Splitting Methods in Convex Optimization

by

Xuetong Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2022, 1:23pm

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this report, we study splitting methods for solving optimization problems that are modelled as the sum of two convex functions. Because solving the sum of two functions can sometimes be difficult, splitting methods split the original objective into two parts and solve two separate convex optimization problems, often an easier task. The splitting methods provide an iterative update scheme that deals with the two functions separately. In addition, it can allow a simpler form for the objective function. Moreover, we often need to consider the dual of the original problem in order to solve the problem efficiently. We will compare several popular splitting algorithms in the report, such as the the Forward-Backward method, the Douglas-Rachford method, the Peaceman-Rachford method, and Alternating Direction Method of Multiplier.

Acknowledgements

I would like to sincerely thank Prof. Henry Wolkowicz and Prof. Walaa Moursi. Thank you for guiding my report with patience and extensive knowledge. This work would not have been possible without your help.

Table of Contents

List of Figures

Chapter 1

Introduction

Splitting methods are first order iterative methods that have their roots in partial differential equations. Perhaps the most simple, and yet most famous, instance of these methods dates back to Cauchy and his introduction of the steepest descent method, where one of the functions is set to zero and the other is assumed to be smooth.

Applications of splitting methods in optimization include: image processing, e.g., medical imaging and inverse problems; data science and machine learning e.g., empirical risk minimization, support vector machine, and the least absolute shrinkage and selection operator (LASSO) problems; and physics e.g., computerized tomography and electron microscopy. See, e.g., [3], [5], [6], [9], and the references therein.

Popular splitting algorithms include:

HenryW ♣ no consistency in bold or in using index textdef for the index; when are things in the index? not in the index? why? why are some things bold here? with commas and others not?

the Douglas–Rachford and the Peaceman–Rachford algorithms, e.g., [8], **(projected) gradient methods**; e.g., the celebrated Fast Iterative Shrinkage-Thresholding Algorithm (**FISTA**) [1], the method of **alternating projections** [4], the **Dykstra** algorithm [2] and the popular Alternating Direction Method of Multipliers (**ADMM**) [7].

Chapter 2

Convex Analysis

In this chapter, we will review basic definitions and facts that we need from convex analysis.

HenryW ♣ Define X here ... perhaps as a Euclidean space

2.1 Convex Sets

Definition 2.1 (affine subspace). *Let $C \subseteq X$. Then C is an affine subspace if $C \neq \varnothing$ and*

$$x, y \in C,\ \lambda \in \mathbb{R} \implies \lambda x + (1-\lambda) y \in C.$$

Examples of affine sets are: a point; a line; a plane; a hyperplane.

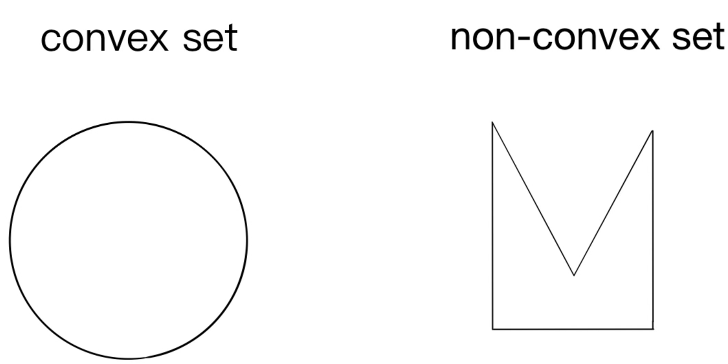

Figure 2.1: Example of convex and non-convex set

Definition 2.2 (convex set)**.** *Let $C \subseteq X$. Then C is convex if*

$$\lambda \in]0,1[,\ x, y \in C \implies \lambda x + (1-\lambda)y \in C.$$

Examples of convex sets in $\mathbb{R}^d$ are: $\varnothing \subseteq \mathbb{R}^d$, a ball $C = \{y \in \mathbb{R}^d \mid \|y - x\| \leq \gamma\}$, an affine subspace, a half-space $C = \{x \in \mathbb{R}^d \mid \langle x, u\rangle \leq \eta\}$ where $u \in \mathbb{R}^d$, $\eta \in \mathbb{R}$ are fixed.

HenryW ♣ if u, η are fixed, what about γ?

Let C be a subset of X. The *affine hull* of C, denoted by aff C, is the intersection of all affine subspaces containing C (smallest affine set containing C). The convex hull of C, denoted by conv C is the intersection of all convex sets containing C (smallest convex set containing C).

HenryW ♣ why are affine/convex hull not in the index?

Theorem 2.3. *The intersection of an arbitrary collection of convex subsets of X is convex.*

Proof. Let I be an index set (not necessarily finite). Let $(C_i)_{i\in\mathbb{I}}$ be a collection of convex subsets of X. Set $C := \bigcap_{i\in\mathbb{I}} C_i$. Let $\lambda \in]0,1[$ and let $(x,y) \in C \times C$. Because each C_i is convex, we learn that $(\forall i \in \mathbb{I})\lambda x + (1-\lambda)y \in C_i$. Hence, $\lambda x + (1-\lambda)y \in \bigcap_{i\in\mathbb{I}} C_i = C$. Thus, C is convex. □

Definition 2.4 (convex combination)**.** *A linear combination $\lambda_1 x_1 + \cdots + \lambda_m x_m$ is called a convex combination of the vectors $x_1, \cdots, x_m$, if $\sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0, \forall i$.*

2.2 Convex Functions

Convex functions are particularly important in the study of optimization problems because they have many convenient properties. In particular, any local minimizer of a convex function is a global minimizer.

HenryW ♣ epi should be epi i.e. it is a math operator also ... if epi, epigraph is entered into the index then also epigraph, epi must be entered

Definition 2.5 (epigraph)**.** *Let $f : X \to]-\infty, \infty]$. The epigraph of f is* $\operatorname{epi}(f) = \{(x,\alpha) \mid f(x) \leq \alpha\} \subseteq X \times \mathbb{R}$.

Definition 2.6 (domain)**.** *Let $f : X \to]-\infty, \infty]$. Then domain of f is* $\operatorname{dom} f = \{x \in X \mid f(x) < +\infty\}$.

Definition 2.7 (proper)**.** *Let* $f : X \to]-\infty, \infty]$. *Then* f *is proper if* $\operatorname{dom} f \neq \varnothing$

HenryW ♣ why was this needed?????????

and $(\forall x \in X)\ f(x) > -\infty$.

Definition 2.8 (convex function)**.** *Let* $f : X \to]-\infty, \infty]$. *Then* f *is convex if*

$$x, y \in X, \lambda \in]0,1[\implies f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Fact 2.9. *Let* $f : X \to]-\infty, \infty]$. *Then* f *is convex if and only if* $\operatorname{epi} f$ *is convex.*

Corollary 2.10. *Let* $f : X \to]-\infty, \infty]$ *be convex. Then* $\operatorname{dom} f$ *is convex.*

Proof. If $\operatorname{dom} f = \varnothing$ then the result is clear.

HenryW ♣ You have sentences that begin with no capitalization ... sentences that do not make any sense.

Now suppose that $x, y \in \operatorname{dom} f$. Let $\lambda \in]0,1[, z = \lambda x + (1-\lambda)y$. Then $f(z) = f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) < +\infty$. Hence, $z \in \operatorname{dom} f$. ☐

Definition 2.11 (indicator function)**.** *Let* $C \subseteq X$. *Then the indicator function* $\iota_C(x) : X \to]-\infty, +\infty]$ *of* C *is defined by*

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.1}$$

Definition 2.12 (local and global minimizers)**.** *Let* $f : X \to]-\infty, \infty]$ *be proper and let* $\bar{x} \in X$. *Then,* $\bar{x}$ *is a local minimizer of* f *if* $(\exists \delta > 0)$ *such that* $\|x - \bar{x}\| < \delta \Rightarrow f(\bar{x}) \leq f(x)$; *and* $\bar{x}$ *is a global minimizer of* f *if* $(\forall x \in \operatorname{dom} f)\ f(\bar{x}) \leq f(x)$.

Fact 2.13. *Let* $f : X \to]-\infty, \infty]$ *be convex and proper. Then every local minimizer of* f *is a global minimizer.*

2.3 Subdifferential Operators and Normal Cones

In many optimization problems, the functions are not necessarily smooth, differentiable, which leads us to study the subdifferential operator to help in algorithmic development.

Definition 2.14 (lower semicontinuous function)**.** *Let* $f : X \to]-\infty, \infty]$, *and let* $x \in X$. *Then* f *is a*

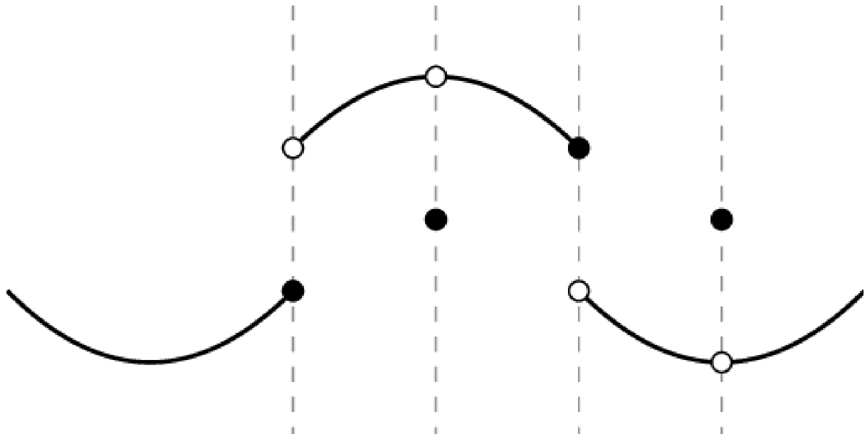

Figure 2.2: Lower Semicontinuous Function

HenryW ♣ why is lsc bold in the alias? why is the alias used sometimes and not others? please ensure consistency.

I have read up to here aug31 1;22PM. Please fix the material above and remove comments once they are fixed! Please read further on your own and try and be correct and consistent.

lower semicontinuous function (*lsc*) *at* x *if for every sequence* $(x_n)_{n\in\mathbb{N}}$ *in* X*,* $x_n \longrightarrow x \Rightarrow f(x) \leq \liminf f(x_n)$*. Moreover,* f *is lsc if* f *is lsc at every point in* X*.*

Definition 2.15 (subgradient and subdifferential)**.** *Let* $f : X \to\,]-\infty, \infty]$ *be proper, let* $x \in \operatorname{dom} f$ *and* $u \in X$*. Then* u *is a subgradient of* f *at* x *if* $(\forall y \in X) f(y) \geq f(x) + \langle u, y-x\rangle$*. The subdifferential of* f *is* $\partial f : x \mapsto \{u \in X \mid f(y) \geq f(x) + \langle u, y-x\rangle\}$*.*

Theorem 2.16 (Fermat's theorem)**.** *Let* $f : X \to\,]-\infty, \infty]$ *be proper. Then* $\operatorname{argmin} f = \{x \in X \mid 0 \in \partial f(x)\}$*.*

Proof. Let $x \in X$. Then

$$\begin{aligned} x \in \operatorname{argmin} f &\iff (\forall y \in X)\, f(x) \leq f(y) \\ &\iff (\forall y \in X)\, \langle 0, y-x\rangle + f(x) \leq f(y) \\ &\iff 0 \in \partial f(x). \end{aligned}$$

□

Definition 2.17 (cone)**.** *Let* $C \subseteq X$*. Then* C *is a cone if for every* $c \in C$*, for every* $\lambda \geq 0$ *we have* $\lambda c \in C$*.*

Definition 2.18 (normal cone)**.** *Let C be a nonempty convex subset of X and let $x \in X$. The normal cone of C at x is*

$$N_C(x) = \begin{cases} \{u \in X \mid \sup_{c \in C}\langle c - x, v\rangle \leq 0\}, & \text{if } x \in C; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.2}$$

Fact 2.19. *Let I be a finite indexed set, let $(f_i)_{i\in\mathbb{I}}$ be a family of functions from X to $]-\infty,\infty]$.*

1. *Suppose $(\forall i \in \mathbb{I})$ f_i is convex. Then $\sum_{i\in\mathbb{I}} fi$ is convex.*
2. *Suppose $(\forall i \in \mathbb{I})$ f_i is lsc. Then $\sum_{i\in\mathbb{I}} fi$ is lsc.*

Fact 2.20. *Let I be an indexed set and let $(f_i)_{i\in\mathbb{I}}$ be a family of convex and lsc functions on X. Then,* $\operatorname{epi} F = \bigcap\limits_{i\in\mathbb{I}} \operatorname{epi} f_i$.

Proposition 2.21. *Let I be an indexed set and let $(f_i)_{i\in\mathbb{I}}$ be a family of convex and lsc functions on X. Then $\sup_{i\in\mathbb{I}} f_i$ is convex and lsc.*

Proof. Set F $= \sup_{i\in\mathbb{I}} f_i$. We have $\operatorname{epi} F = \bigcap\limits_{i\in\mathbb{I}} \operatorname{epi} f_i$ by Fact 2.20. Since $(\forall i \in \mathbb{I}) f_i$ is convex and lsc, we conclude that $(\forall i \in \mathbb{I}) \operatorname{epi} f_i$ is convex and closed. Since the intersection of an arbitrary collection of convex sets in X is convex. We learn that $\operatorname{epi} F = \bigcap\limits_{i\in\mathbb{I}} \operatorname{epi} f_i$ is convex. Similarly, $\operatorname{epi} F = \bigcap\limits_{i\in\mathbb{I}} \operatorname{epi} f_i$ is closed. Then, $F = \bigcap\limits_{i\in\mathbb{I}} \operatorname{epi} f_i$ is lsc. □

Definition 2.22 (The support function)**.** *Let C be a subset of $\mathbb{R}^d$. The support function of C is*

$$\sigma_c(x) : u \longrightarrow \sup_{c\in C}\langle c, u\rangle.$$

Example 2.23. *Let $C = [a, b] \subseteq \mathbb{R}_+$. Then $(\forall x \in \mathbb{R})$*

$$\sigma_c(x) = \sup_{c\in[a,b]} cx = \begin{cases} bx, & \text{if } x \geq 0; \\ ax, & \text{otherwise.} \end{cases}$$

Example 2.24. *Let $f : \mathbb{R} \to \mathbb{R} : x \mapsto |x|$. Then, by Theorem 2.34 we have*

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

Proof. Let $x \in \operatorname{dom} f$ and let $u \in X$ which is a subgradient of f at X. Then, $(\forall x \in X)$ we have $f(y) \geq f(x) + \langle u, y - x\rangle$ where $u = \partial f(x)$. We can rewrite as $u \in [a, b]$ where $a = \lim_{y\to x^-} \frac{f(y)-f(x)}{y-x}$ and $b = \lim_{y\to x^+} \frac{f(y)-f(x)}{y-x}$. Consider,

1. When $x > 0$, $a = \lim_{y \to x^-} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^-} \frac{y-x}{y-x} = 1$ and $b = \lim_{y \to x^+} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^+} \frac{y-x}{y-x} = 1$. Then, $u = \{1\}$.

2. When $x < 0$, $a = \lim_{y \to x^-} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^-} \frac{-y+x}{y-x} = -1$ and $b = \lim_{y \to x^+} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^+} \frac{-y+x}{y-x} = -1$. Then, $u = \{-1\}$.

3. When $x = 0$, $a = \lim_{y \to x^-} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^-} \frac{-y+x}{y-x} = -1$ and $b = \lim_{y \to x^+} \frac{f(y)-f(x)}{y-x} = \lim_{y \to x^+} \frac{y-x}{y-x} = 1$. Then, $u = [-1, 1]$.

□

Fact 2.25. *Let $\varnothing \neq C \subseteq X$, let $x \in C$. Then, $N_C(x)$ is a nonempty closed convex cone.*

We denote S^n to be the set of all $n \times n$ symmetric matrices.

Example 2.26. *Let $f : S^n \to \mathbb{R} : X \mapsto \lambda_{max}(x)$ (the maximum eigenvalue of X). Let $X \in S^n$, let v be a normalized eigenvector of X (i.e., $\|v\| = 1$) associated with $\lambda_{max}(x)$. Then $vv^T \in \partial f(x)$.*

Proof. $(\forall y \in S^n)$ $f(y) \geq f(x) + \langle vv^T, y - x\rangle$. Then $\lambda_{max}(y) \geq \lambda_{max}(x) + tr(vv^T(y-x))$.

$$\begin{aligned}
\lambda_{max}(y) &= \max_{\|u\|=1} u^T y u \\
&\geq v^T y v \\
&= v^T(y - X_+ X)v \\
&= v^T(y - x) + v^T x v \\
&= tr(v^T(y-x)v) + \lambda_{max}(x)\|v\|^2 \\
&= tr(vv^T(y-x)) + \lambda_{max}(x) \\
&= \lambda_{max}(X) + tr(vv^T(y-x)).
\end{aligned}$$

□

Example 2.27. *Let $f : \mathbb{R} \to]-\infty, \infty]$ Then,*

$$x \mapsto \begin{cases} -\sqrt{x}, & \text{if } x \geq 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Suppose for eventual contradiction that $\partial f(0) \neq \varnothing$. Let $u \in \partial f(0)$. By the Definition 2.15. $(\forall y \in \mathbb{R})$ $f(y) \geq f(0) + u(y-0) = uy \iff (\forall y \geq 0) - \sqrt{y} \geq uy$. If $y = 1 \implies u \leq -1 < 0 \implies u^2 > 0$. If $y = \frac{1}{2u^2} \implies -\sqrt{\frac{1}{2u^2}} \geq \frac{1}{2u}$. Squaring both sides yields $\frac{1}{2u^2} \leq \frac{1}{ru^2} \iff 2u^2 \leq 1 \iff |u| \leq \frac{1}{\sqrt{2}}$.

Fact 2.28. *Let $f : X \to \,]-\infty, \infty]$ convex and proper. Thus,* $\text{int}\,(\text{dom } f) \subseteq \text{dom } \partial f \subseteq \text{dom } f$.

Fact 2.29. *Let $f : X \to \mathbb{R}$ be convex. Then f is subdifferentiable over X, i.e.,* $(\forall x \in X)$ $\partial f(x) \neq \varnothing$.

Fact 2.30. *Let $f : X \to \,]-\infty, \infty]$ be proper, $\alpha > 0$. Then,* $(\forall x \in \text{dom } f)$ $\partial(\alpha f)(x) = \alpha \partial f(x)$.

Fact 2.31. *Let $f_1, f_2 : X \to \,]-\infty, \infty]$ be proper and convex and suppose that $x \in \text{dom } f_1 \cap \text{dom } f_2$. Then,*

1. $\partial f_1(x) + \partial f_2(x) \subseteq \partial(f_1 + f_2)(x)$.
2. *Suppose that* $x \in \text{int}\,(\text{dom } f_1) \cap \text{int}\,(\text{dom } f_2)$. *Then* $\partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x)$.
3. *Suppose that* $x \in \text{ri}\,(\text{dom } f_1) \cap \text{ri}\,(\text{dom } f_2)$. *Then,* $\partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x)$.

Fact 2.32. *Let $f, g : X \to \,]-\infty, \infty]$ be convex lsc and proper. Suppose one of the following holds:*

1. $\text{int}\,(\text{dom } f) \cap (\text{dom } g) \neq \varnothing$.
2. $\text{ri}\,(\text{dom } f) \cap \text{ri}\,(\text{dom } g) \neq \varnothing$.

Then, $\partial(f + g) = \partial f + \partial g$.

Fact 2.33. *Let $f : X \to \,]-\infty, +\infty]$ be convex and proper. Let $x \in X$ and let $u \in X$. Then*

$$u \in \partial f(x) \iff (u, -1) \in N_{\text{epi}}(x, f(x)).$$

Theorem 2.34. *$f : X \to \,]-\infty, \infty]$ convex and proper. Suppose that $x \in \text{int}\,(\text{dom } f)$. If f is differentiable at x then $\partial f(x) = \{\nabla f(x)\}$.*

Proof. Since f is convex, proper, $x \in \text{int}\,(\text{dom } f) \subseteq \text{dom } \partial f$ we have $\partial f(x) \neq \varnothing$. Let $x^* \in \partial f(x)$. Then,

$$(\forall z \in X)\ f(z) \geq f(x) + \langle x^*, z - x \rangle.$$

Fix $h \in X$, consider $Z = x + th, t > 0$. Then $f(x + th) \geq f(x) + \langle x*, x + th - x \rangle = f(x) + t\langle x^*, h \rangle$. Rearranging,

$$\begin{aligned} \langle x^*, h \rangle &\leq \frac{f(x + th) - f(x)}{t} \\ &\leq \lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} = \langle \nabla f(x), h \rangle. \end{aligned}$$

Thus, we have $\langle x^* - \nabla f(x), h \rangle \leq 0$. Setting $h = x^* - \nabla f(x)$ yields $\|x^* - \nabla f(x)\|^2 \leq 0 \iff x^* = \nabla f(x)$. □

Fact 2.35. *$f : X \to]-\infty, \infty]$ convex and proper, and let $x \in \operatorname{int}(\operatorname{dom} f)$. If f has a unique subgradient at x then it is differentiable at x and $\partial f(x) = \{\nabla f(x)\}$.*

Example 2.36. *$f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$. Then,*

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2}, & \text{if } x \neq 0; \\ ball(0; 1), & \text{if } x = 0. \end{cases}$$

2.4 Conjugate of Convex Functions

Definition 2.37. *Let $f : X \to]-\infty, \infty]$. The Fenchel-Legendre convex conjugate of f is*

$$\begin{aligned} f^* &: X \to]-\infty, \infty] \\ &: u \mapsto \sup_{x \in X}(\langle x, u\rangle - f(x)). \end{aligned}$$

Example 2.38. *Let C be a nonempty closed convex subset of X and set $f = \iota_C$. Then $f^* = \sigma_C$.*

Proof. Let $u \in X$. By definition, we have

$$\begin{aligned} f^*(u) &= \sup_{x \in X}(\langle x, u\rangle - \iota_C(x)) \\ &= \sup_{x \in X}(\langle x, u\rangle) = \sigma_C(u). \end{aligned}$$

□

Theorem 2.39. *Let $f : X \to -]\infty, \infty]$. Then f^* is convex and lsc.*

Proof. Indeed, let $u \in X$. $f^*(u) = \sup_{x \in X}(\langle x, u\rangle - f(x))$. Set $(\forall x \in X)$ $h_x = \langle x, u\rangle - f(x)$. Then h_x is affine, hence lsc and convex. Consequently, f^* is a supremum of convex, lsc functions which means f^* is convex and lsc, by Proposition 2.21. □

Exercise 2.40. *Let $f : \mathbb{R} \to \mathbb{R} : x \mapsto e^x$. Then, $f^* = \begin{cases} u\ln(u) - u, & \text{if } u \geq 0; \\ 0, & \text{if } u = 0; \\ +\infty, & \text{otherwise.} \end{cases}$*

Proof. Let $u \in X$. Then,

$$f^*(u) = \sup_{x \in X}(xu - e^x).$$

1. if $u = 0$: $f^*(u) = sup_{x \in \mathbb{R}}(-e^x) = 0$.

2. if $u > 0$: $f^*(u) = \sup_{x\in\mathbb{R}}(xu - e^x)$. Set $g(x) = xu - e^x$ then $g'(x) = u - e^x$. Setting $g'(x) = 0$ yields $x = \ln u$ which implies $f^*(u) = u(\ln u) - e^{\ln u} = u\ln u - u$.

3. if $u < 0$: we learn that $g'(x) = u - e^x < 0$. Therefore,

$$\sup_{x\in\mathbb{R}}(xu - e^x) = \lim_{x\to-\infty}(xu - e^x) = +\infty.$$

□

Fact 2.41. *Let $f : X \to \,]-\infty, \infty]$ be proper and convex. Then f^* is proper.*

Theorem 2.42 (Fenchel Young inequality)**.** *Let $f : X \to \,]-\infty, \infty]$ be proper.* $\forall(x \in X)$

$$\forall(u \in X)\ f(x) + f^*(u) \geq \langle x, u\rangle.$$

Proof. By definition of f^* we have

$$\begin{aligned} f^*(u) &= \sup_{y\in X}(\langle u, y\rangle - f(y)) \\ &\geq \langle u, x\rangle - f(x) \\ \text{since } f(x) \neq -\infty \text{ then } f^*(u) + f(x) &\geq \langle u, x\rangle. \end{aligned}$$

$f^*(u) = \sup_{y\in X}(\langle u, y\rangle - f(y) \geq \langle u, x\rangle - f(x)$. Since we have $f(x) \neq -\infty$, then we can have $f^*(u) + f(x) \geq \langle u, x\rangle$. □

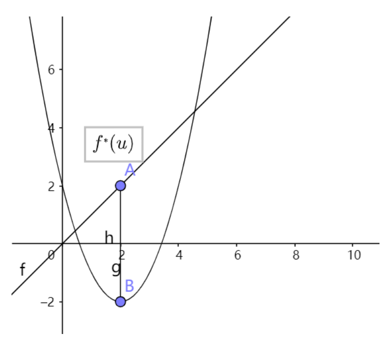

Figure 2.3: Geometrically Explanation

Definition 2.43. *The biconjugate of a function f is defined* $(\forall x \in X)$ $(f^*)^*(x) = \sup_{x\in X}(\langle x, y\rangle - f^*(y)) = (f^*)^*(x)$.

Fact 2.44. *Let $f : X \to \,]-\infty, \infty]$ be convex, lsc and proper. Let $x \in X$, let $u \in X$. Then*

$$u \in \partial f(x) \iff f(x) + f^*(u) = \langle x, u\rangle.$$

Fact 2.45. *Let $f : X \to \,]-\infty, \infty]$ be convex, lsc and proper. Then $f^{**} = f$.*

Proposition 2.46. *Let $f : X \to]-\infty, \infty]$ be convex, lsc and proper. Then*

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

Proof. We have $u \in \partial f(x) \Longrightarrow f(x) + f^*(u) = \langle x, u \rangle$ by Fact 2.44. Set $g = f^*$. Then g is convex lsc and proper. Moreover, $g^* = f^{**} = f$ by Fact 2.45. Hence,

$$\begin{aligned} u \in \partial f(x) &\Longleftrightarrow f(x) + f^*(u) = \langle x, u \rangle \\ &\Longleftrightarrow g^* + g(u) = \langle x, u \rangle \\ &\Longleftrightarrow x \in \partial g(u) = \partial f^*(u). \end{aligned}$$

□

Definition 2.47. *The primal problem associated with the sum of function is*

$$\min_{x \in X} f(x) + g(x),$$

and its Fenchel dual problem is

$$\min_{u \in X} f^*(-u) + g^*(u).$$

We set $\mu = \min(f+g)(X)$ and $\mu^* = \min(f^* \circ (-\operatorname{Id}) + g^*)(X)$. Observe that by Theorem 2.42 we have $\mu \geq -\mu^*$.

Definition 2.48 (Fenchel-Rockafellar Duality)**.** *Let Y be a Euclidean space and let $A : X \longrightarrow Y$ be linear, $f : X \longrightarrow]-\infty, +\infty]$, $g : Y \longrightarrow]-\infty, +\infty]$ and proper. The primal problem associated with the sum of two proper function is*

$$\min_{x \in X} f(x) + g(Ax),$$

and its dual problem is

$$\min_{u \in X} f^*(-A^T u) + g^*(u).$$

We set $\mu = \min(f + g \circ A)$ and $\mu^* = \min(f^* \circ (-A^T) + g^*)$. Thus, we can have $\mu \geq -\mu^*$. The duality gap is $\mu + \mu^*$.

2.5 Differentiability of Convex Functions

Fact 2.49. *Let $f : X \to]-\infty, \infty]$ be convex, lsc, and proper. Suppose that* dom f *is open and convex and that f is differentiable on* dom f*. Then ∇f is monotone, i.e., $(\forall x \in X)$ $(\forall y \in X)$ $\langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq 0$.*

Definition 2.50. *Let $L \geq 0$. A function $f : X \to]-\infty, \infty]$ is said to be L-smooth over a set $D \in X$ if it is differentiable over D and ∇f is L-Lipschitz continuous over D, i.e., $(\forall x \in D)$ $(\forall y \in D)$ $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$.*

Lemma 2.51 (The descent lemma)**.** *Let $L \geq 0$ and $f : X \to]-\infty, \infty]$ be L-smooth, i.e., ∇f is L-Lipschitz over $D \in X$. Then*

$$(\forall x \in D)\ (\forall y \in D)\ f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2.$$

Proof.

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x\rangle dt \\
\implies |f(y) - f(x) - \langle \nabla f(x), y - x\rangle| &= \left| \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x\rangle dt \right| \\
&\leq \left| \int_0^1 \|\nabla f(x + t(y-x)) - \nabla f(x)\| \|y - x\| dt \right| \\
\text{(since } \nabla f \text{ is } L\text{-Lipschitz)} &\leq \left| \int_0^1 L\|x + t(y-x) - x\| \|y - x\| dt \right| \\
&= \left| \int_0^1 tL\|y - x\|^2 dt \right| \\
&= L\|y - x\|^2 \frac{t^2}{2}|_0^1 \\
&= \frac{L}{2}\|y - x\|^2.
\end{aligned}$$

□

Fact 2.52. *Let $f : X \to \mathbb{R}$ be differentiable and convex and let $L > 0$. The following are equivalent:*

1. *∇f is L-Lipschitz continuous.*
2. *$(\forall x \in X)\ (\forall y \in X)\ f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2$.*
3. *$(\forall x \in X)\ (\forall y \in X)\ f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$.*
4. *$(\forall x \in X)\ (\forall y \in X)\ \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$.*

Remark 2.53. *By the Lemma 2.51 we can get $(i) \to (ii)$. Indeed, consider $f = -\frac{1}{2}\|\cdot\|^2$. $(\forall x \in X)\ \nabla f(x) = -x$, $\nabla f(x)$ is 1-Lipschitz. So $(i) \implies (ii)$. Now, $-f$ is convex since f is concave, we have $(\forall x \in X)\ (\forall y \in X)\ \frac{1}{2}\|y\|^2 \geq \frac{1}{2}\|x\|^2 + \langle x, y - x\rangle$. Observe that:*

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

But $\nabla f(x)$ is not 0-Lipschitz.

Fact 2.54. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be twice continuously differentiable and convex. Then the following are equivalent:*

1. *∇f is L-Lipschitz for some $L \geq 0$.*
2. *$\lambda_{max}(\nabla^2 f(x)) \leq L$ for any $x \in \mathbb{R}^m$.*

Definition 2.55. *Let $f : X \to \,]-\infty, \infty]$ be proper. Then f is β-strongly convex for some $\beta > 0$ if*

$$(\forall x \in X)\ (\forall y \in X)\ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\beta}{2}\lambda(1-\lambda)\|x-y\|^2.$$

Fact 2.56. *$f : X \to \,]-\infty+\infty]$ convex, lsc, proper, $\beta > 0$. Then the following are equivalent:*

1. *f is β-strongly convex.*
2. *$(\forall x \in \operatorname{dom} \partial f)\ (\forall y \in \operatorname{dom} f)\ (\forall u \in \partial f(x))\ f(y) \geq f(x) + \langle u, y-x\rangle + \frac{\beta}{2}\|y-x\|^2$.*
3. *$(\forall x, y \in \operatorname{dom} \partial f)\ (\forall u \in \partial f(x))\ (\forall v \in \partial f(y))\ \langle x-y, u-v\rangle \geq \beta\|x-y\|^2$.*

Fact 2.57. *Let $\beta > 0$ and let $f : X \to \,]-\infty, \infty]$ be β-strongly convex, lsc and proper. Then the following hold:*

1. *f has a unique minimizer x^*.*
2. *$(\forall x \in \operatorname{dom} f)\ f(x) - f(x^*) \geq \frac{\beta}{2}\|x-x^*\|^2$.*

Fact 2.58. *Let $f : X \to \,]-\infty, \infty]$ be convex, lsc, proper and let $g : X \to \,]-\infty, \infty]$ be β-strongly convex, proper. Then $f+g$ is β-strongly convex.*

Fact 2.59. *Let $\beta > 0$ and let $f : X \to \,]-\infty, \infty]$ be convex, lsc, and proper. Then the following hold,*

1. *∇f is $\frac{1}{\beta}$ Lipschitz and f is convex $\Longrightarrow f^*$ is β-strongly convex.*
2. *f is β-strongly convex $\Longrightarrow \nabla f^*$ is β-Lipschitz.*

2.6 The Proximal Mapping

Definition 2.60. *Let $f : X \to \,]-\infty, \infty]$ be convex, lsc, and proper. The proximal point mapping of f is,*

$$\operatorname{prox}_f(x) = \operatorname{argmin}_{u \in X}(f(u) + \frac{1}{2}\|u-x\|^2).$$

Theorem 2.61. *Let $f : X \to]-\infty, \infty]$ be convex, lsc, and proper. Then, $(\forall x \in X)$ $\operatorname{prox}_f(x)$ is a singleton.*

Proof. Indeed, let $x \in X$. Set $(\forall y \in X)$

$$g_x(y) = f(y) + \frac{1}{2}\|y - x\|^2.$$

Observe that f is proper, hence g_x is proper. Also, f is lsc, $\frac{1}{2}\|\cdot - x\|^2$ is smooth (hence lsc) $\implies g_x$ is lsc by Fact 2.19. In addtion, f is convex, $\frac{1}{2}\|\cdot - x\|^2$ is β-strongly convex for every $\beta \in]0, 1[$. We have $g_x = f + \frac{1}{2}\|\cdot - x\|^2$ is strongly convex by Fact 2.58. Therefore, by Theorem (2.57) we conclude g_x has a unique minimizer over X. □

Example 2.62. *Let C be a nonempty closed convex subset of X. Then $\operatorname{prox}_{\iota_C} = \mathrm{P}_C$.*

Proof. Let $x \in X$, let $p \in X$. Then

$$\begin{aligned} p = \operatorname{prox}_{\iota_C}(x) &\iff p = \operatorname{argmin}_{y \in X}(\iota_C(y) + \frac{1}{2}\|y - x\|^2) \\ &\iff (\forall y \in C)\, \iota_C(p) + \frac{1}{2}\|p - x\|^2 \le \iota_C(y) + \frac{1}{2}\|y - x\|^2 \\ &\iff (p \in C)\, (\forall y \in C)\, \|p - x\|^2 \le \|y - x\|^2 \\ &\iff (p \in C)\, (\forall y \in C)\, \|p - x\| \le \|y - x\| \\ &\iff p = \mathrm{P}_C(x). \end{aligned}$$

□

Fact 2.63. *Let $f : X \to]-\infty, \infty]$ be convex, lsc, and proper. Let $x \in X$ and $p \in X$. Then*

$$p = \operatorname{prox}_f(x) \iff (\forall y \in X)\ f(y) \ge f(p) + \langle y - p, x - p\rangle.$$

Corollary 2.64. *Let C be a nonempty closed convex subset of X, let $x \in X$ and $p \in X$. Then $p = \mathrm{P}_C(x) \iff (p \in C)$ and $(\forall c \in C)\langle p - x, p - c\rangle \le 0$.*

Proof. Recall the Proposition (2.63) we have $\operatorname{prox}_{\iota_C} = \mathrm{P}_C$. Now,

$$p = \mathrm{P}_C(x) \iff (\forall y \in X)\ \iota_C(y) \ge \iota_C(p) + \langle y - p, x - p\rangle \iff (p \in C) \text{ and } (\forall y \in C)\ \langle y - p, x - p\rangle \le 0.$$

□

Proposition 2.65. *Let $f : X \to]-\infty, \infty]$ be convex lsc and proper. Then*

$$\bar{x} \in \operatorname{argmin}_X f(x) \iff \bar{x} = \operatorname{prox}_f(\bar{x}).$$

Proof. Let $\bar{x} \in X$. Recall that by Proposition (2.63) we have

$$\bar{x} = \operatorname{prox}_f(\bar{x}) \iff (\forall y \in X)\ f(y) \ge f(\bar{x}) + \langle y - \bar{x}, \bar{x} - \bar{x}\rangle \iff (\forall y \in X)\ f(y) \ge f(\bar{x}).$$

Thus, $\bar{x}$ minimizes f over X as claimed. □

Example 2.66. *Let* $f : \mathbb{R} \to \mathbb{R} : x \mapsto \lambda|x|$, $\lambda > 0$. *Clearly* f *is convex lsc and proper. Moreover,* $(\forall x \in \mathbb{R})$

$$\operatorname{prox}_f(x) = \begin{cases} x - \lambda, & \textit{if } x > \lambda; \\ 0, & \textit{if } |x| \leq \lambda; \\ x + \lambda, & \textit{if } x < -\lambda. \end{cases}$$

This is known as the soft thresholder.

Fact 2.67. *Let* $f : \mathbb{R}^m \to]-\infty, \infty]$ *be given by* $(\forall x = (x_1, ..., x_m) \in \mathbb{R}^m)$ $f(x_1, ..., x_m) = \sum_{i=1}^{m} f_i(x_i)$, *where* $(\forall i \in \{1, ..., m\})$ $f_i : \mathbb{R} \longrightarrow]-\infty, +\infty]$ *is convex, lsc, and proper. Then* $(\forall x = (x_1, ..., x_m) \in \mathbb{R}^m)$ $\operatorname{prox}_f(x) = (\operatorname{prox}_{f_i}(x_i))_{i=1}^{n} = (\operatorname{prox}_{f_{1}(x_1)}, ..., \operatorname{prox}_{f_{m}(x_m)})$.

Chapter 3

3.1 Zeros of the sum of monotone operators: a static framework

Let $A : X \mapsto X$ be a possibly set-valued operator, i.e., $A(x) \subseteq X$. Then A is *monotone*, if

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

where $\operatorname{gra}(A)$ denotes the graph of A defined by

$$\operatorname{gra}(A) = \{(x,u) \in X \times X \,:\, u \in A(x)\}.$$

It is a *maximally monotone operator* if $\operatorname{gra}(A)$ cannot be properly extended without destroying monotonicity. In the following we assume that

$$A \text{ and } B \text{ are maximally monotone operators on } X.$$

Splitting algorithms have been successfully employed to solve, when a solution exists, various monotone inclusion problems of the type:

$$\text{Find } x \in \operatorname{zer}(A + B) = \big\{x \in X \;\big|\; 0 \in Ax + Bx\big\}. \tag{3.1}$$

We denote the *resolvent of A*, $J_A = (\operatorname{Id} + A)^{-1}$, and the *reflected resolvent of A*, $R_A = 2J_A - \operatorname{Id}$. Both the resolvent and the reflected resolvent are of central importance Let $T \colon X \to X$. Recall that the fixed point set of T, $\operatorname{Fix} T$, is given by $\operatorname{Fix} T = \big\{x \in X \;\big|\; x = Tx\big\}$.

Fact 3.1. *Let $\gamma > 0$ and let $\alpha \in \,]0,1]$. The following hold:*

1. $\operatorname{zer}(A + B) = J_A(\operatorname{Fix}((1-\alpha)\operatorname{Id} + \alpha R_{\gamma B} R_{\gamma A}))$.
2. *Suppose that $B \colon X \to X$. Then* $\operatorname{zer}(A + B) = \operatorname{Fix}(J_{\gamma B}(\operatorname{Id} - \gamma A))$.

Connection to Subdifferential Operators

In the following we assume that

$f, g : X \to \left]-\infty, +\infty\right]$ are proper lower semicontinuous, not necessarily smooth, convex functions.

A classical optimization problem takes the form.

Problem 3.2.

$$\textit{Find } \bar{x} \in \operatorname{argmin}_{x \in X} f(x) + g(x).$$

Recall the subdifferential of f is the the set-valued operator

$$\partial f(x) = \big\{ u \in X \;\big|\; f(y) \geq f(x) + \langle u, y - x \rangle \,,\, \forall x, y \in X \big\}. \tag{3.2}$$

It follows from Rockafellar's fundamental result that ∂f is maximally monotone. The subdifferential operator is a powerful tool in optimization. By Theorem 2.16

$$0 \in \partial f(x) \Leftrightarrow x \text{ is a global minimizer of } f. \tag{3.3}$$

By Fact 2.31, assuming an appropriate constraint qualification to guarantee the sum rule $\partial f + \partial g \neq \varnothing$ holds (e.g., $\partial(f + g) = \partial f + \partial g$), Problem 3.2 reduces to (3.1), where A and B are maximally monotone operators on X, namely the subdifferential operators ∂f and ∂g of the functions under consideration. Constrained optimization problems of minimizing an objective function f over a constraint set C are typically modelled in the form of Problem 3.2. In this case, we set $g = \iota_C$, the **indicator function** of the set C, i.e., it has the value 0 on C, and $+\infty$, otherwise.

3.2 Firmly Nonexpansive and Averaged Mappings: A Dynamic Framework

Definition 3.3. *Let* $T \colon X \to X$*, let* $\alpha \in \left]0, 1\right[$ *and let* $\beta > 0$*. Then*

1. *T is nonexpansive if* $(\forall (x, y) \in X \times X)$ $\|Tx - Ty\| \leq \|x - y\|$*, i.e.,* 1- *Lipschitz continuous;*

2. *T is* firmly nonexpansive *if* $(\forall (x, y) \in X \times X)$ $\|Tx - Ty\|^2 + \|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2 \leq \|x - y\|^2$*;*

3. *T is* α-averaged *if there exists a nonexpansive mapping* $N \colon X \to X$ *such that* $T = (1 - \alpha)\operatorname{Id} + \alpha N$*;*

4. *T is* β-cocoercive *if* βT *is firmly nonexpansive. .*

Remark 3.4. *It is straightforward to verify that T is firmly nonexpansive if and only if T is $\frac{1}{2}$-averaged.*

The class of maximally monotone operators is closely related to the class of (firmly) nonexpansive mappings via the corresponding resolvent (and also reflected resolvent) as we demonstrate in the following fact.

Fact 3.5. *Let $T\colon X \to X$, set $R = 2T - \mathrm{Id}$ and $A = T^{-1} - \mathrm{Id}$. Then the following hold:*

1. $T = J_A$.
2. *T is firmly nonexpansive $\Leftrightarrow$ R is nonexpansive $\Leftrightarrow$ A is maximally monotone.*

Example 3.6. *Let $f\colon X \to \,]-\infty, +\infty]$ be convex lower semicontinuous and proper and let $L > 0$. The following hold:*

1. $\mathrm{prox}_f = J_{\partial f}$. *Hence prox_f is firmly nonexpansive.*
2. *Suppose that f is differentiable and that ∇f is L-Lipschitz contrinuous. Then $\frac{1}{L}\nabla f$ and $\mathrm{Id} - \frac{1}{L}\nabla f$ are firmly nonexpansive.*

Fact 3.7. *Let $m \in \{1, 2 \ldots\}$, set $I = \{1, \ldots, m\}$ and let $(\alpha_i)_{i\in I}$ be a family of real numbers in $]0, 1[$. Suppose that $(\forall i \in I)$ $T_i\colon X \to X$ is α_i-averaged. Set*

$$T = T_m \ldots T_1 \quad \textit{and} \quad \alpha = \frac{\sum_{i\in I} \frac{\alpha_i}{1-\alpha_i}}{1+\sum_{i\in I} \frac{\alpha_i}{1-\alpha_i}}. \tag{3.4}$$

Then $\alpha \in \,]0, 1[$ and T is α-averaged.

The notion of firm nonexpansiveness (and more generally averagedness) is very useful when studying the iterative behaviour of the corresponding operators as we recall in the following fact.

Fact 3.8. *Let $T\colon X \to X$ and let $\alpha \in \,]0, 1[$. Suppose that T is α-averaged and that $\mathrm{Fix}\, T \neq \varnothing$. Let $x_0 \in X$ and $(\forall n \in \mathbb{N})$ update via*

$$x_{n+1} = Tx_n. \tag{3.5}$$

Then $(\exists x^* \in \mathrm{Fix}\, T)$ such that $(x_n)_{n\in\mathbb{N}}$ converges weakly to x^*.

Chapter 4

A Catalogue of Splitting Methods

Recall that A and B are maximally monotone operators on X, a Hilbert space. We assume that the set of zeros

$$\operatorname{zer}(A + B) \neq \varnothing. \tag{4.1}$$

In view of (4.1), we observe that setting $(A, B) = (\partial f, \partial g)$, implies

$$\operatorname{argmin}(f + g) = \operatorname{zer}(\partial f + \partial g) \neq \varnothing. \tag{4.2}$$

In this section we provide a collection of prominent splitting methods. Each of the methods listed below produces a sequence that converges to a point in $\operatorname{zer}(A + B)$.

4.1 The Forward-Backward Method

In view of (4.1), an immediate consequence of Fact 3.5, Fact 3.1(1), and Fact 3.7 is the following convergence result of the forward-backward method. .

Fact 4.1. *Let $\beta > 0$. Suppose that A is β-cocoercive. Let $\gamma \in \left]0, 2\beta\right[$. Then the forward-backward operator $T_{\mathrm{FB}} = J_{\gamma B}(\operatorname{Id} - \gamma A)$ is averaged. Let $x_0 \in X$ and $(\forall n \in \mathbb{N})$ update via:*

$$x_{n+1} = T_{\mathrm{FB}} x_n. \tag{4.3}$$

Then the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A + B) = \operatorname{Fix} T_{\mathrm{FB}}$.

The proximal gradient method

Suppose that f is smooth and that $\boldsymbol{\nabla} f$ is L-Lipschitz continuous for some $L > 0$. Observe that, by the Baillon–Haddad Theorem, $\boldsymbol{\nabla} f$ is $\frac{1}{L}$-cocoercive. Set $(A, B) = (\boldsymbol{\nabla} f, \partial g)$ and let $\gamma \in \left]0, \frac{2}{L}\right[$. The forward-backward operator in this case reduces to the proximal gradient operator $T_{\mathrm{FB}} = \operatorname{prox}_{\gamma g}(\operatorname{Id} - \gamma \boldsymbol{\nabla} f)$. Recalling (4.2), the sequence $(x_n)_{n\in\mathbb{N}}$ defined in (4.3) converges weakly to a point in $\operatorname{argmin}(f + g)$.

4.2 The Douglas–Rachford Method

The Douglas–Rachford operator associated with the ordered pair (A, B) is

$$T_{\rm DR} = \tfrac{1}{2}(\operatorname{Id} + R_B R_A). \tag{4.4}$$

Note that $R_A = 2\operatorname{prox}_f - \operatorname{Id}$ and $R_B = 2\operatorname{prox}_g - \operatorname{Id}$. Also, it follows from the nonexpansiveness of R_A and R_B (see Fact 3.5(2)) that the Douglas–Rachford operator defined in (4.4) is firmly nonexpansive, i.e., $\frac{1}{2}$-averaged. The convergence of the Douglas–Rachford algorithm is summarized in the following result.

Fact 4.2. *Let T be the Douglas–Rachford operator associated with the ordered pair (A, B) and let $x_0 \in X$. $(\forall n \in \mathbb{N})$ update via:*

$$y_n = J_A x_n \tag{4.5a}$$

$$x_{n+1} = T x_n \tag{4.5b}$$

Then the governing *sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in* $\operatorname{Fix} T$ *and the* shadow *sequence $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point in* $\operatorname{zer}(A + B) = J_A(\operatorname{Fix} T)$.

Douglas–Rachford method in optimization settings

Set $(A, B) = (\partial f, \partial g)$. The Douglas–Rachford operator in this case reduces to the operator $T_{\rm DR} = \frac{1}{2}(\operatorname{Id} + (2\operatorname{prox}_g - \operatorname{Id})(2\operatorname{prox}_f - \operatorname{Id}))$. Recalling (4.2), (4.5a) in view of (3.6)(1), the sequence $(\operatorname{prox}_f x_n)_{n\in\mathbb{N}}$ defined in (4.5a) converges weakly to a point in $\operatorname{argmin}(f + g) = \operatorname{zer}(\partial f + \partial g)$.

Parallel Splitting and Pierra's Product Space Technique

Let $m \in \{2, 3, \ldots\}$, set $I = \{1, \ldots, m\}$ and let $(A_i)_{i\in I}$ be a family of maximally monotone operators from X to X. Using Pierra's product space technique, the Douglas–Rachford algorithm can be recasted to find a zero of $\sum_{i\in I} A_i$ (provided that one exists) at the expense of working in the product space X^m. A utility version of this adaptation is stated in the following fact.

Fact 4.3. *Let $m \in \{2, 3, \ldots\}$, set $I = \{1, \ldots, m\}$ and let $(A_i)_{i\in I}$ be a family of maximally monotone operators from X to X. Suppose that* $\operatorname{zer} \sum_{i\in I} A_i \neq \varnothing$. *Let $(y_{i,0})_{i\in I} \in X^m$ and $(\forall n \in \mathbb{N})$ update via:*

$$p_n = \tfrac{1}{m} \sum_{i\in I} y_{i,n} \tag{4.6a}$$

$$x_{i,n} = J_{A_i} y_{i,n}, \quad i \in I \tag{4.6b}$$

$$q_n = \tfrac{1}{m} \sum_{i\in I} x_{i,n} \tag{4.6c}$$

$$y_{i,n+1} = y_{i,n} + 2q_n - p_n - x_{i,n}, \quad i \in I \tag{4.6d}$$

Then $(p_n)_{n\in\mathbb{N}}$ *converges weakly to some point in* $\operatorname{zer}\sum_{i\in I} A_i$.

4.3 The Peaceman–Rachford Method

The Peaceman–Rachford operator associated with the ordered pair (A, B) is

$$T = R_B R_A. \tag{4.7}$$

The convergence of the Peaceman–Rachford algorithm is summarized in the following result.

Fact 4.4. *Suppose that A is uniformly monotone. Let T be the Peaceman–Rachford operator associated with the ordered pair (A, B) and let $x_0 \in X$. $(\forall n \in \mathbb{N})$ update via:*

$$y_n = J_A x_n \tag{4.8a}$$

$$x_{n+1} = T x_n. \tag{4.8b}$$

Then the shadow *sequence* $(y_n)_{n\in\mathbb{N}}$ *converges strongly to a point in* $\operatorname{zer}(A+B) = J_A(\operatorname{Fix} T)$.

Peaceman–Rachford method in optimization settings

Suppose that f is uniformly convex. Observe that ∂f is uniformly monotone. Set $(A, B) = (\partial f, \partial g)$. The Peaceman–Rachford operator in this case reduces to the operator $T = (2\operatorname{prox}_g - \operatorname{Id})(2\operatorname{prox}_f - \operatorname{Id})$. Recalling (4.2), (4.8a) in view of Example 3.6(1), the sequence $(\operatorname{prox}_f x_n)_{n\in\mathbb{N}}$ defined in (4.8a) converges strongly to a point in $\operatorname{argmin}(f + g)$.

4.4 Alternating Direction Method of Multipliers (ADMM)

Let Y be a Hilbert spaces, let $A\colon X \to Y$, be a continuous and linear, and let $h\colon Y \to \left]-\infty, +\infty\right]$ be convex, lower semicontinuous and proper. Consider the convex optimization problem

$$\underset{x\in X}{\text{minimize}} \;\; f(x) + h(Ax), \tag{4.9}$$

and its Fenchel–Rockafellar dual

$$\underset{y\in Y}{\text{minimize}} \;\; f^*(-A^*y) + h^*(y). \tag{4.10}$$

Fix $\gamma > 0$. The augmented Lagrangian associated with (4.9) is

$$L : X \times Y \times Y : (x, y, z) \mapsto f(x) + h(y) + \langle z, Ax - y\rangle + \tfrac{\gamma}{2}\|Ax - y\|^2. \tag{4.11}$$

The ADMM scheme consists in minimizing the augmented Lagrangian (4.11) over x then over y and then update the dual variable. Let $(x_0, y_0, z_0) \in (X \times Y \times Y)$. The ADMM scheme updates (x_0, y_0, z_0) via

$$x_{n+1} \in \operatorname{argmin}_{x \in X}\{f(x) + \langle z_n, Ax\rangle + \tfrac{\gamma}{2}\|Ax - y_n\|^2\}. \tag{4.12a}$$

$$y_{n+1} \in \operatorname{argmin}_{y \in Y}\{h(y) + \langle z_n, y\rangle + \tfrac{\gamma}{2}\|Ax_{n+1} - y\|^2\}. \tag{4.12b}$$

$$z_{n+1} = z_n + \gamma(Ax_{n+1} - y_{n+1}). \tag{4.12c}$$

Convergence of ADMM

Under appropriate constraint qualifications, the dual problem (4.10) is equivalent to the monotone inclusion problem

$$\text{Find } y \in Y \text{ such that } 0 \in \partial(f^* \circ (-A^*))(y) + \partial h^*(y). \tag{4.13}$$

It is well-known that applying Douglas–Rachford method to solve (4.13) reduces to the scheme in (4.12).

Chapter 5

Conclusion

In this report, we presented various splitting methods to solve the problem of the sum of two convex functions. We learned that splitting methods split the original objective into two parts and solve two convex optimization problems. In addition, we compared the differences of various splitting methods and learned what method to use in what case. Also, these methods can be applied to other fields, such as machine learning (e.g., support vector machines, regularization), data science, and image processing.

Index

References

[1] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 1

[2] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959. 1

[3] Jonathan Eckstein and Dimitri P Bertsekas. On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992. 1

[4] René Escalante and Marcos Raydan. *Alternating projection methods*. SIAM, 2011. 1

[5] Masao Fukushima. The primal douglas-rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem. *Mathematical Programming*, 72(1):1–15, 1996. 1

[6] Pontus Giselsson and Stephen Boyd. Linear convergence and metric selection for douglas-rachford splitting and admm. *IEEE Transactions on Automatic Control*, 62(2):532–544, 2016. 1

[7] Xin Luo, MengChu Zhou, Shuai Li, Zhuhong You, Yunni Xia, and Qingsheng Zhu. A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method. *IEEE transactions on neural networks and learning systems*, 27(3):579–592, 2015. 1

[8] Walaa M Moursi and Matthew Saurette. On the douglas-rachford and peaceman-rachford algorithms in the presence of uniform monotonicity and the absence of minimizers. *arXiv preprint arXiv:2201.06661*, 2022. 1

[9] Yunzhang Zhu. An augmented admm algorithm with application to the generalized lasso problem. *Journal of Computational and Graphical Statistics*, 26(1):195–204, 2017. 1