CQ algorithms: theory, computations and nonconvex extensions

by

Yipin Guo

B.Sc., Tongji University, 2011

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

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Interdisciplinary Studies)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

August 2014

Abstract

The split feasibility problem (SFP) is important due to its occurrence in signal processing and image reconstruction, with particular progress in intensity-modulated radiation therapy. Mathematically, it can be formulated as finding a point x^* such that $x^* \in C$ and $Ax^* \in Q$, where A is a bounded linear operator, C and Q are subsets of two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. One particular algorithm for solving this problem is the CQ algorithm.

In this thesis, previous work on CQ algorithm is presented and a new proof of convergence of the relaxed CQ algorithm is given. The CQ algorithm is shown to be a special case of the subgradient projection algorithm. The SFP is extended into two nonconvex cases. The first one is on S-subdifferentiable functions, and the other one is on prox-regular functions. The subgradient projection algorithm and CQ algorithm are proved to converge to a solution of the first and second case respectively.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my supervisor Dr. Shawn Wang for his patient instruction in mathematics and his support for my study in UBCO. I would also like to thank Dr. Heinz Bauschke and Dr. Warren Hare for all their help and suggestions, and Chayne Planiden for his comments on my English.

Chapter 1

Introduction

In this thesis, the split feasibility problem, CQ algorithm and error estimation are studied. After an introduction to Euclidean space and convex analysis, the SFP and the CQ algorithm are introduced. Previous work on the CQ algorithm is presented and a new form of the relaxed CQ algorithm with subgradient projector is studied. This shows that the CQ algorithm with López's stepsize is actually a subgradient projector algroithm. A new proof is given for the new relaxed CQ algorithm.

Then, the S-subdifferential, which is defined with respect to a set S, is introduced. With this new subdifferential, the SFP is extended to a fully nonconvex case where both the two functions in the SFP are only required to be continuous and S-subdifferentiable. Then the SPA with S-subdifferential is shown to converge to a solution of this case.

After that, the SFP is extended to the case that function g is prox-regular but not necessarily convex. A more general CQ algorithm with proximal mapping is shown to converge to a solution of the extended SFP.

In the last chapter, the error estimation is studied and some examples are presented. The error estimation is shown to be related to subgradient projectors.

Chapter 2

Inner Product Spaces and Convex Analysis

2.1 Inner Product Spaces

In this chapter, we will introduce some background material on inner product spaces and some necessary convex analysis used later in the thesis.

Definition 2.1. (Vector Space) A vector space consists of a set V with elements called vectors, along with two operations such that the following properties hold:

(1) Vector addition: For all $u, v \in V$, we have $u + v \in V$ and the following are satisfied.

  (i) Commutativity: $u + v = v + u$, $\forall u, v \in V$.

  (ii) Associativity: $u + (v + w) = (u + v) + w$, $\forall u, v, w \in V$.

  (iii) Zero: there is a vector $\mathbf{0} \in V$ such that $\mathbf{0} + u = u = u + \mathbf{0}$, $\forall u \in V$.

  (iv) Inverses: for each $u \in V$, there is a vector $-u$ such that $u + (-u) = \mathbf{0}$.

(2) Scalar multiplication: For all $u, v \in V$ and for all $r, s \in \mathbb{R}$, we have $ru \in V$ and the following are satisfied.

(i) Left distributivity: $(r+s)v = rv + sv$, $\forall r, s \in \mathbb{R}^n, \forall v \in V$.

(ii) Associativity: $r(sv) = (rs)v$, $\forall r, s \in \mathbb{R}^n, \forall v \in V$.

(iii) Right distributivity: $r(u+v) = ru + rv$, $\forall r, s \in \mathbb{R}^n, \forall v \in V$.

(iv) Neutral element: $1v = v$, $\forall v \in V$.

(v) Absorbing element: $0v = \mathbf{0}$, $\forall v \in V$.

(vi) Inverse neutral element: $(-1)v = -v$, $\forall v \in V$.

Example 2.2. The space $\mathbb{R}^n$ consists of vectors $v = (v_1, \cdots, v_n)$ with $v_i \in \mathbb{R}$ for $1 \leq i \leq n$ and operations defined by

$$(u_1, \cdots, u_n) + (v_1, \ldots, v_n) := (u_1 + v_1, \cdots, u_n + v_n)$$

$$r(v_1, \cdots, v_n) := (rv_1, \cdots, rv_n)$$

where $r \in \mathbb{R}$.

We recall the definition of a norm and an inner product.

Definition 2.3. A norm $\|\cdot\|$ on a vector space V is a function $V \to R$ with the following properties:

(i) Positive definite: $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ if and only if $x = 0$.

(ii) Homogeneous: $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in V$.

(iii) Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

Definition 2.4. (Inner Product Space) An inner product on a vector space V is a function $\langle\cdot,\cdot\rangle$: $V \times V \to R$ with the following properties:

(i) Positive definite: $\langle x, x\rangle \geq 0$, $\forall x \in V$, and $\langle x, x\rangle$=0 if and only if $x = 0$.

(ii) Symmetry: $\langle x, y\rangle = \langle y, x\rangle$, $\forall x, y \in V$.

(iii) Bilinearity: $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ $\forall \alpha, \beta \in \mathbb{R}$, $x, y, z \in V$.

We call a vector space paired with an inner product and norm induced by $\|x\| := \langle x, x\rangle^{\frac{1}{2}}$ an inner product space.

Example 2.5. In $\mathbb{R}^n$, define

$$(\forall x \in \mathbb{R}^n)(\forall y \in \mathbb{R}^n) \quad \langle x, y\rangle = \sum_{i=1}^{n} x_i y_i. \tag{2.1}$$

Then $\mathbb{R}^n$ is an inner product space.

Throughout this thesis we will use the Euclidean norm, which is given by

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}, \tag{2.2}$$

so in this thesis $\|x\|$ refers to the Euclidean norm of x.

Definition 2.6. In a normed vector space $(V, \|\cdot\|)$, a sequence $(v_n)_{n=1}^{\infty}$ converges to $v \in V$ if $\lim_{n\to\infty} \|v_n - v\| = 0$.

Definition 2.7. A sequence $(x_n)_{n=1}^{\infty}$ is called a Cauchy sequence if for every $\epsilon > 0$, there is an integer $N > 0$ such that $\|x_n - x_m\| < \epsilon$ for all $n, m \geq N$.

Definition 2.8. An inner product space $\mathcal{X}$ is called **complete**, or a **Hilbert Space**, if each Cauchy sequence in $\mathcal{X}$ converges to a point in $\mathcal{X}$.

Hereafter, $\mathcal{H}$ denotes a Hilbert space.

Example 2.9. [8, Example 1.5-1] $\mathbb{R}^n$ is complete. Thus, $\mathbb{R}^n$ is a Hilbert space.

The Cauchy-Schwarz Inequality is an important fact and really useful.

Fact 2.10. [7, Theorem 1.2] ***(Cauchy-Schwarz Inequality)*** *Let x and y be in $\mathbb{R}^n$. Then*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|, \text{ i.e. }, |\sum_{i=1}^{n} x_i y_i| \leq \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}. \tag{2.3}$$

Moreover, $\langle x, y\rangle = \|x\| \cdot \|y\| \Leftrightarrow (\exists \alpha \in [0, \infty))\ x = \alpha y$ *or* $y = \alpha x$.

Fact 2.11. *Let $x, y \in \mathbb{R}^n$, and $\langle x, z\rangle = \langle y, z\rangle$ $\forall z \in \mathbb{R}^n$. Then $x = y$.*

Proof. Since

$$\langle x, z\rangle = \langle y, z\rangle \quad \forall z \in \mathbb{R}^n,$$

taking $z = x - y$ gives us

$$\langle x, x - y\rangle = \langle y, x - y\rangle.$$

Thus,

$$\|x - y\|^2 = \langle x - y, x - y\rangle = 0.$$

So $x = y$. ■

Definition 2.12. The identity operator is denoted by $\text{Id} : \mathbb{R}^n \to \mathbb{R}^n$, for which we have

$$(\forall x \in \mathbb{R}^n) \quad \text{Id}\, x = x. \tag{2.4}$$

Definition 2.13. Let $f : \mathbb{R}^n \to [-\infty, +\infty]$, and let x be a point such that $|f(x)| < +\infty$. We say that f is differentiable (or Fréchet differentiable) at

x if and only if there exists a vector x^* with the property

$$\lim_{y \to x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} = 0. \tag{2.5}$$

If such x^* exists, it is called the gradient of f at x and is denoted by $\nabla f(x)$.

Definition 2.14. For a function $f : \mathbb{R}^n \to \mathbb{R}$, we define

$$\operatorname{argmin} f = \{x \in \mathbb{R}^n : f(x) = \inf_{y \in \mathbb{R}^n} f(y)\}. \tag{2.6}$$

2.2 Linear Operators

Definition 2.15. An operator $L : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if and only if

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \tag{2.7}$$

for all $x, y \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$.

Definition 2.16. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, the adjoint of L is the unique linear operator $L^* : \mathbb{R}^m \to \mathbb{R}^n$ that satisfies

$$(\forall x \in \mathbb{R}^n)(\forall y \in \mathbb{R}^m) \quad \langle Lx, y \rangle = \langle x, L^* y \rangle. \tag{2.8}$$

Lemma 2.17. *In $\mathbb{R}^n$, a linear operator $L : \mathbb{R}^n \to \mathbb{R}^m$ can always be represented as an $m \times n$ matrix. In this manner, L^* is the transpose of L, which is an $n \times m$ matrix.*

Proof. Let A be a $m \times n$ matrix. Then it is obvious that A is a linear operator. So we just need to prove that a linear operator $L : \mathbb{R}^n \to \mathbb{R}^m$ is a $m \times n$ matrix.

Let $x \in \mathbb{R}^n$ and

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Let e_k be the unit vector on the kth dimention, for example,

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then

$$\begin{aligned} Lx &= L \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + L \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + L \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \cdot Le_1 + x_2 \cdot Le_2 + \cdots + x_n \cdot Le_n. \end{aligned}$$

Let

$$A = (Le_1, Le_2, \ldots, Le_n).$$

Then

$$Ax = (Le_q, \dots, Le_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$
$$= x_1 Le_1 + x_2 Le_2 + \cdots + x_n Le_n.$$

Thus, $Lx = Ax$ and A is an $m \times n$ matrix, so L can be represented by matrix A.

Since L is an $m \times n$ matrix, we have for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$

$$\langle Lx, y\rangle = \langle Ax, y\rangle = \langle x, A^\top y\rangle. \tag{2.9}$$

Thus, by (2.8) we have for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$

$$\langle Lx, y\rangle = \langle x, L^* y\rangle. \tag{2.10}$$

So $A^\top y = L^* y$ for every $y \in \mathbb{R}^m$ by Fact 2.11. Hence, $L^* = A^\top$, which is an $n \times m$ matrix. ■

Definition 2.18. The induced norm or operator norm on $\mathbb{R}^{m\times n}$ is given by:

$$\|A\| = \max\left\{ \frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n \text{ with } x \neq 0 \right\}.$$

Fact 2.19. [2, Fact 2.18] *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator and let $\ker T = \{x : Tx = 0\}$ be the kernel of T and $\operatorname{ran} T = T(\mathbb{R}^n)$. Then the following hold:*

(i) $T^{**} = T$.

(ii) $\|T^*\| = \|T\| = \sqrt{\|T^*T\|}$.

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

(iv) $(\operatorname{ran} T)^\perp = \ker T^*$.

(v) $\ker T^*T = \ker T$ *and* $\overline{\operatorname{ran}} TT^* = \overline{\operatorname{ran}} T$.

2.3 Sequence and Sets

Definition 2.20. A sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is bounded if $\exists M > 0$ such that $\|x_k\| \leq M$, $\forall k \geq 1$.

Fact 2.21. (Bolzano-Weierstrass Theorem) *Every bounded sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ has a convergent subsequence, i.e., $\exists (x_{k_l})_{l\in\mathbb{N}}$ with $x_{k_l} \to \bar{x} \in \mathbb{R}^n$.*

Definition 2.22. A set $C \subseteq \mathbb{R}^n$ is **closed** if it contains all limit points, i.e. for every sequence $(x_k)_{k=1}^\infty$ in C and $x_k \to \bar{x}$, we have $\bar{x} \in C$.

Definition 2.23. A set $O \subseteq \mathbb{R}^n$ is **open** if $\forall x \in O$, $\exists r > 0$ such that $B(x;r) \subseteq O$, where $B(x;r) = \{y \in \mathbb{R}^n : \|y - x\| \leq r\}$.

Definition 2.24. The **projection** from $\mathbb{R}^n$ onto a nonempty closed subset $C \subseteq \mathbb{R}^n$ defined by

$$P_C x := \operatorname{argmin}_{y\in C} \|x - y\|, \quad x \in \mathbb{R}^n. \tag{2.11}$$

Figure 1.1 gives us an example of projection.

Fact 2.25. *Let C be a closed nonempty subset of $\mathbb{R}^n$. Then $P_C(x)$ exists for every $x \in \mathbb{R}^n$.*

Proof. Let x be any point in $\mathbb{R}^n$. Since $\|x - c\| \geq 0$ for all $c \in C$, so $\inf_{c \in C} \|x - c\|$ exists. Let $(c_k)_{k \in \mathbb{N}}$ be a sequence in C such that

$$\|x - c_k\| \to \inf_{c \in C} \|x - c\|.$$

This means $(c_k)_{k \in \mathbb{N}}$ is bounded, by the Bozano-Weierstrass theorem, $(c_k)_{k \in \mathbb{N}}$ has a convergent sbsequence, say (c_{k_l}), $c_{k_l} \to \bar{c}$. Then

$$\|x - \bar{c}\| = \inf_{c \in C} \|x - c\|.$$

Since C is closed, so $\bar{c} \in C$. Thus, $P_C x = \bar{c}$. ■

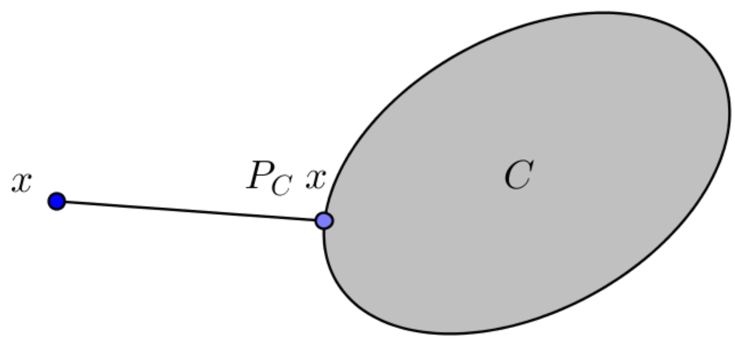

Figure 2.1: An example of projection.

2.4 Convex Analysis

2.4.1 Convex Sets and Convex Functions

Definition 2.26. A set $C \subseteq \mathbb{R}^n$ is convex if for any $x, y \in C$ and $\alpha \in]0,1[$ we have

$$\alpha x + (1 - \alpha) y \in C. \tag{2.12}$$

Graphically, a set C is convex if the line segment between any two points in C is also contained in C, see Figure 1.2.

Figure 2.2: Examples of convex sets.

Example 2.27. Let $r \in \mathbb{R}_{++}$ and $c \in \mathbb{R}^n$. Then the closed ball

$$B(c;r) = \{x \in \mathbb{R}^n : \|x - c\| \leq r\}$$

is convex.

Proof. Let $x, y \in B(c;r)$ and $\alpha \in]0,1[$. Then

$$\begin{aligned}
&\|\alpha x + (1-\alpha)y - c\| \\
&= \|\alpha(x-c) + (1-\alpha)(y-c)\| \\
&\leq \|\alpha(x-c)\| + \|(1-\alpha)(y-c)\| \quad \text{(By the Triangle Inequality)} \\
&= \alpha\|(x-c)\| + (1-\alpha)\|(y-c)\| \\
&\leq r.
\end{aligned}$$

Figure 2.3: Examples of nonconvex sets.

So $\alpha x + (1-\alpha)y \in B(c;r)$. Therefore $B(c;r)$ is convex. ■

Example 2.28. Let $r \in \mathbb{R}_{++}$. Then the cube

$$K = \{x \in \mathbb{R}^n : \|x\|_\infty \le r\}$$

is convex. Here $\|x\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$.

Proof. Let $x, y \in K$ and $\alpha \in \,]0,1[$. Then for all $i \in \{i \in \mathbb{N} : 1 \le i \le n\}$, $x_i, y_i \in \mathbb{R}$ and $|x_i|, |y_i| \le r$. Then

$$\begin{aligned}
&|\alpha x_i + (1-\alpha)y_i| \\
&\le |\alpha x_i| + |(1-\alpha)y_i| \quad \text{(By the Triangle Inequality)} \\
&= \alpha|x_i| + (1-\alpha)|y_i| \\
&\le r.
\end{aligned}$$

So $\alpha x + (1-\alpha)y \in K$. Therefore K is convex. ■

Example 2.29. Let $r_1, r_2 \in \mathbb{R}$, $a \in \mathbb{R}^n$. Then the slab set

$$S = \{x \in \mathbb{R}^n : r_1 \leq a^\top x \leq r_2\} \tag{2.13}$$

is convex.

Proof. Let $x, y \in S$. Then $r_1 \leq a^\top x \leq r_2$, $r_1 \leq a^\top y \leq r_2$. Assume $\alpha \in]0, 1[$. Then

$$\begin{aligned} \alpha r_1 \leq & a^\top \alpha x \leq \alpha r_2, \\ (1-\alpha) r_1 \leq & a^\top (1-\alpha) y \leq (1-\alpha) r_2. \end{aligned}$$

So $r_1 \leq a^\top [\alpha x + (1-\alpha) y] \leq r_2$. Hence $\alpha x + (1-\alpha) y \in S$. Therefore S is convex.

■

Definition 2.30. Let $f : \mathbb{R}^n \to [-\infty, +\infty]$. The domain of f is

$$\operatorname{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}, \tag{2.14}$$

the graph of f is

$$\operatorname{gra} f = \{(x, \xi) \in \mathbb{R}^n \times \mathbb{R} : f(x) = \xi\}, \tag{2.15}$$

the epigraph of f is

$$\operatorname{epi} f = \{(x, \xi) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \xi\}, \tag{2.16}$$

the lower level set of f at height $\xi \in \mathbb{R}$ is

$$\operatorname{lev}_{\leq \xi} f = \{x \in \mathbb{R}^n : f(x) \leq \xi\}, \tag{2.17}$$

and the strict lower level set of f at height $\xi \in \mathbb{R}$ is

$$\operatorname{lev}_{< \xi} f = \{x \in \mathbb{R}^n : f(x) < \xi\}, \tag{2.18}$$

Definition 2.31. Let $f : \mathbb{R}^n \to [-\infty, +\infty]$. Then $f(x)$ is **convex** if its epigraph $\{(x, r) : f(x) \leq r\}$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}$. Moreover, f is concave if $-f$ is convex.

Graphically, a function f is convex if the line segment joining $(x, f(x))$ and $(y, f(y))$ lies above the graph of f, (see Figure 1.4).

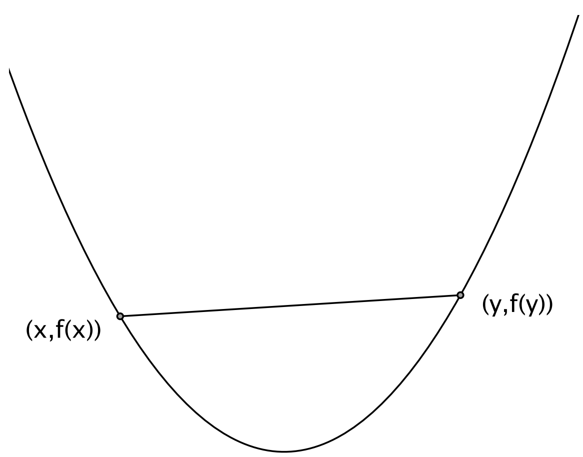

Figure 2.4: An example of convex functions.

Proposition 2.32. [2, Proposition 8.2] *Let* $f : \mathbb{R}^n \to [-\infty, +\infty]$ *be convex. Then its domain* $\operatorname{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ *is convex.*

Definition 2.33. We say that the function $f : \mathbb{R}^n \to [-\infty, +\infty]$ is proper if its domain is nonempty and $f > -\infty$.

Definition 2.34. Let $f : \mathbb{R}^n \to [-\infty, +\infty]$ be a proper function. Then $f(x)$ is **strictly convex** if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[) \tag{2.19}$$

$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y). \tag{2.20}$$

Now let C be a nonempty convex subset of $\operatorname{dom} f$. Then f is convex on C if

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in]0,1[)$$
$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \tag{2.21}$$

and f is strictly convex on C if

$$(\forall x \in C)(\forall y \in C(\forall \alpha \in]0,1[) \tag{2.22}$$

$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y). \tag{2.23}$$

Example 2.35. The function $\|\cdot\|$ is convex, but not strictly convex.

Proof. Let $x, y \in \mathbb{R}^n$, $\alpha \in]0,1[$. Then

$$\begin{aligned} \|\alpha x + (1-\alpha)y\| &\leq \|\alpha x\| + \|(1-\alpha)y\| \quad \text{(By the Triangle Inequality)} \\ &= \alpha\|x\| + (1-\alpha)\|y\| \end{aligned}$$

So $\|\cdot\|$ is convex. Using $y = 0$, we have

$$\|\alpha x + (1-\alpha)0\| = \alpha\|x\| + (1-\alpha)\|0\|.$$

Therefore it is not strictly convex. ■

Example 2.36. The function $\|\cdot\|^2$ is strictly convex.

Proof. Let $x, y \in \mathbb{R}^n$ and $x \neq y$. Let $\alpha \in]0,1]$. Then we have

$$\begin{aligned}
&\|\alpha x-(1-\alpha)y\|^2-(\alpha\|x\|^2+(1-\alpha)\|y\|^2)\\
&=\alpha^2\|x\|^2+2\alpha(1-\alpha)\langle x,y\rangle+(1-\alpha)^2\|y\|^2-\alpha\|x\|^2-(1-\alpha)\|y\|^2\\
&=(\alpha-1)\alpha\|x\|^2+2\alpha(1-\alpha)\langle x,y\rangle-(1-\alpha)\alpha\|y\|^2\\
&=-\alpha(1-\alpha)(\|x\|^2-2\langle x,y\rangle+\|y\|^2)\\
&=-\alpha(1-\alpha)\|x-y\|^2<0 \quad (x\neq y).
\end{aligned}$$

Therefore $\|\cdot\|^2$ is strictly convex. ■

Example 2.37. Let $C \subseteq \mathbb{R}^n$ be convex. Then $\mathrm{d}_C : \mathbb{R}^n \to [0,+\infty] : x \mapsto \|(\mathrm{Id} - P_C)x\|$ is convex.

Proof. Let $x, y \in \mathbb{R}^n$, $z \in C$ and $\alpha \in]0,1[$. Then

$$\mathrm{d}_C(\alpha x+(1-\alpha)y)\le\|\alpha x+(1-\alpha)y-z\|. \tag{2.24}$$

Since C is convex, we have

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

Now take $z = \alpha P_C(x) + (1-\alpha)P_C(y) \in C$. By (2.24) we have

$$\begin{aligned}
\mathrm{d}_C(\alpha x+(1-\alpha)y)&\le\|\alpha x+(1-\alpha)y-(\alpha P_C x+(1-\alpha)P_C y)\|\\
&=\|\alpha(\mathrm{Id}-P_C)x+(1-\alpha)(\mathrm{Id}-P_C)y\|\\
&\le\alpha\|(\mathrm{Id}-P_C)x\|+(1-\alpha)\|(\mathrm{Id}-P_C)y\|
\end{aligned}$$

$$= \alpha d_C(x) + (1-\alpha) d_C(y).$$

Therefore d_C is convex.

■

Fact 2.38. [14, Proposition 2.7] *For a convex function $f : \mathbb{R}^n \to \mathbb{R}$ all the level sets of type* $\mathrm{lev}_{\leq \xi} f$ *and* $\mathrm{lev}_{< \xi} f$ *are convex.*

Proof. Since f is convex, so $\mathrm{epi}\, f$ is convex. Let $x, y \in \mathrm{lev}_{\leq \xi} f$. Let $\alpha \in]0,1[$, we have $\alpha(x, f(x)) \in \mathrm{epi}\, f$, $(1-\alpha)(y, f(y)) \in \mathrm{epi}\, f$. Thus

$$\alpha(x, f(x)) + (1-\alpha)(y, f(y)) \in \mathrm{epi}\, f.$$

Since $f(x) \leq \xi$, $f(y) \leq \xi$, so

$$f(\alpha x + (1-\alpha y)) \leq \alpha f(x) + (1-\alpha) f(y) \leq \xi.$$

So $\alpha x + (1-\alpha y) \in \mathrm{lev}_{\leq \xi} f$. So $\mathrm{lev}_{\leq \xi} f$ is convex. Similarly we can prove that $\mathrm{lev}_{< \xi} f$ is convex.

■

Fact 2.39. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator. Then $g = f \circ A$ is convex.*

Proof. Let $x, y \in \mathbb{R}^m$, $\alpha \in]0,1[$. Then

$$\begin{aligned} & g(\alpha x + (1-\alpha) y) \\ &= f \circ A(\alpha x + (1-\alpha) y) \\ &= f(\alpha A x + (1-\alpha) A y). \end{aligned}$$

Since f is convex, so we have

$$\begin{aligned}
&f(\alpha Ax + (1-\alpha)Ay) \\
&\leq \alpha f(Ax) + (1-\alpha) f(Ay) \\
&= \alpha g(x) + (1-\alpha) g(y).
\end{aligned}$$

Therefore,

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha) g(y).$$

Thus, g is convex. ■

Fact 2.40. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be convex and increasing, $g : \mathbb{R}^n \to \mathbb{R}_+$ be convex. Then $h = f \circ g$ is convex.*

Proof. Let $x, y \in \mathbb{R}^n$, $\alpha \in]0,1[$. Since g is convex, we have

$$\alpha g(x) + (1-\alpha) g(y) \geq g(\alpha x + (1-\alpha) y).$$

Since f is increasing, we have

$$f(\alpha g(x) + (1-\alpha) g(y)) \geq f(g(\alpha x + (1-\alpha) y)).$$

Since f is convex, we have

$$\begin{aligned}
&\alpha f(g(x)) + (1-\alpha) f(g(y)) \\
&\geq f(\alpha g(x) + (1-\alpha) g(y)) \\
&\geq f(g(\alpha x + (1-\alpha) y)).
\end{aligned}$$

So

$$\alpha h(x) + (1-\alpha)h(y) \geq h(\alpha x + (1-\alpha)y).$$

Therefore h is convex. ■

Definition 2.41. A operator $T : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous with constant $\beta \in [0, \infty)$ if

$$(\forall x \in \mathbb{R}^n)(\forall y \in \mathbb{R}^n) \quad \|Tx - Ty\| \leq \beta\|x - y\|. \tag{2.25}$$

The operator T is locally Lipschitz continuous near a point $x_0 \in \mathbb{R}^n$ if there exists $r \in \mathbb{R}_{++}$ such that $T|_{B(x_0;r)}$ is Lipschitz continuous.

Definition 2.42. Let $A \subseteq \mathbb{R}$. The infimum of A is the largest lower bound and denoted by $\inf A$. The supremum of A is the smallest upper bound and denoted by $\sup A$.

Definition 2.43. The lower limit of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $\bar{x}$ is the value in $\mathbb{R}$ defined by

$$\begin{aligned} \liminf_{x \to \bar{x}} f(x) &:= \lim_{\delta \searrow 0} \left[\inf_{x \in B(\bar{x};\delta)} f(x) \right] \\ &= \sup_{\delta > 0} \left[\inf_{x \in B(\bar{x};\delta)} f(x) \right]. \end{aligned} \tag{2.26}$$

Definition 2.44. A function f is lower semicontinous at a point x_0 if

$$\liminf_{x \to x_0} f(x) \geq f(x_0). \tag{2.27}$$

The function is said to be lower semicontinuous on $\mathbb{R}^n$ if it is lower semicontinuous at every point $x_0 \in \mathbb{R}^n$.

Fact 2.45. [2, Theorem 8.29] *Let* $f : \mathcal{H} \to \,]-\infty, +\infty]$ *be proper and convex, and let* $x_0 \in \operatorname{dom} f$*. Then the following are equivalent:*

(i) f *is locally Lipschitz continuous near* x_0*.*

(ii) f *is continuous at* x_0*.*

(iii) f *is bounded on a neighborhood of* x_0*.*

(iv) f *is bounded above on a neighborhood of* x_0*.*

Moreover, if one of these conditions holds, then f *is locally Lipschitz continuous on* $\operatorname{int}\operatorname{dom} f$*.*

Fact 2.46. [2, Corollary 8.30] *Let* $f : \mathcal{H} \to \,]-\infty, +\infty]$ *be proper and convex, and suppose that one of the following holds:*

(i) f *is bounded above on some neighborhood.*

(ii) f *is lower semicontinuous.*

(iii) $\mathcal{H}$ *is finite-dimensional.*

Then $\operatorname{cont} f = \operatorname{int}\operatorname{dom} f$*.*

Here $\operatorname{cont} f$ is the set of points at which the function f is continuous.

Theorem 2.47. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is lower semicontinuous and convex, then* f *is locally Lipschitz continuous.*

Proof. By Fact 2.46(iii) we have f is continuous. Then by Fact 2.45(ii) we have f is locally Lipchitz in $\mathbb{R}^n$. ■

Definition 2.48. Let f and g be functions from $\mathbb{R}^n$ to $[-\infty, +\infty]$. The infimal convolution of f and g is

$$f \Box g : \mathbb{R}^n \to [-\infty, +\infty] : x \mapsto \inf_{y \in \mathbb{R}^n} (f(y) + g(x-y)),$$

and it is exact at a point $x \in \mathbb{R}^n$ if $(f \Box g)(x) = \min_{y \in \mathbb{R}^n} f(y) + g(x-y)$, i.e.,

$$(\exists y \in \mathcal{H}) \quad (f \Box g)(x) = f(y) + g(x-y) \in \left]-\infty, +\infty\right] ;$$

$f \Box g$ is exact if it is exact at every point of its domain, in which case it is denoted by $f \boxdot g$.

Definition 2.49 (Moreau envelope and Proximal mapping)**.** Let $f : \mathbb{R}^n \to \left]-\infty, +\infty\right]$ and let $\lambda \in \mathbb{R}_{++}$. The Moreau envelope of f of parameter λ is

$$e_\lambda f = f \Box \big(\frac{1}{2\lambda} \| \cdot \|^2\big).$$

The proximal mapping of f of λ is

$$\mathrm{prox}_{\lambda f}(x) := \operatorname*{argmin}_{w} \big\{ f(w) + \frac{1}{2\lambda} \|w - x\|^2 \big\},$$

where for a function g, $\operatorname{argmin} g$ denotes the set of minimizers of g.

Fact 2.50. [2, Proposition 12.29] *Let $f : \mathbb{R}^n \to \left]-\infty, +\infty\right]$ be proper, lower semicontinuous and convex, and let $\lambda \in \mathbb{R}_{++}$. Then $e_\lambda f$ is Fréchet differentiable on $\mathbb{R}^n$, and its gradient is*

$$\nabla(e_\lambda f) = \lambda^{-1}(\mathrm{Id} - \mathrm{prox}_{\lambda f}) \tag{2.28}$$

is λ^{-1} Lipschitz continuous.

2.4.2 Projections on Convex Sets

The first result is a characterization of projection.

Fact 2.51. [2, Theorem 3.14] *Let C be a closed convex subset of $\mathbb{R}^n$, then for every x and p in $\mathbb{R}^n$,*

$$p = P_C x \Leftrightarrow [p \in C \quad \textit{and} \quad (\forall y \in C) \quad \langle y - p, x - p \rangle \leq 0]. \tag{2.29}$$

Example 2.52. Let $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ and $H = \{x \in \mathbb{R}^n : a^\top x = b\}$. Then

$$P_H x = \begin{cases} x + \frac{b - a^\top x}{\|a\|^2} a, & x \notin H, \\ x, & x \in H. \end{cases} \tag{2.30}$$

Let $H_s = \{x \in \mathbb{R}^n : a^\top x \leq b\}$. Then

$$P_{H_s} x = \begin{cases} x + \frac{b - a^\top x}{\|a\|^2} a, & x \notin H_s, \\ x, & x \in H_s. \end{cases} \tag{2.31}$$

Example 2.53. Let $B(0;1) = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$. Then

$$P_{B(0;1)x} = \begin{cases} \frac{x}{\|x\|}, & \|x\| > 1, \\ x, & \|x\| \leq 1. \end{cases} \tag{2.32}$$

For proofs of Example 2.52 and Example 2.53, see [2].

2.4.3 Subgradient and Subdifferential

Definition 2.54. A vector $u \in \mathbb{R}^n$ is said to be a subgradient of a convex function $f : \mathbb{R}^n \to [-\infty, +\infty]$ at the point x if we have

$$(\forall y \in \mathbb{R}^n) \quad \langle y - x, u \rangle + f(x) \leq f(y). \tag{2.33}$$

The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f(x)$.

Fact 2.55. [2, Theorem 16.2, (Fermat's rule)] *Let $f : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then*

$$\operatorname{argmin} f = \{x \in \mathbb{R}^n | 0 \in \partial f(x)\}. \tag{2.34}$$

Fact 2.56. [2, Proposition 16.5] *Let $f : \mathbb{R}^n \to]-\infty, +\infty]$ and $g : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, let $L : \mathbb{R}^n \to \mathbb{R}^m$ be an linear operator, and let $\lambda \in (0, \infty)$. Then the following hold:*

(i) $\partial(\lambda f) = \lambda \partial f$.

(ii) *Suppose that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \emptyset$, where $\operatorname{dom} f$ denotes the domain of f. Then $\partial f + L^* \circ (\partial g) \circ L \subset \partial(f + g \circ L)$.*

Fact 2.57. [2, Proposition 16.37] *Let $\mathcal{K}$ be a real Hilbert space. Let $f : \mathcal{H} \to]-\infty, +\infty]$ and $g : \mathcal{K} \to]-\infty, +\infty]$ be proper, lsc and convex, and let $L : \mathcal{H} \to \mathcal{K}$ be a linear operator. Suppose f and g are polyhedral, and $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \emptyset$. Then*

$$\partial(f + g \circ L) = \partial f + L^* \circ (\partial g) \circ L. \tag{2.35}$$

Fact 2.58. [2, Proposition 17.26] *Let $f : \mathbb{R}^n \to]-\infty, +\infty]$ be proper and* *convex, and let $x \in \operatorname{dom} f$. Suppose that f is differentiable. Then*

$$\partial f(x) = \{\nabla f(x)\}. \tag{2.36}$$

Example 2.59. Assume $f : \mathbb{R}^m \to \mathbb{R}$ is convex, A is a $m \times n$ matrix, and $x \in \mathbb{R}^n$. Then

$$\partial e_\lambda f(Ax) = \{A^* \nabla e_\lambda f(Ax)\}. \tag{2.37}$$

Proof. By Fact 2.58, we have $\partial e_f(Ax) = \{\nabla(e_\lambda f \circ A)(x)\}$. Then by Fact 2.57 and Fact 2.50 we have $\partial e_\lambda f(Ax) = \{A^* \nabla e_\lambda f(Ax)\}$. ■

Example 2.60. Let $f(x) = \frac{1}{2}\|x\|^2$. Then

$$\partial f(x) = \{x\}. \tag{2.38}$$

Proof. Since $f(x) = \frac{1}{2}\|x\|^2$ is differentiable, by Fact 2.58 we have

$$\partial f(x) = \{\nabla f(x)\} = \{x\}.$$

■

Example 2.61. Let $f(x) = |x|$. Then

$$\partial f(x) = \begin{cases} \{-1\}, & x < 0, \\ [-1, 1], & x = 0, \\ \{1\}, & x > 0. \end{cases} \tag{2.39}$$

Proof. If $x < 0$, $f(x) = -x$ and f is differentiable. So for $x < 0$, $\partial f(x) = \nabla f(x) = \{-1\}$. Similarly, if $x > 0$, $f(x) = x$ and f is differentiable. So for $x > 0$, $\partial f(x) = \nabla f(x) = \{1\}$.

If $x = 0$, let $u \in \partial f(0)$. Then we have $\forall y \in \mathbb{R}^n$,

$$\langle y - 0, u\rangle + f(0) \leq f(y).$$

So

$$(\forall y \in \mathbb{R}) \quad y \cdot u \leq |y|.$$

If $y < 0$, then $u \geq -1$. If $y > 0$, then $u \leq 1$. Thus we have $u \in [-1, 1]$.

On the other hand, let $u \in [-1, 1]$. Then,

$$\begin{aligned}
&\langle u, x\rangle + f(0)\\
&= ux\\
&\leq |u| \cdot |x| \quad \text{(By the Cauchy-Schwarz Inequality)}\\
&\leq |x| \quad (u \in [-1, 1])\\
&= f(x).
\end{aligned}$$

Thus, $u \in \partial f(0)$. Therefore, $\partial f(0) = [-1, 1]$. ■

Example 2.62. Let $f(x, y) = |x| + |y|$. Then

$$\partial f(x,y) = \begin{cases} (\frac{x}{|x|}, \frac{y}{|y|}), & x \neq 0, y \neq 0,\\ [-1,1] \times \frac{y}{|y|}, & x = 0, y \neq 0,\\ \frac{x}{|x|} \times [-1,1], & x \neq 0, y = 0,\\ [-1,1] \times [-1,1], & x = 0, y = 0. \end{cases} \tag{2.40}$$

Fact 2.63. [14, Example 10.2] *Let* $f : \mathbb{R}^n \to]-\infty, +\infty]$ *be proper, convex* *and* $\lambda > 0$. *Then*

$$\operatorname{prox}_{\lambda f} = (\operatorname{Id} + \lambda \partial f)^{-1}. \tag{2.41}$$

For a set-valued mapping $T : \mathbb{R}^n \to \mathbb{R}^n$, we define

$$T^{-1}(y) = \{x \in \mathbb{R}^n : y \in T(x)\}, \tag{2.42}$$

for each $y \in \mathbb{R}^n$.

Fact 2.64. [2, Proposition 16.35] *Let* $f : \mathbb{R}^n \to]-\infty, +\infty]$ *be proper, convex and continuous. Then* $\operatorname{ran}(\operatorname{Id} + \partial f) = \mathbb{R}^n$.

Fact 2.65. [2, Corollary 16.38] *Let* $f : \mathbb{R}^n \to]-\infty, +\infty]$, $g : \mathbb{R}^m \to]-\infty, +\infty]$ *be proper, convex and continuous. If* $\operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g \neq \emptyset$, *then* $\partial(f + g) = \partial f + \partial g$.

Fact 2.66. [2, Proposition 16.17] *Let* $f : \mathcal{H} \to \mathbb{R}$ *be lower semicontinuous and convex. Then the following are equivalent:*

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

(ii) f *is Lipschitz continuous relative to every bounded subset of* $\mathcal{H}$.

(iii) $\operatorname{dom} \partial f = \mathcal{H}$ *and* ∂f *maps every bounded subset of* $\mathcal{H}$ *to a bounded set.*

If $\mathcal{H}$ *is finite-dimensional, then* f *satisfies these properties.*

Theorem 2.67. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is continuous and convex, then* f *is subdifferentiable, i.e.,* $\partial f(x) \neq \emptyset, \forall x \in \mathbb{R}^n$.

Proof. Since $\mathbb{R}^n$ is finite-dimensional, then by Fact 2.66(iii), we have $\operatorname{dom} \partial f = \mathbb{R}^n$. The proof is done. ■

Definition 2.68. (Fenchel Subdifferential) For (not necessarily convex) $f : \mathbb{R}^n \to [-\infty, +\infty]$, define its Fenchel subdifferential at $\bar{x}$

$$\partial f(\bar{x}) := \{\nu \in \mathbb{R}^n : f(x) \geq f(\bar{x}) + \langle \nu, x - \bar{x} \rangle \text{ for all } x \in \mathbb{R}^n\}. \tag{2.43}$$

When f is convex, $\partial f(\bar{x})$ is the usual subdifferential.

2.4.4 Nonexpansive Operators

Definition 2.69. Let D be a nonempty subset of $\mathbb{R}^n$. Let $T : D \to \mathbb{R}^n$. Then T is

(i) nonexpansive if

$$(\forall x, y \in D) \quad \|Tx - Ty\| \leq \|x - y\|; \tag{2.44}$$

(ii) firmly nonexpansive if

$$(\forall x, y \in D) \quad \|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2; \tag{2.45}$$

(iii) quasinonexpansive if

$$(\forall x \in D)(y \in \operatorname{Fix} T) \quad \|Tx - y\| \leq \|x - y\|; \tag{2.46}$$

(iv) and strictly quasinonexpansive if

$$(\forall x \in D \setminus \operatorname{Fix} T)(y \in \operatorname{Fix} T) \quad \|Tx - y\| < \|x - y\|; \tag{2.47}$$

where the fixed points set is defined by

$$\operatorname{Fix} T = \{x \in \mathbb{R}^n : Tx = x\}. \tag{2.48}$$

Proposition 2.70. *Let D be a nonempty subset of $\mathbb{R}^n$. Let $T : D \to \mathbb{R}^n$.*

(i) *T is firmly nonexpansive $\Rightarrow$ T is nonexpansive.*

(ii) *T is nonexpansive $\Rightarrow$ T is quasinonexpansive.*

(iii) *T is strictly quasinonexpansive $\Rightarrow$ T is quasinonexpansive.*

(iv) *T is firmly nonexpansive $\Rightarrow$ T is strictly quasinonexpansive.*

Proof. (i): If T is firmly nonexpansive, then $\forall x, y \in D$ we have

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

So $\|Tx - Ty\|^2 \leq \|x - y\|^2$. Hence $\|Tx - Ty\| \leq \|x - y\|$. So T is nonexpansive.

(ii): If T is nonexpansive, then $\forall x, y \in D$ we have

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

So

$$(\forall x \in D)(y \in \operatorname{Fix} T) \quad \|Tx - y\| \leq \|x - y\|.$$

Hence, T is quasinonexpansive.

(iii): Let T be strictly quasinonexpansive. Then

$$(\forall x \in D \setminus \operatorname{Fix} T)(y \in \operatorname{Fix} T) \quad \|Tx - y\| < \|x - y\|.$$

So if $x \in D \setminus \operatorname{Fix} T$, $y \in \operatorname{Fix} T$, then $\|Tx - y\| < \|x - y\|$. If $x, y \in \operatorname{Fix} T$. then $\|Tx - y\| = \|x - y\|$. Therefore, $\forall x \in D$, $\forall y \in \operatorname{Fix} T$, $\|Tx - Ty\| \leq \|x - y\|$. So T is quasinonexpansive.

(iv): If T is firmly nonexpansive, then $\forall x, y \in D$ we have

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

Then $\forall x \in D \setminus \operatorname{Fix} T$, $y \in \operatorname{Fix} T$,

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

So

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

So $\forall x \in D \setminus \operatorname{Fix} T$, $y \in \operatorname{Fix} T$, $\|Tx - Ty\| < \|x - y\|$. Thus T is strictly quasinonexpansive. ■

Lemma 2.71. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an operator. The following statements are equivalent.*

(i) *T is firmly nonexpansive.*

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

(iii) $\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle, \quad x, y \in \mathbb{R}^n.$

For proof of this lemma, see [2].

A typical example of a firmly nonexpansive mapping is the orthogonal projection.

Fact 2.72. *Let C be a convex closed subset of $\mathbb{R}^n$. Then P_C is firmly nonexpansive.*

Proof. Fix x and y in $\mathbb{R}^n$. By (2.29) we have that

$$\langle P_C y - P_C x, x - P_C x\rangle \leq 0$$

and

$$\langle P_C x - P_C y, y - P_C y\rangle \leq 0.$$

So we have

$$\begin{aligned} &\langle P_C y - P_C x, x - P_C x\rangle + \langle P_C x - P_C y, y - P_C y\rangle \\ &= \langle P_C x - P_C y, y - P_C y - (x - P_C x)\rangle \\ &= \langle P_C x - P_C y, (P_C x - P_C y) - (x - y)\rangle \\ &= \langle P_C x - P_C y, P_C x - P_C y\rangle - \langle P_C x - P_C y, (x - y)\rangle \\ &= \|P_C x - P_C y\|^2 - \langle x - y, P_C x - P_C y\rangle \\ &\leq 0. \end{aligned}$$

So

$$\|P_C x - P_C y\|^2 \leq \langle x - y, P_C x - P_C y\rangle. \tag{2.49}$$

The claim follows from Lemma 2.71. ■

Definition 2.73. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an operator, $\epsilon \geq 0$. We say that T is ϵ-relaxed firmly nonexpansive on $U \subseteq \mathbb{R}^n$ if

$$\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle + \epsilon\|x - y\|^2 \quad \forall x, y \in U. \tag{2.50}$$

2.4.5 Fejér Monotonicity and Quasi Fejér Monotonicity

Definition 2.74. Let C be a nonempty subset of $\mathbb{R}^n$ and let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then $(x_k)_{k\in\mathbb{N}}$ is Fejér Monotone with respect to C if

$$(\forall c \in C)(\forall k \in \mathbb{N}) \quad \|x_{k+1} - c\| \leq \|x_k - c\|. \tag{2.51}$$

The following result concerning Fejér monotone sequence is crucial for the proof of a later result.

Lemma 2.75. [10, Lemma 2.3] *Let C be a nonempty closed convex subset in $\mathbb{R}^n$. If the sequence (x_k) is Féjer monotone w.r.t. C, then the following hold:*

(i) $x_k \to x^* \in C$ if and only if every cluster point of (x_k) is in C;

(ii) the sequence $(P_C x_k)$ converges;

(iii) if $x_k \to x^* \in C$, then $x^* = \lim_{k\to\infty} P_C x_k$.

Proof. (i): Since (x_k) is Féjer monotone, $(\|x_k - c\|)_{k\in\mathbb{N}}$ converges for every $c \in C$. Suppose c_1, c_2 are cluster points of (x_k) in C. Then we have that

$$\|x_k - c_1\|^2 - \|x_k - c_2\|^2 = 2\langle x_k, c_2 - c_1\rangle + \|c_1\|^2 - \|c_2\|^2$$

converges. That is to say, $\langle x_k, c_1 - c_2\rangle$ converges. So $\langle c_1, c_1 - c_2\rangle = \langle c_2, c_1 - c_2\rangle$, i.e., $\langle c_1 - c_2, c_1 - c_2\rangle = 0$. Thus $c_1 = c_2$.

(*ii*): For every m and n in $\mathbb{N}$,

$$\|P_C x_k - P_C x_{k+m}\|^2$$
$$= \|P_C x_k - x_{k+m}\|^2 + \|x_{k+m} - P_C x_{k+m}\|^2 + 2\langle P_C x_k - x_{k+m}, x_{k+m} - P_C x_{k+m}\rangle$$
$$\leq \|P_C x_k - x_k\|^2 + \mathrm{d}_C^2(x_{k+m}) + 2\langle P_C x_k - P_C x_{k+m}, x_{k+m} - P_C x_{k+m}\rangle$$
$$+ 2\langle P_C x_{k+m} - x_{k+m}, x_{k+m} - P_C x_{k+m}\rangle$$
$$\leq \|P_C x_k - x_k\|^2 + \mathrm{d}_C^2(x_{k+m}) + 2\langle P_C x_k - P_C x_{k+m}, x_{k+m} - P_C x_{k+m}\rangle - 2\mathrm{d}_C^2(x_{k+m})$$
$$\leq \mathrm{d}_C^2(x_k) - \mathrm{d}_C^2(x_{k+m}).$$

Taking the infimum over all $x \in C$ on (2.51) gives us

$$d_C(x_{n+1}) \leq d_C(x_n). \tag{2.52}$$

Hence, $(d_C(x_k))_{k\in\mathbb{N}}$ converges. Thus $(P_C x_k)_{k\in\mathbb{N}}$ is a Cauchy sequence. Since $\mathbb{R}^n$ is complete, so $P_C x_k$ converges.

(*iii*): Suppose $(P_C x_k)_{k\in\mathbb{N}}$ converges to some point $y \in C$. Since

$$x^* - P_C x_k \to x^* - y,$$

and

$$x_k - P_C x_k \to x^* - y,$$

we have

$$0 \geq \langle x^* - P_C x_k, x_k - P_C x_k\rangle \to \|x^* - y\|^2.$$

Thus $x^* = y$. ■

Remark 2.76. (i) Lemma 2.75(iii) is easier to prove in $\mathbb{R}^n$, since P_C is continuous. Indeed

$$\lim_{k\to\infty} P_C x_k = P_C x^* = x^*, \tag{2.53}$$

since $x^* \in C$.

(ii) In a general Hilbert space, assuming that all weak cluster points of $(x_k)_{k\in\mathbb{N}}$ lies in C, one can show $x_k \rightharpoonup x^* \in C$ and $x^* = \lim_{k\to\infty} P_C x_k$ (in norm) as above by using the Fejér monotonicity of $(x_k)_{k\in\mathbb{N}}$, see [2, pages 76–77]. However, on should note that P_C is not weak-to-norm continuous.

Theorem 2.77. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be quasinonexpansive. Then the sequence $(x_k)_{k=0}^{\infty}$ generated by iteration $x_{k+1} = Tx_k$ is Fejér monotone respect to* $\operatorname{Fix} T$.

Proof. It is clear by the definition of quasinonexpansiveness. ■

Fact 2.78. *Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and let C be a nonempty subset in $\mathbb{R}^n$. Suppose that $(x_k)_{k\in\mathbb{N}}$ is Fejér monotone respect to C. Then the following hold:*

(i) *$(x_k)_{k\in\mathbb{N}}$ is bounded.*

(ii) *For every $c \in C$, $(\|x_k - c\|_{k\in\mathbb{N}})$ converges.*

(iii) *$(d_C(x_k))_{k\in\mathbb{N}}$ is decreasing and converges.*

Proof. (i): Let $c \in C$. Then by (2.51), we have $(x_k)_{k\in\mathbb{N}}$ lies in $B(c; \|x_0 - c\|)$. So it is bounded.

(ii): By (2.51), $(\|x_k - c\|_{k\in\mathbb{N}})$ is decreasing and bounded below. So $(\|x_k - c\|_{k\in\mathbb{N}})$ converges.

(iii): Taking the infimum in (2.51) over C gives us $(\forall k \in \mathbb{N})$,

$$\mathrm{d}_C(x_{k+1}) \leq \mathrm{d}_C(x_k). \tag{2.54}$$

So $(\mathrm{d}_C(x_k))_{k\in\mathbb{N}}$ is decreasing and converges. ■

Fact 2.79. [2, Theorem 5.5] *Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and C be a nonempty subset of $\mathbb{R}^n$. Suppose that $(x_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to C and every cluster point of $(x_k)_{k\in\mathbb{N}}$ belongs to C. Then $(x_k)_{k\in\mathbb{N}}$ converges to point in C.*

Combettes introduced quasi-Fejér monotonicity in 2001 [6]. A quasi-Fejér sequence is a sequence which satisfies the standard Fejér monotonicity property to within an additional error term.

Definition 2.80. Relative to a nonempty target set $S \in \mathcal{H}$, a sequence $(x_n)_{n\geq 0}$ in $\mathcal{H}$ is

(i) Quasi-Fejér of Type I if

$$(\exists(\epsilon_n)_{n\geq 0} \in l_+ \cap l^1)(\forall x \in S)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leq \|x_n - x\| + \epsilon_n. \tag{2.55}$$

(ii) Quasi-Fejér of Type II if

$$(\exists(\epsilon_n)_{n\geq 0} \in l_+ \cap l^1)(\forall x \in S)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + \epsilon_n. \tag{2.56}$$

(iii) Quasi-Fejér of Type III if

$$(\forall x \in S)(\exists(\epsilon_n)_{n\geq 0} \in l_+ \cap l^1)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + \epsilon_n. \tag{2.57}$$

l_+ denotes the set of all sequences in $[0, +\infty[$ and l^1 denotes the space of all absolutely summable sequences in $\mathbb{R}$.

Fact 2.81. [6, Theorem 3.8] *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type III relative to a nonempty set S in $\mathcal{H}$. Then $(x_n)_{n\geq 0}$ converges weakly to a point in set S if and only if every cluster point of $(x_n)_{n\geq 0}$ is in S.*

Chapter 3

CQ Algorithms for Convex Sets and Convex Functions

In this chapter, the CQ algorithm and relaxed CQ algorithm are reviewed, and a new proof for the convergence of relaxed CQ algorithm is given. The key tool is subgradient projector of a convex function.

3.1 The Split Feasibility Problem

The split feasibility problem (SFP) is the problem of finding a point $\hat{x}$ such that

$$\hat{x} \in C \qquad \text{and} \qquad A\hat{x} \in Q, \tag{3.1}$$

where C and Q are nonempty closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, and A is an $m \times n$ matrix. Many decision problems can be formulated as an SFP.

When $A = \mathrm{Id}$, the SFP becomes a set intersection problem: find $\hat{x} \in C \cap Q$.

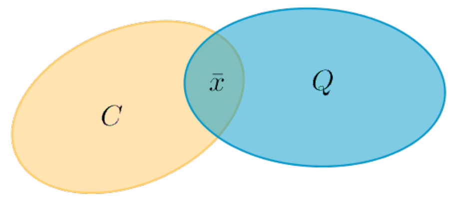

Figure 3.1: An example of the SFP.

3.2 The Proximity Function

Define a proximity function $f : \mathbb{R}^n \to \mathbb{R}_+$ by

$$x \mapsto f(x) := \frac{1}{2}\|P_Q(Ax) - Ax\|^2. \tag{3.2}$$

The function f is convex as a composition of a linear operator A and a convex function $\frac{1}{2}\mathrm{d}_Q^2$.

Lemma 3.1.

$$\min_{x \in C} f(x) = 0 \Longleftrightarrow x \in C \cap A^{-1}(Q). \tag{3.3}$$

Proof. "$\Rightarrow$": If $x \in C$, $f(x) = 0$, then $x \in C$, $P_Q(Ax) = Ax$. So $Ax \in Q$. Therefore, $x \in C \cap A^{-1}Q$.

"$\Leftarrow$": Suppose $x \in C \cap A^{-1}Q$. Then $x \in C$, $Ax \in Q$. So $P_Q(Ax) = Ax$, thus $f(x) = 0$. We have $\min_{x \in C} f(x) = 0$ because $f \geq 0$. ∎

Lemma 3.2. *Let f be defined as in* (3.2) *. Then*

(i) f *is convex and differentiable;*

(ii) $\nabla f(x) = A^*(\mathrm{Id} - P_Q)Ax, \quad \forall x \in \mathbb{R}^n$*;*

(iii) ∇f *is* $\|A\|^2$*-Lipschitz:* $\|\nabla f(x) - \nabla f(y)\| \leq \|A\|^2 \|x - y\|, \quad x, y \in \mathbb{R}^n$.

Proof. (i)&(ii): Let ι_Q be the indicator function defined on Q. The Moreau envelope of ι_Q at y is

$$\begin{aligned} e_1\iota_Q(y) &= \inf_z(\iota_Q(z) + \frac{1}{2}\|y-z\|^2) \\ &= \frac{1}{2}\inf_{z\in Q}\|y-z\|^2 \\ &= \frac{1}{2}\mathrm{d}_Q^2(y), \end{aligned}$$

where d_Q is the distance function to Q.

Since d_Q is convex, by Fact 2.40 we have d_Q^2 is convex. And by Fact 2.50 we have d_Q^2 is also differentiable, so $e_1\iota_Q$ is convex and differentiable. Now let $y = Ax$, then we have

$$e_1\iota_Q(Ax) = \frac{1}{2}\|(\mathrm{Id} - P_Q)(Ax)\|^2 = f(x).$$

Since A is linear, $e_1\iota_Q(Ax)$ is convex and differentiable. Thus, $f(x)$ is convex and differentiable.

By Fact 2.50, we have

$$\nabla e_1\iota_Q(Ax) = A^*(\mathrm{Id} - P_Q)Ax.$$

So $\nabla f(x) = A^*(\mathrm{Id} - P_Q)Ax$.

(iii): We have

$$\begin{aligned} &\|\nabla f(x) - \nabla f(y)\| \\ &= \|A^*(\mathrm{Id} - P_Q)Ax - A^*(\mathrm{Id} - P_Q)Ay\| \\ &\le \|A^*\| \cdot \|(\mathrm{Id} - P_Q)Ax - (\mathrm{Id} - P_Q)Ay\| \\ &\le \|A^*\| \cdot \|Ax - Ay\| \qquad (\mathrm{Id} - P_Q \text{ is firmly nonexpansive}) \end{aligned}$$

$$\leq \|A^*\| \cdot \|A\| \cdot \|x - y\|$$
$$\leq \|A\|^2 \|x - y\|. \qquad (\|A\| = \|A^*\|, \text{ see [8].})$$

■

Theorem 3.3. *Let f be defined as in* (3.2)*. Then*

$$\operatorname{argmin} f = \{x \in \mathcal{H}_1 : A^*(P_Q(Ax) - Ax) = 0\}. \tag{3.4}$$

Proof.

$$x \in \operatorname{argmin} f \Leftrightarrow \nabla f(x) = 0.$$

By Lemma 3.2(ii)

$$\nabla f(x) = A^*(\operatorname{Id} - P_Q)Ax,$$

so the proof is done. ■

3.3 Landweber Algorithms

3.3.1 Landweber Operator

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces, $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a nonzero bounded linear operator and $Q \subseteq \mathcal{H}_2$ be a nonempty closed convex subset.

Definition 3.4. The operator $T : \mathcal{H}_1 \to \mathcal{H}_1$ defined by the equality

$$Tx := x + \frac{1}{\|A\|^2} A^*(P_Q(Ax) - Ax) \tag{3.5}$$

is called the Landweber Operator.

The Landweber operator is closely related to a method for the problem: find $x \in \mathcal{H}_1$ such that $Ax \in Q$. The method was proposed by Landweber [9] for approximating the least-squares solution of a first-kind integral equation.

Note that $\|A\| \neq 0$, because A is a nonzero operator, consequently, the operator T is well-defined. We can also take $\lambda_{\max}(A^*A)$ or $\lambda_{\max}(AA^*)$ instead of $\|A\|^2$ in Definition 3.4 if A is a linear operator and $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite dimensional spaces.

Lemma 3.5. [5, Lemma 4.6.2] *Let $T : \mathcal{H}_1 \to \mathcal{H}$ be the Landweber operator defined by* (3.4) *and the corresponding proximity function $f : \mathcal{H}_1 \to \mathbb{R}_+$ be defined by* (3.2)*. Then*

$$\operatorname{Fix} T = \underset{x \in \mathcal{H}_1}{\operatorname{argmin}} f(x). \tag{3.6}$$

Proof. By Theorem 3.3 we have

$$\begin{aligned} & x = Tx \\ \Longleftrightarrow\ & \frac{1}{\|A\|^2} A^*(P_Q(Ax) - Ax) = 0 \\ \Longleftrightarrow\ & A^*(P_Q(Ax) - Ax) = 0 \\ \Longleftrightarrow\ & x \in \operatorname{argmin} f. \end{aligned}$$

■

Definition 3.6. Let $\lambda \in \mathbb{R}$. The operator $T_\lambda : \mathcal{H}_1 \to \mathcal{H}_1$ defined by the equality

$$T_\lambda x := x + \frac{\lambda}{\|A\|^2} A^*(P_Q(Ax) - Ax) \tag{3.7}$$

is called a λ-relaxation of the Landweber Operator.

Lemma 3.7. *Let $T : \mathcal{H}_1 \to \mathcal{H}_1$ be the Landweber operator and T_λ be a λ-relaxation of T. Then $T_\lambda = (1-\lambda)\operatorname{Id} + \lambda T$.*

Proof. Let $x \in \mathcal{H}_1$. Then

$$\begin{aligned}
&(1-\lambda)x + \lambda Tx \\
&= (1-\lambda)x + \lambda(x + \frac{1}{\|A\|^2} A^*(P_Q(Ax) - Ax)) \\
&= x + \frac{\lambda}{\|A\|^2} A^*(P_Q(Ax) - Ax) \\
&= T_\lambda x.
\end{aligned}$$

■

3.3.2 Landweber Operator for Linear Systems

Let $b \in \mathcal{H}_2$. If we take $Q := \{b\}$, then $\{x \in \mathcal{H}_1 : Ax \in Q\}$ is the solution set of the linear equation $Ax = b$. Of course, $P_Q(Ax) = b$, and the Landweber operator has the form

$$Tx = x + \frac{1}{\|A\|^2} A^*(b - Ax). \tag{3.8}$$

The proximity function $f : \mathcal{H}_1 \to \mathbb{R}_+$ for the linear equation $Ax = b$ by $f(x) := \frac{1}{2}\|Ax - b\|^2$. In this case, $x \in \operatorname{Fix} T \iff x$ is a least-squares solution of $Ax = b$.

Suppose that $\mathcal{H}_1 = \mathbb{R}^n$ and $\mathcal{H}_2 = \mathbb{R}^m$ with the standard inner product, and that $b \in \mathbb{R}^m$ with standard inner product. If we take $Q := \{u \in \mathbb{R}^m : u \leq b\}$ (here $u \leq b$ means $u_i \leq b_i$ for $i = 1, \ldots, m$), then $\{x \in \mathcal{H}_1 : Ax \in Q\}$ is the solution set of a system of linear inequalities $Ax \leq b$, where A is an $m \times n$ matrix, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. Hence, the SFP covers many interesting problems.

3.3.3 Projected Landweber Operator

Definition 3.8. An operator $U : \mathcal{H}_1 \to \mathcal{H}_1$ defined by

$$U := P_C \left(\mathrm{Id} + \frac{1}{\|A\|^2} A^*(P_Q - \mathrm{Id})A \right) \tag{3.9}$$

is called the projected Landweber operator or an oblique projection.

Proposition 3.9. [5, Proposition 4.7.2] *Let* $R_\lambda x := P_C T_\lambda$, *where* $\lambda > 0$,

$$T_\lambda := \mathrm{Id} + \frac{\lambda}{\|A\|^2} A^*(P_Q - \mathrm{Id})A$$

be a projected relaxation of the Landweber operator T *defined by* (3.5), *i.e.,*

$$R_\lambda x := P_C \left(x + \frac{\lambda}{\|A\|^2} A^*(P_Q - \mathrm{Id})Ax \right), \tag{3.10}$$

and $f : C \to \mathbb{R}_+$ *be the proximity function defined by* (3.2)*. Then*

$$\mathrm{Fix}\, R_\lambda = \underset{x \in C}{\mathrm{argmin}} f(x). \tag{3.11}$$

3.3.4 Landweber Method and Projected Landweber Method

Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a nonzero linear bounded operator and $Q \subseteq \mathcal{H}_2$ be a nonempty closed convex subset. The *Landweber method* (**LM**) for finding $x \in \mathcal{H}_1$ satisfying $Ax \in Q$ (if such an x exists) is defined by the recurrence

$$x_{k+1} = x_k + \frac{\lambda_k}{\|A\|^2} A^*(P_Q(Ax_k) - Ax_k), \tag{3.12}$$

where $x_0 \in \mathcal{H}_1$ is arbitrary and $\lambda_k \in [0, 2]$. We can also write $x_{k+1} = T_{\lambda_k} x_k$, where T_λ is a $\lambda - relaxation$ of the Landweber operator $T := \mathrm{Id} + \frac{1}{\|A\|^2} A^*(P_Q - \mathrm{Id})A$. Landweber proposed the method with $\lambda_k \in]0, 2[$, $k \geq 0$.

The *projected Landweber method* (**PLM**) is an iterative method for solving the SFP. The iterative step of the projected Landweber method has the form

$$x_{k+1} = P_C\left(x_k + \frac{\lambda_k}{\|A\|^2} A^*(P_Q(Ax_k) - Ax_k)\right), \tag{3.13}$$

where $\lambda_k \in [0, 2]$. We can also write $x_{k+1} = P_C T_{\lambda_k} x_k$, where T_{λ_k} is a λ_k-relaxation of the Landweber operator T, or $x_{k+1} = R_{\lambda_k} x_k$ where $R_{\lambda_k} := P_C T_{\lambda_k}$ is the projected λ_k-relaxation of the Landweber operator. It is clear that if $C = \mathcal{H}_1$, then the projected Landweber method reduces to the Landweber method.

The projected Landweber method was studied by Byrne for $\mathcal{H}_1$ and $\mathcal{H}_2$ Euclidean spaces, $\lambda_k \in]0, 2[$, $k \geq 0$, and he called the method a ***CQ algorithm*** and proved the convergence of sequences generated by the method to a solution of the split feasibility problem, if such a solution exists [3].

3.4 CQ Algorithms

3.4.1 Byrne's CQ Algorithm

Assume that the solution set S of the SFP is nonempty. Byrne [3, 4] first proposed this so-called CQ algorithm which generates a sequence (x_k) by the recursive procedure,

$$x_{k+1} = P_C(x_k - \tau_k A^*(\mathrm{Id} - P_Q)Ax_k), \tag{3.14}$$

where the stepsize τ_k is chosen in the interval $(0, 2/\|A\|^2)$, A^* is the transpose of A, P_C and P_Q are the orthogonal projections onto C and Q, respectively. The algorithm can be summarized as follows.

Algorithm 1: *CQ* Algorithm

Choose arbitrary x_0;

while $\|x_{k+1} - x_k\| \neq 0$ **do**

  Set $x_k = x_{k+1}$;

  Choose τ_k from $(0, 2/\|A\|^2)$;

  Set $x_{k+1} = P_C(x_k - \tau_k A^*(\mathrm{Id} - P_Q)Ax_k)$;

end

In implementation, we can stop the algorithm when $\|x_{k+1} - x_k\| \leq \epsilon$, where $\epsilon > 0$ is a given tolerance. Byrne also proved the convergence of this *CQ* algorithm in his paper.

Theorem 3.10. [3, Corollary 2.1] *The sequence* (x_k) *defined by* (3.14) *converges to a solution of the SFP, whenever such solution exists.*

The *CQ* algorithm is a special case of the gradient-projection algorithm (GPA) by Lemma 3.2(ii). Let

$$f(x) := \frac{1}{2}\|(\mathrm{Id} - P_Q)Ax\|^2, \tag{3.15}$$

and consider the convex minimization problem

$$\min_{x \in C} f(x). \tag{3.16}$$

Recall that the GPA for the minimization problem (3.16) is

$$x_{k+1} = P_C(x_k - \tau_k \nabla f(x_k)), \tag{3.17}$$

where τ_k is chosen in the interval $(0, 2/L)$, and L is the Lipschitz constant of ∇f. By Lemma 3.2(iii), ∇f has $\|A\|^2$ Lipschitz constant. In (3.14), $0 < \tau_k < \frac{2}{\|A\|^2} = \frac{2}{L}$, so Byne's algorithm is a GPA.

The stepsize in the CQ algorithm (3.14) and GPA algorithm (3.17) depends on the operator (matrix) norm $\|A\|$. However, to compute or even estimate the matrix norm of $\|A\|$ is in general not an easy task in practice.

To overcome this difficulty, Byrne [3] considered using a method to estimate the matrix norm. However, the condition on the method was restrictive. Another way is to construct a different stepsize that does not require the matrix norm. This is given in the following section.

3.4.2 Yang's Improvement on Byrne's CQ Algorithm

Yang [15] proposed the following stepsize:

$$\tau_k = \frac{\rho_k}{\|\nabla f(x_k)\|}, \tag{3.18}$$

where (ρ_k) is a sequence of positive real numbers such that

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty. \tag{3.19}$$

In this way, he constructed a different stepsize that does not require the matrix norm, in order to avoid the difficulty of computing it. Yang proved the convergence of the GPA (3.17), where the sequence (τ_k) of stepsizes satisfies (3.18) and (3.19) with two more conditions satisfied:

(i) Q is a bounded subset;

(ii) A is a matrix with a full column rank.

Theorem 3.11. [15, Theorem 2.1] *Suppose that conditions (i) and (ii) above are satisfied. Assume sequence* (x_k) *is produced by* (3.17) *with* (3.18) *and* (3.19)*, and* (x_k) *is bounded. Then any accumulation point of* (x_k) *is a solution of the SFP. Furthermore, if the SFP has no solution at which* $f(x)$ *vanishes, then the solution of the SFP is unique and* (x_k) *converges to this solution.*

However, these conditions are also restrictive, and some important cases may be excluded, like the least-squares problem.

3.4.3 López, Martín-Márquez, Wang and Xu's Method

Based on Yang's idea, López, Martín-Márquez, Wang and Xu [10] constructed another step size that removed the conditions:

$$\tau_k := \frac{\rho_k f(x_k)}{\|\nabla f(x_k)\|^2}, \tag{3.20}$$

where $0 < \rho_k < 4$. Then they proved the convergence of the CQ algorithm (3.14) with stepsize (3.20).

Theorem 3.12. [10, Theorem 3.5] *Assume that* $\inf_k \rho_k(4 - \rho_k) > 0$*. Then the sequence* (x_k) *generated by* (3.14) *with stepsize* (3.20) *converges weakly to a solution* x^* *of the SFP.*

3.5 Relaxed CQ Algorithms

López, Martín-Márquez, Wang and Xu [10] proposed a relaxed CQ algorithm. Instead of considering general closed convex subsets C and Q, they considered a relaxed CQ algorithm in which the closed convex subsets C and

Q are of the particular structure, i.e., level sets of convex functions given as follows:

$$C = \{x \in \mathbb{R}^n : c(x) \leq 0\} \quad \text{and} \quad Q = \{y \in \mathbb{R}^m : q(y) \leq 0\}, \tag{3.21}$$

where $c : \mathbb{R}^n \to \mathbb{R}$ and $q : \mathbb{R}^m \to \mathbb{R}$ are convex subdifferentiable functions.

Define

$$C_k = \{x \in \mathbb{R}^n : c(x_k) + \langle \xi_k, x - x_k \rangle \leq 0\}, \tag{3.22}$$

where $\xi_k \in \partial c(x_k)$, and

$$Q_k = \{y \in \mathbb{R}^m : q(Ax_k) + \langle \zeta_k, y - Ax_k \rangle \leq 0\}, \tag{3.23}$$

where $\zeta_k \in \partial q(Ax_k)$.

Then we define

$$f_k(x) = \frac{1}{2}\|(\mathrm{Id} - P_{Q_k})Ax\|^2, \quad k \geq 0, \tag{3.24}$$

where Q_k is given in (3.23). We then have

$$\nabla f_k(x) = A^*(\mathrm{Id} - P_{Q_k})Ax. \tag{3.25}$$

Then we define the relaxed CQ algorithm by the recursive procedure,

$$x_{k+1} = P_{C_k}(x_k - \tau_k \nabla f_k(x_k)), \tag{3.26}$$

where

$$\tau_k = \frac{\rho_k f_k(x_k)}{\|\nabla f_k(x_k)\|^2}, \quad 0 < \rho_k < 4. \tag{3.27}$$

Algorithm 2: Relaxed CQ Algorithm

Choose arbitrary x_0;

while $\|\nabla f_k(x_k)\| \neq 0$ **do**

  Set $x_k = x_{k+1}$;

  Set $\tau_k = \frac{\rho_k f_k(x_k)}{\|\nabla f_k(x_k)\|^2}, \quad 0 < \rho_k < 4$;

  Set $x_{k+1} = P_{C_k}(x_k - \tau_k \nabla f_k(x_k))$;

end

Then they proved the convergence of this relaxed CQ algorithm.

Theorem 3.13. [10, Theorem 4.3] *Assume that* $\inf_k \rho_k(4 - \rho_k) > 0$ *and the solution set* $S \neq \emptyset$*. Then the sequence* (x_k) *generated by algorithm 2 converges to a solution* x^* *of the SFP* (3.1)*, where* $x^* = \lim_{k\to\infty} P_S x_k$*.*

3.6 Reformulate CQ Algorithm with Subgradient Projectors

In last section, we can see the key point of the relaxed CQ algorithm is using projection onto a halfspace that contains the set instead of the projection onto the set itself. There is one operator called subgradient projector doing the same job. We can use the subgradient projector help us to find projection onto C_k and Q_k given by (3.22) and (3.23). First let us see the definition of subgradient projector.

3.6.1 Subgradient Projector

Definition 3.14. Let $f : \mathcal{H} \to \mathbb{R}$ be a continuous convex function, let $\xi \in \mathbb{R}$ be such that $C = \text{lev}_{\leq \xi} f \neq \emptyset$, and let s be a selection of ∂f. The subgradient projector onto C associated with (f,ξ,s) is

$$G:\mathcal{H}\to\mathcal{H}:x\mapsto\begin{cases}x+\frac{\xi-f(x)}{\|s(x)\|^2}s(x), & \text{if } f(x)>\xi\\ x, & \text{if } f(x)\leq\xi.\end{cases}\tag{3.28}$$

Note that $s(x)\neq 0$ because $C=\operatorname{lev}_{\leq\xi} f\neq\emptyset$. If f is Gâteaux differentiable on $\mathcal{H}\setminus C$, then the subgradient projector associated with (f,ξ) is

$$G:\mathcal{H}\to\mathcal{H}:x\mapsto\begin{cases}x+\frac{\xi-f(x)}{\|\nabla f(x)\|^2}\nabla f(x), & \text{if } f(x)>\xi\\ x, & \text{if } f(x)\leq\xi.\end{cases}\tag{3.29}$$

The subgradient projector has the properties below. They are due to a lecture of Bauschke, which will be given in the second edition of [2]. See also Cegielski's book [5].

Proposition 3.15. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a continuous convex function, let $\xi\in\mathbb{R}$ be such that $C=\operatorname{lev}_{\leq\xi} f\neq\emptyset$, and let s be a selection of ∂f. Let G be the subgradient projector onto C associated with (f,ξ,s), let $x\in\mathbb{R}^n$, let $z\in C$, and set*

$$H=\{y\in\mathbb{R}^n|\langle y-x,s(x)\rangle+f(x)\leq\xi\}.\tag{3.30}$$

Then the following hold:

(i) $\operatorname{Fix} G=C\subset H$.

(ii) $Gx=P_H x$.

(iii) $\langle z-Gx,x-Gx\rangle\leq 0$.

(iv) $\max\{f(x)-\xi,0\}=\|s(x)\|\|Gx-x\|$.

(v) *Let* $(x_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}^n$ *and let* $\bar{x}$ *be a point in* $\mathbb{R}^n$ *such that* $Gx_n - x_n \to 0$ *and* $x_n \to \bar{x}$. *Then* $\bar{x} \in C$.

(vi) G *is continuous on* C.

(vii) *Suppose that* f *is Fréchet differentiable on* $\mathbb{R}^n \setminus C$. *Then* G *is continuous.*

(viii) *Let* $\lambda \in \,]0,2]$ *and set* $R = \mathrm{Id} + \lambda(G - \mathrm{Id})$. *Then*

$$\|Rx - z\|^2 \le \|x - z\|^2 - \lambda(2-\lambda)\|Gx - x\|^2. \tag{3.31}$$

(ix) $\|Gx - z\|^2 \le \|x - z\|^2 - \|Gx - x\|^2$.

(x) $\operatorname{Fix} R = \operatorname{Fix} G = C$;

(xi) G *and* R *are strictly quasinonexpansive.*

(xii) $2G - \mathrm{Id}$ *is quasinonexpansive.*

Proof. (i): By (3.28), it is obvious that $\operatorname{Fix} G = C$. Since $s(x) \in \partial f(x)$, by (2.33) we have $\langle z - x, s(x)\rangle + f(x) \le f(z) \le \xi$. Hence $z \in H$.

(ii): By (2.31) we have

$$\begin{aligned} P_H x &= x + \frac{(\xi - f(x) + \langle x, s(x)\rangle) - \langle s(x), x\rangle}{\|s(x)\|^2} s(x) \\ &= x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x) \\ &= Gx. \end{aligned}$$

(iii): We deduce from (i) that $z \in H$. In turn, (ii) and (2.29) imply that $\langle z - Gx, x - Gx\rangle = \langle z - P_H x, x - P_H x\rangle \le 0$.

(iv): **Case 1**: $x \in C$. Then $f(x) - \xi \le 0$ and $Gx = x$. So

$$\max\{f(x) - \xi, 0\} = 0$$

and

$$\|s(x)\|\|Gx - x\| = 0.$$

Hence $\max\{f(x) - \xi, 0\} = \|s(x)\|\|Gx - x\|$.

Case 2: $x \notin C$. Then $f(x) > \xi$. Hence $\max\{f(x) - \xi, 0\} = f(x) - \xi$.

$$\begin{aligned}
\|s(x)\|\|Gx - x\| &= \|s(x)\| \left\| (x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x)) - x \right\| \\
&= \left\| \frac{\xi - f(x)}{\|s(x)\|} s(x) \right\| \\
&= \frac{f(x) - \xi}{\|s(x)\|} \|s(x)\| \\
&= f(x) - \xi.
\end{aligned}$$

So $\max\{f(x) - \xi, 0\} = \|s(x)\|\|Gx - x\|$.

(v): By Fact 2.66(iii), there exists $\rho > 0$ such that $(x_k)_{k\in\mathbb{N}}$ lies in $B(\bar{x}; \rho)$ and $\sigma = \sup \|\partial f(B(\bar{x}; \rho))\| < +\infty$. Hence we have

$$(\forall n \in \mathbb{N}) \quad \|s(x_n)\| \le \sigma. \tag{3.32}$$

Hence, by (iv),

$$\max\{f(\bar{x}) - \xi, 0\} \le \lim \max\{f(x_k) - \xi, 0\} \le \sigma \lim \|Gx_k - x_k\| = 0. \tag{3.33}$$

Thus, $f(\bar{x}) \le \xi$.

(vi): Let $\bar{x} \in C$ and let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ converging to $\bar{x}$. If

$(x_{k_l})_{l\in\mathbb{N}}$ is a subsequence of $(x_k)_{k\in\mathbb{N}}$ lying in C, then by (i)

$$Gx_{k_l} = x_{k_l} \to \bar{x} = G\bar{x}.$$

Thus, we suppose that $(x_k)_{k\in\mathbb{N}}$ lies in $\mathbb{R}^n\backslash C$. Then $(\forall k \in \mathbb{N})$

$$\begin{aligned}&f(x_k) - \xi\\ \leq &f(\bar{x}) - \xi - \langle \bar{x} - x_k, s(x_k)\rangle\\ \leq &\|x_k - \bar{x}\|\|s(x_k)\|.\end{aligned}$$

Hence

$$0 \leq \|Gx_k - x_k\| = \frac{f(x_k) - \xi}{\|s(x_k)\|} \leq \|x_k - \bar{x}\| \to 0.$$

Hence, $Gx_k \to \bar{x} = G\bar{x}$.

(vii): By the assumption, we have $s(x) = \nabla f(x)$. Take $\bar{x} \in \mathcal{H}$. If $\bar{x} \in C$, by (vi) done.

Consider $\bar{x} \in \mathcal{H}\backslash C$. Then near $\bar{x}$ we have

$$Gx = x + \frac{\xi - f(x)}{\|\nabla f(x)\|^2}\nabla f(x).$$

Since f is convex, ∇f is continuous. Moreover $Gx = x$ for $x \in C$. So G is continuous.

(viii): By(iii)

$$\begin{aligned}\|Rx - z\|^2 &= \|x - z + \lambda(Gx - x)\|^2\\ &= \|x - z\|^2 + 2\lambda\langle x - z, Gx - x\rangle + \lambda^2\|Gx - x\|^2.\end{aligned}$$

By (ii), $Gx = P_H x$. Since $P_H z = z$,

$$\begin{aligned}
&\|x - z\|^2 + 2\lambda\langle x - z, Gx - x\rangle + \lambda^2\|Gx - x\|^2 \\
&= \|x - z\|^2 + 2\lambda\langle x - z, P_H x - x\rangle + \lambda^2\|Gx - x\|^2 \\
&= \|x - z\|^2 - 2\lambda\langle x - z, x - P_H x\rangle + \lambda^2\|Gx - x\|^2 \\
&= \|x - z\|^2 + 2\lambda\langle Gx - z, Gx - x\rangle - \lambda(2 - \lambda)\|Gx - x\|^2 \\
&\leq \|x - z\|^2 - \lambda(2 - \lambda)\|Gx - x\|^2.
\end{aligned}$$

(ix): Set $\lambda = 1$ in (3.31). Then by (viii) the proof is done.

(x): Since $R = \mathrm{Id} + \lambda(G - \mathrm{Id})$. So $\forall x \in \mathbb{R}^n$

$$Rx = x \Longleftrightarrow \lambda(G - \mathrm{Id})x = 0 \Longleftrightarrow Gx = x.$$

So by (i), $\operatorname{Fix} R = \operatorname{Fix} G = C$.

(xi): By (viii) and (ix) we have R and G are strictly quasinonexpansive.

(xii): Apply (viii) with $\lambda = 2$. ■

The subgradient projectors generalize the classical projectors. See the following example.

Example 3.16. Let C be a nonempty closed convex subset of $\mathbb{R}^n$. Denote the subgradient projector onto C associated with $(\mathrm{d}_C, 0)$ by G. Then $G = P_C$.

Proof. If $x \in C$, then $\mathrm{d}_C(x) = 0$, so $Gx = x = P_C x$.

If $x \notin C$, then using $\nabla \mathrm{d}_C(x) = \frac{(\mathrm{Id} - P_C)x}{\mathrm{d}_C(x)} = \frac{x - P_C x}{\mathrm{d}_C(x)}$,

$$Gx = x - \frac{\mathrm{d}_C(x)}{\|\nabla \mathrm{d}_C(x)\|^2} \nabla \mathrm{d}_C(x)$$

$$= x - \frac{\mathrm{d}_C(x)}{\|\frac{(\mathrm{Id} - P_C)x}{\mathrm{d}_C(x)}\|^2} \frac{(\mathrm{Id} - P_C)x}{\mathrm{d}_C(x)}$$

$$= x - \frac{\mathrm{d}_C^2(x)}{\|(\mathrm{Id} - P_C)(x)\|^2}(\mathrm{Id} - P_C)x$$

$$= x - (\mathrm{Id} - P_C)x = P_C x.$$

■

3.6.2 Subgradient Projection Algorithm

Now we introduce the subgradient projection algorithm (SPA), which is constructed by subgradients projectors.

Let

$$C = \{x \in \mathbb{R}^n : c(x) \leq 0\}, \tag{3.34}$$

$$Q = \{y \in \mathbb{R}^n : q(y) \leq 0\}, \tag{3.35}$$

where $c : \mathbb{R}^n \to \mathbb{R}$ and $q : \mathbb{R}^n \to \mathbb{R}$ are both convex and continuous.

Define the subgradient projectors:

$$G_c : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{c(x)}{\|s_c(x)\|^2} s_c(x), & \text{if } c(x) > 0 \\ x, & \text{if } c(x) \leq 0, \end{cases} \tag{3.36}$$

and

$$G_q : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto \begin{cases} x - \frac{q(x)}{\|s_q(x)\|^2} s_q(x), & \text{if } q(x) > 0 \\ x, & \text{if } q(x) \leq 0. \end{cases} \tag{3.37}$$

$$R_{\lambda q} = \mathrm{Id} + \lambda(G_q - \mathrm{Id}), \tag{3.38}$$

where $\lambda \in \,]0,2]$, $s_c(x)$ is a selection of $\partial c(x)$ and s_q is a selection of $\partial q(x)$.

Suppose $C \cap Q \neq \emptyset$, then we propose the Subgradient Projection Algorithm (SPA) given by the iteration below:

$$x_{k+1} = G_c(R_{\lambda_k q} x_k). \tag{3.39}$$

See the algorithm below.

Algorithm 3: The Subgradient Projection Algorithm

Choose an arbitrary x_0;
while $\|x_{k+1} - x_k\| \neq 0$ **do**
  Set $x_k = x_{k+1}$;
  Choose λ_k from $(0,2)$;
  Set $x_{k+1} = G_c(R_{\lambda_k q} x_k)$;
end

The operator G_c is the subgradient projector of c, and $R_{\lambda_k q}$ is a relaxation of subgradient projector G_q. The convergence of Algorithm 3 is shown by the following theorem.

Theorem 3.17. *Assume* $C \cap Q \neq \emptyset$ *and* $\inf_k \lambda_k(2-\lambda_k) > 0$. *The sequence* (x_k) *generated by* (3.39) *converges to a solution* x^* *such that* $x^* \in C \cap Q$.

Proof. Let $z \in C \cap Q$. By (3.39), we have

$$\begin{aligned}
&\|x_{k+1} - z\|^2 \\
&= \|G_c(R_{\lambda_k q} x_k) - z\|^2 \\
&\leq \|R_{\lambda_k q}(x_k) - z\|^2 - \|G_c(R_{\lambda_k q} x_k) - R_{\lambda_k q} x_k\|^2 \quad \text{(By Proposition 3.15(ix))} \\
&\leq \|x_k - z\|^2 - \lambda_k(2-\lambda_k)\|G_q x_k - x_k\|^2
\end{aligned}$$

$$- \|G_c(R_{\lambda_k q} x_k) - R_{\lambda_k q} x_k\|^2 \quad \text{(By Proposition 3.15(viii))} \tag{3.40}$$

So we have Fejér monotonicity

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2, \quad \forall k \in \mathbb{N}. \tag{3.41}$$

Hence (x_k) is Fejér monotone. So by Fact 2.79 we just need to prove every cluster point of (x_k) is in the solution set. Taking the summation on both sides of (3.40) gives us

$$\begin{aligned}
&\|x_{n+1} - z\|^2 \\
&\leq \|x_0 - z\|^2 - \sum_{k=0}^{n} \lambda_k(2 - \lambda_k)\|G_q x_k - x_k\|^2 \\
&- \sum_{k=0}^{n} \|G_c(R_{\lambda_k q} x_k) - R_{\lambda_k q} x_k\|^2.
\end{aligned}$$

So we have

$$\begin{aligned}
&\sum_{k=0}^{n} [\lambda_k(2 - \lambda_k)\|G_q x_k - x_k\|^2 + \|G_c(R_{\lambda_k q} x_k) - R_{\lambda_k q} x_k\|^2] \\
&\leq \|x_0 - z\|^2 - \|x_{n+1} - z\|^2.
\end{aligned} \tag{3.42}$$

Thus

$$\sum_{k=0}^{n} \lambda_k(2 - \lambda_k)\|G_q x_k - x_k\|^2 \leq \|x_0 - z\|^2. \tag{3.43}$$

Then let $n \to \infty$ we have,

$$\sum_{k=0}^{\infty} \lambda_k(2-\lambda_k)\|G_q x_k - x_k\|^2 \le \|x_0 - z\|^2 < +\infty. \tag{3.44}$$

Let $y_k = R_{\lambda_k q} x_k$. By (3.42) we have

$$\sum_{k=0}^{n} \|G_c(y_k) - y_k\|^2 \le \|x_0 - z\|^2. \tag{3.45}$$

Then let $n \to \infty$, we have

$$\sum_{k=0}^{\infty} \|G_c(y_k) - y_k\|^2 \le \|x_0 - z\|^2 < +\infty. \tag{3.46}$$

Since $\inf_{k\ge 0} \lambda_k(2-\lambda_k) > 0$, $\exists \epsilon > 0$ such that $\lambda_k(2-\lambda_k) > \epsilon \ \forall k \ge 1$. By (3.44) we have

$$\epsilon \cdot \sum_{k=0}^{\infty} \|G_q x_k - x_k\|^2 < +\infty, \tag{3.47}$$

thus $\|G_q x_k - x_k\| \to 0$.

Because $(x_k)_{k\in\mathbb{N}}$ is Fejér monotone, so it is bounded. Now let $\bar{x}$ be a cluster point of (x_k) and (x_{k_l}) be a subsequence of (x_k), $(x_{k_l}) \to \bar{x}$. Then by Proposition 3.15(v) we have $\bar{x} \in Q$. Since $y_{k_l} = R_{\lambda_{k_l} q} x_{k_l} = x_{k_l} + \lambda_{k_l}(G_q x_{k_l} - x_{k_l})$, so

$$\|y_{k_l} - x_{k_l}\| = \lambda_{k_l}\|G_q x_{k_l} - x_{k_l}\| \to 0. \tag{3.48}$$

As $x_{k_l} \to \bar{x}$, $y_{k_l} \to \bar{x}$. By (3.46) we have

$$\|G_c y_{k_l} - y_{k_l}\| \to 0.$$

Again by Proposition 3.15(v) we have $\bar{x} \in C$. Thus $\bar{x} \in C \cap Q$. The proof is done. ■

Corollary 3.18. *Assume $C \cap Q \neq \emptyset$. For every initial point $x_0 \in \mathbb{R}^n$, the sequence given by*

$$x_{k+1} = G_c G_q(x_k) \quad \forall k \geq 1 \tag{3.49}$$

converges to $x^* \in C \cap Q$.

Proof. Apply Theorem 3.17 with $\lambda_k = 1$ for all $k \geq 1$. ■

Lemma 3.19. *Let*

$$\hat{Q} = \{y \in \mathbb{R}^m : \hat{q}(y) \leq 0\},$$

and

$$Q = \{x \in \mathbb{R}^n : q(x) \leq 0\},$$

where $\hat{q} : \mathbb{R}^n \to \mathbb{R}$ is convex, A is an $m \times n$ matrix, and $q = \hat{q} \circ A$. Then

$$Ax \in \hat{Q} \Longleftrightarrow x \in Q. \tag{3.50}$$

Proof.

$$\begin{aligned} & x \in Q \\ \Longleftrightarrow\ & q(x) \leq 0 \\ \Longleftrightarrow\ & \hat{q} \circ A(x) \leq 0 \\ \Longleftrightarrow\ & \hat{q}(Ax) \leq 0 \\ \Longleftrightarrow\ & Ax \in \hat{Q}. \end{aligned}$$

■

Corollary 3.20. *Let C and Q be defined as in* (3.34) *and* (3.35)*, and*

$$\hat{Q} = \{y \in \mathbb{R}^m : \hat{q}(y) \le 0\}, \tag{3.51}$$

where $q = \hat{q} \circ A$, $A : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix. Assume $C \cap Q \neq \emptyset$ and $\inf_k \lambda_k(2 - \lambda_k) > 0$. The sequence (x_k) generated by (3.39) *converges to a point x^* such that $x^* \in C$ and $Ax^* \in \hat{Q}$.*

Proof. By Theorem 3.17 we know the sequence converges to a point $x^* \in C \cap Q$. So $x^* \in Q$, thus by Lemma 3.19 we have $Ax^* \in \hat{Q}$. The proof is done. ■

By Corollary 3.20, we can see the SPA can solve the SFP. Actually, we can deduce the CQ algorithm from the SPA. See the following corollary.

Corollary 3.21. *The sequence (x_k) defined by* (3.14)*, where*

$$\tau_k = \rho_k \frac{f(x_k)}{\|\nabla f(x_k)\|^2},$$

$$f(x) = \frac{1}{2}\|(\mathrm{Id} - P_Q)Ax\|^2,$$

and $0 < \rho_k < 2$, converges to a solution of the SFP, whenever such solution exists.

Proof. Denote the subgradient projector associated with $(\mathrm{d}_C, 0)$ by G_{d_C}. Denote the subgradient projector associated with $(f, 0)$ by G_f. Let

$$R_{\rho_k f} = \mathrm{Id} + \rho_k(G_f - \mathrm{Id}).$$

Then by (3.14), (3.20) and Example 3.16,

$$\begin{aligned} x_{k+1} &= P_C(x_k - \tau_k A^*(\mathrm{Id} - P_Q)Ax_k) \\ &= P_C(x_k - \rho_k \frac{f(x_k)}{\|\nabla f(x_k)\|^2} \nabla f(x_k)) \\ &= G_{\mathrm{d}_C}(R_{\rho_k f}(x_k)). \end{aligned}$$

Then by Theorem 3.17, the sequence $(x_k)_{k\in\mathbb{N}}$ converges to a point x^* such that $x \in \mathrm{lev}_{\leq 0}\, \mathrm{d}_C \cap \mathrm{lev}_{\leq 0}\, f$. Thus, $x^* \in C$ and $\frac{1}{2}\|(\mathrm{Id} - P_Q)Ax^*\|^2 \leq 0$. Hence, $Ax^* = P_Q Ax^*$, i.e., $Ax^* \in Q$. So x^* is in the solution set of the SFP.

■

In the CQ algorithm we use $0 < \rho_k < 4$. But in Corollary 3.21 we assume $0 < \rho_k < 2$ because the SPA works on more general cases and requires more restrictive conditions.

3.6.3 Reformulate the Relaxed CQ Algorithm with Subgradient Projectors

Similar as what we did in Corollary 3.21, we are able to reformulate the relaxed CQ algorithm into a subgradient projection algorithm.

Let

$$C = \{x \in \mathbb{R}^n : c(x) \leq 0\} \quad \text{and} \quad Q = \{y \in \mathbb{R}^m : q(y) \leq 0\}, \tag{3.52}$$

where $c : \mathbb{R}^n \to \mathbb{R}$ and $q : \mathbb{R}^m \to \mathbb{R}$ are convex subdifferentiable functions. Define

$$C_k = \{x \in \mathbb{R}^n : c(x_k) + \langle \xi_k, x - x_k \rangle \leq 0\}, \tag{3.53}$$

where $\xi_k \in \partial c(x_k)$, and

$$Q_k = \{y \in \mathbb{R}^m : q(Ax_k) + \langle \zeta_k, y - Ax_k \rangle \le 0\}, \tag{3.54}$$

where $\zeta_k \in \partial q(Ax_k)$. Then we define

$$f_k(x) = \frac{1}{2}\|(\operatorname{Id} - P_{Q_k})Ax\|^2, \quad k \ge 0, \tag{3.55}$$

where Q_k is given in (3.54). We then have

$$\nabla f_k(x) = A^*(\operatorname{Id} - P_{Q_k})Ax. \tag{3.56}$$

Define

$$\tau_k = \rho_k \frac{f_k(x_k)}{\nabla f_k(x_k)}, \quad 0 < \rho_k < 2. \tag{3.57}$$

Denote the subgradient projector associated with $(c, 0)$ by G_c. Denote the subgradient projector associated with $(q, 0)$ by G_q. Denote the subgradient projector associated with $(f_k, 0)$ by G_{f_k}. Let

$$R_{\rho_k f_k} = \operatorname{Id} + \rho_k(G_{f_k} - \operatorname{Id}). \tag{3.58}$$

Then we have $P_{C_k} = G_c$, $P_{Q_k} = G_q$.

Lemma 3.22. *$P_{C_k} = G_c$, and $P_{Q_k} = G_q$.*

Proof. It is obvious by Proposition 3.15(ii). ■

Thus by (3.26)

$$\begin{aligned} x_{k+1} &= P_{C_k}(x_k - \tau_k A^*(\operatorname{Id} - P_{Q_k})Ax_k) \\ &= G_c(x_k - \rho_k \frac{f_k(x_k)}{\nabla f_k(x_k)} \nabla f_k(x_k)) \end{aligned}$$

$$= G_c(R_{\rho_k f_k}(x_k)).$$

Notice that we use $0 < \rho_k < 2$. Then let us see the following theorem.

Theorem 3.23. *Assume that* $\inf_{k\in\mathbb{N}} \rho_k(2-\rho_k) > 0$*, and* $c : \mathbb{R}^n \to \mathbb{R}$*,* $q : \mathbb{R}^n \to \mathbb{R}$ *are continuous convex functions. Then the sequence* $(x_k)_{k\in\mathbb{N}}$ *generated by the iteration*

$$x_{k+1} = G_c(R_{\rho_k f_k}(x_k)) \tag{3.59}$$

converges to a point x^* *such that* $x^* \in C$ *and* $Ax^* \in Q$*.*

Proof. Let z be any point in the solution set. That is, $z \in C$ and $Az \in Q$. Hence,

$$q(Ax_k) + \langle \zeta_k, Az - Ax_k\rangle \leq q(Az) \leq 0.$$

Thus, $Az \in Q_k$ and $f_k(z) = 0$. So $z \in \operatorname{lev}_{\leq 0} f_k$. Then we have

$$\begin{aligned}
&\|x_{k+1} - z\|^2 \\
&= \|G_c(R_{\rho_k f_k} x_k) - z\|^2 \\
&\leq \|R_{\rho_k f_k}(x_k) - z\|^2 - \|G_c(R_{\rho_k f_k} x_k) - R_{\rho_k f_k} x_k\|^2 \quad \text{(By Proposition 3.15(ix))} \\
&\leq \|x_k - z\|^2 - \rho_k(2-\rho_k)\|G_{f_k} x_k - x_k\|^2 \\
&- \|G_c(R_{\rho_k q} x_k) - R_{\rho_k q} x_k\|^2 \quad \text{(By Proposition 3.15(viii))}
\end{aligned} \tag{3.60}$$

So we have Fejér monotonicity

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2, \quad \forall k \in \mathbb{N}. \tag{3.61}$$

Hence (x_k) is Fejér monotone. So by Fact 2.79 we just need to prove every cluster point of (x_k) is in the solution set. Taking the summation on both side of (3.60) gives us

$$\begin{aligned}&\|x_{n+1} - z\|^2 \\ &\leq \|x_0 - z\|^2 - \sum_{k=0}^{n} \rho_k(2-\rho_k)\|G_{f_k}x_k - x_k\|^2 \\ &- \sum_{k=0}^{n} \|G_c(R_{\rho_k f_k}x_k) - R_{\rho_k f_k}x_k\|^2.\end{aligned}$$

So we have

$$\begin{aligned}&\sum_{k=0}^{n}[\rho_k(2-\rho_k)\|G_{f_k}x_k - x_k\|^2 + \|G_c(R_{\rho_k f_k}x_k) - R_{\rho_k f_k}x_k\|^2] \\ &\leq \|x_0 - z\|^2 - \|x_{n+1} - z\|^2.\end{aligned} \tag{3.62}$$

Thus

$$\sum_{k=0}^{n} \rho_k(2-\rho_k)\|G_{f_k}x_k - x_k\|^2 \leq \|x_0 - z\|^2. \tag{3.63}$$

Then let $n \to \infty$ we have,

$$\sum_{k=0}^{\infty} \rho_k(2-\rho_k)\|G_{f_k}x_k - x_k\|^2 \leq \|x_0 - z\|^2 < +\infty. \tag{3.64}$$

Let $y_k = R_{\rho_k f_k}x_k$. By (3.62) we have

$$\sum_{k=0}^{n} \|G_c(y_k) - y_k\|^2 \leq \|x_0 - z\|^2. \tag{3.65}$$

Then let $n \to \infty$, we have

$$\sum_{k=0}^{\infty} \|G_c(y_k) - y_k\|^2 \leq \|x_0 - z\|^2 < +\infty. \tag{3.66}$$

By (3.64) we have

$$\epsilon \cdot \sum_{k=0}^{\infty} \|G_{f_k} x_k - x_k\|^2 < +\infty, \tag{3.67}$$

thus

$$\|G_{f_k} x_k - x_k\| \to 0. \tag{3.68}$$

Therefore,

$$\begin{aligned} &\|G_{f_k} x_k - x_k\| \\ &= \|x_k - \frac{f_k(x_k)}{\|\nabla f_k(x_k)\|^2} \nabla f_k(x_k) - x_k\| \\ &= \frac{f_k(x_k)}{\|\nabla f_k(x_k)\|} \to 0. \end{aligned} \tag{3.69}$$

Since $\nabla f_k(z) = 0$, by Lemma 3.2(iii),

$$\|\nabla f_k(x_k)\| = \|\nabla f_k(x_k) - \nabla f_k(z)\| \leq \|A\|^2 \|x_k - z\|.$$

Therefore, $(\nabla f_k(x_k))$ is bounded. It follows from (3.69) that $f_k(x_k) \to 0$, because

$$0 \leq f_k(x_k) = \|\nabla f_k(x_k)\| \cdot \|G_{f_k} x_k - x_k\| \leq M \|G_{f_k} x_k - x_k\| \to 0,$$

which means $\|(\mathrm{Id} - P_{Q_k}) A x_k\| \to 0$.

Since ∂q is bounded on bounded sets, there exists $\delta > 0$ such that $\|\zeta_k\| \leq$

δ. Since $P_{Q_k}(Ax_k) \in Q_k$, by Cauchy-Schwarz inequality we have

$$\begin{aligned} q(Ax_k) &\leq \langle \zeta, Ax_k - P_{Q_k}(Ax_k) \rangle \\ &\leq \delta \|(\mathrm{Id} - P_{Q_k})Ax_k\| \to 0. \end{aligned}$$

Now let $\bar{x}$ be a cluster point of (x_k) and (x_{k_l}) be a subsequence of (x_k), $(x_{k_l}) \to \bar{x}$. Then the contunuity of q implies that

$$q(A\bar{x}) = \lim_{l \to \infty} q(Ax_{k_l}) = 0.$$

Hence, $A\bar{x} \in Q$.

Since $y_{k_l} = R_{\rho_{k_l} f_{k_l}} x_{k_l} = x_{k_l} + \rho_{k_l}(G_{f_{k_l}} x_{k_l} - x_{k_l})$, so by (3.68)

$$\|y_{k_l} - x_{k_l}\| = \rho_{k_l} \|G_{f_{k_l}} x_{k_l} - x_{k_l}\| \to 0. \tag{3.70}$$

As $\|x_{k_l} - \bar{x}\| \to 0$, $y_{k_l} \to \bar{x}$. By (3.66) we have

$$\|G_c y_{k_l} - y_{k_l}\| \to 0.$$

So by Proposition 3.15(v) we have $\bar{x} \in C$. Thus $\bar{x} \in C$ and $A\bar{x} \cap Q$. The proof is done. ■

Chapter 4

SPA for S-subdifferentiable Functions

In the chapter, the SFP is extended to a nonconvex case. The sets C and Q are two level sets of two continuous and S-subdifferentiable functions, where S-subdifferential is defined below. Then the SPA with S-subdifferential is proved to converge to a point in the solution set $C \cap Q$.

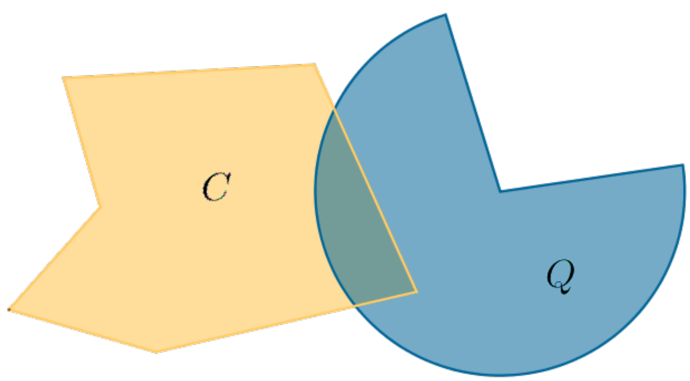

Figure 4.1: An example of the SFP with S-subdifferentiable functions.

Definition 4.1. **(S-subdifferential)** Given a set $S \subseteq \mathbb{R}^n$, $r > 0$, a vector $u \in \mathbb{R}^n$ is said to be a S-subgradient of function $f : \mathbb{R}^n \to \mathbb{R}$ at the point x

if we have

$$(\forall y \in \mathbb{R}^n) \quad \langle y - x, u \rangle + f(x) + \frac{r}{2} \mathrm{d}_S^2(x) \leq f(y) + \frac{r}{2} \mathrm{d}_S^2(y). \tag{4.1}$$

The set of all S-subgradients of f at x is called the S-subdifferential of f at x and is denoted by

$$\begin{aligned} \hat{\partial} f(x) = \{u \in \mathbb{R}^n : \langle y - x, u \rangle + f(x) + \frac{r}{2} \mathrm{d}_S^2(x) \\ \leq f(y) + \frac{r}{2} \mathrm{d}_S^2(y) \quad \forall y \in \mathbb{R}^n\}. \end{aligned} \tag{4.2}$$

Of course, one can use $\partial_{S,r} f(x)$ to indicate that S-subdiffernetial relies on S and r, but we choose $\hat{\partial} f(x)$ instead for the simplicity of notation.

Note that

– If $S = \mathbb{R}^n$, then the S-subdifferential is the Fenchel subdifferential.

– If $r = 0$, then the S-subdifferential is the Fenchel subdifferential.

Definition 4.2. Let S be a closed convex subset of $\mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be S-subdifferentiable on $\mathbb{R}^n$ with respect to S if $\exists r > 0$ such that $f + \frac{r}{2} \mathrm{d}_S^2$ is Fenchel subdifferentiable everywhere.

4.1 Basic Properties of S-subdifferential $\hat{\partial} f(x)$

Fact 4.3. *If $g : \mathbb{R}^n \to \mathbb{R}$ is Fenchel subdifferentiable on $\mathbb{R}^n$, then g is convex.*

Proof. Let $x, y \in \mathbb{R}$, $0 \leq \lambda \leq 1$. Take $v \in \partial g(\lambda x + (1-\lambda) y)$. Then we have

$$g(x) \geq g(\lambda x + (1-\lambda) y) + \langle v, x - (\lambda x + (1-\lambda) y) \rangle, \tag{4.3}$$

and

$$g(y) \geq g(\lambda x + (1-\lambda)y) + \langle v, y - (\lambda x + (1-\lambda)y)\rangle. \tag{4.4}$$

Then we have

$$\lambda g(x) + (1-\lambda)g(y) \geq g(\lambda x + (1-\lambda)y) + \langle v, 0\rangle = g(\lambda x + (1-\lambda)y).$$

■

Theorem 4.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be S-subdifferentiable everywhere on $\mathbb{R}^n$, $r > 0$. Then $f + \frac{r}{2} d_S^2$ is convex.*

Proof. Define $g = f + \frac{r}{2}\mathrm{d}_S^2$. By assumption, g is Fenchel subdifferentiable everywhere on $\mathbb{R}^n$. Then g is convex by Fact 4.3. ■

Lemma 4.5. *Let $S \subseteq \mathbb{R}^n$ be closed and convex. Then $\frac{1}{2} d_S^2$ is convex and differentiable, and*

$$\partial(\frac{1}{2} d_S^2)(x) = \nabla(\frac{1}{2} d_S^2)(x) = (\mathrm{Id} - P_S)(x). \tag{4.5}$$

Proof. Since

$$\frac{1}{2}\mathrm{d}_S^2(x) = \frac{1}{2}\|(\mathrm{Id} - P_S)x\|^2. \tag{4.6}$$

So by Lemma 3.2 (i)&(ii), we have $\frac{1}{2}\mathrm{d}_S^2$ is convex and differentiable, and

$$\partial(\frac{1}{2}\mathrm{d}_S^2)(x) = \nabla(\frac{1}{2}\mathrm{d}_S^2)(x) = (\mathrm{Id} - P_S)(x). \tag{4.7}$$

■

Lemma 4.6. *Assume $S \subseteq \mathbb{R}^n$ is closed and convex, and $f : \mathbb{R}^n \to \mathbb{R}$ is*

S-subdifferentiable everywhere. Then $\exists r \geq 0$ such that

$$u \in \hat{\partial} f(x) \Longleftrightarrow u \in \partial f(x) + r(\mathrm{Id} - P_S)(x). \tag{4.8}$$

Proof. Let

$$g(y) = f(y) + \frac{r}{2}\mathrm{d}_S^2(y) - \langle u, y\rangle. \tag{4.9}$$

Then by Theorem 4.4, $f + \frac{r}{2}d_S^2$ is convex. Thus, g is convex. By (4.2) we know $g(y)$ attains minimum at $y = x$. Hence by Lemma 4.5

$$\begin{aligned} & 0 \in \partial g(x) \\ \Longleftrightarrow\ & 0 \in \partial(f + \frac{r}{2}\mathrm{d}_S^2 - \langle u, \cdot\rangle)(x) \\ \Longleftrightarrow\ & 0 \in \partial f(x) + r(\mathrm{Id} - P_S) - u. \end{aligned}$$

Then

$$u \in \partial f(x) + r(\mathrm{Id} - P_S)(x). \tag{4.10}$$

■

Lemma 4.7. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz, and S be a closed convex subset of $\mathbb{R}^n$. Then $\hat{\partial} f$ with respect to S is bounded on bounded sets, i.e., $\forall l > 0$, $\exists M > 0$ such that $u \in \hat{\partial} f(B(l; 0)) \Rightarrow \|u\| \leq M$.*

Proof. The function f is locally Lipschitz on $\mathbb{R}^n$, so ∂f is locally bounded, by compactness arguments. Thus, ∂f is bounded on bounded sets. Since $r(\mathrm{Id} - P_S)$ is also bounded on bounded sets, $\hat{\partial} f$ is as well.

■

Remark 4.8. (i) Let $f : \mathbb{R} \to \mathbb{R}$, $f''(x) \geq c$ with $c \in \mathbb{R}$ and $\operatorname{argmin} f = \{x_0\}$ be a singleton. Then $f + \frac{r(x-x_0)^2}{2}$ is convex, whenever $r \geq -c$.

This is not true if $\operatorname{argmin} f$ is not a singleton.

(ii) Let $D^2 f(x)$ be the Hessian matrix of f at x. Assume $D^2 f(x) \geq -c\,\mathrm{Id}$ with $c \in \mathbb{R}$ and $\operatorname{argmin} f = \{x_0\}$ is a singleton. Then $f + \frac{r}{2}\|x - x_0\|^2$ is convex, whenever $r \geq -c$.

4.2 SPA with S-subgradient

Now let us define the subgradient projector with the S-subgradient.

Definition 4.9. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous, and let $\xi \in \mathbb{R}$ be such that $C = \operatorname{lev}_{\leq \xi} f \neq \emptyset$. Let $C \subseteq S \subseteq \mathbb{R}^n$ and S be closed and convex. Assume that f is subdifferentiable on $\mathbb{R}^n$. Let $\hat{\partial} f$ be the S-subdifferential of f with respect to S, and let s be a selection of $\hat{\partial} f$. The subgradient projector onto C associated with (f, ξ, s) is

$$\hat{G} : \mathcal{H} \to \mathcal{H} : x \mapsto \begin{cases} x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > \xi \\ x, & \text{if } f(x) \leq \xi. \end{cases} \tag{4.11}$$

Then we define

$$\hat{R} = \mathrm{Id} + \lambda(\hat{G} - \mathrm{Id}), \tag{4.12}$$

where $\lambda \in \,]0, 2]$.

Lemma 4.10. *Assume $C = \operatorname{lev}_{\leq \xi} f \neq \emptyset$ and $\exists r > 0$ such that f is S-subdifferentiable on $\mathbb{R}^n$. If $x \notin C$, then*

$$s(x) \in \hat{\partial} f(x) \Rightarrow s(x) \neq 0. \tag{4.13}$$

Proof. Assume $s(x) = 0$, then $0 \in \partial(f + \frac{r}{2} \mathrm{d}_S^2)(x)$, so $f + \frac{r}{2}\mathrm{d}_S^2$ attains its minimum at x. Then $\forall y \in C$,

$$f(x) + \frac{r}{2}\mathrm{d}_S^2(x) \leq f(y) + \frac{r}{2}\mathrm{d}_S^2(y) \leq \xi + 0 = \xi, \tag{4.14}$$

since $C \subset S$. Then $f(x) \leq \xi$. Thus, $x \in C$, which contradicts with the assumption. ■

Lemma 4.11. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous, and let $\xi \in \mathbb{R}$ be such that $C = \mathrm{lev}_{\leq \xi} f \neq \emptyset$. Let $C \subseteq S \subseteq \mathbb{R}^n$ and S be closed and convex. Assume that f is S-subdifferentiable on $\mathbb{R}^n$. Let $\hat{\partial} f$ be the S-subdifferential of f with respect to S, and let s be a selection of $\hat{\partial} f$. Assume $f + \frac{r}{2} d_S^2$ is convex. Let $\hat{G}$ be the subgradient projector onto C associated with (f, ξ, s), let $x \in \mathbb{R}^n$, and let $z \in C$. Then the following hold:*

(i) $\langle z - \hat{G}x, x - \hat{G}x \rangle \leq 0$.

(ii) $\|\hat{R}x - z\|^2 \leq \|x - z\|^2 - \lambda(2-\lambda)\|\hat{G}x - x\|^2$.

(iii) $\|\hat{G}x - z\|^2 \leq \|x - z\|^2 - \|\hat{G}x - x\|^2$.

(iv) $\max\{f(x) - \xi, 0\} = \|s(x)\| \|\hat{G}x - x\|$.

(v) *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and let $\bar{x}$ be a point in $\mathbb{R}^n$ such that $\hat{G}x_n - x_n \to 0$ and $x_n \to \bar{x}$. Then $\bar{x} \in C$.*

Proof. (i): By assumption, we have $z \in C$ and $x \in \mathbb{R}^n$. If $x \in C$, then

$$\begin{aligned} &\langle z - \hat{G}x, x - \hat{G}x \rangle \\ &= \langle z - \hat{G}x, x - x \rangle \\ &= 0. \end{aligned}$$

If $x \notin C$, then $f(x) > \xi$ and

$$\begin{aligned}
&\langle z - \hat{G}x, x - \hat{G}x\rangle \\
&= \langle z - x + \frac{f(x)-\xi}{\|s(x)\|^2} s(x), \frac{f(x)-\xi}{\|s(x)\|^2} s(x)\rangle \\
&= \langle z - x, \frac{f(x)-\xi}{\|s(x)\|^2} s(x)\rangle + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} \\
&= \frac{f(x)-\xi}{\|s(x)\|^2}\langle z - x, s(x)\rangle + \frac{(f(x)-\xi)^2}{\|s(x)\|^2}.
\end{aligned}$$

Since $\frac{f(x)-\xi}{\|s(x)\|^2} > 0$, we have

$$\begin{aligned}
&\frac{f(x)-\xi}{\|s(x)\|^2}\langle z - x, s(x)\rangle + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} \\
&\leq \frac{f(x)-\xi}{\|s(x)\|^2}\Big(f(z) - f(x) + \frac{r}{2}(\mathrm{d}_S^2(z) - \mathrm{d}_S^2(x))\Big) + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} \\
&\leq \frac{f(x)-\xi}{\|s(x)\|^2}\Big(\xi - f(x) + \frac{r}{2}(\mathrm{d}_S^2(z) - \mathrm{d}_S^2(x))\Big) + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} \\
&\leq \frac{f(x)-\xi}{\|s(x)\|^2}\Big(\xi - f(x) + \frac{r}{2}\mathrm{d}_S^2(z)\Big) + \frac{(f(x)-\xi)^2}{\|s(x)\|^2}.
\end{aligned}$$

Since $z \in C$ and $C \subseteq S$, $d_S(z) = 0$. Hence

$$\begin{aligned}
&\frac{f(x)-\xi}{\|s(x)\|^2}(\xi - f(x) + \frac{r}{2}\mathrm{d}_S^2(z)) + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} \\
&= \frac{-(f(x)-\xi)^2}{\|s(x)\|^2} + \frac{(f(x)-\xi)^2}{\|s(x)\|^2} = 0.
\end{aligned}$$

Therefore, $\langle z - \hat{G}x, x - \hat{G}x\rangle \leq 0$.

(ii): Using (i) we have

$$\begin{aligned}
&\|\hat{R}x - z\|^2 \\
&= \|x + \lambda(\hat{G}x - x) - z\|^2
\end{aligned}$$

$= \|x - z + \lambda(\hat{G}x - x)\|^2$

$= \|x - z\|^2 + 2\lambda\langle x - z, \hat{G}x - x\rangle + \lambda^2\|\hat{G}x - x\|^2$

$= \|x - z\|^2 + 2\lambda\langle x - \hat{G}x, \hat{G}x - x\rangle + 2\lambda\langle \hat{G}x - z, \hat{G}x - x\rangle + \lambda^2\|\hat{G}x - x\|^2$

$= \|x - z\|^2 + 2\lambda\langle \hat{G}x - z, \hat{G}x - x\rangle - \lambda(2-\lambda)\|\hat{G}x - x\|^2$

$\leq \|x - z\|^2 - \lambda(2-\lambda)\|\hat{G}x - x\|^2 \quad \text{(By } (i))$

$= \|x - z\|^2 - \lambda(2-\lambda)\|\hat{G}x - x\|^2.$

(iii): Set $\lambda = 1$ in (ii).

(iv): **Case 1**: $x \in C$. Then $f(x) - \xi \leq 0$ and $\hat{G}x = x$. So

$$\max\{f(x) - \xi, 0\} = 0$$

and

$$\|s(x)\|\|\hat{G}x - x\| = 0.$$

Hence, $\max\{f(x) - \xi, 0\} = \|s(x)\|\|\hat{G}x - x\|$.

Case 2: $x \notin C$. Then $f(x) > \xi$. Hence, $\max\{f(x) - \xi, 0\} = f(x) - \xi$.

$$\begin{aligned}
\|s(x)\|\|\hat{G}x - x\| &= \|s(x)\| \left\|(x + \frac{\xi - f(x)}{\|s(x)\|^2}s(x)) - x\right\| \\
&= \left\|\frac{\xi - f(x)}{\|s(x)\|}s(x)\right\| \\
&= \frac{f(x) - \xi}{\|s(x)\|}\|s(x)\| \\
&= f(x) - \xi.
\end{aligned}$$

So $\max\{f(x) - \xi, 0\} = \|s(x)\|\|\hat{G}x - x\|$.

(v): By Fact 2.66(iii) and Lemma 4.6, there exists $\rho > 0$ such that

$(x_k)_{k\in\mathbb{N}}$ lies in $B(\bar{x};\rho)$ and

$$\sigma = \sup \|\partial f(B(\bar{x};\rho))\| + r \sup_{k\in\mathbb{N}} \|(\mathrm{Id} - P_S)x_k\| < +\infty.$$

Hence, we have

$$(\forall n \in \mathbb{N}) \quad \|s(x_n)\| \leq \sigma. \tag{4.15}$$

Hence, by (iv),

$$\max\{f(\bar{x}) - \xi, 0\} \leq \lim \max\{f(x_k) - \xi, 0\} \leq \sigma \lim \|\hat{G}x_k - x_k\| = 0. \tag{4.16}$$

Thus, $f(\bar{x}) \leq \xi$.

■

Recall

$$C = \{x \in \mathbb{R}^n : c(x) \leq \xi\},$$

$$Q = \{x \in \mathbb{R}^n : q(x) \leq \zeta\},$$

where $\xi, \zeta \in \mathbb{R}_{++}$, and the functions c and q are S-subdifferentiable.

Let U and V be closed convex subsets of $\mathbb{R}^n$ such that $C \subseteq U$, $\hat{Q} \subseteq V$. Assume c and q are S-subdifferentiable on $\mathbb{R}^n$ with respect to U and V (i.e., $\exists r_1 > 0, r_2 > 0$ such that $c + \frac{r_1}{2}\mathrm{d}_U^2$ and $q + \frac{r_2}{2}\mathrm{d}_V^2$ are convex). Let $\hat{\partial}c(x)$ be S-subdifferential of c with respect to U, $\hat{\partial}q(x)$ be S-subdifferetial with respect to V and $\hat{\partial}c(x)$, $\hat{\partial}q(x)$ not empty for all $x \in \mathbb{R}^n$. Let $s_c(x)$ be a selection of $\hat{\partial}c(x)$ and $s_{q(x)}$ be a selection of $\hat{\partial}q(x)$.

Define

$$\hat{G}_c : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto \begin{cases} x + \frac{\xi - c(x)}{\|s_c(x)\|^2} s_c(x), & \text{if } c(x) > \xi \\ x, & \text{if } c(x) \leq \xi, \end{cases} \tag{4.17}$$

$$\hat{G}_q : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto \begin{cases} x + \frac{\xi - q(x)}{\|s_q(x)\|^2} s_q(x), & \text{if } q(x) > \zeta \\ x, & \text{if } q(x) \leq \zeta, \end{cases} \tag{4.18}$$

$$\hat{R}_{\lambda q} = \operatorname{Id} + \lambda(\hat{G}_q - \operatorname{Id}), \tag{4.19}$$

where $\lambda \in]0, 2]$.

Now we can define the subgradient projection algorithm with the new subgradient projector by iteration

$$x_{k+1} = \hat{G}_c(\hat{R}_{\lambda_k q}(x_k)). \tag{4.20}$$

First, we show the sequence converges to a point in $C \cap \hat{Q}$.

Theorem 4.12. *Assume $C \cap Q \neq \emptyset$ and $\inf_k \lambda_k(2 - \lambda_k) > 0$. The sequence (x_k) generated by* (4.20) *converges to a solution x^* such that $x^* \in C \cap Q$.*

Proof. Let $z \in C \cap Q$. By (4.20), we have

$$\begin{aligned} & \|x_{k+1} - z\|^2 && (4.21) \\ &= \|\hat{G}_c(\hat{R}_{\lambda_k q} x_k) - z\|^2 \\ &\leq \|\hat{R}_{\lambda_k q} - z\|^2 - \|\hat{G}_c(\hat{R}_{\lambda_k q} x_k) - \hat{R}_{\lambda_k q} x_k\|^2 \quad \text{(By Proposition 4.11 (iii))} \\ &\leq \|x_k - z\|^2 - \lambda_k(2 - \lambda_k)\|\hat{G}_q x_k - x_k\|^2 \\ &- \|\hat{G}_c(\hat{R}_{\lambda_k q} x_k) - \hat{R}_{\lambda_k q} x_k\|^2 \quad \text{(By Proposition 4.11 (ii))}. && (4.22) \end{aligned}$$

So we have

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2, \tag{4.23}$$

hence, $(x_k)_{k \in \mathbb{N}}$ is Fejér monotone. So by Fact 2.79, we just need to prove every cluster point of $(x_k)_{k\in\mathbb{N}}$ is in the solution set. Taking the summation on both side of (4.21) gives us

$$\begin{aligned}&\|x_{n+1}-z\|^2\\&\le \|x_0-z\|^2-\sum_{k=0}^{n}[\lambda_k(2-\lambda_k)\|\hat{G}_q x_k - x_k\|^2+\|\hat{G}_c(\hat{R}_q x_k)-\hat{R}_{\lambda_k q}x_k\|^2].\end{aligned}$$

So we have

$$\begin{aligned}&\sum_{k=0}^{n}[\lambda_k(2-\lambda_k)\|\hat{G}_q x_k - x_k\|^2+\|\hat{G}_c(\hat{R}_{\lambda_k q} x_k)-\hat{R}_{\lambda_k q}x_k\|^2]\\&\le \|x_0-z\|^2-\|x_{n+1}-z\|^2.\end{aligned} \tag{4.24}$$

Thus

$$\sum_{k=0}^{n}\lambda_k(2-\lambda_k)\|\hat{G}_q x_k - x_k\|^2 \le \|x_0-z\|^2. \tag{4.25}$$

Letting $n\to\infty$, we have

$$\sum_{k=0}^{\infty}\lambda_k(2-\lambda_k)\|\hat{G}_q x_k - x_k\|^2 \le \|x_0-z\|^2 < +\infty. \tag{4.26}$$

Let $y_k=\hat{R}_{\lambda_k q}x_k$. By (4.24) we have

$$\sum_{k=0}^{n}\|\hat{G}_c(y_k)-y_k\|^2 \le \|x_0-z\|^2. \tag{4.27}$$

Then letting $n \to \infty$, we have

$$\sum_{k=0}^{\infty} \|\hat{G}_c(y_k) - y_k\|^2 \leq \|x_0 - z\|^2 < +\infty. \tag{4.28}$$

As $\inf_k \lambda_k(2 - \lambda_k) > 0$, so $\exists \epsilon > 0$ such that $\lambda_k(2 - \lambda_k) > \epsilon$, $\forall k \geq 1$. By (4.26) we have

$$\epsilon \sum_{k=0}^{\infty} \|\hat{G}_q x_k - x_k\|^2 < +\infty.$$

So

$$\sum_{i=0}^{\infty} \|\hat{G}_q x_k - x_k\|^2 < +\infty, \tag{4.29}$$

and $\|\hat{G}_q x_k - x_k\| \to 0$. Now let $\bar{x}$ be a cluster point of (x_k) and (x_{k_l}) be a subsequence of (x_k) and $(x_{k_l}) \to \bar{x}$. Then by Lemma 4.11(v) we have $\bar{x} \in Q$.

Since $y_{k_l} = \hat{R}_q x_{k_l} = x_{k_l} + \lambda_{k_l}(\hat{G}_q x_{k_l} - x_{k_l})$,

$$\|y_{k_l} - x_{k_l}\| = \lambda_{k_l}\|\hat{G}_q x_{k_l} - x_{k_l}\| \to 0. \tag{4.30}$$

Hence, $y_{k_l} \to \bar{x}$. By (4.28) we have $\|\hat{G}_c y_{k_l} - y_{k_l}\| \to 0$. So by Lemma 4.11(v) we have $\bar{x} \in C$. Thus, $\bar{x} \in C \cap Q$. The proof is done. ■

The following corollary shows that the sequence converges to a solution of the SFP.

Corollary 4.13. *Let*

$$\hat{Q} = \{y \in \mathbb{R}^m : \hat{q}(y) \leq 0\}, \tag{4.31}$$

where $q = \hat{q} \circ A$, $A : \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix. Assume $C \cap Q \neq \emptyset$ and $\inf_k \lambda_k(2 - \lambda_k) > 0$. The sequence (x_k) generated by (4.20) converges to a *point* x^* *such that* $x^* \in C$ *and* $Ax^* \in \hat{Q}$.

Proof. By Theorem 4.12 we know the sequence converges to a point $x^* \in C \cap Q$. So $x^* \in Q$, thus by Lemma 3.19 we have $Ax^* \in \hat{Q}$. The proof is done. ■

4.3 A Numerical Example

Let

$$C = \{x : c(x) \leq 0\},$$
$$Q = \{x : q(x) \leq 0\},$$

where

$$c(x) = x^4 - x^2 - 2x + 2,$$
$$q(x) = \ln(x^2 - 2x + 2).$$

Then $C = Q = \{1\}$. Let $S = \{1\}$, $r = 2$. Then by Lemma 4.6, we have

$$\hat{\partial} c(x) = \partial c(x) + 2(\mathrm{Id} - P_S)(x) = 4x^3 - 2x - 2 + 2(x-1) = 4x^3 - 4,$$

$$\hat{\partial} q(x) = \partial q(x) + 2(\mathrm{Id} - P_S)(x) = \frac{2x-2}{x^2 - 2x + 2} + 2x - 2.$$

Function c is not convex because

$$c''(x) = 12x^2 - 2,$$

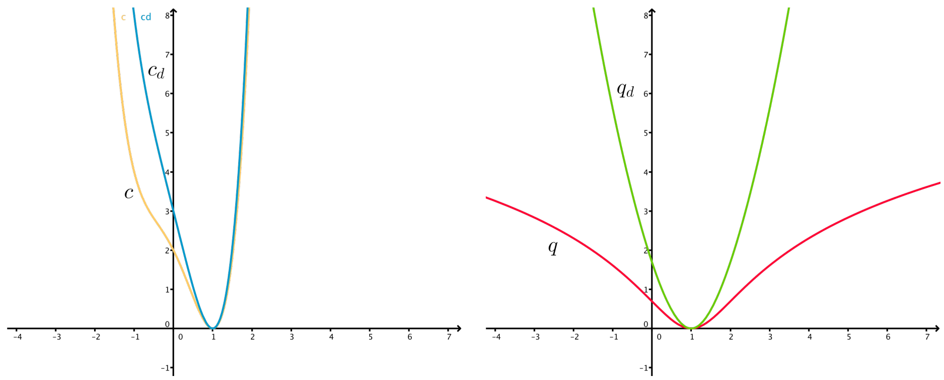

Figure 4.2: Function c and function q

and

$$c''(0) = -2 < 0.$$

Function q is also not convex as

$$q''(x) = \frac{2x(2-x)}{(x^2-2x+2)^2},$$

and

$$q''(-2) = -\frac{4}{25} < 0.$$

However, $c_d = c + \mathrm{d}_S^2$ and $q_d = q + \mathrm{d}_S^2$ are convex because

$$c_d''(x) = 12x^2 \geq 0,$$

and

$$q_d''(x) = \frac{2x(2-x)}{(x^2-2x+2)^2} + 2 = 2\frac{(x-1)^2[(x-1)^2+1]+2}{[(x-1)^2+1]^2} > 0.$$

See Figure 4.2.

Let $\lambda = 1$, $x_0 = 5$. Then the result is shown in Table 4.1. We can see it converges to the solution set $C \cap Q = \{1\}$ in 16 iterations.

Iter	x_k
0	5
1	3.5599
2	2.4981
3	1.7791
4	1.3688
5	1.1673
6	1.0745
⋮	⋮
13	1.0005
14	1.0002
15	1.0001
16	1.0000

Table 4.1: Numerical result for S-subdifferentiable functions.

Chapter 5

CQ Algorithm for Prox-regular Functions

In this chapter, the SFP is extended to another nonconvex case. We assume that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and function $g : \mathbb{R}^n \to \mathbb{R}$ is prox-regular and locally Lipschitz. Then the SFP becomes finding $\bar{x}$ such that

$$0 \in \partial f(\bar{x}) \text{ and } 0 \in \partial_P g(A\bar{x}). \tag{5.1}$$

where $\partial_P g(\bar{x})$ denotes the proximal subdifferential. Moreover, we assume for sequences $(x_k)_{k\in\mathbb{N}}$ and $(\lambda_k)_{k\in\mathbb{N}}$, g satisfies:

$$\begin{cases} \text{whenever} \\ \qquad \|x_k - \operatorname{prox}_{\lambda_k g}(x_k)\| \to 0, \text{ and } \lambda_k \to 0, \\ \text{we have} \\ \qquad \lim_{k\to 0} \frac{\|x_k - \operatorname{prox}_{\lambda_k g}(x_k)\|}{\lambda_k} = 0. \end{cases} \tag{5.2}$$

The definition of proximal subdifferential is given below.

Definition 5.1. (Proximal Subdifferential and Limiting Subdifferential)

(i) A vector v is in the **proximal subdifferential** of $\bar{x}$, denoted $\partial_P g(\bar{x})$, if there exist $r > 0$ and $\varepsilon > 0$ such that for all $x \in B(\bar{x}, \varepsilon)$,

$$\langle v, x - \bar{x} \rangle \leq g(x) - g(\bar{x}) + \frac{r}{2}\|x - \bar{x}\|^2. \tag{5.3}$$

(ii) A vector v is in the **limiting subdifferential** of $\bar{x}$, denoted $\tilde{\partial} g(\bar{x})$, if there are sequences $x_k \to \bar{x}$ and $v_k \in \partial_P g(x_k)$ such that $v_k \to v$.

See [14] for more properties of proximal subdifferentials and limiting subdifferetials.

By taking $f = \max\{c(x), 0\}$, $g = \max\{q(x), 0\}$, we have

$$\{x \in \mathbb{R}^n : c(x) \leq 0\} = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\},$$

and

$$\{x \in \mathbb{R}^n : q(Ax) \leq 0\} = \{x \in \mathbb{R}^n : 0 \in \tilde{\partial} g(x)\}.$$

Then problem (5.1) recovers the original SFP, which says: find $\bar{x}$ such that $\bar{x} \in C$ and $A\bar{x} \in Q$.

When a function is prox-regular and locally Lipschitz, the proximal subdifferential and the limiting subdifferential are actually the same. For the definition of prox-regularity, see Definition 5.6.

Lemma 5.2. *Let f be prox-regular and locally Lipschitz on an open set O. Then $\partial_P f = \tilde{\partial} f$ on O.*

Proof. By the assumption and [14, Proposition 13.33], we have f is lower-C^2 on O. Then by [14, Theorem 10.33], there exists a finite, convex function g

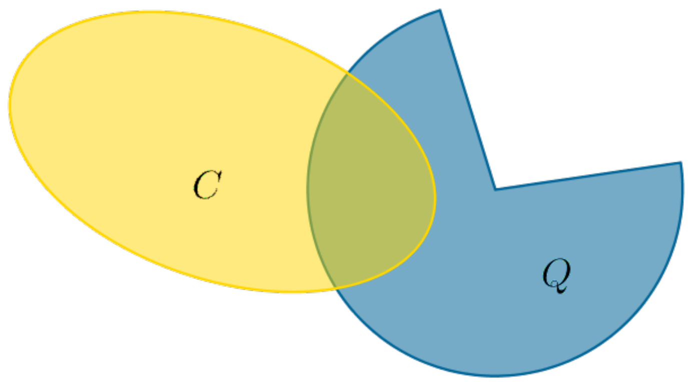

Figure 5.1: An example of the SFP with prox-regular function.

such that for some $\rho > 0$

$$f = g - \frac{1}{2}\rho\|\cdot\|^2. \tag{5.4}$$

Then $\forall x \in O$,

$$\begin{aligned}\partial_P f(x) &= \partial_P g(x) - \rho x \\ &= \tilde{\partial} g(x) - \rho x \\ &= \tilde{\partial}\left(g(x) - \frac{1}{2}\|x\|^2\right) \\ &= \tilde{\partial} f(x).\end{aligned}$$

■

Remark 5.3. If f is twice continuously differentiable, then f verifies the assumption in Lemma 5.2.

Then let us see a simple fact about the proximal mapping.

Fact 5.4. *For every $g : \mathbb{R}^n \to]-\infty, +\infty]$, we have*

$$\frac{x - \text{prox}_{\lambda g}(x)}{\lambda} \in \partial_P g(\text{prox}_{\lambda g}(x)). \tag{5.5}$$

Proof. By the proximal mapping definition

$$y \mapsto g(y) + \frac{1}{2\lambda}\|x - y\|^2 \text{ attains minimum at } \text{prox}_{\lambda g}(x).$$

Then

$$0 \in \partial_P g(\text{prox}_{\lambda g}(x)) + \frac{1}{\lambda}(\text{prox}_{\lambda g}(x) - x), \tag{5.6}$$

Hence,

$$\frac{x - \text{prox}_{\lambda g}(x)}{\lambda} \in \partial_P g(\text{prox}_{\lambda g}(x)). \tag{5.7}$$

■

5.1 *CQ* Algorithm with Proximal Mapping

Let

$$h(x_k) = \frac{1}{2}\|(\text{Id} - \text{prox}_{\lambda_k g})Ax_k\|^2, \tag{5.8}$$

and

$$\hat{h}(x_k) = A^*(\text{Id} - \text{prox}_{\lambda_k g})Ax_k. \tag{5.9}$$

Define

$$\mu_k = \frac{\rho_k h(x_k)}{\|\hat{h}(x_k)\|^2}, \tag{5.10}$$

where $0 < \rho_k < 4$.

Note that $(\mu_k)_{k\in\mathbb{N}}$ is a López-type stepsize, given in section 2.6. In the iteration of the CQ algorithm, using proximal mapping instead of projection,

$$x_{k+1} = \text{prox}_f(x_k - \mu_k A^*(\text{Id} - \text{prox}_{\lambda_k g})Ax_k), \tag{5.11}$$

gives us the algorithm below.

Algorithm 4: CQ Algorithm with Proximal Mapping

Choose arbitrary x_0;

while $x_{k+1} \neq x_k$ **do**

  Set $x_k = x_{k+1}$;

  Set $x_{k+1} = \text{prox}_f(x_k - \mu_k A^*(\text{Id} - \text{prox}_{\lambda_k g})Ax_k)$;

end

Notice that by taking $f = \delta_C$ and $g = \delta_Q$, the indicator functions of two nonempty convex sets of $\mathbb{R}^n$, we have

$$\text{prox}_f = P_C \text{ and } \text{prox}_{\lambda g} = P_Q.$$

Then we recover Algorithm 1 which is given on page 44.

5.2 Prox-regularity

Before stating the definition of prox-regularity of g, let us first see the definition of the $g-$attentive $\varepsilon-$localization of $\partial_P g$.

Definition 5.5. For $\varepsilon > 0$, the $g-$attentive $\varepsilon-$localization of $\partial_P g$ around

$(\bar{x}, \bar{\nu})$, is the mapping $T_\varepsilon : \mathcal{H}_2 \mapsto 2^{\mathcal{H}_2}$ defined by

$$T_\varepsilon(x) = \begin{cases} \{\nu \in \partial_P g(x), \|\nu - \bar{\nu}\| < \varepsilon\} & \text{if } \|x - \bar{x}\| < \varepsilon \text{ and } |g(x) - g(\bar{x})| < \varepsilon, \\ \emptyset & \text{otherwise.} \end{cases}$$

Then the prox-regularity of g is defined as follows:

Definition 5.6. (Prox-regularity of functions) A function g is said to be prox-regular at $\bar{x}$ for $\bar{\nu}$, if g is finite and and locally lsc at $\bar{x}$ with $\bar{\nu} \in \partial_P g(\bar{x})$ and there exist $\varepsilon > 0$ and $r \geq 0$ such that

$$g(x') \geq g(x) + \langle \nu, x' - x \rangle - \frac{r}{2}\|x' - x\|^2 \text{ for all } x' \in B(\bar{x}; \varepsilon)$$
$$\text{when } \nu \in \partial_p g(x), |\nu - \bar{\nu}| < \varepsilon, |x - \bar{x}| < \varepsilon, g(x) < g(\bar{x}) + \varepsilon.$$

When this holds for all $\bar{\nu} \in \partial_P g(\bar{x})$, the function is said to be prox-regular at $\bar{x}$.

Fact 5.7. *Every convex function is prox-regular.*

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for all $x, x' \in \mathbb{R}^n$, $\nu \in \partial f(x)$,

$$f(x') \geq f(x) + \langle \nu, x' - x \rangle. \tag{5.12}$$

So f is prox-regular with $r = 0$ and every $\varepsilon > 0$. ■

Fact 5.8. *Every twice differentiable function is prox-regular.*

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Assume $\bar{x} \in \mathbb{R}^n$, $\bar{v} = \nabla f(\bar{x})$, $\|v - \bar{v}\| < \varepsilon$, $v = \nabla f(x)$. Let $D^2 f$ be the Hessian matrix.

Then by Taylor's formula with Lagrange's form of the remainder [1, page

283], $\exists c \in [x, y]$ such that

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

Since f is twice differentiable, we have $D^2 f$ is continuous. Then there exists $\delta > 0$ such that

$$\|D^2 f(y) - D^2 f(\bar{x})\| < 1$$

when $\|y - \bar{x}\| < \delta$. Then $\exists r > 0$ such that

$$\begin{aligned} \|D^2 f(y)\| &\leq \|D^2 f(y) - D^2 f(\bar{x})\| + \|D^2 f(\bar{x})\| \\ &< 1 + \|D^2 f(\bar{x})\| = r. \end{aligned}$$

when $\|y - \bar{x}\| < \delta$. Hence,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle - \frac{r}{2}\|y - x\|^2.$$

Therefore, f is prox-regular at $\bar{x}$. ■

5.3 Auxiliary Results

Now, let us state a proposition that summarizes some important consequences of prox-regularity.

Proposition 5.9. [14, Theorem 13.37] *Suppose that g is locally lower semicontinuous at $\bar{x}$ and prox-regular at $\bar{x}$ for $\bar{v} = 0$ with respect to r and ε. Let T_ε be the g-attentive ε-localization of $\partial_P g$ around $(\bar{x}, \bar{v})$. Then for each*

$\lambda \in]0, \frac{1}{r}[$ there is a neighborhood U_λ of $\bar{x}$ such that, on U_λ

(i) the mapping $\operatorname{prox}_{\lambda g}$ is single-valued and Lipschitz continuous with constant $\frac{1}{1-\lambda r}$ and $\operatorname{prox}_{\lambda g}(x) = (\operatorname{Id} + \lambda T_\epsilon)^{-1}(x)$ is a singleton.

(ii) $e_\lambda g$ is differentiable with

$$\nabla e_\lambda g(x) = \frac{x - \operatorname{prox}_{\lambda g}(x)}{\lambda} = (\lambda \operatorname{Id} + T_\varepsilon^{-1})^{-1}(x).$$

The following proposition gives us the monotonicity of $T_\varepsilon + r \operatorname{Id}$.

Proposition 5.10. [14, Theorem 13.36] *Suppose $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ is finite and locally lsc at $\bar{x}$, and let $\bar{\nu} \in \partial f(\bar{x})$ be a proximal subgradient. Then the following conditions are equivalent:*

(a) f is prox-regular at $\bar{x}$ for $\bar{\nu}$;

(b) ∂f has an $f-$attentive localization T around $(\bar{x}, \bar{\nu})$ such that $T + \rho \operatorname{Id}$ is monotone for some $\rho \in \mathbb{R}_+$.

Now, let us prove the following key property of the proximal mapping complement. It says that $\operatorname{Id} - \operatorname{prox}_{\lambda g}$ is ε-relaxed firmly nonexpansive on a neighborhood of $\bar{x}$ (see Definition 2.73).

Lemma 5.11. [11, Lemma 2.1] *If the assumptions of Proposition 5.9 hold, then $\forall \lambda \in (0, \frac{1}{r})$ and $\forall x_1, x_2 \in U_\lambda$, one has*

$$\begin{aligned}&\langle (\operatorname{Id} - \operatorname{prox}_{\lambda g})(x_1) - (\operatorname{Id} - \operatorname{prox}_{\lambda g})(x_2), x_1 - x_2 \rangle \\ &\geq \|(\operatorname{Id} - \operatorname{prox}_{\lambda g})(x_1) - (\operatorname{Id} - \operatorname{prox}_{\lambda g})(x_2)\|^2 - \frac{\lambda r}{(1-\lambda r)^2} \|x_1 - x_2\|^2.\end{aligned}$$

Proof. Let $x_i \in U_\lambda$, $i = 1, 2$. Then by Fact 5.4,

$$v_i = \frac{x_i - \operatorname{prox}_{\lambda g}(x_i)}{\lambda} \in T_\varepsilon(\operatorname{prox}_{\lambda g}(x_i)).$$

By the prox-regularity of g and Proposition 5.10 we have $T_\varepsilon + r\,\mathrm{Id}$ is monotone. So for the pairs (x_1, v_1) and (x_2, v_2),

$$\langle (v_1 + r\operatorname{prox}_{\lambda g}(x_1)) - (v_2 + r\operatorname{prox}_{\lambda g}(x_2)), \operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\rangle \geq 0.$$

This gives us

$$\langle v_1 - v_2, \operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\rangle \geq -r\|\operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\|^2. \tag{5.13}$$

We also have

$$\begin{aligned}
&\langle v_1 - v_2, x_1 - x_2\rangle \\
&= \langle v_1 - v_2, (x_1 - \operatorname{prox}_{\lambda g}(x_1)) - (x_2 - \operatorname{prox}_{\lambda g}(x_2))\rangle \\
&+ \langle v_1 - v_2, \operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\rangle \\
&= \langle \frac{x_1 - \operatorname{prox}_{\lambda g}(x_1)}{\lambda} - \frac{x_2 - \operatorname{prox}_{\lambda g}(x_2)}{\lambda}, (x_1 - \operatorname{prox}_{\lambda g}(x_1)) - (x_2 - \operatorname{prox}_{\lambda g}(x_2))\rangle \\
&+ \langle v_1 - v_2, \operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\rangle \\
&\geq \lambda^{-1}\|x_1 - \operatorname{prox}_{\lambda g}(x_1) - (x_2 - \operatorname{prox}_{\lambda g}(x_2))\|^2 \\
&- r\|\operatorname{prox}_{\lambda g}(x_1) - \operatorname{prox}_{\lambda g}(x_2)\|^2.
\end{aligned}$$

Hence

$$\langle (\mathrm{Id} - \operatorname{prox}_{\lambda g})(x_1) - (\mathrm{Id} - \operatorname{prox}_{\lambda g})(x_2), x_1 - x_2\rangle$$

$$\geq \|(\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_1) - (\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_2)\|^2 - \lambda r \| \mathrm{prox}_{\lambda g}(x_1) - \mathrm{prox}_{\lambda g}(x_2)\|^2.$$

By Proposition 5.9,

$$\| \mathrm{prox}_{\lambda g}(x_1) - \mathrm{prox}_{\lambda g}(x_2)\| \leq \frac{1}{1-\lambda r}\|x_1 - x_2\|, \tag{5.14}$$

we obtain

$$\begin{aligned} &\langle (\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_1) - (\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_2), x_1 - x_2\rangle \\ \geq\ &\|(\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_1) - (\mathrm{Id} - \mathrm{prox}_{\lambda g})(x_2)\|^2 - \frac{\lambda r}{(1-\lambda r)^2}\|x_1 - x_2\|^2 \end{aligned}$$

as required. ■

Then we give the following lemma which will be used to prove $\hat{h}(x_k)$ is bounded in a later proof.

Lemma 5.12. *Let $\hat{h}$ be the function defined in* (5.9)*. Let $\bar{x} \in \mathbb{R}^n$ and U_λ be a neighborhood of $\bar{x}$. Assume $x, y \in U_\lambda$ and g is prox-regular. Then*

$$\|\hat{h}(x) - \hat{h}(y)\| \leq \frac{2-\lambda r}{1-\lambda r}\|A\|^2\|x - y\|. \tag{5.15}$$

Proof.

$$\begin{aligned} \|\hat{h}(x) - \hat{h}(y)\| &= \|A^*(\mathrm{Id} - \mathrm{prox}_{\lambda g})Ax - A^*(\mathrm{Id} - \mathrm{prox}_{\lambda g})Ay\| \\ &\leq \|A\| \cdot \|(\mathrm{Id} - \mathrm{prox}_{\lambda g})Ax - (\mathrm{Id} - \mathrm{prox}_{\lambda g})Ay\| \\ &= \|A\| \cdot \|(Ax - Ay) - (\mathrm{prox}_{\lambda g} Ax - \mathrm{prox}_{\lambda g} Ay)\| \\ &\leq \|A\| \cdot (\|Ax - Ay\| + \| \mathrm{prox}_{\lambda g} Ax - \mathrm{prox}_{\lambda g} Ay\|) \\ &\leq \|A\| \cdot (\|Ax - Ay\| + \frac{1}{1-\lambda r}\|Ax - Ay\|) \quad (By\ (5.14)) \end{aligned}$$

$$= \|A\| \cdot \frac{2-\lambda r}{1-\lambda r}\|Ax - Ay\|$$
$$\leq \frac{2-\lambda r}{1-\lambda r}\|A\|^2\|x-y\|.$$

■

Now we state the following lemmas and fact which will be used in the proof of the main result in this chapter.

Lemma 5.13. [13, Lemma 2.2.2] *Let* (α_k), (β_k) *and* (γ_k), $k \in \mathbb{N}$ *be three sequences of nonegative numbers satisfying* $\alpha_{k+1} \leq (1+\beta_k)\alpha_k + \gamma_k$. *If* $\sum_{k=0}^{\infty} \beta_k < +\infty$, *and* $\sum_{k=0}^{\infty} \gamma_k < +\infty$, *then* (α_k) *is convergent.*

Fact 5.14. [2, Proposition 2.19] *Let* $\mathcal{K}$ *be a real Hilbert space and let* $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. *Then* $\operatorname{ran} T$ *is closed* $\Leftrightarrow$ $\operatorname{ran} T^*$ *is closed* $\Leftrightarrow$ $(\exists \alpha \in \mathbb{R}_{++})(\forall x \in (\ker T)^{\perp})\|Tx\| \geq \alpha\|x\|$.

5.4 Main Result

Lemma 5.15. *Let* $\gamma_k \geq 0$, $k \in \mathbb{N}$. *Then*

$$\prod_{k=0}^{\infty}(1+\gamma_k) < +\infty \Longleftrightarrow \sum_{k=0}^{\infty} \gamma_k < +\infty. \tag{5.16}$$

Proof. Write

$$p = \ln \prod_{k=0}^{\infty}(1+\gamma_k) = \sum_{k=0}^{\infty} \ln(1+\gamma_k).$$

Therefore,

$$p < +\infty \Longleftrightarrow \prod_{k=0}^{\infty}(1+\gamma_k) < +\infty. \tag{5.17}$$

Now let us prove (5.16). We can assume $\gamma_k > 0$, $\forall k \geq 0$, since $\gamma_k = 0$ can be ignored.

"$\Rightarrow$": Assume $p < +\infty$. Then $\lim_{k\to\infty} \ln(1+\gamma_k) \to 0$. So $\gamma_k = e^{\ln(1+\gamma_k)} - 1 \to e^0 - 1 = 0$.

As $\lim_{x\to 0^+} \frac{\ln(1+x)}{x} = \lim_{x\to 0^+} \frac{1}{1+x} = 1$, by the L'Hôpital's Rule, we have

$$\lim_{k\to\infty} \frac{\ln(1+\gamma_k)}{\gamma_k} = 1. \tag{5.18}$$

By the limit comparison test for series, we obtain

$$\sum_{k=0}^{\infty} \gamma_k < +\infty. \tag{5.19}$$

"$\Leftarrow$": Assume $\sum_{k=0}^{\infty} \gamma_k < +\infty$. Then $\gamma_k \to 0$. Then $\gamma_k \to 0$. Since

$$\lim_{k\to\infty} \frac{\ln(1+\gamma_k)}{\gamma_k} = \lim_{k\to\infty} \frac{1}{1+\gamma_k} = 1, \tag{5.20}$$

again by the comparison test,

$$\sum_{k=0}^{\infty} \ln(1+\gamma_k) < +\infty, \tag{5.21}$$

i.e., $p < +\infty$.

■

Lemma 5.16. *Assume $(x_k)_{k\in\mathbb{N}}$ is a sequence in $\mathbb{R}^n$, and $\sum_{k=0}^{\infty} \gamma_k < +\infty$, $\bar{x} \in \mathbb{R}^n$.*

Then the following hold:

(i) *If*

$$\|x_{k+1} - \bar{x}\|^2 \leq (1+\gamma_k)\|x_k - \bar{x}\|^2, \quad \forall k \in \mathbb{N}, \tag{5.22}$$

then

$$\|x_{k+1} - \bar{x}\|^2 \leq \left[\prod_{i=0}^{k}(1+\gamma_i)\right] \|x_0 - \bar{x}\|^2. \tag{5.23}$$

(ii) *Let* $\prod_{i=0}^{\infty}(1+\gamma_i) = M$. *Then* $M < +\infty$.

(iii) *If* $\|x_0 - \bar{x}\|^2 < \frac{r^2}{M}$, *and (i) and (ii) hold, then* $\|x_k - \bar{x}\|^2 < r^2$ *for* $k \in \mathbb{N}$.

Proof. (i): We have

$$\begin{aligned}
\|x_{k+1} - \bar{x}\|^2 &\leq (1+\gamma_k)\|x_k - \bar{x}\|^2 \\
&\leq (1+\gamma_k)(1+\gamma_{k-1})\|x_{k-1} - \bar{x}\|^2 \\
&\vdots \\
&\leq (1+\gamma_k)(1+\gamma_{k-1})\dots(1+\gamma_0)\|x_0 - \bar{x}\|^2 \\
&= \prod_{i=0}^{k}(1+\gamma_i)\|x_0 - \bar{x}\|^2.
\end{aligned}$$

(ii): As $\sum_{k=0}^{\infty}\gamma_k < +\infty$, by Lemma 5.15, it follows that

$$M = \prod_{k=0}^{\infty}(1+\gamma_k) < +\infty. \tag{5.24}$$

(iii): Since

$$\begin{aligned}
M = \prod_{i=0}^{\infty}(1+\gamma_k) &= \prod_{i=0}^{k}(1+\gamma_i) \cdot \prod_{i=k+1}^{\infty}(1+\gamma_i) \\
&\geq \prod_{i=0}^{k}(1+\gamma_i),
\end{aligned}$$

by (i), $\forall k \in \mathbb{N}$,

$$\begin{aligned}\|x_{k+1} - \bar{x}\|^2 &\leq \prod_{i=0}^{k}(1+\gamma_i)\|x_0 - \bar{x}\|^2 \\ &\leq M\|x_0 - \bar{x}\|^2.\end{aligned}$$

This gives us

$$\|x_{k+1} - \bar{x}\| \leq \sqrt{M}\|x_0 - \bar{x}\|.$$

If $\|x_0 - \bar{x}\| < \frac{r}{\sqrt{M}}$, then $\|x_{k+1} - \bar{x}\| < \sqrt{M} \cdot \frac{r}{\sqrt{M}} = r$. This completes the proof.

■

Fact 5.17. [14, Proposition 8.46(f)] *Assume that f is prox-bounded. Then points at which f has proximal subgradients are the points that are λ-proximal for some $\lambda > 0$, i.e., belong to $\cup_{\lambda>0} \operatorname{ran} \operatorname{prox}_{\lambda f}$. That is,* $\operatorname{dom} \partial_P f = \cup_{\lambda>0} \operatorname{ran}(\operatorname{prox}_{\lambda f})$. *Indeed, if $x \in \operatorname{prox}_{\lambda f}(w)$, the vector $v = \frac{w-x}{\lambda}$ is a proximal subgradient at x and satisfies*

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{1}{2\lambda}\|x' - x\|^2 \tag{5.25}$$

for all $x' \in \mathbb{R}^n$.

Lemma 5.18. *Let $\bar{x}$ satisfy $0 \in \partial f(\bar{x})$, $0 \in \partial_P g(\bar{x})$. Then the following hold.*

(i) *If f is convex, then $\bar{x} = \operatorname{prox}_{\lambda f}(\bar{x})$, $\forall \lambda > 0$.*

(ii) *If g is prox-bounded, then* $\exists\bar{\lambda} > 0$ *such that* $\forall 0 < \lambda < \bar{\lambda}$*, we have*

$$\bar{x} \in \operatorname{prox}_{\lambda g}(\bar{x}). \tag{5.26}$$

When $\operatorname{prox}_{\lambda g}$ *is singleton around* $\bar{x}$*, we have* $\bar{x} = \operatorname{prox}_{\lambda g}(\bar{x})$.

Proof. (i): Since $0 \in \partial f(\bar{x})$, by Fact 2.55, $\bar{x}$ minimizes f. So $f(\bar{x}) \leq f(x)$ for any $x \in \mathbb{R}^n$. Thus, $\forall \lambda > 0$,

$$f(x) + \frac{1}{2\lambda}\|x - \bar{x}\|^2 \geq f(\bar{x}) + \frac{1}{2\lambda}\|\bar{x} - \bar{x}\|^2.$$

Hence, $\bar{x}$ minimizes the function $f(x) + \frac{1}{2\lambda}\|x - \bar{x}\|^2$ for any x. It follows that $\bar{x} = \operatorname{prox}_{\lambda f}(\bar{x})$.

(ii): By Fact 5.17, $\exists\bar{\lambda} > 0$ such that

$$\bar{x} \in \operatorname{prox}_{\bar{\lambda} g}(\bar{x} + \bar{\lambda} \cdot 0) = \operatorname{prox}_{\bar{\lambda} g}(\bar{x}). \tag{5.27}$$

More specifically,

$$\begin{aligned} f(x) &\geq g(\bar{x}) + \langle 0, x - \bar{x}\rangle - \frac{1}{2\bar{\lambda}}\|x - \bar{x}\|^2 \\ &= g(\bar{x}) - \frac{1}{2\bar{\lambda}}\|x - \bar{x}\|^2, \end{aligned}$$

for all $x \in \mathbb{R}^n$. For $0 < \lambda < \bar{\lambda}$, we have $\frac{1}{\lambda} > \frac{1}{\bar{\lambda}}$, so

$$g(x) \geq g(\bar{x}) - \frac{1}{2\lambda}\|x - \bar{x}\|^2 \quad \forall x \in \mathbb{R}^n,$$

so

$$g(x) + \frac{1}{2\lambda}\|x - \bar{x}\|^2 \geq g(\bar{x}) \quad \forall x \in \mathbb{R}^n.$$

Hence

$$\bar{x} \in \operatorname{prox}_{\lambda g}(\lambda \cdot 0 + \bar{x}) = \operatorname{prox}_{\lambda g}(\bar{x}).$$

Therefore, when $\operatorname{prox}_{\lambda g}$ is singleton around $\bar{x}$, we have $\bar{x} = \operatorname{prox}_{\lambda g}(\bar{x})$.

■

Remark 5.19. Notice that by Proposition 5.9(i), when g is prox-regular at $\bar{x}$, $\operatorname{prox}_{\lambda g}$ is singleton near $\bar{x}$. Thus, by Lemma 5.18(ii), if $0 \in \partial_P g(\bar{x})$, $\bar{x} = \operatorname{prox}_{\lambda g}(\bar{x})$.

Now let us see the main result and its proof. Part of the proof is motivated by Moudafi and Thakur [12].

Theorem 5.20. *Assume that f is a proper convex lower-semicontinuous function, g is locally Lipschitz at $A\bar{x}$, prox-bounded and prox-regular at $A\bar{x}$ for $\bar{v} = 0$ with $\bar{x}$ a point which solves* (5.1)*, and A a bounded linear operator which is surjective.*

If the parameters satisfy the following conditions: $\sum_{k=0}^{\infty} \lambda_k < +\infty$ and $\inf_k \rho_k(4 - \rho_k) > 0$, and if $\|x_0 - \bar{x}\|$ is small enough, then the sequence (x_k) converges to a solution of (5.1).

Proof. Let z be any point in the solution set S. Since $0 \in \partial f(z)$, by Lemma 5.18(i) $\operatorname{prox}_{\lambda f}(z) = z$, $\forall \lambda > 0$.

$$\begin{aligned} &\|x_{k+1} - z\|^2 \\ &= \|\operatorname{prox}_f(x_k - \mu_k A^*(\mathrm{Id} - \operatorname{prox}_{\lambda_k g})(Ax_k)) - z\|^2 \\ &= \|\operatorname{prox}_f(x_k - \mu_k \hat{h}(x_k)) - \operatorname{prox}_f z\|^2. \end{aligned}$$

Now since prox_f is firmly nonexpansive,

$$\begin{aligned}
&\|\operatorname{prox}_f(x_k - \mu_k \hat{h}(x_k)) - \operatorname{prox}_f z\|^2 \\
&\leq \|x_k - \mu_k \hat{h}(x_k) - z\|^2 - \|(\mathrm{Id} - \operatorname{prox}_f)(x_k - \mu_k \hat{h}(x_k)) - (\mathrm{Id} - \operatorname{prox}_f)z\|^2 \\
&= \|(x_k - z) - \mu_k \hat{h}(x_k)\|^2 - \|(\mathrm{Id} - \operatorname{prox}_f)(x_k - \mu_k \hat{h}(x_k))\|^2 \\
&= \|x_k - z\|^2 + \mu_k^2 \|\hat{h}(x_k)\|^2 - 2\mu_k \langle \hat{h}(x_k), x_k - z\rangle \\
&- \|(\mathrm{Id} - \operatorname{prox}_f)(x_k - \mu_k \hat{h}(x_k))\|^2.
\end{aligned}$$

Since $z \in S$, so $0 \in \partial_P g(Az)$. As g is prox-regular, by Proposition 5.9(i) $\operatorname{prox}_{\lambda g}$ is singleton. Thus, by Lemma 5.18(ii) $\operatorname{prox}_{\lambda_k g} Az = Az$. Hence,

$$\begin{aligned}
&\langle \hat{h}(x_k), x_k - z\rangle \\
&= \langle A^*(\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k, x_k - z\rangle \\
&= \langle (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k, Ax_k - Az\rangle \\
&= \langle (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k - (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Az, Ax_k - Az\rangle.
\end{aligned}$$

Then by Lemma 5.11,

$$\begin{aligned}
&\langle (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k - (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Az, Ax_k - Az\rangle \\
&\geq \|(\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k - (\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Az\|^2 - \frac{\lambda_k r}{(1-\lambda_k r)^2}\|Ax_k - Az\|^2 \\
&= \|(\mathrm{Id} - \operatorname{prox}_{\lambda_k g})Ax_k\|^2 - \frac{\lambda_k r}{(1-\lambda_k r)^2}\|Ax_k - Az\|^2 \\
&= 2h(x_k) - \frac{\lambda_k r}{(1-\lambda_k r)^2}\|Ax_k - Az\|^2 \\
&\geq 2h(x_k) - \frac{\lambda_k r\|A\|^2}{(1-\lambda_k r)^2}\|x_k - z\|^2.
\end{aligned}$$

So

$$\begin{aligned}
&\|x_k - z\|^2 + \mu_k^2\|\hat{h}(x_k)\|^2 - 2\mu_k\langle\hat{h}(x_k), x_k - z\rangle \\
&- \|(\mathrm{Id} - \mathrm{prox}_f)(x_k - \mu_k\hat{h}(x_k))\|^2 \\
&\leq \|x_k - z\|^2 + \mu_k^2\|\hat{h}(x_k)\|^2 - 2\mu_k\left(2h(x_k) - \frac{\lambda_k r\|A\|^2}{(1-\lambda_k r)^2}\|x_k - z\|^2\right) \\
&- \|(\mathrm{Id} - \mathrm{prox}_f)(x_k - \mu_k\hat{h}(x_k))\|^2 \\
&= \|x_k - z\|^2 + 2\mu_k\frac{\lambda_k r\|A\|^2}{(1-\lambda_k r)^2}\|x_k - z\|^2 - 4\mu_k h(x_k) + \mu_k^2\|\hat{h}(x_k)\|^2 \\
&- \|(\mathrm{Id} - \mathrm{prox}_f)(x_k - \mu_k\hat{h}(x_k))\|^2 \\
&\leq \|x_k - z\|^2 + 2\rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\frac{\lambda_k r\|A\|^2}{(1-\lambda_k r)^2}\|x_k - z\|^2 - \rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \\
&- \|(\mathrm{Id} - \mathrm{prox}_f)(x_k - \rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\|^2 \\
&\leq \left(1 + 2\rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\frac{\lambda_k r\|A\|^2}{(1-\lambda_k r)^2}\right)\|x_k - z\|^2 - \rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \\
&- \|(\mathrm{Id} - \mathrm{prox}_f)(x_k - \rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\|^2.
\end{aligned}$$

Since A is surjective, A has full range, which means $\mathrm{ran}\, A$ is closed. Hence by Fact 5.14 and Fact 2.19(iii), $\exists \gamma \in \mathbb{R}_{++}$ such that

$$\|A^* y\| \geq \gamma\|y\| \quad \forall y \in \mathbb{R}^n.$$

Hence,

$$\begin{aligned}
\frac{2h(x_k)}{\|\hat{h}(x_k)\|^2} &= \frac{\|(\mathrm{Id} - \mathrm{prox}_{\lambda_k g})(Ax_k)\|^2}{\|A^*(\mathrm{Id} - \mathrm{prox}_{\lambda_k g})(Ax_k)\|^2} \\
&\leq \frac{\|(\mathrm{Id} - \mathrm{prox}_{\lambda_k g})Ax_k\|^2}{\gamma^2\|(\mathrm{Id} - \mathrm{prox}_{\lambda_k g}\, Ax_k)\|^2} \leq \frac{1}{\gamma^2}.
\end{aligned}$$

Conditions on the parameters λ_k and ρ_k assure the existence of a positive constant M such that

$$\begin{aligned}\|x_{k+1}-z\|^2 \leq &(1+M\lambda_k)\|x_k-z\|^2 - \rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2}\\ &- \|(\mathrm{Id}-\mathrm{prox}_f)(x_k-\rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\|^2, \end{aligned} \tag{5.28}$$

which gives

$$\|x_{k+1}-z\|^2 \leq \|x_k-z\|^2 + M\lambda_k\|x_k-z\|^2. \tag{5.29}$$

By Lemma (5.13) we have that the sequence $(\|x_k-z\|^2)$ is convergent, which implies it is bounded. Set $\epsilon_k = M\lambda_k\|x_k-z\|^2$, by (2.57) we can see that the sequence $(x_k)_{k\geq 0}$ is quasi-Fejér of Type III. Then by Theorem 2.81, to prove $(x_k)_{k\geq 0}$ converges to a point in solution set S, we just need to show every cluster point of $(x_k)_{k\geq 0}$ is in S. By (5.28) we have

$$\begin{aligned}&\rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} + \|(\mathrm{Id}-\mathrm{prox}_f)(x_k-\rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\|^2\\ &\leq (1+M\lambda_k)\|x_k-z\|^2 - \|x_{k+1}-z\|^2. \end{aligned} \tag{5.30}$$

By assumption, we have $\sum_{k=0}^{\infty}\lambda_k < +\infty$, which implies $\lambda_k \to 0$. Hence $M\lambda_k \to 0$. Therefore,

$$(1+M\lambda_k)\|x_k-z\|^2 - \|x_{k+1}-z\|^2 \to 0. \tag{5.31}$$

So

$$\rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} + \|(\mathrm{Id}-\mathrm{prox}_f)(x_k - \rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\|^2 \to 0. \tag{5.32}$$

So

$$\rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \to 0, \tag{5.33}$$

and

$$\left\|(\mathrm{Id}-\mathrm{prox}_f)(x_k - \rho_k\frac{h(x_k)}{\|\hat{h}(x_k)\|^2}\hat{h}(x_k))\right\|^2 \to 0. \tag{5.34}$$

From (5.33), we have $\exists\varepsilon > 0$ such that

$$\varepsilon\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \le \rho_k(4-\rho_k)\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \to 0. \tag{5.35}$$

Thus

$$\frac{h^2(x_k)}{\|\hat{h}(x_k)\|^2} \to 0. \tag{5.36}$$

By (5.30) we have

$$\|x_{k+1} - z\|^2 \le (1 + M\lambda_k)\|x_k - z\|^2. \tag{5.37}$$

By assumption $\bar{x} \in S$, so

$$\|x_{k+1} - \bar{x}\|^2 \le (1 + M\lambda_k)\|x_k - \bar{x}\|^2. \tag{5.38}$$

Since $\sum_{k=0}^{\infty}\lambda_k < +\infty$ and M is a constant. So $\sum_{k=0}^{\infty}M\lambda_k < +\infty$. Thus by

Lemma 5.16 (i) we have

$$\|x_{k+1} - \bar{x}\|^2 \leq \left[\prod_{i=0}^{k-1}(1 + M\lambda_i)\right] \|x_0 - \bar{x}\|^2. \tag{5.39}$$

Let $\lambda > 0$ be a constant. By assumption we know $\|x_0 - \bar{x}\|$ is small enough. So if

$$\|x_0 - \bar{x}\| < \frac{\lambda}{\sqrt{\prod_{i=0}^{\infty}(1 + M\lambda_i)}}, \tag{5.40}$$

then by Lemma 5.16 (iii) we have $\|x_k - \bar{x}\| < \lambda$ for all $k \geq 0$. So we have $x_k \in U_\lambda$ for all $k \geq 0$, where U_λ is a neighborhood of $\bar{x}$. Since $(x_k)_{k\geq 0}$ is bounded, by Lemma 5.12 we have

$$\|\hat{h}(x_k)\| = \|\hat{h}(x_k) - \hat{h}(\bar{x})\| \leq \frac{2 - \lambda_k r}{1 - \lambda_k r}\|A\|^2 \|x_k - \bar{x}\|. \tag{5.41}$$

Thus, $\|\hat{h}(x_k)\|$ is bounded. By (5.36), we have $\frac{h(x_k)}{\|\hat{h}(x_k)\|} \to 0$. Then

$$h(x_k) = \frac{h(x_k)}{\|\hat{h}(x_k)\|}\|\hat{h}(x_k)\| \to 0, \tag{5.42}$$

i.e.,

$$\|Ax_k - \operatorname{prox}_{\lambda_k g}(Ax_k)\| \to 0. \tag{5.43}$$

Let x^* be a cluster point of (x_k), there exist a subsequence (x_{k_l}) which converges to x^*. Then $Ax_{k_l} \to Ax^*$. Since $\|Ax_k - \operatorname{prox}_{\lambda_k g} Ax_k\| \to 0$ by (5.43), $\operatorname{prox}_{\lambda_{k_l} g} Ax_{k_l} \to Ax^*$. By the assumption, we have

$$\lim_{l\to\infty} \frac{Ax_{k_l} - \operatorname{prox}_{\lambda_{k_l} g}(Ax_{k_l})}{\lambda_{k_l}} = 0.$$

As by Fact 5.4,

$$\frac{Ax_{k_l} - \operatorname{prox}_{\lambda_{k_l} g}(Ax_{k_l})}{\lambda_{k_l}} \in \partial_P g(\operatorname{prox}_{\lambda_{k_l} g}(Ax_k)),$$

hence, $0 \in \tilde{\partial} g(Ax^*)$. Therefore, by Lemma 5.2, we have $0 \in \partial_P g(Ax^*)$.

Since $\frac{h(x_{k_l})}{\|\hat{h}(x_{k_l})\|} \to 0$ and $x_{k_l} \to x^*$,

$$\begin{aligned}
&\|(\operatorname{Id} - \operatorname{prox}_f)(x_{k_l} - \rho_{k_l} \frac{h(x_{k_l})}{\|\hat{h}(x_{k_l})\|^2} \hat{h}(x_{k_l})) - (\operatorname{Id} - \operatorname{prox}_f)(x^*)\| \\
&\leq \|x_{k_l} - \rho_{k_l} \frac{h(x_{k_l})}{\|\hat{h}(x_{k_l})\|^2} \hat{h}(x_{k_l}) - x^*\| \\
&\leq \|x_{k_l} - x^*\| + \rho_{k_l} \frac{h(x_{k_l})}{\|\hat{h}(x_{k_l})\|^2} \|\hat{h}(x_{k_l})\| \\
&= \|x_{k_l} - x^*\| + \rho_{k_l} \frac{h(x_{k_l})}{\|\hat{h}(x_{k_l})\|} \\
&\to 0 + 0 = 0.
\end{aligned}$$

Thus,

$$\|(\operatorname{Id} - \operatorname{prox}_f)(x_k - \rho_k \frac{h(x_k)}{\|\hat{h}(x_k)\|^2} \hat{h}(x_k))\|^2 \to \|(\operatorname{Id} - \operatorname{prox}_f)x^*\|^2. \tag{5.44}$$

Therefore, by (5.34)

$$\begin{aligned}
0 &\leq \|(\operatorname{Id} - \operatorname{prox}_f)x^*\|^2 \\
&\leq \lim_{k \to +\infty} \|(\operatorname{Id} - \operatorname{prox}_f)(x_k - \rho_k \frac{h(x_k)}{\|\hat{h}(x_k)\|^2} \hat{h}(x_k))\|^2 = 0.
\end{aligned} \tag{5.45}$$

So we have $0 \in \partial f(x^*)$. So $x^* \in S$. ■

Remark 5.21. A linear operator between two finite dimensional spaces is always bounded. Therefore, in Theorem 5.20 one only needs A to be surjective in finite dimensional spaces.

We end up with a result concerning (5.2).

Proposition 5.22. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then the following statements are equivalent:*

(i) *If $x \to \bar{x}$, $x - \operatorname{prox}_{\lambda f}(x) \to 0$ and $\lambda \to 0$, then $\frac{\|x-\operatorname{prox}_{\lambda f}(x)\|}{\lambda} \to 0$.*

(ii) *If $x \to \bar{x}$, $x - \operatorname{prox}_{\lambda f}(x) \to 0$ and $\lambda \to 0$, then $\nabla f(\bar{x}) = 0$.*

Proof. (i)⇒(ii): Suppose $x \to \bar{x}$, $x - \operatorname{prox}_{\lambda f}(x) \to 0$ and $\lambda \to 0$. By (i), we have

$$\frac{\|x - \operatorname{prox}_{\lambda f}(x)\|}{\lambda} \to 0.$$

As

$$\frac{x - \operatorname{prox}_{\lambda f}(x)}{\lambda} \in \partial_P f(\operatorname{prox}_{\lambda f}(x)),$$

hence, $0 \in \partial f(\bar{x})$. As f is continuously differentiable, we have $\partial f(x) = \{\nabla f(x)\}$. Therefore, $\nabla f(\bar{x}) = 0$.

(ii)⇒(i): Assume $\nabla f(\bar{x}) = 0$. As

$$e_\lambda f(x) = f(\operatorname{prox}_{\lambda f}(x)) + \frac{1}{2\lambda}\|x - \operatorname{prox}_{\lambda f} x\|^2,$$

we have

$$\begin{aligned} 0 &\le \frac{1}{2}\frac{\|x - \operatorname{prox}_{\lambda f} x\|}{\lambda} \\ &= \frac{e_\lambda f(x) - f(\operatorname{prox}_{\lambda f}(x))}{\|x - \operatorname{prox}_{\lambda f}(x)\|} \\ &\le \frac{f(x) - f(\operatorname{prox}_{\lambda f}(x))}{\|x - \operatorname{prox}_{\lambda f}(x)\|} \quad \text{(since } e_\lambda f \le f) \end{aligned}$$

$$= \frac{f(x) - f(\text{prox}_{\lambda f}(x)) - \langle \nabla f(\bar{x}), x - \text{prox}_{\lambda f}(x) \rangle}{\|x - \text{prox}_{\lambda f}(x)\|}$$

$$\to 0,$$

since f is continuously differentiable and $\nabla f(\bar{x}) = 0$. Hence,

$$\lim_{x \to \bar{x}, \lambda \to 0} \frac{x - \text{prox}_{\lambda f}(x)}{\lambda} = 0.$$

■

5.5 A Numerical Example

Now we will implement Algorithm 4 on a simple example where

$$f(x) = 0, \tag{5.46}$$

and

$$g(x) = \begin{cases} x(1-x), & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.47}$$

To do this, we need the following proposition.

Proposition 5.23.

$$\partial_P g(x) = \begin{cases} \{0\}, & \textit{if } x < 0, \\ [0,1], & \textit{if } x = 0, \\ \{1-2x\}, & \textit{if } 0 < x < 1, \\ [-1,0], & \textit{if } x = 1, \\ \{0\}, & \textit{if } x > 1. \end{cases} \tag{5.48}$$

Proof. Let $v \in \partial_P g(\bar{x})$.

Case $\bar{x} < 0$:

$$\begin{aligned} v(x-\bar{x}) &\leq g(x) - g(\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ &= g(x) - 0 + \frac{r}{2}(x-\bar{x})^2, \\ &= g(x) + \frac{r}{2}(x-\bar{x})^2. \end{aligned} \tag{5.49}$$

When $|x-\bar{x}| < \delta$, $x-\bar{x} > 0$, $g(x) = 0$, by (5.49)

$$\begin{aligned} v(x-\bar{x}) &\leq g(x) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\leq 0 + \frac{r}{2}(x-\bar{x})^2, \\ v &\leq \frac{r}{2}(x-\bar{x}). \end{aligned}$$

When $x \to \bar{x}$, we have $v \leq 0$.

When $|x-\bar{x}| < \delta$, $x-\bar{x} < 0$, $g(x) = 0$ by (5.49)

$$\begin{aligned} v(x-\bar{x}) &\leq g(x) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\leq 0 + \frac{r}{2}(x-\bar{x})^2, \\ v &\geq \frac{r}{2}(x-\bar{x}). \end{aligned}$$

When $x \to \bar{x}$, we have $v \geq 0$. Therefore $v = 0$. Because g is twice continuously differentiable, we know $\partial_P g(x) = \{g'(x)\}$. Therefore $\partial_P g(\bar{x})$ is not empty.

Case $\bar{x} = 0$:

$v \in \partial_P g(\bar{x}) \Longleftrightarrow \exists r \geq 0,\ \delta > 0$ such that when $|x - \bar{x}| < \delta$,

$$\begin{aligned} v(x-\bar{x}) &\leq g(x) - g(\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ vx &\leq g(x) - 0 + \frac{r}{2}x^2, \\ vx &\leq g(x) + \frac{r}{2}x^2. \end{aligned} \tag{5.50}$$

When $|x - \bar{x}| < \delta$, $x - \bar{x} > 0$, $g(x) = x(1-x)$, by (5.50)

$$\begin{aligned} vx &\leq g(x) + \frac{r}{2}x^2, \\ vx &\leq x(1-x) + \frac{r}{2}x^2, \\ v &\leq 1 + (\frac{r}{2} - 1)x. \end{aligned}$$

When $x \to \bar{x}$, we have $v \leq 1$.

When $|x - \bar{x}| < \delta$, $x - \bar{x} < 0$, $g(x) = 0$, by (5.50)

$$\begin{aligned} vx &\leq g(x) + \frac{r}{2}x^2, \\ vx &\leq 0 + \frac{r}{2}x^2, \\ v &\geq \frac{r}{2}x. \end{aligned}$$

When $x \to \bar{x}$, we have $v \geq 0$. Therefore $v \in [0,1]$. That is $\partial_P g(0) \subseteq [0,1]$. Moreover, for $v \in [0,1]$, letting $r = 2$, we have

$$\begin{aligned} \langle v, x-0 \rangle &\leq x - x^2 - 0 + x^2, \quad \text{when } 0 \leq x < 1, \\ \langle v, x-0 \rangle &\leq 0 - 0 + x^2, \quad \text{when } x \geq 1, \\ \langle v, x-0 \rangle &\leq 0 - 0 + x^2, \quad \text{when } x < 0. \end{aligned}$$

Hence, $[0,1] \subseteq \partial_P g(0)$.

Case $0 < \bar{x} < 1$:

$$\begin{aligned} v(x-\bar{x}) &\le g(x) - g(\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\le g(x) - \bar{x}(1-\bar{x}) + \frac{r}{2}(x-\bar{x})^2. \end{aligned} \tag{5.51}$$

When $|x-\bar{x}| < \delta$, $x - \bar{x} > 0$, by (5.51)

$$\begin{aligned} v(x-\bar{x}) &\le x(1-x) - \bar{x}(1-\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\le (x-\bar{x}) - (x+\bar{x})(x-\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v &\le 1 - x - \bar{x} + \frac{r}{2}(x-\bar{x}). \end{aligned}$$

When $x \to \bar{x}$, we have $v \le 1 - 2\bar{x}$.

When $|x-\bar{x}| < \delta$, $x - \bar{x} < 0$, by (5.51)

$$\begin{aligned} v(x-\bar{x}) &\le x(1-x) - \bar{x}(1-\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\le (x-\bar{x}) - (x+\bar{x})(x-\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v &\ge 1 - x - \bar{x} + \frac{r}{2}(x-\bar{x}). \end{aligned}$$

When $x \to \bar{x}$, we have $v \ge 1 - 2\bar{x}$. Therefore $v = 1 - 2\bar{x}$.

Case $\bar{x} = 1$:

$v \in \partial_P g(\bar{x}) \iff \exists r \ge 0$, $\delta > 0$ such that when $|x - \bar{x}| < \delta$,

$$\begin{aligned} v(x-\bar{x}) &\le g(x) - g(\bar{x}) + \frac{r}{2}(x-\bar{x})^2, \\ v(x-1) &\le g(x) - 0 + \frac{r}{2}(x-1)^2, \\ v(x-1) &\le g(x) + \frac{r}{2}(x-1)^2. \end{aligned} \tag{5.52}$$

When $|x-\bar{x}|<\delta$, $x-\bar{x}>0$, $g(x)=0$, by (5.52)

$$\begin{aligned} v(x-1) &\leq g(x)+\frac{r}{2}(x-1)^2, \\ v(x-1) &\leq 0+\frac{r}{2}(x-1)^2, \\ v &\leq \frac{r}{2}(x-1). \end{aligned}$$

When $x\to\bar{x}$, we have $v\leq 0$.

When $|x-\bar{x}|<\delta$, $x-\bar{x}<0$, $g(x)=x(1-x)$, by (5.52)

$$\begin{aligned} v(x-1) &\leq g(x)+\frac{r}{2}(x-1)^2, \\ v(x-1) &\leq x(1-x)+\frac{r}{2}(x-1)^2, \\ v &\geq -x+\frac{r}{2}(x-1). \end{aligned}$$

When $x\to\bar{x}$, we have $v\geq -1$. Therefore $v\in[-1,0]$.

Case $\bar{x}>1$:

$$\begin{aligned} v(x-\bar{x}) &\leq g(x)-g(\bar{x})+\frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\leq g(x)-0+\frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\leq g(x)+\frac{r}{2}(x-\bar{x})^2. \end{aligned} \tag{5.53}$$

When $|x-\bar{x}|<\delta$, $x-\bar{x}>0$, $g(x)=0$, by (5.53)

$$\begin{aligned} v(x-\bar{x}) &\leq g(x)+\frac{r}{2}(x-\bar{x})^2, \\ v(x-\bar{x}) &\leq 0+\frac{r}{2}(x-\bar{x})^2, \\ v &\leq \frac{r}{2}(x-\bar{x}). \end{aligned}$$

When $x \to \bar{x}$, we have $v \leq 0$.

When $|x - \bar{x}| < \delta$, $x - \bar{x} < 0$, $g(x) = 0$, by (5.53)

$$v(x - \bar{x}) \leq g(x) + \frac{r}{2}(x - \bar{x})^2,$$
$$v(x - \bar{x}) \leq 0 + \frac{r}{2}(x - \bar{x})^2,$$
$$v \geq \frac{r}{2}(x - \bar{x}).$$

When $x \to \bar{x}$, we have $v \geq 0$. Therefore $v = 0$.

■

Let $A = \mathrm{Id}$, $\rho_k = 2$. Then

$$\mu_k = \frac{\rho_k h(x_k)}{\|\hat{h}(x_k)\|^2} = \frac{2 \cdot \frac{1}{2}\|(\mathrm{Id} - \mathrm{prox}_{\lambda_k g})x\|^2}{\|(\mathrm{Id} - \mathrm{prox}_{\lambda_k g})x\|^2} = 1.$$

Let $\lambda_k = (\frac{1}{4})^{k+1}$. Then $\sum_{k=0}^{\infty} \lambda_k = \frac{4}{3}$.

Fact 5.24. *Let $C \in \mathbb{R}$ be constant and $f(x) \equiv C$. Then* $\mathrm{prox}_f(x) = \{x\}$.

Proof.

$$\mathrm{prox}_f(x) = \underset{y \in \mathbb{R}}{\mathrm{argmin}}\{C + \frac{1}{2}\|x - y\|^2\} = \{x\}. \tag{5.54}$$

■

Since $f(x) = 0$, so by Fact 5.24 (5.11) reduces to

$$x_{k+1} = \mathrm{prox}_{\lambda_k g} x_k. \tag{5.55}$$

Proposition 5.25. *Let* $0 < \lambda < \frac{1}{2}$. *Then*

$$\operatorname{prox}_{\lambda g}(x) = \begin{cases} 0, & 0 < x \leq \lambda, \\ \frac{x-\lambda}{1-2\lambda}, & \lambda < x < 1-\lambda, \\ 1, & 1-\lambda \leq x < 1 \\ x, & \textit{otherwise}. \end{cases} \tag{5.56}$$

Proof. If $x \leq 0$, or $x \geq 1$, then $\operatorname{prox}_{\lambda g} x = x$. Now let us consider the case $0 < x < 1$.

If $y \leq 0$, then $g(y) = 0$.

$$\begin{aligned} &\inf_{y\leq 0}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} \\ =&\inf_{y\leq 0}\{\frac{1}{2\lambda}(x-y)^2\} \\ =&\frac{1}{2\lambda}x^2. \end{aligned}$$

So $\operatorname{argmin}_{y\leq 0}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = 0$.

If $y \geq 1$, then $g(y) = 0$.

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

So $\operatorname{argmin}_{y\geq 1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = 1$.

If $0 < y < 1$, then $g(y) = y(1-y)$.

$$\inf_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\}$$

$$= \inf_{0<y<1}\{y(1-y) + \frac{1}{2\lambda}(x-y)^2\}$$
$$= \inf_{0<y<1}\{\frac{1-2\lambda}{2\lambda}y^2 + \frac{\lambda - x}{\lambda}y + \frac{1}{2\lambda}x^2\}.$$

Let

$$l(y) = \frac{1-2\lambda}{2\lambda}y^2 + \frac{\lambda - x}{\lambda}y + \frac{1}{2\lambda}x^2.$$

Then l is a quadratic function of y. As $0 < \lambda < \frac{1}{2}$, $\frac{1-2\lambda}{2\lambda} > 0$. Since l is convex, it attains its minimum at

$$-\frac{1}{2}\frac{\frac{\lambda-x}{\lambda}}{\frac{1-2\lambda}{2\lambda}} = \frac{x-\lambda}{1-2\lambda}.$$

Thus, if $0 < x \leq \lambda$, then $\frac{x-\lambda}{1-2\lambda} \leq 0$. Hence l is increasing on $]0,1[$. Thus,

$$\operatorname*{argmin}_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = 0,$$

$$\inf_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = g(0) + \frac{1}{2\lambda}(x-0)^2 = \frac{1}{2\lambda}x^2.$$

Since $x \leq \lambda < \frac{1}{2}$, so $\frac{1}{2\lambda}x^2 < \frac{1}{2\lambda}(1-x)^2$. Hence, $\operatorname{prox}_{\lambda g}(x) = 0$.

If $\lambda < x < 1-\lambda$, then $0 < \frac{x-\lambda}{1-2\lambda} < 1$. Hence

$$\operatorname*{argmin}_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = \frac{x-\lambda}{1-2\lambda},$$

and

$$\inf_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\}$$
$$= g(\frac{x-\lambda}{1-2\lambda}) + \frac{1}{2}(x - \frac{x-\lambda}{1-2\lambda})^2$$
$$= \frac{-2x^2 + 2x - \lambda}{2(1-2\lambda)}.$$

Since

$$\frac{1}{2\lambda}x^2 - \frac{-2x^2+2x-\lambda}{2(1-2\lambda)} = \frac{(\lambda-x)^2}{2\lambda(1-2\lambda)} > 0,$$

and

$$\frac{1}{2\lambda}(1-x)^2 - \frac{-2x^2+2x-\lambda}{2(1-2\lambda)} = \frac{(1-x-\lambda)^2}{\lambda(1-2\lambda)} > 0,$$

we have $\text{prox}_{\lambda g}(x) = \frac{x-\lambda}{1-2\lambda}$.

If $1-\lambda \le x < 1$, then $\frac{x-\lambda}{1-2\lambda} \ge 1$. Hence l is decreasing on $]0,1[$. Thus,

$$\operatorname*{argmin}_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = 1,$$

$$\inf_{0<y<1}\{g(y) + \frac{1}{2\lambda}\|x-y\|^2\} = g(1) + \frac{1}{2\lambda}\|x-1\|^2 = \frac{1}{2\lambda}(1-x)^2.$$

Since $x \ge 1-\lambda > \frac{1}{2}$, so $\frac{1}{2\lambda}(1-x)^2 < \frac{1}{2\lambda}x^2$. Hence, $\text{prox}_{\lambda g}(x) = 1$.

■

The function g is prox-regular because $g(x) = \max\{x(x-1), 0\}$ and each function in the brackets is at least twice continuously differentiable, see [14, Proposition 13.34].

Six initial numbers are used in the numerical experiment and the result is shown in Table 5.1. The result shows that if the initial point is near 0.5

Iter	x_k					
0	0.0001	0.25	0.499999	0.5	0.500001	0.75
1	0	0	0.4999980	-	0.5000020	1
2	-	-	0.4999973	-	0.5000026	-
3	-	-	0.4999969	-	0.5000030	-
4	-	-	0.49999967	-	0.5000033	-
5	-	-	0.49999966	-	0.5000034	-
6	-	-	-	-	0.5000034	-
7	-	-	-	-	0.5000034	-
8	-	-	-	-	0.5000034	-

Table 5.1: Numerical result for prox-regular functions.

then it will converges to a point near 0.5. Otherwise, if it's on the left side of 0.5 it will converge to 0, if it's on the right side of 0, it will converge to 1.

Chapter 6

Error Estimation

In this chapter, we estimate $f(x) - \inf f$. It turns out that it is useful for estimating the difference between consecutive iterations.

6.1 Minimization Estimate for a Convex Function

First, we introduce a useful fact about convex functions given by Rockafellar and Wets.

Fact 6.1. [14, Proposition 10.59] *For a proper, lsc, convex function* $f : \mathbb{R}^n \to [-\infty, +\infty]$ *with* $\operatorname{argmin} f \neq \emptyset$ *and any* $x \in \mathbb{R}^n$*, one has*

$$f(x) - \min f \leq \mathrm{d}(x, \operatorname{argmin} f)\, \mathrm{d}(0, \partial f(x)).$$

We'd like to extend Fact 6.1 to some nonconvex functions, and study the function $g(x) = \mathrm{d}(x, \operatorname{argmin} f) \cdot \mathrm{d}(0, \partial f(x))$.

Assume $\xi \geq \min f$. We can also estimate $f(x) - \xi$. The following result is new.

Theorem 6.2. *Let* $f : \mathbb{R}^n \to [-\infty, +\infty]$ *be proper, convex and lsc function. Let* $\xi \in \mathbb{R}$*. Then the following statements holds.*

(i) *If* $\xi = \inf f$ *and* $\operatorname{argmin} f \neq \emptyset$*, then for any* $x \in \mathbb{R}^n$*,*

$$f(x) - \xi \leq \mathrm{d}(x, \mathrm{lev}_{\leq \xi} f)\, \mathrm{d}(0, \partial f(x)). \tag{6.1}$$

(ii) *If* $\xi > \inf f$*, then for any* $x \in \mathbb{R}^n$*,*

$$f(x) - \xi \leq \mathrm{d}(x, \mathrm{lev}_{\leq \xi} f)\, \mathrm{d}(0, \partial f(x)). \tag{6.2}$$

Proof. (i): By Fact 6.1, the proof is done.

(ii): Fix any point $\bar{x} \in \mathbb{R}^n$. Let $\rho = \mathrm{d}(0, \partial f(\bar{x}))$. Let $\bar{v} \in \partial f(\bar{x})$ such that $\|\bar{v}\| = \rho$. Then

$$\bar{v} \in \partial f(\bar{x}) \cap \rho B(0; 1).$$

Let

$$g(x) = f(x) + \rho\|x - \bar{x}\|.$$

Then by Fact 2.57 we have

$$\partial g(\bar{x}) = \partial f(\bar{x}) + \rho B(0; 1).$$

So $0 \in \partial g(\bar{x})$. Therefore, $g(x)$ attains its minimum at $\bar{x}$. So we have

$$(\forall x \in \mathbb{R}^n) \quad g(\bar{x}) \leq g(x). \tag{6.3}$$

Therefore, for every $x \in \mathbb{R}^n$, we have

$$f(\bar{x}) \leq f(x) + \rho\|x - \bar{x}\|. \tag{6.4}$$

Let $z \in \operatorname{lev}_{\leq \xi} f$. Then by (6.4),

$$\begin{aligned} f(\bar{x}) &\leq f(z) + \rho\|z - \bar{x}\| \\ &\leq \xi + \rho\|z - \bar{x}\|. \end{aligned}$$

Taking the infimum over $z \in \operatorname{lev}_{\leq \xi} f$ we obtain

$$f(\bar{x}) - \xi \leq \rho \mathrm{d}(\bar{x}, \operatorname{lev}_{\leq \xi} f). \tag{6.5}$$

So

$$f(\bar{x}) - \xi \leq \mathrm{d}(0, \partial f(\bar{x}))\mathrm{d}(\bar{x}, \operatorname{lev}_{\leq \xi} f). \tag{6.6}$$

■

Corollary 6.3. *Let* $f : \mathbb{R}^n \to [-\infty, +\infty]$ *be proper, convex and lsc with* $\operatorname{argmin} f \neq \emptyset$*. Then for any* $x \in \mathbb{R}^n$,

$$f(x) - \min f \leq \mathrm{d}(x, \operatorname{argmin} f)\, \mathrm{d}(0, \partial f(x)). \tag{6.7}$$

Proof. Let $\xi = \min f$. By Theorem 6.2 we have

$$f(x) - \min f \leq \mathrm{d}(x, \operatorname{argmin} f)\mathrm{d}(0, \partial f(x)). \tag{6.8}$$

■

6.2 Why Do We Consider $f(x) - \xi$?

The error esitimation is related to the subgradient projector. Recall

$$G : \mathcal{H} \to \mathcal{H} : x \mapsto \begin{cases} x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > \xi \\ x, & \text{if } f(x) \leq \xi. \end{cases} \tag{6.9}$$

When $f(x) > \xi$,

$$Gx = x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x) = x - \left(\frac{f(x) - \xi}{\|s(x)\|} \right) \frac{s(x)}{\|s(x)\|}, \tag{6.10}$$

which means with direction $-\frac{s(x)}{\|s(x)\|}$ and stepsize $\frac{f(x)-\xi}{\|s(x)\|}$. Now let us take a look at the stepsize $\frac{f(x)-\xi}{\|s(x)\|}$. By Theorem 6.2 we have

$$\begin{aligned} & f(x) - \xi \\ & \leq \mathrm{d}(x, \mathrm{lev}_{\leq \xi} f) \mathrm{d}(0, \partial f(x)) \\ & \leq \mathrm{d}(x, \mathrm{lev}_{\leq \xi} f) \cdot \|s(x)\|. \end{aligned}$$

So when $f(x) > \xi$, we have

$$\frac{f(x) - \xi}{\|s(x)\|} \leq \mathrm{d}(x, \mathrm{lev}_{\leq \xi} f). \tag{6.11}$$

Therefore, the subgradient projector uses a stepsize related to the error estimation.

6.3 Minimization Estimate for a Nonconvex Function

For a nonconvex function, we can use the Fenchel subdifferential. The following theorem shows that Fact 6.1 works for nonconvex functions as well.

Theorem 6.4. *For a proper, lsc function* $f : \mathbb{R}^n \to [-\infty, +\infty]$ *with* $\operatorname{argmin} f \neq \emptyset$ *and any* $x \in \mathbb{R}^n$*, one has*

$$f(x) - \min f \leq \mathrm{d}(x, \operatorname{argmin} f)\, \mathrm{d}(0, \partial f(x)).$$

Proof. **Proof 1:**

Fix $\bar{x} \in \mathbb{R}^n$, and let $\rho = \mathrm{d}(0, \partial f(\bar{x}))$. The aim is to show $f(\bar{x}) \leq \inf f + \rho \mathrm{d}(\bar{x}, \operatorname{argmin} f)$.

When $\rho = +\infty$, done, so suppose $\rho < +\infty$, i.e., $\partial f(\bar{x}) \neq \emptyset$. Since $\partial f(\bar{x})$ is closed, $\exists \bar{\nu}$ such that

$$\bar{\nu} \in \partial f(\bar{x}),$$

$$\|\bar{\nu}\| = \rho = \mathrm{d}(0, \partial f(\bar{x})).$$

Consider

$$g(x) = f(x) + \rho \|x - \bar{x}\|.$$

As $\bar{\nu} \in \partial f(\bar{x})$,

$$f(x) \geq f(\bar{x}) + \langle \bar{\nu}, x - \bar{x} \rangle \quad \text{for all } x,$$

so

$$f(x) + \langle -\bar{\nu}, x - \bar{x} \rangle \geq f(\bar{x}),$$

$$f(x) + \rho \|x - \bar{x}\| \geq f(\bar{x}), \quad \forall x \in \mathbb{R}^n.$$

Hence,

$$g(x) \geq f(\bar{x}) = g(\bar{x}), \quad \forall x \in \mathbb{R}^n.$$

In particular, $f(x) + \rho\|x - \bar{x}\| \geq f(\bar{x})$ for $x \in \mathrm{argmin} f$. Then

$$\inf f + \rho\|x - \bar{x}\| \geq f(\bar{x}) \quad \forall x \in \mathrm{argmin} f.$$

Taking the infimum over x,

$$\inf f + \rho \inf\{\|x - \bar{x}\| : x \in \mathrm{argmin} f\} \geq f(\bar{x}),$$

which gives

$$\inf f + \rho \mathrm{d}(\bar{x}, \mathrm{argmin} f) \geq f(\bar{x}),$$

as required, because $\rho = \mathrm{d}(0, \partial f(\bar{x}))$, and $\inf f = \min f$.

Proof 2:

Let $x \in \mathbb{R}^n$, $y \in \mathrm{argmin}\, f$ and $v \in \partial f(x)$. By (2.43) we have

$$\begin{aligned} f(y) &\geq f(x) + \langle v, y - x\rangle, \\ f(x) - f(y) &\leq \langle v, x - y\rangle, \\ f(x) - f(y) &\leq \|v\|\|x - y\|, \\ f(x) - \min f &\leq \|v\|\|x - y\|. \end{aligned}$$

Taking the infimum over y,

$$\begin{aligned} f(x) - \min f &\leq \inf_y \|v\|\|x - y\|, \\ f(x) - \min f &\leq \|v\| \mathrm{d}(x, \mathrm{argmin}\, f). \end{aligned}$$

Then taking the infimum over $v \in \partial f(x)$ yields,

$$
\begin{aligned}
f(x) - \min f &\leq \|v\| \mathrm{d}(x, \operatorname{argmin} f), \\
f(x) - \min f &\leq \inf_v \|v\| \mathrm{d}(x, \operatorname{argmin} f), \\
f(x) - \min f &\leq \mathrm{d}(0, \partial f(x)) \mathrm{d}(x, \operatorname{argmin} f).
\end{aligned}
$$

■

6.4 Properties of the Estimation Function

Recall

$$
\begin{aligned}
f(x) - \min f &\leq \mathrm{d}(x, \operatorname{argmin} f) \cdot \mathrm{d}(0, \partial f(x)) \\
&= \mathrm{d}(x, \partial f^{-1}(0)) \cdot \mathrm{d}(0, \partial f(x)).
\end{aligned}
$$

Define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$
g(x) = \mathrm{d}(x, (\partial f)^{-1}(0)) \cdot \mathrm{d}(0, \partial f(x)) \quad \forall x \in \mathbb{R}^n.
$$

We'd like to consider when the function g is convex. To make the analysis simple, let us assume

- $(\partial f)^{-1}(0) = \{0\}$, so f has a unique minimizer at $x = 0$;
- $\partial f(x) = \{f'(x)\}$, which means f is continuously differentiable when $x \neq 0$;
- $f'(x) \geq 0$, hence f is increasing when $x \geq 0$.

(6.12)

Then $g(x) = x \cdot f'(x)$, when $x > 0$.

From now on, we assume that $f : \mathbb{R} \to \mathbb{R}$ has continuous third order derivatives.

Theorem 6.5. *The function g is convex on $[0, +\infty) \Longleftrightarrow 2f''(x)+xf'''(x) \geq 0$, when $x \geq 0$.*

Proof. We have

$$\begin{aligned} g'(x) &= (x \cdot f'(x))' \\ &= f'(x) + x \cdot f''(x), \\ g''(x) &= f''(x) + f''(x) + x \cdot f'''(x) \\ &= 2f''(x) + x \cdot f'''(x). \end{aligned}$$

Therefore,

$$\text{g is convex } \Longleftrightarrow g'' \geq 0, \text{ i.e., } 2f''(x) + xf'''(x) \geq 0 \quad \forall x \geq 0.$$

■

Theorem 6.6. *Assume f is convex, even, differentiable, and has a unique minimizer and f is increasing on $[0, +\infty)$. Then $g(x) = x \cdot f'(x)$ is increasing on $[0, +\infty)$.*

Proof. As f is convex on $[0, +\infty)$, f' is non-decreasing. Hence $f''(x) \geq 0$. Since f' is non-decreasing, $f'(x) \geq f'(0) = 0$, when $x \geq 0$. So $g'(x) = f'(x) + x \cdot f''(x) \geq 0$ on $[0, +\infty)$. Thus $g(x)$ is increasing on $[0, +\infty)$. ■

Theorem 6.7. *The function*

$$g(x) = \mathrm{d}(x, (\partial f)^{-1}(0)) \cdot \mathrm{d}(0, \partial f(x))$$

satisfies:

(i) $g(x) \geq 0$,

(ii) $g(x) > 0$ *if* $x \notin (\partial f)^{-1}(0)$, *i.e.,* $x \notin \operatorname{argmin} f$.

Proof. (i): It's obvious.

(ii): If $x \notin (\partial f)^{-1}(0)$, $\mathrm{d}(x, (\partial f)^{-1}(0)) > 0$, since $(\partial f)^{-1}(0)$ is closed.
Also, $\mathrm{d}(0, \partial f(x)) > 0$, since $\partial f(x)$ is closed.
Hence $g(x) > 0$, if $x \notin (\partial f)^{-1}(0)$. ∎

Theorem 6.8. *If f is even and f is increasing on $[0, +\infty[$, then g is even.*

Proof. By assumption

$$f(x) = f(-x) \quad \forall x \in \mathbb{R},$$

and

$$f'(x) \geq 0 \quad \text{when } x \geq 0.$$

Note that $g(x) = x \cdot f'(x)$ when $x > 0$.
When $x < 0$, we have

$$\begin{aligned} g(x) &= \mathrm{d}(x, 0) \cdot \mathrm{d}(0, \partial f(x)) \\ &= -x \cdot \mathrm{d}(0, -f'(-x)). \end{aligned}$$

Since $f'(-x) \geq 0$, $f(x) = f(-x)$, so $f'(x) = -f'(-x)$. So

$$\begin{aligned} g(x) &= -x \cdot f'(-x) \\ &= g(-x). \end{aligned}$$

Hence g is even. ■

6.5 Examples for the Error Estimation Function

We illustrate our function g by a few examples.

Example 6.9. Let $f(x) = |x|$, which is even on $\mathbb{R}$. Then

$$g(x) = x \cdot 1, \quad \text{if } x \geq 0.$$

$$g(x) = -x \cdot 1, \quad \text{if } x \leq 0.$$

Hence $g(x) = |x|$.

Example 6.10. Let $f(x) = x^2$, which is even on $\mathbb{R}$. Then

$$\begin{aligned} g(x) &= |x| \cdot |2x| \\ &= 2x^2. \end{aligned}$$

Example 6.11. If $f(x) = \frac{1}{p}|x|^p$, $p > 1$, then for $x > 0$

$$f(x) = \frac{1}{p}x^p,$$

$$f'(x) = p \cdot \frac{1}{p}x^{p-1} = x^{p-1},$$

$$f''(x) = (p-1)x^{p-2},$$

$$f'''(x) = (p-1)(p-2)x^{p-3}.$$

Thus,

$$\begin{aligned} & 2f''(x) + xf''' \\ =& 2(p-1)x^{p-2} + x \cdot (p-1)(p-2)x^{p-3} \\ =& 2(p-1)x^{p-2} + (p-1)(p-2)x^{p-2} \\ =& (p-1)x^{p-2}(2+p-2) \\ =& p(p-1)x^{p-2} \\ \geq& 0 \quad (\text{when } x \geq 0). \end{aligned}$$

This means

$$g(x) = |x| \cdot |x^{p-1}| = |x|^p,$$

which is convex.

The three examples above indicate that g is convex. However, is g always convex? We need to consider

$$2f''(x) + xf'''(x) \geq 0 \quad \text{when } x \geq 0. \tag{6.13}$$

If (6.13) fails, then g is not convex.

A good way to make (6.13) fail is letting $f''(x) \geq 0$ be bounded above, f''' be oscillating. The following example is given by Dr. Bauschke.

Example 6.12. Let $k \geq 1$. Define an even function f on $\mathbb{R}$ with

$$f(x) = \frac{x^2}{2} + (-\cos x) + kx \quad \text{when } x \geq 0.$$

Then the function g is not convex.

Proof. We have

$$f''(x) = 1 + \cos x \geq 0.$$

Hence f is convex. We also have

$$f''(x) \leq 2, \quad f'''(x) = -\sin x.$$

Then $2f''(x) + x \cdot f'''(x) = 2(1 + \cos x) + x(-\sin x) < 0$, for some sufficientlly large x, say

$$x = 2n\pi + \frac{\pi}{2} \quad \text{with } n \geq 1.$$

Therefore,

$$\begin{aligned} & 2f''(x) + xf'''(x) \\ =& 2(1 + \cos(2n\pi + \frac{\pi}{2})) + (2n\pi + \frac{\pi}{2})(-1) \\ =& 2 + 0 - 2n\pi - \frac{\pi}{2} \quad (2(1 + \cos x) \leq 4) \\ <& 0. \end{aligned}$$

This means that g is not convex. ■

Remark 6.13. Since

$$f(x) = \frac{x^2}{2} + (-\cos x) + kx \quad \text{for } x \geq 0,$$

$$f'(x) = x + \sin x + k \geq 0,$$

so

$$\begin{aligned} g(x) &= x \cdot (x + \sin x + k) \\ &= x^2 + x \sin x + kx, \end{aligned}$$

where $k \geq 1$.

Example 6.14. Let $k \geq 1$, $f(x) = kx + \frac{x^2}{2} - \sin x$. then

$$g(x) = x \cdot (k + x - \cos x) \quad \text{when } x \geq 0$$

is not convex.

Proof. We have

$$\begin{aligned} &f'(x) = k + x - \cos x \geq 0, \text{ as } k \geq 1, \\ &2 \geq f''(x) = 1 + \sin x \geq 0. \end{aligned}$$

Thus, f is a convex function.

As $f'''(x) = \cos x$, we have

$$\begin{aligned} &2f''(x) + xf'''(x) \\ =&2(1 + \sin x) + x \cdot \cos x \end{aligned}$$

is negative when x is large, e.g., $x = 2k\pi + n\pi$, $k \geq 1$, and $\cos x = -1$.

This means g(x) is not convex on $[0, +\infty)$. ∎

Example 6.15. Let $k \geq \frac{1}{2}$. Define an even function with

$$f(x) = \frac{x^2}{2} + kx - \frac{\cos 2x}{4}, \quad \text{when } x \geq 0.$$

Then

$$g(x) = x \cdot (x + \frac{\sin 2x}{2} + k)$$

is not convex on $[0, +\infty)$.

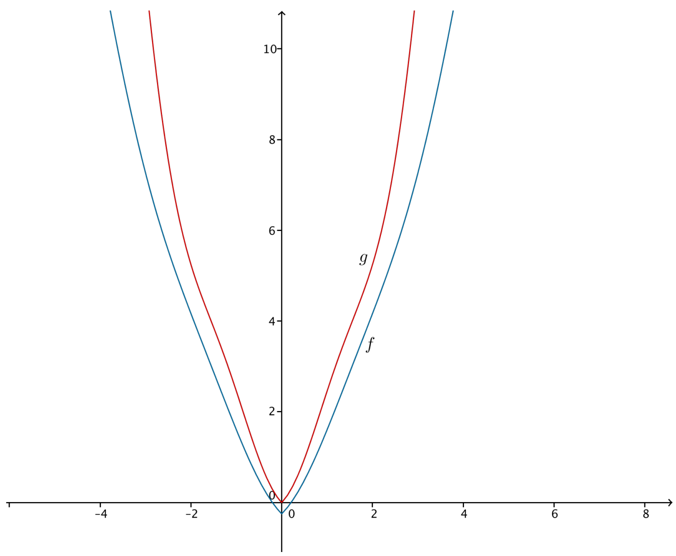

Figure 6.1: Example 6.15 with $k = 1$.

Proof. When $x \geq 0$,

$$f'(x) = x + k + \frac{\sin 2x}{2} \geq 0 \quad \text{when } k \geq \frac{1}{2}.$$

Since

$$2 \geq f''(x) = 1 + \cos 2x \geq 0,$$

So f is convex. We also have

$$f'''(x) = -2 \sin 2x.$$

So

$$\begin{aligned} & 2f''(x) + x \cdot f'''(x) \\ =& 2(1 + \cos 2x) + x \cdot (-2 \sin 2x) \\ =& 2(1 + \cos 2x) - 2x \sin 2x. \end{aligned}$$

Hence, when x is large and $\sin 2x = +1$, e.g.,

$$2x = 2n\pi + \frac{\pi}{2},$$

we have

$$2f''(x) + x \cdot f'''(x) = 2(1 + \cos 2x) - 2x \sin 2x < 0.$$

Hence g is not convex. ■

We can also use a non-trigonometric function with $0 \leq f''(x) \leq M$.

Example 6.16. Let $f(x) = x + e^{-x} - 1$ for $x > 0$. Then

$$g(x) = x \cdot (1 - e^{-x})$$

is not convex.

Proof. We have

$$f'(x) = 1 - e^{-x} \geq 0 \quad \text{if } x \geq 0,$$

$$f''(x) = e^{-x} \geq 0.$$

Hence, f is convex. Now, $f'''(x) = -e^{-x}$, we have

$$2f''(x) + x \cdot f'''(x)$$

$$= e^{-x}(2-x) < 0 \quad \text{when } x > 2.$$

Hence g is not convex. ■

Is there a systematic way to construct a convex function f such that its error estimation function g is not convex? Here is one way to it.

Example 6.17. Define $f : \mathbb{R} \to]-\infty, +\infty]$ by

$$f(x) = \int_0^x (\int_0^u \sqrt{1-s^2}\mathrm{d}s)\mathrm{d}u, \quad \text{for } 0 \le x \le 1,$$

otherwise $f(x) = +\infty$. Then f is convex, but

$$g(x) = x \int_0^x \sqrt{1-s^2}\mathrm{d}s \quad \text{on } [0,1]$$

is not convex.

Proof. We have

$$f'(x) = \int_0^x \sqrt{1-s^2}\ \mathrm{d}s \ge 0,$$

$$f'' = \sqrt{1-x^2} \ge 0, \quad \text{when } 0 \le x \le 1,$$

$$f'''(x) = \frac{-x}{\sqrt{1-x^2}}.$$

Hence f is convex on $\mathbb{R}$. However,

$$\begin{aligned}
&2f''(x) + x \cdot f'''(x) \\
=&2 \cdot \sqrt{1-x^2} + \frac{x \cdot (-x)}{\sqrt{1-x^2}} \\
=&\frac{2(1-x^2) - x^2}{\sqrt{1-x^2}} \\
=&\frac{2-3x^2}{\sqrt{1-x^2}}
\end{aligned}$$

<0

if $x > \sqrt{\frac{2}{3}}$.

Hence g is not convex on $[0, 1]$. ■

6.6 When Does $f(x) = g(x)$?

With f being convex, continuous, $\operatorname{argmin} f \neq \emptyset$, our error estimation gives

$$f(x) \leq g(x) = \mathrm{d}(x, \operatorname{argmin} f)\mathrm{d}(0, \partial(x)).$$

Under assumption (6.12), $g(x) = x \cdot f'(x)$ if $x > 0$. It is natural to ask:

(i) When does $g(x) = f(x)\ \forall x \in \mathbb{R}$?

Now $g(x) = f(x) \Longleftrightarrow xf'(x) = f(x)$. This implies

$$\frac{f'(x)}{f(x)} = \frac{1}{x} \quad (x > 0),$$
$$\ln f(x) = \ln x + k \quad (x > 0),$$
$$f(x) = x \cdot e^k = x \cdot C.$$

Hence we have the following theorem.

Theorem 6.18. *$g(x) = f(x) \Longleftrightarrow f(x) = C|x|$, where $C > 0$.*

(ii) When does $g(x) = kf(x) \quad \forall x \in \mathbb{R}$?

We have

$$xf'(x) = kf(x),$$
$$\frac{f'(x)}{f(x)} = k \cdot \frac{1}{x} \quad (x > 0),$$

$$\ln f(x) = k \ln x + C = \ln x^k + C,$$
$$f(x) = x^k \cdot e^C = x^k \cdot M.$$

where $M > 0$. Therefore, we have the theorem below.

Theorem 6.19.

$$g(x) = kf(x) \quad \forall x \in \mathbb{R} \Longleftrightarrow f(x) = |x|^k \cdot M$$

for some $M > 0$.

Chapter 7

Conclusions and Future Work

In this thesis, the CQ algorithm is studied. This is an algorithm solving the split feasibility problem: find $\hat{x}$ such that

$$\hat{x} \in C \qquad \text{and} \qquad A\hat{x} \in Q.$$

The CQ algorithm can be formed in another way with the subgradient projector

$$G : \mathcal{H} \to \mathcal{H} : x \mapsto \begin{cases} x + \frac{\xi - f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > \xi \\ x, & \text{if } f(x) \leq \xi \end{cases},$$

which shows it is a special case of the subgradient projection algorithm. A new proof of the convergence is given. This thesis shows that the CQ alorithm works for the SFP with prox-regular functions and the SPA works for S-subdifferentiable functions. The error estimation $f(x) - \min f$ gives us the stepsize in the algorithm.

Possible future work:

– Can the original CQ algorithm work for general nonconvex cases?

- Can the CQ alorithm work for two prox-regular functions?
- How can we effectively use error estimation in the CQ algorithm?
- More numerical experiments on the CQ algorithms for nonconvex functions are required.
- Explore the CQ algorithm with subgradients of the Moreau envelope.
- What is the convergence rate when the CQ algorithm works?

Bibliography

[1] Apostol, T. *Calculus, Jon Wiley & Sons*. 1967. → pages 86

[2] Bauschke, H. H., and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011. → pages 8, 14, 20, 21, 22, 23, 26, 29, 33, 34, 49, 91

[3] Byrne, C. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems 18*, 2 (2002), 441. → pages 43, 44, 45

[4] Byrne, C. A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Problems 20*, 1 (2004), 103. → pages 43

[5] Cegielski, A. *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, vol. 2057. Springer Berlin, 2012. → pages 40, 42, 49

[6] Combettes, P. L. Quasi-Fejérian analysis of some optimization algorithms. *Studies in Computational Mathematics 8* (2001), 115–152. → pages 34, 35

[7] Deutsch, F. *Best Approximation in Inner Product Spaces*, vol. 7. Springer, 2001. → pages 5

[8] Kreyszig, E. *Introductory Functional Analysis with Applications*, vol. 81. Wiley New York, 1989. → pages 4, 39

[9] Landweber, L. An iteration formula for Fredholm integral equations of the first kind. *American Journal of Mathematics* (1951), 615–624. → pages 40

[10] López, G., Martín-Márquez, V., Wang, F., and Xu, H.-K. Solving the split feasibility problem without prior knowledge of matrix norms. *Inverse Problems 28*, 8 (2012), 085004. → pages 31, 46, 48

[11] Moudafi, A. A proximal iterative approach to a non-convex optimization problem. *Nonlinear Analysis: Theory, Methods & Applications 72*, 2 (2010), 704–709. → pages 88

[12] Moudafi, A., and Thakur, B. Solving proximal split feasibility problems without prior knowledge of operator norms. *Optimization Letters* (2013), 1–12. → pages 96

[13] Polyak, B. *Introduction to Optimization.* Optimization Software, Publications Division (New York), 1987. → pages 91

[14] Rockafellar, R., and Wets, R. *Variational Analysis*, vol. 317. Springer, 1998. → pages 17, 25, 82, 87, 88, 94, 112, 114

[15] Yang, Q. On variable-step relaxed projection algorithm for variational inequalities. *Journal of Mathematical Analysis and Applications 302*, 1 (2005), 166–179. → pages 45, 46