

2016

# Convex and Nonconvex Optimization Techniques for the Constrained Fermat-Torricelli Problem

Nathan Lawrence  
*Portland State University*

Let us know how access to this document benefits you.

Follow this and additional works at: <https://pdxscholar.library.pdx.edu/honorsthesis>

---

## Recommended Citation

Lawrence, Nathan, "Convex and Nonconvex Optimization Techniques for the Constrained Fermat-Torricelli Problem" (2016).  
*University Honors Theses*. Paper 317.

[10.15760/honors.319](https://pdxscholar.library.pdx.edu/honors.319)

This Thesis is brought to you for free and open access. It has been accepted for inclusion in University Honors Theses by an authorized administrator of PDXScholar. For more information, please contact [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

# Convex and Nonconvex Optimization Techniques for the Constrained Fermat-Torricelli Problem

by

Nathan Lawrence

An undergraduate honors thesis submitted in partial fulfillment of the  
requirements for the degree of

Bachelor of Arts  
in  
University Honors  
and  
Mathematics

Thesis Advisor  
Dr. Mau Nam Nguyen

Portland State University  
2016

# Convex and Nonconvex Optimization Techniques for the Constrained Fermat-Torricelli Problem

Nathan Lawrence  
Advisor: Dr. Mau Nam Nguyen  
Portland State University, 2016

## ABSTRACT

The Fermat-Torricelli problem asks for a point that minimizes the sum of the distances to three given points in the plane. This problem was introduced by the French mathematician Fermat in the 17th century and was solved by the Italian mathematician and physicist Torricelli. In this thesis we introduce a constrained version of the Fermat-Torricelli problem in high dimensions that involves distances to a finite number of points with both positive and negative weights. Based on the distance penalty method, Nesterov's smoothing technique, and optimization techniques for minimizing differences of convex functions, we provide effective algorithms to solve the problem.

# Contents

<b>Introduction</b>	1
<b>Ch. 1. Preliminaries</b>	2
<b>Ch. 2. Optimization Methods</b>	10
2.1 The Difference of Convex Functions . . . . .	10
2.1.1 An Introduction DC programming . . . . .	12
2.1.2 Convergence of the DC Algorithm . . . . .	13
2.2 Nesterov's Smoothing Technique . . . . .	16
<b>Ch. 3. The Fermat-Torricelli Problem</b>	20
3.1 The Original Problem . . . . .	20
3.1.1 Existence and Uniqueness of Solutions . . . . .	20
3.1.2 Weiszfeld's Algorithm . . . . .	22
3.2 The Constrained Problem . . . . .	24
3.2.1 Solving with Nesterov's Accelerated Gradient Method . . . . .	29
3.2.2 Solving the Fermat-Torricelli Problem via the DCA . . . . .	32
<b>Bibliography</b>	38

# Introduction

Pierre de Fermat proposed a problem in the 17th century that sparked interest in the location sciences: given three points in the plane, find a point such that the sum of its Euclidean distances to the three points is minimal. This problem was solved by Evangelista Torricelli, and is now referred to as the *Fermat-Torricelli problem*. In 1937 Endre Weiszfeld developed the first numerical algorithm to solve this problem. It was Harold Kuhn in 1972 who asserted and proved the necessary and sufficient conditions for Weiszfeld's algorithm to converge. Since then this problem has been generalized to handle any finite number of points in  $\mathbb{R}^n$  and the Weiszfeld algorithm has also been improved and modified.

In this undergraduate thesis we introduce the *constrained Fermat-Torricelli problem* and solve it using the distance penalty method. The paper is organized as follows: Chapter 1 provides the necessary definitions and results from convex analysis in order to understand the latter sections. Chapter 2 is a brief introduction to DC programming and Nesterov's smoothing technique. Chapter 3 has several key parts. We describe the known results surrounding the Fermat-Torricelli problem, then introduce its constrained analog with detailed proofs. A new algorithm for solving the problem is introduced. Finally, we discuss Nesterov's accelerated gradient method and DC programming in light of the constrained problem. Throughout the text we provide remarks, examples, and figures to develop the reader's intuition on the subject.

# Chapter 1

## Preliminaries

The quintessential optimization problem begins with an *objective function*  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  and attains a solution if there exists  $\bar{x} \in \mathbb{R}^n$  such that the value  $f(\bar{x})$  is minimal or maximal. A *constrained* problem is defined similarly, but with the caveat that the solution must adhere to the properties of what is called the *constraint set*. The existence and uniqueness of an optimal solution as well as numerical methods to find an optimal solution are vital concerns when approaching an optimization problem. It is this stipulation that makes convex analysis desirable insofar that properties surrounding the notion of *convexity* align with the necessary conditions of a solvable optimization problem.

The goal of this chapter is to shed light on the “desirable” properties in convex analysis relevant for later chapters of this thesis. We survey key terms and results for understanding later sections of the thesis. Figures and examples are given to guide the reader’s conceptual intuition. For a more comprehensive study of convex analysis the reader is referred to [5, 10, 16].

Let  $\mathbb{R}^n$  denote the set of  $n$ -tuples of real numbers, where each  $x \in \mathbb{R}^n$  is of the form  $(x_1, \dots, x_n)$ . Also, let the *extended real number line* be defined by  $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ . Given  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  and  $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ , the *Euclidean norm* (2-norm) of  $x$  is defined by

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2},$$

and the *inner product* of  $x$  and  $y$  is defined by

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i.$$

It follows immediately that  $\|x\| = \sqrt{\langle x, x \rangle}$ . With this and the property  $\langle x, y \rangle = \langle y, x \rangle$ , we have the identity  $\|x - y\|^2 = \|x\|^2 - 2\langle x, y \rangle + \|y\|^2$ .

**Definition 1.0.1.** The **closed ball** centered at  $\bar{x}$  with radius  $r \geq 0$  and the **closed unit ball** of  $\mathbb{R}^n$  are defined, respectively, by

$$\mathbb{B}(\bar{x}; r) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq r\} \text{ and } \mathbb{B} := \{x \in \mathbb{R}^n : \|x\| \leq 1\}.$$

**Remark 1.0.2.** While the definition is in terms of the Euclidean norm, the notion of a closed ball or closed unit ball makes sense in terms of the other norms.

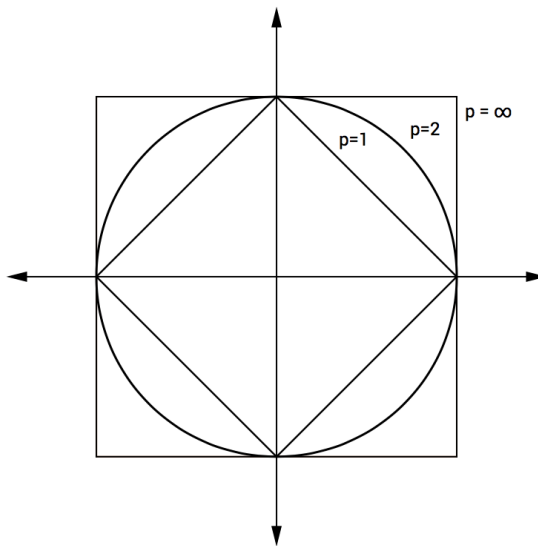


FIGURE 1.0.1: Closed unit ball with respect to various  $p$ -norms  $\|\cdot\|_p$

**Definition 1.0.3.** A set  $\Omega \subset \mathbb{R}^n$  is **convex** if  $\lambda a + (1 - \lambda)b \in \Omega$  for all  $a, b \in \Omega$  and  $\lambda \in [0, 1]$ . That is, the interval  $[a, b] \subset \Omega$  whenever  $a, b \in \Omega$ .

**Definition 1.0.4.** Given elements  $a_1, \dots, a_m \in \mathbb{R}^n$ , the element  $x = \sum_{i=1}^m \lambda_i a_i$ , where  $\sum_{i=1}^m \lambda_i = 1$  and  $\lambda_i \geq 0$  for each  $1 \leq i \leq m$ , is a **convex combination** of these elements.

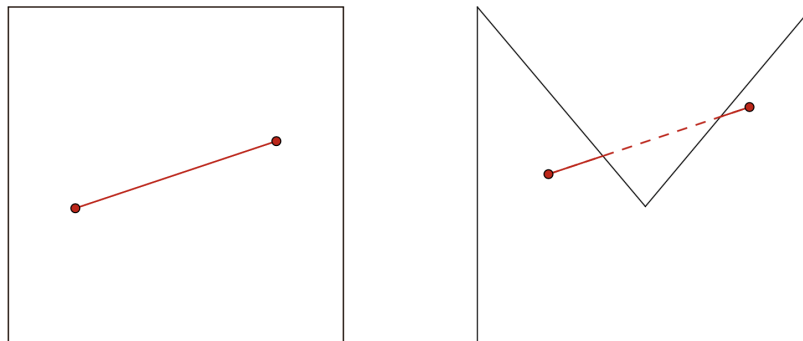


FIGURE 1.0.2: A convex set on the left, and a non-convex set on the right. The line connecting any two points in the set must also be contained.

**Proposition 1.0.5.** A set  $\Omega \subset \mathbb{R}^n$  is convex if and only if it contains all convex combinations of its elements.

*Proof.* Given  $\Omega \subset \mathbb{R}^n$  to be convex, it is straightforward that it contains all convex combinations of its elements. The converse is shown by induction. That is, we must show  $x := \sum_{i=1}^{m+1} \lambda_i \omega_i \in \Omega$ , where each  $\omega_i \in \Omega$ ,  $\sum_{i=1}^{m+1} \lambda_i = 1$  with  $\lambda_i \geq 0$ . Notice the result follows by definition for  $m=1,2$ . So fix  $m \geq 3$  and suppose every convex combination up to  $m$  of elements in  $\Omega$  is contained in  $\Omega$ . It is also worth pointing out that if  $\lambda_{m+1} = 1$ , then  $y = \omega_{m+1}$  and we are done. Otherwise, we see

$$y := \sum_{i=1}^m \frac{\lambda_i}{1 - \lambda_{m+1}} \omega_i \in \Omega$$

Further, we get the reformulation  $x = (1 - \lambda_{m+1})y + \lambda_{m+1}\omega_{m+1} \in \Omega$  by definition.  $\square$

**Definition 1.0.6.** Let  $\Omega \subset \mathbb{R}^n$ . The **convex hull** of  $\Omega$  is defined and denoted by

$$\text{co}(\Omega) := \bigcap \{C : C \text{ is convex and } \Omega \subset C\}.$$

**Proposition 1.0.7.** Let  $\Omega \subset \mathbb{R}^n$ . The convex hull  $\text{co}(\Omega)$  is the minimal convex set containing  $\Omega$ .

*Proof.* Define  $\mathcal{C} := \{C : C \text{ is convex and } \Omega \subset C\}$ . Let  $a, b \in \text{co}(\Omega) = \bigcap \mathcal{C}$ , then  $a, b \in C$  for all  $C \in \mathcal{C}$ . Then for any  $\lambda \in (0, 1)$  we see  $\lambda a + (1 - \lambda)b \in C$ . This shows  $\text{co}(\Omega)$  is convex. Further, for any  $x \in \text{co}(\Omega)$  we see  $x \in C$  whenever  $\Omega \subset C$ , by properties of set intersection. Thus,  $\text{co}(\Omega)$  is minimal.  $\square$



**Definition 1.0.8.** The **domain** of a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is defined by

$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

Further,  $f$  is **proper** if  $\text{dom } f \neq \emptyset$ .

**Definition 1.0.9.** Given a nonempty set  $\Omega \subset \mathbb{R}^n$ , the **distance function** associated with  $\Omega$  is defined by

$$d(x; \Omega) = \inf\{\|x - \omega\| : \omega \in \Omega\}.$$

**Remark 1.0.10.** For a nonempty close set, it is an important fact about the distance function that  $d(\bar{x}; \Omega) = 0$  whenever  $\bar{x} \in \Omega$  and  $d(\bar{x}; \Omega) > 0$  whenever  $\bar{x} \notin \Omega$ .

**Definition 1.0.11.** For any  $x \in \mathbb{R}^n$ , the **Euclidean projection** from  $x$  to  $\Omega$  is given by

$$P(x; \Omega) = \{\omega \in \Omega : \|x - \omega\| = d(x; \Omega)\}.$$

**Definition 1.0.12.** Let  $f : \Omega \rightarrow \overline{\mathbb{R}}$  be an extended-real-valued function defined on a convex set  $\Omega \subset \mathbb{R}^n$ . Then  $f$  is **convex** on  $\Omega$  if for all  $x, y \in \Omega$  and  $\lambda \in (0, 1)$

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

If this inequality is strict whenever  $x \neq y$ , we say  $f$  is **strictly convex**.

**Example 1.0.13.** We show that the norm function  $f(x) = \|x\|$ ,  $x \in \mathbb{R}^n$ , is convex. Indeed, for any  $x, y \in \mathbb{R}^n$  and  $\lambda \in (0, 1)$  we have

$$f(\lambda x + (1 - \lambda)y) = \|\lambda x + (1 - \lambda)y\| \leq \lambda \|x\| + (1 - \lambda)\|y\| = \lambda f(x) + (1 - \lambda)f(y)$$

**Proposition 1.0.14.** Let  $f_i : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be convex functions for all  $i = 1, \dots, m$ . Then the maximum function  $\max_{1 \leq i \leq m} f_i$  is convex.

*Proof.* Let  $g := \max\{f_i : i = 1, \dots, m\}$  and  $x, y \in \mathbb{R}^n$  be given. Then for any  $\lambda \in (0, 1)$  we see

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

for  $i = 1, \dots, m$ . It follows that

$$g(\lambda x + (1 - \lambda)y) = \max\{f_i(\lambda x + (1 - \lambda)y) : i = 1, \dots, m\} \leq \lambda g(x) + (1 - \lambda)g(y).$$

Thus, the maximum function is convex. □

**Proposition 1.0.15.** Let  $f_i : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be convex functions for all  $i = 1, \dots, m$ . Then  $\sum_{i=1}^m f_i$  is convex.

*Proof.* We show the result for  $m = 2$ . The general result follows from mathematical induction. For ease in notation, let  $f$  and  $g$  be convex functions on  $\mathbb{R}^n$ . Fix any  $x, y \in \mathbb{R}^n$  and  $\lambda \in (0, 1)$ . Then we see that

$$\begin{aligned} [f + g](\lambda x + (1 - \lambda)y) &= f(\lambda x + (1 - \lambda)y) + g(\lambda x + (1 - \lambda)y) \\ &\leq \lambda f(x) + (1 - \lambda)f(y) + \lambda g(x) + (1 - \lambda)g(y) \\ &= \lambda[f + g](x) + (1 - \lambda)[f + g](y). \end{aligned}$$

Thus,  $f + g$  is convex. □

**Definition 1.0.16.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is **Fréchet differentiable** at  $\bar{x} \in \text{int}(\text{dom}f)$  if there exists an element  $v \in \mathbb{R}^n$  such that

$$\lim_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.$$

Then  $v$  is called the **Fréchet derivative** of  $f$  at  $\bar{x}$  and is denoted by  $\nabla f(\bar{x})$ .

**Definition 1.0.17.** Let  $C^1(\Omega)$  denote the set of all functions whose partial derivatives exist and are continuous on  $\Omega$ . Then given a real-valued function  $f \in C^1(\mathbb{R}^n)$ , the **gradient** of  $f$  at  $x$  is given as

$$\nabla f(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)$$

**Definition 1.0.18.** Given a convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\bar{x} \in \text{dom}f$ , and element  $v \in \mathbb{R}^n$  is called a **subgradient** of  $f$  at  $\bar{x}$  if

$$\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n.$$

Moreover, the *subdifferential* refers to the collection of all the subgradients of  $f$  at  $\bar{x}$  and is denoted by  $\partial f(\bar{x})$ .

**Remark 1.0.19.** If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is Fréchet differentiable at  $\bar{x}$ , then  $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ .

**Example 1.0.20.** Let  $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$  be defined by  $\rho(x) := \|x\|$ . Then

$$\partial \rho(x) = \begin{cases} \mathbb{B} & \text{if } x = 0, \\ \frac{x}{\|x\|} & \text{otherwise.} \end{cases}$$

The equality is routine if  $x \neq 0$  because the function is differentiable in this case. Our main concern is calculating the subdifferential at  $x = 0$ . Let  $v \in \partial \rho(0)$ , then by definition we have that  $\langle v, x \rangle \leq \|x\|$  for all  $x \in \mathbb{R}^n$ . This implies that  $\langle v, v \rangle \leq \|v\|$ .

It then follows that  $\|v\| \leq 1$ , so  $v \in \mathbb{B}$ . The opposite inclusion follows from the Cauchy-Schwarz inequality. Indeed, take any  $v \in \mathbb{B}$ , then

$$\langle v, x - 0 \rangle = \langle v, x \rangle \leq \|v\| \|x\| \leq \|x\| = p(x) - p(0) \text{ for all } x \in \mathbb{R}^n.$$

Therefore  $v \in \partial\rho(0)$  and we have  $\partial\rho(0) = \mathbb{B}$ .

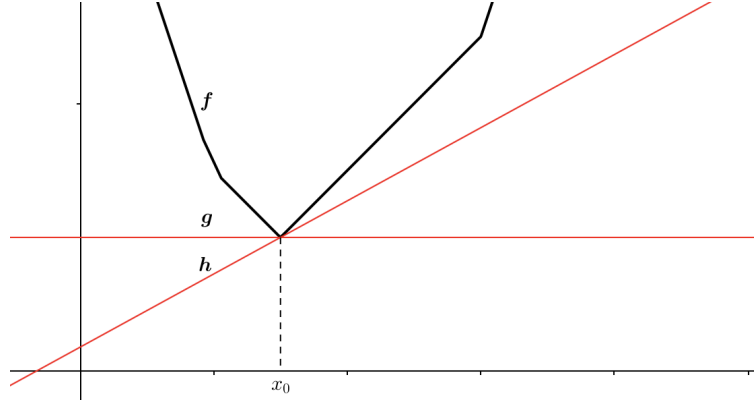


FIGURE 1.0.3: Choosing two values from the subdifferential of  $f$  at  $x_0$  we attain the *subderivatives*  $g$  and  $h$  at  $x_0$ . Notice  $g$  is a constant.

**Proposition 1.0.21.** Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be convex and  $\bar{x} \in \text{dom}f$  be a local minimizer of  $f$ . Then  $f$  attains its global minimum at this point.

*Proof.* By our assumptions, for some  $\delta > 0$  we have

$$f(\bar{x}) \leq f(u) \text{ for all } u \in \mathbb{B}(\bar{x}; \delta).$$

Now let  $x \in \mathbb{R}^n$  be given. Then any  $z \in \mathbb{B}(\bar{x}; \delta)$  is expressible as a convex combination of  $x$  and  $\bar{x}$ . That is, for  $\lambda \in [0, 1]$  we have  $z = \lambda x + (1 - \lambda)\bar{x}$ . Now since  $f$  is convex,

$$f(\bar{x}) \leq f(z) \leq \lambda f(x) + (1 - \lambda)f(\bar{x}).$$

This inequality implies that  $\lambda f(\bar{x}) \leq \lambda f(x)$ . Further,  $f(\bar{x}) \leq f(x)$  for all  $x \in \mathbb{R}^n$ .  $\square$

**Proposition 1.0.22.** Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be convex and  $\bar{x} \in \text{dom}f$ . Then  $f$  attains its local/global minimum at  $\bar{x}$  if and only if  $0 \in \partial f(\bar{x})$ .

*Proof.* First suppose  $f$  attains its global minimum at  $\bar{x}$ . Then

$$0 = \langle 0, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

which implies  $0 \in \partial f(\bar{x})$  by the definition of subdifferential. The sufficient condition follows by a similar argument.  $\square$

**Definition 1.0.23.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is **lower semicontinuous** at  $\bar{x} \in \mathbb{R}^n$  if for every  $\alpha \in \mathbb{R}$  with  $f(\bar{x}) > \alpha$  there is  $\delta > 0$  such that

$$f(x) > \alpha \text{ for all } x \in B(\delta; \bar{x}).$$

**Definition 1.0.24.** Given a proper function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  its **Fenchel conjugate**  $f^* : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is

$$f^*(v) := \sup\{\langle v, x \rangle - f(x) : x \in \mathbb{R}^n\} = \sup\{\langle v, x \rangle - f(x) : x \in \text{dom} f\}.$$

**Remark 1.0.25.** It is worth noting that  $f$  is not necessarily convex on  $\mathbb{R}^n$ , but its conjugate  $f^*$  is. Further, given  $f$  is proper, convex, and lower semicontinuous, then  $x \in \partial g(y)$  if and only if  $y \in \partial g^*(x)$ .

**Definition 1.0.26.** A function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is **Lipschitz continuous** on a set  $\Omega \subset \text{dom} f$  if there is a constant  $\ell \geq 0$  such that

$$|f(x) - f(y)| \leq \ell \|x - y\| \text{ for all } x, y \in \Omega. \quad (1.0.1)$$

Also,  $f$  is called **locally Lipschitz continuous** around  $\bar{x} \in \text{dom} f$  if there are constants  $\ell \geq 0$  and  $\delta > 0$  such that (1.0.1) holds, where  $\Omega = \mathbb{B}(\bar{x}; \delta)$ .

The following theorem is a nice result between the concept of Lipschitz continuity and its local analog. We see how local boundedness of a convex function is used to show its local Lipschitz continuity.

**Theorem 1.0.27.** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex and  $\bar{x} \in \text{dom} f$ . If  $f$  is bounded above on  $\mathbb{B}(\bar{x}; \delta)$  for some  $\delta > 0$ , then  $f$  is Lipschitz continuous on  $\mathbb{B}(\bar{x}; \frac{\delta}{2})$ .

*Proof.* The proof is omitted here and the reader is referred to [10]. □

**Corollary 1.0.28.** Any finite convex function on  $\mathbb{R}^n$  is locally Lipschitz continuous on  $\mathbb{R}^n$ .

*Proof.* Fix  $x \in \mathbb{R}^n$  and let  $\varepsilon > 0$ . Define  $\mathcal{A} := \{x \pm \varepsilon e_i : i = 1, \dots, n\} \subset \mathbb{R}^n$  where  $\{e_i : i = 1, \dots, n\}$  is the standard orthonormal basis of  $\mathbb{R}^n$ . Further,  $M := \max\{f(a) : a \in \mathcal{A}\}$  is finite. We express the convex combination of any  $\bar{x} \in \mathbb{B}(x; \frac{\varepsilon}{n})$  as

$$x = \sum_{i=1}^m \lambda_i a_i \text{ with } \sum_{i=1}^m \lambda_i = 1, \text{ where } a_i \in \mathcal{A} \text{ and } \lambda_i \geq 0.$$

The following is then obtained

$$f(x) \leq \sum_{i=1}^m \lambda_i f(a_i) \leq \sum_{i=1}^m \lambda_i M = M. \quad (1.0.2)$$

Thus,  $f$  is bounded above on  $\mathbb{B}(x; \frac{\varepsilon}{n})$ . By Theorem 1.0.27,  $f$  is Lipschitz continuous on  $\mathbb{B}(x; \frac{\varepsilon}{2n})$ , which implies it is locally Lipschitz on  $\mathbb{R}^n$ .  $\square$

**Remark 1.0.29.** We made the assumption that  $\mathbb{B}(\bar{x}; \frac{\varepsilon}{n}) \subset \text{co}(\mathcal{A})$ . Also, (1.0.2) relies on *Jensen's inequality*. The reader is referred to [10] for more information.

**Definition 1.0.30.** A function  $f: \mathbb{R}^n \rightarrow (\infty, \infty]$  is **coercive** if

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

**Definition 1.0.31.** Let  $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$  be an extended-real-valued function. We say that  $\varphi$  is  **$\gamma$ -convex** with parameter  $\gamma \geq 0$  if the function  $g: \mathbb{R}^n \rightarrow (-\infty, \infty]$  given by  $g(x) := f(x) - \frac{\gamma}{2}\|x\|^2$  is convex. We say  $f$  is **strongly convex** if the convexity of  $g$  holds for some  $\gamma > 0$ .

# Chapter 2

## Optimization Methods

This chapter outlines basic properties of functions expressible as the difference of convex functions (DC functions). We begin with basic properties of such functions, then introduce the reader to DC programming and the DCA. Nesterov's smoothing technique and accelerated gradient method are also discussed.

### 2.1 The Difference of Convex Functions

**Definition 2.1.1.** Given a convex set  $\Omega$ , a function  $f : \Omega \rightarrow \mathbb{R}$  is called a **DC function** if there exists a pair of convex functions  $g, h$  on  $\Omega$  such that  $f = g - h$ .

**Proposition 2.1.2.** Let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  be DC with the DC decomposition  $f_i = g_i - h_i$  for  $i = 1, \dots, m$ . Then the following functions are also DC:

(a)  $\sum_{i=1}^m \lambda_i f_i$

(d)  $|f_i|$  for  $i = 1, \dots, m$

(b)  $\max_{i=1, \dots, m} f_i$

(c)  $\min_{i=1, \dots, m} f_i$

(e)  $\prod_{i=1}^m f_i$ .

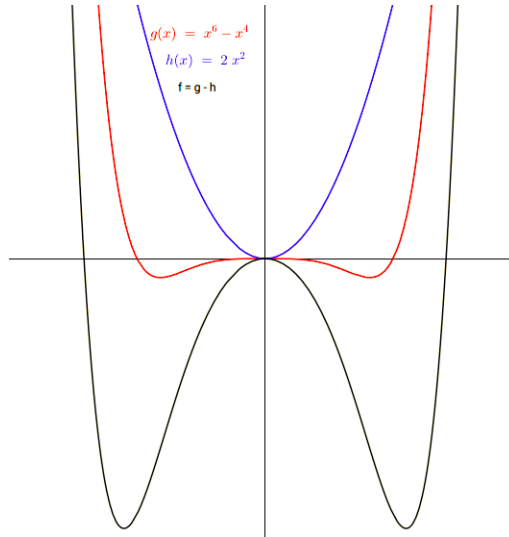


FIGURE 2.1.1: A DC function and its decomposition.

*Proof.* These results are shown using propositions 1.0.14 and 1.0.15.

- (a) Let  $I = \{i = 1, \dots, m : \lambda_i \geq 0\}$  and  $J = \{1, \dots, m\} \setminus I$ . Also let  $\bar{\lambda}_j = -\lambda_j$  for  $j \in J$ . Then the following holds

$$\begin{aligned} \sum_{i=1}^m \lambda_i f_i &= \sum_{i \in I} \lambda_i (g_i - h_i) + \sum_{j \in I} \lambda_j (g_j - h_j) \\ &= \left[ \sum_{i \in I} \lambda_i g_i + \sum_{i \in I} \lambda_i (-h_i) \right] + \left[ \sum_{j \in J} \lambda_j g_j + \sum_{j \in J} \bar{\lambda}_j h_j \right] \\ &= \left[ \sum_{i \in I} \lambda_i g_i + \sum_{j \in J} \bar{\lambda}_j h_j \right] - \left[ \sum_{i \in I} \lambda_i h_i + \sum_{j \in J} \bar{\lambda}_j g_j \right]. \end{aligned}$$

- (b) Notice that for any  $i = 1, \dots, m$  that  $f_i = g_i + \sum_{j=1, j \neq i}^m h_j - \sum_{j=1}^m h_j$ . Then we have

$$\max_{i=1, \dots, m} f_i = \max_{i=1, \dots, m} \left( g_i + \sum_{j=1, j \neq i}^m h_j \right) - \sum_{j=1}^m h_j.$$

- (c) This is shown in a similar fashion as (b).  
 (d) For  $i = 1, \dots, m$  notice that  $|f_i| = \max\{f_i, -f_i\}$ . The result follows by part (b).  
 (e) The case of  $m = 2$  is given by Philip Hartman in [3] as a corollary to the main theorems. The general result follows by mathematical induction.

□

### 2.1.1 An Introduction DC programming

Here we provide a brief introduction to DC programming and the DC algorithm; for a comprehensive study the reader is referred to [8, 17, 6]. Convexity is a desirable property of functions in that local optimal solutions are global solutions. As convex optimization has been studied for a quite long time, going beyond convexity is of great interest in research of mathematical optimization. One of the first approaches in this direction is to consider minimizing differences of convex functions. DC programming utilizes the convexity in the decomposition of the original objective function. Consider the following optimization problem:

$$\text{minimize } f(x) = g(x) - h(x), x \in \mathbb{R}^n, \quad (2.1.1)$$

where  $g: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  and  $h: \mathbb{R}^n \rightarrow \mathbb{R}$  are convex. If  $f$  is our objective function, then  $g - h$  is called a *DC decomposition* of the function  $f$ . The problem illustrated by (2.1.1) is the *primal* DC program and it is always expressible as an unconstrained optimization problem. Indeed, if the objective function is constrained to a convex set  $\Omega$ , adding the indicator function  $\delta_\Omega$ , where  $\delta_\Omega(x) = 0$  if  $x \in \Omega$  and  $\delta_\Omega(x) = \infty$  otherwise, to  $g$  yields an unconstrained DC program.

In a similar fashion, we define the *dual* DC program of (2.1.1) in terms of the Fenchel conjugates of  $g$  and  $h$

$$\text{minimize } h^*(y) - g^*(y) \text{ subject to } y \in \text{dom} h^* \quad (2.1.2)$$

The DC algorithm introduced by Tao and An [17, 18] uses components from both of these objective functions; it is given as follows:

#### The DCA

```

INPUT:  $x_1 \in \mathbb{R}^n, N \in \mathbb{N}$ 
for  $k = 1, \dots, N$  do
    Find  $y_k \in \partial h(x_k)$ 
    Find  $x_{k+1} \in \partial g^*(y_k)$ 
end for
OUTPUT:  $x_{N+1}$ 

```

We now discuss the sufficient conditions for the sequence  $\{x_k\}$  to be generated.

**Proposition 2.1.3.** Let  $g: \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a proper lower semicontinuous convex function. Then

$$\partial g(\mathbb{R}^n) := \bigcup_{x \in \mathbb{R}^n} \partial g(x) = \text{dom } \partial g^* = \{x \in \mathbb{R}^n \mid \partial g^*(x) \neq \emptyset\}.$$



*Proof.* Let  $y \in \partial g(x)$  for some  $x \in \mathbb{R}^n$ . Then  $x \in \partial g^*(y)$ , which implies  $\partial g^*(y) \neq \emptyset$ , and so  $y \in \text{dom } \partial g^*$ . A similar argument gives the opposite inclusion.  $\square$

**Remark 2.1.4.** If we assume  $g$  is coercive and level-bounded, then  $\text{dom } \partial g^* = \text{dom } g^* = \mathbb{R}^n$ . The proof is given in [13].

**Proposition 2.1.5.** Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  be a proper lower semicontinuous convex function. Then  $v \in \partial g^*(y)$  if and only if

$$v \in \text{argmin}\{g(x) - \langle y, x \rangle : x \in \mathbb{R}^n\} \quad (2.1.3)$$

Further,  $w \in \partial h(x)$  if and only if

$$w \in \text{argmin}\{h^*(y) - \langle y, x \rangle : y \in \mathbb{R}^n\} \quad (2.1.4)$$

*Proof.* Suppose  $v \in \text{argmin}\{g(x) - \langle y, x \rangle : x \in \mathbb{R}^n\}$ . Then we see

$$0 \in \partial g(v) - y$$

This implies  $y \in \partial g(v)$ , and we conclude  $v \in \partial g^*(y)$ . Showing the other implication, if  $v \in \partial g^*(y)$ , then  $0 \in \partial g(v) - y$ , which implies (2.1.3).

Similarly, we see that if  $w \in \text{argmin}\{h^*(y) - \langle y, x \rangle : y \in \mathbb{R}^n\}$ , then we have

$$0 \in \partial h^*(w) - x$$

Therefore,  $w \in \partial h(x)$ . The proof that (2.1.4) implies  $w \in \partial h(x)$  is justified as before.  $\square$

## 2.1.2 Convergence of the DC Algorithm

In this section we develop some fundamental results regarding the convergence of the DC Algorithm. The reader is referred to [6, 8, 17, 18] for a more extensive study on DC programming.

**Proposition 2.1.6.** Let  $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be  $\gamma$ -convex with  $\bar{x} \in \text{dom}(h)$ . Then  $v \in \partial h(\bar{x})$  if and only if

$$\langle v, x - \bar{x} \rangle + \frac{\gamma}{2} \|x - \bar{x}\|^2 \leq h(x) - h(\bar{x}) \quad (2.1.5)$$

*Proof.* Let  $k : \mathbb{R}^n \rightarrow (-\infty, \infty]$  defined as  $k(x) = h(x) - \frac{\gamma}{2} \|x\|^2$  be a convex function. Given  $v \in \partial h(\bar{x})$  we apply the subdifferential sum rule to attain  $v \in \partial k(\bar{x}) + \gamma \bar{x}$ ,

which implies  $v - \gamma\bar{x} \in \partial k(\bar{x})$ . By definition of subgradient we see

$$\langle v - \gamma\bar{x}, x - \bar{x} \rangle \leq k(x) - k(\bar{x}) \text{ for all } x \in \mathbb{R}^n.$$

Therefore,

$$\begin{aligned} \langle v, x - \bar{x} \rangle &\leq \gamma\langle \bar{x}, x \rangle - \gamma\langle \bar{x}, \bar{x} \rangle + h(x) - \frac{\gamma}{2}\|x\|^2 - (h(\bar{x}) - \frac{\gamma}{2}\|\bar{x}\|^2) \\ &\leq h(x) - h(\bar{x}) - \frac{\gamma}{2}(\|x\|^2 - 2\langle x, \bar{x} \rangle + \|\bar{x}\|^2) \\ &= h(x) - h(\bar{x}) - \frac{\gamma}{2}\|x - \bar{x}\|^2 \end{aligned}$$

We then obtain

$$\langle v, x - \bar{x} \rangle + \frac{\gamma}{2}\|x - \bar{x}\|^2 \leq h(x) - h(\bar{x}).$$

Conversely,

$$\langle v, x - \bar{x} \rangle \leq \langle v, x - \bar{x} \rangle + \frac{\gamma}{2}\|x - \bar{x}\|^2 \leq h(x) - h(\bar{x})$$

This implies  $v \in \partial h(\bar{x})$ . □

**Proposition 2.1.7.** Consider the  $f$  defined by (2.1.1) and the sequence  $\{x_k\}$  generated by the DC algorithm. Let  $g$  be  $\gamma_1$ -convex and  $h$  be  $\gamma_2$ -convex. Then

$$\frac{\gamma_1 + \gamma_2}{2}\|x_{k+1} - x_k\|^2 \leq f(x_k) - f(x_{k+1}) \text{ for all } k \in \mathbb{N} \quad (2.1.6)$$

*Proof.* Given  $y_k \in \partial h(x_k)$ , we apply Proposition 2.1.6 to obtain

$$\langle y_k, x_{k+1} - x_k \rangle + \frac{\gamma_2}{2}\|x_{k+1} - x_k\|^2 \leq h(x_{k+1}) - h(x_k)$$

Likewise, we have  $x_{k+1} \in \partial g^*(y_k)$ , which implies  $y_k \in \partial g(x_{k+1})$ . Therefore,

$$\langle y_k, x_k - x_{k+1} \rangle + \frac{\gamma_1}{2}\|x_k - x_{k+1}\|^2 \leq g(x_k) - g(x_{k+1})$$

The sum of these inequalities implies (2.1.6). □

**Lemma 2.1.1.** Suppose  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  is a convex function with  $w_k \in \partial h(x_k)$  where  $\{x_k\}$  is a bounded sequence. Then  $\{w_k\}$  is bounded.

*Proof.* Fix any  $\bar{x} \in \mathbb{R}^n$ . Since  $\{h(x_k)\}$  is bounded, by Corollary 1.0.28 there is  $\delta > 0$  and  $\ell > 0$  such that

$$|h(x) - h(y)| \leq \ell\|x - y\| \text{ whenever } x, y \in \mathbb{B}(\bar{x}; \delta).$$

Now we show  $\|w\| \leq \ell$  whenever  $w \in \partial h(u)$  for  $u \in \mathbb{B}(\bar{x}; \frac{\delta}{2})$ .

Indeed,

$$\langle w, x - u \rangle \leq h(x) - h(u) \text{ for all } x \in \mathbb{R}^n$$

Now choose positive  $\eta$  such that  $\mathbb{B}(u; \eta) \subset \mathbb{B}(\bar{x}; \delta)$ . Then

$$\langle w, x - u \rangle \leq h(x) - h(u) \leq \ell \|x - u\| \text{ whenever } \|x - u\| \leq \eta.$$

Thus,  $\|w\| \leq \ell$ .

Let us suppose by way of contradiction that  $\{w_k\}$  is not bounded. Then assume  $\|w_k\| \rightarrow \infty$ . By the Bolzano-Weierstrass theorem, the boundedness of  $\{x_k\}$  implies there is a subsequence  $\{x_{k_p}\}$  that converges to  $x_0 \in \mathbb{R}^n$ . But for Lipschitz constant  $\ell' > 0$  of  $f$  around  $x_0$ , we have

$$\|w_{k_p}\| \leq \ell' \text{ for some } p.$$

This contradicts our assumption, and a similar argument is given if  $\|w_k\| \rightarrow -\infty$ .  $\square$

**Theorem 2.1.8.** Let  $f$  be the function defined in (2.1.1) and the sequence  $\{x_k\}$  be generated by the DC algorithm. Then  $\{f(x_k)\}$  is a decreasing sequence. Also suppose  $f$  is bounded below,  $g$  is  $\gamma_1$ -convex, and  $h$  is  $\gamma_2$ -convex. If  $\{x_k\}$  is bounded, then every subsequential limit  $\bar{x}$  of the sequence  $\{x_k\}$  is a stationary point of  $f$ . That is,  $\partial g(\bar{x}) \cap \partial h(\bar{x}) \neq \emptyset$ .

*Proof.* Since  $f$  is bounded below, from (2.1.6) it follows that  $f$  converges to a real number. That is,  $f(x_k) - f(x_{k+1}) \rightarrow 0$  as  $k \rightarrow \infty$ . Applying (2.1.6) again we see  $\|x_{k+1} - x_k\| \rightarrow 0$ . Suppose  $x_{k_\ell} \rightarrow \bar{x}$  as  $\ell \rightarrow \infty$ . Let

$$y_k \in \partial g(x_{k+1}) \text{ for all } k \in \mathbb{N}.$$

Then by Lemma 2.1.1,  $\{y_k\}$  is a bounded sequence. Suppose  $y_{k_\ell} \rightarrow \bar{y}$  as  $\ell \rightarrow \infty$ . By observing that  $y_{k_\ell} \in \partial h(x_{k_\ell})$  for all  $\ell \in \mathbb{N}$ , we now show

$$\bar{y} \in \partial h(\bar{x}).$$

Since  $\langle y_{k_\ell}, x - x_{k_\ell} \rangle \leq h(x) - h(x_{k_\ell})$  for all  $x \in \mathbb{R}^n$  and  $\ell \in \mathbb{N}$ , we have

$$\langle y_{k_\ell}, \bar{x} - x_{k_\ell} \rangle \leq h(\bar{x}) - h(x_{k_\ell})$$

Then  $h(x_{k_\ell}) \leq \langle y_{k_\ell}, x_{k_\ell} - \bar{x} \rangle + h(\bar{x})$ , which implies  $\limsup h(x_{k_\ell}) \leq h(\bar{x})$ . Since  $h$  is lower semicontinuous,  $h(x_{k_\ell}) \rightarrow h(\bar{x})$ . Letting  $\ell \rightarrow \infty$  we obtain  $\bar{y} \in \partial h(\bar{x})$ .

Confirming that  $\bar{y} \in \partial g(\bar{x})$  follows a similar argument after observing  $x_{k_\ell+1} \rightarrow \bar{x}$  and  $y_{k_\ell} \in \partial g(x_{k_\ell+1})$ . Thus,  $\bar{x}$  is a stationary point of  $f$ .  $\square$

## 2.2 Nesterov's Smoothing Technique

In this section we study Nesterov's smoothing technique with respect to nonsmooth functions in  $\mathbb{R}^n$ . The results in this section are based on Nesterov's 2005 paper [14] and are a part of a current project where we discuss the following results in Hilbert spaces [12].

For the duration of this section, let  $A \in \mathbb{R}^{m \times n}$ . Given a nonempty closed bounded convex subset  $Q$  of  $\mathbb{R}^m$  and a continuous convex function  $\phi: \mathbb{R}^m \rightarrow \mathbb{R}$ , consider the function  $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$f_0(x) := \max\{\langle Ax, u \rangle - \phi(u) \mid u \in Q\}, \quad x \in \mathbb{R}^n.$$

We define the norm of  $A$  as usual:

$$\|A\| := \sup \{\|Ax\| \mid \|x\| \leq 1\}.$$

It follows from the definition that  $\|Ax\| \leq \|A\|\|x\|$  for all  $x \in \mathbb{R}^n$ . The transpose of  $A$  denoted by  $A^\top: \mathbb{R}^m \rightarrow \mathbb{R}^n$  satisfies the identity

$$\langle x, A^\top(y) \rangle = \langle Ax, y \rangle \text{ for all } x \in \mathbb{R}^n, y \in \mathbb{R}^m,$$

Let us prove a well-known result that involves the norm of an  $m \times n$  matrix.

**Proposition 2.2.1.** Given  $A \in \mathbb{R}^{m \times n}$ , we have  $\|A\| = \|A^\top\|$ .

*Proof.* For any  $x \in \mathbb{R}^n$ , one has

$$\|Ax\| = \sup_{\|y\| \leq 1} \langle Ax, y \rangle = \sup_{\|y\| \leq 1} \langle x, A^\top(y) \rangle \leq \sup_{\|y\| \leq 1} (\|x\| \|A^\top\| \|y\|) \leq \|A^\top\| \|x\|.$$

This implies  $\|A\| \leq \|A^\top\|$ . In addition,  $\|A^\top\| \leq \|(A^\top)^\top\| = \|A\|$ . □

**Proposition 2.2.2.** Let  $\varphi: \mathbb{R}^n \rightarrow (-\infty, \infty]$  be strongly convex with parameter  $\sigma > 0$ . Then

$$\sigma \|x_1 - x_2\|^2 \leq \langle v_1 - v_2, x_1 - x_2 \rangle \tag{2.2.1}$$

whenever  $v_i \in \partial\varphi(x_i)$  for  $i = 1, 2$ .

*Proof.* Define  $g(x) := \varphi(x) - \frac{\sigma}{2}\|x\|^2$  for  $x \in \mathbb{R}^n$ . Then  $g$  is convex and by the subdifferential sum rule

$$v_i \in \partial g(x_i) + \sigma x_i \text{ for } i = 1, 2.$$

By the definition of subdifferentials in the sense of convex analysis,

$$\begin{aligned}\langle v_1 - \sigma x_1, x_2 - x_1 \rangle &\leq g(x_2) - g(x_1), \\ \langle v_2 - \sigma x_2, x_1 - x_2 \rangle &\leq g(x_1) - g(x_2).\end{aligned}$$

Adding these inequalities yields

$$\langle v_1 - v_2 - \sigma(x_1 - x_2), x_2 - x_1 \rangle \leq 0,$$

which implies (2.2.1).  $\square$

**Lemma 2.2.1.** Let  $A \in \mathbb{R}^{m \times n}$ . Suppose that  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  is a strongly convex function with parameter  $\mu > 0$ . Then the function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$f(x) := \max\{\langle Ax, u \rangle - \varphi(u) \mid u \in Q\}$$

is well-defined and Fréchet differentiable with Lipschitz gradient with constant  $\ell = \frac{\|A\|^2}{\mu}$ .

*Proof.* The subdifferential of the function  $f$  is given by

$$\partial f(x) = \{A^\top u(x)\} \text{ for all } x \in \mathbb{R}^n,$$

where  $u(x) \in \mathbb{R}^m$  denotes the unique element such that the maximum is attained in the definition of  $f(x)$ .

Let  $g_x(u) = -\langle Ax, u \rangle + \varphi(u) + \delta(x; Q)$ . Then

$$0 \in \partial g_x(u(x)) = -Ax + \partial g(u(x)),$$

where  $g(x) := h(x) + \delta(x; Q)$  is strongly convex with parameter  $\mu$ . It follows that

$$Ax \in \partial g(u(x)). \tag{2.2.2}$$

The function  $g$  is strongly convex with parameter  $\mu$ , so its subdifferential is strongly monotone in the sense that

$$\mu \|x_1 - x_2\|^2 \leq \langle w_1 - w_2, x_1 - x_2 \rangle \text{ whenever } w_i \in \partial g(x_i).$$

Using (2.2.2) gives

$$\begin{aligned}\mu \|u(x_1) - u(x_2)\|^2 &\leq \langle A(x_1) - A(x_2), u(x_1) - u(x_2) \rangle \\ &= \langle x_1 - x_2, A^\top(u(x_1)) - A^\top(u(x_2)) \rangle \\ &\leq \|x_1 - x_2\| \|A^\top(u(x_1)) - A^\top(u(x_2))\|.\end{aligned}$$

It follows that

$$\begin{aligned} \|A^\top(u(x_1)) - A^\top(u(x_2))\|^2 &\leq \|A^\top\|^2 \|u(x_1) - u(x_2)\|^2 \\ &\leq \frac{\|A\|^2}{\mu} \|x_1 - x_2\| \|A^\top(u(x_1)) - A^\top(u(x_2))\|. \end{aligned}$$

This implies

$$\|A^\top(u(x_1)) - A^\top(u(x_2))\| \leq \frac{\|A\|^2}{\mu} \|x_1 - x_2\|.$$

Thus, the subdifferential mapping  $\partial\mu(\cdot)$  is continuous, and so  $f$  is Fréchet differentiable with  $\partial f(x) = \{\nabla f(x)\} = \{A^\top u(x)\}$ . In addition,

$$\|\nabla\mu(x_1) - \nabla\mu(x_2)\| \leq \frac{\|A\|^2}{\mu} \|x_1 - x_2\| \text{ for all } x_1, x_2 \in \mathbb{R}^n.$$

The proof is now complete.  $\square$

Fix a constant  $\mu > 0$  and  $u_0 \in Q$ . Consider the function  $f_\mu$  defined on  $\mathbb{R}^n$  be

$$f_\mu(x) := \max\{\langle Ax, u \rangle - \phi(u) - \frac{\mu}{2} \|u - u_0\|^2 \mid u \in Q\}. \quad (2.2.3)$$

**Proposition 2.2.3.** The function  $f_\mu$  given by (2.2.3) is Fréchet differentiable with Lipschitz gradient on  $\mathbb{R}^n$ . Its gradient is given by

$$\nabla f_\mu(x) = A^\top u_\mu(x).$$

The gradient  $f_\mu$  is Lipschitz continuous on  $\mathbb{R}^n$  with Lipschitz constant  $\ell = \frac{\|A\|^2}{\mu}$ . In addition,

$$f_\mu(x) \leq f_0(x) \leq f_\mu(x) + \frac{\mu}{2} [D(u_0; Q)]^2 \text{ for all } x \in \mathbb{R}^n. \quad (2.2.4)$$

*Proof.* Observe that the function  $\varphi(u) := \phi(u) + \frac{\mu}{2} \|u - u_0\|^2$  is strongly convex with parameter  $\mu$ . Thus, the fact that  $f_\mu$  is Fréchet differentiable with Lipschitz continuous gradient follows from Lemma 2.2.1. The estimate (2.2.4) is straightforward. Obviously,  $f_\mu(x) \leq f_0(x)$  for all  $x \in \mathbb{R}^n$ . In addition,

$$\begin{aligned} f_0(x) &= \max\{\langle Ax, u \rangle - \phi(u) \mid u \in Q\} \\ &= \max\{\langle Ax, u \rangle - \phi(u) - \frac{\mu}{2} \|u - u_0\| + \frac{\mu}{2} \|u - u_0\| \mid u \in Q\} \\ &= \max\{\langle Ax, u \rangle - \phi(u) - \frac{\mu}{2} \|u - u_0\| \mid u \in Q\} + \max\{\frac{\mu}{2} \|u - u_0\| \mid u \in Q\} \\ &= f_\mu(x) + [D(u_0; Q)]^2. \end{aligned}$$

The proof is now complete.  $\square$

In the next proposition, we consider a particular case of the function  $f_0$  where  $\phi$  is a linear function  $\phi(u) = \langle b, u \rangle$ , where  $b \in \mathbb{R}^n$ .

**Proposition 2.2.4.** In the setting of Proposition 2.2.3, consider the function  $\phi$  given by  $\phi(u) = \langle b, u \rangle$ , where  $b \in \mathbb{R}^n$ . Then the function  $f_\mu$  in has the explicit representation

$$f_\mu(x) = \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u} \rangle - \frac{\mu}{2} \left[ d\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right) \right]^2$$

and is continuously differentiable on  $\mathbb{R}^n$  with its gradient given by

$$\nabla f_\mu(x) = A^\top u_\mu(x),$$

where  $u_\mu$  can be expressed in terms of the Euclidean projection

$$u_\mu(x) = P\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right).$$

*Proof.* We have

$$\begin{aligned} f_\mu(x) &= \sup\left\{ \langle Ax - b, u \rangle - \frac{\mu}{2} \|u - \bar{u}\|^2 \mid u \in Q \right\} \\ &= \sup\left\{ -\frac{\mu}{2} (\|u - \bar{u}\|^2 - \frac{2}{\mu} \langle Ax - b, u \rangle) \mid u \in Q \right\} \\ &= -\frac{\mu}{2} \inf\left\{ \left\| u - \bar{u} - \frac{Ax - b}{\mu} \right\|^2 - \frac{\|Ax - b\|^2}{\mu^2} - \frac{2}{\mu} \langle Ax - b, \bar{u} \rangle \mid u \in Q \right\} \\ &= \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u} \rangle - \frac{\mu}{2} \inf\left\{ \left\| u - \bar{u} - \frac{Ax - b}{\mu} \right\|^2 \mid u \in Q \right\} \\ &= \frac{\|Ax - b\|^2}{2\mu} + \langle Ax - b, \bar{u} \rangle - \frac{\mu}{2} \left[ d\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right) \right]^2. \end{aligned}$$

Since the function  $\psi(u) := [d(u; Q)]^2$  is continuously differentiable with  $\nabla\psi(u) = 2[u - P(u; Q)]$  for all  $u \in \mathbb{R}^m$  (see, e.g., [5, p. 186]), it follows from the chain rule that

$$\begin{aligned} \nabla f_\mu(x) &= \frac{1}{\mu} A^\top (Ax - b) + A^\top \bar{u} - \frac{\mu}{2} \left[ \frac{2}{\mu} A^\top \left( \bar{u} + \frac{Ax - b}{\mu} - P\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right) \right) \right] \\ &= A^\top P\left(\bar{u} + \frac{Ax - b}{\mu}; Q\right). \end{aligned}$$

The proof is now complete.  $\square$

# Chapter 3

## The Fermat-Torricelli Problem

In this section we give a brief survey of the classical Fermat-Torricelli problem then introduce the constrained version. Detailed proofs are given and we solve the problem using the distance penalty method, Nesterov's smoothing technique, and DC programming. In addition, we introduce an algorithm for solving the constrained problem.

### 3.1 The Original Problem

The precise statement of the Fermat-Torricelli problem is as follows: Given a finite set  $\mathcal{A} := \{a_1, \dots, a_m\} \subset \mathbb{R}^n$ ,

$$\text{minimize } f(x) := \sum_{i=1}^m \|x - a_i\| \text{ subject to } x \in \mathbb{R}^n. \quad (3.1.1)$$

**Remark 3.1.1.** Recall  $f$  is convex. Further,  $f$  is continuous.

#### 3.1.1 Existence and Uniqueness of Solutions

Let us discuss the existence and uniqueness of solutions to this problem.



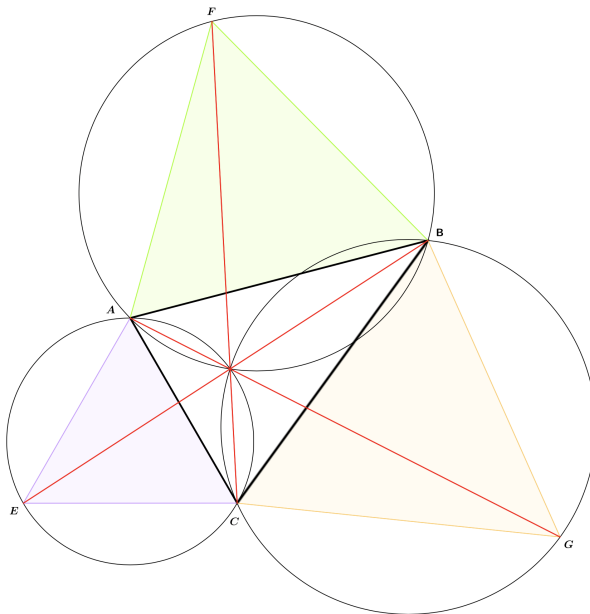


FIGURE 3.1.1: The geometric construction for solving the classical problem raised by Fermat. If  $\triangle ABC$  has an angle greater than or equal to  $120^\circ$ , then the corresponding vertex is the solution.

**Proposition 3.1.2.** The solution set of the Fermat-Torricelli problem (3.1.1) is nonempty.

*Proof.* First, let  $m = \inf\{f(x) : x \in \mathbb{R}^n\}$ . Then choose a sequence  $\{x_k\}_{k=1}^\infty \subset \mathbb{R}^n$  such that  $\lim_{k \rightarrow \infty} f(x_k) = m$ . By the definition of convergence, we can find  $K \in \mathbb{N}$  such that  $\|x_k - a_1\| \leq f(x_k) < m + 1$  whenever  $k \geq K$ . Then  $\|x_k\| \leq m + \|a_1\| + 1$  whenever  $k \geq K$ . Thus,  $\{x_k\}_{k=1}^\infty$  is bounded, so it has a convergent subsequence  $\{x_{k_l}\}_{l=1}^\infty$  that converges to some  $\bar{x} \in \mathbb{R}^n$ . Then by the continuity of  $f$ , we have

$$f(\bar{x}) = \lim_{l \rightarrow \infty} f(x_{k_l}) = m.$$

Thus,  $\bar{x}$  is an optimal solution to the problem. □

**Proposition 3.1.3.** If  $\mathcal{A}$  is not collinear, then the function  $f$  from (3.1.1) is strictly convex and the Fermat-Torricelli problem has a unique solution.

*Proof.* Define  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  by  $f_i(x) := \|x - a_i\|$  for each  $a_i \in \mathcal{A}$ . Then  $f = \sum_{i=1}^m f_i$  is convex. By way of contradiction, suppose  $f$  is not strictly convex. Then there exist  $x, y \in \mathbb{R}^n$ , where  $x \neq y$ , such that for any  $\lambda \in (0, 1)$  we have  $f(\lambda x + (1 - \lambda)y) =$

$\lambda f(x) + (1 - \lambda)f(y)$ . It follows that

$$\|\lambda(x - a_i) + (1 - \lambda)(y - a_i)\| = \|\lambda(x - a_i)\| + \|(1 - \lambda)(y - a_i)\| \text{ for } a_i \in \mathcal{A}.$$

Suppose  $x - a_i \neq 0$  and  $y - a_i \neq 0$ . Then for each  $i = 1, \dots, m$  there is  $t_i > 0$  such that  $t_i \lambda(x - a_i) = (1 - \lambda)(y - a_i)$ . Let  $\gamma_i = \frac{1 - \lambda}{t_i \lambda}$ . Solving for  $a_i$  we obtain

$$a_i = \frac{1}{1 - \gamma_i}x - \frac{\gamma_i}{1 - \gamma_i}y \in \mathcal{L}(x, y),$$

where  $\mathcal{L}(x, y)$  is the line connecting  $x$  and  $y$ . In the case where  $x = a_i$  or  $y = a_i$ , it is obvious that  $a_i \in \mathcal{L}(x, y)$ . Thus, we have arrived at a contradiction.  $\square$

### 3.1.2 Weiszfeld's Algorithm

For the duration of this section we assume the points in  $\mathcal{A}$  are not collinear. Given  $f$  from (3.1.1) we have

$$\nabla f(x) = \sum_{i=1}^m \frac{x - a_i}{\|x - a_i\|}, \quad x \notin \mathcal{A}.$$

Solving the equation  $\nabla f(x) = 0$  we obtain

$$x = \frac{\sum_{i=1}^m \frac{a_i}{\|x - a_i\|}}{\sum_{i=1}^m \frac{1}{\|x - a_i\|}}. \quad (3.1.2)$$

Then we define  $F(x) := x$  as in (3.1.2) for  $x \in \mathcal{A}$ .

**Weiszfeld's Algorithm.**

```

INPUT: Given  $\mathcal{A} = \{a_1, a_2, \dots, a_m\} \subset \mathbb{R}^n$ ,  $x_0 \in \mathbb{R}^n$  and  $N \in \mathbb{N}$ 
for  $k = 0, 1, \dots, N$  do
     $x_{k+1} = F(x_k)$ 
end for
OUTPUT:  $x_{N+1}$ 

```

**Proposition 3.1.4.** If  $F(x) \neq x$ , then  $f(F(x)) < f(x)$ .

*Proof.* We assume  $x$  is not a vertex. Further, we see  $F(x)$  is the unique minimizer of the  $g(z) := \sum_{i=1}^m \frac{\|z - a_i\|^2}{\|x - a_i\|}$ . Note  $g$  is strictly convex and that  $F$  is the unique minimizer.

Thus,  $g(F(x)) < g(x) = f(x)$ . But we need to show  $f(F(x)) < f(x)$ . Indeed,

$$\begin{aligned} g(F(x)) &= \sum_{i=1}^m \frac{\|F(x) - a_i\|^2}{\|x - a_i\|} = \sum_{i=1}^m \frac{(\|x - a_i\| + \|F(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|} \\ &= 2f(F(x)) + \sum_{i=1}^m \frac{(\|F(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|}. \end{aligned}$$

This implies  $f(F(x)) < f(x)$ .  $\square$

**Proposition 3.1.5.** Let  $R_k := \sum_{i=1, i \neq k}^m \frac{a_i - a_k}{\|a_i - a_k\|}$ , where each  $a_j \in \mathcal{A}$ . Then the vertex  $a_k$  is the optimal solution of the problem (3.1.1) if and only if

$$\|R_k\| \leq 1.$$

*Proof.* By proposition 1.0.22, we have that  $a_k$  is an optimal solution if and only if

$$0 \in \partial f(a_k) = -R_k + \mathbb{B}.$$

Equivalently, we have  $\|R_k\| \leq 1$ .  $\square$

**Proposition 3.1.6.** Suppose  $a_k$  is not the optimal solution and  $x$  is not a vertex. Then there exists  $\delta > 0$  such that  $0 < \|x - a_i\| \leq \delta$ . Further, this implies there exists a positive integer  $n$  such that

$$\|F^{n-1}(x) - a_k\| \leq \delta < \|F^n(x) - a_k\|.$$

*Proof.* It follows from the mapping defining  $F$  that

$$F(x) - a_k = \frac{\sum_{i=1, i \neq k}^m \frac{a_i - a_k}{\|x - a_i\|}}{\sum_{i=1}^m \frac{1}{\|x - a_i\|}}$$

It follows that  $\lim_{x \rightarrow a_k} \frac{F(x) - a_k}{\|x - a_k\|} = R_k$ . Further, by proposition 3.1.5 we obtain

$$\lim_{x \rightarrow a_k} \frac{\|F(x) - a_k\|}{\|x - a_k\|} = \|R_k\| > 1.$$

Therefore, we have  $\epsilon > 0$  and  $\delta > 0$  such that

$$\frac{\|F(x) - a_k\|}{\|x - a_k\|} > 1 + \epsilon \text{ whenever } 0 < \|x - a_k\| < \delta.$$

$\square$

We now have the necessary tools for understanding Kuhn's statement and proof on the convergence of Weiszfeld's algorithm.

**Theorem 3.1.7.** Let  $\{x_k\}_{k=1}^\infty$  be the sequence formed by Weiszfeld's algorithm. Suppose that  $x_k \notin \mathcal{A}$  for  $k \in \mathbb{N}$ . Then  $\{x_k\}_{k=1}^\infty$  converges to the unique solution of the problem (3.1.1).

*Proof.* The proof is omitted and the reader is referred to [10]. □

## 3.2 The Constrained Problem

Now that we are familiar with the Fermat-Torricelli problem, let us introduce the *constrained* Fermat-Torricelli problem. This is a new problem, as such the results in the remainder of this chapter still being developed. The temporary citation for our work is [12] in the bibliography.

Let  $\{\Omega_i\}_{i=1}^p \subset \mathbb{R}^n$  be a collection of convex sets. Our new problem is as follows:

$$\text{minimize } f(x) := \sum_{i=1}^m \|x - a_i\| \text{ subject to } x \in \bigcap_{i=1}^p \Omega_i. \quad (3.2.1)$$

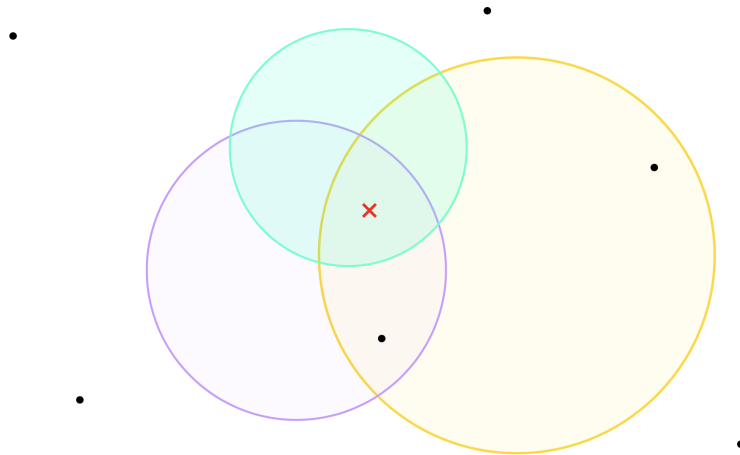


FIGURE 3.2.1: We solve for a point  $x$  contained in the intersection of a finite collection of convex sets such that the sum of distances between  $x$  and each  $a_i \in \mathcal{A}$  is minimal with respect to the intersection.

The distance penalty method allows us to reformulate (3.2.1) into an unconstrained optimization problem (see, e.g., [15, sec.17.1], [7]). Let  $\lambda > 0$  be the penalty constant. Then we have the following:

$$\text{minimize } f_\lambda(x) = \sum_{i=1}^m \|x - a_i\| + \frac{\lambda}{2} \sum_{i=1}^p d(x; \Omega_i)^2 \quad x \in \mathbb{R}^n. \quad (3.2.2)$$

Computationally, this is much more feasible. Recall that  $\nabla[d(x; \Omega)]^2 = 2[x - P(x; \Omega)]$ . We then obtain

$$\nabla f_\lambda(x) = \sum_{i=1}^m \frac{x - a_i}{\|x - a_i\|} + \lambda \sum_{i=1}^p [x - P(x; \Omega_i)]$$

Solving the equation  $\nabla f_\lambda(x) = 0$ , we have

$$x = \frac{\sum_{i=1}^m \frac{a_i}{\|x - a_i\|} + \lambda \sum_{i=1}^p P(x; \Omega_i)}{\sum_{i=1}^m \frac{1}{\|x - a_i\|} + \lambda p}$$

Then for  $x \in \mathcal{A} = \{a_1, \dots, a_m\}$  define  $F_\lambda(x) := x$ .

We introduce the following algorithm: choose a starting point  $x_1 \in \mathbb{R}^n$  and define

$$x_{k+1} = F_\lambda(x_k) \text{ for } k \in \mathbb{N}. \quad (3.2.3)$$

To increase the accuracy of the algorithm, we incrementally increase  $\lambda$  with each iteration.

### New Weiszfeld's Algorithm.

INITIALIZE:  $x_0 \in H$ ,  $\mu_0 > 0$ ,  $\sigma > 1$ .  
 Set  $k = 0$ .  
**Repeat the following**  
 Let  $\lambda := \mu_k$  and let  $y := x_k$   
   Repeat  
      $y := F_\lambda(y)$   
   Until a stopping criterion is satisfied  
 Set  $k := k + 1$   
 Update  $x_k := y$  and  $\mu_k := \sigma \lambda$ .  
**Until a stopping criterion is satisfied.**

Define  $g : (\mathbb{R}^n \setminus \mathcal{A}) \times \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$g_\lambda(x, y) = \sum_{i=1}^m \frac{\|y - a_i\|^2}{\|x - a_i\|} + \lambda \sum_{j=1}^p \|y - P(x; \Omega_j)\|^2 - \frac{\lambda}{2} \sum_{j=1}^p [d(x; \Omega_j)]^2$$

**Lemma 3.2.1.** We have the following properties:

- (i)  $g_\lambda(x, x) = f(x)$ .
- (ii)  $y = \operatorname{argmin}_{y \in \mathbb{R}^n} g_\lambda(x, y)$  if and only if  $y = F_\lambda(x)$ .
- (iii)  $g_\lambda(x, y) \geq 2f_\lambda(y) - f_\lambda(x)$ .

*Proof.* The assertion (i) follows from the identity  $\|x - P(x; \Omega_j)\| = d(x; \Omega_j)$  for  $x \in \mathbb{R}^n$ . To show (ii), note that for a fixed  $x \notin \mathcal{A}$ , the mapping  $y \mapsto g_\lambda(x, y)$  is smooth and strictly convex. Therefore,  $y$  is its unique minimizer on  $\mathbb{R}^n$  if and only if  $\nabla g_\lambda(x, y) = 0$ . To show the converse, note

$$\nabla_y g_\lambda(x, y) = 2 \sum_{i=1}^m \frac{y - a_i}{\|x - a_i\|} + 2\lambda \sum_{j=1}^p (y - P(x; \Omega_j)).$$

Solving  $\nabla_y g_\lambda(x, y) = 0$  yields  $y = F_\lambda(x)$ . It remains to prove (iii). Using the inequality  $\frac{a^2}{b} \geq 2a - b$  for any two positive numbers  $a$  and  $b$ , we have

$$\sum_{i=1}^m \frac{\|y - a_i\|^2}{\|x - a_i\|} \geq 2 \sum_{i=1}^m \|y - a_i\| - \sum_{i=1}^m \|x - a_i\|. \quad (3.2.4)$$

Observe that  $\|y - P(x; \Omega_j)\| \geq \|y - P(y; \Omega_j)\| = d(y; \Omega_j)$ , then

$$\lambda \sum_{j=1}^p \|y - P(x; \Omega_j)\|^2 - \frac{\lambda}{2} \sum_{j=1}^p [d(x; \Omega_j)]^2 \geq \frac{\lambda}{2} \sum_{j=1}^p 2[d(y; \Omega_j)]^2 - \frac{\lambda}{2} \sum_{j=1}^p [d(x; \Omega_j)]^2. \quad (3.2.5)$$

The result from summing (3.2.4) and (3.2.5) implies (iii).  $\square$

**Proposition 3.2.1.** Let  $\{x_k\}$  be the sequence generated by the method given in (3.2.3). Assume that  $\{x_k\} \cap \mathcal{A} = \emptyset$ , then the sequence  $\{f_\lambda(x_k)\}$  is monotone decreasing. Moreover,  $f_\lambda(x_{k+1}) = f_\lambda(x_k)$  if and only if  $x_k$  is a minimizer of  $f_\lambda$  on  $\mathbb{R}^n$ .

*Proof.* First we show that if  $F_\lambda(x) \neq x$  for each  $x \in \mathbb{R}^n \setminus \mathcal{A}$ , then  $f_\lambda(F_\lambda(x)) < f_\lambda(x)$ . Note  $y \mapsto g_\lambda(x, y)$  is a strictly convex; then by Lemma 3.2.1(ii), its unique minimizer on  $\mathbb{R}^n$  is  $F_\lambda(x)$ . Now suppose  $F_\lambda(x) \neq x$ . Then

$$g_\lambda(x, F_\lambda(x)) < g_\lambda(x, x) = f_\lambda(x),$$

where the equality follows from Lemma 3.2.1(i). Now by invoking Lemma 3.2.1(iii), we have

$$g_\lambda(x, F_\lambda(x)) \geq 2f_\lambda(F_\lambda(x)) - f_\lambda(x).$$

Combining the above estimates yields the desired strict monotonicity. Recall that for  $x \notin \mathcal{A}$ ,  $\nabla f_\lambda(x) = 0$  if and only if  $x = F_\lambda(x)$ , the conclusion now follows.  $\square$

**Proposition 3.2.2.** Define,

$$R_k = \sum_{i=1, i \neq k}^m \frac{a_i - a_k}{\|a_k - a_i\|} + \lambda \sum_{j=1}^p [P(a_k; \Omega_j) - a_k].$$

The point  $a_k$  is a minimizer of  $f_\lambda$  if and only if

$$\|R_k\| \leq 1.$$

*Proof.* By the subdifferential Fermat rule in convex analysis,  $a_k$  is a minimizer of  $f_\lambda$  if and only if

$$0 \in \partial f_\lambda(a_k) = -R_k + \mathbb{B}.$$

This is equivalent to  $\|R_k\| \leq 1$ . □

**Lemma 3.2.2.** If  $a_k$  is not a minimizer of  $f_\lambda$  on  $\mathbb{R}^n$ , then there exists  $\delta > 0$  such that for any  $x \in \mathbb{B}(a_k; \delta) \setminus \{a_k\}$ . Further, there exists a positive integer  $s$  satisfying

$$\|F_\lambda^{s-1}(x) - a_k\| \leq \delta \quad \text{and} \quad \|F_\lambda^s(x) - a_k\| > \delta,$$

with convention that  $F_\lambda^0(x) = x$ .

*Proof.* For any  $x \notin \mathcal{A}$ , we have

$$F_\lambda(x) - a_k = \frac{\sum_{i=1, i \neq k}^m \frac{a_i - a_k}{\|x - a_i\|} + \lambda \sum_{j=1}^p [P(x; \Omega_j) - a_k]}{\lambda p + \sum_{i=1}^m \frac{1}{\|x - a_i\|}}.$$

By the continuity of the projection mappings onto convex sets, we have

$$\lim_{x \rightarrow a_k} \frac{F_\lambda(x) - a_k}{\|x - a_k\|} = \lim_{x \rightarrow a_k} \frac{\sum_{i=1, i \neq k}^m \frac{a_i - a_k}{\|x - a_i\|} + \lambda \sum_{j=1}^p [P(x; \Omega_j) - a_k]}{\lambda p \|x - a_k\| + 1 + \sum_{i=1, i \neq k}^m \frac{\|x - a_k\|}{\|x - a_i\|}} = R_k. \quad (3.2.6)$$

By our assumption  $a_k$  is not a minimizer of  $f_\lambda$ ; we invoke Proposition 3.2.2 to obtain the inequality

$$\lim_{x \rightarrow a_k} \frac{\|F_\lambda(x) - a_k\|}{\|x - a_k\|} = \|R_k\| > 1.$$

Now set  $\varepsilon = \frac{\|R_k\| - 1}{2} > 0$ , then there exists  $\delta > 0$  such that

$$\|F_\lambda(x) - a_k\| > (1 + \varepsilon)\|x - a_k\| \quad \text{whenever} \quad 0 < \|x - a_k\| \leq \delta.$$

It follows that for each  $x \in \mathbb{B}(a_k; \delta) \setminus \{a_k\}$ , there must exist a positive integer  $s$  such

that  $F^s(x) \notin \mathbb{B}(a_k; \delta)$ . Suppose the contrary, then by induction we have

$$\|F_\lambda^s(x) - a_k\| > (1 + \varepsilon)\|F_\lambda^{s-1}(x) - a_k\| \geq \dots \geq (1 + \varepsilon)^s\|x - a_k\| \leq \delta \text{ for each } s.$$

This is a contradiction due to  $(1 + \varepsilon)^s \rightarrow \infty$  as  $s \rightarrow \infty$ . This completes the proof.  $\square$

**Proposition 3.2.3.** Let  $\{x_k\}$  be the sequence generated by the method given in (3.2.3). If  $x_k \notin \mathcal{A}$  for all  $k$ , then every cluster point of  $\{x_k\}$  is a minimizer of  $f_\lambda$  on  $\mathbb{R}^n$ . Moreover, if the anchors are not collinear, then  $\{x_k\}$  converges to the unique minimizer of  $f_\lambda$  on  $\mathbb{R}^n$ .

*Proof.* Let  $S$  be the solution set of (3.2.2). In the case where  $x_K = x_{K+1}$  for some  $K$ , we have that  $x_k$  is a constant sequence for  $k \geq K$ . Thus, it converges to  $x_K$ . Since  $F(x_K) = x_K$  and  $x_K$  is not a vertex,  $x_K$  is a minimizer. So we can assume that  $x_{k+1} \neq x_k$  for every  $k$ . By Proposition 3.2.1, the sequence  $\{f_\lambda(x_k)\}$  is nonnegative and decreasing, so it converges. It follows that

$$\lim_{k \rightarrow \infty} (f_\lambda(x_k) - f_\lambda(x_{k+1})) = 0. \quad (3.2.7)$$

Observe that  $f_\lambda$  is coercive, for any initial point  $x_0$ , the sequence  $\{x_k\}$  is bounded because  $\{x_k : k \in \mathbb{N}\}$  is a subset of the bounded set  $\{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ . Let  $\{x_{k_\ell}\}$  be any convergent subsequence of  $\{x_k\}$  whose limit denoted by  $\bar{x}$ . It remains to prove that  $\bar{x} \in S$ . By (3.2.7),

$$\lim_{k \rightarrow \infty} (f_\lambda(x_{k_\ell}) - f_\lambda(F_\lambda(x_{k_\ell}))) = 0.$$

Using the continuity of  $f_\lambda$  and the mapping  $F_\lambda$ , we have  $f_\lambda(\bar{x}) = f_\lambda(F_\lambda(\bar{x}))$ . If  $\bar{x} \notin \mathcal{A}$ , from the proof of Proposition 3.2.1, we can conclude that  $\bar{x} \in S$ . If  $\bar{x} \in \mathcal{A}$ , without loss of generality, assume  $\bar{x} = a_1$ . Suppose by contradiction that  $\bar{x} = a_1 \notin S$ . Choose  $\delta > 0$  sufficiently small such that the property in Lemma 3.2.2 holds and  $\mathbb{B}(a_1; \delta) \cap [S \cup \{a_2, \dots, a_m\}] = \emptyset$ . Since  $x_{k_\ell} \rightarrow a_1$ , we can assume without loss of generality that this subsequence is contained in the ball  $\mathbb{B}(a_1; \delta)$ .

Recall that the original sequence  $\{x_k\}$  is defined by  $x_{k+1} = F_\lambda(x_k)$ . For  $x = x_{k_1}$ , we can choose  $q_1$  such that  $x_{q_1} \in \mathbb{B}(a_1; \delta)$  and  $F(x_{q_1}) \notin \mathbb{B}(a_1; \delta)$ . Choose an index  $k_\ell$  with  $k_\ell > q_1$  and apply Lemma 3.2.2, we find  $q_2 > q_1$  such that  $x_{q_2} \in \mathbb{B}(a_1; \delta)$  and  $F(x_{q_2}) \notin \mathbb{B}(a_1; \delta)$ . Repeating this procedure, we construct another subsequence  $\{x_{q_\ell}\}$  of  $\{x_k\}$  satisfying  $x_{q_\ell} \in \mathbb{B}(a_1; \delta)$  and  $F(x_{q_\ell}) \notin \mathbb{B}(a_1; \delta)$  for all  $\ell$ . Extracting a further subsequence, we can assume that  $x_{q_\ell} \rightarrow \bar{y} \in \mathbb{B}(a_1; \delta)$ . By the argument that has been used above,  $f_\lambda(\bar{y}) = f_\lambda(F_\lambda(\bar{y}))$ . From this, if  $\bar{y} \notin \mathcal{A}$ , then  $\bar{y} \in S$  which is a contradiction because  $\mathbb{B}(a_1; \delta) \cap S = \emptyset$ . Thus  $\bar{y} \in \mathcal{A}$  and thus it must be  $a_1$ . Then by



the boundedness of  $\{x_k\}$ , we have

$$\lim_{\ell \rightarrow \infty} \frac{\|F(x_{q_\ell}) - a_1\|}{\|x_{q_\ell} - a_1\|} = \infty.$$

This contradicts (3.2.6). Thus  $\bar{x} = a_1 \in S$  and the proof is complete.  $\square$

### 3.2.1 Solving with Nesterov's Accelerated Gradient Method

We now use an alternative algorithm for solving the auxiliary problem (3.2.2). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function with Lipschitz gradient. That is, there exists  $\ell \geq 0$  such that  $\|\nabla f(x) - \nabla f(y)\| \leq \ell \|x - y\|$  for all  $x, y \in \mathbb{R}^n$ . Let  $\Omega$  be a nonempty closed convex set. Yu. Nesterov (1983, 2005) considered the optimization problem

$$\min \{f(x) : x \in \Omega\}. \quad (3.2.8)$$

Define  $T_\Omega(x) := \operatorname{argmin} \{ \langle \nabla f(x), y - x \rangle + \frac{\ell}{2} \|x - y\|^2 : y \in \Omega \}$ . Let  $d$  be a continuous and strongly convex function on  $\Omega$  with modulus  $\sigma > 0$ . The function  $d$  is called a *prox-function* of the set  $\Omega$ . Since  $d$  is a strongly convex function on the set  $\Omega$ , it has a unique minimizer on this set. Denote  $x^0 = \operatorname{argmin} \{d(x) : x \in \Omega\}$ . Then Nesterov's accelerated gradient algorithm for solving (3.2.8) is outlined as follows.

```

INPUT:  $f, \ell, x^0 \in \Omega$ 
set  $k = 0$ 
repeat
  find  $y^k := T_\Omega(x^k)$ 
  find  $z^k := \operatorname{argmin} \{ \frac{\ell}{\sigma} d(x) + \sum_{i=0}^k \frac{i+1}{2} [f(x^i) + \langle \nabla f(x^i), x - x^i \rangle] : x \in \Omega \}$ 
  set  $x^k := \frac{2}{k+3} z^k + \frac{k+1}{k+3} y^k$ 
  set  $k := k + 1$ 
until a stopping criterion is satisfied.
OUTPUT:  $y^k$ .

```

In the case where the objective function in (3.2.8) is nonsmooth but has a particular form, Nesterov made use of the special structure of  $f$  to approximate it by a convex function with Lipschitz continuous gradient and then applied his accelerated gradient method to minimize the smooth approximation.

Observe that

$$f_i(x) = \|x - a_i\| = \sup \{ \langle x - a_i, u \rangle : u \in \mathbb{B} \}.$$

Applying Proposition 2.2.4 with  $\bar{u} = 0 \in Q = \mathbb{B}$ ,  $\varphi(u) = \langle a_i, u \rangle$  and  $A$  is the identity mapping, a smooth approximation of  $f_i$  is given by

$$f_{i,\mu}(x) = \frac{\|x - a_i\|^2}{2\mu} - \frac{\mu}{2} \left[ d\left(\frac{x - a_i}{\mu}; \mathbb{B}\right) \right]^2.$$

Therefore, the function  $f_\lambda$  in (3.2.2) has the following smooth approximation

$$f_{\lambda,\mu}(x) = \frac{1}{2\mu} \sum_{i=1}^m \|x - a_i\|^2 - \frac{\mu}{2} \sum_{i=1}^m \left[ d\left(\frac{x - a_i}{\mu}; \mathbb{B}\right) \right]^2 + \frac{\lambda}{2} \sum_{j=1}^p [d(x; \Omega_j)]^2,$$

with its gradient given by

$$\nabla f_{\lambda,\mu}(x) = \sum_{i=1}^m P\left(\frac{x - a_i}{\mu}; \mathbb{B}\right) + \lambda \sum_{j=1}^p [x - P(x; \Omega_j)].$$

This  $\nabla f_{\lambda,\mu}$  is a Lipschitz function whose constant is

$$\ell = \frac{m}{\mu} + \lambda p.$$

Moreover, we have the following estimate

$$f_{\lambda,\mu}(x) \leq f_\lambda(x) \leq f_{\lambda,\mu}(x) + \frac{m\mu}{2}.$$

Applying the Nesterov's accelerated gradient method, we have the following pseudo code for solving the auxiliary problem (3.2.2).

---

**Algorithm 1** AGM( $f_{\lambda,\mu}, x^0$ )

---

INPUT:  $\mathcal{A} = \{a_1, \dots, a_m\}$ , constrained sets  $\Omega_j$  for  $j = 1, \dots, p$ ,  
 $\lambda > 0$ , sufficiently small  $\mu > 0$  and initial point  $x^0 \in \mathbb{R}^n$ .

OUTPUT:  $y^k$  that approximately solve (3.2.2).

Compute  $\ell = \frac{m}{\mu} + \lambda p$ .

Set  $k := 0$ .

**repeat**

    Compute  $\nabla f_{\lambda,\mu}(x^k) = \sum_{i=1}^m P\left(\frac{x^k - a_i}{\mu}; \mathbb{B}\right) + \lambda \sum_{j=1}^p [x^k - P(x^k; \Omega_j)]$ .

    Find  $y^k := x^k - \frac{1}{\ell} \nabla f_{\lambda,\mu}(x^k)$ .

    Find  $z^k := x^0 - \frac{1}{\ell} \sum_{i=0}^k \frac{i+1}{2} \nabla f_{\lambda,\mu}(x^i)$ .

    Set  $x^{k+1} := \frac{2}{k+3} z^k + \frac{k+1}{k+3} y^k$ .

    Set  $k := k + 1$ .

**until a stopping criterion is satisfied.**

**return:**  $y^k$ .

---

To solve the original problem (3.2.1), we often gradually decrease the smooth parameter  $\mu$  and increase the penalty parameter  $\lambda$ . The general scheme is outlined as follows.

---

**Algorithm 2** AGMDPM( $f$ )

---

INPUT:  $\mathcal{A} = \{a_1, \dots, a_m\}$ , constrained sets  $\Omega_j$  for  $j = 1, \dots, p$ ,  
parameters  $\mu_0 > 0$ ,  $\lambda_0 > 0$ ,  $\sigma > 1$ ,  $\gamma \in (0, 1)$  and initial point  $x^0$ .

OUTPUT:  $x^k$  that approximately solve (3.2.1).

Set  $k = 0$ .

**repeat**

    Set  $\lambda := \lambda_k$ ,  $\mu := \mu_k$ ,  $\ell = \frac{m}{\mu} + \lambda p$ ,  $y := x^k$ ,  $x^0 := y$

    Compute  $y := \text{AGM}(f_{\lambda,\mu}, x^0)$ .

    Set  $k := k + 1$

    Set  $x^k := y$ ,  $\mu_k := \gamma \mu$ ,  $\lambda_k := \sigma \lambda$ .

**until a stopping criterion is satisfied.**

**return:**  $x^k$ .

---

Let  $f(x) := \sum_{i=1}^m \|x - a_i\|$ ,  $P(x) := \frac{1}{2} \sum_{j=1}^p [d(x; \Omega_j)]^2$ .

$$f_\mu(x) := \frac{1}{2\mu} \sum_{i=1}^m \|x - a_i\|^2 - \frac{\mu}{2} \sum_{i=1}^m \left[ d\left(\frac{x - a_i}{\mu}; \mathbb{B}\right) \right]^2.$$

---

**Algorithm 2** SDPM( $f$ )
 

---

**Input:** Parameters sequences  $\{\lambda_k\}$ ,  $\{\mu_k\}$ .

Set  $k:=0$

**repeat**

Let  $Q_k(x) = f_{\mu_k}(x) + \lambda_k P(x)$ ,  $x_k^* = \operatorname{argmin}_{x \in \mathbb{R}^n} Q_k(x)$  and  $L_k = \frac{m}{\mu_k} + p\lambda_k$ .

Using Nesterov's accelerated gradient method to compute  $x_k$  such that  $Q_k(x_k) \leq Q_k(x_k^*) + \frac{\tau_k^2}{2L_k}$  with the initial point  $x_{k-1}$ .

Set  $k := k + 1$

**until a stopping criterion is satisfied.**

**Output:**  $x_k$ .

---

### 3.2.2 Solving the Fermat-Torricelli Problem via the DCA

Let  $F$  be a nonempty closed convex set that contains the origin in its interior. Consider the Minkowski function associated with  $F$  defined by

$$\rho_F(x) := \inf\{t > 0 : x \in tF\}.$$

Given a finite number of points  $a_i$  with the associated weights  $c_i$  for  $i = 1, \dots, m$ , consider the function

$$\varphi(x) := \sum_{i=1}^m c_i \rho_F(x - a_i). \quad (3.2.9)$$

Then the mathematical model of the weighted constrained Fermat-Torricelli problem is

$$\text{minimize } \varphi(x) \text{ subject to } x \in \bigcap_{i=1}^p \Omega_i, \quad (3.2.10)$$

where  $\Omega_i$  for  $i = 1, \dots, p$  are nonempty closed subsets of  $\mathbb{R}^n$ .

We use this general version of the constrained Fermat-Torricelli problem for the remainder of this thesis. Our methods are based on the DCA and the Distance Penalty Method.

The objective function (3.2.9) can be rewritten as follows

$$\varphi(x) = \sum_{i \in I} \alpha_i \rho_F(x - a_i) - \sum_{j \in J} \beta_j \rho_F(x - a_j),$$

where  $I := \{i = 1, \dots, m : c_i > 0\}$  and  $J := \{i = 1, \dots, m : c_i < 0\}$ ,  $\alpha_i = c_i$  for  $i \in I$  and  $\beta_j = -c_j$  for  $j \in J$ .

We reformulate the problem (3.2.10) using the Distance Penalty Method as follows

$$\text{minimize } \psi(x) := \sum_{i \in I} \alpha_i \rho_F(x - a_i) - \sum_{j \in J} \beta_j \rho_F(x - a_j) + \sum_{i=1}^p \frac{\lambda}{2} [d(x; \Omega_i)]^2. \quad (3.2.11)$$

Given a nonempty bounded set  $K$ , the *support function* associated with  $K$  is given by

$$\sigma_K(x) := \sup\{\langle x, y \rangle : y \in K\}.$$

It follows from the definition of the Minkowski function (see, e.g., [4, Proposition 2.1]) that  $\rho_F(x) = \sigma_{F^\circ}(x)$ , where

$$F^\circ := \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 1 \text{ for all } x \in F\}.$$

In the proposition below, we consider a version of Nesterov's smooth approximation for the Minkowski function in  $\mathbb{R}^n$ .

**Proposition 3.2.4. (DC decomposition of the Minkowski gauge).** Given any  $a \in \mathbb{R}^n$  and  $\mu > 0$ , a Nesterov smoothing approximation of  $\varphi(x) := \rho_F(x - a)$  has the representation

$$\varphi_\mu(x) = \frac{1}{2\mu} \|x - a\|^2 - \frac{\mu}{2} \left[ d\left(\frac{x - a}{\mu}; F^\circ\right) \right]^2.$$

Moreover,  $\nabla \varphi_\mu(x) = P\left(\frac{x - a}{\mu}; F^\circ\right)$  and

$$\varphi_\mu(x) \leq \varphi(x) \leq \varphi_\mu(x) + \frac{\mu}{2} \|F^\circ\|^2, \quad (3.2.12)$$

where  $\|F^\circ\| := \sup\{\|u\| : u \in F\}$ .

*Proof.* The function  $\varphi$  can be represented as

$$\varphi(x) = \sigma_{F^\circ}(x - a) = \sup\{\langle x - a, u \rangle \mid u \in F^\circ\}.$$

Using the prox-function  $d(x) = \frac{1}{2}\|x\|^2$  in [14], one obtains a smooth approximation

of  $\varphi$  given by

$$\begin{aligned}
\varphi_\mu(x) &:= \sup\{\langle x - a, u \rangle - \frac{\mu}{2}\|u\|^2 \mid u \in F^\circ\} \\
&= \sup\{-\frac{\mu}{2}(\|u\|^2 - \frac{2}{\mu}\langle x - a, u \rangle) \mid u \in F^\circ\} \\
&= \sup\{-\frac{\mu}{2}\|u - \frac{1}{\mu}(x - a)\|^2 + \frac{1}{2\mu}\|x - a\|^2 \mid u \in F^\circ\} \\
&= \frac{1}{2\mu}\|x - a\|^2 - \frac{\mu}{2} \inf\{\|u - \frac{1}{\mu}(x - a)\|^2 \mid u \in F^\circ\} \\
&= \frac{1}{2\mu}\|x - a\|^2 - \frac{\mu}{2} [d(\frac{x - a}{\mu}; F^\circ)]^2.
\end{aligned}$$

The formula for computing the gradient of  $\varphi_\mu$  follows from the gradient formulas for the squared Euclidean norm and the squared distance function generated by a nonempty closed convex set:  $\nabla d^2(x; \Omega) = 2[x - P(x; \Omega)]$ . Estimate (3.2.12) can be proved directly; see also [14]. This completes the proof.  $\square$

**Proposition 3.2.5.** Consider the function  $f$  define by

$$f(x) := \sum_{i=1}^m \rho_F(x - a_i).$$

Then a smooth approximation of the function  $f$  is given by

$$f_\mu(x) := \sum_{i=1}^m \frac{1}{2\mu} \|x - a_i\|^2 - \sum_{i=1}^m \frac{\mu}{2} [d(\frac{x - a_i}{\mu}; F^\circ)]^2.$$

Moreover,

$$f_\mu(x) \leq f(x) \leq f_\mu(x) + \frac{m\mu}{2} \|F^\circ\|^2.$$

**Proposition 3.2.6. (DC decomposition of the square distance function).**

Let  $\Omega$  be a nonempty closed convex set in  $\mathbb{R}^n$ . Then the square distance function associated with  $\Omega$  has the following DC decomposition

$$[d(x; \Omega)]^2 = \|x\|^2 - \varphi_\Omega(x),$$

where  $\varphi_\Omega(x) := 2 \sup\{\langle x, w \rangle - \frac{1}{2}\|w\|^2 : w \in \Omega\}$  is a differentiable function with  $\nabla \varphi_\Omega(x) = 2P(x; \Omega)$ .

*Proof.* We have

$$\begin{aligned}
[d(x; \Omega)]^2 &= \inf\{\|x - w\|^2 : w \in \Omega\} \\
&= \inf\{\|x\|^2 - 2\langle x, w \rangle + \|w\|^2 : w \in \Omega\} \\
&= \|x\|^2 + \inf\{\|w\|^2 - 2\langle x, w \rangle : w \in \Omega\} \\
&= \|x\|^2 - \sup\{\langle 2x, w \rangle - \|w\|^2 : w \in \Omega\}
\end{aligned}$$

It follows from the representation of  $[d(x; \Omega)]^2$  above that

$$\varphi_\Omega(x) = \|x\|^2 - [d(x; \Omega)]^2.$$

Note that the function  $\psi(x) := [d(x; \Omega)]^2$  is differentiable with  $\nabla\psi(x) = 2[x - P(x; \Omega)]$ . Then function  $\varphi_\Omega$  is differentiable with

$$\nabla\varphi_\Omega(x) = 2x - 2[x - P(x; \Omega)] = 2P(x; \Omega),$$

which completes the proof.  $\square$

We are now ready to obtain a DC decomposition based on Nesterov's smoothing technique for the problem.

**Proposition 3.2.7.** The objective function in (3.2.11) admits the following DC smooth approximation

$$f_\mu(x) = g_\mu(x) - h_\mu(x),$$

where

$$g_\mu(x) := \frac{p\lambda}{2}\|x\|^2 + \sum_{i \in I} \frac{\alpha_i}{2\mu}\|x - a_i\|^2$$

and

$$h_\mu(x) := \frac{\lambda}{2} \sum_{i=1}^p \varphi_{\Omega_i}(x) + \sum_{j \in J} c_j \frac{\mu}{2} [d(\frac{x - a_j}{\mu}; F^\circ)]^2 + \sum_{j \in J} \beta_j \rho_F(x - a_j).$$

In this formulation, it is ready to apply the DCA.

---

**Algorithm 3** DCA

---

INPUT: points  $a_i \in \mathcal{A}$  for  $i \in I$  and  $a_j \in \mathcal{A}$  for  $j \in J$ , constrained sets  $\Omega_j$  for  $j = 1, \dots, p$ ,  
 $\lambda > 0$ , sufficiently small  $\mu > 0$  and initial point  $x^0 \in \mathbb{R}^n$ .

OUTPUT:  $x^k$  that approximately solve (3.2.2).

Set  $k:=0$ .

**repeat**

    Compute  $y^k \in \partial h_\mu(x^k)$

    Find  $x^{k+1} := \nabla g_\mu^*(y^k) = \frac{y^k + \sum_{i \in I} \frac{\alpha_i a_i}{\mu}}{p\lambda \sum_{i \in I} \frac{\alpha_i}{\mu}}$

    Set  $k := k + 1$ .

**until a stopping criterion is satisfied.**

**return:**  $x^k$ .

---

To implement this algorithm, it is required to find  $y^k \in \partial h_\mu(x^k)$ . By the subdifferential sum rule,

$$\partial h_\mu(x) = \lambda \sum_{i=1}^p P(x; \Omega_i) + \sum_{j \in J} c_j \left[ \frac{x - a_j}{\mu} - P\left(\frac{x - a_j}{\mu}; F^\circ\right) \right] + \sum_{j \in J} \beta_j \partial \rho_F(x - a_j).$$

The subdifferential formula of the Minkowski function is given by

$$\partial \rho_F(x) = \begin{cases} \{v \in \mathbb{R}^n : \rho_{F^\circ}(v) = 1, \langle v, x \rangle = \rho_F(x)\} & x \neq 0, \\ F^\circ & x = 0. \end{cases}$$

For example, in the case where  $F = \mathbb{B}$ , the closed unit ball of  $\mathbb{R}^n$ ,

$$\partial \rho_F(x) = \begin{cases} \left\{ \frac{x}{\|x\|} \right\} & x \neq 0, \\ F & x = 0. \end{cases}$$



# Bibliography

- [1] An, N. T. and Nam, N. M. (2015). Convergence analysis of a proximal point algorithm for minimizing differences of functions.
- [2] Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.
- [3] Hartman, P. (1959). On functions representable as a difference of convex functions. *Pacific J. Math*, 9(3):707–713.
- [4] He, Y. and Ng, K. F. (2006). Subdifferentials of a minimum time function in banach spaces. *Journal of mathematical analysis and applications*, 321(2):896–910.
- [5] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2012). *Fundamentals of convex analysis*. Springer Science & Business Media.
- [6] Horst, R. and Thoai, N. V. (1999). Dc programming: overview. *Journal of Optimization Theory and Applications*, 103(1):1–43.
- [7] Lange, K. and Keys, K. L. (2015). The proximal distance algorithm. *arXiv preprint arXiv:1507.07598*.
- [8] Le Thi, H. A., Belghiti, M. T., and Tao, P. D. (2007). A new efficient algorithm based on dc programming and dca for clustering. *Journal of Global Optimization*, 37(4):593–608.
- [9] McDonald, J. N. and Weiss, N. A. (2012). *A course in real analysis*. Academic Press.
- [10] Mordukhovich, B. S. and Nam, N. M. (2013). An easy path to convex analysis and applications. *Synthesis Lectures on Mathematics and Statistics*, 6(2):1–218.
- [11] Nam, N. M. (2013). The fermat-torricelli problem in the light of convex analysis. *arXiv preprint arXiv:1302.5244*.

- [12] Nam, N. M., An, N. T., and Lawrence, N. (2016). Effective optimization techniques for constrained fermat-torricelli problem. *In progress*.
- [13] Nam, N. M., Rector, R. B., and Giles, D. (2015). Minimizing differences of convex functions and applications to facility location and clustering. *arXiv preprint arXiv:1511.07595*.
- [14] Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1):127–152.
- [15] Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.
- [16] Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.
- [17] Tao, P. D. and An, L. T. H. (1997). Convex analysis approach to dc programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355.
- [18] Tao, P. D. and An, L. T. H. (1998). A d.c. optimization algorithm for solving the trust-region subproblem. *SIAM J. Optim.*, 8:476–505.