [13, Example 5.1] Suppose that $\mathcal{H} = C = \mathbb{R}$, that $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2 - 1$, that $r_n = (n+1)^{-1}$ if n is even and $r_n = n^{-1/2}$ if n is odd, and that $\eta_n \equiv 1$. Clearly, $r_n \to 0$ and $\sum_{n\in\mathbb{N}} r_n^2 = +\infty$. However, $(\varepsilon_n)_{n\in\mathbb{N}} := (r_n|f'(x_n)|)_{n\in\mathbb{N}}$ is not strictly decreasing.

Proof. The sequence $(x_n)_{n=1}^{\infty}$ is bounded. Suppose that $f(x_n) > 0$ for some $n \in \mathbb{N}$. By Example 3.4,

$$x_{n+1} = U_{r_n} x_n = \frac{x_n}{2} + \frac{1}{2x_n} - r_n \operatorname{sgn}(x_n). \tag{3.59}$$

Assume that n is even, say $n = 2m$, where $m \geq 2$, and that $1 < x_{2m} < (2m+1)/2$. Then $x_{2m} > 2x_{2m}/\sqrt{2m+1}$ and

$$\varepsilon_{2m} = r_{2m}|f'(x_{2m})| = 2r_{2m}x_{2m} = \frac{2x_{2m}}{2m+1}. \tag{3.60}$$

Hence, using (3.59),

$$x_{2m+1} = \frac{x_{2m}}{2} + \frac{1}{2x_{2m}} - r_{2m} > \frac{x_{2m}}{2} + \frac{1}{2m+1} - \frac{1}{2m+1} = \frac{x_{2m}}{2}, \tag{3.61}$$

and therefore

$$2x_{2m+1} > x_{2m} > \frac{2x_{2m}}{\sqrt{2m+1}}. \tag{3.62}$$

Thus $\varepsilon_{2m+1} = r_{2m+1}|f'(x_{2m+1})| = 2r_{2m+1}x_{2m+1}$. It follows that

$$\varepsilon_{2m+1} = \frac{2x_{2m+1}}{\sqrt{2m+1}} > \frac{2x_{2m}}{2m+1} = \varepsilon_{2m} \tag{3.63}$$

and the proof is complete. □

We conclude this chapter with a simple experiment.

Remark 3.25. (Numerical experiment) We compare the performance of our algorithm to Censor's modified cyclic subgradient projection algorithm (MCSPA), which is another popular method and which has similar update formulas. Suppose that $\mathcal{H} = C = \mathbb{R}$ and that $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto x^2 - 1$. Let T be the subgradient projector associated with f.

Then our algorithm (3.38) is reduced to

$$x_{n+1} = \begin{cases} x_n - \eta_n \frac{f(x_n)+r_n\|\nabla f(x_n)\|}{\|\nabla f(x_n)\|^2}\nabla f(x_n), & \text{if } f(x_n) > 0, \\ x_n, & \text{otherwise.} \end{cases} \tag{3.64}$$

Let $(\varepsilon_n)_{n=1}^{\infty} \subseteq \mathbb{R}_+$ be a monotonically decreasing sequence such that $\varepsilon_n \to 0$ and $\sum_{n=1}^{\infty} \varepsilon_n = +\infty$. Censor's MCSPA in [21] is defined by

$$x_{n+1} = \begin{cases} x_n - \eta_n \frac{f(x_n)+\varepsilon_n}{\|\nabla f(x_n)\|^2}\nabla f(x_n), & \text{if } f(x_n) > 0, \\ x_n, & \text{otherwise.} \end{cases} \tag{3.65}$$

We randomly choose 100 starting points in the interval $[1, 10^6]$. In the following table, we record the performance of (3.64) and (3.65) with different selections of parameters. The choices of (r_n, η_n) are for (3.64), and ε_n is for MCSPA. Mean and median refer to the number of iterations until the current iterate is 10^{-6} feasible. See Table 3.1.

Table 3.1: Performance of the algorithm for $f(x) = x^2 - 1$.

Algorithm for $x^2 - 1$	Mean	Median
$(r_n, \eta_n) = \big(1/(n+1), 1\big)$	11.49	13
$(r_n, \eta_n) = \big(1/(n+1), 2\big)$	2	2
$(r_n, \eta_n) = \big(1/\sqrt{n+1}, 1\big)$	10.83	12
$(r_n, \eta_n) = \big(1/\sqrt{n+1}, 2\big)$	2	2
$\varepsilon_n = 1/(n+1)$	11.81	13
$\varepsilon_n = 1/\sqrt{n+1}$	12.19	13

Now let us instead consider $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto 100x^2 - 1$. The corresponding data are in the Table 3.2.

Table 3.2: Performance of the algorithm for $f(x) = 100x^2 - 1$.

Algorithm for $100x^2 - 1$	Mean	Median
$(r_n, \eta_n) = \big(1/(n+1), 1\big)$	13.29	14
$(r_n, \eta_n) = \big(1/(n+1), 2\big)$	12	12
$(r_n, \eta_n) = \big(1/\sqrt{n+1}, 1\big)$	17.52	19
$(r_n, \eta_n) = \big(1/\sqrt{n+1}, 2\big)$	105	105
$\varepsilon_n = 1/(n+1)$	15.27	16
$\varepsilon_n = 1/\sqrt{n+1}$	15.76	17

We observe that both the step lengths r_n and ε_n and the relaxation parameter η_n have significant impact on the performance of the algorithms.

This chapter gave a new method for finding a fixed point of a cutter, motivated by Polyak and Crombez's works. More precisely, our assumptions on the parameters are more general than existing results. Furthermore, we provided limiting examples to illustrate that our assumptions can not be weakened.

Chapter 4

On Subgradient Projectors

4.1 Overview

This chapter is based on the paper [12]. Most results in this chapter are new, and some are complementary to work of B. Pauwels [44]. The results in this chapter are central for the understanding of the subgradient projector. Fundamental properties such as continuity, nonexpansiveness and monotonicity are investigated. We also point out that the Yamagishi-Yamada operator is a subgradient projector of another convex function. We start with the definition of our key notion: subgradient projector.

Throughout this chapter, we assume that $f\colon \mathcal{H} \to \mathbb{R}$ is convex and continuous, and $C := \{x \in \mathcal{H} : \ f(x) \leq 0\} \neq \varnothing$.

Unless stated otherwise, we assume that $s\colon \mathcal{H} \to \mathcal{H}$ is a selection of ∂f, i.e., $(\forall x \in \mathcal{H})\ s(x) \in \partial f(x)$. Then the associated subgradient projector is defined by

$$(\forall x \in X) \quad G_f(x) = \begin{cases} x - \dfrac{f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > 0; \\ x, & \text{otherwise.} \end{cases} \tag{4.1}$$

The subgradient projector is the key ingredient in Polyak's seminal work [48] on subgradient projection algorithms, which have since found many applications; see, e.g., [2] (which deals with subgradient algorithms viewed as projection algorithms with variable supersets), [6] (which provides a framework for transforming weakly into strongly convergent cutter methods), [20] (which provides a very nice and recent monograph on cutters, subgradient projectors and related algorithms), [22] (which introduces the cyclic subgradient projections method for solving systems of convex inequalities), [23] (which deals with an almost cyclic sequential algorithm for solving a common fixed point problem of cutters), [24] (which is a book that not only features subgradient projection algorithms but it also won their authors the *INFORMS Computing Society prize*), [26] (which presents a general framework for algorithmically solving feasibility problems especially in signal processing), [28] (which is a broad survey on subgradient projection algorithms

for solving image recovery problems), [29] (which develops the mathematical theory of the general and fast parallel projection method for image recovery based on subgradient projectors), [30] (which presents an adaptive level set method for constrained image recovery relying upon subgradient projectors), [43] (which introduces a cut that is tighter than the subgradient projector cut when applied to certain quadratic functions), [49] (which is a seminal book that includes subgradient methods for the constrained minimization of nonsmooth functions), [50] (which employs randomly chosen subgradient projectors to solve convex inequalities), [55] (which provides an adaptive projected subgradient method with applications to machine learning), [56] (which is not only a review on subgradient projection algorithms for adaptive learning but it also won the *IEEE Signal Processing Society Signal Processing Magazine Best Paper Award*), [58] (which presents a subgradient-projector based method for solving robust adaptive signal processing problems), [59] (which features a very general hybrid steepest descent for solving variational inequalities with applications to convex optimization via subgradient projectors), [60] (which introduces an adaptive filtering algorithm based on parallel subgradient projection methods), and the references therein. This impressive body of (including even award-winning) research, which has also attracted a significant number of citations and which clearly brought out the importance of the subgradient projector as an algorithmic building block, warrants a mathematical study of the subgradient projector and its properties.

4.2 Basic Properties

Let us record some basic results on subgradient projectors, which are essentially contained already in [48] and the proofs of which we provide for completeness.

Fact 4.1. *Let $x \in \mathcal{H}$, and set*

$$H = \{y \in \mathcal{H} : \langle s(x), y - x\rangle + f(x) \leq 0\}. \tag{4.2}$$

Then the following hold:

(i) $f^+(x) + \langle s(x), G_f(x) - x\rangle = 0$, *where* $f^+(x) = \max\{f(x), 0\}$.

(ii) $\operatorname{Fix} G_f = C \subseteq H$.

(iii) $G_f(x) = P_H x$.

(iv) (G_f is a cutter) $(\forall c \in C)\ \langle c - G_f(x), x - G_f(x)\rangle \leq 0$.

(v) $(\forall c \in C)\ \|x - G_f(x)\|^2 + \|G_f(x) - c\|^2 \leq \|x - c\|^2$.

(vi) $f^+(x) = \|s(x)\|\|x - G_f(x)\|$.

(vii) *If* $x \notin C$, *then* $(\forall c \in C)\ \frac{f^2(x)}{\|s(x)\|^2} + \|G_f(x) - c\|^2 \leq \|x - c\|^2$.

(viii) $f^+(x)(x - G_f(x)) = \|x - G_f(x)\|^2 s(x)$.

(ix) *Suppose that* f *is Fréchet differentiable at* $x \in \mathcal{H} \smallsetminus C$. *Then* $g = \ln \circ f \colon \mathcal{H} \smallsetminus C \to \mathbb{R}$ *is Fréchet differentiable at* x *and* $G_f(x) = x - \nabla g(x)/\|\nabla g(x)\|^2$.

(x) *Suppose that* $\min f(\mathcal{H}) = 0$ *and* f *is Fréchet differentiable on* $\mathcal{H}$ *with* ∇f *being Lipschitz continuous*[9] *with constant* L, *that* $x \notin C$, *and that there exists* $\alpha > 0$ *such that* $f(x) \geq \alpha d_C^2(x)$. *Then* $d_C^2(G_f(x)) \leq (1 - \alpha^2/L^2) d_C^2(x)$.

(xi) *Suppose that* $\min f(\mathcal{H}) = 0$, *that* $x \notin C$, *and that there exists* $\alpha > 0$ *such that* $f(x) \geq \alpha d_C(x)$. *Then* $d_C^2(G_f(x)) \leq (1 - \alpha^2/\|s(x)\|^2) d_C^2(x)$.

Proof. (i): This follows directly from the definition of G_f. Indeed, suppose $x \notin C$, we have $f(x) > 0$, then $f^+(x) = f(x)$.

$$\langle s(x), G_f(x) - x\rangle = \langle s(x), -\frac{f(x)}{\|s(x)\|^2} s(x)\rangle = -f(x), \tag{4.3}$$

therefore $f^+(x) + \langle s(x), G_f(x) - x\rangle = 0$. When $x \in C$, we have $G_f(x) = x$ and $f(x) \leq 0$, then $f^+(x) = 0$, so $f^+(x) + \langle s(x), G_f(x) - x\rangle = 0$.

(ii): The equality is clear from the definition of G_f. Let $z \in C$. By convexity of function or definition of subgradient, we have $\langle s(x), z - x\rangle + f(x) \leq f(z) \leq 0$ and hence $z \in H$.

(iii): Assume first that $x \in C$. Then $x \in \operatorname{Fix} G_f \subseteq H$ by (ii) and hence $G_f(x) = x = P_H x$. Now assume that $x \notin H$, then $x \notin C$. Then

[9] We say an operator $T : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous with a constant L if $\forall x, y \in \mathcal{H}$,

$$\|Tx - Ty\| \leq L\|x - y\|.$$

$0 < f(x) = f^+(x)$ and $s(x) \neq 0$. Using the formula of projection onto a half-space (see [7, Example 28.16]) we have

$$P_H x = x - \frac{\langle s(x), x\rangle - \big(\langle s(x), x\rangle - f(x)\big)}{\|s(x)\|^2} s(x) = x - \frac{f(x)}{\|s(x)\|^2} s(x) = G_f(x). \tag{4.4}$$

(iv): In view of (iii) and Fact 2.40, we have $(\forall h \in H)$ $\langle h - P_H x, x - P_H x\rangle = \langle h - G_f(x), x - G_f(x)\rangle \leq 0$. Now invoke (ii), for $\forall c \in C \subseteq H$, we have $\langle c - G_f(x), x - G_f(x)\rangle \leq 0$.

(v): This is equivalent to (iv) by definition of cutter.

(vi): Assume first that $x \in C$. Then $f(x) \leq 0$, i.e., $f^+(x) = 0$, and $x = G_f(x)$ by (ii). Hence the identity is true. Now assume that $x \notin C$. Then $0 < f(x) = f^+(x)$ and $x - G_f(x) = \frac{f(x)}{\|s(x)\|^2} s(x)$. Taking the norm, we learn that $\|x - G_f(x)\| = \frac{f(x)}{\|s(x)\|} = \frac{f^+(x)}{\|s(x)\|}$.

(vii): Combine (ii), (v), and (vi). If $x \notin C$, then $\|x - G_f(x)\| = \frac{f^+(x)}{\|s(x)\|} = \frac{f(x)}{\|s(x)\|}$. Substitute $\|x - G_f(x)\|$ into (v).

(viii): This follows from (vi) and the definition of G_f. When $x \in C$, we have $x = G_f(x)$, then equation in (viii) holds. When $x \notin C$, we have $f^+(x) = f(x) > 0$ and $\|x - G_f(x)\| = \frac{f^+(x)}{\|s(x)\|}$. By definition of G_f we have

$$f^+(x)(x - G_f(x)) = f^+(x) \frac{f^+(x)}{\|s(x)\|^2} s(x) = \|x - G_f(x)\|^2 s(x). \tag{4.5}$$

(ix): The chain rule implies that $\nabla g(x) = \frac{\nabla f(x)}{f(x)}$. Hence $\|\nabla g(x)\|^2 = \frac{\|\nabla f(x)\|^2}{f^2(x)}$ and thus $x - \frac{\nabla g(x)}{\|\nabla g(x)\|^2} = x - \frac{f(x)}{\|\nabla f(x)\|^2} \nabla f(x) = G_f(x)$.

(x): Let $c \in C$. Then $\nabla f(c) = 0$ and hence $\|\nabla f(x)\| = \|\nabla f(x) - \nabla f(c)\| \leq L\|x - c\|$. Hence $\|\nabla f(x)\| \leq L d_C(x)$ and therefore, using (vii), we obtain

$$\|G_f(x) - c\|^2 \leq \|x - c\|^2 - \frac{f^2(x)}{\|\nabla f(x)\|^2} \leq \|x - c\|^2 - \frac{\alpha^2 d_C^4(x)}{L^2 d_C^2(x)}. \tag{4.6}$$

Now take the minimum over $c \in C$.

(xi): Using (vii), we have

$$d_C^2(G_f(x)) \leq \|G_f(x) - P_C x\|^2 \leq \|x - P_C x\|^2 - \frac{f^2(x)}{\|s(x)\|^2} \leq d_C^2(x) - \frac{\alpha^2 d_C^2(x)}{\|s(x)\|^2}. \tag{4.7}$$

The proof is complete. □

Example 4.2. Suppose that $f(x) = \|x\|$, $x \in \mathcal{H}$. Then $\nabla f(x) = \frac{x}{\|x\|}$ and $G_f(x) = 0$ when $x \neq 0$.

4.3 Calculus

We now turn to basic calculus rules. It is convenient to introduce the operator $\mathcal{G}\colon \mathcal{H} \rightrightarrows \mathcal{H}$, defined by

$$(\forall x \in \mathcal{H}) \quad \mathcal{G}x = \mathcal{G}_f x = \{G_s(x) : s \text{ is a selection of } \partial f\}, \tag{4.8}$$

where G_f is the operator occurring in (4.1) for a particular subgradient s. When f is Gâteaux differentiable outside C, then we will identify $\mathcal{G}$ with G.

Proposition 4.3 (Calculus)**.** *Let $\alpha > 0$, let $A\colon \mathcal{H} \to \mathcal{H}$ be continuous and linear such that $A^*A = AA^* = \mathrm{Id}$, and let $z \in \mathcal{H}$. Furthermore, let $(f_i)_{i\in I}$ be a finite family of convex continuous functions on $\mathcal{H}$ such that $\bigcap_{i\in I} C_{f_i} \neq \varnothing$. Then the following hold:*

(i) Suppose that $g = \alpha f$. Then $C_g = C_f$ and $\mathcal{G}_g = \mathcal{G}_f$.

(ii) Suppose that $g = f \circ \alpha\,\mathrm{Id}$. Then $C_g = \alpha^{-1} C_f$ and $\mathcal{G}_g = \alpha^{-1}\mathcal{G}_f \circ \alpha\,\mathrm{Id}$.

(iii) Suppose that $f \geq 0$ and that $g = f^\alpha$ is convex. Then $C_g = C_f$ and $\mathcal{G}_g = (1 - \alpha^{-1})\,\mathrm{Id} + \alpha^{-1}\mathcal{G}_f$.

(iv) Suppose that $g = f \circ A$. Then $C_g = A^* C_f$ and $\mathcal{G}_g = A^* \circ \mathcal{G}_f \circ A$.

(v) Suppose that $g\colon x \mapsto f(x - z)$. Then $C_g = z + C_f$ and $\mathcal{G}_g\colon x \mapsto z + \mathcal{G}_f(x - z)$.

(vi) Suppose that $g = \max_{i\in I} f_i$. Then $C_g = \bigcap_{i\in I} C_{f_i}$ and if $g(x) > 0$ and $I(x) = \{i \in I : f_i(x) = g(x)\}$, then $\mathcal{G}_g(x) = \{x - \frac{g(x)}{\|x^*\|^2} x^* : x^* \in \operatorname{conv} \bigcup_{i\in I(x)} \partial f_i(x)\}$.

(vii) Suppose that $g = f^+$. Then $\mathcal{G}_g = \mathcal{G}_f$.

(viii) (Moreau envelope) Suppose that $\min f(\mathcal{H}) = 0$ *and that* $g = f\square(1/2)\|\cdot\|^2$ *is the Moreau envelope of* f*. Then* $C_g = C_f$ *and*

$$(\forall x \in \mathcal{H})\ G_g(x) = \begin{cases} x - \dfrac{g(x)}{\|x - \operatorname{Prox}_f x\|^2}(x - \operatorname{Prox}_f x), & \text{if } f(x) > 0; \\ x, & \text{if } f(x) = 0. \end{cases} \tag{4.9}$$

Proof. Let $x \in \mathcal{H}$. We shall only prove one inclusion for the subgradient projector as the remaining one is proved similarly.

(i): Since $\alpha > 0$, then $g(x) \leq 0 \Leftrightarrow f(x) \leq 0$, it follows that $C_g = C_f$. Suppose that $f(x) > 0$. For $s_f(x) \in \partial f(x)$, it is easy to have $\alpha s_f(x) \in \partial g(x)$, we obtain

$$G_f(x) = x - \frac{f(x)}{\|s_f(x)\|^2} s_f(x) = x - \frac{g(x)}{\|\alpha s_f(x)\|^2}(\alpha s_f(x)). \tag{4.10}$$

This implies $\mathcal{G}_f(x) \subseteq \mathcal{G}_g(x)$.

(ii): For $\forall x \in C_g$, then $g(x) = f(\alpha x) \leq 0$, we have $\alpha x \in C_f$. On the other side, for $x \in C_f$, then $f(x) = f(\alpha \frac{x}{\alpha}) = g(\frac{x}{\alpha}) \leq 0$, then $\frac{x}{\alpha} \in C_g$. So we have $C_g = C_f$. Suppose that $g(x) > 0$, i.e., $f(\alpha x) > 0$. Then

$$\alpha^{-1} G_f(\alpha x) = \alpha^{-1}(\alpha x - \frac{f(\alpha x)}{\|s_f(\alpha x)\|^2} s_f(\alpha x)) \tag{4.11}$$

$$= x - \alpha^{-1} \frac{f(\alpha x)}{\|s_f(\alpha x)\|^2} s_f(\alpha x) \tag{4.12}$$

$$= x - \frac{f(\alpha x)}{\|\alpha s_f(\alpha x)\|^2}(\alpha s_f(\alpha x)) \in \mathcal{G}_g(x). \tag{4.13}$$

Hence $\alpha^{-1}\mathcal{G}_f(\alpha x) \subseteq \mathcal{G}_g(x)$.

(iii): It is easy to check that $C_g = C_f$. Suppose that $g(x) > 0$. Then $f(x) > 0$ and

$$\alpha^{-1}(x - G_f(x)) = \alpha^{-1} \frac{f(x)}{\|s_f(x)\|^2} s_f(x) \tag{4.14}$$

$$= \frac{f^{\alpha}(x)}{\|\alpha f^{\alpha-1}(x) s_f(x)\|^2} \alpha f^{\alpha-1}(x) s_f(x) \in x - \mathcal{G}_g(x). \tag{4.15}$$

(iv): We have $x \in C_g \Leftrightarrow g(x) = f(Ax) \leq 0 \Leftrightarrow Ax \in C_f \Leftrightarrow A^*Ax \in A^*C_f$ $\Leftrightarrow x \in A^*C_f$. Suppose that $g(x) > 0$. Then $f(Ax) > 0$, for $s_f(Ax) \in \partial f(Ax)$, we have $A^*s_f(Ax) \in \partial g(x)$ and

$$A^*G_f(Ax) = A^*(Ax - \frac{f(Ax)}{\|s_f(Ax)\|^2}s_f(Ax)) \tag{4.16}$$

$$= x - \frac{f(Ax)}{\|A^*s_f(Ax)\|^2}A^*s_f(Ax) \in \mathcal{G}_g(x). \tag{4.17}$$

The last equation is using $\|Ax\| = \|A^*x\| = \|x\|$. Since

$$\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^*Ax, x\rangle = \langle x, x\rangle = \|x\|^2, \tag{4.18}$$

$$\|A^*x\|^2 = \langle A^*x, A^*x\rangle = \langle AA^*x, x\rangle = \langle x, x\rangle = \|x\|^2. \tag{4.19}$$

(v): We have $x \in C_g \Leftrightarrow g(x) = f(x-z) \leq 0 \Leftrightarrow x-z \in C_f \Leftrightarrow x \in z+C_f$. Suppose that $0 < g(x) = f(x-z)$. Then

$$z + G_f(x-z) = z + \Big(x - z - \frac{f(x-z)}{\|s_f(x-z)\|^2}s_f(x-z)\Big) \in \mathcal{G}_g(x). \tag{4.20}$$

(vi): This follows from the well known formula for the subdifferential of a maximum; see, e.g., [45, Proposition 3.38].

(vii): This follows from (vi) since $f^+ = \max\{0, f\}$.

(viii): When $x \in C_f$, then $f(x) \leq 0$.

$$g(x) = \inf_{y\in\mathcal{H}}\{f(y) + \frac{1}{2}\|x-y\|^2\} \leq f(x) + \frac{1}{2}\|x-x\|^2 = f(x) \leq 0.$$

Then $x \in C_g$. On the other side, let $x \in C_g$, $g(x) = \inf_{y\in\mathcal{H}}\{f(y) + \frac{1}{2}\|x-y\|^2\} \leq 0$. For $\frac{1}{2}\|x-y\|^2$, it attains its minimum at $y = x$, we have $\inf_{y\in\mathcal{H}}\{f(y) + \frac{1}{2}\|x-y\|^2\} = f(x) \leq 0$. So $x \in C_f$. When $g \geq 0$, $\nabla g = \mathrm{Id} - \mathrm{Prox}_f$ (see, e.g., [7, Proposition 12.29]), and $\operatorname{argmin} g = \operatorname{argmin} f$ (see, e.g., [7, Corollary 17.5]). □

4.4 Examples

In this section, we present several illustrative examples.

Example 4.4. Suppose that $f = \|\cdot\|^2$. Then $\nabla f = 2\,\mathrm{Id}$ and $G_f = \frac{1}{2}\,\mathrm{Id}$.

Example 4.5. Suppose that

$$(\forall x \in \mathcal{H}) \quad f(x) = \begin{cases} \frac{1}{2}\|x\|^2, & \text{if } x \in \operatorname{ball}(0;1); \\ \|x\| - \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{4.21}$$

Then $G_f = \frac{1}{2}P_{\operatorname{ball}(0;1)}$ and G_f is firmly nonexpansive.

Proof. Let $x \in \mathcal{H}$. By [9, Example 1.1.1], we have $f = \|\cdot\|\Box(1/2)\|\cdot\|^2$ is the Moreau envelope of the norm. Hence it follows from Proposition 4.3 (viii) that

$$G_f(x) = x - \frac{f(x)}{\|x - \operatorname{Prox}_{\|\cdot\|} x\|^2}(x - \operatorname{Prox}_{\|\cdot\|} x) \tag{4.22}$$

provided that $x \neq 0$, and $G_f(x) = 0 = \frac{1}{2}P_{\operatorname{ball}(0;1)}x$ if $x = 0$. Using [9, Example 1.1.1] and [7, Example 12.25], we have $\|\cdot\|^{*}$[10] $= \iota_{\operatorname{ball}(0;1)}$ and $\operatorname{Prox}_{\|\cdot\|^*} = P_{\operatorname{ball}(0;1)}$. Therefore we have $\operatorname{Prox}_{\|\cdot\|} = \operatorname{Id} - \operatorname{Prox}_{\|\cdot\|^*} = \operatorname{Id} - \operatorname{Prox}_{\iota_{\operatorname{ball}(0;1)}} = \operatorname{Id} - P_{\operatorname{ball}(0;1)}$. Thus, $\operatorname{Id} - \operatorname{Prox}_{\|\cdot\|} = P_{\operatorname{ball}(0;1)}$. Assume now $x \neq 0$. If $0 < \|x\| \leq 1$, then

$$G_f(x) = x - \frac{\frac{1}{2}\|x\|^2}{\|P_{\operatorname{ball}(0;1)}x\|^2}P_{\operatorname{ball}(0;1)}(x) = x - \frac{\|x\|^2}{2\|x\|^2}x = \tfrac{1}{2}x = \tfrac{1}{2}P_{\operatorname{ball}(0;1)}x; \tag{4.23}$$

and if $1 < \|x\|$, then

$$G_f(x) = x - \frac{\|x\| - \frac{1}{2}}{\|P_{\operatorname{ball}(0;1)}x\|^2}P_{\operatorname{ball}(0;1)}(x) \tag{4.24}$$

$$= x - \frac{\|x\| - \frac{1}{2}}{\left\|x/\|x\|\right\|^2}\frac{x}{\|x\|} = \tfrac{1}{2}\frac{x}{\|x\|} = \tfrac{1}{2}P_{\operatorname{ball}(0;1)}x. \tag{4.25}$$

Now $P_{\operatorname{ball}(0;1)}$ is firmly nonexpansive, and hence so is $\operatorname{Id} - P_{\operatorname{ball}(0;1)}$. It follows that $2G_f - \operatorname{Id} = -(\operatorname{Id} - P_{\operatorname{ball}(0;1)})$ is nonexpansive, and therefore that G_f is firmly nonexpansive. $\Box$

Proposition 4.6. *Let $(C_i)_{i\in I}$ be a finite family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$ and $f = \max_{i\in I} d_{C_i}$. Let $x \in \mathcal{H} \smallsetminus C$, set*

[10] Let $f : \mathcal{H} \to [-\infty, +\infty]$. The Fenchel conjugate of f is

$$f^*(u) = \sup_{x\in\mathcal{H}} \left(\langle x, u\rangle - f(x)\right).$$

$I(x) = \{i \in I : f(x) = d_{C_i}(x)\}$, and set $Q(x) = \operatorname{conv}\{P_{C_i}x\}_{i\in I(x)}$. Then

$$\mathcal{G}(x) = \bigcup_{q(x)\in Q(x)} \left\{x - \frac{f^2(x)}{\|x-q(x)\|^2}\big(x-q(x)\big)\right\} \text{ and } Q(x) \subseteq \operatorname{conv}\big(\{x\}\cup\mathcal{G}(x)\big). \tag{4.26}$$

If $I(x) = \{i\}$ is a singleton, then $\mathcal{G}(x) = \{P_{C_i}x\}$.

Proof. This follows from Proposition 4.3 (vi) and the fact that $\nabla d_{C_i}(x) = (x - P_{C_i}x)/d_{C_i}(x)$ when $x \in \mathcal{H}\smallsetminus C_i$. When $q(x) \in Q(x) = \operatorname{conv}\{P_{C_i}x\}_{i\in I(x)}$, □

Proposition 4.7. *Let $(C_i)_{i\in I}$ be a finite family of nonempty closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$. Let $(\lambda_i)_{i\in I}$ be a family in $]0,1]$ such that $\sum_{i\in I}\lambda_i = 1$. Let $p \geq 1$ and suppose that $f = \sum_{i\in I}\lambda_i d^p_{C_i}$. Set $(\forall x \in \mathcal{H})$ $I(x) = \{i \in I : x \notin C_i\}$. Then $\forall x \in \mathcal{H}$*

$$G_f(x) = x - \frac{\sum_{i\in I(x)}\lambda_i d^p_{C_i}(x)}{p\big\|\sum_{i\in I(x)}\lambda_i d^{p-2}_{C_i}(x)(x-P_{C_i}x)\big\|^2}\sum_{i\in I(x)}\lambda_i d^{p-2}_{C_i}(x)(x-P_{C_i}x) \tag{4.27}$$

and if $p = 2$, we rewrite this as

$$G_f(x) = \begin{cases} x - \dfrac{\sum_{i\in I}\lambda_i\|x-P_{C_i}x\|^2}{2\big\|\sum_{i\in I}\lambda_i(x-P_{C_i}x)\big\|^2}\Big(x - \sum_{i\in I}\lambda_i P_{C_i}x\Big), & \text{if } x \notin C; \\ x, & \text{otherwise.} \end{cases} \tag{4.28}$$

Proof. Let $x \in \mathcal{H}$, and let $i \in I$. Then $\nabla d_{C_i}(x) = \frac{x-P_{C_i}x}{d_{C_i}(x)}$ if $x \notin C_i$ and $0 \in \partial d_{C_i}(x)$ otherwise. Hence

$$\nabla d^p_{C_i}(x) = p d^{p-2}_{C_i}(x)(x - P_{C_i}x) \tag{4.29}$$

if $x \notin C_i$, and $0 \in \partial d^p_{C_i}(x)$ otherwise. The result follows. □

Example 4.8. Let $p \geq 1$ and suppose that $f = d^p_C$. Then

$$G_f = (1 - \tfrac{1}{p})\operatorname{Id} + \tfrac{1}{p}P_C.$$

Proof. This follows from Proposition 4.7 when I is a singleton. More precisely, $\nabla f(x) = pd_C^{p-1}\frac{x-P_Cx}{d_Cx} = pd_C^{p-2}(x-P_Cx)$. When $x \in \mathcal{H}\backslash C$,

$$\begin{aligned}
G_f(x) &= x - \frac{d_C^p(x)}{\|pd_C^{p-2}(x-P_Cx)\|^2}pd_C^{p-2}(x-P_Cx)\\
&= x - \frac{d_C^{2p-2}(x)}{p^2d_C^{2p-4}(x)\|x-P_Cx\|^2}p(x-P_Cx)\\
&= x - \frac{d_C^{2p-2}(x)}{p^2d_C^{2p-4}(x)d_C^2(x)}p(x-P_Cx)\\
&= x - \frac{1}{p}(x-P_Cx)\\
&= (1-\frac{1}{p})x + \frac{1}{p}P_Cx.
\end{aligned}$$

□

Example 4.9. Suppose that $u \in \mathcal{H}$ satisfies $\|u\| = 1$, and let $\beta \in \mathbb{R}$. Then the following hold:

(i) If $f\colon x \mapsto |\langle u, x\rangle - \beta|$, then $C = \{x \in \mathcal{H} : \langle u, x\rangle = \beta\}$ and $G_f\colon x \mapsto x - (\langle u, x\rangle - \beta)u$.

(ii) If $f\colon x \mapsto \langle u, x\rangle - \beta$, then $C = \{x \in \mathcal{H} : \langle u, x\rangle \leq \beta\}$ and $G_f\colon x \mapsto x - (\langle u, x\rangle - \beta)^+u$.

Proof. (i):By using the formula of projection onto a hyperplane in [7, Example 3.21], then $d_C(x) = \|x - P_Cx\| = \|x - (x - \frac{\langle u,x\rangle-\beta}{\|u\|^2}u)\| = |\langle u, x\rangle - \beta|$, so we have $f = d_C$ and hence $G_f = P_C$ by Example 4.8.
(ii): Note that $f^+ = d_C$ and hence $G_f = P_C$ by Proposition 4.3 (vii) and Example 4.8. □

We now give an example in which G_f is positively homogenenous but not necessarily linear.

Example 4.10. Let K be a nonempty closed convex cone with polar cone $K^\ominus$, and suppose that $f\colon x \mapsto \frac{1}{2}\langle x, P_Kx\rangle$. Then $G_f = \mathrm{Id} - \frac{1}{2}P_K = P_{K^\ominus} + \frac{1}{2}P_K$.

Proof. By Moreau's conical decomposition (see [7, Theorem 6.29]), let $x \in \mathcal{H}$, we have

$$x = P_Kx + P_{K^\ominus}x \text{ and } \langle P_Kx, P_{K^\ominus}x\rangle = 0.$$

Then

$$\langle x, P_K x\rangle = \langle P_K x + P_{K^\ominus} x, P_K x\rangle = \|P_K x\|^2.$$

So $f(x) = \frac{1}{2}\|P_K x\|^2 = \frac{1}{2}\|x - P_{K^\ominus} x\|^2 = \frac{1}{2} d^2_{K^\ominus}(x)$. By using [7, Corollary 12.30] it follows that $\nabla f(x) = x - P_{K^\ominus} x = P_K x$. Then

$$\begin{aligned} G_f(x) &= x - \frac{f(x)}{\|\nabla f(x)\|^2}\nabla f(x) \\ &= x - \frac{\frac{1}{2}\|P_K x\|^2}{\|P_K x\|^2} P_K x \\ &= x - \frac{1}{2} P_K x \\ &= (P_K x + P_{K^\ominus} x) - \frac{1}{2} P_K x \\ &= P_{K^\ominus} x + \tfrac{1}{2} P_K x. \end{aligned}$$

□

A direct verification yields the following result which is known when $p = 2$ (see [27] and [29]).

Proposition 4.11. *Let Y be another real Hilbert space, let $A\colon \mathcal{H} \to Y$ be continuous and linear, let $b \in Y$, and let $\varepsilon \geq 0$, and let $p \geq 1$. Suppose that $(\forall x \in \mathcal{H})\ f(x) = \|Ax - b\|^p - \varepsilon^p$ and that $C = \{x \in \mathcal{H} : \|Ax - b\| \leq \varepsilon\} \neq \varnothing$. Then $\forall x \in \mathcal{H}$*

$$G_f(x) = \begin{cases} x - \dfrac{\|Ax - b\|^p - \varepsilon^p}{p\|Ax - b\|^{p-2}\|A^*(Ax - b)\|^2} A^*(Ax - b), & \text{if } \|Ax - b\| > \varepsilon; \\ x, & \text{otherwise.} \end{cases} \tag{4.30}$$

4.5 Continuity of G_f vs Fréchet differentiability of f

In this section, we investigate the continuity of G_f, which is a desirable property when G_f is used as an operator in an algorithm. It turns out that strong-to-strong continuity of G_f corresponds precisely to Fréchet differentiability of f. We start with a technical result.

Lemma 4.12. *Let $(x_n)_{n=1}^\infty$ be a sequence in $\mathcal{H}$ converging weakly to $\bar{x}$ such that $x_n - G_f(x_n) \to 0$. Suppose that one of the following holds:*

(i) there exists $\delta > 0$ such that $(x_n)_{n=1}^{\infty}$ lies eventually in $\operatorname{ball}(\bar{x}, \delta)$ *and* $\sigma := \sup \|\partial f(\operatorname{ball}(\bar{x}; \delta))\| < +\infty$.

(ii) $x_n \to \bar{x}$.

(iii) f is bounded on every bounded subset of $\mathcal{H}$.

Then $\bar{x} \in C$.

Proof. If (i) is true : Without loss of generality, we assume that

$$(\forall n \in \mathbb{N}) \quad \|s(x_n)\| \leq \sigma. \tag{4.31}$$

Since f^+ is weakly lower semicontinuous, by using Fact 4.1(vi), we have

$$f^+(\bar{x}) \leq \liminf f^+(x_n) \leq \sigma \liminf \|x_n - G_f(x_n)\| = 0,$$

then $f(\bar{x}) \leq 0$, i.e., $\bar{x} \in C$.
Suppose (ii) holds: Since $x_n \to \bar{x}$, by definition of convergence of the sequence, $\forall \epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$x_n \in \operatorname{ball}(\bar{x}, \epsilon) \tag{4.32}$$

when $n > N$. Moreover, f is continuous at $\bar{x}$, by using [7, Proposition 16.14(iii)], there exists $\delta \in \mathbb{R}_{++}$ such that $\partial f(\operatorname{ball}(\bar{x}; \delta))$ is bounded. In (4.32), let $\epsilon = \delta$, we have (ii) $\Rightarrow$ (i).
Assume (iii) holds: By assumption, $(x_n)_{n=1}^{\infty}$ converges weakly to $\bar{x}$. Fact 2.24 gives (x_n) is bounded. Then there exists $\delta > 0$ such that $x_n \in \operatorname{ball}(\bar{x}, \delta)$ because f is bounded on every bounded subset of $\mathcal{H}$. Applying the equivalence of (i) and (iii) in [7, Proposition 16.17], we have $\operatorname{dom} \partial f = \mathcal{H}$ and ∂f maps every bounded subset of $\mathcal{H}$ to a bounded set. For $\operatorname{ball}(\bar{x}, \delta)$, it is a bounded subset of $\mathcal{H}$. Thus $\partial f(\operatorname{ball}(\bar{x}; \delta))$ is also bounded. Hence (iii)$\Rightarrow$ (i). □

Remark 4.13. Lemma 4.12 and Fact 4.1(ii) imply that G_f is *fixed-point closed* at $\bar{x}$ (see, e.g., also [20, Theorem 4.2.7] or [4]), i.e., if $x_n \to \bar{x}$ and $x_n - G_f(x_n) \to 0$, then $\bar{x} = G_f(\bar{x})$.

Proposition 4.14. *G_f is continuous at every point in C.*

Proof. Let $\bar{x} \in C$, and let $(x_n)_{n=1}^{\infty}$ be a sequence in $\mathcal{H}$ converging to $\bar{x}$. The result is clear if $(x_n)_{n=1}^{\infty}$ lies in C, so we can and do assume that $(x_n)_{n=1}^{\infty}$ lies in $\mathcal{H} \smallsetminus C$. Then $(\forall n \in \mathbb{N})$ $f(x_n) \leq f(\bar{x}) - \langle s(x_n), \bar{x} - x_n \rangle \leq \langle s(x_n), x_n - \bar{x} \rangle \leq \|s(x_n)\| \|\bar{x} - x_n\|$. Hence $0 < f(x_n)/\|s(x_n)\| \leq \|\bar{x} - x_n\| \to 0$. By Fact 4.1(vi), $x_n - G_f(x_n) \to 0$. Thus $\lim G_f(x_n) = \lim x_n = \bar{x} = G_f(\bar{x})$ by using Fact 4.1(ii). □

The continuity of G_f outside C is more delicate.

Fact 4.15. (Smulyan) (See, e.g., [18, Proposition 6.1.4].) The following hold:

(i) f is Fréchet differentiable at $\bar{x} \Leftrightarrow s$ is (strong-to-strong) continuous at $\bar{x}$.

(ii) f is Gâteaux differentiable at $\bar{x} \Leftrightarrow s$ is strong-to-weak continuous at $\bar{x}$.

Lemma 4.16. *Suppose that $\bar{x} \in \mathcal{H} \smallsetminus C$, that G_f is strong-to-weak continuous at $\bar{x}$, but G_f is not strong-to-strong continuous at $\bar{x}$. Then f is not Gâteaux differentiable at $\bar{x}$.*

Proof. There exists a sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H} \smallsetminus C$ such that $x_n \to \bar{x}$, $G_f(x_n) \rightharpoonup G_f(\bar{x})$ yet $G_f(x_n) \not\to G_f(\bar{x})$. It follows that

$$x_n - G_f(x_n) \rightharpoonup \bar{x} - G_f(\bar{x}) \quad \text{and} \quad x_n - G_f(x_n) \not\to \bar{x} - G_f(\bar{x}). \tag{4.33}$$

By the Kadec–Klee property[11] of $\mathcal{H}$, $\|x_n - G_f(x_n)\| \not\to \|\bar{x} - G_f(\bar{x})\|$. Since $\|\cdot\|$ is weakly lower semicontinuous, we assume (after passing to a subsequence and relabeling if necessary) that

$$\|\bar{x} - G_f(\bar{x})\| < \eta := \lim_{n \in \mathbb{N}} \|x_n - G_f(x_n)\|. \tag{4.34}$$

Using Fact 4.1 (viii), it follows that

$$s(x_n) = f(x_n) \frac{x_n - G_f(x_n)}{\|x_n - G_f(x_n)\|^2} \tag{4.35}$$

$$\rightharpoonup f(\bar{x}) \frac{\bar{x} - G_f(\bar{x})}{\eta^2} \tag{4.36}$$

$$\neq f(\bar{x}) \frac{\bar{x} - G_f(\bar{x})}{\|\bar{x} - G_f(\bar{x})\|^2} \tag{4.37}$$

$$= s(\bar{x}). \tag{4.38}$$

Thus, s is not strong-to-weak continuous at $\bar{x}$. It follows now from Fact 4.15 (ii) that f is not Gâteaux differentiable at $\bar{x}$. □

Theorem 4.17. *Let $\bar{x} \in \mathcal{H} \smallsetminus C$. Then the following are equivalent:*

[11] Kadec–Klee property is a sequence $(y_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$ converges to $\bar{y}$ if and only if $y_n \rightharpoonup \bar{y}$ and $\|y_n\| \to \|\bar{y}\|$.

(i) *f is Fréchet differentiable at $\bar{x}$.*

(ii) *G_f is (strong-to-strong) continuous at $\bar{x}$.*

(iii) *f is Gâteaux differentiable at $\bar{x}$ and G_f is strong-to-weak continuous at $\bar{x}$.*

Proof. (i)⇒(ii): By Fact 4.15 (i), s is continuous at $\bar{x}$. It follows from the definition of G_f that G_f is continuous at $\bar{x}$ as well.

(i)⇐(ii): In view of Fact 4.1 (viii), we have $s(x) = f(x)(x - G_f(x))/\|x - G_f(x)\|^2$ for all x sufficiently close to $\bar{x}$. Hence s is continuous at $\bar{x}$ and therefore f is Fréchet differentiable at $\bar{x}$ by Fact 4.15 (i).

(i)⇒(iii) and (ii)⇒(iii): This is clear since (i)⇔(ii) by the above.

(iii)⇒(ii): Suppose to the contrary that G_f is not strong-to-strong continuous. Then, by Lemma 4.16, f is not Gâteaux differentiable at $\bar{x}$ which is absurd. □

Corollary 4.18. *(*continuity) G_f is continuous everywhere if and only if f is Fréchet differentiable on $\mathcal{H} \smallsetminus C$.

Proof. Combine Proposition 4.14 with Theorem 4.17. □

Example 4.19. Suppose that $\mathcal{H} = \mathbb{R}$ and that $f(x) = \max\{-x, x, 2x - 1\}$ ($\forall x \in \mathbb{R}$). Then $C = \{0\}$ and f is not differentiable at 1; consequently, by Corollary 4.18, G_f is not continuous at 1.

Remark 4.20. (weak-to-weak continuity) It is unrealistic to expect that G_f is weak-to-weak continuous even when f is Fréchet differentiable; see [4, Example 3.2 and Remark 3.3.(ii)].

4.6 Continuity of G_f vs Gâteaux differentiability of f

In view of Fact 4.15 and Corollary 4.18, it is now tempting to conjecture that G_f is strong-to-weak continuous if and only if f is Gâteaux differentiable on $\mathcal{H} \smallsetminus C$. Perhaps somewhat surprisingly, this turns out to be wrong. The counterexample is based on an ingenious construction by Borwein and Fabian [15].

Example 4.21. (Borwein–Fabian) (See [15, Proof of Theorem 4].) Suppose that $\mathcal{H}$ is infinite-dimensional. Then there exists a function $b\colon \mathcal{H} \to \mathbb{R}$ such that the following hold:

(i) b is continuous, convex and $\min b(\mathcal{H}) = b(0) = 0$.

(ii) b is Fréchet differentiable on $\mathcal{H} \smallsetminus \{0\}$.

(iii) b is Gâteaux differentiable at 0, and $\nabla b(0) = 0$.

(iv) b is not Fréchet differentiable at 0.

Example 4.22. (lack of strong-to-weak continuity) Let b be as in Example 4.21. Then there exists $y \in \mathcal{H}$ such that $\nabla b(y) \neq 0$. Suppose that

$$(\forall x \in \mathcal{H}) \quad f(x) = b(x) - \langle \nabla b(y), x\rangle - \tfrac{1}{2}\big(b(y) - \langle \nabla b(y), y\rangle\big). \tag{4.39}$$

Then the following hold:

(i) f is Gâteaux differentiable (but not Fréchet differentiable) at 0, and G_f is not strong-to-weak continuous at 0.

(ii) f is Fréchet differentiable on $\mathcal{H} \smallsetminus \{0\}$, and G_f is continuous on $\mathcal{H} \smallsetminus \{0\}$.

Proof. By Example 4.21 (iii), $0 \in \operatorname{ran} \nabla b$. If $\{0\} = \operatorname{ran} \nabla b$, then we would deduce that b is constant and therefore Fréchet differentiable; in turn, this would contradict Example 4.21 (iv). Hence $\{0\} \subsetneqq \operatorname{ran} \nabla b$ and there exists $y \in \mathcal{H}$ such that

$$v = \nabla b(y) \neq 0. \tag{4.40}$$

Now set

$$g\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto b(x) - \langle v, x\rangle. \tag{4.41}$$

Then

$$(\forall x \in \mathcal{H}) \quad f(x) = g(x) - \tfrac{1}{2}g(y), \tag{4.42}$$

and $g(0) = b(0) - \langle v, 0\rangle = 0$ by Example 4.21 (i). Example 4.21 (iii) and (4.40) yield $\nabla g(0) = \nabla b(0) - v = -v \neq 0$ while $\nabla g(y) = \nabla b(y) - v = 0$. Hence $\min g(\mathcal{H}) = g(y) < g(0) = 0$ and therefore

$$f(y) = \min f(\mathcal{H}) = \min g(\mathcal{H}) - \tfrac{1}{2}g(y) = \tfrac{1}{2}g(y) < 0 < 0 - \tfrac{1}{2}g(y) = f(0). \tag{4.43}$$

Thus $y \in C$ while $0 \notin C$.

(i): On the one hand, since b is not Fréchet differentiable at 0 (Example 4.21 (iv)), neither is f. On the other hand, since b is Gâteaux differentiable at 0 (Example 4.21 (iii)), so is f. Altogether, f is Gâteaux differentiable, but not Fréchet differentiable at 0. Therefore, by Theorem 4.17, G_f is not strong-to-weak continuous at 0.

(ii): Since b is Fréchet differentiable on $\mathcal{H} \smallsetminus \{0\}$ (Example 4.21 (ii)), so is f. Now apply Theorem 4.17. $\square$

4.7 G_f as an accelerated mapping

In this section, we consider the case when f is a power of a quadratic form. We link G_f to the accelerated mapping corresponding to a linear operator. This mapping can significantly speed up convergence of algorithms (see [8]).

Proposition 4.23. *Suppose that $f\colon x \mapsto \sqrt{\langle x, Mx\rangle^p}$, where $p \geq 1$ and $M\colon \mathcal{H} \to \mathcal{H}$ be continuous, linear, self-adjoint, and positive. Then G_f is continuous everywhere and*

$$(\forall x \in \mathcal{H}) \quad G_f(x) = \begin{cases} x - \dfrac{\langle x, Mx\rangle}{p\|Mx\|^2}Mx, & \text{if } Mx \neq 0; \\ x, & \text{if } Mx = 0. \end{cases} \tag{4.44}$$

Proof. Assume first that $p = 1$. Since M has a unique positive square root, i.e., there exists[12] $B\colon \mathcal{H} \to \mathcal{H}$ such that B is continuous, linear, self-adjoint, and positive, and $\ker B = \ker M$. Hence $(\forall x \in \mathcal{H})$ $f(x) = \sqrt{\langle x, Mx\rangle} = \|Bx\|$ so f is indeed convex and continuous. If $x \in \mathcal{H} \smallsetminus \ker M = \mathcal{H} \smallsetminus \ker B$, then f is Fréchet differentiable at x with $\nabla f(x) = B^*Bx/\|Bx\| = Mx/\|Bx\|$; hence,

$$G_f(x) = x - \frac{\|Bx\|}{\|Mx\|^2/\|Bx\|^2}\frac{Mx}{\|Bx\|} = x - \frac{\|Bx\|^2}{\|Mx\|^2}Mx = x - \frac{\langle x, Mx\rangle}{\|Mx\|^2}Mx \tag{4.45}$$

and G_f is continuous everywhere by Corollary 4.18. If $p > 1$, then the result follows from the above and Proposition 4.3 (iii). $\square$

[12] See, e.g., [36, Theorem 9.4-2], where this is stated in a complex Hilbert space; however, the proof works unchanged in our real setting as well.

Example 4.24. Let $A\colon \mathcal{H} \to \mathcal{H}$ be linear, self-adjoint, and nonexpansive. Suppose that $(\forall x \in \mathcal{H})\ f(x) = \sqrt{\langle x, x - Ax\rangle}$. Then G_f is positively homogeneous, continuous everywhere, and

$$(\forall x \in \mathcal{H}) \quad G_f(x) = \begin{cases} x - \dfrac{\langle x, x - Ax\rangle}{\|x - Ax\|^2}(x - Ax), & \text{if } Ax \neq x; \\ x, & \text{if } Ax = x. \end{cases} \tag{4.46}$$

Proof. Use Proposition 4.23 with $M = \mathrm{Id} - A$ and $p = 1$. □

Remark 4.25. (accelerated mapping) Let $A\colon \mathcal{H} \to \mathcal{H}$ be linear, nonexpansive, and self-adjoint. In [8], the authors study the accelerated mapping[13] of A, i.e.,

$$x \mapsto t_x Ax + (1 - t_x)x, \quad \text{where } t_x = \begin{cases} \dfrac{\langle x, x - Ax\rangle}{\|x - Ax\|^2}, & \text{if } x \neq Ax; \\ 1, & \text{otherwise.} \end{cases} \tag{4.47}$$

In view of the Example 4.24, the accelerated mapping of A is precisely the subgradient projector G_f of the function $x \mapsto \sqrt{\langle x, x - Ax\rangle}$. Now suppose that $\mathcal{H} = \ell^2(\mathbb{N})$, let $(e_n)_{n\in\mathbb{N}}$ be the standard orthonormal basis of $\mathcal{H}$, and suppose that

$$A\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \sum_{n\in\mathbb{N}} \tfrac{n}{n+1}\langle e_n, x\rangle e_n. \tag{4.48}$$

This A is linear, self-adjoint and nonexpansive. Indeed, linearity is obviously from the constrcution of A. A is self-adjoint because $(\forall x, y \in \mathcal{H})$

$$\langle Ax, y\rangle = \langle \sum_{n\in\mathbb{N}} \tfrac{n}{n+1}\langle e_n, x\rangle e_n, y\rangle = \sum_{n\in\mathbb{N}} \tfrac{n}{n+1} x_n y_n = \langle x, Ay\rangle.$$

For the nonexpansiveness of A, we will use Bessel inequality. For every

[13] In fact, the operator A in [8] need not necessarily be self-adjoint.

$x, y \in \mathcal{H}$, then

$$\begin{aligned}
\|Ax - Ay\|^2 &= \|A(x-y)\|^2 \qquad \text{(since } A \text{ is linear)} \\
&= \left\| \sum_{n\in\mathbb{N}} \tfrac{n}{n+1} \langle e_n, x-y\rangle e_n \right\|^2 \\
&= \langle \sum_{n\in\mathbb{N}} \tfrac{n}{n+1} \langle e_n, x-y\rangle e_n, \sum_{n\in\mathbb{N}} \tfrac{n}{n+1} \langle e_n, x-y\rangle e_n \rangle \\
&= \sum_{n\in\mathbb{N}} \left(\tfrac{n}{n+1}\right)^2 \langle e_n, x-y\rangle^2 \\
&\leq \sum_{n\in\mathbb{N}} \langle e_n, x-y\rangle^2 \\
&\leq \|(x-y)\|^2 \quad \text{(using Bessel inequality)} \\
&\leq \|x-y\|^2.
\end{aligned}$$

Thus A defined in (4.48) is nonexpansive. By using Example 4.24, then G_f is continuous; however, G_f is neither linear nor uniformly continuous (see the [8, Remark following Lemma 3.8]).

4.8 Nonexpansiveness

We now discuss when G_f is (firmly) nonexpansive or monotone. These properties occur when studying resolvents, subdifferentials and gradients. For instance, every resolvent and every proximal mapping is firmly nonexpansive; subdifferential (or gradient) operators of convex functions are monotone. However, these properties are not automatic for subgradient projectors as we will see in this section.

Proposition 4.26. *Suppose that f is Gâteaux differentiable on $\mathcal{H} \smallsetminus C$ and that G_f is firmly nonexpansive. Then G_g is likewise in each of the following situations:*

(i) $\alpha > 0$, *and* $g = f \circ \alpha\,\mathrm{Id}$ *is convex.*

(ii) $f \geq 0$, $\alpha \geq 1$, *and* $g = f^\alpha$ *is convex.*

(iii) $A\colon \mathcal{H} \to \mathcal{H}$ *is continuous and linear,* $AA^* = A^*A = \mathrm{Id}$, *and* $g = f \circ A$.

(iv) $z \in \mathcal{H}$ *and* $g\colon x \mapsto f(x-z)$.

The analogous statement holds when G_f is assumed to be nonexpansive.

Proof. This follows from the corresponding items in Proposition 4.3, which do preserve (firm) nonexpansiveness. □

On the real line, we obtain a simpler test.

Proposition 4.27. *Suppose that $\mathcal{H} = \mathbb{R}$ and that f is twice differentiable on $\mathcal{H} \smallsetminus C$. Then G_f is monotone. Moreover, G_f is (firmly) nonexpansive if and only if*

$$(\forall x \in \mathbb{R}) \quad f(x)f''(x) \leq \big(f'(x)\big)^2. \tag{4.49}$$

Proof. By Corollary 4.18, G_f is continuous. Let $x \in \mathbb{R} \smallsetminus C$. Then $G_f(x) = x - f(x)/f'(x)$ and hence $G_f'(x) = f(x)f''(x)/(f'(x))^2 \geq 0$. It follows that G_f is increasing on $\mathbb{R} \smallsetminus C$ and hence on $\mathbb{R}$. Furthermore, G_f is (firmly) nonexpansive if and only if $G_f'(x) \leq 1$, which gives the remaining characterization. □

Example 4.28. Suppose that $\mathcal{H} = \mathbb{R}$, let $\alpha > 0$, and suppose that $(\forall x \in \mathbb{R})$ $f(x) = x^n - \alpha$, where $n \in \{2, 4, 6, 8, \ldots\}$. Then G_f is firmly nonexpansive.

Proof. If $x \in \mathbb{R} \smallsetminus C$, then $(f'(x))^2 - f(x)f''(x) = nx^{n-2}(\alpha n + x^n - \alpha) > 0$ and we are done by Proposition 4.27. □

Example 4.29. Suppose that $\mathcal{H} = \mathbb{R}$ and that $f\colon x \mapsto \exp(|x|) - 1$. Then $(\forall x \in \mathcal{H})$ $G_f(x) = x - \operatorname{sgn}(x)(1 - \exp(-|x|))$ and $G_f'(x) = 1 - \exp(-|x|) \in [0, 1[$. It follows that G_f is firmly nonexpansive[14].

Example 4.30. Suppose that $\mathcal{H} = \mathbb{R}$ and that $f\colon x \mapsto \exp(x^2) - 1$. Then G_f is not (firmly) nonexpansive. Indeed, we compute $(f'(x))^2 - f(x)f''(x) = 4x^2\exp(x^2) + 2\exp(x^2) - 2\exp(2x^2)$, which strictly negative when $|x| > 1.2$. Now apply Proposition 4.27.

Proposition 4.31. *Suppose that $\mathcal{H} = \mathbb{R}$ and that f is twice differentiable, that $\min f(\mathcal{H}) = 0$, that $g = f\Box(1/2)|\cdot|^2$, and that $2ff'' \leq (2 + f'')(f')^2$. Then G_g is firmly nonexpansive.*

Proof. We start by observing a couple of facts. First,

$$g' = \operatorname{Id} - \operatorname{Prox}_f\,. \tag{4.50}$$

[14] Since G is monotone by Proposition 4.27, its antiderivative $x \mapsto \frac{1}{2}x^2 - |x| - \exp(-|x|)$ is convex — although this does not look like convex function on first glance! It is interesting to do this also for other instances of f.

Fix $x \in \mathbb{R}$. Recall $y := \operatorname{Prox}_f(x)$ is a minimizer of

$$z \mapsto f(z) + \frac{1}{2}|x - z|^2.$$

Then we have

$$0 \in \partial f(y) - (x - y).$$

Since f is differentiable, so we have

$$x = y + f'(y). \tag{4.51}$$

Because $(y+f'(y))' = 1+f''(y) > 0$. If we view x as a function of y in (4.51), we have $x(y)$ is differentiable. By the Inverse Function Theorem[15], we conclude that y as a function of x is differentiable. Hence implicit differentiation gives

$$1 = y'(x) + f''(y)y'(x) = y'(x)(1 + f''(y(x))).$$

Hence $y' = \frac{1}{1+f''(y(x))}$ and thus

$$g''(x) = \big(\operatorname{Id} - \operatorname{Prox}_f\big)'(x) = 1 - \frac{1}{1 + f''(\operatorname{Prox}_f(x))} = \frac{f''\big(\operatorname{Prox}_f(x)\big)}{1 + f''\big(\operatorname{Prox}_f(x)\big)}. \tag{4.52}$$

In view of Proposition 4.27 and because $g(x) = f(\operatorname{Prox}_f(x)) + (1/2)(x - \operatorname{Prox}_f(x))^2$ we must verify that $gg'' \leq (g')^2$, i.e.,

$$\frac{\big(f(\operatorname{Prox}_f(x)) + \frac{1}{2}(x - \operatorname{Prox}_f(x))^2\big) f''\big(\operatorname{Prox}_f(x)\big)}{1 + f''\big(\operatorname{Prox}_f(x)\big)} \leq \big(x - \operatorname{Prox}_f(x)\big)^2. \tag{4.53}$$

Again writing $y = \operatorname{Prox}_f(x)$ gives $x - y = f'(y)$ and so see that (4.53) is equivalent to

$$\frac{\big(f(y) + \frac{1}{2}(f'(y))^2\big) f''(y)}{1 + f''(y)} \leq \big(f'(y)\big)^2. \tag{4.54}$$

Actually, (4.54) is equivalent to our assumption on f. Indeed, $(\forall y \in \mathbb{R})$

[15]The Inverse Function Theorem: (see [33, Theorem 1A.1] on page 10) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on some open set containing x_0. Suppose that the determinant of Jacobian of $f(x_0)$ does not equal to 0. Then there exists an open set V containing x_0 and an open set U containing $f(x_0)$ such that $f : V \to U$ ha a continuous inverse $f^{-1} : U \to V$ which is differentiable.

$$\frac{\big(f(y)+\frac{1}{2}(f'(y))^2\big)f''(y)}{1+f''(y)} \leq \big(f'(y)\big)^2 \tag{4.55}$$

$$\Leftrightarrow \big(f(y)+\tfrac{1}{2}(f'(y))^2\big)f''(y) \leq (1+f''(y))(f')^2 \tag{4.56}$$

$$\Leftrightarrow f(y)f''(y) \leq (1+\frac{1}{2}f''(y))(f'(y))^2 \tag{4.57}$$

$$\Leftrightarrow 2f(y)f''(y) \leq (2+f''(y))(f'(y))^2. \tag{4.58}$$

Therefore, by Proposition 4.27, G_g is firmly nonexpansive. □

We conclude this section with a result on the range of $\operatorname{Id} - G_f$.

Proposition 4.32. *We have* $\operatorname{ran}(\operatorname{Id} - G_f) \subseteq \operatorname{cone}\operatorname{ran}\partial f \subseteq (\operatorname{rec} C)^{\ominus}$.

Proof. Let $y^* \in \operatorname{ran}\partial f$. Then there exists $y \in \operatorname{dom}\partial f$ such that $y^* \in \partial f(y)$. Let $c \in C$ and $x \in \operatorname{rec} C$. Then $(c+nx)_{n\in\mathbb{N}}$ lies in C. Hence $(\forall n \geq 1)$ $0 \geq f(c+nx) \geq f(y) + \langle y^*, c+nx-y\rangle$. Then

$$\langle y^*, x\rangle \leq \frac{\langle y^*, y-c\rangle - f(y)}{n} \to 0 \quad \text{as } n \to +\infty. \tag{4.59}$$

It follows that $y^* \in (\operatorname{rec} C)^{\ominus}$. Recall that $\operatorname{rec} C$ is actually a cone. Thus, $\operatorname{cone}\operatorname{ran}\partial f \subseteq (\operatorname{rec} C)^{\ominus}$. Let $z \in \mathcal{H} \smallsetminus C$ and $z^* = s(z) \in \partial f(z)$, we have

$$z - G_f(z) = \frac{f(z)}{\|z^*\|^2} z^* \in \operatorname{cone}\operatorname{ran}\partial f.$$

Therefore, $\operatorname{ran}(\operatorname{Id} - G_f) \subseteq \operatorname{cone}\operatorname{ran}\partial f \subseteq (\operatorname{rec} C)^{\ominus}$. □

4.9 The decreasing property

We say that f has the *decreasing property* if

$$(\forall x \in \mathcal{H}) \quad \sup f(\mathcal{G}x) \leq f(x). \tag{4.60}$$

This property is interesting in the context of applying subgradient projectors as building blocks for algorithms. To investigate the decreasing property, it suffices to consider points outside C.

Proposition 4.33. *If* $(\forall x \in \mathcal{H})$ $G_f(x) \in \operatorname{conv}(\{x\} \cup C)$, *then* f *has the decreasing property.*

Proof. Let $x \in \mathcal{H} \smallsetminus C$. Then there exists $c \in C$ and $\lambda \in [0,1]$ such that $G_f(x) = (1-\lambda)x + \lambda c$. It follows that $f(G_f(x)) \leq (1-\lambda)f(x) + \lambda f(c) \leq (1-\lambda)f(x) \leq f(x)$. □

Lemma 4.34. *Let $(x,y,z) \in \mathbb{R}^3$ be such that $x \neq z$ and $(z-y)(x-y) \leq 0$. Then $y \in \operatorname{conv}\{x,z\}$.*

Proof. Suppose first that $z < x$. If $y > x$, then $(z-y)(x-y) > 0$ because it is the product of two strictly negative numbers. Similarly, if $y < z$, then $(z-y)(x-y) > 0$. We deduce that $y \in [z,x]$. Analogously, when $x < z$, we obtain that $y \in [x,z]$. In either case, $y \in \operatorname{conv}\{x,z\}$. □

The next result shows that G is particularly well behaved on the real line.

Corollary 4.35. *Suppose that $\mathcal{H} = \mathbb{R}$. Then f has the decreasing property.*

Proof. Let $x \in \mathbb{R} \smallsetminus C$. Then $x \neq P_C x$ and, by Fact 4.1 (iv), $(P_C x - G_f(x))(x - G_f(x)) \leq 0$. Lemma 4.34 thus yields $G_f(x) \in \operatorname{conv}\{x, P_C x\}$. Hence $G_f(x) \in \operatorname{conv}(\{x\} \cup C)$, and we are done by Proposition 4.33. □

The next example shows that the decreasing property is not automatic.

Example 4.36. Suppose that $\mathcal{H} = \mathbb{R}^2$, that $C_1 = \mathbb{R} \times \{0\}$, that $C_2 = \{(\xi,\xi) \in \mathcal{H} : \xi \in \mathbb{R}\}$, and that $f = \max\{d_{C_1}, d_{C_2}\}$. Then f does not have the decreasing property.

Proof. Set $x = (2,1)$. Then, using Proposition 4.6, we obtain that $G_f(x) = (2,0)$ and $f(x) = 1 < \sqrt{2} = f(G_f(x))$. □

We now illustrate that the sufficient condition of Proposition 4.33 is not necessary:

Example 4.37. Suppose that $\mathcal{H} = \mathbb{R}^2$ and that $(\forall x = (x_1,x_2) \in \mathbb{R}^2)$ $f(x) = |x_1| + |x_2|$. Then f has the decreasing property, $G_f^2(x) = (0,0)$ yet $G_f(x) \notin \operatorname{conv}\{(0,0),x\}$ for almost every $x \in \mathbb{R}^2$. Furthermore, G_f is not monotone.

Proof. We have $\operatorname{lev}_0 f = \{(0,0)\}$ and

$$G_f(x_1,x_2) = \begin{cases} \left(\frac{|x_1|-|x_2|}{2|x_1|}x_1, \frac{|x_2|-|x_1|}{2|x_2|}x_2\right) & \text{if } x_1 \neq 0, x_2 \neq 0, \\ \left(-\frac{s|x_2|}{s^2+1}, \frac{s^2}{s^2+1}x_2\right) & \text{if } x_1 = 0, x_2 \neq 0, \\ \left(\frac{s^2}{s^2+1}x_1, -\frac{s|x_1|}{s^2+1}\right) & \text{if } x_1 \neq 0, x_2 = 0, \\ (0,0) & \text{if } x_1 = x_2 = 0, \end{cases} \tag{4.61}$$

where $s \in [-1, 1]$ depends on x.
In particular, when $x_1 x_2 > 0$,

$$G_f(x_1, x_2) = \left(\frac{x_1 - x_2}{2}, \frac{x_2 - x_1}{2} \right)$$

so it is a projection on $\mathbb{R}(1, -1)$.
When $x_1 x_2 < 0$,

$$G_f(x_1, x_2) = \left(\frac{x_1 + x_2}{2}, \frac{x_1 + x_2}{2} \right)$$

so it is a projection on $\mathbb{R}(1, 1)$.

(i) f has the decreasing property.
When $x_1 \neq 0, x_2 \neq 0$, then

$$f(G_f(x_1, x_2)) = ||x_1| - |x_2|| \leq |x_1| + |x_2| = f(x_1, x_2).$$

When $x_1 = 0, x_2 \neq 0$, we have

$$f(G_f(x_1, x_2)) = \frac{|s x_2|}{s^2 + 1} + \frac{s^2 |x_2|}{s^2 + 1} = \frac{(|s| + s^2)|x_2|}{s^2 + 1} \leq |x_2| = f(0, x_2).$$

When $x_1 \neq 0, x_2 = 0$, we have

$$f(G_f(x_1, x_2)) = \frac{(s^2 + |s|)|x_1|}{s^2 + 1} \leq |x_1| = f(x_1, 0).$$

(ii) $G_f(x) \notin \operatorname{conv}\{(0, 0), x\}$ for almost every $x \in \mathbb{R}^2$.
Indeed, $[(x_1, x_2), (0, 0)] = \{\lambda(x_1, x_2) : \ 0 \leq \lambda \leq 1\}$.
When $(x_1, x_2) \in \mathbb{R}^2 \setminus \Big(\mathbb{R}(1, 0) \cup \mathbb{R}(0, 1) \cup \mathbb{R}(1, 1) \cup \mathbb{R}(1, -1) \Big)$, we have $|x_1| \neq |x_2|$ so that

$$\frac{|x_1| - |x_2|}{2|x_1|} \frac{|x_2| - |x_1|}{2|x_2|} < 0$$

then

$$G_f(x_1, x_2) = \left(\frac{|x_1| - |x_2|}{2|x_1|} x_1, \frac{|x_2| - |x_1|}{2|x_2|} x_2 \right) \notin [(x_1, x_2), (0, 0)].$$

Moreover,

$$G_f^2(x_1, x_2) \tag{4.62}$$

$$= G_f(G_f(x_1, x_2)) = \tag{4.63}$$

$$\left(\frac{\frac{\left||x_1|-|x_2|\right|}{2} - \frac{\left||x_2|-|x_1|\right|}{2}}{\left||x_1| - |x_2|\right|} \frac{|x_1| - |x_2|}{2|x_1|}, \frac{\frac{\left||x_2|-|x_1|\right|}{2} - \frac{\left||x_1|-|x_2|\right|}{2}}{\left||x_2| - |x_1|\right|} \frac{|x_2| - |x_1|}{2|x_2|} \right) \tag{4.64}$$

$$= (0, 0). \tag{4.65}$$

(iii) G_f is not monotone.
Let $x = (-1, 3)$ and $y = (1, 3)$. Referring to (4.61), we have

$$G_f(x) = \left(\frac{1-3}{2}(-1), \frac{3-1}{6}3 \right) = (1, 1)$$

and

$$G_f(y) = \left(\frac{1-3}{2}, \frac{3-1}{6}3 \right) = (-1, 1).$$

Hence $\langle x - y, G_f(x) - G_f(y) \rangle = \langle (-2, 0), (2, 0) \rangle = -4 < 0$, so G_f is not monotone.

□

Remark 4.38. (infeasibility detection) Using the decreasing property, one obtains a sufficient condition for infeasibility: Suppose that $\mathcal{H} = \mathbb{R}$ and we find a point x such that $f(G_f(x)) > f(x)$. Then C must be empty because of Corollary 4.35. For instance, suppose that $f \colon x \mapsto x^2 + 1$. Then

$$(\forall x \in \mathbb{R} \smallsetminus \{0\}) \quad G_f(x) = (x^2 - 1)/(2x). \tag{4.66}$$

Now set $x = 1/2$. Then $G_f(x) = -3/4$ and $f(G_f(x)) = 25/16 > 5/4 = f(x)$. It is known since the 19th century that the concrete instance (4.66) exhibits chaotic behaviour; see, e.g., [38, Problem 7-a on page 72].

Remark 4.39. (Newton iteration) Suppose that $\mathcal{H} = \mathbb{R}$ and that f is differentiable on $X \smallsetminus C$. Then

$$(\forall x \in \mathbb{R} \smallsetminus C) \quad G_f(x) = x - \frac{f(x)}{\big(f'(x)\big)^2} f'(x) = x - \frac{f(x)}{f'(x)} \tag{4.67}$$

is the same as the Newton operator for finding a zero of f!

The decreasing property is preserved in certain cases:

Proposition 4.40. *Suppose that f has the decreasing property. Then the following hold:*

(i) If $\alpha > 0$, then αf has the decreasing property.

(ii) If $\alpha \geq 1$, then $(f^+)^\alpha$ has the decreasing property.

Proof. Let $x \in \mathcal{H} \smallsetminus C$.
(i): Let $\alpha > 0$. $(\alpha f)\mathcal{G}_{\alpha f}(x) = (\alpha f)\mathcal{G}_f(x)$ and hence $\sup(\alpha f)(\mathcal{G}_{\alpha f}(x)) = \alpha \sup f(\mathcal{G}_f(x)) \leq \alpha f(x) = (\alpha f)(x)$ by Proposition 4.3 (i).

(ii): Set $g = (f^+)^\alpha$ and $\beta = 1/\alpha$. Then $0 < \beta \leq 1$ and $\mathcal{G}_g(x) = (1-\beta)x + \beta\mathcal{G}_f(x)$ by Proposition 4.3 (iii). Hence $\sup g(\mathcal{G}_g x) \leq (1-\beta)g(x) + \beta \sup g(\mathcal{G}_f x)$. On the other hand, since f has the decreasing property, then $f(\mathcal{G}_f x) \leq f(x)$, then $g(\mathcal{G}_f x) = \Big(f^+(\mathcal{G}_f x)\Big)^\alpha \leq \Big(f^+(x)\Big)^\alpha = g(x)$, thus $\sup g(\mathcal{G}_f(x)) \leq g(x)$. Altogether, $\sup g(\mathcal{G}_g x) \leq g(x)$, i.e., g is decreasing. $\square$

The following result is complementary to the decreasing property.

Proposition 4.41. *Suppose that f is strictly convex at $x \in \mathcal{H}$ and $f(x) > 0$. Then $f(G_f(x)) > 0$.*

Proof. Recall that f is strictly convex at x if $(\forall y \in \mathcal{H} \smallsetminus \{x\})$ $(\forall \lambda \in \,]0,1[)$ $f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y)$. Arguing as in [18, proof of Proposition 5.3.4 (a)], we see that $\frac{1}{2}\langle s(x), G_f(x) - x\rangle = \langle s(x), (\frac{1}{2}x + \frac{1}{2}G_f(x)) - x\rangle \leq f(\frac{1}{2}x + \frac{1}{2}G_f(x)) - f(x) < \frac{1}{2}f(x) + \frac{1}{2}f(G_f(x)) - f(x) = \frac{1}{2}(f(G_f(x)) - f(x))$. Therefore, $f(G_f(x)) > f(x) + \langle s(x), G_f(x) - x\rangle = 0$ using Fact 4.1 (i). $\square$

Remark 4.42. Suppose that f is strictly convex. Then Proposition 4.41 shows that iterating G_f starting at a point outside C will never reach C in finitely many steps. This is clearly illustrated by Example 4.8, which shows that the function d_C, even though it is neither strictly convex nor differentiable everywhere, performs best because $G_f = P_C$ yields a solution after just one step.

4.10 The subgradient projector of $f(x_1, x_2) = |x_1|^p + |x_2|^p$

This section contains a case study. It reveals that the various properties of G_f can exhibit complicated behaviour.

The following result complements Example 4.37.

Proposition 4.43. *Suppose that $\mathcal{H} = \mathbb{R}^2$ and that $f\colon (x_1, x_2) \mapsto |x_1|^p + |x_2|^p$, where $p > 1$, and let $x = (x_1, x_2) \in \mathbb{R}^2 \smallsetminus \{(0,0)\}$. Then $G_f(x)$ is*

$$\left(x_1 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_1|^{p-1}\operatorname{sgn}(x_1)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}, x_2 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_2|^{p-1}\operatorname{sgn}(x_2)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}\right) \tag{4.68}$$

and the following hold:

(i) If $p \geq 2$, then $f(x) \geq f(G_f(x)) \geq (1 - 2p^{-1})^p f(x)$.

(ii) If $1 < p \leq 2$, then $f(x) \geq f(G_f(x)) \geq 2^{-1}(1 - p^{-1})^p f(x)$.

(iii) If $1 < p < 2$, then G_f is not monotone.

Proof. The formula (4.68) is a direct verification. More precisely, since

$$\nabla f(x_1, x_2) = (p|x_1|^{p-1}\operatorname{sgn}(x_1), p|x_2|^{p-1}\operatorname{sgn}(x_2)). \tag{4.69}$$

When $x_1 = 0$, $x_2 \neq 0$, we have $\nabla f(0, x_2) = (0, p|x_2|^{p-1}\operatorname{sgn}(x_2))$ from above formula, then we have

$$G_f(0, x_2) = (0, (1 - 1/p)x_2).$$

As $p > 1$, so we have

$$f(G_f(0, x_2)) = 0 + (1 - 1/p)^p|x_2|^p = (1 - 1/p)^p f(0, x_2) \leq |x_2|^p \leq f(x_1, x_2).$$

The proof for $x_1 \neq 0$ and $x_2 = 0$ is similar. Hence (i) (ii) hold when $x_1 = 0$ or $x_2 = 0$. Consequently, we assume that $x_1 \neq 0$ and $x_2 \neq 0$. Thus (4.69) gives $G_f(x)$, which is

$$\left(x_1 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_1|^{p-1}\operatorname{sgn}(x_1)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}, x_2 - \frac{\big(|x_1|^p + |x_2|^p\big)|x_2|^{p-1}\operatorname{sgn}(x_2)}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)}\right) \tag{4.70}$$

Now (4.70) gives $f(G_f(x_1, x_2)) =$

$$|x_1|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_1|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p + |x_2|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_2|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p. \tag{4.71}$$

Define

$$c_i(x_1, x_2) = \frac{\big(|x_1|^p + |x_2|^p\big)|x_i|^{p-2}}{p\big(|x_1|^{2p-2} + |x_2|^{2p-2}\big)},$$

for $i = 1, 2$.

(i): If $p \geq 2$. Note that

$$f(G_f(x)) = |x_1|^p\big|1 - c_1\big|^p + |x_2|^p\big|1 - c_2\big|^p. \tag{4.72}$$

When $|x_1| \geq |x_2|$, we have

$$c_1(x_1, x_2) \leq \frac{2|x_1|^p|x_1|^{p-2}}{p|x_1|^{2p-2}} = \frac{2}{p}.$$

When $|x_1| \leq |x_2|$, because $p \geq 2$ we have $|x_1|^{p-2} \leq |x_2|^{p-2}$ so that

$$c_1(x_1, x_2) \leq \frac{2|x_2|^p|x_2|^{p-2}}{p|x_2|^{2p-2}} = \frac{2}{p}.$$

Similarly, we have

$$c_2(x_1, x_2) \leq \frac{2}{p}.$$

Thus, for $i \in \{1, 2\}$ we have

$$0 < c_i(x_1, x_2) \leq \frac{2}{p}.$$

Therefore,

$$0 \leq 1 - \frac{2}{p} \leq 1 - c_i(x_1, x_2) \leq 1.$$

Together with (4.72), we obtain

$$\left(1 - \frac{2}{p}\right)^p f(x_1, x_2) \leq f(G_f(x_1, x_2)) \leq f(x_1, x_2).$$

(ii): If $1 < p \leq 2$. We assume that $|x_1| \leq |x_2|$, the other case is treated analogously. Write $f(G_f(x_1, x_2))$

$$= |x_1|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_1|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p + |x_2|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_2|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p \tag{4.73}$$

$$= |x_2|^p \left(\frac{|x_1|^p}{|x_2|^p}\left|1 - \frac{(|x_1|^p + |x_2|^p)|x_1|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p\right) \tag{4.74}$$

$$+ |x_2|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_2|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p \tag{4.75}$$

$$= |x_2|^p \left|\frac{|x_1|}{|x_2|} - \frac{|x_1|}{|x_2|}\frac{(|x_1|^p + |x_2|^p)|x_1|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p \tag{4.76}$$

$$+ |x_2|^p \left|1 - \frac{(|x_1|^p + |x_2|^p)|x_2|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})}\right|^p. \tag{4.77}$$

Set $t = |x_1|/|x_2|$. Then $0 < t \leq 1$. Define

$$c_1(t) = \frac{|x_1|}{|x_2|} - \frac{|x_1|}{|x_2|}\frac{(|x_1|^p + |x_2|^p)|x_1|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})} = t - \frac{t^{2p-1} + t^{p-1}}{p(t^{2p-2} + 1)},$$

$$c_2(t) = 1 - \frac{(|x_1|^p + |x_2|^p)|x_2|^{p-2}}{p(|x_1|^{2p-2} + |x_2|^{2p-2})} = 1 - \frac{t^p + 1}{p(t^{2p-2} + 1)}.$$

Then

$$f(G_f(x_1, x_2)) = |x_2|^p(|c_1(t)|^p + |c_2(t)|^p). \tag{4.78}$$

Since $p - 2 \leq 0$, we have $t^{p-2} \geq 1$ and $t^p > 0$, hence

$$1 \geq c_2 \geq 1 - \frac{1 + t^p}{p\big(1 + t^p\big)} = 1 - \tfrac{1}{p} \geq 0. \tag{4.79}$$

Thus $c_2 \geq 0$. We now claim that

$$|c_1| + c_2 \leq 1. \tag{4.80}$$

This will implies $|c_i(t)| \leq 1$ $(i = 1, 2)$ so that $|c_i(t)|^p \leq |c_i(t)|$ as $p > 1$. Then from (4.78) we will have

$$f(G_f(x_1, x_2)) = |x_2|^p(|c_1(t)|^p + |c_2(t)|^p) \tag{4.81}$$

$$\leq |x_2|^p(|c_1(t)| + |c_2(t)|) \tag{4.82}$$

$$\leq |x_2|^p \tag{4.83}$$

$$\leq f(x_1, x_2). \tag{4.84}$$

Observe that (4.80) is equivalent to

$$c_1 + c_2 \leq 1 \tag{4.85a}$$
$$-c_1 + c_2 \leq 1 \tag{4.85b}$$

Since

$$c_1 + c_2 = t - \frac{t^{2p-1} + t^{p-1}}{p(t^{2p-2} + 1)} + 1 - \frac{t^p + 1}{p(t^{2p-2} + 1)}.$$

Thus (4.85a) is equivalent to

$$t - \frac{t^{2p-1} + t^{p-1}}{p(t^{2p-2} + 1)} + 1 - \frac{t^p + 1}{p(t^{2p-2} + 1)} \leq 1.$$

That is

$$t \leq \frac{(1 + t^p)(1 + t^{p-1})}{p(1 + t^{2p-2})}. \tag{4.86}$$

And (4.85b) is equivalent to

$$-c_1 + c_2 = -t + \frac{t^{2p-1} + t^{p-1}}{p(t^{2p-2} + 1)} + 1 - \frac{t^p + 1}{p(t^{2p-2} + 1)} \leq 1,$$

we have

$$\frac{t^{p-1}(1 + t^p)}{p(1 + t^{2p-2})} \leq t + \frac{1 + t^p}{p(1 + t^{2p-2})}. \tag{4.87}$$

Now check that (4.86) holds by using $t^{p-1} \leq 1$ and the convexity of $h\colon \xi \mapsto 1 + \xi^p$, which implies $h(t) \geq h(1) + h'(1)(t - 1)$, i.e., $pt \leq 1 + t^p$. Hence

$$t \leq \frac{t^p + 1}{p} \leq \frac{(t^p + 1)(t^{p-1} + 1)}{p(t^{2p-2} + 1)}.$$

Since $0 < t^{p-1} \leq 1$ when $0 < t \leq 1$and $1 < p \leq 2$, this implies that

$$\frac{t^{p-1}(t^p + 1)}{p(t^{2p-2} + 1)} \leq \frac{t^p + 1}{p(t^{2p-2} + 1)}$$

from which (4.87) follows. Therefore, we have proved that $|c_1| + c_2 \leq 1$ and hence f has decreasing property.
Furthermore, using (4.78), (4.79) and the assumption that $|x_2| \geq |x_1|$, we obtain

$$f(G_f(x)) \geq c_2^p |x_2|^p \geq \left(1 - \tfrac{1}{p}\right)^p |x_2|^p \geq \left(1 - \tfrac{1}{p}\right)^p \frac{|x_1|^p + |x_2|^p}{2} = \frac{\left(1 - \frac{1}{p}\right)^p}{2} f(x). \tag{4.88}$$

(iii): Consider the points $y = (1, \xi)$ and $z = (-1, \xi)$, where $\xi > 0$. Then $y - z = (2, 0)$ and

$$G_f(y) = \left(1 - \frac{1+\xi^p}{p(1+\xi^{2p-2})}, \xi - \frac{(1+\xi^p)\xi^{p-1}}{p(1+\xi^{2p-2})}\right) \tag{4.89a}$$

$$G_f(z) = \left(-1 + \frac{1+\xi^p}{p(1+\xi^{2p-2})}, \xi - \frac{(1+\xi^p)\xi^{p-1}}{p(1+\xi^{2p-2})}\right). \tag{4.89b}$$

It follows that

$$\langle G_f(y) - G_f(z), y - z\rangle = 4\left(1 - \frac{1+\xi^p}{p(1+\xi^{2p-2})}\right) < 0 \quad \text{as } \xi \to +\infty \tag{4.90}$$

because $\lim_{\xi\to+\infty} \frac{1+\xi^p}{p(1+\xi^{2p-2})} = \lim_{\xi\to+\infty}(2p-2)^{-1}\xi^{2-p} = +\infty$ using l'Hôpital's rule. Therefore, G_f is not monotone. □

4.11 G_f and the Yamagishi–Yamada operator

In this last section we study the accelerated version[16] of G_f proposed by Yamagishi and Yamada in [61]. Their operator, which has shown improved performance compared to G_f, is actually a subgradient projector of a variant of f when $\mathcal{H}$ is the real line[17]. For fixed $L > 0$ and $r > 0$, we assume in addition that

$$f \text{ is Fréchet differentiable and } \nabla f \text{ is Lipschitz continuous with constant } L, \tag{4.91}$$

and that

$$f \text{ is bounded below with } \inf f(\mathcal{H}) \geq -\rho \tag{4.92}$$

where $\rho > 0$ and we set

$$(\forall x \in \mathcal{H}) \quad \theta(x) = \frac{\|\nabla f(x)\|^2}{2L} - \rho. \tag{4.93}$$

By [61, Lemma 1], we have

$$f \geq \theta. \tag{4.94}$$

The Yamagishi–Yamada operator [61] is

$$Z\colon \mathcal{H} \to \mathcal{H}, \tag{4.95}$$

[16] See also [43] for another accelerated version of G_f.

[17] Unfortunately our proof does not seem to extend to $\mathcal{H}$ (e.g., the set D may not be convex and there is no obvious counterpart to (4.98)).

defined at $x \in \mathcal{H}$ by

$$Zx = \begin{cases} x, & \text{if } f(x) \leq 0; \\ x - \dfrac{\nabla f(x)}{\|\nabla f(x)\|^2} f(x), & \text{if } f(x) > 0 \text{ and } \theta(x) \leq 0; \\ x - \dfrac{\nabla f(x)}{\|\nabla f(x)\|^2} \Big(f(x) + \big(\sqrt{\theta(x)+\rho} - \sqrt{\rho}\big)^2 \Big), & \text{if } f(x) > 0 \text{ and } \theta(x) > 0. \end{cases} \tag{4.96}$$

Note that if $f(x) \leq 0$ or $\theta(x) \leq 0$, then $Zx = G_f(x)$.

We now prove that if $\mathcal{H} = \mathbb{R}$, then Z is itself a subgradient projector.

Theorem 4.44. *Suppose that $\mathcal{H} = \mathbb{R}$ and that f is also twice differentiable. Then for every $x \in \mathbb{R}$, (4.96) can be rewritten as*

$$Zx = \begin{cases} x, & \text{if } f(x) \leq 0; \\ x - \dfrac{1}{f'(x)} f(x), & \text{if } f(x) > 0 \text{ and } |f'(x)| \leq \sqrt{2L\rho}; \\ x - \dfrac{1}{f'(x)} \left(f(x) + \left(\dfrac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho} \right)^2 \right), & \text{if } f(x) > 0 \text{ and } |f'(x)| > \sqrt{2L\rho}. \end{cases} \tag{4.97}$$

Set $D = \{x \in \mathbb{R} : \theta(x) \leq 0\}$ and assume that $\operatorname{bdry} D \subseteq \mathbb{R} \smallsetminus C$. Then D is a closed convex superset of C, and Z is a subgradient projector of a function y, defined as follows. On D, we set y equal to f. The set $\mathbb{R} \smallsetminus D$ is empty, or an open interval, or the disjoint union of two open intervals. Assume that I is one of these nonempty intervals, and let q be defined on I such that

$$(\forall x \in I) \quad q'(x) = \frac{1}{x - Zx}. \tag{4.98}$$

Now set $d = P_D(I) \in D \smallsetminus C$ and

$$(\forall x \in I) \quad y(x) = \frac{f(d)}{e^{q(d)}} e^{q(x)}. \tag{4.99}$$

The so-constructed function $y \colon \mathbb{R} \to \mathbb{R}$ is convex, and it satisfies $Z = G_y$.

Proof. It is easy to check that (4.97) is the same as (4.96). Let $x \in \mathbb{R}$ such that $f(x) > 0$ and $\theta(x) \geq 0$, and set

$$z(x) = \frac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho} = \frac{\operatorname{sgn}\big(f'(x)\big) f'(x)}{\sqrt{2L}} - \sqrt{\rho} = \sqrt{\theta(x)+\rho} - \sqrt{\rho} \geq 0. \tag{4.100}$$

Then

$$z'(x) = \frac{\operatorname{sgn}\big(f'(x)\big) f''(x)}{\sqrt{2L}}. \tag{4.101}$$

Using the convexity of f, (4.94), (4.100), and (4.101), we obtain

$$\begin{aligned}
0 &\leq f''(x)\big(f(x) - \theta(x)\big) && \text{(4.102a)}\\
&= f''(x)\left(f(x) - \left(\frac{|f'(x)|}{\sqrt{2L}} + \sqrt{\rho}\right)\left(\frac{|f'(x)|}{\sqrt{2L}} - \sqrt{\rho}\right)\right) && \text{(4.102b)}\\
&= f''(x)\left(f(x) + z(x)\left(z(x) - \frac{2|f'(x)|}{\sqrt{2L}}\right)\right) && \text{(4.102c)}\\
&= f''(x)\left(f(x) + z^2(x) - 2z(x)\frac{|f'(x)|}{\sqrt{2L}}\right) && \text{(4.102d)}\\
&= f''(x)\left(f(x) + z^2(x)\right) - 2f''(x)z(x)\frac{|f'(x)|}{\sqrt{2L}} && \text{(4.102e)}\\
&= f''(x)\left(f(x) + z^2(x)\right) - 2f''(x)z(x)\frac{\operatorname{sgn}\big(f'(x)\big) f'(x)}{\sqrt{2L}} && \text{(4.102f)}\\
&= f''(x)\big(f(x) + z^2(x)\big) - f'(x)2z(x)\frac{\operatorname{sgn}\big(f'(x)\big) f''(x)}{\sqrt{2L}} && \text{(4.102g)}\\
&= f''(x)\big(f(x) + z^2(x)\big) - f'(x)\big(2z(x)z'(x)\big). && \text{(4.102h)}
\end{aligned}$$

Because $x - Zx = \frac{f(x)+z^2(x)}{f'(x)}$ is continuous, it is clear that there is an antiderivative q on I such that

$$q'(x) = \frac{1}{x - Zx} = \frac{f'(x)}{f(x) + z^2(x)}. \tag{4.103}$$

Calculus and (4.102) now result in

$$\begin{aligned}
q''(x) &= \frac{f''(x)\big(f(x) + z^2(x)\big) - f'(x)\big(f'(x) + 2z(x)z'(x)\big)}{\big(f(x) + z^2(x)\big)^2} && \text{(4.104a)}\\
&= \frac{f''(x)\big(f(x) - \theta(x)\big) - \big(f'(x)\big)^2}{\big(f(x) + z^2(x)\big)^2}. && \text{(4.104b)}
\end{aligned}$$

Observe that y is clearly continuous everywhere. Furthermore, $y'(x) =$

$\frac{f(d)}{e^{q(d)}}e^{q(x)}q'(x)$ and hence, using (4.103), (4.104) and again (4.102), we obtain

$$y''(x) = \frac{f(d)}{e^{q(d)}}\Big(e^{q(x)}\big(q'(x)\big)^2 + e^{q(x)}q''(x)\Big) \tag{4.105}$$

$$= \frac{f(d)}{e^{q(d)}}e^{q(x)}\Big(\big(q'(x)\big)^2 + q''(x)\Big) \tag{4.106}$$

$$= y(x)\frac{f''(x)\big(f(x)-\theta(x)\big)}{\big(f(x)+z^2(x)\big)^2} \tag{4.107}$$

$$\geq 0. \tag{4.108}$$

Hence y is convex on I. As $x \in I$ approaches d, we deduce (because $d \notin C$, i.e., $f(d) > 0$) that $q'(x) \to \frac{f'(d)}{f(d)+z^2(d))} = \frac{f'(d)}{f(d)}$ and hence that $y'(x) \to \frac{f(d)}{e^{q(d)}}e^{q(d)}\frac{f'(d)}{f(d)} = f'(d)$. It follows that y is convex on $\mathbb{R}$. Finally, if $x \notin D$, then $G_y(x) = x - y(x)/y'(x) = x - 1/q'(x) = x - (x - Zx) = Zx$. □

Example 4.45. Consider Theorem 4.44 and assume that $f\colon x \mapsto x^2 - 1$, that $L = 3$, and that $\rho = 1$. Then (4.97) turns into

$$Zx = \begin{cases} x, & \text{if } |x| \leq 1; \\ \dfrac{x^2+1}{2x}, & \text{if } 1 < |x| \leq \sqrt{6}/2; \\ \dfrac{x^2+2\sqrt{6}|x|}{6x}, & \text{if } |x| > \sqrt{6}/2. \end{cases} \tag{4.109}$$

Hence $D = \big[-\sqrt{6}/2, \sqrt{6}/2\big]$. Using elementary manipulations, we obtain

$$(\forall x \in \mathbb{R} \smallsetminus D) \quad q(x) = \tfrac{6}{5}\ln\big(\tfrac{5}{6}|x| - \tfrac{\sqrt{6}}{3}\big); \tag{4.110}$$

Indeed, since $q'(x)$ is defined in I such that $q'(x) = \frac{1}{x-Zx}$.

$$\begin{aligned} q'(x) &= \frac{1}{x - Zx} \\ &= \frac{1}{x - \frac{x^2+2\sqrt{6}|x|}{6x}} \\ &= \frac{6x}{6x^2 - x^2 - 2\sqrt{6}|x|} \\ &= \frac{6x}{5x^2 - 2\sqrt{6}|x|} \\ &= \frac{1}{\frac{5x}{6} - \frac{\sqrt{6}}{3}\frac{|x|}{x}}. \end{aligned}$$

When $x < -\frac{\sqrt{6}}{2}$, then $q'(x) = \frac{1}{\frac{5x}{6}+\frac{\sqrt{6}}{3}}$. When $x > \frac{\sqrt{6}}{2}$, we have $q'(x) = \frac{1}{\frac{5x}{6}-\frac{\sqrt{6}}{3}}$. Applying integration, we have

$$\begin{aligned} q(x) &= \begin{cases} \frac{6}{5}\ln\left|\frac{5}{6}x+\frac{\sqrt{6}}{3}\right|, & \text{if } x < -\frac{\sqrt{6}}{2}; \\ \frac{6}{5}\ln\left|\frac{5}{6}x-\frac{\sqrt{6}}{3}\right|, & \text{if } x > \frac{\sqrt{6}}{2}. \end{cases} \\ &= \begin{cases} \frac{6}{5}\ln\left(-\frac{5}{6}x-\frac{\sqrt{6}}{3}\right), & \text{if } x < -\frac{\sqrt{6}}{2}; \\ \frac{6}{5}\ln\left(\frac{5}{6}x-\frac{\sqrt{6}}{3}\right), & \text{if } x > \frac{\sqrt{6}}{2}. \end{cases} \end{aligned}$$

Thus, we have

$$(\forall x \in \mathbb{R} \setminus D) \quad q(x) = \tfrac{6}{5}\ln\left(\tfrac{5}{6}|x| - \tfrac{\sqrt{6}}{3}\right); \tag{4.111}$$

By definition of function y in Theorem 4.44: on D, $y = f$; on $\mathbb{R} \setminus D$, I is one of two open intervals which are disjoint included in $\mathbb{R} \setminus D$,

$$(\forall x \in I) \quad y(x) = \frac{f(d)}{e^{q(d)}} e^{q(x)}.$$

Recall that $D = \left[\frac{-\sqrt{6}}{2}, \frac{\sqrt{6}}{2}\right]$. Then we have $y(x) = x^2 - 1$ when $|x| \leq \frac{\sqrt{6}}{2}$. Now when $|x| > \frac{\sqrt{6}}{2}$, we have $x \in I$, thus $d = P_D(x) = \left|\frac{\sqrt{6}}{2}\right|$.

$$\begin{aligned} y(x) &= \frac{f(d)}{e^{q(d)}} e^{q(x)} \\ &= \frac{\frac{6}{4}-1}{e^{\ln\left(\frac{5}{6}\frac{\sqrt{6}}{2}-\frac{\sqrt{6}}{3}\right)^{\frac{6}{5}}}} e^{\ln\left(\frac{5}{6}|x|-\frac{\sqrt{6}}{3}\right)^{\frac{6}{5}}} \\ &= \frac{\frac{1}{2}}{\left(\frac{\sqrt{6}}{12}\right)^{\frac{6}{5}}} \left(\frac{5}{6}|x| - \frac{\sqrt{6}}{3}\right)^{\frac{6}{5}} \\ &= \frac{\frac{1}{2}}{\left(\frac{\sqrt{6}}{12}\right)^{\frac{6}{5}}} \left(\frac{5|x| - 2\sqrt{6}}{6}\right)^{\frac{6}{5}} \\ &= \frac{1}{2}\frac{1}{\left(\frac{\sqrt{6}}{12}\right)^{\frac{6}{5}}} \left(\frac{1}{6}\right)^{\frac{6}{5}} \left(5|x| - 2\sqrt{6}\right)^{\frac{6}{5}} \\ &= \frac{1}{2}\left(\frac{2}{\sqrt{6}}\right)^{\frac{6}{5}} \left(5|x| - 2\sqrt{6}\right)^{\frac{6}{5}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{2^{-1}2^{\frac{6}{5}}}{6^{\frac{3}{5}}}\left(5|x|-2\sqrt{6}\right)^{\frac{6}{5}} \\
&= \frac{2^{\frac{1}{5}}6^{\frac{2}{5}}}{6}\left(5|x|-2\sqrt{6}\right)^{\frac{6}{5}} \\
&= \frac{2^{\frac{1}{5}}(2\cdot 3)^{\frac{2}{5}}}{6}\left(5|x|-2\sqrt{6}\right)^{\frac{6}{5}} \\
&= \frac{(2\cdot 36)^{\frac{1}{5}}}{6}\left(5|x|-2\sqrt{6}\right)^{\frac{6}{5}}.
\end{aligned}
$$

Thus we have $y(x) = \frac{72^{1/5}}{6}\big(5|x|-2\sqrt{6}\big)^{6/5}$. Consequently, the function y, given by

$$
(\forall x \in \mathbb{R}) \quad y(x) = \begin{cases} x^2-1, & \text{if } |x| \le \sqrt{6}/2; \\ \dfrac{72^{1/5}}{6}\big(5|x|-2\sqrt{6}\big)^{6/5}, & \text{if } |x| > \sqrt{6}/2, \end{cases} \tag{4.112}
$$

satisfies $G_y = Z$ by Theorem 4.44.

This chapter gave a comprehensive study of properties of the subgradient projector, which included calculus rules, characterization of strong-to-strong and strong-to-weak continuity, nonexpansiveness, monotonicity and the decreasing property. We also considered the Yamagishi-Yamada operator and discovered it is a subgradient projector of a different convex function.

Chapter 5

Characterizations of Subgradient Projectors

5.1 Motivation

This chapter is based on the manuscript [10]. The results in this chapter extend some of those found in previous chapters. Chapter 4 mainly focuses on the subgradient projector of a convex function. In this chapter, we want to explore the subgradient projector of nonconvex functions. We investigate when a subgradient projector is a cutter or when a mapping is a subgradient projector of a convex function. Moreover, we present calculus rules for the subgradient projector of nonconvex functions.

Studies of optimization problems and convex feasibility problems have led in recent years to the development of a theory of subgradient projectors. Rather than finding projections on the level sets of the original functions, the iterative algorithms find projection on half spaces containing the former. Polyak developed the subgradient projector algorithms for convex functions [48–50]; this was further developed by Censor, Combettes, Yamada and others, and also applied to optimization problems [19, 21, 22, 29, 30, 44, 58].

We want to extend the basic theory of subgradient projectors to nonconvex functions. As far as nonconvex functions are concerned, the cutter theory or $\mathcal{T}$-class operators developed by Cegielski [20] or Bauschke, Borwein and Combettes [3] furnishes a new approach to subgradient projectors without appealing to the existence theory of subgradient projectors for convex functions.

5.2 Preliminary and definitions

Throughout the following two chapters, we assume

$$X = \mathbb{R}^n$$

and $f : X \to]-\infty, +\infty]$ is a lower semicontinuous function, its 0-level set is denoted by

$$\mathrm{lev}_0 f = \{x \in X : f(x) \leq 0\}. \tag{5.1}$$

$T : X \to X$ is a mapping.

To introduce subgradient projectors for possible non-convex functions, we need the following subgradients [40, 42, 53].

Definition 5.1. Suppose that $\bar{x} \in X$ such that $f(\bar{x})$ is finite. For a vector $v \in X$, one says that

(i) v is a *regular subgradient* of f at $\bar{x}$, written $v \in \hat{\partial} f(\bar{x})$, if

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|), \tag{5.2}$$

(ii) v is a *limiting (or Mordukhovich) subgradient* of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if there are sequences $x_n \to \bar{x}$, $f(x_n) \to f(\bar{x})$ and $v_n \in \hat{\partial} f(x_n)$ with $v_n \to v$.

(iii) f is *subdifferentially regular* at $\bar{x}$ if $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$.

Remark 5.2. The regular subgradient inequality (5.2) in o form is shorthand for the one-side limit condition

$$\liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0. \tag{5.3}$$

This is because if v satisfies (5.2), which is equivalent to

$$f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle \geq o(\|x - \bar{x}\|), \tag{5.4}$$

Dividing $\|x - \bar{x}\|$ on both sides of above inequality and applying the definition of o form, we have (5.3). Conversely, suppose that v satisfies (5.3). Then by definition of limit inferior, we have

$$\frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq -\varepsilon, \tag{5.5}$$

where ε is any positive number and x sufficiently close to $\bar{x}$. Simplifying this inequality, we have

$$f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle \geq -\varepsilon \|x - \bar{x}\|. \tag{5.6}$$

Let $\phi(x) = \min\{f(x) - f(\bar{x}) - \langle v, x - \bar{x}\rangle, 0\}$. Obviously, we have $\phi(x) \leq 0$. From (5.6) we have

$$\phi(x) = \min\{f(x) - f(\bar{x}) - \langle v, x - \bar{x}\rangle, 0\} \geq \min\{-\varepsilon\|x - \bar{x}\|, 0\}. \tag{5.7}$$

Dividing $\|x - \bar{x}\|$ on both sides of above equation, we have

$$\frac{\phi(x)}{\|x - \bar{x}\|} \geq -\varepsilon. \tag{5.8}$$

So we have

$$\liminf_{x \to \bar{x}} \frac{\phi(x)}{\|x - \bar{x}\|} \geq -\varepsilon. \tag{5.9}$$

Letting $\varepsilon \to 0$, then

$$\liminf_{x \to \bar{x}} \frac{\phi(x)}{\|x - \bar{x}\|} \geq 0. \tag{5.10}$$

Moreover, $\phi(x) \leq 0$, so

$$\limsup_{x \to \bar{x}} \frac{\phi(x)}{\|x - \bar{x}\|} \leq 0. \tag{5.11}$$

Then we have $\lim_{x \to \bar{x}} \frac{\phi(x)}{\|x - \bar{x}\|} = 0$. Thus (5.2) holds.

Next, we provide an example and show how to calculate corresponding subdifferentials for a specific function.

Example 5.3. Consider function $f(x) = -|x|$ on $\mathbb{R}$. Then

(i) $\hat{\partial} f(0) = \varnothing$.

(ii) $\hat{\partial} f(\bar{x}) = \{-1\}$ if $\bar{x} > 0$; $\hat{\partial} f(\bar{x}) = \{1\}$ if $\bar{x} < 0$.

(iii) $\partial f(x) = \{-1\}$ if $x > 0$; $\partial f(x) = \{1\}$ if $x < 0$.

(iv) $\partial f(0) = \{-1, 1\}$.

Proof. (i) We prove it by contradiction. Assume there exists $v \in \hat{\partial} f(0)$ such that v satisfying (5.2). Then we have

$$\liminf_{x \to 0, x \neq 0} \frac{-|x| - vx}{|x|} = \liminf_{x \to 0, x \neq 0} (-1 - v\,\mathrm{sgn}(x)) \geq 0. \tag{5.12}$$

When $x \to 0^+$, v must satisfy $v \leq -1$ such that (5.12) holds; when $x \to 0^-$, v has to be greater or equal to 1 such that (5.12) holds. Hence here is no proper v such that (5.12) holds for any $x \in \mathbb{R}$. Therefore, $\hat{\partial} f(0) = \varnothing$.

(ii) Let $\bar{x} < 0$ and $v \in \hat{\partial} f(\bar{x})$. Then we have

$$\liminf_{x \to \bar{x}, x \neq \bar{x}} \left(\frac{|\bar{x}| - |x|}{|x - \bar{x}|} - v \operatorname{sgn}(x - \bar{x}) \right) \geq 0. \tag{5.13}$$

When $x \to \bar{x}^-$, we have $v \geq 1$; when $x \to \bar{x}^+$, we have $v \leq 1$. Therefore $\hat{\partial} f(\bar{x}) = \{1\}$. The proof is similar when $\bar{x} > 0$.

(iii) It is clear from definition of limiting subdifferential.

(iv) From (ii), we can see when $x_n \to 0^+$, $v_n = -1 \in \hat{\partial} f(x_n)$ and $x_n \to 0^-$, $v_n = 1 \in \hat{\partial} f(x_n)$. There are just two sides: $\lim v_n = -1$ and $\lim v_n = 1$. Thus $\partial f(0) = \{-1, 1\}$.

□

Definition 5.4. [53, Definition 9.1] We say $f : X \to]-\infty, +\infty[$ is *locally Lipschitz continuous* near a point $x \in X$ if $\exists \rho \in \mathbb{R}_{++}$ and $L \in \mathbb{R}_+$ such that

$$|f(y) - f(z)| \leq L\|y - z\|, \tag{5.14}$$

for any $y, z \in \operatorname{ball}(x, \rho)$.

Remark 5.5. [53, Corollary 8.10] When f is lower semi-continuous, the set of points at which ∂f is nonempty-valued is at least dense in the domain of f. When f is locally Lipschitz, ∂f is nonempty-valued everywhere.

Remark 5.6. When function f is convex, recall v is a subgradient of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle. \tag{5.15}$$

We use the same notation since all subdifferentials are the same for convex functions. (For more details, see [40, Theorem 1.93].)

Definition 5.7. For a general lower semicontinuous function $f : X \to]-\infty, +\infty]$, the *subgradient projector* of f is defined by

$$G_f : X \to X : x \mapsto \begin{cases} x - \frac{f(x)}{\|s(x)\|^2} s(x) & \text{if } f(x) > 0 \text{ and } 0 \notin \partial f(x), \\ x & \text{otherwise,} \end{cases} \tag{5.16}$$

where $s : X \to X$ is a selection of ∂f with $s(x) \in \partial f(x)$.

Proposition 5.8. *(i) When f is continuously differentiable on $X \setminus lev_0 f$, G_f reduces to*

$$G_f : X \to X : x \mapsto \begin{cases} x - \frac{f(x)}{\|\nabla f(x)\|^2} \nabla f(x) & \textit{if } f(x) > 0 \textit{ and } \nabla f(x) \neq 0, \\ x & \textit{otherwise.} \end{cases} \tag{5.17}$$

(ii) When f is convex and $\inf_X f \leq 0$, G_f reduces to

$$G_f : X \to X : x \mapsto \begin{cases} x - \frac{f(x)}{\|s(x)\|^2} s(x) & \text{if } f(x) > 0, \\ x & \text{otherwise,} \end{cases} \tag{5.18}$$

where $s : X \to X$ is a selection of ∂f with $s(x) \in \partial f(x)$. Referring to Lemma 3.13, $f(x) > 0$ automatically implies $s(x) \neq 0$.

Recall that the projection on half space:

Fact 5.9. *([7] or [20, page 133]) For the half space*

$$H(a, \beta) := \{z \in X : \langle a, z \rangle \leq \beta\}, \tag{5.19}$$

where $a \in X$, $a \neq 0$ and $\beta \in \mathbb{R}$, its metric projection is given by

$$P_{H(a,\beta)} x = \begin{cases} x - \frac{\langle a, x \rangle - \beta}{\|a\|^2} a & \text{if } \langle a, x \rangle > \beta, \\ x & \text{if } \langle a, x \rangle \leq \beta. \end{cases} \tag{5.20}$$

Proposition 5.10. *(i) For a locally Lipschitz $f : X \to \mathbb{R}$, when $f(x) > 0$, $0 \notin \partial f(x)$, we have*

$$G_f(x) = P_{H(s(x), \beta(x))}(x) \tag{5.21}$$

where $s(x) \in \partial f(x)$ and the half space

$$H(s(x), \beta(x)) : = \{z \in X : f(x) + \langle s(x), z - x \rangle \leq 0\} \tag{5.22}$$

$$= \{z \in X : \langle s(x), z \rangle \leq \langle s(x), x \rangle - f(x)\}. \tag{5.23}$$

(ii) If $f : X \to \,]-\infty, +\infty]$, then

$$\operatorname{Fix} G_f = \{x \in X : 0 \in \partial f(x)\} \cup lev_0 f. \tag{5.24}$$

If f is locally Lipschitz, $\operatorname{Fix} G_f$ is closed.

(iii) If $f : X \to \,]-\infty, +\infty]$ is convex and $\inf_X f \leq 0$, then

$$\operatorname{Fix} G_f = lev_0 f. \tag{5.25}$$

Proof. (i): Apply Fact 5.9 with $a := s(x), \beta := \beta(x) = \langle s(x), x \rangle - f(x)$.

(ii): This follows from the definition of G_f. When f is locally Lipschitz, ∂f is outer-semicontinuous, so $\{x \in X : 0 \in \partial f(x)\}$ is closed.[18] Being a union of two closed sets, $\operatorname{Fix} G$ is closed.

[18] (See [53]) For the definition of outer-semicontinuity of subdifferentials (page 151). For the relevant result, see Proposition 8.7 on page 302.

(iii): When f is convex, $0 \in \partial f(x)$ gives $f(x) = \min_X f$, so $f(x) \leq 0$. Then

$$\{x \in X : 0 \in \partial f(x)\} \subset \mathrm{lev}_0 f. \tag{5.26}$$

Thus (iii) follows from (ii). □

Fact 5.11. *[46, Proposition 1.6] If the convex function f is continuous, then it is locally Lipschitz. Thus Proposition 5.10 is also true for convex functions.*

A simple example illustrates the difference between convex and nonconvex case.

Example 5.12. Consider $f : X \to \mathbb{R}$ by $f(x) = \sqrt[k]{\|x\|} = \|x\|^{1/k}$ where $k > 0$. Then

(i) When $k \leq 1$, f is convex, $G_f = (1-k)\,\mathrm{Id}$ is firmly nonexpansive.

(ii) When $k > 1$, f is not convex, $G_f = (1-k)\,\mathrm{Id}$ may be neither monotone nor nonexpansive, e.g, $k = 3$.

Proof. This follows from definition directly or use $G_f = (1-k)\,\mathrm{Id} + kG_{\|\cdot\|}$ by Theorem 5.27 and $G_{\|\cdot\|} \equiv 0$. Indeed,
(i): If $0 < k \leq 1$. We want to show function f is convex. Define $h(t) = t^{\frac{1}{k}}$ where $t \in \mathbb{R}$. Then $f(x) = h(\|x\|)$. Since $h'(t) = \frac{1}{k}t^{\frac{1}{k}-1} > 0$ and $h''(t) = \frac{1}{k}(\frac{1}{k}-1)t^{\frac{1}{k}-2} > 0$. Thus function h as defined is an increasing convex function on the real line. Moreover, $\|\cdot\|$ is also convex. By [41, Proposition 1.39], we have f is convex. Notice that

$$\|G_f(x) - G_f(y)\|^2 = \|(1-k)x - (1-k)y\|^2 \tag{5.27}$$
$$= \|(1-k)(x-y)\|^2 \tag{5.28}$$
$$= (1-k)^2\|x-y\|^2, \tag{5.29}$$

and

$$\langle x-y, G_f(x) - G_f(y)\rangle = \langle x-y, (1-k)(x-y)\rangle \tag{5.30}$$
$$= (1-k)\|x-y\|^2. \tag{5.31}$$

Since $0 < k \leq 1$, then $(1-k)^2 \leq (1-k)$, thus

$$\|G_f(x) - G_f(y)\|^2 \leq \langle x-y, G_f(x) - G_f(y)\rangle. \tag{5.32}$$

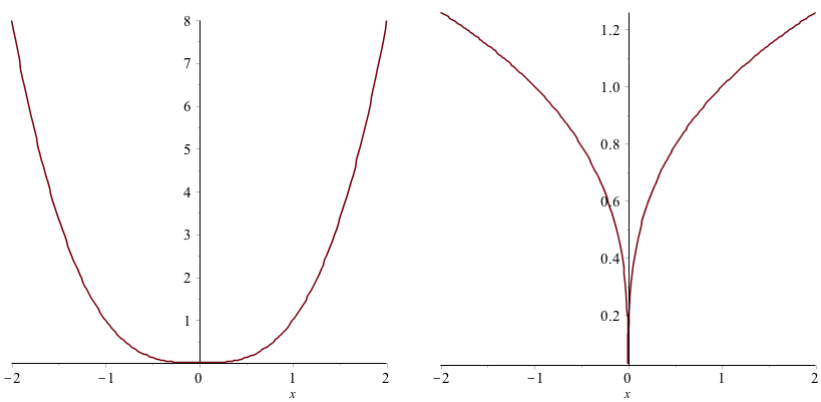
Hence G_f is firmly nonexpansive.
(ii) Define $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(t) = f(td)$ where d is a direction with $d \neq 0$. Then $\phi(t) = |t|^{\frac{1}{k}}\|d\|^{\frac{1}{k}}$. Since $\|d\|^{\frac{1}{k}}$ is a constant, we assume $g(t) = |t|^{\frac{1}{k}}$. Then $g'(t) = \frac{1}{k}|t|^{\frac{1}{k}-2}t$ and $g''(t) = \frac{1}{k}|t|^{\frac{1}{k}-2}(\frac{1}{k} - 1) < 0$ when $k > 1$. Thus f is not convex. Fig. 5.1(b) shows that f is not convex. For $x, y \in X$, when $k > 1$, we have

$$\langle x-y, G_f(x)-G_f(y)\rangle = \langle x-y, (1-k)(x-y)\rangle = (1-k)\|x-y\|^2 < 0. \quad (5.33)$$

Then G_f is not monotone when $k > 1$.
For $x, y \in X$,

$$\|G_f(x) - G_f(y)\| = \|(1-k)(x-y)\| = |1-k|\|x-y\|. \quad (5.34)$$

When $k > 2$, it is obvious that G_f is not nonexpansive. □



(a) Graph of f is defined when $k = \frac{1}{3}$

(b) Graph of f is given when $k = 3$.

Figure 5.1: Illustration for Example 5.12

5.3 Calculus for subgradient projectors

We start with a stronger differentiability than Fréchet differentiability (see Definition 2.62).

Definition 5.13. [40, Definition 1.13] A function $f : X \to]-\infty, +\infty]$ is *strictly differentiable* at $\bar{x} \in X$ if

$$\lim_{x \to \bar{x}, u \to \bar{x}} \frac{f(x) - f(u) - \nabla f(\bar{x})(x - u)}{\|x - u\|} = 0, \tag{5.35}$$

where $\nabla f(\bar{x})$ is gradient of f at $\bar{x}$.

It is clear that when $u = \bar{x}$, Definition 5.13 coincides with definition of Fréchet differentiability. However, Fréchet differentiable functions are not always strictly differentiable, see the following example.

Example 5.14. A function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is rational,} \\ 0 & \text{otherwise.} \end{cases} \tag{5.36}$$

Then f is Fréchet differentiable but not strictly differentiable at $\bar{x} = 0$.

Proof. Since $\bar{x} = 0$ is a rational number, so $f'(\bar{x}) = 2\bar{x} = 0$.

$$0 \leq \frac{f(y)}{|y|} \leq |y|. \tag{5.37}$$

Then no matter whether y is a rational or an irrational number, we still have

$$\lim_{0 \neq |y| \to 0} \frac{f(y)}{|y|} = 0, \tag{5.38}$$

thus f is Fréchet differentiable at $\bar{x} = 0$. It remains to show that f is not strictly differentiable at $\bar{x} = 0$. Let $x > 0$ be rational and set $u = -\pi x^2 + x \notin \mathbb{Q}$. Note that if $x \to 0^+$, then $u \to 0^+$. Furthermore, since $x - u = \pi x^2 > 0$, we have

$$\frac{f(x) - f(u) - 0 \cdot (x - u)}{|x - u|} = \frac{x^2}{|x - u|} = \frac{x^2}{\pi x^2} = \frac{1}{\pi} \not\to 0 \tag{5.39}$$

as $x \to 0^+$. Therefore, f is not strictly differentiable at $\bar{x} = 0$. □

The following facts are crucial to study the calculus of subgradient projectors.

Fact 5.15. *Assume that $\Phi : X \to X$ is locally Lipschitz at $\bar{x} \in X$, and $\phi : X \to \mathbb{R}$ be strictly differentiable at $\Phi(\bar{x})$. Then for $m(x) = \phi(\Phi(x))$, one has*

$$\partial m(\bar{x}) = \partial \langle \phi'(\bar{y}), \Phi \rangle (\bar{x}) \text{ with } \bar{y} = \Phi(\bar{x}). \tag{5.40}$$

Proof. See [42, Theorem 6.5]. □

Fact 5.16. *[53, Theorem 10.6] Suppose $f(x) = g(F(x))$ for a proper lower semicontinuous $g : X \to]-\infty, +\infty]$ and a strictly differentiable $F : X \to X$, and let $\bar{x}$ be a point where f is finite and the Jacobian $\nabla F(\bar{x})$ has full rank. Then*

$$\partial f(\bar{x}) = \nabla F(\bar{x})^* \partial g(\bar{y}) \text{ with } \bar{y} = F(\bar{x}). \tag{5.41}$$

Fact 5.17. *Let $f_j : X \to \mathbb{R}$ be locally Lipschitz at $\bar{x}$ and $J(\bar{x}) = \{j : f_j(\bar{x}) = \max\{f_1, f_2\}(\bar{x})\}$. Then*

$$\partial \max\{f_1, f_2\}(\bar{x}) \subseteq \operatorname{conv}\{\partial f_j(\bar{x}) : \ j \in J(\bar{x})\}, \tag{5.42}$$

where the equality holds and $\max\{f_1, f_2\}$ is subdifferentially regular at $\bar{x}$ if the functions f_j is subdifferentially regular at $\bar{x}$ for $j \in J(\bar{x})$.

Proof. See [42, Theorem 7.5(ii)]. □

These facts allow us to establish a calculus of subgradient projectors.

Fact 5.18. *Let $f : X \to \mathbb{R}$ be a lower semicontinuous function.*

(i) If $k > 0$ then $G_{kf} = G_f$.

(ii) Let $\alpha \in \mathbb{R}$. Define $G_{f,\alpha} : X \to X$ as follow:

$$G_{f,\alpha}(x) := \begin{cases} x - \frac{f(x)-\alpha}{\|s(x)\|^2} s(x) & \text{if } f(x) > \alpha \text{ and } 0 \notin \partial f(x) \\ x & \text{otherwise.} \end{cases} \tag{5.43}$$

Then $G_{f,\alpha} = G_{f-\alpha}$.

(iii) Let $\alpha > 0$. Then

$$G_{f-\alpha}(x) = \begin{cases} G_f(x) + \frac{\alpha s(x)}{\|s(x)\|^2} & \text{if } f(x) > \alpha \text{ and } 0 \notin \partial f(x) \\ x & \text{otherwise.} \end{cases} \tag{5.44}$$

Proof. (i) By [53, D. Rescaling 10(6)], we have $\partial(kf) = k\partial f$ when $k > 0$. Note that $kf(x) > 0$ if and only if $f(x) > 0$, and $0 \notin \partial(kf)(x)$ if and only if $0 \notin \partial f(x)$. When $kf(x) > 0$ and $0 \notin \partial(kf)(x)$, for $s(x) \in \partial f(x)$, we have $ks(x) \in k\partial f(x) = \partial(kf)(x)$ so that

$$G_{kf}(x) = x - \frac{kf(x)}{\|ks(x)\|^2} ks(x) = x - \frac{f(x)}{\|s(x)\|^2} s(x) = G_f(x). \tag{5.45}$$

When $kf(x) \leq 0$ or $0 \in \partial(kf)(x)$, we have $f(x) \leq 0$ or $0 \in \partial f(x)$, so $G_{kf}(x) = x = G_f(x)$.

(ii) It suffices to note that $\partial(f - \alpha) = \partial f$. Indeed, let $\Phi = f$ and $\phi(x) = x - \alpha$. Applying Fact 5.15, we have $\partial(f - \alpha)(x) = \partial(\phi(\Phi))(x) = \partial\langle\phi'(y), \Phi\rangle(x) = \partial\Phi(x) = \partial f(x)$, where $y = \Phi(x)$.

(iii) When $f(x) > \alpha > 0$ and $0 \notin \partial f(x)$, we have $f(x) > 0$ and $0 \notin \partial f(x)$. For $s(x) \in \partial f(x)$, we have $s(x) \in \partial(f - \alpha)(x)$. Then

$$G_{f-\alpha}(x) = x - \frac{f(x) - \alpha}{\|s(x)\|^2} s(x) \tag{5.46}$$

$$= x - \frac{f(x)}{\|s(x)\|^2} s(x) + \frac{\alpha}{\|s(x)\|^2} s(x) \tag{5.47}$$

$$= G_f(x) + \frac{\alpha}{\|s(x)\|^2} s(x) \tag{5.48}$$

When $f(x) \leq \alpha$ or $0 \in \partial f(x)$, $G_{f-\alpha}(x) = x$ by definition. □

Remark 5.19. Fact 5.18 (i) is also true when f is convex. Some literature has already contained this result, like[12, Proposition 3.1 (i)] and [44, Proposition 2.1].

Proposition 5.20. *Assume that $f_1, f_2 : X \to \mathbb{R}$ are locally Lipschitz and subdifferentially regular. For the maximum function $g = \max(f_1, f_2)$, one has*

$$G_g(x) = \begin{cases} G_{f_1}(x) & \text{if } g(x) > \max(f_2(x), 0),\ 0 \notin \partial f_1(x), \\ G_{f_2}(x) & \text{if } g(x) > \max(f_1(x), 0),\ 0 \notin \partial f_2(x), \\ s(x) & \text{if } g(x) = f_1(x) = f_2(x) > 0,\ 0 \notin \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}, \\ x & \text{if } g(x) \leq 0, \text{ or } 0 \in \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\} \end{cases} \tag{5.49}$$

where

$$s(x) = x - \frac{f_i(x)}{\|v\|^2} v \tag{5.50}$$

with $v \in \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}$. In particular, when $g(x) = f_1(x) = f_2(x) > 0$ and $0 \notin \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}$, one may choose $s(x) = G_{f_1}(x)$ or $G_{f_2}(x)$. Consequently, one may think G_g as two piecewise defined function, namely,

$$G_g(x) = \begin{cases} G_{f_1}(x) & \text{if } g(x) > \max(f_2(x), 0),\ 0 \notin \partial f_1(x), \\ G_{f_2}(x) & \text{if } g(x) > \max(f_1(x), 0),\ 0 \notin \partial f_2(x), \\ G_{f_1}(x) & \text{if } g(x) = f_1(x) = f_2(x) > 0,\ 0 \notin \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}, \\ x & \text{if } g(x) \leq 0, \text{ or } 0 \in \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}, \end{cases} \tag{5.51}$$

or

$$G_g(x) = \begin{cases} G_{f_1}(x) & \text{if } g(x) > \max(f_2(x), 0),\ 0 \notin \partial f_1(x), \\ G_{f_2}(x) & \text{if } g(x) > \max(f_1(x), 0),\ 0 \notin \partial f_2(x), \\ G_{f_2}(x) & \text{if } g(x) = f_1(x) = f_2(x) > 0,\ 0 \notin \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}, \\ x & \text{if } g(x) \leq 0 \text{ or } 0 \in \operatorname{conv}\{\partial f_1(x), \partial f_2(x)\}. \end{cases} \tag{5.52}$$

Proof. When $g(x) > 0$, we consider three cases: (i) $g(x) > f_2(x)$; (ii) $g(x) > f_1(x)$; (iii) $g(x) \leq \min(f_2(x), f_1(x))$ which is $g(x) = f_1(x) = f_2(x)$. Also note that

$$\partial g(x) = \operatorname{conv}(\partial f_1(x), \partial f_2(x)) \tag{5.53}$$

when $f_1(x) = f_2(x)$ by Fact 5.17. □

Proposition 5.21. *Assume that $k \in \mathbb{R} \setminus \{0\}$, and $g(x) = f(kx)$. Then*

$$G_g(x) = \frac{1}{k} G_f(kx) \tag{5.54}$$

for every $x \in X$. Moreover, $\operatorname{Fix} G_g = \frac{1}{k} \operatorname{Fix} G_f$.

Proof. By Fact 5.16, $\partial g(x) = k\partial f(y)$ where $y = kx$. Indeed, $\partial g(x) = \partial\langle k, f\rangle(y) = k\partial f(y)$ where $y = kx$.

And $0 \notin \partial g(x)$ if and only if $0 \notin \partial f(y)$ with $y = kx$. When $g(x) > 0$ and $0 \notin \partial g(x)$, we have $f(kx) > 0$ and $0 \notin \partial f(kx)$. Suppose $s(y) \in \partial f(y)$ with $y = kx$. Then $ks(y) \in k\partial f(y) = \partial g(x)$. Therefore

$$G_g(x) = x - \frac{f(kx)}{\|ks(y)\|^2} ks(y) = x - \frac{1}{k} \frac{f(kx)}{\|s(y)\|^2} s(y) \tag{5.55}$$

$$= \frac{1}{k}\left(kx - \frac{f(kx)}{\|s(y)\|^2} s(y)\right) = \frac{1}{k} G_f(kx) \tag{5.56}$$

When $g(x) \leq 0$ or $0 \in \partial g(x)$, we have $f(kx) \leq 0$ or $0 \in \partial f(y)$ with $y = kx$, thus

$$G_g(x) = x = \frac{1}{k} kx = \frac{1}{k} G_f(kx). \tag{5.57}$$

This establishes the result. □

Remark 5.22. Proposition 5.21 can be guaranteed if f is a continuous and convex function, see [12, Proposition 3.1 (ii)].

For an operator $A : X \to X$, A^* is the adjoint of A.

Proposition 5.23. *Assume that $A : X \to X$ is unitary, i.e., $A^*A = \mathrm{Id}$, and let $b \in X$, $f : X \to]-\infty, +\infty]$. Define $x \mapsto g(x) := f(Ax + b)$. Then*

$$G_g(x) = A^*\big(G_f(Ax + b) - b\big) \tag{5.58}$$

for every $x \in X$. Furthermore,

$$\operatorname{Fix} G_g = A^*(\operatorname{Fix} G_f - b). \tag{5.59}$$

Proof. By Fact 5.16, $\partial g(x) = A^*\partial f(y)$ where $y = Ax + b$. As A is unitary, $\forall s(y) \in \partial f(y)$, we have $\|A^*s(y)\| = \|s(y)\|$. Indeed,

$$\|A^*s(y)\|^2 = \langle A^*s(y), A^*s(y)\rangle = \langle s(y), AA^*s(y)\rangle = \langle s(y), s(y)\rangle = \|s(y)\|^2. \tag{5.60}$$

When $g(x) > 0$ and $0 \notin \partial g(x)$, we have $f(Ax + b) > 0$ and $0 \notin \partial f(y)$ with $y = Ax + b$. Let $s(y) \in \partial f(y)$, we have $A^*s(y) \in \partial g(x)$.

$$G_g(x) = x - \frac{f(Ax + b)}{\|A^*s(y)\|^2}A^*s(y) = A^*\left(Ax + b - \frac{f(Ax + b)}{\|s(y)\|^2}s(y) - b\right) \tag{5.61}$$

$$= A^*(G_f(Ax + b) - b). \tag{5.62}$$

When $g(x) \leq 0$ or $0 \in \partial g(x)$, we have $f(Ax + b) \leq 0$ or $0 \in \partial f(y)$ with $y = Ax + b$, thus

$$G_g(x) = x = A^*(Ax + b - b) = A^*(G_f(Ax + b) - b). \tag{5.63}$$

Hence (5.58) holds. Finally (5.59) follows from (5.58). □

Remark 5.24. Combining [12, Proposition 3.1 (iii)] with [12, Proposition 3.1 (iv)], we have a result related to Proposition 5.23 if f is continous and convex. For another similar result, see [44, Proposition 2.9]

Corollary 5.25. *Let $a \in X$, $f : X \to]-\infty, +\infty]$ and $g(x) = f(x - a)$. Then*

$$(\forall\, x \in X) \quad G_g(x) = G_f(x - a) + a. \tag{5.64}$$

Moreover, $\operatorname{Fix} G_g = a + \operatorname{Fix} G_f$.

Proof. Let $A = \mathrm{Id}$ in Proposition 5.23. The result is straightforward. □

Remark 5.26. When f is a continuous convex function, Corollary 5.25 was shown in [12, Proposition 3.1 (v)].

Theorem 5.27. *Assume that $f \geq 0$ and $g = f^k$ with $k > 0$. Then*

$$G_g = \left(1 - \frac{1}{k}\right) \mathrm{Id} + \frac{1}{k} G_f.$$

Proof. By Fact 5.15, $\partial g(x) = kf(x)^{k-1}\partial f(x)$ when $f(x) \geq 0$.
When $g(x) > 0$ and $0 \notin \partial g(x)$, we have $f(x) > 0$ and $0 \notin \partial f(x)$, therefore

$$(\mathrm{Id} - G_g)(x) = \frac{f(x)^k}{\|kf(x)^{k-1}s(x)\|^2} kf(x)^{k-1}s(x) \tag{5.65}$$

$$= \frac{1}{k}\frac{f(x)}{\|s(x)\|^2}s(x) \tag{5.66}$$

$$= \frac{1}{k}(\mathrm{Id} - G_f)(x) \tag{5.67}$$

where $s(x) \in \partial f(x)$.
When $g(x) = 0$ or $0 \in \partial g(x)$, we have $f(x) = 0$ or $0 \in \partial f(x)$, thus

$$(\mathrm{Id} - G_g)(x) = 0 = (\mathrm{Id} - G_f)(x). \tag{5.68}$$

Therefore, $\mathrm{Id} - G_g = \frac{1}{k}(\mathrm{Id} - G_f)$ which gives $G_g = \left(1 - \frac{1}{k}\right) \mathrm{Id} + \frac{1}{k} G_f$. □

Remark 5.28. If f and g are convex functions, the conclusion of Theorem 5.27 can be found in [12, Proposition 3.1 (iii)] and in [44, Corollary 2.3].

The following is immediate from the definition of subgradient projectors.

Theorem 5.29. *Let f, g be two functions such that $f \equiv g$ on an open set $U \subseteq X$. Then $\mathcal{G}_f = \mathcal{G}_g$ on U.*

Remark 5.30. For calculus of subgradient projectors of convex functions, see [12, 44].

5.4 Basic properties of subgradient projectors

Based on the definition of subgradient projector, we have the following fundamental properties which give an alternative expression of G_f in terms of $g(x) = \ln f(x)$.

Theorem 5.31. *(i) We have*

$$\|x - G_f(x)\| = \frac{f(x)}{\|s(x)\|} \quad \textit{and} \tag{5.69}$$

$$\frac{x - G_f(x)}{\|x - G_f(x)\|^2} = \frac{s(x)}{f(x)} \tag{5.70}$$

for every x satisfying $f(x) > 0$ and $0 \notin \partial f(x)$. In particular, when f is continuously differentiable, one has

$$\frac{x - G_f(x)}{\|x - G_f(x)\|^2} = \nabla(\ln f(x)).$$

(ii) Set

$$g(x) = \begin{cases} \ln f(x) & \text{if } f(x) > 0, \\ -\infty & \text{if } f(x) \leq 0. \end{cases} \tag{5.71}$$

Then g is lower-semicontinuous on the open set $\{x \in X : f(x) > 0\}$. Then whenever $f(x) > 0$ and $0 \notin \partial f(x)$ we have

$$G_f(x) = x - \frac{c(x)}{\|c(x)\|^2} \tag{5.72}$$

where $c(x) \in \partial g(x)$.

If f is continuously differentiable on $X \setminus lev_0 f$, then

$$G_f(x) = x - \frac{\nabla g(x)}{\|\nabla g(x)\|^2},$$

whenever $f(x) > 0$ and $\nabla f(x) \neq 0$.

Proof. (i) When $f(x) > 0$ and $0 \notin \partial f(x)$, let $s(x) \in \partial f(x)$, we have

$$x - G_f(x) = \frac{f(x)}{\|s(x)\|^2} s(x).$$

Therefore,

$$\|x - G_f(x)\| = \frac{f(x)}{\|s(x)\|}.$$

If follows that

$$x - G_f(x) = \frac{f(x)^2}{\|s(x)\|^2} \frac{s(x)}{f(x)} = \|x - G_f(x)\|^2 \frac{s(x)}{f(x)}. \tag{5.73}$$

Hence

$$\frac{x - G_f(x)}{\|x - G_f(x)\|^2} = \frac{s(x)}{f(x)}.$$

When f is continuously differentiable, $s(x) = \nabla f(x)$. The result follows as $\nabla(\ln f(x)) = \frac{\nabla f(x)}{f(x)}$.
(ii) By Fact 5.15, we have $\partial g(x) = \frac{1}{f(x)}\partial f(x)$ when $f(x) > 0$. Then

$$c(x) = \frac{s(x)}{f(x)} \in \partial g(x)$$

because $s(x) \in \partial f(x)$. Since

$$\frac{1}{\|c(x)\|^2} = \frac{f(x)^2}{\|s(x)\|^2},$$

when $f(x) > 0$ and $0 \in \partial f(x)$. So we have

$$\frac{c(x)}{\|c(x)\|^2} = \frac{s(x)}{f(x)}\frac{f(x)^2}{\|s(x)\|^2} = \frac{f(x)}{\|s(x)\|^2}s(x).$$

Thus

$$G_f(x) = x - \frac{f(x)}{\|s(x)\|^2}s(x) = x - \frac{c(x)}{\|c(x)\|^2}, \tag{5.74}$$

when $f(x) > 0$ and $0 \notin \partial f(x)$. Hence (5.72) is true.
When f is continuously differentiable and $f(x) > 0$ and $\nabla f(x) \neq 0$, the result follows from

$$\partial g(x) = \left\{\frac{\nabla f(x)}{f(x)}\right\} \quad \Rightarrow \quad \nabla g(x) = c(x) = \frac{\nabla f(x)}{f(x)}.$$

From (5.74), we have

$$G_f(x) = x - \frac{\nabla g(x)}{\|\nabla g(x)\|^2},$$

whenever $f(x) > 0$ and $\nabla f(x) \neq 0$. □

5.4.1 When is a mapping T a subgradient projector?

Theorem 5.32. *A mapping $T : X \to X$ is a subgradient projector G_f of a locally Lipschitz function $f : X \to \mathbb{R}$ if and only if*

$$\frac{x - Tx}{\|x - Tx\|^2} \in \partial(\ln f(x)) \quad \textit{whenever } f(x) > 0 \textit{ and } 0 \notin \partial f(x), \tag{5.75}$$

$$Tx = x \quad \textit{whenever } f(x) \leq 0 \textit{ or } 0 \in \partial f(x). \tag{5.76}$$

Proof. $\Rightarrow$: Suppose that $T = G_f$. Applying Theorem 5.31(i), we obtain (5.75).

$\Leftarrow$: Assume that (5.75) and (5.76) hold. By Theorem 5.31(ii), $\partial \ln f = \frac{\partial f}{f}$. When $f(x) > 0$ and $0 \notin \partial f(x)$, (5.75) gives

$$\frac{1}{\|x - Tx\|} = \frac{\|s(x)\|}{f(x)} \quad \text{i.e., } \|x - Tx\| = \frac{f(x)}{\|s(x)\|},$$

where $s(x) \in \partial f(x)$. By (5.75), when $f(x) > 0$ and $0 \notin \partial f(x)$, we have

$$x - Tx = \|x - Tx\|^2 \frac{s(x)}{f(x)},$$

so that Tx is equal to

$$x - \|x - Tx\|^2 \frac{s(x)}{f(x)} = x - \left(\frac{f(x)}{\|s(x)\|}\right)^2 \frac{s(x)}{f(x)} = x - \frac{f(x)}{\|s(x)\|^2} s(x) = G_f(x)$$

when $f(x) > 0$ and $0 \notin \partial f(x)$. □

5.4.2 Recovering f from its subgradient projector G_f

Can one determine the function f if G_f is known?

Definition 5.33. A locally Lipschitz function $f : X \to \mathbb{R}$ is called *essentially strictly differentiable* on an open set $A \subset X$ if f is strictly differentiable everywhere on A except possibly on a Lebesgue null set.

Such a class of functions has been extensively studied by Borwein and Moors [17]. This class of functions include finite-valued convex functions, Clarke regular locally Lipschitz functions, semismooth locally Lipschitz functions, C^1 functions and others, [17, pages 323-328]. If a locally Lipschitz function is essentially smooth, then ∂f is single-valued almost everywhere. Moreover, ∂f can be recovered by every densely defined selection $s \in \partial f$.

Fact 5.34. *Let f, g be locally Lipschitz on a polygonally connected*[19] *and open subset O of $\mathbb{R}^n$. If $\nabla f = \nabla g$ almost everywhere on O, then $h := f - g$ is a constant on O.*

[19] (See [35, Definition 2] on page 189) A set $W \subset \mathbb{R}^n$ is said to be polygonally connected if given any two points $x, y \in W$, there are points $x_0 = x, x_1, \cdots, x_m = y$ such that $\bigcup_{i=1}^m \overline{x_{i-1}x_i} \subseteq W$, where $\overline{x_{i-1}x_i}$ is the closed segment joining x_{i-1} and x_i.

Proof. We prove this by contradiction. By the assumption, h is locally Lipschitz and $\nabla h = 0$ almost everywhere on O. Suppose that $x, y \in O$ and $h(x) \neq h(y)$. As O is polygonally connected, there exists $z \in O$ such that either $[x, z] \subseteq O$ with $h(x) \neq h(z)$ or $[z, y] \subseteq O$ with $h(z) \neq h(y)$. Without loss of generality, assume $[z, y] \subseteq O$ and $h(z) \neq h(y)$. As h is differentiable almost everywhere, by Fubini's Theorem [52, Theorem 6.2.2, page 110], we can choose $\tilde{z}$ nearby z and $\tilde{y}$ nearby y so that both h is differentiable and $\nabla h = 0$ almost everywhere on $[\tilde{z}, \tilde{y}] \subseteq O$, and $h(\tilde{z}) \neq h(\tilde{y})$. Then

$$h(\tilde{y}) - h(\tilde{z}) = \int_0^1 \langle \nabla h(\tilde{z} + t(\tilde{y} - \tilde{z})), \tilde{y} - \tilde{z} \rangle dt = \int_0^1 0 dt = 0$$

which contradicts $h(\tilde{z}) \neq h(\tilde{y})$. □

Theorem 5.35. *Let $T : X \to X$ be a subgradient projector. More precisely, if*

$$G_f = T = G_{f_1}$$

and f, f_1 are essentially strictly differentiable, then there exists $k > 0$ such that

$$f = k f_1$$

on each polygonally connected component.

Proof. If $G_f = T$ then $\operatorname{dom} f \supset \operatorname{dom} T$. As T has full domain, we have $\operatorname{dom} f = X$. Assume that there exist two essentially strictly differentiable and locally Lipschitz functions such that $T = G_f = G_{f_1}$. By Theorem 5.32 (ii), we have

$$\frac{x - T(x)}{\|x - T(x)\|^2} \in \partial(\ln f(x)) \quad \text{whenever } f(x) > 0,$$

$$\frac{x - T(x)}{\|x - T(x)\|^2} \in \partial(\ln f_1(x)) \quad \text{whenever } f_1(x) > 0.$$

As f, f_1 are locally Lipschitz, both $\ln f, \ln f_1$ are locally Lipschitz on $X \setminus \operatorname{Fix} T$. Then

$$\partial \ln f = \frac{1}{f} \partial f, \quad \partial \ln f_1 = \frac{1}{f_1} \partial f_1$$

by Fact 5.15 or [25, Theorem 2.3.9(ii)]. Because f, f_1 are essentially strictly differentiable and locally Lipschitz, $\partial f, \partial f_1$ are single-valued almost everywhere [46], thus

$$\partial(\ln f_1(x)) = \frac{x - T(x)}{\|x - T(x)\|^2} = \partial(\ln f(x)) \quad \text{almost everywhere.}$$

By Fact 5.34, $\ln f - \ln f_1 = c$ for some $c \in \mathbb{R}$, which implies that $f_1 = kf$ for some $k > 0$. □

5.4.3 Fixed point closed property and continuity

Theorem 5.36. *(fixed-point closed property) For a locally Lipschitz function f, G_f is fixed-point closed at every $x \in \mathbb{R}^n$, i.e.,*

$$\|y - G_f(y)\| \to 0 \text{ and } y \to x \quad \Rightarrow \quad x = G_f(x). \tag{5.77}$$

Proof. Assume that a sequence $(y_n)_{n\in}$ in X satisfies

$$\|y_n - G_f(y_n)\| \to 0 \text{ and } y_n \to x. \tag{5.78}$$

Consider three cases.

Case 1: If there exists infinitely many y_n's, say $(y_{n_k})_{k\in\mathbb{N}}$, such that $0 \in \partial f(y_{n_k})$. Since ∂f is upper semicontinuous, taking limit when $k \to \infty$ gives $0 \in \partial f(x)$. Hence $x = G_f(x)$.

Case 2: If there exists infinitely many y_n's, say $(y_{n_k})_{k\in}$, such that $f(y_{n_k}) \leq 0$. Taking limit when $k \to \infty$ and using the continuity of f at x gives

$$f(x) = \lim_{k\to\infty} f(y_{n_k}) \leq 0.$$

Hence $x = G_f(x)$.

Case 3: There exists $N \in \mathbb{N}$ such that $f(y_n) > 0$ and $0 \notin \partial f(y_n)$ when $n > N$. Then by (5.69),

$$f(y_n) = \|y_n - G_f(y_n)\|\|s(y_n)\|. \tag{5.79}$$

As f is continuous at x, f is locally Lipschitz around x, so ∂f is locally bounded around x. Therefore,

$$f(x) = \lim_{n\to\infty} f(y_n) = \lim_{n\to\infty} (\|y_n - G_f(y_n)\|\|s(y_n)\|) = 0$$

since $\|y_n - G_f(y_n)\| \to 0$. Hence $x = G_f(x)$. Altogether, $x \in \operatorname{Fix} G_f$. This establishes (5.77) because $(y_n)_{n\in\mathbb{N}}$ was an arbitrary sequence satisfying (5.78). □

Theorem 5.37. *Let f be a locally Lipschitz function and essentially strictly differentiable. Then G_f is continuous at $x \in X \setminus \operatorname{Fix} G_f$ if and only if f is strictly differentiable at x.*

Proof. Let $x \in X \setminus \operatorname{Fix} G_f$.
$\Leftarrow$: Assume that f is strictly differentiable at x. By [39, page 258], the limiting subdifferential is reduced to the singleton. Thus we have $\nabla f(x) \neq 0$ which is continuous at x. The result follows from the definition of G_f, which is

$$y \mapsto G_f(y) = y - \frac{f(y)}{\|\nabla f(y)\|^2} \nabla f(y). \tag{5.80}$$

$\Rightarrow$: Assume that G_f is continuous at x. By (5.70),

$$s(y) = f(y) \frac{y - G_f(y)}{\|y - G_f(y)\|^2} \tag{5.81}$$

so s is (norm to norm) continuous at x. As s is a selection of ∂f and f is essentially differentiable, f is strictly differentiable at x. □

The following example illustrates that G_f cannot be continuous if f is not differentiable.

Example 5.38. Define

$$f(x) = \begin{cases} |x| & \text{if } x \leq 1, \\ 2x - 1 & \text{if } x > 1. \end{cases} \tag{5.82}$$

Then

$$G_f(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1/2 & \text{if } x > 1, \\ 1 - \frac{1}{s(x)} \text{ where } s(x) \in [1, 2] & \text{if } x = 1, \end{cases} \tag{5.83}$$

which is discontinuous at $x = 1$

Proof. When $x < 0$, $G_f(x) = x - \frac{-x}{(-1)^2}(-1) = 0$; When $x = 0$, $f(0) = 0$, so $G_f(0) = 0$; When $0 < x < 1$, $G_f(x) = x - \frac{x}{1^2}(1) = 0$; When $x > 1$, $G_f(x) = x - \frac{2x-1}{2^2} 2 = 1/2$; When $x = 1$, $\partial f(1) = [1, 2]$, so

$$G_f(x) = x - \frac{1}{s(x)^2}(s(x)) = 1 - \frac{1}{s(x)} \tag{5.84}$$

where $s(x) \in [1, 2]$. □

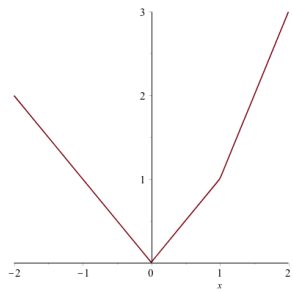


Figure 5.2: Graph of f, where f is defined in Example 5.38.

5.5 When is the subgradient projector G_f a cutter?

Our first result characterizes the class of functions f for which its G_f is a cutter.

Theorem 5.39. *(level sets of tangent plane including the target set) Let $f : X \to \mathbb{R}$ be locally Lipschitz. Then G_f is a cutter if and only if whenever x satisfies $f(x) > 0$, $0 \notin \partial f(x)$ and u satisfies $f(u) \leq 0$ or $0 \in \partial f(u)$ one has*

$$f(x) + \langle s(x), u - x\rangle \leq 0, \tag{5.85}$$

where $s(x) \in \partial f(x)$. That is, when $f(x) > 0$ and $0 \notin \partial f(x)$,

$$\{u \in X : f(u) \leq 0 \text{ or } 0 \in \partial f(u)\} \subseteq \{u \in X : f(x) + \langle s(x), u - x\rangle \leq 0\}. \tag{5.86}$$

Proof. When $f(x) \leq 0$ or $0 \in \partial f(x)$, $x = G_f x$, then G_f satisfies the inequality in the definition of cutter. Assume that $f(x) > 0$, $0 \notin \partial f(x)$ and $s(x) \in \partial f(x), u \in \operatorname{Fix} G_f$. We have

$$\langle x - G_f x, u - G_f x\rangle = \langle \frac{f(x)}{\|s(x)\|^2} s(x), u - x + \frac{f(x)}{\|s(x)\|^2} s(x)\rangle \tag{5.87}$$

$$= \frac{f(x)}{\|s(x)\|^2}\langle s(x), u - x\rangle + \frac{f^2(x)}{\|s(x)\|^2} \tag{5.88}$$

$$= \frac{f(x)}{\|s(x)\|^2}\big(f(x) + \langle s(x), u - x\rangle\big). \tag{5.89}$$

As $f(x) > 0$,

$$\langle x - G_f x, u - G_f x\rangle \leq 0 \quad \Leftrightarrow \quad f(x) + \langle s(x), u - x\rangle \leq 0. \tag{5.90}$$

□

Remark 5.40. When function f is locally Lipschitz (not necessarily convex), if it satisfies

$$\{u \in X : f(u) \leq 0 \text{ or } 0 \in \partial f(u)\} \subseteq \{u \in X : f(x) + \langle s(x), u - x\rangle \leq 0\}, \tag{5.91}$$

when $f(x) > 0$ and $0 \notin \partial f(x)$. By Theorem 5.39, we have G_f is a cutter. Moreover, if the interior of $\operatorname{lev}_0 f$ is nonempty, we take $T = G_f$ into finite convergent algorithm in [13], where f is locally Lipschitz. Then the sequence generated by G_f will converge finitely.

Theorem 5.41. *Let $k \geq 1$ and $f : X \to [0, +\infty)$ be locally Lipschitz. If $g = f^k$ and G_f is a cutter, then G_g is a cutter.*

Proof. By Theorem 5.27, $G_g = (1 - 1/k)\operatorname{Id} + 1/k G_f$. As Id and G_f are both cutters, and $\operatorname{Fix} G_f \cap \operatorname{Fix}\operatorname{Id} = \operatorname{Fix} G_f \neq \varnothing$, being a convex combination of cutters, G_g is a cutter by [20, Corollary 2.1.49, page 62]. □

Theorem 5.42. *Assume that $A : X \to X$ is unitary, i.e., $A^*A = \operatorname{Id}$, and let $b \in X$, $f : X \to \,]-\infty, +\infty]$. Define $x \mapsto g(x) := f(Ax + b)$. If G_f is a cutter, then G_g is a cutter.*

Proof. Let $x \in X$, $u \in \operatorname{Fix} G_g$. Proposition 5.23 gives

$$G_g(x) = A^*\big(G_f(Ax + b) - b\big) \tag{5.92}$$

and $Au + b \in \operatorname{Fix} G_f$. Since A is unitary and G_f is a cutter, we have

$$\begin{aligned}
\|x - G_g(x)\|^2 &= \|x - A^*\big(G_f(Ax + b) - b\big)\|^2 && (5.93)\\
&= \|Ax + b - G_f(Ax + b)\|^2 && (5.94)\\
&\leq \|Ax + b - (Au + b)\|^2 - \|G_f(Ax + b) - (Au + b)\|^2 && (5.95)\\
&= \|x - u\|^2 - \|G_f(Ax + b) - Au - b\|^2 && (5.96)\\
&= \|x - u\|^2 - \|A^*\big(G_f(Ax + b) - b\big) - u\|^2 && (5.97)\\
&= \|x - u\|^2 - \|G_g(x) - u\|^2. && (5.98)
\end{aligned}$$

Hence G_g is a cutter by Fact 2.27(i). □

Theorem 5.43. *Assume that $\mathbb{R} \ni k \neq 0$, and $g(x) = f(kx)$. If G_f is a cutter, then G_g is a cutter.*

Proof. Proposition 5.21 gives $G_g(x) = \frac{1}{k}G_f(kx)$, and $\operatorname{Fix} G_g = \frac{1}{k}\operatorname{Fix} G_f$. Let $x \in X$ and $u \in \operatorname{Fix} G_g$, we have $ku \in \operatorname{Fix} G_f$. Then

$$\langle x - G_g(x), u - G_g(x)\rangle = \langle x - \frac{1}{k}G_f(kx), u - \frac{1}{k}G_f(kx)\rangle \tag{5.99}$$

$$= \frac{1}{k^2}\langle kx - G_f(kx), ku - G_f(kx)\rangle \leq 0 \tag{5.100}$$

since G_f is a cutter. Therefore, G_g is a cutter. □

It is easy to show that

Fact 5.44. *[20, Corollary 4.2.6] Let $f : X \to \mathbb{R}$ be convex and $\operatorname{lev}_0 f \neq \varnothing$. Then G_f is a cutter. Consequently, G_f is continuous at every $x \in \operatorname{lev}_0 f$.*

Proof. As $\operatorname{lev}_0 f \neq \varnothing$, $\operatorname{Fix} G_f = \operatorname{lev}_0 f$. Assume that $f(x) > 0$. For $u \in \operatorname{Fix} G_f$, $f(u) \leq 0$. By the convexity of f we have

$$f(x) + \langle s(x), u - x\rangle \leq f(u) \leq 0.$$

Hence Theorem 5.39 applies. Therefore, G_f is a cutter. The remaining result follows from Fact 2.27(ii). □

Remark 5.45. If the functions f defined in Theorems 5.41, 5.42 and 5.43 are convex, then functions g are also convex, respectively. By [13, Example 2.1], we have G_f and G_g are cutters. Moreover, one more result [13, Proposition 5.3] supports Fact 5.44.

In Fact 5.44, $\operatorname{lev}_0 f \neq \varnothing$ is required, as the following example shows.

Example 5.46. (1). Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \exp|x| \quad \forall\, x \in \mathbb{R}.$$

Then $\operatorname{lev}_0 f = \varnothing$ and

$$G_f(x) = \begin{cases} x - 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ x + 1 & \text{if } x < 0. \end{cases}$$

In particular, this G_f is discontinuous at $x = 0$ and not a cutter. Moreover, G_f is not monotone.

(2). Consider

$$x \mapsto f(x) = \exp(\|x\|^2/2).$$

We have $\operatorname{lev}_0 f = \varnothing$ and

$$G_f(x) = \begin{cases} x - \frac{x}{\|x\|^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

In particular, G_f is not continuous at 0, so not a cutter.

One might ask: If each function $f_i : X \to \mathbb{R}$ has G_{f_i} being a cutter, must the maximum $g := \max\{f_1, f_2\}$ has G_g being a cutter? The answer is negative as the following example shows.

Example 5.47. Let $f_1(x) = 1 + x$ and $f_2(x) = 1 - x$ on $\mathbb{R}$. Each G_{f_i} is a cutter by Fact 5.44. The function $g(x) := \max\{f_1(x), f_2(x)\}$ has

$$G_g(x) := \begin{cases} -1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x < 0 \end{cases}$$

which is not continuous at $x = 0$, so G_g is not a cutter.

Another question that may arise is if the convexity of the function is necessary for G_f to be a cutter? Following examples illustrate that the convexity of the function is independent of G_f to be a cutter.

Example 5.48. If f is not convex, G_f need not be a cutter. Consider $f : \mathbb{R} \to \mathbb{R}$ defined by $\mathbb{R} \ni x \mapsto f(x) = 1 - \exp(-x^2)$. Then the subgradient projector of f is

$$G_f(x) = \begin{cases} x - \left(\frac{1}{2x\exp(-x^2)} - \frac{1}{2x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases} \tag{5.101}$$

and $\operatorname{Fix} G_f = \{0\}$. However G_f is not a cutter.

Indeed, when $|x| > \sqrt{2}$ we have

$$f(x) + \nabla f(x)(0-x) = 1 - \exp(-x^2) + (2x\exp(-x^2))(0-x) \tag{5.102}$$

$$= 1 - \frac{1+2x^2}{\exp(x^2)} \tag{5.103}$$

$$= \frac{\exp(x^2) - (1+2x^2)}{\exp(x^2)} \tag{5.104}$$

$$\geq \frac{1 + x^2 + \frac{x^4}{2} - (1+2x^2)}{\exp(x^2)} \tag{5.105}$$

$$= \frac{x^2(x^2-2)}{2\exp(x^2)} > 0. \tag{5.106}$$

By Theorem 5.39, G_f is not a cutter.

Example 5.49. Even when f is not convex, G_f may still be a cutter. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 20/7, \\ 8(x-2.5) & \text{if } 20/7 \leq x \leq 3, \\ 2(x-1) & \text{if } x > 3. \end{cases} \tag{5.107}$$

Then f is not convex since $f'(x)$ is not monotone on $[20/7, +\infty)$. However, its subgradient projector

$$G_f(x) = \begin{cases} x & \text{if } x \leq 0, \\ 0 & \text{if } 0 < x < 20/7, \\ \frac{20}{7} - \frac{20}{7}\frac{1}{s(x)} & \text{if } x = 20/7, \text{ where } s(x) \in [1,8], \\ 2.5 & \text{if } 20/7 < x < 3, \\ 3 - \frac{4}{s(x)} & \text{if } x = 3, \text{ where } s(x) \in [2,8], \\ 1 & \text{if } x > 3. \end{cases} \tag{5.108}$$

is a cutter.

To see this, by Theorem 5.39, it suffices to consider zero level sets of tangent planes. Indeed, Let $f(u) \leq 0$, i.e., $u \leq 0$. When $x_0 > 3$,

$$f(x_0) + s(x_0)(u - x_0) = 2(u-1) \leq 0; \tag{5.109}$$

when $x_0 = 3$,

$$f(x_0) + s(x_0)(u - x_0) = 4 + s(3)(u-3) \leq 4 + 2(u-3) \leq 0; \tag{5.110}$$

where $2 \leq s(3) \leq 8$; when $20/7 < x_0 < 3$,

$$f(x_0) + s(x_0)(u - x_0) = 8(u - 2.5) \leq 0; \tag{5.111}$$

when $x_0 = 20/7$,

$$f(x_0) + s(x_0)(x - x_0) = 20/7 + s(20/7)(u - 20/7) \leq u \leq 0; \tag{5.112}$$

where $1 \leq s(20/7) \leq 8$; when $0 < x_0 < 20/7$,

$$f(x_0) + s(x_0)(u - x_0) = u \leq 0. \tag{5.113}$$

Remark 5.50. Note that even if G_f is continuous, it does not mean that G_f is a cutter, e.g., see Example 5.12(ii). (See Corollary 6.14(ii) for an example on $\mathbb{R}^2$). In [20], Cegielski developed a systematic theory for cutters. The theory of cutters can be used to study the class of functions (Theorem 5.39) whose subgradient projectors are cutters.

5.6 Characterization of subgradient projectors of convex functions

The following result is of independent interest.

Proposition 5.51. *Assume that $C \subseteq X$ is closed and convex. The function $f : X \to \mathbb{R}$ satisfies*

(i) $f \geq 0$ on $X \setminus C$;

(ii) f is convex on every convex subsets of $X \setminus C$;

(iii) $\lim_{\substack{y \to x \\ y \in X \setminus C, x \in C}} f(y) = 0$, i.e., $\lim_{i \to \infty} f(y_i) = 0$ whenever $(y_i)_{i=1}^{\infty}$ is a sequence in $X \setminus C$ converging to a boundary point x of $X \setminus C$.

Define

$$g(x) = \begin{cases} f(x) & \text{if } x \notin C \\ 0 & \text{if } x \in C. \end{cases}$$

Then g is convex on X.

Proof. Let $x, y \in X$, $0 \leq \lambda \leq 1$. We consider three cases.

(i) If $[x, y] \subseteq X \setminus C$, $g = f$ is convex on $[x, y]$ by assumption.

(ii) If $\lambda x + (1-\lambda)y \in C$, then

$$g(\lambda x + (1-\lambda)y) = 0 \leq \lambda g(x) + (1-\lambda)g(y)$$

since $g(x), g(y) \geq 0$.

(iii) $\lambda x + (1-\lambda)y \notin C$ and $[x, y] \cap C \neq \varnothing$. In particular, both x, y cannot both be in C. We consider two subcases. (1). $x \in C$ and $y \notin C$. We need to show

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y).$$

As $y \notin C$, there exists $z \in \operatorname{bdry}(C)$ such that

$$\lambda x + (1-\lambda)y \in [z, y] \subset X \setminus C \quad \text{and} \quad f(z) = 0.$$

As

$$\lambda x + (1-\lambda)y = \alpha z + (1-\alpha)y \quad \text{for some } 0 \leq \alpha \leq 1.$$

and f is convex on $[z, y]$, we have

$$f(\lambda x + (1-\lambda)y) = f(\alpha z + (1-\alpha)y) \leq \alpha f(z) + (1-\alpha)f(y) = (1-\alpha)f(y). \tag{5.114}$$

Now

$$z = \beta x + (1-\beta)y \quad \text{for some } 0 \leq \beta \leq 1,$$

$$\lambda x + (1-\lambda)y = \alpha z + (1-\alpha)y = \alpha(\beta x + (1-\beta)y) + (1-\alpha)y$$

give $\lambda = \alpha\beta$. Therefore, by (5.114), $g(x) = 0$ and $g(y) = f(y) \geq 0$,

$$g(\lambda x + (1-\lambda)y) = f(\lambda x + (1-\lambda)y) \tag{5.115}$$

$$\leq (1-\alpha\beta)f(y) = (1-\lambda)g(y) + \lambda g(x). \tag{5.116}$$

(2). $x \notin C$ and $y \notin C$. We need to show

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y).$$

There exists $z \in \operatorname{bdry}(C)$ such that

$$\lambda x + (1-\lambda)y \in [z, y] \quad \text{or} \quad \lambda x + (1-\lambda)y \in [x, z],$$

say $\lambda x + (1-\lambda)y \in [z, y]$. Then

$$\lambda x + (1-\lambda)y = \alpha z + (1-\alpha)y \quad \text{for some } 0 \leq \alpha \leq 1.$$

As f is convex on $[z, y]$, $f(z) = 0$,

$$g(\lambda x + (1-\lambda)y) = f(\alpha z + (1-\alpha)y) \le \alpha f(z) + (1-\alpha)f(y) = (1-\alpha)f(y). \tag{5.117}$$

Now

$$z = \beta x + (1-\beta)y \quad \text{for some } 0 \le \beta \le 1,$$

$$\lambda x + (1-\lambda)y = \alpha z + (1-\alpha)y = \alpha(\beta x + (1-\beta)y) + (1-\alpha)y$$

give $\lambda = \alpha\beta$. Then by (5.117), using $g(x) = f(x) \ge 0$, $g(y) = f(y) \ge 0$,

$$g(\lambda x + (1-\lambda)y) \le (1-\alpha\beta)f(y) = (1-\lambda)f(y) \tag{5.118}$$

$$\le (1-\lambda)f(y) + \lambda f(x) \tag{5.119}$$

$$= (1-\lambda)g(y) + \lambda g(x). \tag{5.120}$$

Combining (i)–(iii), we conclude that g is convex on X. □

Theorem 5.52. *Let $T : X \to X$ and*

$$C := \{x \in X : Tx = x\}.$$

Then T is a subgradient projector of a convex function $f : X \to \mathbb{R}$ with $\mathrm{lev}_0 f = C$ if and only if there exists $g : X \to [-\infty, +\infty)$ such that $g : X \setminus C \to \mathbb{R}$ is locally Lipschitz, $g(x) = -\infty$ for every $x \in C$, and

(i) for every $x \in X \setminus C$, $\frac{x-Tx}{\|x-Tx\|^2} \in \partial g(x)$;

(ii) The function defined by

$$f(x) := \begin{cases} \exp(g(x)) & \text{if } x \notin C, \\ 0 & \text{if } x \in C \end{cases}$$

is convex.

In this case, $T = G_f$.

Proof. $\Rightarrow$: Assume that T is a subgradient projector, say $T = G_{f_1}$ with $f_1 : X \to \mathbb{R}$ being convex and $\mathrm{lev}_0 f_1 = C$. Then $f = \max\{0, f_1\}$ is convex and $G_f = G_{f_1}$. Put $g = \ln f$ and $C = \mathrm{lev}_0 f$. Since f is locally Lipschitz, g is locally Lipschitz on $X \setminus C$. Note that $\partial g(x) = (\partial f(x))/f(x)$ when $f(x) > 0$. Apply Theorem 5.31(i) to obtain (i).

$\Leftarrow$: Assume that (i), (ii) hold. When $x \notin C$, (i) and (ii) give

$$\|x - Tx\| = \frac{1}{\|c(x)\|}, \quad \partial g(x) = \frac{\partial f(x)}{f(x)}$$

where $c(x) \in \partial g(x)$. Using (i) again, we have

$$Tx = x - \|x - Tx\|^2 c(x) = x - \frac{c(x)}{\|c(x)\|^2} = G_f(x) \tag{5.121}$$

by Theorem 5.31(ii). Moreover, when $x \in C$, $Tx = x = G_f(x)$. Hence $T = G_f$. □

Theorem 5.53. *Let $T : X \to X$ be differentiable and*

$$C = \{x \in X : Tx = x\}$$

be closed and convex. Define $T_1 : X \setminus C \to X$ by

$$x \mapsto T_1(x) = \frac{x - Tx}{\|x - Tx\|^2}.$$

Then T is a subgradient projector of a convex function $f : X \to \mathbb{R}$ with $\mathrm{lev}_0 f = C$ and being differentiable on $X \setminus C$ if and only if

(i) For every $x \in X \setminus C$,

$$T_1(x)(T_1(x))^* + \nabla T_1(x) \succeq 0$$

i.e., positive semidefinite.

(ii) There exists a function $g : X \to [-\infty, +\infty]$ such that

$$(\forall\ x \in X \setminus C) \quad \nabla g(x) = T_1(x),$$

$$(\forall\ x \in C) \quad \lim_{\substack{y \to x \\ y \in X \setminus C}} g(y) = -\infty,$$

and

$$(\forall\ x \in C) \quad g(x) = -\infty.$$

Proof. $\Rightarrow$: Assume that $T = G_f$ with f being convex and $\mathrm{lev}_0 f = C$. By Theorem 5.31(i), we can put $g = \ln f$ to obtain (ii). Indeed, for $x \in X \setminus C$,

$$T_1(x) = \frac{x - Tx}{\|x - Tx\|^2} = \frac{x - G_f(x)}{\|x - G_f(x)\|^2} = \nabla(\ln(f(x))) = \nabla g(x). \tag{5.122}$$

For $x \in C$, we have $f(x) \leq 0$. Moreover, let $(y_n) \subseteq X \setminus C$ and $y_n \to x$. Then $f(y_n) > 0$. Since f is continuous, thus we have $f(y_n) \to f(x) \geq 0$. Therefore $f(x) = 0$ and $\ln(f(x)) = -\infty$ for $x \in C$. By Theorem 5.52, $g = \ln(f)$ is locally Lipschitz, thus we have

$$\lim_{\substack{y \to x \\ y \in X \setminus C}} g(y) = \lim_{\substack{y \to x \\ y \in X \setminus C}} \ln(f(y)) = -\infty. \tag{5.123}$$

Furthermore, when $x \in C$, $f(x) \leq 0$ and $f(x) = \exp(g(x)) \geq 0$ as well. Thus $f(x) = \exp(g(x)) = 0$ when $x \in C$. Therefore $g(x) = -\infty$ when $x \in C$. As $f = \exp(g)$, for every $x \notin C$ we have

$$\nabla f(x) = e^{g(x)} \nabla g(x) = e^{g(x)} T_1(x),$$

$$\nabla^2 f(x) = e^{g(x)} T_1(x)(T_1(x))^* + e^{g(x)} \nabla T_1(x) = e^{g(x)} \big(T_1(x)(T_1(x))^* + \nabla T_1(x)\big).$$

Since f is convex, $\nabla^2 f(x) \succeq 0$, and this is equivalent to

$$T_1(x)(T_1(x))^* + \nabla T_1(x) \succeq 0$$

which is (i).

$\Leftarrow$: Assume that (i) and (ii) hold. Put $f = \exp(g)$. Then $\operatorname{lev}_0 f = C$. Indeed, let $x \in \operatorname{lev}_0 f$, we have $f(x) = \exp(g(x)) \leq 0$. That is $g(x) = -\infty$. If $x \notin C$, by (ii), we have $\nabla g(x) = T_1(x)$. This is not true when $g(x) = -\infty$. Thus $x \in C$. So $\operatorname{lev}_0 f \subseteq C$. Conversely, let $x \in C$, from (ii), we have $g(x) = -\infty$. Hence $f(x) = \exp(g(x)) = 0$. So $x \in \operatorname{lev}_0 f$. Then $C \subseteq \operatorname{lev}_0 f$. For $x \in X \setminus C$,

$$\nabla f(x) = e^{g(x)} \nabla g(x) = e^{g(x)} T_1(x),$$

$$\nabla^2 f(x) = e^{g(x)} T_1(x)(T_1(x))^* + e^{g(x)} \nabla T_1(x) = e^{g(x)} \big(T_1(x)(T_1(x))^* + \nabla T_1(x)\big).$$

(i) and (ii) imply that f is differentiable and convex on convex subsets of $X \setminus C$, and $f \equiv 0$ on C. By Proposition 5.51, f is convex on X.

Moreover, when $x \neq Tx$ we have

$$G_f(x) = x - \left(\frac{f(x)}{\|\nabla f(x)\|}\right)^2 \frac{\nabla f(x)}{f(x)} = x - \frac{T_1(x)}{\|T_1(x)\|^2} \tag{5.124}$$

$$= x - \frac{\frac{x - Tx}{\|x - Tx\|^2}}{\left(\frac{1}{\|x - Tx\|}\right)^2} = x - (x - Tx) = Tx. \tag{5.125}$$

□

Let's reduce $X = \mathbb{R}^n$ to the real line $\mathbb{R}$.

Corollary 5.54. *Let $T : \mathbb{R} \to \mathbb{R}$ be differentiable and*

$$C = \{x \in \mathbb{R} : x = Tx\}$$

be a closed interval. Then T is a subgradient projector of a convex function $f : \mathbb{R} \to \mathbb{R}$ with $\mathrm{lev}_0 f = C$ and being differentiable on $\mathbb{R} \setminus C$ if and only if

(i) T is monotonically increasing on convex subsets of $\mathbb{R} \setminus C$;

(ii) The function

$$g(x) = \int_a^x \frac{1}{s - Ts} ds$$

satisfies

$$\lim_{x \downarrow \sup(C)} g(x) = -\infty$$

for some $a > \sup(C)$; and

$$\lim_{x \uparrow \inf(C)} g(x) = -\infty$$

for some $a < \inf(C)$.

Proof. Define $n : \mathbb{R} \to \mathbb{R}$ by $x \mapsto n(x) = x - Tx$. Then for every $x \notin C$, $T_1(x) = \frac{1}{n(x)}$. Theorem 5.53 (i) is equivalent to

$$\frac{1}{n^2(x)} - \frac{n'(x)}{n^2(x)} \geq 0.$$

This is the same as $n'(x) \leq 1$, which implies that $T'(x) \geq 0$. Conclusions in (ii) are actually equivalent to Theorem 5.53(ii). □

Corollary 5.55. *Let $T : \mathbb{R} \to \mathbb{R}$ be differentiable and*

$$C = \{x \in \mathbb{R} : x = Tx\}$$

be a closed interval. Define $N : \mathbb{R} \to \mathbb{R}$ by

$$x \mapsto N(x) := x - Tx.$$

Suppose that

(i) N is nonexpansive;

(ii) The function

$$g(x) = \int_a^x \frac{1}{s - Ts} ds$$

satisfies $\lim_{x \downarrow \sup(C)} g(x) = -\infty$ *for some* $a > \sup(C)$*; and* $\lim_{x \uparrow \inf(C)} g(x) = -\infty$ *for some* $a < \inf(C)$*.*

Then T is a subgradient projector of a convex function $f : \mathbb{R} \to \mathbb{R}$ with $\mathrm{lev}_0 f = C$ and being differentiable on $\mathbb{R} \setminus C$. In particular, the assumption (i) holds when T is firmly nonexpansive.

Proof. It suffices to observe that

$$T = \mathrm{Id} - N.$$

Since N is nonexpansive, T is monotone. Also note that T is firmly nonexpansive if and only if N is. Moreover, $N'(x) \leq 1$ is equivalent to $T'(x) \geq 0$. Thus (i) is the same as Corollary 5.54 (i). □

We illustrate Corollary 5.54 with three examples. They demonstrate that both conditions (i) and (ii) in Corollary 5.54 are needed. More precisely, (i) is for convexity of function f; (ii) is for $\mathrm{lev}_0 f = C$.

In the first example, T fails to be monotone but verified condition (ii) in Corollary 5.54. Then T is not the subgradient projector of a convex function.

Example 5.56. Define $T : \mathbb{R} \to \mathbb{R}$ by

$$T(x) := \begin{cases} x - \sqrt{x} + \sqrt{x} e^{-2\sqrt{x}} & \text{if } x > 0, \\ x & \text{if } x \leq 0. \end{cases}$$

Then T is a subgradient projector of the nonconvex function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) := \begin{cases} e^{2\sqrt{x}} - 1 & \text{if } x > 0 \\ x & \text{if } x \leq 0 \end{cases}$$

In this case, T fails to be monotone nor to be a cutter, but T verifies condition (ii) of Corollary 5.54.

Proof. When $x > 0$, $f'(x) = e^{2\sqrt{x}} x^{-1/2}$, so that $f''(x) = \frac{e^{2\sqrt{x}}\left(1 - \frac{1}{2\sqrt{x}}\right)}{x}$. Since $f''(x) < 0$ when $0 < x < 1/4$, f is not convex on $\mathbb{R}$. Now we show that (i)

T fails to be monotone. This is equivalent to verify that for some x we have $N'(x) > 1$ where $N(x) = x - Tx$. Indeed,

$$N'(x) = \left(\sqrt{x} - \frac{\sqrt{x}}{e^{2\sqrt{x}}}\right)' = \frac{1}{2}\frac{e^{2\sqrt{x}} - 1}{e^{2\sqrt{x}}\sqrt{x}} + e^{-2\sqrt{x}}.$$

L'Hospital rule gives $\lim_{x\to 0^+} \frac{e^{2\sqrt{x}}-1}{e^{2\sqrt{x}}\sqrt{x}} = 2$, so $\lim_{x\to 0^+} N'(x) = 2$. Therefore, T is not monotone.

(ii) T fails to be a cutter. From the definition of T, we have $\operatorname{Fix} T = \mathbb{R}^-$. Let $x = \frac{1}{4}$ and $u = 0 \in \operatorname{Fix} T$, then

$$\langle x - Tx, u - Tx\rangle \approx (\frac{1}{4} - (-0.066))(0 - (-0.066)) \approx 0.020856 > 0.$$

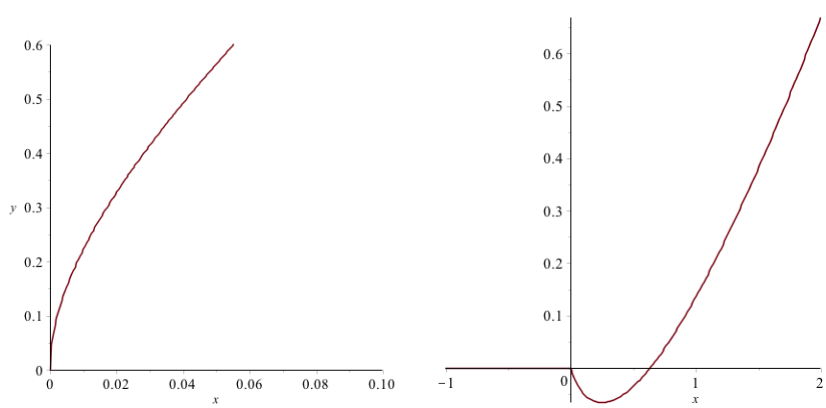
(iii) T verified condition (ii) of Corollary 5.54. For $x > 0$,

$$N(x) = \frac{e^{2\sqrt{x}} - 1}{e^{2\sqrt{x}}x^{-1/2}}.$$

With $a > 0$, we have

$$g(x) = \int_a^x \frac{1}{N(s)}ds = \int_a^x \frac{e^{2\sqrt{x}}x^{-1/2}}{e^{2\sqrt{x}} - 1}dx = \ln(e^{2\sqrt{x}} - 1) - \ln(e^{2\sqrt{a}} - 1).$$

Clearly, $\lim_{x\to 0^+} g(x) = -\infty$. Hence (ii) holds. □



(a) Nonconvex function f.

(b) The plot of T which is not monotonically increasing.

Figure 5.3: Illustration for Example 5.56

In the following example, both two Ts are the subgradient projectors of convex functions but Ts do not satisfy Corollary 5.54(ii).

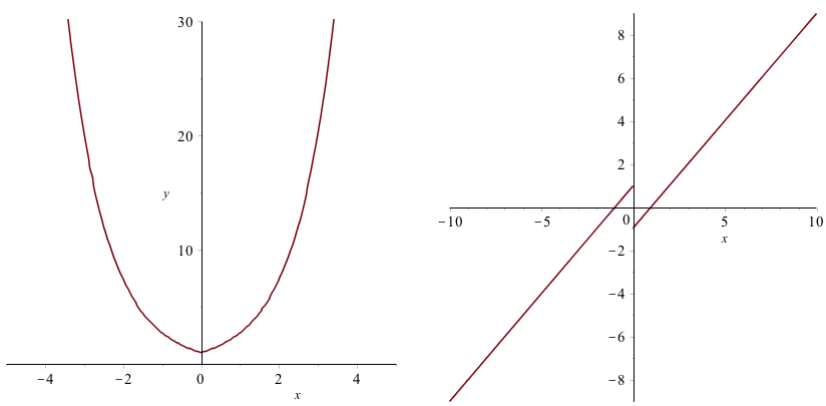
Example 5.57. (1) Define $T : \mathbb{R} \to \mathbb{R}$ by

$$x \mapsto T(x) = \begin{cases} x - 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ x + 1 & \text{if } x < 0. \end{cases}$$

Then $T = G_f$ where the convex function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$x \mapsto f(x) = e^{|x|}.$$

However, $\operatorname{lev}_0 f = \varnothing$ but $\operatorname{Fix} T = \{0\}$. In this case, in Corollary 5.54 condition (i) holds but condition (ii) fails.



(a) curve of convex function f.

(b) the plot of T which is not monotonically increasing.

Figure 5.4: Illustration for Example 5.57 (1)

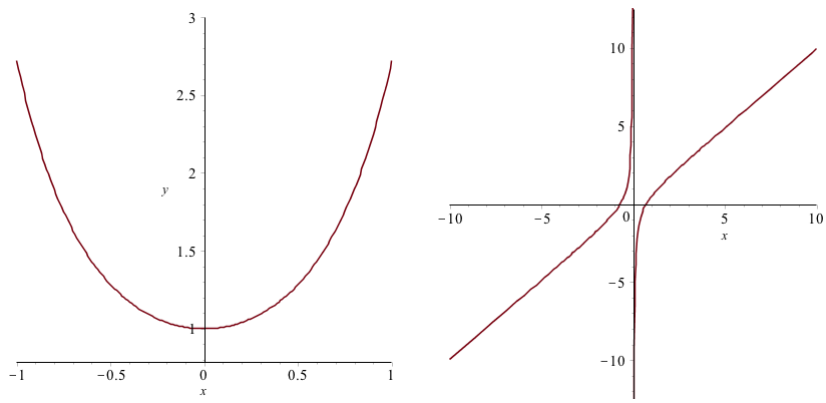
(2) Define $T : \mathbb{R} \to \mathbb{R}$ by

$$x \mapsto T(x) = \begin{cases} x - \frac{1}{2x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $T = G_f$ where $f : \mathbb{R} \to \mathbb{R}$ is given by

$$x \mapsto f(x) = e^{x^2}.$$

However, $\operatorname{lev}_0 f = \varnothing$ but $\operatorname{Fix} T = \{0\}$. In this case, in Corollary 5.54 condition (i) holds but condition (ii) fails.



(a) convex function f.

(b) the plot of T which is not monotonically increasing.

Figure 5.5: Illustration for Example 5.57 (2)

Proof. (1) We have

$$N(x) = x - G_f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Let $a > 0$. For $x > 0$ we have

$$g(x) = \int_a^x \frac{1}{N(s)} ds = \int_a^x 1 ds = x - a.$$

Thus, $\lim_{x \to a^+} g(x) = 0$. Therefore, (ii) of Corollary 5.54 fails.

(2) We have

$$N(x) = x - T(x) = \frac{1}{2x}$$

and $N'(x) = -\frac{1}{2x^2}$. Therefore, T is monotone on $(0, +\infty)$ and $(-\infty, 0)$. This says that condition (i) of Corollary 5.54 holds.

However, when $a > 0$, for $x > 0$ we have

$$g(x) = \int_a^x \frac{1}{N(s)} ds = \int_a^x 2s ds = x^2 - a^2.$$

Then $\lim_{x\to 0^+} g(x) = -a^2$, so condition (ii) of Corollary 5.54 fails. □

Note that T defined in Example 5.57(1) is a subgradient projector of a convex function, but is not a cutter. Actually, it is easy to see that $\operatorname{Fix} T = \{0\}$. Let $x = \frac{1}{2} > 0$, we have $Tx = x - 1 = -\frac{1}{2}$. Let $0 = y \in \operatorname{Fix} T$, we have

$$\langle x - Tx, y - Tx\rangle = \langle \frac{1}{2} - (-\frac{1}{2}), 0 - (-\frac{1}{2})\rangle = \frac{1}{2} > 0. \tag{5.126}$$

Then T is not a cutter.

Remark 5.58. Example 5.57(1) gives us an example of a subgradient projector of a convex function which is not a cutter. On the other hand, there is an example. $f(x) = 1$. f is convex on X. By Proposition 5.8(i), we have the subgradient projector of f is identity, which is a cutter. Thus we cannot conclude that the subgradient projector of a convex function is a cutter if we don't have condition of $\operatorname{lev}_0 f \neq \varnothing$.

Example 5.59. Define $T : \mathbb{R} \to \mathbb{R}$ by

$$x \mapsto T(x) = \begin{cases} x - \sqrt{x} & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ x - \sqrt{-x} & \text{if } x < 0 \end{cases} \tag{5.127}$$

Then $T = G_f$ where the nonconvex function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$x \mapsto f(x) = \begin{cases} e^{2\sqrt{x}} & \text{if } x \geq 0 \\ e^{-2\sqrt{-x}} & \text{if } x < 0. \end{cases} \tag{5.128}$$

However, $\operatorname{lev}_0 f = \varnothing$ but $\operatorname{Fix} T = \{0\}$. In this case, both conditions (i) and (ii) in Corollary 5.54 fail.

Proof. $f(x) = e^{2\sqrt{x}}$ is nonconvex on $[0, +\infty)$, see Example 5.56. $G_f = T$ follows by direct calculations.

Condition (i) of Corollary 5.54 fails: T is not monotonically increasing on $[0, +\infty)$ since $T'(x) = 1 - \frac{1}{2\sqrt{x}} < 0$ when $x > 0$ is sufficiently near 0.

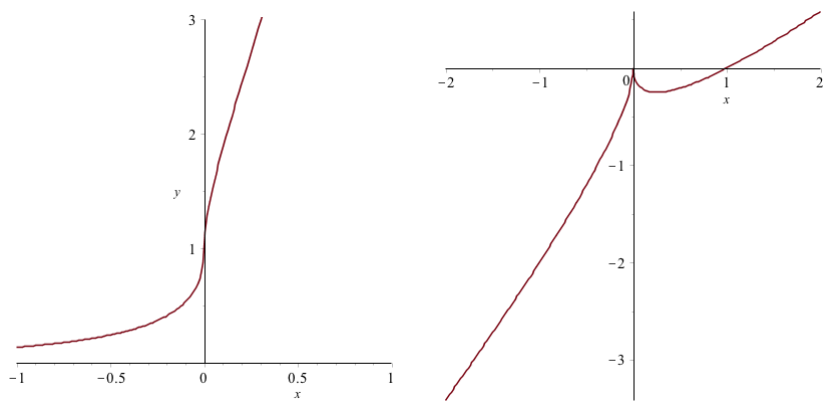
Condition (ii) of Corollary 5.54 fails. Indeed,

$$N(x) = x - T(x) = \sqrt{x} \tag{5.129}$$

when $x \geq 0$. When $a > 0$, for $x > 0$ we have

$$g(x) = \int_a^x \frac{1}{\sqrt{s}} ds = 2\sqrt{x} - 2\sqrt{a}, \tag{5.130}$$

so that $\lim_{x \to 0^+} g(x) = -2\sqrt{a}$. □



(a) Plot of nonconvex function f.

(b) The plot of T which is not monotonically increasing.

Figure 5.6: Illustration for Example 5.59.

This chapter presented basic properties of subgradient projectors of nonconvex functions. We also investigated when a subgradient projector is a cutter and when a mapping is a subgradient projector of a convex function.

Chapter 6

Linear Subgradient Projectors

6.1 Overview

In this chapter, based on the manuscript [11], we mainly focus on the properties of G_f when it is linear. We characterize symmetric linear subgradient projectors and obtain a closed form. It turns out that the set of subgradient projectors is not convex.

6.2 Characterizations of G_f when G_f is linear

Proposition 6.1. *Let T be a linear operator. Then T is a cutter if and only if T is firmly nonexpansive.*

Proof. $\Rightarrow$: Assume that T is a cutter. Then for every $x \in X$ and $u \in \operatorname{Fix} T$,

$$\langle x - Tx, u - Tx \rangle = \langle Tx - x, Tx - u \rangle \leq 0. \tag{6.1}$$

Put $u = 0$. Since T is linear, thus T mapps 0 to 0. Then $0 \in \operatorname{Fix} T$. So

$$\langle Tx - x, Tx - 0 \rangle \leq 0 \quad \Rightarrow \quad \|Tx\|^2 \leq \langle x, Tx \rangle. \tag{6.2}$$

By (2.44), T is firmly nonexpansive.

$\Leftarrow$: Assume that T is firmly nonexpansive. Let $u \in \operatorname{Fix} T$. Then $Tu = u$ and

$$\begin{aligned}
\langle Tx - x, Tx - u \rangle &= \langle Tx - x, Tx - Tu \rangle && (6.3)\\
&= \langle Tx - Tu + Tu - x, Tx - Tu \rangle && (6.4)\\
&= \|Tx - Tu\|^2 + \langle Tu - x, Tx - Tu \rangle && (6.5)\\
&= \|Tx - Tu\|^2 - \langle x - u, Tx - Tu \rangle \leq 0. && (6.6)
\end{aligned}$$

Hence T is a cutter. $\square$

The following example says that Proposition 6.1 fails if T is not linear.

Example 6.2. Define a continuous nonlinear $T : \mathbb{R} \to \mathbb{R}$ by

$$T(x) = \begin{cases} x/2 & \text{if } -2 \le x \le 2, \\ 3 - x & \text{if } 2 \le x \le 3, \\ -(3+x) & \text{if } -3 \le x \le -2, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

Then T is a cutter but not firmly nonexpansive as T is not monotone (referring to Theorem 2.32).

Proof. (i) It is easy to obtain that $\operatorname{Fix} T = \{0\}$. This means that T is a cutter if and only if $(T(x))^2 \le xT(x)$.
When $-2 \le x \le 2$, we have

$$(T(x))^2 = \frac{x^2}{4} \le x\frac{x}{2} = xT(x); \tag{6.8}$$

When $2 \le x \le 3$, we have

$$(T(x))^2 = (3-x)^2 = (3-x)(3-x) \le x(3-x); \tag{6.9}$$

when $-3 \le x \le -2$, we have

$$(T(x))^2 = [-(x+3)]^2 = [-(x+3)][-(x+3)] \le x[-(x+3)]; \tag{6.10}$$

when $|x| > 3$, we have

$$(T(x))^2 = 0 = xT(x). \tag{6.11}$$

Hence T is a cutter.

(ii) T is not monotone. Actually, $\forall x, y \in [2, 3]$, the following is true.

$$\langle Tx - Ty, x - y\rangle = \langle (3-x) - (3-y), x - y\rangle = -(x-y)^2 \le 0. \tag{6.12}$$

Therefore we conclude that T is not monotone, either firmly nonexpansive.

□

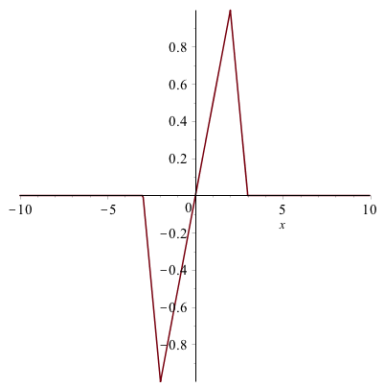


Figure 6.1: Graph of T defined in Example 6.2.

Observe that Example 6.2 is much simpler than [20, Example 2.2.8, page 68].

Theorem 6.3. *Let $a > 0$ and $B \neq 0$ being an $n \times n$ symmetric and positive semidefinite matrix. Consider the function*

$$(\forall x \in X) \quad f(x) = (x^\intercal Bx)^{1/(2a)} \tag{6.13}$$

(i) We have

$$G_f(x) = \begin{cases} x - a\frac{x^\intercal Bx}{\|Bx\|^2}Bx & \text{if } Bx \neq 0, \\ x & \text{if } Bx = 0. \end{cases} \tag{6.14}$$

(ii) G_f is linear if and only if $B = \lambda P_L$ where $\lambda > 0$ and $L \subseteq X$ is a subspace. In this case

$$\ker B = L^\perp, \tag{6.15}$$

$$f(x) = \lambda^{1/(2a)}\big(d_{L^\perp}(x)\big)^{1/a} \text{ and} \tag{6.16}$$

$$G_f = \operatorname{Id} - aP_L = (1-a)\operatorname{Id} + aP_{L^\perp}. \tag{6.17}$$

(iii) Assume that G_f is linear. Then G_f is a cutter if and only if $0 < a \leq 1$.

Proof. (i) As B is positive semidefinite, i.e., $(\forall x \in X)\ x^\intercal Bx \geq 0$. $f(x) \leq 0$ if and only if $f(x) = 0$, which is equivalent to $Bx = 0$. G_f follows from direct calculations.

(ii) $\Rightarrow$: Assume that G_f is linear. The mapping

$$x \mapsto T_1(x) := a^{-1}\big(x - G_f(x)\big) = \begin{cases} \frac{x^\intercal Bx}{\|Bx\|^2} Bx & \text{if } Bx \neq 0 \\ 0 & \text{if } Bx = 0 \end{cases} \tag{6.18}$$

is linear. Let λ_i be eigenvalues of B, where $i = 1, 2$. Assume that $\lambda_1, \lambda_2 > 0$. We want to show that $\lambda_1 = \lambda_2$. We prove it by contradiction. Suppose that $\lambda_1 \neq \lambda_2$. Take unit length eigenvector v_i associated with λ_i. Note that $\langle v_1, v_2\rangle = 0$, $Bv_i \neq 0$ and $B(v_1 + v_2) = \lambda_1 v_1 + \lambda_2 v_2 \neq 0$. As T_1 is linear, we have $T_1(v_1 + v_2) = T_1 v_1 + T_1 v_2$. Now

$$\begin{aligned}
T_1(v_1 + v_2) &= \frac{(v_1 + v_2)^\intercal B(v_1 + v_2)}{\|B(v_1 + v_2)\|^2} B(v_1 + v_2) && (6.19)\\
&= \frac{(v_1 + v_2)^\intercal(\lambda_1 v_1 + \lambda_2 v_2)}{\|\lambda_1 v_1 + \lambda_2 v_2)\|^2}(\lambda_1 v_1 + \lambda_2 v_2) && (6.20)\\
&= \frac{\lambda_1 v_2^\intercal v_1 + \lambda_2 v_2^\intercal v_2 + \lambda_1 v_1^\intercal v_1 + \lambda_2 v_1^\intercal v_2}{\lambda_1^2 + \lambda_2^2}(\lambda_1 v_1 + \lambda_2 v_2) && (6.21)\\
&= \frac{\lambda_1 + \lambda_2}{\lambda_1^2 + \lambda_2^2}(\lambda_1 v_1 + \lambda_2 v_2), && (6.22)\\
T_1 v_1 + T_1 v_2 &= \frac{v_1^\intercal B v_1}{\|Bv_1\|^2} Bv_1 + \frac{v_2^\intercal B v_2}{\|Bv_2\|^2} Bv_2 && (6.23)\\
&= \frac{\lambda_1 \|v_1\|^2}{\|\lambda_1 v_1\|^2}\lambda_1 v_1 + \frac{\lambda_2 \|v_2\|^2}{\|\lambda_2 v_2\|^2}\lambda_2 v_2 && (6.24)\\
&= v_1 + v_2 && (6.25)
\end{aligned}$$

As $\{v_1, v_2\}$ are linearly independent, the above gives $\lambda_1 = \lambda_2$ which contradicts $\lambda_1 \neq \lambda_2$. Therefore, all positive eigenvalues of B have to be equal. Hence, we have

$$B = \lambda U^\intercal \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix} U \tag{6.26}$$

where U is an orthogonal matrix, $\lambda > 0$, Id is an $m \times m$ identity matrix for some $m \geq 0$. The matrix

$$U^\intercal \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix} U \tag{6.27}$$

is idempotent[20] and symmetric, so it is a matrix associated with an orthogonal projection on a closed subspace, say P_L, [37, page 430, page 433]. Hence

$$B = \lambda P_L \tag{6.28}$$

[20] We say A is an idempotent operator if and only if $AAx = Ax$ for $x \in X$.

which implies that $Bx = 0$ if and only if $P_L x = 0$, i.e., $\ker B = L^\perp$. Then when $P_L x \neq 0$,

$$T_1(x) = \frac{x^\intercal Bx}{\|Bx\|^2}Bx = \frac{\lambda x^\intercal P_L x}{x^\intercal \lambda P_L \lambda P_L x}\lambda P_L x = \frac{\lambda x^\intercal P_L x}{\lambda^2 x^\intercal P_L x}\lambda P_L x = P_L x; \quad (6.29)$$

when $P_L x = 0$, $T_1 x = 0 = P_L x$. Hence $T_1 = P_L$. It follows that

$$G_f = \operatorname{Id} - aT_1 = \operatorname{Id} - aP_L = (1-a)\operatorname{Id} + a(\operatorname{Id} - P_L) = (1-a)\operatorname{Id} + aP_{L^\perp}. \quad (6.30)$$

We proceed to find the expression of $f(x)$:

$$\begin{aligned} f(x) &= (x^\intercal Bx)^{1/(2a)} = (x^\intercal \lambda P_L x)^{1/(2a)} && (6.31)\\ &= \lambda^{1/(2a)}(x^\intercal P_L P_L x)^{1/(2a)} = \lambda^{1/(2a)}(\|P_L x\|^2)^{1/(2a)} && (6.32)\\ &= \lambda^{1/(2a)}(\|x - P_{L^\perp}x\|^2)^{1/(2a)} = \lambda^{1/(2a)}(d_{L^\perp}(x)^2)^{1/(2a)} && (6.33)\\ &= \lambda^{1/(2a)}\big(d_{L^\perp}(x)\big)^{1/a} && (6.34) \end{aligned}$$

$\Leftarrow$: Assume that $B = \lambda P_L$ for $\lambda > 0$ and some subspace $L \subseteq X$. The assumption gives

$$f(x) = \lambda^{1/(2a)}\big(d_{L^\perp}(x)\big)^{1/a}. \quad (6.35)$$

By Fact 5.18,

$$G_f = G_{\left(d_{L^\perp}\right)^{1/a}}. \quad (6.36)$$

By Theorem 5.27,

$$G_{\left(d_{L^\perp}\right)^{1/a}} = (1-a)\operatorname{Id} + aG_{d_{L^\perp}}. \quad (6.37)$$

By Fact 6.5,

$$G_{\left(d_{L^\perp}\right)^{1/a}} = (1-a)\operatorname{Id} + aP_{L^\perp}. \quad (6.38)$$

Since $P_{L^\perp}$ is linear as $L^\perp$ is a subspace. Hence G_f is linear.

(iii) $\Rightarrow$: Assume that G_f is a linear cutter. By Fact 6.1, G_f is firmly nonexpansive, so is $\operatorname{Id} - G_f$. By (ii), $\operatorname{Id} - G_f = aP_L$, aP_L has to be nonexpansive. Take $0 \neq x \in L$. The nonexpansiveness requires

$$\|aP_L x - aP_L 0\| = \|ax\| \leq \|x\| \quad (6.39)$$

so that $a \leq 1$.

$\Leftarrow$: Assume that $0 < a \leq 1$. Since $x \mapsto (x^\intercal Bx)^{1/2}$ is convex, and the function

$$[0, +\infty) \ni t \mapsto t^{1/a} \quad (6.40)$$

is convex and inceasing when $0 < a \leq 1$. By [41, Proposition 1.39], we have that

$$x \mapsto f(x) = \big((x^\intercal Bx)^{1/2}\big)^{1/a} \tag{6.41}$$

is convex. By Example 2.26, we have G_f is a cutter. □

We illustrate Theorem 6.3(iii) with the following example.

Example 6.4. Let $a > 1$. Consider

$$X \ni x \mapsto f(x) = (x^\intercal x)^{1/(2a)} = \|x\|^{1/a}. \tag{6.42}$$

Then f is not convex, and

$$G_f(x) = (1-a)x \tag{6.43}$$

for every $x \in X$. Although G_f is linear, but it is not a cutter since it is not monotone, see, e.g., Proposition 6.1.

For a set $C \subset X$, its distance function $d_C : X \to [0, +\infty)$ is defined by

$$X \ni x \mapsto d_C(x) := \inf\{\|x - y\| : y \in C\}. \tag{6.44}$$

Fact 6.5. ([5]) *Let $C \subseteq X$ be closed and convex, and $f = d_C$. Then $G_f = P_C$.*

The following result completely characterizes symmetric and linear subgradient projectors.

Theorem 6.6. *Assume that $T : X \to X$ is linear and symmetric. Then the following are equivalent.*

(i) T is a subgradient projector of a convex function.

(ii) $T = G_f$ where $f : X \to \mathbb{R}$ is given by

$$f(x) = K(x^\intercal P_L x)^{1/(2\lambda)} = K\big(d_{L^\perp}(x)\big)^{1/\lambda} \tag{6.45}$$

where $0 < \lambda \leq 1$, $K > 0$ and $L \subseteq X$ is a subspace. In this case, $G_f = (1-\lambda)\,\mathrm{Id} + \lambda P_{L^\perp}$.

Proof. (i)⇒(ii): Assume that $T = G_f$ for some convex function. Since T is linear and a cutter, T is firmly nonexpansive by Proposition 6.1. Then $T_1 = \mathrm{Id} - T$ is firmly nonexpansive. We consider two cases.
Case 1: $\operatorname{int}\operatorname{lev}_0 f \neq \varnothing$. Let $x_0 \in \operatorname{int}\operatorname{lev}_0 f$, i.e., there is $\varepsilon > 0$ such that

$\operatorname{ball}(x_0, \varepsilon) \subseteq \operatorname{lev}_0 f$. We now show that $T_1 \equiv 0$ on $\operatorname{ball}(x_0, \varepsilon)$. Actually, $\forall b \in X$ with $\|b\| \leq \varepsilon$, we have $x_0 + b \in \operatorname{ball}(x_0, \varepsilon) \subseteq \operatorname{lev}_0 f$, thus $T(x_0 + b) = x_0 + b$, so $T_1(x_0 + b) = 0$. Moreover, since T is linear, so is T_1, thus $T_1(b) = T_1(x_0 + b) - T_1(x_0) = 0$. Then $T_1 \equiv 0$ on $\operatorname{ball}(0, \varepsilon)$. Next we show $T_1 \equiv 0$ outside of $\operatorname{ball}(0, \varepsilon)$. Indeed, let $x \in X \setminus \operatorname{ball}(0, \varepsilon)$. Then $\frac{x}{\|x\|}\varepsilon \in \operatorname{ball}(0, \varepsilon)$, thus we have $T_1(\frac{x}{\|x\|}\varepsilon) = \frac{\varepsilon}{\|x\|}T_1(x) = 0$. Therefore $T_1 \equiv 0$ on X. Thus, $T = \operatorname{Id}$ on X. Then $T = G_f$ with $f \equiv 0$. This that means (ii) holds with $L = \{0\}$, $\lambda = 1$ and $K > 0$.

Case 2: $\operatorname{int} \operatorname{lev}_0 f = \varnothing$. As T_1 is continuous, we only need to consider

$$T_1(x) = \frac{f(x)}{\|\nabla f(x)\|^2} \nabla f(x) \quad \text{when } f(x) > 0. \tag{6.46}$$

Then

$$\|T_1(x)\| = \frac{f(x)}{\|\nabla f(x)\|} \tag{6.47}$$

and

$$\frac{\nabla f(x)}{f(x)} = \frac{T_1(x)}{\|T_1(x)\|^2}. \tag{6.48}$$

Since T is symmetric, T_1 is symmetric, so there exists an orthogonal matrix Q such that $Q^\intercal T_1 Q = D$ where D is an diagonal matrix and $Q^\intercal$ denotes the transpose of Q. Put $g = \ln f$ and $x = Qy$. We have

$$(\nabla g)(Qy) = \frac{T_1 Qy}{\|T_1 Qy\|^2}. \tag{6.49}$$

Multiplying both sides by $Q^\intercal$ and using $Q^\intercal$ being an isometry (i.e., $\|Q^\intercal z\| = \|z\|$ for every $z \in X$) give

$$Q^\intercal(\nabla g)(Qy) = \frac{Q^\intercal T_1 Qy}{\|T_1 Qy\|^2} = \frac{Dy}{\|Q^\intercal T_1 Qy\|^2} = \frac{Dy}{\|Dy\|^2}. \tag{6.50}$$

If we put $h = g \circ Q$, then $\nabla h(y) = Q^\intercal \nabla g(Qy)$ for every $y \in X$, so

$$\nabla h(y) = \frac{Dy}{\|Dy\|^2}. \tag{6.51}$$

Write

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{6.52}$$

When $\lambda_1 = \cdots = \lambda_n = 0$, this is covered in Case 1. We can assume that $T_1 \not\equiv 0$. As T_1 is monotone, we can and do assume that $\lambda_1, \cdots, \lambda_m > 0$ and $\lambda_{m+1} = \cdots = \lambda_n = 0$. Then

$$\nabla h(y) = \left(\frac{\lambda_1 y_1}{\sum_{k=1}^m \lambda_k^2 y_k^2}, \cdots, \frac{\lambda_m y_m}{\sum_{k=1}^m \lambda_k^2 y_k^2}, 0, \cdots, 0 \right). \tag{6.53}$$

Since h has continuous second order derivatives,

$$\frac{\partial^2 h}{\partial y_i \partial y_j} = \frac{\partial^2 h}{\partial y_j \partial y_i} \tag{6.54}$$

which gives

$$\frac{-2\lambda_j \lambda_i^2 y_i y_j}{\sum_{k=1}^m \lambda_k^2 y_k^2} = \frac{-2\lambda_i \lambda_j^2 y_i y_j}{\sum_{k=1}^m \lambda_k^2 y_k^2} \tag{6.55}$$

when $1 \leq i, j \leq m$, $i \neq j$. As $\operatorname{int} \operatorname{lev}_0 f = \varnothing$, (6.55) holds on open subset of X, so we have $\lambda_i = \lambda_j$. Because $1 \leq i, j \leq m$ were arbitrary, we obtain that $\lambda_1 = \cdots = \lambda_m$. Hence

$$T_1 = Q \begin{pmatrix} \lambda \operatorname{Id}_m & 0 \\ 0 & 0 \end{pmatrix} Q^\intercal = \lambda Q \begin{pmatrix} \operatorname{Id}_m & 0 \\ 0 & 0 \end{pmatrix} Q^\intercal = \lambda P_L \tag{6.56}$$

where $L \subseteq X$ is a linear subspace, see [37, page 430]. Namely, T_1 is a positive multiple of an orthogonal projector.

Now T_1 is firmly nonexpansive and $T_1^\intercal = T_1$, this implies that

$$\frac{T_1 + T_1^\intercal}{2} - T_1^\intercal T_1 = T_1 - T_1^2 = Q \begin{pmatrix} (\lambda - \lambda^2) \operatorname{Id} & 0 \\ 0 & \end{pmatrix} Q \tag{6.57}$$

is positive semidefinite, so $0 \leq \lambda \leq 1$. As $T_1 \neq 0$ in this case, $0 < \lambda \leq 1$.

Therefore,

$$\nabla \ln f(x) = \frac{T_1 x}{\|T_1 x\|^2} = \frac{\lambda P_L x}{\|\lambda P_L x\|^2} = \frac{1}{\lambda} \frac{P_L x}{\|P_L x\|^2}. \tag{6.58}$$

Note that $P_L^\intercal = P_L$, $P_L^2 = P_L$,

$$\nabla \ln \|P_L x\| = \frac{1}{\|P_L x\|} \nabla \|P_L x\| = \frac{1}{\|P_L x\|} P_L^\intercal \frac{P_L x}{\|P_L x\|} = \frac{P_L x}{\|P_L x\|^2}. \tag{6.59}$$

It follows that

$$\nabla \ln f(x) = \frac{1}{\lambda} \nabla \ln \|P_L x\| = \nabla \ln \|P_L x\|^{1/\lambda} \tag{6.60}$$

which is equivalent to

$$\ln f(x) = \ln \|P_L x\|^{1/\lambda} + c \tag{6.61}$$

for some constant $c \in \mathbb{R}$. Taking exp both sides gives

$$f(x) = K\|P_L x\|^{1/\lambda} = K(\|P_L x\|^2)^{1/(2\lambda)} = K(x^\intercal P_L x)^{1/(2\lambda)} \tag{6.62}$$

where $K = \exp(c) > 0$. As $P_L = \mathrm{Id} - P_{L^\perp}$,

$$f(x) = K\|x - P_\perp x\|^{1/\lambda} = K(d_{L^\perp}(x))^{1/\lambda}. \tag{6.63}$$

Finally,

$$T = G_f = \mathrm{Id} - T_1 = \mathrm{Id} - \lambda P_L = \mathrm{Id} - \lambda(\mathrm{Id} - P_{L^\perp}) = (1-\lambda)\,\mathrm{Id} + \lambda P_{L^\perp}. \tag{6.64}$$

□

Theorem 6.7. *Subgradient projectors of convex functions are not closed under convex combination.*

Proof. Let $L := \{0\} \times \mathbb{R} \subseteq \mathbb{R}^2$ and $M := \{(x,y) \in \mathbb{R}^2 : \langle \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \rangle = 0\}$. Define two functions

$$(\forall x \in \mathbb{R}^2)\ f(x) = K_1\big(d_{L^\perp}(x)\big)^{1/\lambda_1},\ g(x) = K_2\big(d_{M^\perp}(x)\big)^{1/\lambda_2} \tag{6.65}$$

where $0 < \lambda_1 \neq \lambda_2 < 1$. By Theorem 6.6, we have

$$G_f = (1-\lambda_1)\,\mathrm{Id} + \lambda_1 P_{L^\perp},\ G_g = (1-\lambda_2)\,\mathrm{Id} + \lambda_2 P_{M^\perp}. \tag{6.66}$$

Now consider $\lambda_3 G_f + (1-\lambda_3)G_g$ where $0 < \lambda_3 < 1, \lambda_3 \neq \lambda_1, \lambda_2$. Then

$$\lambda_3 G_f + (1-\lambda_3)G_g \tag{6.67}$$

$$= \lambda_3\big((1-\lambda_1)\,\mathrm{Id} + \lambda_1 P_{L^\perp}\big) + (1-\lambda_3)\big((1-\lambda_2)\,\mathrm{Id} + \lambda_2 P_{M^\perp}\big) \tag{6.68}$$

$$= (1 - \lambda_2 + \lambda_2\lambda_3 - \lambda_1\lambda_3)\,\mathrm{Id} + \lambda_1\lambda_3 P_{L^\perp} + \lambda_2(1-\lambda_3)P_{M^\perp} \tag{6.69}$$

If $\lambda_3 G_f + (1-\lambda_3)G_g$ is a subgradient projector, by Theorem 6.6, there are $0 < \lambda < 1$ and S which is a subspace of $\mathbb{R}^2$ such that

$$\lambda_3 G_f + (1-\lambda_3)G_g = (1-\lambda)\,\mathrm{Id} + \lambda P_{S^\perp}. \tag{6.70}$$

Therefore we have

$$(1 - \lambda_2 + \lambda_2\lambda_3 - \lambda_1\lambda_3)\,\mathrm{Id} + \lambda_1\lambda_3 P_{L^\perp} + \lambda_2(1-\lambda_3)P_{M^\perp} = (1-\lambda)\,\mathrm{Id} + \lambda P_{S^\perp}. \tag{6.71}$$

Natually, the set of fixed points of left-hand side is equal to the set of fixed points of right-hand side. Thus we have

$$\begin{aligned} &\operatorname{Fix}((1-\lambda)\operatorname{Id}+\lambda P_{S^\perp}) && (6.72)\\ =&\operatorname{Fix}((1-\lambda_2+\lambda_2\lambda_3-\lambda_1\lambda_3)\operatorname{Id}+\lambda_1\lambda_3 P_{L^\perp}+\lambda_2(1-\lambda_3)P_{M^\perp}). && (6.73)\end{aligned}$$

By [7, Proposition 4.34], we have

$$\operatorname{Fix}((1-\lambda_2+\lambda_2\lambda_3-\lambda_1\lambda_3)\operatorname{Id}+\lambda_1\lambda_3 P_{L^\perp}+\lambda_2(1-\lambda_3)P_{M^\perp}) = L^\perp\cap M^\perp. \tag{6.74}$$

Also,

$$\operatorname{Fix}((1-\lambda)\operatorname{Id}+\lambda P_{S^\perp}) = S^\perp. \tag{6.75}$$

Hence

$$\begin{aligned} \{(0,0)\} &= L^\perp\cap M^\perp && (6.76)\\ &= \operatorname{Fix}((1-\lambda_2+\lambda_2\lambda_3-\lambda_1\lambda_3)\operatorname{Id}+\lambda_1\lambda_3 P_{L^\perp}+\lambda_2(1-\lambda_3)P_{M^\perp}) && (6.77)\\ &= \operatorname{Fix}((1-\lambda)\operatorname{Id}+\lambda P_{S^\perp}) && (6.78)\\ &= S^\perp && (6.79)\end{aligned}$$

Therefore $S^\perp = \{(0,0)\}$, it imples $S = \mathbb{R}^2$. Now we view a projection operator as a matrix. Then we have

$$P_{L^\perp} = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ P_{M^\perp} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix},\ P_{S^\perp}\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}. \tag{6.80}$$

So we can see that $P_{L^\perp}, P_{S^\perp}$ are diagonal matrices, but $P_{M^\perp}$ is not. Hence, (6.71) is not true. This means $\lambda_3 G_f + (1-\lambda_3)G_g$ is not a subgradient projector. □

Theorem 6.8. *Assume $0<\lambda_1<1, 0<\lambda_2<1, 0<\lambda<1$. Suppose that L, M are linear subspaces of X satisfying $L = M^\perp, M = L^\perp$ and L, M are not $\{0\}$ nor X. Let*

$$G_f = (1-\lambda_1)\operatorname{Id}+\lambda_1 P_{L^\perp},\ G_g = (1-\lambda_2)\operatorname{Id}+\lambda_2 P_{M^\perp}. \tag{6.81}$$

If $\frac{1-\lambda}{\lambda}\neq\frac{\lambda_2}{\lambda_1}$, then $(1-\lambda)G_f+\lambda G_g$ is not a subgradient projector of a convex function.

Proof. We have

$$
\begin{aligned}
&(1-\lambda)G_f + \lambda G_g && (6.82)\\
=&(1-\lambda)\left((1-\lambda_1)\operatorname{Id} + \lambda_1 P_{L^\perp}\right) + \lambda\left((1-\lambda_2)\operatorname{Id} + \lambda_2 P_{M^\perp}\right) && (6.83)\\
=&\left[(1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2)\right]\operatorname{Id} + \lambda_1(1-\lambda)P_{L^\perp} + \lambda\lambda_2 P_{M^\perp} && (6.84)\\
=&\beta\operatorname{Id} + (1-\beta)(\gamma P_{L^\perp} + (1-\gamma)P_{M^\perp}) && (6.85)
\end{aligned}
$$

where $\beta = (1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2)$ and $\gamma = \frac{\lambda_1(1-\lambda)}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)}$. We next show that $0 < \beta < 1$ and $\gamma \neq \frac{1}{2}$. To see this,

$$
0 < \beta = (1-\lambda)(1-\lambda_1) + \lambda(1-\lambda_2) < (1-\lambda) + \lambda = 1. \tag{6.86}
$$

Also,

$$
\begin{aligned}
&\gamma = \frac{1}{2} && (6.87)\\
\Longleftrightarrow&\frac{\lambda_1(1-\lambda)}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)} = \frac{\lambda\lambda_2}{1-(1-\lambda)(1-\lambda_1)-\lambda(1-\lambda_2)} && (6.88)\\
\Longleftrightarrow&\lambda_1(1-\lambda) = \lambda\lambda_2 && (6.89)\\
\Longleftrightarrow&\frac{1-\lambda}{\lambda} = \frac{\lambda_2}{\lambda_1}. && (6.90)
\end{aligned}
$$

By the assumption: $\frac{1-\lambda}{\lambda} \neq \frac{\lambda_2}{\lambda_1}$, so $\gamma \neq \frac{1}{2}$. Since G_f, G_g are linear and symmetric, so is $(1-\lambda)G_f + \lambda G_g$.
If $(1-\lambda)G_f + \lambda G_g$ is a subgradient projector of a convex function, by Theorem 6.6, we have

$$
(1-\lambda)G_f + \lambda G_g = (1-\alpha)\operatorname{Id} + \alpha P_{S^\perp} \tag{6.91}
$$

where $0 < \alpha < 1$ and S is a subspace of X. Note that G_f, G_g are averaged mapping[21], so is $(1-\lambda)G_f + \lambda G_g$. As

$$
\operatorname{Fix} G_f = L^\perp,\ \operatorname{Fix} G_g = M^\perp. \tag{6.92}
$$

By [7, Proposition 4.34], we obtain

$$
\operatorname{Fix}((1-\lambda)G_f + \lambda G_g) = \operatorname{Fix} G_f \bigcap \operatorname{Fix} G_g = L^\perp \bigcap M^\perp = \{0\}. \tag{6.93}
$$

[21](See [7, Definition 4.23]) Let $t \in \,]0,1[$ and $T : X \to X$. Then T is averaged with constant t if there exists a nonexpansive operator N such that $T = (1-t)\operatorname{Id} + tN$.

As

$$\operatorname{Fix}((1-\alpha)\operatorname{Id}+\alpha P_{S^\perp}) = S^\perp. \tag{6.94}$$

Thus we have $S^\perp = \{0\}$. Therefore, refer to (6.91), we have

$$(1-\lambda)G_f + \lambda G_g = (1-\alpha)\operatorname{Id}. \tag{6.95}$$

Equivalently,

$$\beta\operatorname{Id}+(1-\beta)(\gamma P_{L^\perp} + (1-\gamma)P_{M^\perp}) = (1-\alpha)\operatorname{Id}. \tag{6.96}$$

We proceed to analyze α, β. Take $x \in M, x \neq 0$. Then $P_{M^\perp}x = 0, P_{L^\perp}x = P_M x = x$. (6.96) gives

$$\beta x + (1-\beta)(\gamma x) = (1-\alpha)x. \tag{6.97}$$

It implies

$$\beta + (1-\beta)\gamma = 1-\alpha. \tag{6.98}$$

Take $x \in L, x \neq 0$. Then $P_{L^\perp}x = 0, P_{M^\perp}x = P_L x = x$. (6.96) gives

$$\beta x + (1-\beta)(1-\gamma)x = (1-\alpha)x. \tag{6.99}$$

It implies

$$\beta + (1-\beta)(1-\gamma) = 1-\alpha. \tag{6.100}$$

Substracting (6.98) from (6.100), we have

$$(1-\beta)(1-2\gamma) = 0 \tag{6.101}$$

which implies $\beta = 1$ or $\gamma = \frac{1}{2}$. This contradicts the choices of $\lambda, \lambda_1, \lambda_2$. □

Question: Why must G_f be symmetric when G_f is linear and f is convex? When f is not convex, we know that G_f can be nonsymmetric, see Example 6.14 (2).

Proposition 6.9. *Let $M : X \to X$ be linear, monotone and*

$$(\forall\ x \in X \text{ with } Mx \neq 0) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2} \tag{6.102}$$

where the function $h : \{x \in X : Mx \neq 0\} \to \mathbb{R}$. If $\dim \operatorname{ran} M \neq 2$, then M is symmetric.

Proof. If $\dim \operatorname{ran} M = 0$, then $M = 0$, so it is symmetric. Let us assume that $\dim \operatorname{ran} M > 0$ and $\dim \operatorname{ran} M \neq 2$. Since h has continuous mixed second order derivatives at x whenever $Mx \neq 0$, the Hessian matrix $\nabla^2 h(x)$ is symmetric. As

$$\nabla^2 h(x) = \frac{\|Mx\|^2 M - Mx(\nabla \|Mx\|^2)^\intercal}{\|Mx\|^4} = \frac{\|Mx\|^2 M - 2Mxx^\intercal M^\intercal M}{\|Mx\|^4}, \tag{6.103}$$

the symmetric property means that

$$\|Mx\|^2 M - 2Mxx^\intercal M^\intercal M = (\|Mx\|^2 M - 2Mxx^\intercal M^\intercal M)^\intercal \tag{6.104}$$

$$= M^\intercal \|Mx\|^2 - 2M^\intercal Mxx^\intercal M^\intercal \tag{6.105}$$

whenever $Mx \neq 0$. Put $y = Mx$. The above is simplified to

$$\frac{M - M^\intercal}{2} = \frac{yy^\intercal}{\|y\|^2} M - M^\intercal \frac{yy^\intercal}{\|y\|^2}. \tag{6.106}$$

Denote the projection operator on the line spanned by $\{y\}$, $\operatorname{span}(y)$, by $P_y = \frac{yy^\intercal}{\|y\|^2}$. We have

$$(\forall\ y \in \operatorname{ran} M) \quad \frac{M - M^\intercal}{2} = P_y M - M^\intercal P_y. \tag{6.107}$$

Since M is monotone, $\operatorname{ran} M = \operatorname{ran} M^\intercal$, see, e,g, [14, Theorem 3.2]. Let $\{e_i : i = 1, \ldots, m\}$ be an orthonormal basis of $\operatorname{ran} M$. Then

$$P_{\operatorname{ran} M} = \sum_{i=1}^{m} e_i e_i^\intercal. \tag{6.108}$$

Note that

$$M^\intercal P_{\operatorname{ran} M} = M^\intercal P_{\operatorname{ran} M^\intercal} = M^\intercal (P_{\operatorname{ran} M^\intercal} + P_{(\operatorname{ran} M^\intercal)^\perp}) = M^\intercal \tag{6.109}$$

because $M^\intercal P_{(\operatorname{ran} M^\intercal)^\perp} = 0$. To see this, let $y \in (\operatorname{ran} M^\intercal)^\perp$. For every $z \in X$,

$$\langle M^\intercal y, z \rangle = \langle y, Mz \rangle = 0 \tag{6.110}$$

because $Mz \in \operatorname{ran} M = \operatorname{ran} M^\intercal$. As $z \in X$ was arbitrary, we must have $M^\intercal y = 0$. Since

$$\frac{M - M^\intercal}{2} = P_{e_i} M - M^\intercal P_{e_i} \tag{6.111}$$

by (6.107), summing up (6.111) from from $i = 1$ to $i = m$, followed by using (6.108) and (6.109), we obtain

$$\frac{m}{2}(M-M^\intercal) = (\sum_{i=1}^{m} P_{e_i})M-M^\intercal(\sum_{i=1}^{m} P_{e_i}) = P_{\operatorname{ran} M}M-M^\intercal P_{\operatorname{ran} M} = M-M^\intercal, \tag{6.112}$$

that is,

$$(\frac{m}{2}-1)(M-M^\intercal) = 0. \tag{6.113}$$

Hence $M - M^\intercal = 0$, and so M is symmetric. □

Proposition 6.9 fails when $\dim \operatorname{ran} M = 2$ as the following example shows.

Example 6.10. When $\dim \operatorname{ran} M = 2$, although $M : \mathbb{R}^2 \to \mathbb{R}^2$ is linear, monotone and

$$(\forall\ Mx \neq 0) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2}, \tag{6.114}$$

one cannot guarantee that M is symmetric.

Let $\mathbf{x} = (x, y)^\intercal \in \mathbb{R}^2$.

(1). Define

$$M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{6.115}$$

Then M is linear, monotone, $\dim \operatorname{ran} M = 2$ and

$$\nabla \arctan(y/x) = \frac{\begin{pmatrix} -y \\ x \end{pmatrix}}{x^2+y^2} = \frac{M\mathbf{x}}{\|M\mathbf{x}\|^2} \tag{6.116}$$

whenever $\mathbf{x} \neq 0$. However, M is not symmetric.

(2). Define

$$M = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{6.117}$$

Then M is linear and monotone and

$$\nabla \left(\frac{\ln(x^2+y^2)}{2} + \arctan(y/x) \right) = \frac{\begin{pmatrix} x-y \\ y+x \end{pmatrix}}{x^2+y^2} = \frac{M\mathbf{x}}{\|M\mathbf{x}\|^2} \tag{6.118}$$

whenever $\mathbf{x} \neq 0$. However, $\dim \operatorname{ran} M = 2$ and M is not symmetric.

Conjecture: Let $M : \mathbb{R}^n \to \mathbb{R}^n$ be linear, monotone and

$$(\forall\, x \in \mathbb{R}^n \text{ with } Mx \neq 0) \quad \nabla h(x) = \frac{Mx}{\|Mx\|^2} \tag{6.119}$$

where the function $h : \{x \in \mathbb{R}^n : Mx \neq 0\} \to \mathbb{R}$. If $\dim \operatorname{ran} M = 2$ and $\exp(h)$ is convex on convex subsets of $\{x \in \mathbb{R}^n : Mx \neq 0\}$, then M is symmetric.

Combining Theorem 6.6 and Proposition 6.9, we obtain the following characterization of linear subgradient projectors.

Theorem 6.11. *Assume that $T : X \to X$ is linear and $\dim \operatorname{ran}(\operatorname{Id} - T) \neq 2$. Then the following are equivalent.*

(i) T is a subgradient projector of a convex function.

(ii) $T = G_f$ where $f : X \to \mathbb{R}$ is given by

$$f(x) = K(x^\intercal P_L x)^{1/(2\lambda)} = K\big(d_{L^\perp}(x)\big)^{1/\lambda} \tag{6.120}$$

where $0 < \lambda \leq 1$, $K > 0$ and $L \subseteq X$ is a subspace. In this case, $G_f = (1-\lambda)\operatorname{Id} + \lambda P_{L^\perp}$.

Proof. (i)⇒(ii): Assume that $T = G_f$ for some convex function. Then T is a cutter by Fact 5.44. As T is linear, in view of Proposition 6.1, T is firmly nonexpansive, so $M := \operatorname{Id} - T$ is firmly nonexpansive, in particular, monotone. By Theorem 5.31,

$$\nabla h(x) = \frac{Mx}{\|Mx\|^2} \tag{6.121}$$

where $h(x) = \ln f(x)$, $f(x) > 0$. This is equivalent to

$$\nabla h(x) = \frac{Mx}{\|Mx\|^2} \tag{6.122}$$

when $Mx \neq 0$. Proposition 6.9 shows that M is symmetric, so is $T = \operatorname{Id} - M$. It suffices to apply Theorem 6.6 to obtain (ii).
(ii)⇒(i): Clear. □

6.3 A complete analysis of linear subgradient projectors on $\mathbb{R}^2$

We start with a simple result about essentially strictly differentiable functions (for definition, see Definition 5.33).

Lemma 6.12. *Let $O \subseteq X$ be a nonempty open set and $f : O \subseteq X \to \mathbb{R}$ be an essentially strictly differentiable function. If there exists a continuous selection $s : O \to X$ with $s(x) \in \partial f(x)$ for every $x \in O$, then f is strictly differentiable on O.*

Proof. By [17, Theorem 2.4, Corollary 4.2], f has a minimal Clarke subdifferential $\partial_c f$, and $\partial_c f$ can be recovered by every dense selection of $\partial_c f$. Since $s(x) \in \partial f(x) \subset \partial_c f(x)$, and s is continuous on O, we have $\partial_c f(x) = \partial f(x) = s(x)$ for every $x \in O$, which implies that f is strictly differentiable at x. Hence f is strictly differentiable on O. □

We consider linear operator in $\mathbb{R}^2$:

$$T = \begin{pmatrix} 1-a & -b \\ -c & 1-d \end{pmatrix} \tag{6.123}$$

where $a^2 + b^2 + c^2 + d^2 \neq 0$. Note that when $a = b = c = d = 0$, $T = G_f$ with $f \equiv 0$.

Theorem 6.13. *Let T be given by* (6.123)*. Then T is a subgradient projector of an essentially strictly differentiable function on $\mathbb{R}^2 \setminus \operatorname{Fix} T$ if and only if one of the following holds*

(i) *(1)* $a = b = c = 0, d \neq 0$*: $T = G_f$ where $f(x_1, x_2) = K|x_2|^{1/d}$ for some $K > 0$;*

(2) $b = c = d = 0, a \neq 0$*: $T = G_f$ where $f(x_1, x_2) = K|x_1|^{1/a}$ for some $K > 0$;*

(3) $b = c = 0, a = d \neq 0$*: $T = G_f$ where $f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{1}{2a}}$ for some $K > 0$.*

(ii) $a \neq 0, b = c, d = c^2/a$*: $T = G_f$ where $f(x_1, x_2) = K|ax_1 + cx_2|^{a/(a^2+c^2)}$ for some $K > 0$.*

(iii) $a = d \neq 0, b = -c \neq 0$*: $T = G_f$ where*

$$f(x_1, x_2) = \begin{cases} K(x_1^2 + x_2^2)^{\frac{a}{2(a^2+c^2)}} \exp\left(-\frac{c}{a^2+c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) & \text{if } x_2 \neq 0, \\ 0 & \text{if } (x_1, x_2) = (0, 0), \\ K|x_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2} \frac{\pi}{2}\right) & \text{if } x_1 \neq 0,\ x_2 = 0, \end{cases} \tag{6.124}$$

for some $K > 0$, and f is lower semicontinuous. In particular, when $c \neq 0$, f is nonconvex since f is not continuous when $x_1 \neq 0$ and $x_2 = 0$.

Proof. Assume that T is a subgradient projector. By Theorem 5.31 and Lemma 6.12, we can find a differentiable function $g : X \to \mathbb{R}$ such that

$$\text{for every } x \in X, \frac{x - Tx}{\|x - Tx\|^2} = \nabla g(x).$$

As

$$x - Tx = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix},$$

we have

$$\begin{aligned} \frac{\partial g}{\partial x_1} &= \frac{ax_1 + bx_2}{(ax_1 + bx_2)^2 + (cx_1 + dx_2)^2}, \\ \frac{\partial g}{\partial x_2} &= \frac{cx_1 + dx_2}{(ax_1 + bx_2)^2 + (cx_1 + dx_2)^2}. \end{aligned}$$

Since

$$\frac{\partial^2}{\partial x_1 \partial x_2} g(x_1, x_2) \tag{6.125}$$

$$= - \frac{(a^2c + c^3)x_1^2 + (cd^2 - b^2c + 2abd)x_2^2 + 2(c^2d + a^2d)x_1x_2}{((ax_1 + bx_2)^2 + (cx_1 + dx_2)^2)^2}, \tag{6.126}$$

$$\frac{\partial^2}{\partial x_2 \partial x_1} g(x_1, x_2) \tag{6.127}$$

$$= - \frac{(a^2b - bc^2 + 2acd)x_1^2 + (b^3 + bd^2)x_2^2 + 2(ab^2 + ad^2)x_1x_2}{((ax_1 + bx_2)^2 + (cx_1 + dx_2)^2)^2} \tag{6.128}$$

on a nonempty open set of $\mathbb{R}^2$. Then we have

$$\frac{\partial^2}{\partial x_1 \partial x_2} g(x_1, x_2) = \frac{\partial^2}{\partial x_2 \partial x_1} g(x_1, x_2).$$

This leads to

$$\begin{cases} a^2b - bc^2 + 2acd &= a^2c + c^3, & (1) \\ b^3 + bd^2 &= cd^2 - b^2c + 2abd, & (2) \\ ab^2 + ad^2 &= c^2d + a^2d. & (3) \end{cases}$$

Now $a * (2) - b * (3) \Rightarrow (ad - bc)(ab + cd) = 0$. It follows that

(A) If $ad = bc$. (1) implies $(b-c)(a^2+c^2) = 0$. Then the following two case could happen.

i. $b = c = 0$. Then (3) $\Rightarrow ad(a-d) = 0$. That means

$$a = b = c = 0, d \neq 0,$$

or

$$b = c = d = 0, a \neq 0,$$

or

$$a = d \neq 0, b = c = 0.$$

ii. $b = c \neq 0$ and $d = c^2/a$.

(B) If $ab + cd = 0$. (1) implies $(b+c)(a^2+c^2) = 0$. We only consider the case $b = -c \neq 0$. Then (2) and (3) implies $a = d$.

Thus we get the following three cases.

(i) $a = b = c = 0, d \neq 0$. Then we get

$$g(x_1, x_2) = \frac{\ln|x_2|}{d} + C_1, \text{ if } x_2 \neq 0.$$

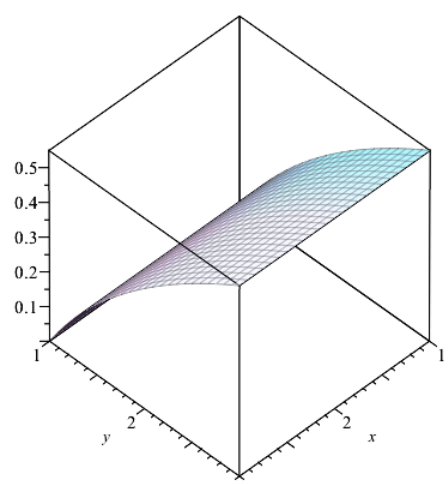


Figure 6.2: Graph of function g.

$$f(x_1, x_2) = K|x_2|^{1/d} \text{ for some } K > 0.$$

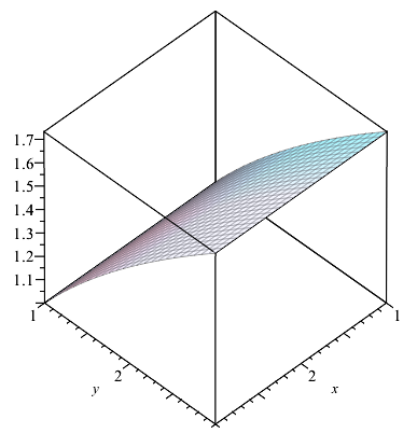


Figure 6.3: Graph of function f.

Or $b = c = d = 0, a \neq 0$. Then we get

$$g(x_1, x_2) = \frac{\ln|x_1|}{a} + C_1, \text{ if } x_1 \neq 0.$$

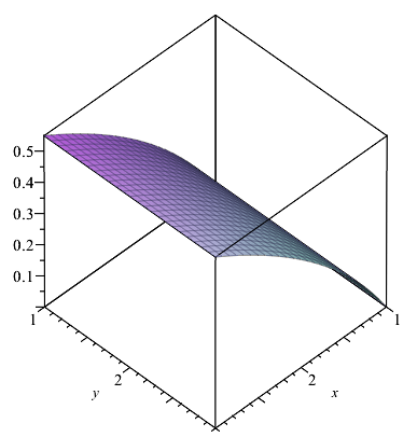


Figure 6.4: Graph of function g.

$$f(x_1, x_2) = K|x_1|^{1/a} \text{ for some } K > 0.$$

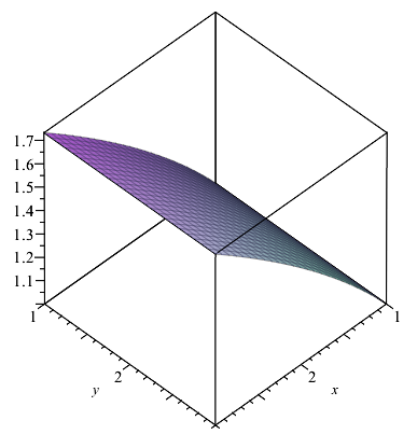


Figure 6.5: Graph of function f.

Or $a = d \neq 0, b = c = 0$. Then we get

$$g(x_1, x_2) = \frac{1}{2a} \ln (x_1^2 + x_2^2) + C_1, \text{ if } x_1 \neq 0 \text{ and } x_2 \neq 0.$$

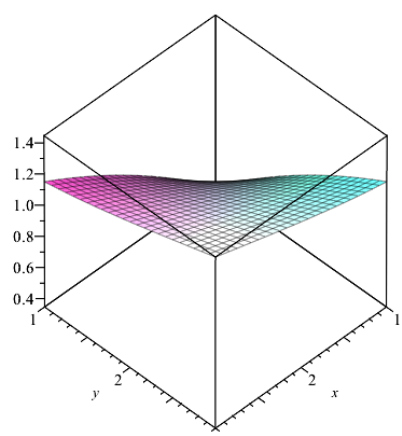


Figure 6.6: Graph of function g

$$f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{1}{2a}} \text{ for some } K > 0.$$

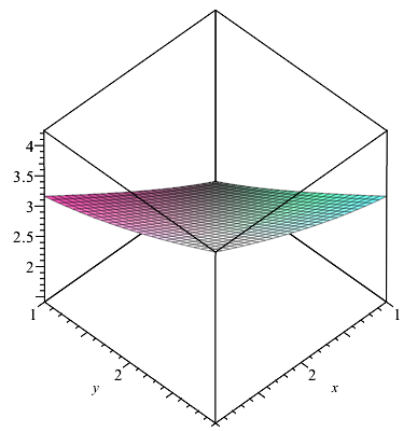


Figure 6.7: Graph of function f

(ii) $a \neq 0, b = c \neq 0, d = c^2/a$. Then we get

$$g(x_1, x_2) = \frac{a}{a^2 + c^2} \ln |ax_1 + cx_2| + C_2, \text{ if } ax_1 + cx_2 \neq 0.$$

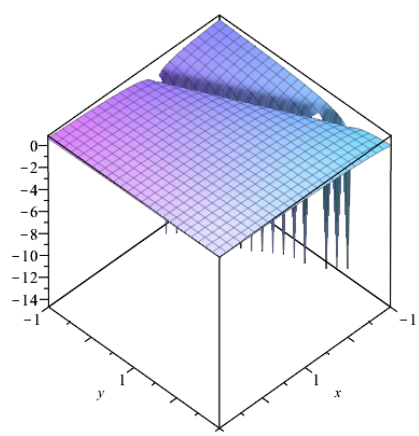


Figure 6.8: Graph of function g

$$f(x_1, x_2) = K|ax_1 + cx_2|^{a/(a^2+c^2)} \text{ for some } K > 0.$$

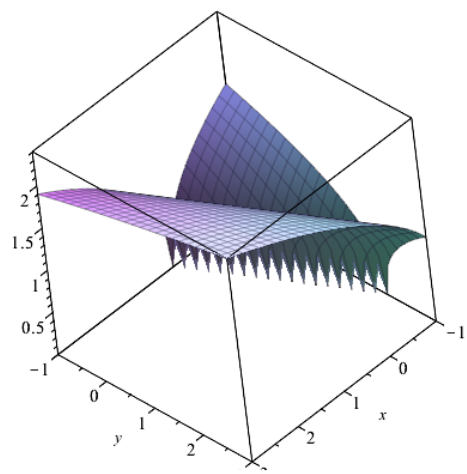


Figure 6.9: Graph of function f

(iii) $a = d \neq 0, b = -c \neq 0$. Then we get

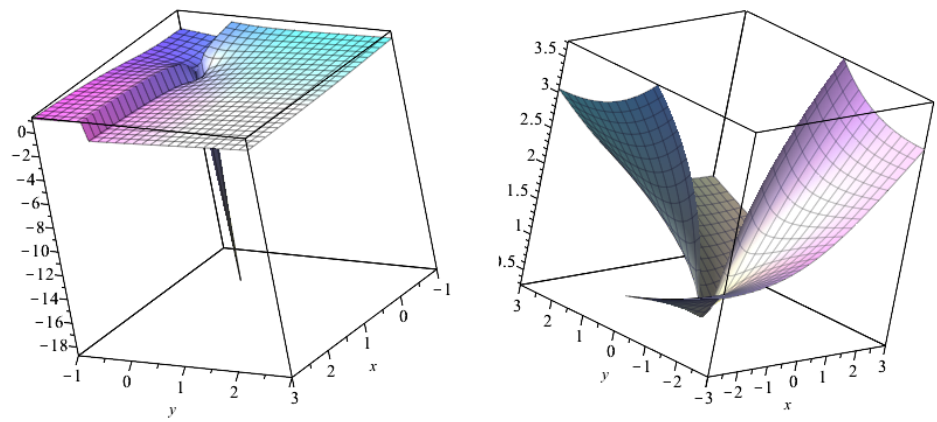
$$g(x_1, x_2) = \frac{a}{2(a^2+c^2)} \ln(x_1^2+x_2^2) - \frac{c}{a^2+c^2} \arctan\left(\frac{x_1}{x_2}\right) + C_3, \quad (6.129)$$

if $x_2 \neq 0$. Since $g = \ln f$, we obtain $f = \exp(g)$ by using (i) and (ii).

For (iii), we obtain

$$f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{a}{2(a^2+c^2)}} \exp\left(-\frac{c}{a^2+c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) \quad \text{if } x_2 \neq 0 \quad (6.130)$$

for some $K > 0$.



(a) Graph of function g.

(b) Graph of function f.

Figure 6.10: Plots of g and f when $a = 1, c = 1$ and $K = 1$ in Theorem 6.13(iii).

However, when $c \neq 0$, f is not continuous at $(\bar{x}_1, 0)$ with $\bar{x}_1 \neq 0$ since

$$\frac{\pi}{2} = \lim_{x_1 \to \bar{x}_1, x_2 \downarrow 0} \arctan \frac{x_1}{x_2} \neq \lim_{x_1 \to \bar{x}_1, x_2 \uparrow 0} \arctan \frac{x_1}{x_2} = -\frac{\pi}{2}. \tag{6.131}$$

The function given by (6.124) is lower semicontinuous but not continuous at every $(\bar{x}_1, 0)$. Moreover, f is not convex on $\mathbb{R}^2$ since a finite-valued convex function is continuous.

It is natural to ask for what selection $s \in \partial f$, we have $G_f = T$ on $\mathbb{R}^2$. On $\mathbb{R}^2 \setminus \{(x_1, x_2) : x_2 \neq 0\}$, one clearly chooses $s = \nabla f$. It remains to determine the subgradient of f at $(\bar{x}_1, 0)$. Indeed, when $x_2 \neq 0$, $f(x_1, x_2) = \exp(g(x_1, x_2))$, so that $\nabla f(x_1, x_2) = f(x_1, x_2) \nabla g(x_1, x_2)$, i.e.

$$\nabla f(x_1, x_2) \tag{6.132}$$

$$= K(x_1^2 + x_2^2)^{\frac{a}{2(a^2+c^2)}} \exp\left(-\frac{c}{a^2+c^2} \arctan\left(\frac{x_1}{x_2}\right)\right) \frac{1}{a^2+c^2} \frac{(ax_1 - cx_2, ax_2 + cx_1)}{x_1^2 + x_2^2}. \tag{6.133}$$

When $(x_1, x_2) \to (\bar{x}_1, 0)$, $cx_1/x_2 > 0$, we have

$$f(x_1, x_2) \to K|\bar{x}_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2} \frac{\pi}{2}\right) = f(\bar{x}_1, 0) \tag{6.134}$$

and

$$\nabla f(x_1, x_2) \to K|\bar{x}_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2}\frac{\pi}{2}\right)\frac{1}{a^2+c^2}\left(\frac{a}{\bar{x}_1}, \frac{c}{\bar{x}_1}\right). \quad (6.135)$$

Therefore, by the definition of limiting subdifferential (see Definition 5.1),

$$K|\bar{x}_1|^{\frac{a}{(a^2+c^2)}} \exp\left(-\frac{|c|}{a^2+c^2}\frac{\pi}{2}\right)\frac{1}{a^2+c^2}\left(\frac{a}{\bar{x}_1}, \frac{c}{\bar{x}_1}\right) \in \partial f(\bar{x}_1, 0). \quad (6.136)$$

Hence we can choose $s(\bar{x}_1, 0)$ to be the limiting subgradient given by (6.136). □

Corollary 6.14. *(1). Define*

$$T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (6.137)$$

The skew linear mapping T is not firmly nonexpansive, so not a cutter. However, T is a subgradient projector of a nonconvex, discontinuous but lower semicontinuous function f given by

$$f(x,y) = \begin{cases} (x^2+y^2)^{1/4} * \exp(-(1/2) * \arctan(x/y)) & \text{if } y \neq 0, \\ 0 & \text{if } (x,y) = (0,0), \\ |x|^{1/2}\exp(-\pi/4) & \text{if } x \neq 0, y = 0. \end{cases} \quad (6.138)$$

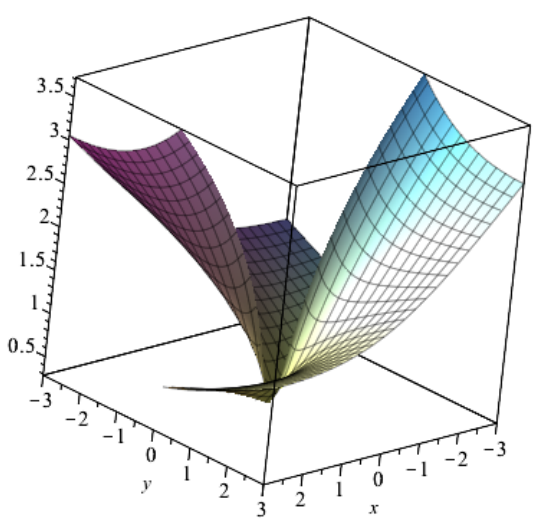


Figure 6.11: Graph of f defined in (6.138).

(2). The linear mapping

$$T := \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \tag{6.139}$$

is firmly nonexpansive and a cutter. However, T is a subgradient projector of a nonconvex, discontinuous but lower semicontinuous function f given by

$$f(x,y) = \begin{cases} (x^2+y^2)^{1/2} * \exp(-\arctan(x/y)) & \text{if } y \neq 0, \\ 0 & \text{if } (x,y) = (0,0), \\ |x| \exp(-\pi/2) & \text{if } x \neq 0, y = 0. \end{cases} \tag{6.140}$$

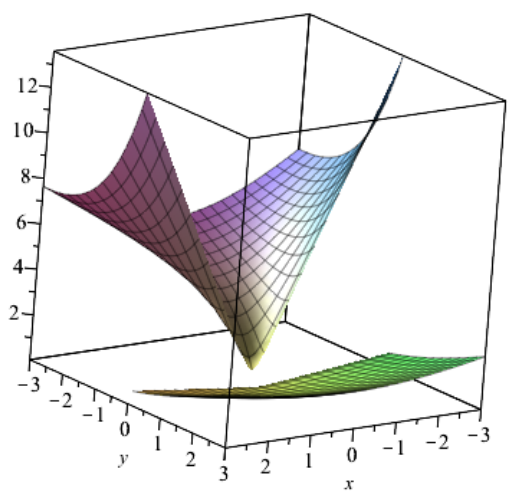


Figure 6.12: Graph of f defined in (6.140).

By Theorem 5.35, there exists no continuous convex function f such that $G_f = T$ in either case. Corollary 6.14 says that $T = G_f$ being linear and firmly nonexpansive does not implies that f is convex. A key point below is that if $T = G_f$ is linear and f is convex, then T has to be firmly nonexpansive and symmetric.

Corollary 6.15. *Let T be given by* (6.123)*. Then T is a subgradient projector of a convex function if and only if one of the following holds*

(i) $a = b = c = 0, d \neq 0$, $0 < d \leq 1$*:* $T = G_f$ *where* $f(x_1, x_2) = K|x_2|^{1/d}$ *for some* $K > 0$*; or* $b = c = d = 0, a \neq 0$, $0 < a \leq 1$*:* $T = G_f$ *where* $f(x_1, x_2) = K|x_1|^{1/a}$ *for some* $K > 0$*.*

(ii) $a \neq 0, b = c, d = c^2/a$, $a \geq a^2 + c^2$*:* $T = G_f$ *where* $f(x_1, x_2) = K|ax_1 + cx_2|^{a/(a^2+c^2)}$ *for some* $K > 0$*.*

(iii) $a = d, b = c = 0$, $0 < a \leq 1$*:* $T = G_f$ *where* $f(x_1, x_2) = K(x_1^2 + x_2^2)^{\frac{1}{2a}}$ *for some* $K > 0$*.*

In this chapter, we not only characterized symmetric linear subgradient projectors (see Theorem 6.6), but also provided a criterion when a linear monotone operator is a subgradient projector (see Theorem 6.11).

Chapter 7

Conclusion

7.1 Key Results

This thesis has provided a complete study of charaterizations of subgradient projectors, as well as a new method for finding a fixed point of a cutter. Let us list the key results and some crucial examples.

Theorem 3.14 gives a finite convergence result provided that $\operatorname{int}(C \cap \operatorname{Fix} T) \neq \varnothing$ and $\sum \eta_n r_n = +\infty$.

Theorem 3.15 shows the finite convergence is also true if we have a less restrictive assumption on $(\operatorname{Fix} T, C)$ but a more restrictive one on parameters (r_n, η_n).

Example 3.20 and Example 3.21 show that both the divergent-series condition and the nonempty-interior condition are necessary and cannot be omitted for Theorem 3.14 and Theorem 3.15.

Theorem 4.17 lists the corresponding relationship between the continuity of subgradient projectors and Fréchet differentiability of the function f.

Example 4.22 provides an example that subgradient projectors lack strong-to-weak continuity when the function f is only Gâteaux differentiable.

Example 4.24 links the subgradient projector to the accelerated mapping corresponding to a linear operator when the function f is a power of a quadratic form.

Proposition 4.43 exhibits surprisingly complicated properties of the subgradient projector of the function $f(x_1, x_2) = |x_1|^p + |x_2|^p$ with different choices of the parameter $p \geq 1$.

Theorem 4.44 shows that, on the real line, the Yamagishi-Yamada operator is actually a subgradient projector of another convex function.

Theorem 5.32 gives a criterion when a mapping is a subgradient projector.

Theorem 5.39 characterizes the class of functions f for which its subgradient projector is a cutter.

Theorem 6.3 lists several interesting results of subgradient projectors of functions defined by a symmetric, positive semidefinite matrix. We obtained

that its subgradient projector is linear if and only if the involved matrix can be represented as a positive multiple of the projector onto a linear subspace.

Theorem 6.6 completely characterizes symmetric linear subgradient projectors.

It is known that the set of cutters is convex. Even though subgradient projectors are cutters, the set of subgradient projectors is not convex: see Theorem 6.7 and Theorem 6.8.

Proposition 6.9 provides a criterion when a linear monotone operator is a subgradient projector.

7.2 Future Work and Open Questions

7.2.1 Open Questions

(1) Consider Theorem 6.6. Suppose the subgradient projector of a convex function is linear. Must it be symmetric?

(2) Consider Theorem 6.11. Suppose that the rank of the displacement matrix is equal to 2. What conditions are sufficient to guarantee that the original matrix is a subgradient projector of a convex function?

7.2.2 S-cutter

In this section, we consider a new operator: S-cutter.

Definition 7.1. Let S be a nonempty subset of $\operatorname{Fix} T$. We call $T : \mathcal{H} \to \mathcal{H}$ a *S-cutter* if for $\forall x \in \mathcal{H},\ \forall y \in S$,

$$\langle y - Tx, x - Tx \rangle \leq 0. \tag{7.1}$$

Clearly, if $S = \operatorname{Fix} T$, then the definition of S-cutters matches the definition of cutters.

In [20], there is a systematic study of the properties of cutters. A natural question is: which properties of cutters also hold for S-cutters? For example, [20, Theorem 2.1.39] shows the relationship between cutters and strongly quasi-nonexpansive operators. It would be interesting to find connections between S-cutters and strongly quasi-nonexpansive operators, and to consider the following list of questions:

(1) Is the set of S-cutters closed under convex combinations and compositions?

(2) What desirable properties do the relaxations of S-cutters have?

(3) In [20, Theorem 3.4.3], we see that a strongly quasi-nonexpansive operator with a fixed point is asymptotically regular. If there is some relation between S-cutters and strongly quasi-nonexpansive operators, does a similar property hold for S-cutters? That is: are S-cutters or their relaxations asymptotically regular? If so, what is the consequence of [20, Theorem 3.5.2] when applied to S-cutters?

(4) Considering [20, Corollary 3.7.1, Corollary 3.7.3], what are the Opial-type theorems for S-cutters?

(5) In [13], Theorems 3.1 and 3.2 hold if the interior of $\operatorname{Fix} T$ is nonempty. If $S \subseteq \operatorname{Fix} T$, we may ask: do Theorems 3.1 and 3.2 hold for S-cutters under some sufficient conditions?

7.2.3 Analysis of Parameters

From the numerical experiments at the end of Chapter 3, we can see that the selection of parameters is crucial and plays an important role in the performance of algorithms. A question for future work is how to predefine parameters so that our algorithm will perform better.

7.2.4 Generalized Monotonicities

Chapters 4–6 provide some properties and characterizations of subgradient projectors. During the last twenty-five years, generalized monotonicities have been studied a lot. They play an important role in variational inequality problems. Let $C \subseteq \mathcal{H}$ be closed convex and $\mathcal{F} : \mathcal{H} \to \mathcal{H}$. The *variational inequality problem* is to find $x \in C$ such that

$$\langle \mathcal{F}x, y - x \rangle \geq 0 \tag{7.2}$$

holds for all $y \in C$. It would be interesting to investigate the generalized monotonicities of subgradient projectors in the context of solving the variational inequality problem.

7.2.5 Recognition Convexity of f from G_f

In Section 6.3, we provide a complete analysis of linear subgradient projectors on $\mathbb{R}^2$. Outside $\mathbb{R}^2$, is it possible to detect convexity of f from G_f?

Bibliography

[1] H. H. Bauschke, *Projection Algorithms and Monotone Operators*, PhD thesis, Simon Fraser University, Burnaby, BC, Canada, August 1996. → pages

[2] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Review, 38 (1996), pp. 367–426. → pages

[3] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms*, SIAM Journal on Control Optimization, 42 (2003), pp. 596–636. → pages

[4] H. H. Bauschke, J. Chen, and X. Wang, *A projection method for approximating fixed points of quasi nonexpansive mappings without the usual demiclosedness condition*, Journal of Nonlinear and Convex Analysis, 15 (2014), pp. 129–135. → pages

[5] H. H. Bauschke, J. Chen, and X. Wang, *A Bregman projection method for approximating fixed points of quasi-Bregman nonexpansive mappings*, Applicable Analysis, 94 (2015), pp. 75–84. → pages

[6] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert space*, Mathematics of Operations Research, 26 (2001), pp. 248–264. → pages

[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics / Ouvrages de Mathématiques de la SMC, Springer, New York, 2011. With a foreword by Hédy Attouch. → pages

[8] H. H. Bauschke, F. Deutsch, H. Hundal, and S. H. Park, *Accelerating the convergence of the method of alternating projections*, Transactions of the American Mathematical Society, 355 (2003), pp. 3433–3461. → pages

[9] H. H. Bauschke, S. M. Moffat, and X. Wang, *Self-dual smooth approximations of convex functions via the proximal average, Fixed-Point Algorithms for Inverse Problems in Science and Engineering (Banff 2009)*, Springer, (2011), pp. 23–32. → pages

[10] H. H. Bauschke, C. Wang, X. Wang, and J. Xu, *Characterizations of subgradient projectors*, in progress. → pages

[11] H. H. Bauschke, C. Wang, X. Wang, and J. Xu, *The linearity of subgradient projectors*, in progress. → pages

[12] H. H. Bauschke, C. Wang, X. Wang, and J. Xu, *On subgradient projectors*, SIAM Journal on Optimization, 25 (2015), pp. 1064–1082. → pages

[13] H. H. Bauschke, C. Wang, X. Wang, and J. Xu, *On the finite convergence of a projected cutter method*, Journal of Optimization Theory and Applications, 165 (2015), pp. 901–916. → pages

[14] H. H. Bauschke, X. Wang, and L. Yao, *Monotone linear relations: maximality and Fitzpatrick functions*, Journal of Convex Analysis, 16 (2009), pp. 673–686. → pages

[15] J. M. Borwein and M. Fabián, *On convex functions having points of Gateaux differentiability which are not points of Fréchet differentiability*, Canadian Journal of Mathematics, 45 (1993), pp. 1121–1134. → pages

[16] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag New York, 2 ed., 2006. → pages

[17] J. M. Borwein and W. B. Moors, *Essentially smooth Lipschitz functions*, Journal of Functional Analysis, 149 (1997), pp. 305–351. → pages

[18] J. M. Borwein and J. D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010. → pages

[19] T. D. Capricelli, *Projections Convexes Généralisées et Applications en Imagerie Ménicale*, PhD thesis, LÚniversité Pierre et Marie Curie Pairs VI, 2008. → pages

[20] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Springer, 2012. → pages

[21] Y. Censor, W. Chen, and H. Pajoohesh, *Finite convergence of a subgradient projection method with expanding controls*, Applied Mathematics & Optimization, 64 (2011), pp. 273–285. → pages

[22] Y. Censor and A. Lent, *Cyclic subgradient projections*, Mathematical Programming, 24 (1982), pp. 233–235. → pages

[23] Y. Censor and A. Segal, *Sparse string-averaging and split common fixed points, in "Nonlinear Analysis and Optimization I", A. Leizarowitz, B.S. Mordukhovich, I. Shafrir, and A.J. Zaslavski (editors)*, Contemporary Mathematics, 513 (2010), pp. 125–142. → pages

[24] Y. Censor and S. A. Zenios, *Parallel Optimization*, Oxford University Press, 1997. → pages

[25] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990. → pages

[26] P. L. Combettes, *The foundations of set theoretic estimation*, Proceedings of the IEEE, 81 (1993), pp. 182–208. → pages

[27] P. L. Combettes, *Restauration ensembliste d'images par itérations parallèles extrapolées de sous-gradients*, Proceedings of the 15th GRETSI Symposium, (1995), pp. 447–450. → pages

[28] P. L. Combettes, *The convex feasibility problem in image recovery, in Advances in Imaging and Electron Physics, P. Hawkes (editor)*, Academic Press, 95 (1996). → pages

[29] P. L. Combettes, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Transactions on Image Processing, 6 (1997), pp. 493–506. → pages

[30] P. L. Combettes and J. Luo, *An adaptive level set method for nondifferentiable constrained image recovery*, IEEE Transactions on Image Processing, 11 (2002), pp. 1295–1304. → pages

[31] G. Crombez, *Finding common fixed points of a class of paracontractions*, Acta Mathematica Hungarica, 103 (2004), pp. 233–241. → pages

[32] K. R. Davidson and A. P. Donsig, *Real Analysis and Applications*, Springer, 2010. → pages

[33] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Monographs in Mathematics, 2009. → pages

[34] M. Fukushima, *A finite convergent algorithm for convex inequalities*, IEEE Transactions on Automatic Control, 27 (1982), pp. 1126–1127. → pages

[35] M. C. Gemignani, *Elementary Topology*, Dover Publications Inc, second ed., 1972. → pages

[36] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1989. → pages

[37] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. → pages

[38] J. Milnor, *Dynamics in One Complex Variable*, Princeton University Press, third ed., 2006. → pages

[39] B. S. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, Journal of Mathematical Analysis and Applications, (1994), pp. 250–288. → pages

[40] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I*, Springer-Verlag, 2006. → pages

[41] B. S. Mordukhovich and N. M. Nam, *An Easy Path to Convex Analysis and Applications*, Morgan and Claypool Publishers, 2014. → pages

[42] B. S. Mordukhovich and Y. Shao, *Nonsmooth sequential analysis in asplund spaces*, Transactions of the American Mathematical Society, 348 (1996), pp. 1235–1280. → pages

[43] N. Ogura and I. Yamada, *A deep outer approximating half space of the level set of certain quadratic functions*, Journal of Nonlinear and Convex Analysis, 6 (2005), pp. 187–201. → pages

[44] B. Pauwels, *Subgradient projection operators*, Master's thesis, Universite Pierre et Marie Curie, France, September 2012. → pages

[45] J. P. Penot, *Calculus Without Derivatives*, Springer, 2013. → pages

[46] R. R. Phelps, *Convex functions, Monotone Operators and Differentiability*, no. 1364 in Lecture Notes in Mathematics, Springer-Verlag, Berlin, second ed., 1993. → pages

[47] A. R. D. Pierro and A. N. Iusem, *A finitely convergent "row-action" method for the convex feasibility problem*, Applied Mathematics and Optimization, 17 (1988), pp. 225–235. → pages

[48] B. T. Polyak, *Minimization of unsmooth functionals*, U.S.S.R. Computational Mathematics and Mathematical Physics, 9 (1969), pp. 14–29. → pages

[49] B. T. Polyak, *Introduction to Optimization*, Optimization Software, 1987. → pages

[50] B. T. Polyak, *Random algorithms for solving convex inequalities. In D. Butnariu, Y. Censor, S. Reich(eds.): Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, Elsevier Science Publishers, (2001), pp. 409–422. → pages

[51] E. Raik, *A class of iterative methods with Fejér-monotone sequences*, Eesti NSV Teaduste Akadeemia Toimetised. Füüsika-Matemaatika, 18 (1969), pp. 22–26. → pages

[52] L. F. Richardson, *Measure and Integration : a Concise Introduction to Real Analysis*, Wiley, 2009. → pages

[53] R. T. Rockafellar and J. B. Wets, *Variational Analysis*, Springer, third ed., 2009. → pages

[54] K. A. Ross, *Elementary Analysis*, Springer, second ed., 2013. → pages

[55] K. Slavakis and I. Yamada, *The adaptive projected subgradient method constrained by families of quasi-nonexpansive mappings and its application to online learning*, SIAM Journal on Optimization, 23 (2013), pp. 126–152. → pages

[56] S. Theodoridis, K. Slavakis, and I. Yamada, *Adaptive learning in a world of projections*, IEEE Signal Processing Magazine, 97 (2011), pp. 97–123. → pages

[57] S. X. Wang, *Fine and Pathological Properties of Subdifferentials*, PhD thesis, Simon Fraser University, August 1999. → pages

[58] I. Yamada and N. Ogura, *Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions*, Numerical Functional Analysis and Optimization, 25 (2004), pp. 593–617. → pages

[59] I. Yamada and N. Ogura, *Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings*, Numerical Functional Analysis and Optimization, 25 (2004), pp. 619–655. → pages

[60] I. Yamada, K. Slavakis, and K. Yamada, *An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques*, IEEE Transactions on Signal Processing, 50 (2002), pp. 1091–1101. → pages

[61] M. Yamagishi and I. Yamada, *A deep monotone approximation operator based on the best quadratic lower bound of convex functions*, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E91 (2008), pp. 1858–1866. → pages

Appendix

Appendix A

Maple Code

A.1 Code to verify Remark 3.24 when $f(x) = x^2 - 1$

$$f := x \mapsto x^2 - 1$$

$$df := \mathrm{D}(f)$$

$$G_f := x \mapsto x - 1/2\,\frac{x^2 - 1}{x}$$

$$r_1 := n \mapsto \frac{1}{\sqrt{n+1}}$$

$$r_2 := n \mapsto (n+1)^{-1}$$

$$\eta_1 := 1$$

$$\eta_2 := 2$$

$$\epsilon_1 := n \mapsto \frac{1}{\sqrt{n+1}}$$

$$\epsilon_2 := n \mapsto (n+1)^{-1}$$

$$U[1] := (x, r) \mapsto \mathbf{piecewise}(f(x) \leq 0, x, f(x) > 0, x + \frac{\eta_1 * (r + |G[f](x) - x)| * (G[f](x) - x)}{|G[f](x) - x|})$$

$$U[2] := (x, r) \mapsto \mathbf{piecewise}(f(x) \leq 0, x, f(x) > 0, x + \frac{\eta_2 * (r + |G[f](x) - x)| * (G[f](x) - x)}{|G[f](x) - x|})$$

$$T[1] := y \mapsto \mathbf{piecewise}(f(y) \leq 0, y, f(y) > 0, y - \frac{f(y) + \epsilon_1(n)}{df(y)})$$

$$T[2] := y \mapsto \mathbf{piecewise}(f(y) \leq 0, y, f(y) > 0, y - \frac{f(y) + \epsilon_2(n)}{df(y)})$$

$with(RandomTools)$

$with(Statistics)$

$with(LinearAlgebra)$

$M := 100$

$h := Vector(M)$

$g := Vector(M)$

$k := Vector(M)$

$l := Vector(M)$

$u := Vector(M)$

$p := Vector(M)$

$SetState(state = 12345)$

for v from 1 by 1 to M do

x[0] := **Generate(float(range = 1 .. 1000000))**;

y[0]:= x[0]; z[0]:= x[0]; w[0] := x[0]; a[0] := x[0]; b[0] := x[0];

n := 0; m := 0; i := 0; j := 0; s := 0; q := 0; ;

while $f(x[n]) > 10^{-6}$ **do** $x[n+1] := evalf(U[1](x[n], r[1](n)))$**;** $n := n + 1$**; end do**

while $f(y[m]) > 10^{-6}$ **do** $y[m+1] := evalf(U[1](y[m], r[2](m)))$**;** $m := m + 1$**; end do**

while $f(z[i]) > 10^{-6}$ **do** $z[i+1] := evalf(U[2](z[i], r[1](i)))$**;** $i := i + 1$**; end do**

while $f(w[j]) > 10^{-6}$ **do** $w[j+1] := evalf(U[2](w[j], r[2](j)))$**;** $j := j + 1$**; end do**

while $f(a[s]) > 10^{-6}$ **do** $a[s+1] := evalf(T[1](a[s]))$**;** $s := s + 1$**; end do**

while $f(b[q]) > 10^{-6}$ **do** $b[q+1] := evalf(T[2](b[q]))$**;** $q := q+1$**; end do**

h[v] := n; g[v] := m; k[v] := i; l[v] := j; u[v] := s; p[v] := q

$mean_{U_1(x,r_1)} := Mean\,(h)$

$$10.8300000000000$$

$mn_1 := evalf\left(mean_{U_1(x,r_1)}, 4\right)$

$$10.83$$

$mean_{U_1(x,r_2)} := Mean\,(g)$

$$11.4900000000000$$

$mn_2 := evalf\left(mean_{U_1(x,r_2)}, 4\right)$

$$11.49$$

$mean_{U_2(x,r_1)} := Mean\,(k)$

$$2.0$$

$mn_3 := evalf\left(mean_{U_2(x,r_1)}, 4\right)$

$$2.0$$

$mean_{U_2(x,r_2)} := Mean\,(l)$

$$2.0$$

$mn_4 := evalf\left(mean_{U_2(x,r_2)}, 4\right)$

$$2.0$$

$mean_{T_1} := Mean\,(u)$

$$11.8100000000000$$

$mn_5 := evalf\,(mean_{T_1}, 4)$

$$11.81$$

$mean_{T_2} := Mean\,(p)$

$$12.1900000000000$$

$mn_6 := evalf\,(mean_{T_2}, 4)$

$$12.19$$

$median_{U_1(x,r_1)} := Median\,(h)$

$$12.0$$

$d_1 := evalf\left(median_{U_1(x,r_1)}, 4\right)$

$$12.0$$

$median_{U_1(x,r_2)} := Median\,(g)$

$$13.0$$

$d_2 := evalf\left(median_{U_1(x,r_2)}, 4\right)$

$$13.0$$

$$median_{U_2(x,r_1)} := Median\,(k)$$

$$2.0$$

$$d_3 := evalf\left(median_{U_2(x,r_1)}, 4\right)$$

$$2.0$$

$$median_{U_2(x,r_2)} := Median\,(l)$$

$$2.0$$

$$d_4 := evalf\left(median_{U_2(x,r_2)}, 4\right)$$

$$2.0$$

$$median_{T_1} := Median\,(u)$$

$$13.0$$

$$d_5 := evalf\,(median_{T_1}, 4)$$

$$13.0$$

$$median_{T_2} := Median\,(p)$$

$$13.0$$

$$d_6 := evalf\,(median_{T_2}, 4)$$

$$13.0$$

$N := Matrix(6, 2, [mn[2], d[2], mn[4], d[4], mn[1], d[1], mn[3], d[3], mn[5], d[5], mn[6], d[6]])$

$$N := \begin{bmatrix} 11.49 & 13.0 \\ 2.0 & 2.0 \\ 10.83 & 12.0 \\ 2.0 & 2.0 \\ 11.81 & 13.0 \\ 12.19 & 13.0 \end{bmatrix}$$

A.2 Code to verify Remark 3.24 when $f(x) = 100x^2 - 1$

$f := x \mapsto 100 * x^2 - 1$

$df := \mathrm{D}(f)$

$G_f := x \mapsto x - \dfrac{100 * x^2 - 1}{200 * x}$

$r_1 := n \mapsto \dfrac{1}{\sqrt{n+1}}$

$r_2 := n \mapsto (n+1)^{-1}$

$\eta_1 := 1$

$\eta_2 := 2$

$\epsilon_1 := n \mapsto \dfrac{1}{\sqrt{n+1}}$

$\epsilon_2 := n \mapsto (n+1)^{-1}$

$U[1] := (x, r) \mapsto \mathbf{piecewise}(f(x) \leq 0, x, f(x) > 0, x + \frac{\eta_1 * (r + |G[f](x) - x)| * (G[f](x) - x)}{|G[f](x) - x|})$

$U[2] := (x, r) \mapsto \mathbf{piecewise}(f(x) \leq 0, x, f(x) > 0, x + \frac{\eta_2 * (r + |G[f](x) - x)| * (G[f](x) - x)}{|G[f](x) - x|})$

$T[1] := y \mapsto \mathbf{piecewise}(f(y) \leq 0, y, f(y) > 0, y - \dfrac{f(y) + \epsilon_1(n)}{df(y)})$

$T[2] := y \mapsto \mathbf{piecewise}(f(y) \leq 0, y, f(y) > 0, y - \dfrac{f(y) + \epsilon_2(n)}{df(y)})$

$with(RandomTools)$

$with(Statistics)$

$with(LinearAlgebra)$

$M := 100$

$h := Vector(M)$

$g := Vector(M)$

$k := Vector(M)$

$l := Vector(M)$

$u := Vector(M)$

$p := Vector(M)$

$SetState(state = 12345)$

for v from 1 by 1 to M do

x[0] := **Generate(float(range = 1 .. 1000000))**;

y[0]:= x[0]; z[0]:= x[0]; w[0] := x[0]; a[0] := x[0]; b[0] := x[0];

n := 0; m := 0; i := 0; j := 0; s := 0; q := 0; ;

while $f(x[n]) > 10^{-6}$ **do** $x[n+1] := evalf(U[1](x[n], r[1](n)))$**;** $n := n+1$**; end do**

while $f(y[m]) > 10^{-6}$ **do** $y[m+1] := evalf(U[1](y[m], r[2](m)))$**;** $m := m+1$**; end do**

while $f(z[i]) > 10^{-6}$ **do** $z[i+1] := evalf(U[2](z[i], r[1](i)))$**;** $i := i+1$**; end do**

while $f(w[j]) > 10^{-6}$ **do** $w[j+1] := evalf(U[2](w[j], r[2](j)))$**;** $j := j+1$**; end do**

while $f(a[s]) > 10^{-6}$ **do** $a[s+1] := evalf(T[1](a[s]))$**;** $s := s+1$**; end do**

while $f(b[q]) > 10^{-6}$ **do** $b[q+1] := evalf(T[2](b[q]))$**;** $q := q+1$**; end do**

h[v] := n; g[v] := m; k[v] := i; l[v] := j; u[v] := s; p[v] := q

$mean_{U_1(x,r_1)} := Mean\,(h)$

$$17.5200000000000$$

$mn_1 := evalf\left(mean_{U_1(x,r_1)}, 4\right)$

$$17.52$$

$$mean_{U_1(x,r_2)} := Mean\,(g)$$

$$13.2900000000000$$

$$mn_2 := evalf\left(mean_{U_1(x,r_2)}, 4\right)$$

$$13.29$$

$$mean_{U_2(x,r_1)} := Mean\,(k)$$

$$105.0$$

$$mn_3 := evalf\left(mean_{U_2(x,r_1)}, 4\right)$$

$$105.0$$

$$mean_{U_2(x,r_2)} := Mean\,(l)$$

$$12.0$$

$$mn_4 := evalf\left(mean_{U_2(x,r_2)}, 4\right)$$

$$12.0$$

$$mean_{T_1} := Mean\,(u)$$

$$15.2700000000000$$

$$mn_5 := evalf\,(mean_{T_1}, 4)$$

$$15.27$$

$$mean_{T_2} := Mean\,(p)$$

$$15.7600000000000$$

$$mn_6 := evalf\,(mean_{T_2}, 4)$$

$$15.76$$

$$median_{U_1(x,r_1)} := Median\,(h)$$

$$19.0$$

$$d_1 := evalf\left(median_{U_1(x,r_1)}, 4\right)$$

$$19.0$$

$$median_{U_1(x,r_2)} := Median\,(g)$$

$$14.0$$

$$d_2 := evalf\left(median_{U_1(x,r_2)}, 4\right)$$

$$14.0$$

$$median_{U_2(x,r_1)} := Median\,(k)$$

$$105.0$$

$$d_3 := evalf\left(median_{U_2(x,r_1)}, 4\right)$$

$$105.0$$

$$median_{U_2(x,r_2)} := Median\,(l)$$

$$12.0$$

$$d_4 := evalf\left(median_{U_2(x,r_2)}, 4\right)$$

$$12.0$$

$$median_{T_1} := Median\,(u)$$

$$16.0$$

$$d_5 := evalf\,(median_{T_1}, 4)$$

$$16.0$$

$$median_{T_2} := Median\,(p)$$

$$17.0$$

$$d_6 := evalf\,(median_{T_2}, 4)$$

$$17.0$$

$N := Matrix(6, 2, [mn[2], d[2], mn[4], d[4], mn[1], d[1], mn[3], d[3], mn[5], d[5], mn[6], d[6]])$

$$N := \begin{bmatrix} 13.29 & 14.0 \\ 12.0 & 12.0 \\ 17.52 & 19.0 \\ 105.0 & 105.0 \\ 15.27 & 16.0 \\ 15.76 & 17.0 \end{bmatrix}$$