

Abstract Convexity: Duality, Primal-Dual and Proximal based Algorithms

by

The Hung Tran

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

Supervised by:

- Prof.Dr. Ewa Bednarczuk
System Research Institute - Polish Academy of Sciences
Poland
- Prof.Dr. Dirk Lorenz
University of Bremen
Germany

System Research Institute - Polish Academy of Sciences
Poland



Abstract

In this thesis, we investigate a class of nonconvex functions with special structure, which we call abstract convex functions. They are defined on the basis of generalized Fenchel conjugation and bi-conjugation with arbitrary coupling functions (or elementary functions). When the coupling functions are affine, we recover the original convex functions.

Abstract convex functions are accompanied by abstract conjugate and abstract subdifferentials which are global. Consequently, duality theory for minimization of abstract convex functions is not the classical one, in general, it is not related directly to the dual space. This fact is at the core of our studies of conjugate and Lagrange dualities for minimization problems involving abstract convex functions, which is the first task of the thesis.

While the theory of abstract convexity has been developing since the ninetieth, there is a lack of algorithms and real-life applications of abstract convex functions. One reason is the non-linearity of elementary functions which makes it difficult to design practically workable algorithms and investigate their convergence. The second task of this thesis is to generalize and adapt some of the algorithmic ideas functioning efficiently in the context of convex optimization to a specific class of abstract convex functions which we call $\Phi_{lsc}^{\mathbb{R}}$ -convex.

Finally, we connect the first and the second part of our results to propose $\Phi_{lsc}^{\mathbb{R}}$ -primal-dual algorithm which exploits the idea of Chambolle-Pock algorithm.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor Prof. Ewa Bednarczuk and co-supervisor Prof. Dirk Lorenz who guided me through my PhD journey. I have learned a lot from our discussions together and am impressed by your persistence when solving problems and rich experience from the field. Your meticulous attention to the details and visualizing perspective when examining the problems have inspired me to think of new solutions as well as new problems to walk on. I hope I can utilise these traits for my future career.

I also express my appreciation to the committee members for their time and efforts evaluating my work and giving remarks and questions.

Throughout my PhD, I had the opportunities to collaborate and be a part of several research groups. I had a great time working and solving problems together and extending my view on the field of Optimization.

A heartfelt thanks to the group from Braunschweig, Max, Felix, Niklas, Felix, Nanda, Lionel, Emanuele for the wonderful time I had during my secondment. Thank you Dirk for organizing a lot of social events for the team, they made my days during the cloudy weather in Braunschweig.

I would like to extend my appreciation to the Polish team, Krzysztof, Krzysztof, Monika, Rafał, Agnieszka for the cosy atmosphere at the institute. My heartfelt thanks to Agnieszka, Rafał, Bartek and Olesya for supporting me through various administration matters and emergency situations.

Thanks to the TraDE-OPT program, I get to connect with my fellow

colleagues from the network through workshops and conferences. We had a lot of great time together over a tower glass of beer. Especially to Gabriele and Giovanni, I remembered our heated discussion during work as well as wonderful nighttime in Warsaw.

Last but not least, I am grateful for the support of my family and my best friends Khoa Vo, Mai Nhat Truong, Tai Tran, Khuong Vu for accompanied me through this journey in both good and challenging times.

Funding: I would like to acknowledge the financial support from ITN-ETN, project TraDE-OPT, funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 861137.

Contents

1	Introduction	1
2	Structure of Thesis	6
I	Duality for Composite Abstract Convex Optimization Problems	8
3	Abstract Convexity	9
3.1	Abstract Convex Functions	9
3.2	$\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions	12
3.3	Weakly Convex Functions	25
4	Duality within Abstract Convexity Framework	29
4.1	Conjugate Dual	30
4.1.1	Examples of Conjugate dual for specific classes Φ, Ψ . .	34
4.2	Zero Duality Gap	37
4.2.1	Zero duality gap for specific classes Φ, Ψ	46
4.2.2	Moreau Envelope	53
4.3	Strong Duality	56
4.4	Lagrange Duality	67
4.4.1	Construction of Lagrangian Primal-Dual Problems . .	67
4.4.2	Lagrange Zero Duality Gap	69

II	Algorithms for Composite Optimization Problems involving $\Phi_{lsc}^{\mathbb{R}}$-Convex Functions	85
5	$\Phi_{lsc}^{\mathbb{R}}$-Proximal Algorithms	86
5.1	$\Phi_{lsc}^{\mathbb{R}}$ -Proximal Operator	87
5.2	$\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point	95
5.2.1	Auxiliary Convergence Results	95
5.2.2	$\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point Algorithm	98
5.3	$\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward	105
5.4	$\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient	113
5.5	Numerical Examples	124
6	Primal-Dual algorithms for $\Phi_{lsc}^{\mathbb{R}}$-Convex Functions	134
6.1	Primal-Dual Weakly Convex Functions	136
6.1.1	Sharpness	139
6.2	Convergence Analysis	148
6.2.1	Primal-Dual algorithm with Dual Update	149
6.2.2	Primal-Dual algorithm with primal update first	158
6.3	Primal-Dual algorithm using $\Phi_{lsc}^{\mathbb{R}}$ -prox	162
6.4	Numerical Examples	170
7	Conclusion and Future Work	179

List of Figures

3.1	$\Phi_{lsc}^{\mathbb{R}}$ -Subgradients of f at different points (dotted lines). Observe that the purple curve is the $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f at every point between -1 and 1	15
5.1	From left to right: distance between the current iteration and the minimizer; distance to the optimal value; distance between two successive iterates.	125
5.2	Example 16, case 1, constant stepsize. From left to right: distance between the current iteration and the solution; distance to the optimal value; distance between two successive iterates.	127
5.3	Example 16, case 1, adaptive stepsize. From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.	128
5.4	Example 16, case 2, constant stepsize, initialization x_{01} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.	130
5.5	Example 16, case 2, constant stepsize, initialization x_{02} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.	130

5.6	Example 16, case 2, adaptive stepsize, initialization x_{02} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.	131
5.7	$\Phi_{lsc}^{\mathbb{R}}$ -(PSG) for Example 17. From left to right: distance between the current iterate and the solution; function value at each iteration.	133
6.1	Lagrange function from Example 19	144
6.2	From left to right: contour plots of $\mathcal{H}(x, y) - \mu\sqrt{x^2 + y^2}$ at $\mu = 1, 0.5, 0.1$	145
6.3	From left to right: contour plots of $\mathcal{H}(x, y) - \mu\text{dist}((x, y), (0, 0))$ at $\mu = 0.5, 0.9, 1$	146
6.4	Lagrange function from Example 21.	147
6.5	From left to right: contour plots of $\mathcal{H}(x, y) - \mu\text{dist}((x, y), (0, 0))$ at $\mu = 0.5, 0.9, 1$	147
6.6	Example 22. From left to right: distance to the solution; primal-dual iterates.	171
6.7	Example 23 first case. From left to right: distance to the solution; primal-dual iterates.	172
6.8	Example 23 second case. From left to right: distance to the solution; primal-dual iterates.	172
6.9	Objective function in Example 24.	173
6.10	From left to right: distance from the iterate to the minimizer; distance to the optimal value; primal and dual iterates.	175
6.11	Comparisons between different solutions. From left to right: original image, least squares solution, thresholded least squares solution, solution obtained with the dual method proposed in [61], solution obtained with our model.	178

Chapter 1

Introduction

Over the last sixty years, convexity has become an essential part of the development of optimization and variational analysis. With the surge of computational power, interests in machine learning and data science have risen, which makes convexity the backbone of many theories and applications. There existed many attempts to go beyond convexity by investigating functions like strongly convex [90], quasi-convex, pseudo-convex, strongly para-convex and para-convex [78], delta-convex [55], approximately convex [71] etc.

Abstract convexity encompasses many of the above-mentioned classes of functions. Abstract convexity, presented in the monographs of Rubinov [82], Pallaschke and Rolewicz [72] is based on the idea of bringing convexity outside the range of linearity, by introducing the nonlinear environment where a function becomes an upper envelope of a subset of functions with specific rules. The term “abstract convexity” was used by Rubinov [82] to describe functions which are upper envelopes of a given class of function Φ , i.e.

$$f(x) = \sup_{\phi \in \Phi} \{ \phi(x) : X \rightarrow \mathbb{R}, \phi \leq f \},$$

where X is a nonempty set. On the other hand, Rolewicz suggested the terminology "convexity without linearity" in [72].

Consequently, the concept of abstract convexity works well with classical convex functions. As the supremum operation is retained, many global properties of convex analysis are still in effect for abstract convexity. Till now, the books of Pallaschke and Rolewicz [72], and the monograph of Singer [85] gathered many results about abstract convexity, especially about conjugation, subdifferential, and duality. They also have historical results of the main ideas of abstract convexity and its applications. In [82], Rubinov presented basic notions of abstract convexity and applications to global optimization problems.

It has been fifty years since the theory of abstract convexity started. It has made great progress in many disciplines with applications in both theory and computation. In [83], the authors investigated the class of increasing star-shaped functions and constructed an algorithm to solve the global optimization problem using abstract convexity. They also generalized the cutting plane algorithm for nonconvex global minimization problems [83, 2]. While in [22, 43, 44], a general version of monotone operator is developed based on abstract convexity.

In this work, we aim at using abstract convexity in order to solve the following composite minimization problem,

$$\inf_{x \in X} f(x) + g(Lx), \quad (\text{CP})$$

where X and Y are the domain sets (spaces) of the functions f and g , respectively, the operator $L : X \rightarrow Y$. The composite optimization problem is taken as our main interest as it appears as optimization model in many data analysis problems and image processing e.g. [18]. This problem has been studied: convex setting in Hilbert space [6], from the point of view of variational analysis in finite dimensional setting [77]. Composite problems can also be considered as regularized problems by treating $f(x)$ as a data fidelity term and $g(Lx)$ as a penalty term. Such problems are known as total variation model in image deblurring and denoising [8].

Duality By exploiting properties of the underlying concepts of abstract convexity, namely, abstract conjugate and abstract subdifferentials, we investigate optimality conditions of (CP) via suitably formulated dual problem. By constructing a conjugated dual problem of (CP), we investigate optimal conditions for both the primal and dual problems which leads us to obtain zero duality gap and even strong duality.

Several efforts have been made for the last twenty years to investigate dualities for optimization problems within the framework of abstract convexity. In [56], strong duality is proved for infimal convolution of Fenchel's duality. In [21], zero duality gap and exact multiplier of augmented Lagrangian are investigated by using the framework of abstract convexity. While in [86, 87] necessary and sufficient conditions are given to achieve minimax equality from a general Lagrangian using the definition of abstract convexity. In the works of Dolgopoliik [40], Gorokhovik and Tykoun [50, 49], the authors applied abstract convexity to approximate the class of nonsmooth functions and formulate necessary optimal conditions for global nonsmooth nonconvex problem. The authors in [19] investigated the problem of minimizing the finite sum of arbitrary functions and provided conditions for zero duality gap through infimal convolution dual. On the other hand, in [12], the problem of zero duality gap is studied with the help of perturbation function.

In the first part of the thesis, entitled **Duality for Composite Abstract Convex Optimization Problems**, we define conjugate and Lagrange dual for problem (CP) and provide necessary and sufficient conditions for zero duality gap. We also investigate conditions for strong duality for (CP). In our approach, we generalize the results from [12] by combining our *optimality conditions* and *intersection property* to derive Lagrange zero duality gap. Our motivation for studying this problem comes from the fact that nonconvex regularizers, such as ℓ_0, ℓ_p ($0 < p < 1$), smoothly clipped absolute deviation (SCAD) and minimax concave penalty (MCP), have been gaining a lot of attention and significant improvement over convex regularizers recently [57,

92, 96, 58].

Algorithms While there are plenty of works on abstract convexity, not many works are devoted to algorithms for optimization problems involving abstract convex functions. One reason may come from the fact that elementary functions are essentially nonlinear and do not belong to the dual space like gradient or subgradient in convex analysis. Therefore, it is difficult to design a gradient descent algorithm using abstract subdifferentials. On the other hand, convexity plays an important role in the convergence of many algorithms. For example, a lower semicontinuous convex function is also weakly lower semicontinuous, which implies the equivalence between weak convergence and strong convergence of the sequence through the function. Even though gradient descent algorithm is popular, there are many restrictions to ensure convergence (smoothness, stepsize, etc.). Therefore, in convex analysis, proximal-based algorithms work well for nonsmooth functions and even nonconvex functions with special structures. In particular, many primal-dual algorithms are proximal-based.

In the second part of the thesis, entitled **Algorithms for Composite Optimization Problems involving $\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions**, we aim at introducing an abstract convex proximal operator with abstract subdifferentials, formulating and investigating the convergence of: Proximal point algorithm, Forward-Backward algorithm, Projected Subgradient and Primal-Dual algorithm by using this abstract proximal operator.

We focus on the class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions. The elements ϕ of the class $\Phi_{lsc}^{\mathbb{R}}$ are quadratic functions defined on Hilbert space X i.e.

$$\phi(x) := -a\|x\|^2 + \langle u, x \rangle + c,$$

where $u \in X, a, c \in \mathbb{R}$. The main advantage of choosing this class is that the iterative schemes we propose are practically realizable, i.e. we can calculate efficiently $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials and $\Phi_{lsc}^{\mathbb{R}}$ -conjugate. The framework of $\Phi_{lsc}^{\mathbb{R}}$ -

convexity includes the class of so-called weakly convex functions (semiconvex, see e.g. [23, Chapter 2]) which appear naturally in optimization models related to applications such as source localization, [7], phase retrieval, [65], and discrete tomography [84, 61] or distributed network optimization [31].

Once we have the $\Phi_{lsc}^{\mathbb{R}}$ -proximal mapping defined and investigated, we are able to propose $\Phi_{lsc}^{\mathbb{R}}$ -proximal point algorithm and extend it to splitting algorithms like forward-backward or projected subgradient algorithm and primal-dual algorithm for (CP) problem with $\Phi_{lsc}^{\mathbb{R}}$ -convex functions.

Chapter 2

Structure of Thesis

The thesis is divided into two parts: **Duality for Composite Abstract Convex Optimization Problems** (Chapter 3 and Chapter 4) and **Algorithms for Composite Optimization Problems involving $\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions** (Chapter 5 and Chapter 6).

In Part I, we start by recalling the concept of abstract convex function and its properties. We will focus more on the specific class $\Phi_{lsc}^{\mathbb{R}}$ -convex and weakly convex functions. These are presented in Chapter 3.

Chapter 4 contains the main results of Part I: in Section 4.1, we introduce the conjugate dual for the composite minimization problem (CP) within the framework of abstract convexity. We provide some examples. In Section 4.2, we prove Theorem 3 and 4 containing sufficient and necessary conditions for zero duality gap for problem (CP) and its conjugate dual. In Section 4.3, we investigate strong duality for (CP) and its conjugate dual in Theorem 6. Lastly, we discuss the relationship between conjugate dual and Lagrange dual and discuss about Lagrange zero duality gap in Theorem 9.

In Part II, in Chapter 5, for the class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions, we construct $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator (Section 5.1), formulate and investigate the convergence of $\Phi_{lsc}^{\mathbb{R}}$ -proximal point algorithm in Section 5.2 (Theorem 13). In Section 5.3, we propose $\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward algorithm and analyse their

convergence. In Section 5.4, we discuss about $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient .

In the last chapter of Part II, Chapter 6, we introduce primal-dual algorithm for weakly convex functions. We also propose and investigate the convergence of $\Phi_{lsc}^{\mathbb{R}}$ -primal-dual algorithm based on Chambolle-Pock algorithm.

Finally, we summarize our work and give some possible directions for future research in Chapter 7.

(i) The basis of Chapter 3 and 4 comes from the following paper:

- Ewa Bednarczuk and The Hung Tran. “Duality for composite optimization problem within the framework of abstract convexity”. In: *Optimization* 72.1 (2023), pp. 37–80

(ii) Chapter 5 is the main work of the preprint:

- Dirk Lorenz, Ewa Bednarczuk, and The Hung Tran. “Proximal Algorithms for a class of abstract convex functions”. In: *arXiv preprint arXiv:2402.17072* (2024)

(iii) Chapter 6: the basic ideas come from the following works:

- Ewa Bednarczuk et al. “Calculus rules for proximal ε -subdifferentials and inexact proximity operators for weakly convex functions”. In: *2023 European Control Conference (ECC)*. IEEE. 2023, pp. 1–8
- Ewa Bednarczuk et al. “Convergence analysis of an inexact Forward-Backward algorithm for problems involving weakly convex functions”. In: *arXiv preprint arXiv:2303.14021* (2023)

Part I

Duality for Composite Abstract Convex Optimization Problems

Chapter 3

Abstract Convexity

In this chapter, we present the concept of abstract convex function and its properties, together with abstract conjugate and abstract subdifferentials. This will serve as a foundation for subsequent chapters below. In Section 3.2, we pay more attention to the class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions with some examples of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials. Finally, in Section 3.3, we state some properties of weakly convex function and connection between abstract subdifferentials and proximal subdifferentials of this type of functions.

3.1 Abstract Convex Functions

Let X be a nonempty set. A function $f : X \rightarrow (-\infty, +\infty]$ is proper if its domain, denoted by $\text{dom } f = \{x \in X : f(x) < +\infty\}$ is nonempty.

Let $\Phi := \{\phi : X \rightarrow \mathbb{R}\}$ be a collection of real-valued functions, closed under addition of a constant. The support set of f with respect to Φ is defined as

$$\text{supp}_{\Phi} f := \{\phi \in \Phi : f(x) \geq \phi(x) (\forall x \in X)\}.$$

For $f, g : X \rightarrow (-\infty, +\infty]$, we write $f \leq g \Leftrightarrow f(x) \leq g(x)$ for all $x \in X$. Elements of Φ are called elementary functions.

Definition 1. [82, 72] A function $f : X \rightarrow (-\infty, +\infty]$ is Φ -convex if

$$f(x) = \sup_{\phi \in \text{supp}_\Phi f} \phi(x), \quad \forall x \in X. \quad (3.1)$$

A function f is Φ -convex at $x_0 \in X$ if $f(x_0) = \sup_{\phi \in \text{supp}_\Phi f} \phi(x_0)$.

Note that when a function $f : X \rightarrow (-\infty, +\infty]$ is Φ -convex, then $\text{supp } f$ is nonempty, since by convention, $f \equiv -\infty$.

Definition 2. [82, 72] A Φ -conjugate $f^* : \Phi \rightarrow (-\infty, +\infty]$ of f is defined as

$$f_\Phi^*(\phi) := \sup_{x \in X} \{\phi(x) - f(x)\}. \quad (3.2)$$

Analogously, we define Φ -biconjugate $f^{**} : X \rightarrow (-\infty, +\infty]$ of function f as

$$f_\Phi^{**}(x) := \sup_{\phi \in \Phi} \{\phi(x) - f_\Phi^*(\phi)\}. \quad (3.3)$$

As in convex analysis, the following result holds.

Theorem 1. [82, 72] A function $f : X \rightarrow (-\infty, +\infty]$ is Φ -convex if and only if

$$f(x) = f_\Phi^{**}(x), \quad \forall x \in X. \quad (3.4)$$

In view of this theorem, f is Φ -convex at a point $x_0 \in X$ if $f_\Phi^{**}(x_0) = f(x_0)$.

Example 1. Let X be a Banach space with the topological dual X^* and

$$\Phi_{\text{conv}} := \{\phi : X \rightarrow \mathbb{R} \mid \phi(x) = \langle u, x \rangle + c, u \in X^*, c \in \mathbb{R}\}. \quad (3.5)$$

If a function $f : X \rightarrow (-\infty, +\infty]$ is Φ_{conv} -convex, then it is proper, convex and lower semi-continuous. When X is a locally convex space, [45], then f is Φ_{conv} -convex if and only if f is proper, convex and lower semi-continuous.

Definition 3. [82, 72] Let X be a nonempty set and $f : X \rightarrow (-\infty, +\infty]$. A Φ -subgradient of f at a point $x \in \text{dom } f$ is an elementary function $\phi \in \Phi$ satisfying

$$f(y) - f(x) \geq \phi(y) - \phi(x), \quad \forall y \in X. \quad (3.6)$$

The collection of all Φ -subgradients of f at x is called the Φ -subdifferentials of f at x and is denoted by $\partial_\Phi f(x)$,

$$\partial_\Phi f(x) := \{\phi \in \Phi \mid (\forall y \in X) \ f(y) - f(x) \geq \phi(y) - \phi(x)\}. \quad (3.7)$$

Furthermore, for $\varepsilon \geq 0$, the ε - Φ -subdifferentials of f at $x \in \text{dom } f$, $\partial_{\varepsilon, \Phi} f(x)$, is defined as

$$\partial_{\varepsilon, \Phi} f(x) := \{\phi \in \Phi \mid f(y) - f(x) \geq \phi(y) - \phi(x) - \varepsilon, \quad \forall y \in X\}, \quad (3.8)$$

an elements of $\partial_{\varepsilon, \Phi} f(x)$ are called (ε, Φ) -subgradients of f at $x \in \text{dom } f$.

Clearly, $\partial_\Phi f(x_0)$ is a set-valued mapping $\partial_\Phi f : X \rightrightarrows \Phi$ and $\text{dom } \partial_\Phi f = \{x \in X : \partial_\Phi f(x) \neq \emptyset\}$. For $\phi \in \Phi$, we define $(\partial_\Phi f)^{-1}(\phi) = \{x \in X : \phi \in \partial_\Phi f(x)\}$. Trivially, if $x \notin \text{dom } f$ then $\partial_\Phi f(x) = \emptyset$. For an Φ -convex function f and $x_0 \in \text{dom } f$, $\partial_\Phi f(x_0) \neq \emptyset$ if there exists $\phi \in \Phi$, $\phi \leq f$ and $\phi(x_0) = f(x_0)$. Moreover, f is Φ -convex if and only if $\text{dom } f = \bigcap_{\varepsilon > 0} \text{dom } \partial_{\varepsilon, \Phi} f$ while $\text{dom } \partial_\Phi f \subset \text{dom } f$ [19, Proposition 2.7].

Additional facts and results of ε - Φ subdifferentials can be found in [56, 22]. The following properties follow directly from the definitions.

Proposition 1. *Let X be a nonempty set and $f : X \rightarrow (-\infty, +\infty]$.*

- (i) *For all $x \in \text{dom } f$ and $\varepsilon \geq 0$, an element $\phi \in \Phi$ is a (ε, Φ) -subgradient of f at x if and only if*

$$f(x) + f_\Phi^*(\phi) \leq \phi(x) + \varepsilon. \quad (3.9)$$

- (ii) *$\text{dom } f^* = \bigcap_{\varepsilon > 0} \partial_{\varepsilon, \Phi} f(X)$.*

The proof for (i) can be found in [82, Proposition 7.10] and (ii) in [19, Proposition 2.4].

3.2 $\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions in Hilbert Space

Let X be a Hilbert space with norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$. One important instance of the class of elementary functions is the following

$$\Phi_{lsc}^{\mathbb{R}} := \{ \phi : \phi(x) = -a\|x\|^2 + \langle u, x \rangle + c \text{ where } a, c \in \mathbb{R}, u \in X \}.$$

Notice that $\Phi_{lsc}^{\mathbb{R}}$ includes the class of affine functions, so a proper lsc convex function is $\Phi_{lsc}^{\mathbb{R}}$ -convex. In fact, $\Phi_{lsc}^{\mathbb{R}}$ also contains the class of strongly convex, weakly convex functions (see [72, 82, 95]).

We also consider a subclass of $\Phi_{lsc}^{\mathbb{R}}$, defined as

$$\Phi_{lsc}^a := \{ \phi : \phi(x) = -a\|x\|^2 + \langle u, x \rangle + c \text{ where } c \in \mathbb{R}, u \in X \}, \quad (3.10)$$

by fixing the coefficient $a \in \mathbb{R}$. When $a > 0$, Φ_{lsc}^a -convexity is equivalent to weak convexity ($2a$ -weakly convex, also called $2a$ -semiconvex [23, Proposition 1.1.3]), if $a < 0$, Φ_{lsc}^a -convexity is equivalent to strong convexity ($2a$ -strongly convex) (see [95, 90]). By [6, Corollary 11.17], any Φ_{lsc}^a -convex function, $a < 0$, is supercoercive and has a unique minimiser. Another subclass of $\Phi_{lsc}^{\mathbb{R}}$ is, where one considers the coefficients $a \geq 0$ i.e.

$$\Phi_{lsc}^{\geq} := \{ \phi : \phi(x) = -a\|x\|^2 + \langle u, x \rangle + c \text{ where } a \geq 0, c \in \mathbb{R}, u \in X \},$$

which has been studied extensively in [82] and further investigated in [12]. Since the classes $\Phi_{lsc}^{\mathbb{R}}$, Φ_{lsc}^a , $\Phi_{lsc}^{\geq}$ consist of continuous function. Any $\Phi_{lsc}^{\mathbb{R}}$ -convex is also lower semicontinuous. In fact, the class of $\Phi_{lsc}^{\geq}$ -convex functions has been proved to coincide with the set of all lower semi-continuous functions minorized by a function from $\Phi_{lsc}^{\geq}$ on X [82, Proposition 6.3]. This fact

motives the subscript lsc in $\Phi_{lsc}^{\mathbb{R}}$ notation. Moreover, we show the following.

Proposition 2. *Let X be a Hilbert space, $f : X \rightarrow (-\infty, +\infty]$ be proper $\Phi_{lsc}^{\mathbb{R}}$ -convex. If there exists $\phi \in \text{supp}_{\Phi_{lsc}^{\mathbb{R}}} f$ with $a_\phi < 0$ then $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ and there exists $\psi \in \Phi_{lsc}^{\geq}$ such that $f(y) \geq \psi(y)$ for all $y \in X$, i.e. f is $\Phi_{lsc}^{\geq}$ -convex. Moreover, the converse inclusion $\Phi_{lsc}^{\geq}$ -convex $\subset \Phi_{lsc}^{\mathbb{R}}$ -convex is obvious.*

Proof. By assumption, we know that

$$f(y) \geq \phi(y) = -a_\phi \|y\|^2 + \langle u_\phi, y \rangle + c_\phi, \quad (3.11)$$

for all $y \in X$ with $a_\phi < 0$. By taking the limit both sides of (3.11) with $\|y\| \rightarrow +\infty$, we have

$$\lim_{\|y\| \rightarrow +\infty} f(y) \geq \lim_{\|y\| \rightarrow +\infty} \phi(y) = +\infty. \quad (3.12)$$

The second assertion can be seen as

$$f(y) \geq \phi(y) = -a_\phi \|y\|^2 + \langle u_\phi, y \rangle + c_\phi \geq \langle u_\phi, y \rangle + c_\phi \in \Phi_{lsc}^{\geq},$$

for all $y \in X$, $\Phi_{lsc}^{\geq}$ -convexity of f follows from [82, Proposition 6.3]. $\square$

Example 2. An even degree polynomials with a positive leading coefficient is coercive which has a global minimum. Moreover, it is differentiable so it is continuous. Hence, by [82, Proposition 6.3], it is $\Phi_{lsc}^{\geq}$ -convex so it is $\Phi_{lsc}^{\mathbb{R}}$ by Proposition 2.

Recall that according to the general definition (Definition 1), a function $f : X \rightarrow (-\infty, +\infty]$ is Φ_{lsc}^a -convex, if for every $x \in X$, we have

$$f(x) = \sup_{\phi \in \Phi_{lsc}^a} \{\phi(x) \mid \phi \leq f\}.$$

When $a = 0$ then we get the affine elementary functions as in Example 1, so f is lsc proper convex iff f is Φ_{lsc}^0 -convex. While for other cases of a , it has been discussed in (3.10).

In [82, Example 6.2], the set $\Phi_{lsc}^{\geq}$ -convex functions coincides with the set of all lower semi-continuous functions defined on X and minorized by a function from $\Phi_{lsc}^{\geq}$, see [82, Proposition 6.3]. Quadratically minorized functions have also been investigated in [5] by approximating them with Moreau envelope and Lasry-Lions double envelope. The starting point of numerous variants of convex functions stems from the work of Hyers and Ulam [55], where they defined δ -convex function for $\delta \geq 0$, next approximately convex functions were investigated in e.g. [37, 71, 79] and the references therein. Rolewicz has coined the term "paraconvexity", [78], by turning δ into a non-negative function. The definitions of strong and weak convexity were also given in [90] by Vial.

For the sake of completeness, we give a definition of abstract subdifferentials with respect to the class $\Phi_{lsc}^{\mathbb{R}}$ which fully agrees to the definition of Φ -subdifferentials in Definition 3.

Definition 4. [72, 82] The $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f at $x_0 \in \text{dom } f$ is an element $\phi \in \Phi_{lsc}^{\mathbb{R}}$ such that

$$(\forall y \in X) \quad f(y) - f(x_0) \geq \phi(y) - \phi(x_0). \quad (3.13)$$

The collection of all such ϕ satisfying (3.13) is called $\Phi_{lsc}^{\mathbb{R}}$ -subdifferential and is denoted by $\partial_{lsc}^{\mathbb{R}} f(x_0)$. We give an illustration of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials in Figure 3.1. Analogously, we define $\Phi_{lsc}^{\geq}$ -subdifferentials and Φ_{lsc}^a -subdifferentials of f and denote as $\partial_{lsc}^{\geq} f$ and $\partial_{lsc}^a f$, respectively. In particular, $\partial_{lsc}^0 f$ denotes the subdifferentials in the sense of convex analysis.

By Proposition 2, we see that the class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions coincides with the class of $\Phi_{lsc}^{\geq}$ -convex functions. On the other hand, the set $\Phi_{lsc}^{\mathbb{R}}$ is larger than $\Phi_{lsc}^{\geq}$ which implies $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials can be larger than $\Phi_{lsc}^{\geq}$ -

subdifferentials. This is the reason why we decide to include the class $\Phi_{lsc}^{\mathbb{R}}$ -convex functions into our consideration in the sequel.

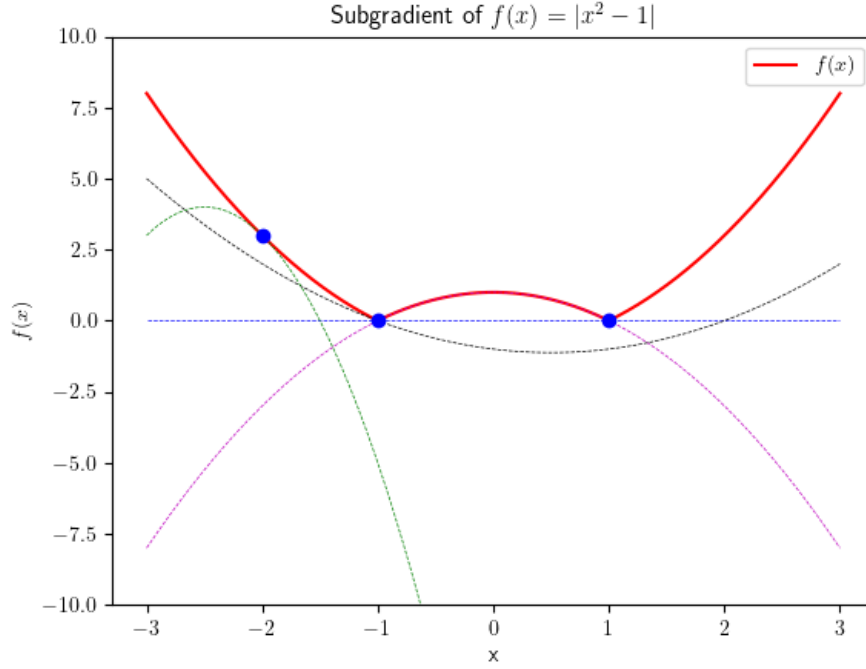


Figure 3.1: $\Phi_{lsc}^{\mathbb{R}}$ -Subgradients of f at different points (dotted lines). Observe that the purple curve is the $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f at every point between -1 and 1 .

Remark 1.

- (i) From the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, the constant c cancels in (4). Therefore, for $x_0 \in \text{dom } f$, $\phi \in \partial_{lsc}^{\mathbb{R}} f(x_0)$ and any real $c \in \mathbb{R}$, the function

$$\phi(\cdot) - c = -a\|\cdot\|^2 + \langle u, \cdot \rangle \in \partial_{lsc}^{\mathbb{R}} f(x_0).$$

It is shown in [12] that the constant c is not important in the study of conjugate duality and can be neglected.

- (ii) The above remark does not imply that the constant term is insignificant. In fact, its purpose is to serve as a connection between $\Phi_{lsc}^{\mathbb{R}}$ -convexity and $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentiability. Let us consider $x_0 \in \text{dom } f$ and $\phi \in \partial_{lsc}^{\mathbb{R}} f(x_0)$

$$(\forall y \in X) \quad f(y) - f(x_0) \geq \phi(y) - \phi(x_0).$$

The function $h(x) := \phi(x) - \phi(x_0) + f(x_0) \leq f(x)$ for all $x \in X$ and $f(x_0) = h(x_0)$. Hence, h is a $\Phi_{lsc}^{\mathbb{R}}$ -minorant of f , i.e. $h \in \Phi_{lsc}^{\mathbb{R}}$ with the constant term $-a\|x_0\|^2 + \langle u, x_0 \rangle + f(x_0)$.

- (iii) We point out that the $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentiability of f on X implies $\Phi_{lsc}^{\mathbb{R}}$ -convexity of f on X but not conversely: Consider the function $f(x) = -\|x\|$, it is $\Phi_{lsc}^{\mathbb{R}}$ -convex but its $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials is empty at $x = 0$.
- (iv) In general, for any class Φ , we have $f \in \partial_{\Phi} f(x)$ for all $x \in X$ if $f \in \Phi$. Even if $f \in \Phi$ is differentiable and Φ -convex then ∂_{Φ} is not unique (in convex analysis, subdifferentials is unique and coincide with the gradient when the function is differential and convex, see Lemma 6 below).

Since we mostly deal with $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, in the sequel, we identify elements $\phi \in \Phi_{lsc}^{\mathbb{R}}$ with pairs (a, u) where $a \in \mathbb{R}, u \in X$.

We demonstrate some examples of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials which will be important in the following sections.

Example 3. Let X be Hilbert and $\gamma > 0$ and consider the function $g_{\gamma}(x) = \frac{1}{2\gamma} \|x\|^2$. We calculate the $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of g_{γ} at $x_0 \in X$. By definition, $\phi = (a, u) \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x_0)$ has to satisfy

$$g_{\gamma}(y) - g_{\gamma}(x_0) \geq \phi(y) - \phi(x_0),$$

for all $y \in X$. Simplifying both sides gives

$$\left(\frac{1}{2\gamma} + a\right) \|y\|^2 - \langle u, y \rangle \geq \left(\frac{1}{2\gamma} + a\right) \|x_0\|^2 - \langle u, x_0 \rangle. \quad (3.14)$$

- If $a = -1/(2\gamma)$, then $\langle u, y - x \rangle \leq 0$ for all $y \in X$, by (3.14) u must be zero. Hence, $(-\frac{1}{2\gamma}, 0) \in \partial_{\Phi_{LSC}}^{\mathbb{R}} g_{\gamma}(x)$ for all $x \in X$.
- If $a > -1/(2\gamma)$, from (3.14) we have, for any $y \in X$ (since we are in Hilbert space X)

$$\left(\frac{1}{2\gamma} + a\right) \left\|y - \frac{u}{\frac{1}{\gamma} + 2a}\right\|^2 \geq \left(\frac{1}{2\gamma} + a\right) \left\|x_0 - \frac{u}{\frac{1}{\gamma} + 2a}\right\|^2. \quad (3.15)$$

This means $u = \left(\frac{1}{\gamma} + 2a\right) x_0$. In particular, $\partial_{LSC}^a g_{\gamma}(x_0) = \{\phi = (a, u) \in \Phi_{LSC}^a : u = \left(\frac{1}{\gamma} + 2a\right) x_0\}$.

- If $a < -1/(2\gamma)$, from (3.15) it should be

$$(\forall y \in X) \quad \left\|y - \frac{u}{\frac{1}{\gamma} + 2a}\right\|^2 \leq \left\|x_0 - \frac{u}{\frac{1}{\gamma} + 2a}\right\|^2,$$

which is impossible. Therefore, a cannot take any values less than $-\frac{1}{2\gamma}$.

Hence, the $\Phi_{LSC}^{\mathbb{R}}$ -subdifferentials of g_{γ} at x_0 takes the form

$$\partial_{LSC}^{\mathbb{R}} g_{\gamma}(x_0) = \left\{ \phi \in \Phi_{LSC}^{\mathbb{R}} : \phi = \left(a, \left(\frac{1}{\gamma} + 2a\right) x_0\right), 2\gamma a \geq -1 \right\}, \quad (3.16)$$

and similarly, we calculate

$$\partial_{LSC}^{\geq} g_{\gamma}(x_0) = \left\{ \phi \in \Phi_{LSC}^{\mathbb{R}} : \phi = \left(a, \left(\frac{1}{\gamma} + 2a\right) x_0\right), a \geq 0 \right\}.$$

Notice that $g_{\gamma} \in \partial_{LSC}^{\mathbb{R}} g_{\gamma}(x_0)$ for any $x_0 \in X$ as $g_{\gamma} = \left(-\frac{1}{2\gamma}, 0\right) \in \Phi_{LSC}^{\mathbb{R}}$.

Example 4. Let X be a Hilbert space. We calculate the inverse of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials of g_{γ} , $(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1} : \Phi_{lsc}^{\mathbb{R}} \rightrightarrows X$ which will be useful in Chapter 5. Equivalently, $(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1} : \mathbb{R} \times X \rightrightarrows X$, with $\text{dom}(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1} = \{\phi \in \Phi_{lsc}^{\mathbb{R}} : (\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi) \neq \emptyset\}$,

$$(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi) := \{x_0 \in X : \phi \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x_0)\}.$$

Consider the following cases.

- If $2\gamma a < -1$, then $(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi) = \emptyset$ as we need $2\gamma a \geq -1$ as in Example 3.
- If $2\gamma a > -1$, for $\phi \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x_0)$ must take the form $\phi = (a, u) = \left(a, \left(\frac{1}{\gamma} + 2a\right)x_0\right)$. We can find x_0 by letting $u = \left(\frac{1}{\gamma} + 2a\right)x_0$ i.e.

$$x_0 = \frac{1}{\frac{1}{\gamma} + 2a}u = \frac{\gamma}{1 + 2\gamma a}u.$$

- If $a = -\frac{1}{2\gamma}$, then u must be zero and $\left(-\frac{1}{2\gamma}, 0\right) \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x)$ for all $x \in X$.

In conclusion, for any $\phi \in \Phi_{lsc}^{\mathbb{R}}$, we have

$$(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi) = \begin{cases} \left\{\frac{\gamma}{1+2\gamma a}u\right\} & \text{if } 2\gamma a > -1, \\ X & \text{if } a = -\frac{1}{2\gamma}, u = 0, \\ \emptyset & \text{otherwise.} \end{cases} \quad (3.17)$$

Observe that

$$\phi \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}((\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi)), \quad (3.18)$$

for all $\phi \in \text{dom}(\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}$. When $a = -1/(2\gamma)$, $u = 0$, (3.18) holds because

of Example 3 and Example 4. When $\phi = (a, u) \in \Phi_{lsc}^{\mathbb{R}}$ with $2\gamma a > -1$, then

$$\begin{aligned} \partial_{lsc}^{\mathbb{R}} g_{\gamma} ((\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi)) &= \partial_{lsc}^{\mathbb{R}} g_{\gamma} \left(\frac{\gamma u}{1 + 2\gamma a} \right) \\ &= \left\{ \phi_1 \in \Phi_{lsc}^{\mathbb{R}} : \phi_1 = \left(a_1, \frac{1 + 2\gamma a_1}{1 + 2\gamma a} u \right), a_1 \geq -\frac{1}{2\gamma} \right\}. \end{aligned}$$

Since $a_1 \geq -1/2\gamma$, it is clear that ϕ lies inside the above set. Hence, (3.18) holds.

Example 5. Let $X = \mathbb{R}^n$, consider the function $f(x) = \langle x, Qx \rangle$ where $Q \in \mathbb{R}^{n \times n}$ is a real symmetric matrix and Id is the identity matrix. We compute the $\Phi_{lsc}^{\mathbb{R}}$ -subdifferential of f at $x \in \mathbb{R}^n$. We need to find $\phi \in \Phi_{lsc}^{\mathbb{R}}$ such that $\phi = (a, u) \in \partial_{lsc}^{\mathbb{R}} f(x)$. By definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferential of f at x , we have

$$(\forall y \in \mathbb{R}^n) \quad \langle y, (Q + aId)y \rangle - \langle x, (Q + aId)x \rangle \geq \langle u, y - x \rangle. \quad (3.19)$$

For any matrix $A \in \mathbb{R}^{n \times n}$, the following identity holds

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

where $A^{\top}$ is the transpose of A . Since Q is real symmetric, we apply the above identity for $A = (Q + aId)$ which is also real symmetric i.e. $A^{\top} = A$. Then inequality (3.19) becomes

$$\langle y - x, (Q + aId)(y - x) \rangle \geq \langle u - 2(Q + aId)x, y - x \rangle. \quad (3.20)$$

We can diagonalize $(Q + aId)$ into $P(D + aId)P^{\top}$ with a diagonal matrix D with the eigenvalues of Q on the diagonal and P an orthogonal matrix which contains the corresponding eigenvectors, which is known as *Schur*

decomposition [48, Section 2.1.7]. Hence, we have

$$\begin{aligned}\langle y - x, (Q + aId)(y - x) \rangle &= \langle y - x, P(D + aId)P^\top(y - x) \rangle \\ &= \langle P^\top(y - x), (D + aId)P^\top(y - x) \rangle.\end{aligned}\quad (3.21)$$

Plugging (3.21) back into (3.20), we obtain

$$\langle P^\top(y - x), (D + aId)P^\top(y - x) \rangle \geq \langle u - 2(Q + aId)x, y - x \rangle. \quad (3.22)$$

As $PP^\top = Id$ is the identity matrix, by changing the variables $\bar{y} = P^\top y, \bar{x} = P^\top x$, (3.22) turns into

$$\langle \bar{y} - \bar{x}, (D + aId)(\bar{y} - \bar{x}) \rangle \geq \langle P^\top u - 2(D + aId)\bar{x}, \bar{y} - \bar{x} \rangle. \quad (3.23)$$

Because (3.22) holds for all $y \in \mathbb{R}^n$ and P is orthogonal, so (3.23) has to hold for all $\bar{y} \in \mathbb{R}^n$.

As we are in $\mathbb{R}^n$, by using standard inner product, we can write (3.23) explicitly

$$\sum_{i=1}^n (d_i + a)(\bar{y}_i - \bar{x}_i)^2 \geq \sum_{i=1}^n (P_i^\top u - 2(d_i + a)\bar{x}_i)(\bar{y}_i - \bar{x}_i), \quad (3.24)$$

where d_i is the i -th eigenvalue of Q and $P_i^\top$ is the i -th row of matrix $P^\top$.

- If there is $0 < j \leq n$ such that $d_j + a = 0$ then we have, from (3.24)

$$\begin{aligned}&\sum_{i \neq j=1}^n (d_i + a) \left[(\bar{y}_i - \bar{x}_i) - \frac{P_i^\top u - 2(d_i + a)\bar{x}_i}{2(d_i + a)} \right]^2 \\ &\geq \sum_{i \neq j=1}^n \frac{(P_i^\top u - 2(d_i + a)\bar{x}_i)^2}{4(d_i + a)} + (P_j^\top u)(\bar{y}_j - \bar{x}_j).\end{aligned}$$

Note that all components of $\bar{y}$ are separable. This holds for all $\bar{y} \in \mathbb{R}^n$

so we can choose $\bar{y}_i = \bar{x}_i$

$$(\forall \bar{y}_j \in \mathbb{R}) \quad 0 \geq (P_j^\top u)(\bar{y}_j - \bar{x}_j),$$

which implies $P_j^\top u = 0$. We can extended this to the case where there are repeated eigenvalues i.e. $d_j + a = d_{j+1} + a = \dots = d_k + a = 0$ for $1 \leq j < k \leq n$.

- Now we assume that $d_i + a \neq 0$ for all $i = 1, \dots, n$, we have

$$\sum_{i=1}^n (d_i + a) \left[(\bar{y}_i - \bar{x}_i) - \frac{P_i^\top u - 2(d_i + a)\bar{x}_i}{2(d_i + a)} \right]^2 \geq \sum_{i=1}^n \frac{(P_i^\top u - 2(d_i + a)\bar{x}_i)^2}{4(d_i + a)}. \quad (3.25)$$

If there is $1 \leq j \leq n$ such that $d_j + a < 0$ while $i \neq j \leq n, d_i + a > 0$

$$\begin{aligned} & \sum_{i \neq j=1}^n (d_i + a) \left[(\bar{y}_i - \bar{x}_i) - \frac{P_i^\top u - 2(d_i + a)\bar{x}_i}{2(d_i + a)} \right]^2 - \frac{(P_j^\top u - 2(d_j + a)\bar{x}_j)^2}{4(d_j + a)} \\ & \geq \sum_{i \neq j=1}^n \frac{(P_i^\top u - 2(d_i + a)\bar{x}_i)^2}{4(d_i + a)} - (d_j + a) \left[(\bar{y}_j - \bar{x}_j) - \frac{P_j^\top u - 2(d_j + a)\bar{x}_j}{2(d_j + a)} \right]^2. \end{aligned} \quad (3.26)$$

Both sides are non-negative and as $\bar{y}$ is separable, we follow the same argument as in the previous case to obtain, for all $\bar{y}_j \in \mathbb{R}$,

$$-\frac{(P_j^\top u - 2(d_j + a)\bar{x}_j)^2}{4(d_j + a)} \geq -(d_j + a) \left[(\bar{y}_j - \bar{x}_j) - \frac{P_j^\top u - 2(d_j + a)\bar{x}_j}{2(d_j + a)} \right]^2.$$

This implies $d_j + a = 0$ or the above inequality cannot hold. Therefore, we only need to consider $d_i + a > 0$ for all $1 \leq i \leq n$. From (3.25), we derive that RHS is zero which is $P_i^\top u - 2(d_i + a)\bar{x}_i = 0$ for all $1 \leq i \leq n$.

Combining all the cases above, we obtain that $(D + aId)$ is non-negative semi-definite and $P^\top u - 2(D + aId)\bar{x} = 0$. Hence, a has to take a value

$a \geq -\min \lambda_Q$ where λ_Q are the eigenvalues of Q . In conclusion,

$$\partial_{lsc}^{\mathbb{R}} f(x) = \left\{ \phi = (a, u) \in \Phi_{lsc}^{\mathbb{R}} : a \geq -\min \lambda_Q, u = 2(Q + aId)x \right\}.$$

Example 6. Let us consider the indicator function of a nonempty set C of Hilbert space X ,

$$f(x) = \iota_C(x) = \begin{cases} 0 & x \in C, \\ +\infty & \text{otherwise,} \end{cases}$$

and calculate $\Phi_{lsc}^{\mathbb{R}}$ -subgradients of f at $x \in X$. If $x \notin C$ then $\partial_{lsc}^{\mathbb{R}} f(x) = \emptyset$, so we consider $x \in C$. Let us take $(a, u) \in \partial_{lsc}^{\mathbb{R}} f(x)$

$$(\forall y \in X) \quad f(y) \geq -a(\|y\|^2 - \|x\|^2) + \langle u, y - x \rangle.$$

When $y \notin C$, the LHS becomes $+\infty$ and the inequality is fulfilled for all a and u . Hence, the above inequality reduces to

$$(\forall y \in C) \quad a\|y\|^2 - \langle u, y \rangle \geq a\|x\|^2 - \langle u, x \rangle. \quad (3.27)$$

- If $a = 0$ then we have $\langle u, y - x \rangle \leq 0$ for all $y \in C$, this means u belongs to the normal cone of C at x in the sense of convex analysis.
- If $a > 0$, we complete the square in (3.27) to obtain

$$a\left\|y - \frac{u}{2a}\right\|^2 \geq a\left\|x - \frac{u}{2a}\right\|^2. \quad (3.28)$$

Since $a > 0$ and this holds for all $y \in C$, we can remove a and taking the infimum with respect to $y \in C$

$$\inf_{y \in C} \left\|y - \frac{u}{2a}\right\|^2 \geq \left\|x - \frac{u}{2a}\right\|^2.$$

This means that x is a projection of $u/(2a)$ onto the set C .

- If $a < 0$, then (3.28) changes sign when dividing by a

$$\left\| y - \frac{u}{2a} \right\|^2 \leq \left\| x - \frac{u}{2a} \right\|^2. \quad (3.29)$$

This leads to $x \in \operatorname{argmax}_{y \in C} \left\| y - \frac{u}{2a} \right\|$.

In conclusion, for $x \in C$

$$\partial_{lsc}^{\mathbb{R}} \iota_C(x) = \left\{ \phi = (a, u) \in \Phi_{lsc}^{\mathbb{R}} : \begin{cases} (\forall y \in C) \quad \langle u, y - x \rangle \leq 0 & a = 0 \\ x \in \operatorname{argmin}_{y \in C} \left\| y - \frac{u}{2a} \right\| & a > 0 \\ x \in \operatorname{argmax}_{y \in C} \left\| y - \frac{u}{2a} \right\| & a < 0 \end{cases} \right\}.$$

Below, we show that there is a connection between $\partial_{lsc}^{\mathbb{R}} \iota_C(x)$ and the proximal normal cone [32, Chapter 1.1] which is defined as follows: For a point $x \in C$, the proximal normal cone to the set C at x is the set

$$N_p(x, C) = \{v \in X : \exists t > 0, \operatorname{dist}(x + tv, C) = t\|v\|\}. \quad (3.30)$$

Proposition 3. *Let X be a Hilbert space. For every nonempty set $C \subset X$ and $x \in C$, if $\phi = (a, u) \in \partial_{lsc}^{\mathbb{R}} \iota_C(x)$ then*

$$\nabla \phi(x) = u - 2ax \in N_p(x, C).$$

On the other hand, for any $v \in N_p(x, C)$ there exists $\phi \in \partial_{lsc}^{\mathbb{R}} \iota_C(x)$ such that $\nabla \phi(x) = v$.

Proof. Let $x \in C$ and $\phi = (a, u) \in \partial_{lsc}^{\mathbb{R}} \iota_C(x)$. If $a \leq 0$, from (3.27) we have for all $y \in C$,

$$a\|y\|^2 - \langle u, y \rangle \geq a\|x\|^2 - \langle u, x \rangle \Rightarrow 0 \geq a\|y - x\|^2 \geq \langle u - 2ax, y - x \rangle. \quad (3.31)$$

The RHS can be further expressed

$$0 \geq \langle u - 2ax, y - x \rangle = \frac{1}{2} [\|u - 2ax\|^2 + \|y - x\|^2 - \|y - x - (u - 2ax)\|^2], \quad (3.32)$$

and finally for all $y \in C$,

$$\|y - x - (u - 2ax)\|^2 \geq \|u - 2ax\|^2. \quad (3.33)$$

This means x is a projection of $(x + u - 2ax)$ onto C which means $u - 2ax \in N_p(x, C)$ with $t = 1$.

If $a > 0$, then

$$0 \geq \left\langle \frac{u - 2ax}{a}, y - x \right\rangle - \|y - x\|^2 = \frac{1}{2} \left(\left\| \frac{u - 2ax}{2a} \right\|^2 - \left\| y - x - \frac{u - 2ax}{2a} \right\|^2 \right),$$

or

$$\left\| y - x - \frac{u - 2ax}{2a} \right\|^2 \geq \left\| \frac{u - 2ax}{2a} \right\|^2 \quad (\forall y \in C).$$

Thus, x is the projection of $x + \frac{u - 2ax}{2a}$ onto C which means $u - 2ax \in N_p(x, C)$ with $t = \frac{1}{2a}$. In both cases, $\nabla \phi(x) = u - 2ax \in N_p(x, C)$.

Conversely, let $v \in N_p(x, C)$. There exists $t > 0$ such that $x \in \text{Proj}_C(x + tv)$ or

$$(\forall y \in C) \quad \|tv\|^2 \leq \|y - x - tv\|^2,$$

which can be expressed as

$$-\frac{1}{2t}\|y\|^2 + \left\langle v + \frac{x}{t}, y \right\rangle \leq -\frac{1}{2t}\|x\|^2 + \left\langle v + \frac{x}{t}, x \right\rangle,$$

which implies that $\phi(\cdot) = -\frac{1}{2t}\|\cdot\|^2 + \left\langle v + \frac{x}{t}, \cdot \right\rangle \in \partial_{lsc}^{\mathbb{R}} \iota_C(x)$ and $\nabla \phi(x) = v$. $\square$

From the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, it is obvious that $\phi = (0, 0) \in \partial_{lsc}^{\mathbb{R}} f(x_0)$, if and only if x_0 is a global minimizer of f (c.f. [82, Proposition 7.13]). In Example 3, the only case where $(0, 0) \in \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x_0)$ is $x_0 = 0$. In

the sequel, we use the concept of a_0 -critical points as defined below.

Definition 5. Let $f : X \rightarrow (-\infty, +\infty]$, a point $x_0 \in X$ is a_0 -critical point if $(a_0, 2a_0x_0) \in \partial_{lsc}^{\mathbb{R}} f(x_0)$, in other words

$$(\forall x \in X) \quad f(x) - f(x_0) \geq -a_0 \|x - x_0\|^2. \quad (3.34)$$

If $a_0 > 0$, then a_0 -criticality has been applied in [38, 11], to analyze algorithms involving the class of weakly convex functions.

If $a_0 \leq 0$, then (3.34) implies x_0 is the global minimizer of f .

3.3 Weakly Convex Functions

For the sake of completeness, we give the definition of weakly convex function and its subdifferentials and connections to the $\Phi_{lsc}^{\mathbb{R}}$ -convex functions.

Definition 6 (ρ -weak convexity). Let X be a Hilbert space. A function $h : X \rightarrow (-\infty, +\infty]$ is said to be ρ -weakly convex with $\rho \geq 0$ such that for $\lambda \in [0, 1]$ the following inequality holds:

$$(\forall (x, y) \in X^2) \quad h(\lambda x + (1 - \lambda)y) \leq \lambda h(x) + (1 - \lambda)h(y) + \lambda(1 - \lambda)\frac{\rho}{2}\|x - y\|^2. \quad (3.35)$$

We refer to ρ as the *modulus of weak convexity* of the function h .

This class of functions has been investigated e.g. [5, 59, 90]. Rockafellar introduced and investigated a larger class of *lower- C^k functions* (see [76, Formula 1.6]). Weakly convex functions can be equivalently characterised in the following manner.

Proposition 4 ([23, Proposition 1.1.3]). *Let X be a Hilbert space. Function $h : X \rightarrow (-\infty, +\infty]$ is ρ -weakly convex if and only if $h + \frac{\rho}{2}\|\cdot\|^2$ is a convex function.*

The following proposition provides a characterisation of weak convexity for Fréchet differentiable functions.

Proposition 5. *Let X be a Hilbert space and let $h : X \rightarrow (-\infty, +\infty]$ be Fréchet differentiable on a convex set $U \subset X$. Then, the following are equivalent:*

(i) h is ρ -weakly convex on U ;

(ii) for every $x \in U$

$$(\forall y \in X) \quad h(y) \geq h(x) + \langle \nabla h(x) \mid y - x \rangle - \frac{\rho}{2} \|x - y\|^2. \quad (3.36)$$

Proof. By Proposition 4 and [6, Proposition 17.7], for all $x \in U$

$$\begin{aligned} (i) \text{ holds} &\Leftrightarrow (\forall y \in X) \quad h(y) + \frac{\rho}{2} \|y\|^2 \geq h(x) + \frac{\rho}{2} \|x\|^2 + \langle \nabla h(x) + \rho x \mid y - x \rangle \\ &\Leftrightarrow (\forall y \in X) \quad h(y) \geq h(x) + \langle \nabla h(x) \mid y - x \rangle - \frac{\rho}{2} \|x\|^2 + \langle \rho x \mid y \rangle - \frac{\rho}{2} \|y\|^2 \\ &\Leftrightarrow (\forall y \in X) \quad h(y) \geq h(x) + \langle \nabla h(x) \mid y - x \rangle - \frac{\rho}{2} \|y - x\|^2 \\ &\Leftrightarrow (ii) \text{ holds.} \end{aligned} \quad (3.37)$$

□

We report the following result, commonly known as *Descent Lemma*, that allows to provide sufficient conditions for weak convexity.

Lemma 1 ([6, Lemma 2.64]). *Let X be a Hilbert space, let U be a non-empty open convex subset of X . Let $h : U \rightarrow \mathbb{R}$ be a Fréchet differentiable function such that ∇h is L_h -Lipschitz continuous on U . Let x and y be in U . Then the following holds:*

$$(i) \quad |h(x) - h(y) - \langle x - y \mid \nabla h(y) \rangle| \leq \frac{L_h}{2} \|x - y\|^2.$$

$$(ii) \quad |\langle x - y \mid \nabla h(x) - \nabla h(y) \rangle| \leq L_h \|x - y\|^2.$$

As an immediate consequence of the Descent Lemma, we obtain the following proposition:

Proposition 6. *Let X be a Hilbert space. Let $h : X \rightarrow \mathbb{R}$ be a Fréchet differentiable function such that ∇h is L_h -Lipschitz continuous on X . Then h is L_h -weakly convex on X .*

Proof. By Descent Lemma, for all $x, y \in X$

$$-\frac{L_h}{2}\|x - y\|^2 \leq h(y) - h(x) - \langle y - x \mid \nabla h(x) \rangle,$$

which, by Proposition 5, is equivalent to L_h -weak convexity of h on X . $\square$

We define the *global proximal subdifferential* for weakly convex functions.

Definition 7 (Global proximal subdifferentials). Let X be a Hilbert space. The global proximal subdifferentials of a function $h : X \rightarrow (-\infty, +\infty]$ at $x_0 \in \text{dom } h$ for constant $C \geq 0$ is defined as follows:

$$\partial_C h(x_0) = \{v \in X \mid h(x) \geq h(x_0) + \langle v \mid x - x_0 \rangle - C\|x - x_0\|^2, \forall x \in X\}. \quad (3.38)$$

In view of (3.38), $\partial_0 h(x)$ denotes the subdifferential in the sense of convex analysis.

By [59, Proposition 3.1], for a proper lsc ρ -weakly convex function h , the global proximal subdifferential $\partial_{\rho/2} h(x_0)$ defined by (3.38) coincides with the set of local proximal subgradients which satisfy the proximal subgradient inequality (3.38) locally in a neighbourhood of x_0 . This important property is called *the globalisation property*, see e.g. [56]. Moreover, when h is ρ -weakly convex, by [59, Theorem 3.1], the proximal subdifferential $\partial_{\rho/2} h(x_0)$ coincides with the Clarke subdifferential $\partial_c h(x_0)$ (see also [4]). An extensive study of local proximal subdifferentials in Hilbert spaces can be found in [32, Chapter 2] and in [77, 13].

Proposition 7 ([10, Proposition 2]). *Let X be a Hilbert space and $h : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function with $\rho \geq 0$. Then $\text{dom } \partial_{\rho/2} h \subset \text{dom } h$.*

We show the relation between $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials and proximal subdifferentials in the following lemma.

Lemma 2. *Let X be a Hilbert space, for a $\Phi_{lsc}^{\mathbb{R}}$ -convex function f , let $x \in \text{dom } f$ such that $\partial_{lsc}^{\mathbb{R}} f(x) \neq \emptyset$. Then we have*

$$\partial_{lsc}^{\mathbb{R}} f(x) \neq \emptyset \Leftrightarrow \bigcup_{a \geq 0} \{a\} \times \partial_a f(x) \neq \emptyset. \quad (3.39)$$

Proof. Let $\phi = (a, u) \in \partial_{lsc}^{\mathbb{R}} f(x)$, by the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials

$$\begin{aligned} (\forall y \in X) \quad f(y) - f(x) &\geq -a(\|y\|^2 - \|x\|^2) + \langle u, y - x \rangle \\ &= \langle u - 2ax, y - x \rangle - a\|y - x\|^2. \end{aligned} \quad (3.40)$$

If $a < 0$ then $a\|y - x\|^2 \leq 0$ so we can further estimate (3.40)

$$(\forall y \in X) \quad f(y) - f(x) \geq \langle u - 2ax, y - x \rangle,$$

which implies that $u - 2ax \in \partial_0 f(x)$ and $(0, u - 2ax) \in \bigcup_{a \in \mathbb{R}} \{a\} \times \partial_a f(x)$. Otherwise, for $a \geq 0$, we can deduce $(a, u - 2ax) \in \bigcup_{a \in \mathbb{R}} \{a\} \times \partial_a f(x)$ directly from (3.40).

Conversely, if $(a, w) \in \bigcup_{a \in \mathbb{R}} \{a\} \times \partial_a f(x)$ then

$$\begin{aligned} (\forall y \in X) \quad f(y) - f(x) &\geq \langle w, y - x \rangle - a\|y - x\|^2 \\ &= -a(\|y\|^2 - \|x\|^2) + \langle w + 2ax, y - x \rangle. \end{aligned} \quad (3.41)$$

This means $(a, w + 2ax) \in \partial_{lsc}^{\mathbb{R}} f(x)$. □

Chapter 4

Duality within Abstract Convexity Framework

This chapter presents the main results of Part I which are conjugate and Lagrange duality for composite minimization problem (CP) within the framework of abstract convexity. It includes:

- Section 4.1: we construct conjugate dual problem (4.4) by using perturbation function. We also give several examples of the dual problems in Subsection 4.1.1, when the class of elementary functions are specified.
- Section 4.2: we give necessary and sufficient conditions for zero duality gap in Theorem 3 and its variant Theorem 4. In Corollary 1, we provide sufficient and necessary conditions for zero duality gap when considering special class of elementary functions Φ_{lsc}^a and Ψ_{lsc}^b .
- Section 4.3: we investigate strong duality for conjugate dual (Theorem 6 and Corollary 2).
- Section 4.4: we relate the conjugate dual to Lagrange dual problem. We combine the results in Section 4.2 and *intersection property* (see

Theorem 8) to obtain zero duality gap characterization for Lagrange dual problem (Theorem 9 and Corollary 3).

4.1 Construction of the conjugate dual

We consider the composite minimization problem

$$\inf_{x \in X} f(x) + g(Lx), \quad (\text{CP})$$

where X is a nonempty set and Y is a vector space, $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ and $L : X \rightarrow Y$ is a mapping from X to Y . Note that here we do not assume that L is linear and continuous. Let Φ be a set of elementary functions $\phi : X \rightarrow \mathbb{R}$ and let Ψ be a set of elementary functions $\psi : Y \rightarrow \mathbb{R}$. Assume that both classes are closed under addition of constants and $0 \in \Phi, 0 \in \Psi$.

We introduce the conjugate dual of (CP) by considering perturbed minimization problems of the form

$$\inf_{x \in X} f(x) + g(Lx + y), \quad y \in Y. \quad (4.1)$$

Following the classical ideas coming from convex analysis, see e.g. [16, 15], we calculate the conjugate of the function $\beta : X \times Y \rightarrow (-\infty, +\infty]$, where

$$\beta(x, y) := f(x) + g(Lx + y), \quad x \in X, y \in Y.$$

Note that, in the convex case, conjugation is taken with respect to linear function, while here we are considering elementary functions ϕ, ψ from given classes Φ and Ψ , respectively, see e.g. [69].

We define the conjugate of β with respect to the coupling function

$$\begin{aligned} c : \Phi \times \Psi \times X \times Y &\rightarrow \mathbb{R} \\ (\phi, \psi, x, y) &\mapsto \phi(x) + \psi(Lx + y) - \psi(Lx). \end{aligned} \quad (4.2)$$

For other types of coupling functions defined on Cartesian products of elementary function sets, see e.g., [70].

The c-conjugate of β , $\beta^c : \Phi \times \Psi \rightarrow (-\infty, +\infty]$, is defined as

$$\beta^c(\phi, \psi) = \sup_{x \in X} \sup_{y \in Y} \{c(\phi, \psi, x, y) - \beta(x, y)\}. \quad (4.3)$$

The conjugate dual to problem (CP) is defined as

$$\sup_{\psi \in \Psi} -\beta^c(0, \psi). \quad (4.4)$$

Investigation of the relationship between (CP) and problem (4.4) is the main focus of this chapter. To write (4.4) in a more explicit form, observe that the objective function of (4.4) (the dual objective) is equal to

$$\begin{aligned} -\beta^c(0, \psi) &= -\sup_{x \in X} \sup_{y \in Y} \{c(0, \psi, x, y) - \beta(x, y)\} \\ &= -\sup_{x \in X} \sup_{y \in Y} \{\psi(Lx + y) - \psi(Lx) - f(x) - g(Lx + y)\} \\ &= -\sup_{x \in X} \sup_{z = Lx + y} \{\psi(z) - \psi(Lx) - f(x) - g(z)\} \\ &= -\sup_{x \in X} \left\{ -\psi(Lx) - f(x) + \sup_{z \in Y} \{\psi(z) - g(z)\} \right\} \\ &= -\sup_{x \in X} \{-\psi(Lx) - f(x)\} - g_{\Psi}^*(\psi). \end{aligned} \quad (4.5)$$

Hence, the conjugate dual problem takes the form

$$\sup_{\psi \in \Psi} -\beta^c(0, \psi) = \sup_{\psi \in \Psi} - \sup_{x \in X} \{-\psi(Lx) - f(x)\} - g_{\Psi}^*(\psi). \quad (\text{DCP})$$

Notice that we have kept the formulation of $\sup_{x \in X} \{-\psi(Lx) - f(x)\}$ in the dual problem (DCP), since we do not know if $\psi \circ L$ belongs to the class Φ in general. Below, we give some examples where we further investigate the dual problem.

Example 7. In some cases, we can rewrite the conjugate dual (DCP) in an equivalent and more convenient way.

- If, for any $\psi \in \Psi$, $-\psi \circ L \in \Phi$ (alternatively, we can assume that $\psi \circ L \in \Phi$ and Φ is symmetric i.e. $-\phi \in \Phi$ for all $\phi \in \Phi$) we obtain

$$\sup_{x \in X} \{-f(x) - \psi(Lx)\} = f_{\Phi}^*(-\psi \circ L), \quad (4.6)$$

and the conjugate dual problem (DCP) takes the form

$$\sup_{\psi \in \Psi} -\beta^c(0, \psi) = \sup_{\psi \in \Psi} -f_{\Phi}^*(-\psi \circ L) - g_{\Psi}^*(\psi). \quad (4.7)$$

With the help of the operator $L : X \rightarrow Y$ we define another class of elementary functions as follows

$$\Phi_L := \{-\psi \circ L : X \rightarrow \mathbb{R}, \psi \in \Psi\},$$

and the dual (4.7) can be written as

$$\sup_{\substack{\phi \in \Phi_L, \psi \in \Psi \\ \phi + \psi \circ L = 0}} -f_{\Phi_L}^*(\phi) - g_{\Psi}^*(\psi). \quad (4.8)$$

- If Φ is symmetric ($-\phi \in \Phi$ for all $\phi \in \Phi$), and for any $\psi \in \Psi$, $\psi \circ L$ is

Φ -convex, we have

$$\psi(Lx) = \sup \{ \phi(x) : \phi \in \text{supp } \psi \circ L \}.$$

The optimal value of the dual problem (DCP) (obtained with the help of the coupling function c from (4.2)) satisfies

$$\begin{aligned} \sup_{\psi \in \Psi} -\beta^c(0, \psi) &= \sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi) \\ &= \sup_{\psi \in \Psi} \left\{ \inf_{x \in X} \left(f(x) + \sup_{\phi \in \text{supp } \psi \circ L} \phi(x) \right) - g_{\Psi}^*(\psi) \right\} \\ &\geq \sup_{\psi \in \Psi} \sup_{\phi \in \text{supp } \psi \circ L} \left\{ \inf_{x \in X} (f(x) + \phi(x)) - g_{\Psi}^*(\psi) \right\} \\ &\geq \sup_{\psi \in \Psi, \phi \in \text{supp } \psi \circ L} -f_{\Phi}^*(-\phi) - g_{\Psi}^*(\psi). \end{aligned} \quad (4.9)$$

Clearly, the Φ -convexity of $\psi \circ L$ is more general than the condition $\psi \circ L \in \Phi$. However, assuming the latter, we obtain, from (DCP) i.e.

$$\sup_{\psi \in \Psi} -\beta^c(0, \psi) = \sup_{\psi \in \Psi} -f_{\Phi}^*(-\psi \circ L) - g_{\Psi}^*(\psi).$$

If $X = Y$, $\Phi = \Psi$ are symmetric ($-\phi \in \Phi$ for all $\phi \in \Phi$), and $L : X \rightarrow X$ is such that $\psi \circ L \in \Phi$, for any $\psi \in \Psi$ then

$$\sup_{x \in X} \{-f(x) - \psi(Lx)\} = f_{\Psi}^*(-\psi \circ L),$$

and the conjugate dual (DCP) takes the form

$$\sup_{\psi \in \Psi} -\beta^c(0, \psi) = \sup_{\psi \in \Psi} -f_{\Psi}^*(-\psi \circ L) - g_{\Psi}^*(\psi). \quad (4.10)$$

If L is the identity operator, then (CP) becomes the minimization problem

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

for which the conjugate dual coincides with (4.10) has been discussed in [12, 19].

4.1.1 Examples of Conjugate dual for specific classes Φ, Ψ

Now, we discuss the conjugate dual problem (DCP) when Φ, Ψ are the sets of linear and quadratic functions i.e. Φ_{conv}, Ψ_{conv} and $\Phi_{lsc}^a, \Psi_{lsc}^b$ given by (3.5) and (3.10), respectively.

Example 8. Let X and Y be Banach spaces with the topological duals X^*, Y^* and their bilinear forms $\langle \cdot, \cdot \rangle_X, \langle \cdot, \cdot \rangle_Y$, respectively. Let $L : X \rightarrow Y$ be a linear continuous operator with the conjugate $L^* : Y^* \rightarrow X^*$. We define the sets Φ and Ψ of elementary functions as follows.

- (i) For the case of affine elementary functions, (cf. (3.5) above),

$$\begin{aligned}\Phi_{conv} &:= \{ \phi : X \rightarrow \mathbb{R} \mid \phi(x) = \langle u, x \rangle_X + c, u \in X^*, c \in \mathbb{R} \}, \\ \Psi_{conv} &:= \{ \psi : Y \rightarrow \mathbb{R} \mid \psi(y) = \langle v, y \rangle_Y + d, v \in Y^*, d \in \mathbb{R} \}.\end{aligned}$$

We have $\psi(Lx + y) - \psi(Lx) = \psi(y)$ which leads to the coupling function as defined in the convex case [15]. We have

$$\begin{aligned}\sup_{x \in X} \{ -f(x) - \psi(Lx) \} &= \sup_{x \in X} \{ -f(x) - \langle v, Lx \rangle_Y \} - d \\ &= \sup_{x \in X} \{ -f(x) - \langle L^*v, x \rangle_X \} - d.\end{aligned}$$

Since $L^*v \in X^*$, we have $-\psi \circ L \in \Phi_{conv}$ so

$$\sup_{x \in X} \{ -f(x) - \langle L^*v, x \rangle_X - d \} = f_{\Phi_{conv}}^*(-\psi \circ L),$$

and the dual problem

$$\sup_{\psi \in \Psi_{conv}} -\beta^c(0, \psi) = \sup_{\psi \in \Psi_{conv}} -f_{\Phi_{conv}}^*(-\psi \circ L) - g_{\Psi_{conv}}^*(\psi). \quad (4.11)$$

Note that the condition $\psi \circ L \in \Phi_{conv}$ is satisfied thanks to the form of Φ_{conv} , Ψ_{conv} , and (4.11) coincides with the classical Fenchel dual investigated e.g. in [16, Chapter 1, Section 2].

- (ii) In the case of quadratic elementary functions ($a, b \in \mathbb{R}$, cf. formula (3.10) above),

$$\Phi_{lsc}^a =: \{ \phi : X \rightarrow \mathbb{R} \mid \phi(x) = -a \|x\|_X^2 + \langle u, x \rangle_X + c, u \in X^*, c \in \mathbb{R} \}, \quad (4.12)$$

$$\Psi_{lsc}^b =: \{ \psi : Y \rightarrow \mathbb{R} \mid \psi(y) = -b \|y\|_Y^2 + \langle v, y \rangle_Y + d, v \in Y^*, d \in \mathbb{R} \}, \quad (4.13)$$

one cannot express $\psi \circ L$ as a function in Φ_{lsc}^a . However, we have

$$\begin{aligned} \psi(Lx) &= -b \|Lx\|_Y^2 + \langle v, Lx \rangle_Y + d \\ &= -b \|x\|_L^2 + \langle L^*v, x \rangle_X + d, \end{aligned}$$

with $\|x\|_L = \langle L^*Lx, x \rangle_X$, [6]. Thus, the dual problem (DCP) takes the form

$$\begin{aligned} \sup_{\psi \in \Psi_{lsc}^b} -\beta^c(0, \psi) &= \sup_{\psi \in \Psi_{lsc}^b} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi_{lsc}^b}^*(\psi) \\ &= \sup_{\psi \in \Psi_{lsc}^b} \inf_{x \in X} f(x) - b \|x\|_L^2 + \langle L^*v, x \rangle_X + d - g_{\Psi_{lsc}^b}^*(\psi) \\ &= \sup_{\psi \in \Psi_{lsc}^b} - (f - b \|\cdot\|_L^2)_{\Phi_{lsc}^a}^*(-\langle L^*v, \cdot \rangle_X - d) - g_{\Psi_{lsc}^b}^*(\psi). \end{aligned} \quad (4.14)$$

- (iii) In the construction of dual problem (4.4), the roles of Φ and Ψ are

not symmetric. We can see this by considering the following pair of elementary functions,

$$\begin{aligned}\Phi_{lsc}^a &:= \{ \phi : X \rightarrow \mathbb{R} \mid \phi(x) = -a \|x\|_X^2 + \langle u, x \rangle_X + c, u \in X^*, c \in \mathbb{R} \}, \\ \Psi_{conv} &:= \{ \psi : Y \rightarrow \mathbb{R} \mid \psi(y) = \langle v, y \rangle_Y + d, v \in Y^*, d \in \mathbb{R} \}.\end{aligned}$$

In this case, since $\psi \in \Psi_{conv}$ is an affine function we have

$$-\psi \circ L(\cdot) = -\langle v, L\cdot \rangle_Y - d = -\langle L^*v, \cdot \rangle_X - d := \phi(\cdot) \in \Phi_{lsc}^a,$$

and the dual problem has the same form as (4.11)

$$\sup_{\psi \in \Psi_{conv}} -\beta^c(0, \psi) = \sup_{\psi \in \Psi_{conv}} -f_{\Phi_{lsc}^a}^*(-\psi \circ L) - g_{\Psi_{conv}}^*(\psi). \quad (4.15)$$

However, when reversing the roles of Φ and Ψ i.e. by taking

$$\begin{aligned}\Phi_{conv} &:= \{ \phi : X \rightarrow \mathbb{R} \mid \phi(x) = \langle u, x \rangle_X + c, u \in X^*, c \in \mathbb{R} \}, \\ \Psi_{lsc}^b &:= \{ \psi : Y \rightarrow \mathbb{R} \mid \psi(y) = -b \|y\|_Y^2 + \langle v, y \rangle_Y + d, v \in Y^*, b, d \in \mathbb{R} \}.\end{aligned}$$

We cannot write the dual problem in the form of (4.11), but it is possible to write, for any $\psi \in \Psi_{lsc}^b$,

$$\psi(Lx) = -b \|Lx\|_Y^2 + \langle v, Lx \rangle_Y + d = \psi_1(x) + \psi_2(x),$$

where $\psi_1(x) = -b \|Lx\|_Y^2$, $\psi_2(x) = \langle L^*v, x \rangle_X + d$ and the dual problem takes the form

$$\sup_{\substack{\psi \in \Psi_{lsc}^b \\ \psi \circ L = \psi_1 + \psi_2}} -\beta^c(0, \psi) = \sup_{\substack{\psi \in \Psi_{lsc}^b \\ \psi \circ L = \psi_1 + \psi_2}} -(f + \psi_1)_{\Phi_{conv}}^*(-\psi_2) - g_{\Psi_{lsc}^b}^*(\psi). \quad (4.16)$$

Unlike the dual problems, considered in [12, 19] where $L = Id$, the ap-

pearance of general operators L in the minimization problem (CP) makes it more difficult to present the dual problem without imposing additional assumptions.

Remark 2. It is clear from the formula (4.5) and the above examples that the presence of constants in the definition of elementary functions is inessential in the c -conjugate of β . Thus, in the sequel, we consider classes Φ and Ψ of functions in which constants are omitted. For the general consideration related to this fact, see [82, formula 1.4.4].

4.2 Zero Duality Gap for Conjugate Dual

Having defined the dual problem (DCP), now, we discuss conditions for weak duality and zero duality gap. As in (CP), we assume that the sets Φ and Ψ of elementary functions are defined on X and Y , respectively, and both contain zeros, i.e., $0 \in \Phi, 0 \in \Psi$.

Let X be a nonempty set and let Y be a vector space. Let $L : X \rightarrow Y$ be a mapping from X to Y . The dual problem (DCP) can be equivalently rewritten in the form

$$val(DCP) := \sup_{\psi \in \Psi} -\beta^c(0, \psi) = - \inf_{\psi \in \Psi} \{ (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) \}. \quad (4.17)$$

For the problem (CP), we have

$$val(CP) := \inf_{x \in X} f(x) + g(Lx) = - (f + g \circ L)_\Phi^*(0). \quad (4.18)$$

Weak duality, i.e. the inequality $val(CP) \geq val(DCP)$ means that

$$(f + g \circ L)_\Phi^*(0) \leq \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi). \quad (4.19)$$

In fact, a more general inequality holds true.

Theorem 2. *For any $\phi \in \Phi$, it holds*

$$(f + g \circ L)_{\Phi}^*(\phi) \leq \inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi).$$

Proof. For any $\psi \in \Psi$ and $\phi \in \Phi$, we have

$$\begin{aligned} (f + g \circ L)_{\Phi}^*(\phi) &= \sup_{x \in X} \phi(x) - f(x) - g(Lx) \\ &= \sup_{x \in X} \phi(x) - f(x) - g(Lx) + \psi(Lx) - \psi(Lx) \\ &\leq \sup_{x \in X} \{\phi(x) - f(x) - \psi(Lx)\} + \sup_{x \in X} \{\psi(Lx) - g(Lx)\} \\ &\leq (f + \psi \circ L)_{\Phi}^*(\phi) + \sup_{y \in L(X)} \{\psi(y) - g(y)\} \\ &\leq (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi). \end{aligned}$$

Since $(f + g \circ L)_{\Phi}^*(\phi)$ is a lower bound of $(f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi)$ for arbitrary $\psi \in \Psi$, we get

$$(f + g \circ L)_{\Phi}^*(\phi) \leq \inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi),$$

which completes the proof. $\square$

By taking $\phi = 0$ in the above theorem, we obtain the weak duality (4.19) between (CP) and (DCP). Note that no additional assumptions are imposed on the functions $\psi \circ L$ and ϕ . However, when zero duality gap is considered, we need to assume a relationship between Φ and Ψ .

The following theorem provides sufficient and necessary conditions for zero duality gap for conjugate duality of problem (CP) and (DCP).

Theorem 3. *Let X be a nonempty set and Y be a vector space. Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$, and $L : X \rightarrow Y$ be a mapping. Assume that $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Suppose $0 \in \Phi$ and $0 \in \Psi$. The following are equivalent.*

- (i) For every $\varepsilon > 0$, there exist $x_\varepsilon \in X, \psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that
 $(f + \psi_\varepsilon \circ L)_\Phi^*(0) \leq -f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) + \varepsilon.$
- (ii) $-val(CP) = (f + g \circ L)_\Phi^*(0) = \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) = -val(DCP) < +\infty.$

Proof. (i) $\Rightarrow$ (ii): Thanks to Theorem 2, we only need to prove that

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) \leq (f + g \circ L)_\Phi^*(0) < +\infty.$$

Let $\varepsilon > 0$. From (i), there exist $x_\varepsilon \in X$ and $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) \leq -f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) + \varepsilon.$$

From the definition of infimum, we have

$$\begin{aligned} \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) &\leq (f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) \\ &\leq -f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) + g_\Psi^*(\psi_\varepsilon) + \varepsilon. \end{aligned} \quad (4.20)$$

By using inequality (3.9) applied to g and taking into account $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$, we obtain

$$\begin{aligned} -f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) + g_\Psi^*(\psi_\varepsilon) + \varepsilon &\leq -f(x_\varepsilon) - g(Lx_\varepsilon) + 2\varepsilon \\ &\leq (f + g \circ L)_\Phi^*(0) + 2\varepsilon. \end{aligned} \quad (4.21)$$

Finally, by using (4.20) and (4.21),

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) \leq (f + g \circ L)_\Phi^*(0) + 2\varepsilon.$$

As both sides of the above inequality do not depend on x_ε or ψ_ε , we can let

$\varepsilon \rightarrow 0$ and obtain zero duality gap

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(0) + g_{\Psi}^*(\psi) \leq (f + g \circ L)_{\Phi}^*(0).$$

From (4.20) and (4.21), we have

$$(f + g \circ L)_{\Phi}^*(0) \leq -f(x_{\varepsilon}) - g(Lx_{\varepsilon}) + 2\varepsilon < +\infty,$$

from the assumption of f and g . Thus $(f + g \circ L)_{\Phi}^*(0) < +\infty$.

(ii) $\Rightarrow$ (i): From the definition of supremum, for every $\varepsilon > 0$ there exists an $x_{\varepsilon} \in X$ such that

$$(f + g \circ L)_{\Phi}^*(0) \leq -f(x_{\varepsilon}) - g(Lx_{\varepsilon}) + \varepsilon/2.$$

Also, there exists $\psi_{\varepsilon} \in \Psi$ such that

$$(f + \psi_{\varepsilon} \circ L)_{\Phi}^*(0) + g_{\Psi}^*(\psi_{\varepsilon}) - \varepsilon/2 < \inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(0) + g_{\Psi}^*(\psi).$$

Zero duality gap gives us

$$(f + \psi_{\varepsilon} \circ L)_{\Phi}^*(0) + g_{\Psi}^*(\psi_{\varepsilon}) - \varepsilon/2 \leq -f(x_{\varepsilon}) - g(Lx_{\varepsilon}) + \varepsilon/2.$$

After rearranging both sides, we get

$$[(f + \psi_{\varepsilon} \circ L)_{\Phi}^*(0) + f(x_{\varepsilon}) + \psi_{\varepsilon}(Lx_{\varepsilon})] + [g_{\Psi}^*(\psi_{\varepsilon}) + g(Lx_{\varepsilon}) - \psi_{\varepsilon}(Lx_{\varepsilon})] < \varepsilon.$$

By the definition of conjugate function, each of the two terms is non-negative, and so each of them has to be smaller than ε . Thus

$$\begin{aligned} (f + \psi_{\varepsilon} \circ L)_{\Phi}^*(0) &< -f(x_{\varepsilon}) - \psi_{\varepsilon}(Lx_{\varepsilon}) + \varepsilon \\ g_{\Psi}^*(\psi_{\varepsilon}) + g(Lx_{\varepsilon}) - \psi_{\varepsilon}(Lx_{\varepsilon}) &< \varepsilon, \end{aligned}$$

which implies that $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$. Hence, (i) holds. $\square$

Notice that when f is improper, then condition (i) in Theorem 3 fails to hold as we have

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) = \sup_{x \in X} -f(x) - \psi_\varepsilon(Lx) = \sup_{x \in X} -\infty > -f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) + \varepsilon = -\infty$$

Remark 3. Observe that, in some cases, we cannot find $\phi \in \Phi, \psi \in \Psi$ such that $\phi + \psi = 0$ (this latter condition is used in [19, Theorem 3.5]). In the case $X = Y, \Phi = \Psi, L$ is identity operator and Φ is symmetric (for any $\psi \in \Phi, -\psi \in \Phi$), the problem (CP) takes the form

$$\inf_{x \in X} f(x) + g(x).$$

For this problem, Theorem (3)-(i) means that, for every $\varepsilon > 0$, there exist $x_\varepsilon \in X, \psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(x_\varepsilon)$ such that

$$(f + \psi_\varepsilon)_\Phi^*(0) \leq -f(x_\varepsilon) - \psi_\varepsilon(x_\varepsilon) + \varepsilon,$$

which implies $-\psi_\varepsilon \in \partial_{\varepsilon, \Phi} f(x_\varepsilon)$. This means that $0 \in \bigcap_{\varepsilon > 0} \partial_{\varepsilon, \Phi}(f + g)(X)$, which reduces to the condition (i) used in [19, Theorem 3.5] for proving zero duality gap. However, it is worth mentioning that [19, Theorem 3.5] refers to the minimization of the sum of finite number of functions. In the present work, we only focus on problem (CP), while it is possible to generalize our result to the sum of finite number of composite functions.

Condition (i) of Theorem 3, could be replaced by stronger conditions which are easier to be checked, when we introduce the following assumption: for given $\phi \in \Phi, \psi \in \Psi$,

$$\phi - \psi \circ L \in \Phi. \tag{4.22}$$

Note that condition (4.23) below will be used in the sequel to check zero

duality gap.

Remark 4. When Φ is a convex cone, then (4.22) is satisfied when $-\psi \circ L \in \text{rec } \Phi$ where $\text{rec } \Phi = \{\varphi \in F(X) : \varphi + \Phi \subset \Phi\}$ is the recession cone of Φ and $F(X)$ is a linear space of all functions defined on X .

Theorem 4. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ be such that $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Let $L : X \rightarrow Y$ be a mapping from X into Y . Suppose that $0 \in \Phi$ and $0 \in \Psi$. Consider the following conditions.*

(i) *For every $\varepsilon > 0$, there exist $x_\varepsilon \in X$, $\phi_\varepsilon \in \partial_{\varepsilon, \Phi} f(x_\varepsilon)$, $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that*

$$\begin{cases} \phi_\varepsilon(z) + \psi_\varepsilon(Lz) \geq -\varepsilon & \forall z \in X, \\ \phi_\varepsilon(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) \leq \varepsilon. \end{cases} \quad (4.23)$$

(ii) $-\text{val}(CP) = (f + g \circ L)_\Phi^*(0) = \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) = -\text{val}(DCP) < +\infty$.

We have (i) $\Rightarrow$ (ii). If, for every $\varepsilon > 0$, there exists $\psi_\varepsilon \in \Psi$ such that

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) - \varepsilon/2 < \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi),$$

and $-\psi_\varepsilon \circ L \in \Phi$, then (i) $\Leftrightarrow$ (ii).

Proof. (i) $\Rightarrow$ (ii). For any $\psi \in \Psi$

$$\inf_{\psi \in \Psi} ((f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi)) \leq (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi).$$

By using inequality (3.9), for $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$, $\phi_\varepsilon \in \partial_{\varepsilon, \Phi} f(x_\varepsilon)$, we have

$$\begin{aligned}
(f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) &\leq (f + \psi_\varepsilon \circ L)_\Phi^*(0) + \psi_\varepsilon(Lx_\varepsilon) - g(Lx_\varepsilon) + \varepsilon \\
&= \sup_{x \in X} \{-\psi_\varepsilon(Lx) - f(x)\} + \psi_\varepsilon(Lx_\varepsilon) - g(Lx_\varepsilon) + \varepsilon \\
&\leq f_\Phi^*(\phi_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) - g(Lx_\varepsilon) + 2\varepsilon \\
&\leq \phi_\varepsilon(x_\varepsilon) - f(x_\varepsilon) + \varepsilon + \psi_\varepsilon(Lx_\varepsilon) - g(Lx_\varepsilon) + 2\varepsilon \\
&\leq 3\varepsilon - f(x_\varepsilon) - g(Lx_\varepsilon) \\
&\leq 3\varepsilon + (f + g \circ L)_\Phi^*(0),
\end{aligned}$$

where in the third estimation, we use (4.23) to obtain $f_\Phi^*(\phi_\varepsilon)$.

By letting $\varepsilon \rightarrow 0$ we obtain zero duality gap. We can use the same argument as in Theorem 3 to prove that

$$(f + g \circ L)_\Phi^*(0) < +\infty.$$

For (ii) $\Rightarrow$ (i): We have

$$\inf_{\psi \in \Psi} ((f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi)) = (f + g \circ L)_\Phi^*(0) < +\infty,$$

so for every $\varepsilon > 0$, there exists $\psi_\varepsilon \in \Psi$ such that

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) - \varepsilon/2 < (f + g \circ L)_\Phi^*(0),$$

and $-\psi_\varepsilon \circ L \in \Phi$. Moreover, there exists an $x_\varepsilon \in X$ such that

$$(f + g \circ L)_\Phi^*(0) < \varepsilon/2 - f(x_\varepsilon) - g(Lx_\varepsilon).$$

Combining the two inequalities gives us

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) + f(x_\varepsilon) + g(Lx_\varepsilon) < \varepsilon.$$

By adding and subtracting $\psi_\varepsilon(Lx_\varepsilon)$, we get

$$\left[(f + \psi_\varepsilon \circ L)_\Phi^*(0) + f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon)\right] + [g_\Psi^*(\psi_\varepsilon) + g(Lx_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon)] < \varepsilon.$$

Since each term is nonnegative, we obtain

$$\begin{aligned} (f + \psi_\varepsilon \circ L)_\Phi^*(0) + f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) &\leq \varepsilon, \\ g_\Psi^*(\psi_\varepsilon) + g(Lx_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) &\leq \varepsilon. \end{aligned}$$

Since $-\psi_\varepsilon \circ L \in \Phi$, and thus $(f + \psi_\varepsilon \circ L)_\Phi^*(0) = f_\Phi^*(-\psi_\varepsilon \circ L)$. By the above inequalities, $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ and

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) = f_\Phi^*(-\psi_\varepsilon \circ L) + f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) \leq \varepsilon,$$

which is equivalent to $-\psi_\varepsilon \circ L \in \partial_{\varepsilon, \Phi} f(x_\varepsilon)$. Hence, (i) is proved. $\square$

The condition $-\psi_\varepsilon \circ L \in \Phi$ is equivalent to (4.22) since $\phi \in \Phi$ and Φ is closed under addition. Moreover, to obtain the equivalence in Theorem 4, we just need to assume that $-\psi_\varepsilon \circ L \in \Phi$ because the other condition is the property of infimum.

We close this section with existence of solution to problem (CP) and (DCP) in the spirit of [19, Theorem 3.6].

Theorem 5. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ be such that $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$ and let $L : X \rightarrow Y$ be a mapping from X into Y . Suppose that $0 \in \Phi$, $0 \in \Psi$ and $x \in \text{dom } g \cap L(\text{dom } f)$. Consider the following conditions.*

(i) *For all $\varepsilon > 0$, there exist $\phi_\varepsilon \in \partial_{\varepsilon, \Phi} f(x)$, $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx)$ such that*

$$\begin{cases} \phi_\varepsilon(z) + \psi_\varepsilon(Lz) \geq -\varepsilon & \text{for all } z \in X, \\ \phi_\varepsilon(x) + \psi_\varepsilon(Lx) \leq \varepsilon. \end{cases}$$

$$(ii) \quad -val(CP) = (f + g \circ L)_\Phi^*(0) = \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) = \\ val(DCP) < +\infty \text{ and } x \text{ is an optimal solution to (CP)}.$$

We have (i) $\Rightarrow$ (ii). For every $\varepsilon > 0$, if there exists $\psi_\varepsilon \in \Psi$ such that

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) - \varepsilon/2 < \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi) \text{ and } -\psi_\varepsilon \circ L \in \Phi, \quad (4.24)$$

then the two statements are equivalent.

Proof. The lines of the proof coincide with the ones of Theorem 4. We only need to prove that x is an optimal solution to (CP). Now as (i) holds for all $\varepsilon > 0$, we have

$$\begin{aligned} \inf_{\psi \in \Psi} ((f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi)) &\leq (f + \psi_\varepsilon \circ L)_\Phi^*(0) + g_\Psi^*(\psi_\varepsilon) \\ &\leq 3\varepsilon - f(x) - g(Lx) \\ &\leq 3\varepsilon + (f + g \circ L)_\Phi^*(0). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ in the above inequality as it holds for all $\varepsilon > 0$. We obtain zero duality gap

$$\inf_{\psi \in \Psi} ((f + \psi \circ L)_\Phi^*(0) + g_\Psi^*(\psi)) = (f + g \circ L)_\Phi^*(0).$$

Notice that

$$\inf_{z \in X} f(z) + g(Lz) = -(f + g \circ L)_\Phi^*(0) \geq f(x) + g(Lx).$$

As $(f + g \circ L)_\Phi^*(0) < +\infty$, the primal problem (CP) is finite with x as an optimal solution. $\square$

Remark 5. Let us observe that the condition (i) of Theorem 4.3 and Theorem 4.4 can be replaced by the condition

$$\phi_\varepsilon(x) + \psi_\varepsilon(Lx) = 0, \quad \forall x \in X. \quad (4.25)$$

Indeed, if (4.25) holds, which means $-\psi_\varepsilon \circ L = \phi_\varepsilon \in \Phi$, then (4.23) and (4.22) are satisfied.

4.2.1 Zero duality gap for specific classes Φ , Ψ

Consider the classes of elementary functions Φ_{lsc}^a and Ψ_{lsc}^b defined in (4.12) and (4.13) for $a, b \geq 0$. We let $c = d = 0$ because they do not affect the dual problem and the condition for zero duality gap. The dual problem (4.14) takes the form

$$\sup_{\substack{\psi \in \Psi_{lsc}^b \\ \psi(\cdot) = -b\|\cdot\|_Y^2 + \langle v, \cdot \rangle_Y}} - (f - b\|L\cdot\|_Y^2)_{\Phi_{lsc}^a}^* (-\langle L^*v, \cdot \rangle) - g_{\Psi_{lsc}^b}^*(\psi). \quad (4.26)$$

We start with the following technical fact.

Proposition 8. *Let X and Y be Hilbert spaces with inner products, $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$, respectively. Let $L : X \rightarrow Y$ be a continuous linear operator from X to Y , and $f : X \rightarrow (-\infty, +\infty]$ be a weakly convex function with modulus a . Then $f(x) - b\|Lx\|_Y^2$ is weakly convex with modulus $a + b\|L\|_{\mathcal{L}(X,Y)}^2$.*

Proof. By Definition 6, a function $f : X \rightarrow (-\infty, +\infty]$ is weakly convex on X with modulus $a > 0$ if $f + a\|x\|_X^2$ is a convex function, or equivalently for any $x_1, x_2 \in X$ and $\lambda \in [0, 1]$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) + a\lambda(1 - \lambda)\|x_1 - x_2\|_X^2.$$

Now we analyze the weak convexity of the function $f(x) - b\|Lx\|_Y^2$, where $b \geq 0$. We have

$$\begin{aligned} \|L(\lambda x_1 + (1 - \lambda)x_2)\|_Y^2 &= \|\lambda Lx_1 + (1 - \lambda)Lx_2\|_Y^2 \\ &= \lambda^2\|Lx_1\|_Y^2 + (1 - \lambda)^2\|Lx_2\|_Y^2 + 2\lambda(1 - \lambda)\langle Lx_1, Lx_2 \rangle_Y \\ &= \lambda\|Lx_1\|_Y^2 + (1 - \lambda)\|Lx_2\|_Y^2 - \lambda(1 - \lambda)\|Lx_1 - Lx_2\|_Y^2. \end{aligned}$$

Denote $x_\lambda := \lambda x_1 + (1 - \lambda) x_2$ for $\lambda \in [0, 1]$. Therefore,

$$\begin{aligned} f(x_\lambda) - b \|Lx_\lambda\|_Y^2 &\leq \lambda (f(x_1) - b \|Lx_1\|_Y^2) + (1 - \lambda) (f(x_2) - b \|Lx_2\|_Y^2) \\ &\quad + b\lambda(1 - \lambda) \|Lx_1 - Lx_2\|_Y^2 + a\lambda(1 - \lambda) \|x_1 - x_2\|_X^2. \end{aligned}$$

As L is linear and continuous, we obtain

$$\begin{aligned} f(x_\lambda) - b \|Lx_\lambda\|_Y^2 &\leq \lambda (f(x_1) - b \|Lx_1\|_Y^2) + (1 - \lambda) (f(x_2) - b \|Lx_2\|_Y^2) \\ &\quad + (a + b \|L\|_{\mathcal{L}(X,Y)}^2) \lambda(1 - \lambda) \|x_1 - x_2\|_X^2, \end{aligned}$$

where $\|L\|_{\mathcal{L}(X,Y)} = \sup_{\|x\|_X \leq 1} \|Lx\|_Y$. This means $f(x) - b\|Lx\|_Y^2$ is weakly convex with modulus $a + b\|L\|_{\mathcal{L}(X,Y)}^2$. $\square$

The following corollary follows from Theorem 4.

Corollary 1. *Let X, Y be Hilbert spaces. Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ and $L : X \rightarrow Y$ be a continuous linear operator. Assume that $a \leq 0, b \in \mathbb{R}$, $c = d = 0$ in the definitions of the classes of elementary functions Φ_{lsc}^a and Ψ_{lsc}^b ((4.12) and (4.13)), respectively, and $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. The following are equivalent.*

- (i) *For every $\varepsilon > 0$, there exist $x_\varepsilon \in X, \psi_\varepsilon \in \partial_{\varepsilon, \Psi_{lsc}^b} g(Lx_\varepsilon)$, $\phi_\varepsilon \in \partial_{\varepsilon, \Phi_{lsc}^a} (f - b\|L\cdot\|_Y^2)(x_\varepsilon)$ such that*

$$\begin{cases} \phi_\varepsilon(x) + \psi_\varepsilon(Lx) \geq -b\|Lx\|_Y^2 - \varepsilon & \forall x \in X, \\ \phi_\varepsilon(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) \leq -b\|Lx_\varepsilon\|_Y^2 + \varepsilon, \end{cases}$$

where $\psi_\varepsilon(y) = -b\|y\|_Y^2 + \langle v_\varepsilon, y \rangle_Y$.

- (ii) *$(f + g \circ L)_{\Phi_{lsc}^a}^*(0) = \inf_{\psi \in \Psi_{lsc}^b} (f - b\|L\cdot\|_Y^2)_{\Phi_{lsc}^a}^*(-\langle L^*v, \cdot \rangle_X) + g_{\Psi_{lsc}^b}^*(\psi) < +\infty$, where $\psi(y) = -b\|y\|_Y^2 + \langle v, y \rangle_Y$.*

Proof. Let $\varepsilon > 0$. By (i), there exist $x_\varepsilon \in X, \psi_\varepsilon \in \partial_{\varepsilon, \Psi_{lsc}^b} g(Lx_\varepsilon), \phi_\varepsilon \in \partial_{\varepsilon, \Phi_{lsc}^a} (f - b\|L\cdot\|_Y^2)(x_\varepsilon)$ such that for all $x \in X$,

$$\phi_\varepsilon(x) + \psi_\varepsilon(Lx) = -a\|x\|_X^2 + \langle u_\varepsilon, x \rangle_X - b\|Lx\|_Y^2 + \langle v_\varepsilon, Lx \rangle_Y \geq -b\|Lx\|_Y^2 - \varepsilon,$$

where $u_\varepsilon \in X, v_\varepsilon \in Y$, i.e.,

$$-a\|x\|_X^2 + \langle u_\varepsilon + L^*v_\varepsilon, x \rangle_X + \varepsilon \geq 0, \quad \forall x \in X. \quad (4.27)$$

Hence, it must be $u_\varepsilon = -L^*v_\varepsilon$. Thus, for $\phi_\varepsilon \in \partial_{\varepsilon, \Phi_{lsc}^a} (f + b\|L\cdot\|_Y^2)(x_\varepsilon)$, we can write

$$\begin{aligned} f(x_\varepsilon) - b\|Lx_\varepsilon\|_Y^2 + (f - b\|L\cdot\|_Y^2)_{\Phi_{lsc}^a}^*(\phi_\varepsilon) &\leq \phi_\varepsilon(x_\varepsilon) + \varepsilon \\ (f + \psi \circ L)_{\Phi_{lsc}^a}^*(0) &\leq -f(x_\varepsilon) - \psi(Lx_\varepsilon) + 3\varepsilon, \end{aligned}$$

which is similar to the proof of Theorem 3, so the assertion follows from Theorem 3. $\square$

Observe that the above corollary does not hold for Φ_{lsc}^a with $a < 0$ due to formula (4.27).

We give simple examples illustrating Theorem 4.

Example 9.

- Let $f(x) = (x+1)^2, g(x) = 4x^2, L = Id$ and

$$\begin{aligned} \Phi &= \{\phi(x) = -ax^2 + bx, a \geq 0, b \in \mathbb{R}\}, \\ \Psi &= \{\psi(x) = cx, c \in \mathbb{R}\}, \end{aligned}$$

(cf. Remark 2).

The conjugates of f and g are given as

$$f_{\Phi}^*(\phi) = \frac{(b-2)^2}{4(a+1)} - 1, \quad a \geq 0, b \in \mathbb{R}.$$

$$g_{\Psi}^*(\psi) = \frac{c^2}{16}, \quad c \in \mathbb{R}.$$

The ε -subdifferentials of f and g at x_0 are given as

$$\partial_{\varepsilon, \Phi} f(x_0) = \left\{ \phi(x) = -ax^2 + bx : (a+1) \left(x_0 - \frac{b-2}{2(a+1)} \right)^2 \leq \varepsilon, a \geq 0, b \in \mathbb{R} \right\},$$

$$\partial_{\varepsilon, \Psi} g(x_0) = \left\{ \psi(x) = cx : \left(2x_0 - \frac{c}{4} \right)^2 \leq \varepsilon, c \in \mathbb{R} \right\}.$$

To verify condition (i) of Theorem 4, we find $x_{\varepsilon} \in \mathbb{R}, \psi_{\varepsilon} \in \partial_{\varepsilon, \Psi} g(x_{\varepsilon}), \phi_{\varepsilon} \in \partial_{\varepsilon, \Phi} f(x_{\varepsilon})$ such that

$$\begin{cases} \phi_{\varepsilon}(x) + \psi_{\varepsilon}(x) \geq -\varepsilon & \forall x \in \mathbb{R}, \\ \phi_{\varepsilon}(x_{\varepsilon}) + \psi_{\varepsilon}(x_{\varepsilon}) \leq \varepsilon. \end{cases}$$

Now consider the first inequality

$$\phi_{\varepsilon}(x) + \psi_{\varepsilon}(x) = -ax^2 + (b+c)x \geq -\varepsilon, \quad \forall x \in \mathbb{R},$$

which gives $a = 0, b+c = 0$, so the second inequality i.e. $\phi_{\varepsilon}(x_{\varepsilon}) + \psi_{\varepsilon}(x_{\varepsilon}) \leq \varepsilon$, holds for $x_{\varepsilon} \in \mathbb{R}$. Hence, Theorem 4-(i) holds.

Moreover, as $\psi_{\varepsilon} \in \partial_{\varepsilon, \Psi} g(x_{\varepsilon}), \phi_{\varepsilon} \in \partial_{\varepsilon, \Phi} f(x_{\varepsilon})$, we have

$$\begin{cases} \left(x_{\varepsilon} - \frac{b-2}{2} \right)^2 \leq \varepsilon, \\ \left(2x_{\varepsilon} + \frac{b}{4} \right)^2 \leq \varepsilon. \end{cases}$$

Let $\varepsilon \rightarrow 0$, we obtain $b = 2x_0 + 2, c = 8x_0$. Since $b+c = 0$, we get

$x_0 = -\frac{1}{5}$. With x_0, a, b, c calculated above we get

$$\inf_{x \in X} f(x) + g(x) = f(x_0) + g(x_0) = \frac{4}{5},$$

and

$$\sup_{\psi \in \Psi} -(f + \psi)_{\Phi}^*(0) - g_{\Psi}^*(\psi) = -f_{\Phi}^*(\phi_{\varepsilon}) - g_{\Psi}^*(\psi_{\varepsilon}) = 1 - \frac{1}{5} = \frac{4}{5}.$$

Hence, we achieve zero duality gap.

- Now we reverse the roles of Φ and Ψ i.e.

$$\begin{aligned}\Phi &= \{\phi(x) = cx, c \in \mathbb{R}\}, \\ \Psi &= \{\psi(x) = -ax^2 + bx, a \geq 0, b \in \mathbb{R}\}.\end{aligned}$$

Then the conjugates of f and g are

$$\begin{aligned}f_{\Phi}^*(\phi) &= \frac{(c-2)^2}{4} - 1, \\ g_{\Psi}^*(\psi) &= \frac{b^2}{4(a+4)}.\end{aligned}$$

The ε -subdifferentials of f and g are

$$\begin{aligned}\partial_{\varepsilon, \Phi} f(x_{\varepsilon}) &= \left\{ \phi(x) = cx : \left(x_{\varepsilon} - \frac{c-2}{2}\right)^2 \leq \varepsilon, c \in \mathbb{R} \right\}, \\ \partial_{\varepsilon, \Psi} g(x_{\varepsilon}) &= \left\{ \psi(x) = -ax^2 + bx : \left(x_{\varepsilon} - \frac{b}{2(a+4)}\right)^2 \leq \frac{\varepsilon}{a+4}, a \geq 0, b \in \mathbb{R} \right\}.\end{aligned}$$

We need to find $x_\varepsilon \in \mathbb{R}$, $\phi_\varepsilon \in \partial_{\varepsilon, \Phi} f(x_\varepsilon)$, $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(x_\varepsilon)$ such that

$$\begin{cases} \phi_\varepsilon(x) + \psi_\varepsilon(x) \geq -\varepsilon & \forall x \in \mathbb{R}, \\ \phi_\varepsilon(x_\varepsilon) + \psi_\varepsilon(x_\varepsilon) \leq \varepsilon. \end{cases}$$

The first inequality, for all $x \in \mathbb{R}$,

$$\phi_\varepsilon(x) + \psi_\varepsilon(x) = -ax^2 + (b+c)x \geq -\varepsilon,$$

also implies $a = 0, b + c = 0$, so the second inequality is automatically satisfied. Repeating the same steps as before, we achieve zero duality gap at value $4/5$ which is the same in the previous case. However, condition (4.22) does not hold for any $\psi \in \Psi$ and 0 , because there is no $\phi \in \Phi$ such that

$$\phi(x) = -\psi(x) = ax^2 - bx,$$

unless $a = 0$.

In some cases, if Theorem 4-(i) is not satisfied, we need to check for zero duality gap in another way. Let us consider the following example.

Example 10. Let $X = Y = \mathbb{R}$, $f(x, y) = 3x^2 + 2y^2$, $g(x) = -(x - 1)^2$, $L(x, y) = x - y$. The classes of elementary functions are defined as

$$\begin{aligned} \Phi &= \{ \phi(x, y) = -a(x^2 + y^2) + b_1x + b_2y, \ a \geq 0, b_1, b_2 \in \mathbb{R} \}, \\ \Psi &= \{ \psi(x) = -cx^2 + dx, \ c \geq 0, d \in \mathbb{R} \} \end{aligned}$$

(cf. Remark 2).

We have

$$\begin{aligned}
 f_{\Phi}^*(\phi) &= \frac{b_1^2}{4(a+3)} + \frac{b_2^2}{4(a+2)}, \\
 \partial_{\varepsilon, \Phi} f(x_{\varepsilon}, y_{\varepsilon}) &= \begin{cases} \phi(x, y) = -a(x^2 + y^2) + b_1x + b_2y, \\ \text{s.t. } (a+3)\left(x_{\varepsilon} - \frac{b_1}{2(a+3)}\right)^2 + (a+2)\left(y_{\varepsilon} - \frac{b_2}{2(a+2)}\right)^2 \leq \varepsilon, \end{cases} \\
 g_{\Psi}^*(\psi) &= \begin{cases} +\infty & 0 \leq c < 1 \text{ or } c = 1, d \neq 2, \\ 1 & c = 1, d = 2, \\ \frac{(d-2)^2}{4(c-1)} + 1 & c > 1, \end{cases} \\
 \partial_{\varepsilon, \Psi} g(x_{\varepsilon}) &= \begin{cases} \psi(x) = -cx^2 + dx & (c-1)\left(x_{\varepsilon} - \frac{d-2}{2(c-1)}\right)^2 \leq \varepsilon, c > 1, \\ \psi(x) = -x^2 + 2x & c = 1, d = 2, \forall x_{\varepsilon}. \end{cases}
 \end{aligned}$$

One option is to find, for $\varepsilon > 0$, $x_{\varepsilon}, y_{\varepsilon} \in \mathbb{R}$, $\phi_{\varepsilon} \in \partial_{\varepsilon, \Phi} f(x_{\varepsilon}, y_{\varepsilon})$, $\psi_{\varepsilon} \in \partial_{\varepsilon, \Psi} g(L(x_{\varepsilon}, y_{\varepsilon}))$ such that

$$\begin{cases} \phi_{\varepsilon}(x, y) + \psi_{\varepsilon}(L(x, y)) \geq -\varepsilon & \forall x, y \in \mathbb{R}, \\ \phi_{\varepsilon}(x_{\varepsilon}, y_{\varepsilon}) + \psi_{\varepsilon}(L(x_{\varepsilon}, y_{\varepsilon})) \leq \varepsilon. \end{cases} \quad (4.28)$$

The first inequality is

$$\begin{aligned}
 &\phi_{\varepsilon}(x, y) + \psi_{\varepsilon}(L(x, y)) \geq -\varepsilon \\
 &-a(x^2 + y^2) + b_1x + b_2y - c(x - y)^2 + d(x - y) + \varepsilon \geq 0 \\
 &-(a + c)(x^2 + y^2) + 2cxy + (b_1 + d)x + (b_2 - d)y + \varepsilon \geq 0,
 \end{aligned}$$

which gives us $a = c = 0$, $b_1 + d = b_2 - d = 0$, so the second inequality of (4.28) holds for any $x, y \in \mathbb{R}$. However, there are no element in $\partial_{\varepsilon, \Psi} g(L(x_{\varepsilon}, y_{\varepsilon}))$ such that $c = 0$, and we cannot applied condition (i) of Corollary 4 to determine zero duality gap.

One option is that instead of solving (4.28), we take $\psi_{\varepsilon} \in \partial_{\varepsilon, \Psi} g(L(x_{\varepsilon}, y_{\varepsilon}))$

and calculate $\partial_{\varepsilon, \Phi}(f + \psi_\varepsilon \circ L)(x_\varepsilon, y_\varepsilon)$. We consider the simplest case where $\psi_\varepsilon(x) = -x^2 + 2x$ and calculate

$$(f + \psi_\varepsilon \circ L)_\Phi^*(\phi) = -(a+2)A^2 - (a+1)B^2 - 2AB + (b_1-2)A + (b_2+2)B,$$

where $A = \frac{ab_1-2a+b_1-b_2-4}{2(a^2+3a+1)}$, $B = \frac{ab_2+2a-b_1+2b_2+6}{2(a^2+3a+1)}$, for $a \geq 0$. We examine whether $0 \in \partial_{\varepsilon, \Phi}(f + \psi_\varepsilon \circ L)(x_\varepsilon, y_\varepsilon)$ by checking the inequality

$$(f + \psi_\varepsilon \circ L)_\Phi^*(0) + (f + \psi_\varepsilon \circ L)(x_\varepsilon, y_\varepsilon) \leq \varepsilon.$$

Substituting $\phi = 0$ or $a = b_1 = b_2 = 0$, we have

$$2x_\varepsilon^2 + y_\varepsilon^2 + 2x_\varepsilon y_\varepsilon + 2(x_\varepsilon - y_\varepsilon) + 5 \leq \varepsilon.$$

This inequality has solution $x_\varepsilon, y_\varepsilon$ as the left hand side is a parabola with minimum value $-\varepsilon$. Thus, condition (i) of Theorem 3 is satisfied, and we have zero duality gap for the conjugate duality.

Theorem 3 is more general than Theorem 4 but it requires more calculations than the latter. Sometimes, it is more convenient to calculate ε -subdifferential of f and g than ε -subdifferential of $(f + \psi_\varepsilon \circ L)$ where $\psi_\varepsilon \in \partial_{\varepsilon, \Psi}g(Lx_\varepsilon)$.

4.2.2 Moreau Envelope

In connection with Corollary 1, we are going to construct conjugate dual problem for the following problem which can be viewed as *generalized Moreau envelope*

$$e_{\gamma f}(x_0) := \inf_{x \in X} f(x) + \frac{1}{2\gamma} \|x - x_0\|^2,$$

where X is Hilbert and $f : X \rightarrow (-\infty, +\infty]$ is proper Φ -convex. We recall the conjugate dual problem (DCP) where operator $L = Id$ is identity

$$\sup_{\psi \in \Psi} \left\{ -\sup_{x \in X} \{-\psi(x) - f(x)\} - g_{\Psi}^*(\psi) \right\} = \sup_{\psi \in \Psi} -(f + \psi)_{\Phi}^*(0) - g_{\Psi}^*(\psi).$$

Since g is lsc convex, we can consider two classes of elementary functions Ψ : affine and $\Phi_{lsc}^{\mathbb{R}}$. We stress that the constant in elementary functions can omitted when constructing conjugate dual as well as calculating the conjugate function (see Remark 2), so we will not write the constant in the class Ψ .

- $\Psi = \{\psi : X \rightarrow \mathbb{R} \mid \psi(x) = \langle y, x \rangle, y \in X\}$. We can identify $\psi \in \Psi$ with $y \in X$ i.e. $\Psi = X$. The Ψ -conjugate of g is

$$g_{\Psi}^*(\psi) = g_{\Psi}^*(y) = \frac{\gamma}{2}\|y\|^2 + \langle x_0, y \rangle.$$

The conjugate dual problem is

$$\sup_{y \in X} \left\{ -\sup_{x \in X} \{-\langle x - x_0, y \rangle - f(x)\} - \frac{\gamma}{2}\|y\|^2 \right\}.$$

If $\Phi = X$, then the conjugate dual takes the form

$$\sup_{y \in X} -f_{\Phi}^*(-y) - \langle x_0, y \rangle - \frac{\gamma}{2}\|y\|^2.$$

By Theorem 3, zero duality gap holds iff for every $\varepsilon > 0$, there exists $x_{\varepsilon}, y_{\varepsilon} \in X$ such that

$$y_{\varepsilon} \in \partial_{\Psi, \varepsilon} \left(\frac{1}{2\gamma} \|\cdot - x_0\|^2 \right)(x_{\varepsilon}) \quad \text{and} \quad 0 \in \partial_{\Phi, \varepsilon} (f + \langle y_{\varepsilon}, \cdot \rangle)(x_{\varepsilon}).$$

As we know the explicit form of Ψ , we solve for y_{ε} in the first term and obtain

$$y_{\varepsilon} = \frac{x_{\varepsilon} - x_0}{\gamma}, \quad f(x) - f(x_{\varepsilon}) \geq \langle y_{\varepsilon}, x - x_{\varepsilon} \rangle - \varepsilon,$$

for all $x \in X$. x_ε can be found by solving the second inequality above.

- $\Psi = \Phi_{lsc}^{\mathbb{R}} = \{\phi = (a, u) : \phi(x) = -a\|x\|^2 + \langle u, x \rangle \text{ where } a \in \mathbb{R}, u \in X\}$.
The Ψ -conjugate of g is

$$g_\Psi^*(\psi) = \begin{cases} (1 + 2\gamma a) \left[\frac{\gamma\|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right] & 2\gamma a > -1, \\ -\frac{1}{2\gamma}\|x_0\|^2 & a = -\frac{1}{2\gamma}, u = -\frac{x_0}{\gamma}, \\ +\infty & \text{otherwise.} \end{cases}$$

The conjugate dual problem is

$$\begin{aligned} & \sup_{\psi \in \Psi} \left\{ -\sup_{x \in X} \{-\psi(x) - f(x)\} - g_\Psi^*(\psi) \right\} \\ &= \sup_{\substack{(a,u) \in \Psi \\ 2\gamma a > -1}} \left\{ -\sup_{x \in X} \{a\|x\|^2 - \langle u, x \rangle - f(x)\} - (1 + 2\gamma a) \left[\frac{\gamma\|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right] \right\}. \end{aligned}$$

Again, if $\Phi = \Phi_{lsc}^{\mathbb{R}}$ then we can simplify the first term of the dual problem

$$\sup_{\substack{(a,u) \in \Psi \\ 2\gamma a > -1}} -f_{\Phi_{lsc}^{\mathbb{R}}}^*(-a, -u) - (1 + 2\gamma a) \left[\frac{\gamma\|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right].$$

Similarly, we have zero duality iff for every $\varepsilon > 0$, there exists $x_\varepsilon \in X, \psi_\varepsilon \in \Psi$ such that

$$\psi_\varepsilon \in \partial_{\Psi, \varepsilon} \left(\frac{1}{2\gamma} \|\cdot - x_0\|^2 \right)(x_\varepsilon) \quad \text{and} \quad 0 \in \partial_{\Phi, \varepsilon}(f + \psi_\varepsilon)(x_\varepsilon).$$

We solve for ψ_ε in the first term and obtain such that

$$\psi_\varepsilon = (a_\varepsilon, u_\varepsilon), \quad 2\gamma a_\varepsilon > -1, \quad u_\varepsilon - 2a_\varepsilon = \frac{x_\varepsilon - x_0}{\gamma}.$$

x_ε can be found by solving

$$\begin{aligned} f(x) - f(x_\varepsilon) &\geq -\psi_\varepsilon(x) - \psi_\varepsilon(x_\varepsilon) - \varepsilon \\ &= a_\varepsilon \|x - x_\varepsilon\|^2 + \langle 2a_\varepsilon x_\varepsilon - u_\varepsilon, x - x_\varepsilon \rangle - \varepsilon \\ &= a_\varepsilon \|x - x_\varepsilon\|^2 + \left\langle \frac{x_0 - x_\varepsilon}{\gamma}, x - x_\varepsilon \right\rangle - \varepsilon, \end{aligned}$$

for all $x \in X$.

- One specific case is $\Psi = \Phi_{lsc}^a = \{\phi : \phi(x) = -a\|x\|^2 + \langle u, x \rangle, u \in X\}$ where $a > 0$ is fixed, then the conjugate of g is The Ψ -conjugate of g is

$$g_\Psi^*(\psi) = (1 + 2\gamma a) \left[\frac{\gamma \|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right].$$

The conjugate dual problem is

$$\begin{aligned} &\sup_{\psi \in \Psi} \left\{ -\sup_{x \in X} \{-\psi(x) - f(x)\} - g_\Psi^*(\psi) \right\} \\ &= \sup_{u \in X} \left\{ -\sup_{x \in X} \{a\|x\|^2 - \langle u, x \rangle - f(x)\} - (1 + 2\gamma a) \left[\frac{\gamma \|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right] \right\}. \end{aligned}$$

If f is Φ_{lsc}^a -convex then we can simplify the first term of the dual problem

$$\sup_{u \in X} -f_{\Phi_{lsc}^a}^*(-u) - (1 + 2\gamma a) \left[\frac{\gamma \|u\|^2}{2} - a\|x_0\|^2 + \langle u, x_0 \rangle \right].$$

4.3 Strong Duality for Conjugate Dual

In this section, we investigate strong duality for the conjugate dual (DCP), i.e., we provide conditions for zero duality gap and the existence of solution to the dual problem (DCP); the respective conditions are expressed in terms of additivity of epigraphs of the conjugate functions.

Definition 8. The epigraph of a function $f : X \rightarrow (-\infty, +\infty]$ is a subset of $X \times \mathbb{R}$ defined as

$$\text{epi } f := \{(x, r) \in X \times \mathbb{R} : f(x) \leq r\}. \quad (4.29)$$

We have

$$\text{epi } f_{\Phi}^* := \{(\phi, r) \in \Phi \times \mathbb{R} : f_{\Phi}^*(\phi) \leq r\}. \quad (4.30)$$

In consequence, if $(\phi, r) \in \text{epi } f_{\Phi}^*$, then $\phi \in \text{dom } f_{\Phi}^*$. To investigate strong duality, together with (4.22), we need the following condition: for a given $\phi \in \Phi$ and $\psi \in \Psi$

$$\phi + \psi \circ L \in \Phi. \quad (4.31)$$

Remark 6. Note that (4.31) is satisfied when $\psi \circ L \in \text{lin } \Phi$ where $\text{lin } \Phi$ is the linear space generated by Φ .

Theorem 6. Let X be a nonempty set and Y be a vector space. Let $f : X \rightarrow (-\infty, +\infty]$ and $g : Y \rightarrow (-\infty, +\infty]$ be proper functions. Let Φ and Ψ be sets of elementary functions defined on X and Y , respectively. Let $L : X \rightarrow Y$ be a mapping from X to Y . Assume that $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$ and consider the following conditions.

- (i) Every $(\phi, c_{\phi}) \in \text{epi } (f + g \circ L)_{\Phi}^*$, can be expressed as $(\phi, c_{\phi}) = (\varphi + \psi \circ L, c_{\varphi} + c_{\psi})$, where $(\varphi, c_{\varphi}) \in \text{epi } f_{\Phi}^*$ and $(\psi, c_{\psi}) \in \text{epi } g_{\Psi}^*$.
- (ii) For any $\phi \in \Phi$, it holds $(f + g \circ L)_{\Phi}^*(\phi) = \inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi)$ and the infimum is attained.

We have (i) $\Rightarrow$ (ii). Moreover, for any $\phi \in \Phi$, if $\bar{\psi} \in \Psi$ is a solution to the problem

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi), \quad (4.32)$$

and conditions (4.31), (4.22) hold for $\bar{\psi}$ and ϕ , then (ii) $\Rightarrow$ (i).

Proof. (i) $\Rightarrow$ (ii): If $\phi \notin \text{dom } (f + g \circ L)_\Phi^*$, then $(f + g \circ L)_\Phi^*(\phi) = +\infty$ and (ii) holds. Consider the case $\phi \in \text{dom } (f + g \circ L)_\Phi^*$. By Theorem 2, for every $\phi \in \Phi$, we have

$$(f + g \circ L)_\Phi^*(\phi) \leq \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi). \quad (4.33)$$

We start by proving the opposite inequality. Since $(\phi, (f + g \circ L)_\Phi^*(\phi)) \in \text{epi } (f + g \circ L)_\Phi^*$, by (i), there exist $(\varphi, c_f) \in \text{epi } f_\Phi^*$ and $(\psi_0, c_g) \in \text{epi } g_\Psi^*$ such that

$$\begin{aligned} \phi(x) &= \varphi(x) + \psi_0(Lx), \quad \forall x \in X, \\ (f + g \circ L)_\Phi^*(\phi) &= c_f + c_g. \end{aligned}$$

Thus

$$\begin{aligned} (f + g \circ L)_\Phi^*(\phi) &= c_f + c_g \\ &\geq f_\Phi^*(\varphi) + g_\Psi^*(\psi_0) \\ &= \sup_{x \in X} \{\varphi(x) - f(x)\} + g_\Psi^*(\psi_0) \\ &= \sup_{x \in X} \{\phi(x) - \psi_0(Lx) - f(x)\} + g_\Psi^*(\psi_0) \\ &= (f + \psi_0 \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi_0) \\ &\geq \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi), \end{aligned}$$

and we have (ii). The infimum is attained from the above inequality, as

$$\begin{aligned} +\infty > c_f + c_g &= (f + g \circ L)_\Phi^*(\phi) \geq (f + \psi_0 \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi_0) \\ &\geq \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi) \\ &\geq (f + g \circ L)_\Phi^*(\phi), \end{aligned}$$

so $\psi_0 \in \Psi$ solves problem (4.32).

(ii) $\Rightarrow$ (i): From (ii), let $(\phi, c_f) \in \text{epi } f_\Phi^*$ and $\bar{\psi}$ be a solution to the problem (4.32) for $\phi \in \Phi$. As $(f + \bar{\psi} \circ L)_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}) < +\infty$, we can find $c_g \in \mathbb{R}$ such that $(\bar{\psi}, c_g) \in \text{epi } g_\Psi^*$. Moreover, from (4.31), $\phi + \bar{\psi} \circ L := \varphi \in \Phi$, we have

$$\begin{aligned} c_f + c_g &\geq f_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}) \\ &= \sup_{x \in X} \{ \varphi(x) - \bar{\psi}(Lx) - f(x) \} + g_\Psi^*(\bar{\psi}) \\ &= (f + \bar{\psi} \circ L)_\Phi^*(\varphi) + g_\Psi^*(\bar{\psi}) \\ &\geq (f + g \circ L)_\Phi^*(\varphi). \end{aligned}$$

This means $(\varphi, c_f + c_g) \in \text{epi } (f + g \circ L)_\Phi^*$.

On the other hand, let $(\phi, r) \in \text{epi } (f + g \circ L)_\Phi^*$. We want to prove that there exist $(\varphi, c_f) \in \text{epi } f_\Phi^*$ and $(\psi, c_g) \in \text{epi } g_\Psi^*$ such that

$$(\phi, r) = (\varphi + \psi \circ L, c_f + c_g).$$

From (ii), for $\phi \in \Phi$, there exists $\bar{\psi} \in \Psi$ such that $\bar{\psi}$ is the solution to the problem (4.32) i.e.

$$(f + g \circ L)_\Phi^*(\phi) = (f + \bar{\psi} \circ L)_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}).$$

Therefore, we have

$$\begin{aligned} r &\geq (f + g \circ L)_\Phi^*(\phi) \\ &\geq (f + \bar{\psi} \circ L)_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}) \\ r - g_\Psi^*(\bar{\psi}) &\geq \sup_{x \in X} \{ \phi(x) - \bar{\psi}(Lx) - f(x) \}. \end{aligned}$$

Condition (4.22) gives us $\varphi := \phi - \bar{\psi} \circ L \in \Phi$. Then

$$r - g_\Psi^*(\bar{\psi}) \geq \sup_{x \in X} \{ \varphi(x) - f(x) \} = f_\Phi^*(\varphi).$$

Therefore, $(\varphi, r - g_{\Psi}^*(\bar{\psi})) \in \text{epi } f_{\Phi}^*$ and $(\bar{\psi}, g_{\Psi}^*(\bar{\psi})) \in \text{epi } g_{\Psi}^*$, which means we can decompose (ϕ, r) into $(\varphi, r - g_{\Psi}^*(\bar{\psi}))$ and $(\bar{\psi}, g_{\Psi}^*(\bar{\psi}))$. Hence, the proof is finished. $\square$

Remark 7.

- Observe that condition (ii) of Theorem 5.2 implies the strong duality relationship between (CP) and (DCP), i.e. zero duality gap holds

$$-val(CP) = (f + g \circ L)_{\Phi}^*(0) = \inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(0) + g_{\Psi}^*(\psi) = -val(DCP),$$

and the dual problem (DCP) is solvable.

- The assumption (i) in Theorem 6 does not coincide with the following

$$\text{epi } (f + g \circ L)_{\Phi}^* = \text{epi } f_{\Phi}^* + \text{epi } (g \circ L)_{\Phi}^*,$$

(see also [56, Theorem 3.1]), because we consider the set

$B := \{(\psi \circ L, r) : (\psi, r) \in \text{epi } g_{\Psi}^*\}$ instead of $\text{epi } (g \circ L)_{\Phi}^*$, where $(\psi, r) \in \text{epi } g_{\Psi}^*$. Then assumption (i) of Theorem 6 is

$$\text{epi } (f + g \circ L)_{\Phi}^* = \text{epi } f_{\Phi}^* + B,$$

whenever $\phi + \psi \circ L \in \Phi$ holds true for any $\psi \in \text{dom } g_{\Psi}^*, \phi \in \text{dom } f_{\Phi}^*$. Examples of classes of elementary functions, where condition $\phi + \psi \circ L \in \Phi$ is satisfied, can be found in Example 8-3.

Corollary 2. *Let X be a nonempty set and Y be a vector space. Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ and $L : X \rightarrow Y$ be a mapping from X to Y . Let Φ, Ψ be sets of elementary functions on X and Y , respectively. Assume that $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Let*

$$(i) \text{ epi } (f + g \circ L)_{\Phi}^* = \text{epi } f_{\Phi}^* + \text{epi } (g \circ L)_{\Phi}^*,$$

- (ii) For $\phi \in \Phi$, $\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi) = (f + g \circ L)_{\Phi}^*(\phi)$ and the infimum is attained.

Consider the following conditions.

- (i) For any pair $(\phi, r) \in \text{epi } (g \circ L)_{\Phi}^*$, there exists $\psi \in \Psi$ such that $\phi \leq \psi \circ L$ and $g_{\Psi}^*(\psi) \leq r$.
- (ii) There exists $\psi \in \Psi$ such that $g_{\Psi}^*(\psi) \leq 0$ and $g \circ L \leq \psi \circ L$.
- (iii) For any pair $(\phi, r) \in \text{epi } (f + g \circ L)_{\Phi}^*$ and a given $\psi \in \Psi$, there exist $\phi_1, \phi_2 \in \Phi$ such that $\phi - \psi \circ L \geq \phi_1$ and $\psi \circ L \geq \phi_2$.
- (iv) $\Phi - \Phi \subset \Phi$, for any $\varepsilon_1, \varepsilon_2 > 0$ and a given $\psi \in \Psi$, there exist $x_{\varepsilon_2} \in X, \phi_{\varepsilon_2} \in \partial_{\varepsilon_2, \Phi}(g \circ L)(x_{\varepsilon_2})$ such that

$$\psi(Lx_{\varepsilon_2}) - \phi_{\varepsilon_2}(x_{\varepsilon_2}) + \varepsilon_1 > \sup_{x \in X} \psi(Lx) - \phi_{\varepsilon_2}(x).$$

If (a) or (b) hold, then (i) $\Rightarrow$ (ii). If (c) or (d) hold for $\bar{\psi} \in \Psi$ at which the infimum in (ii) is attained, then (ii) $\Rightarrow$ (i).

Proof. (i) $\Rightarrow$ (ii): For any $\phi \in \Phi$, $(\phi, (f + g \circ L)_{\Phi}^*(\phi)) \in \text{epi } (f + g \circ L)_{\Phi}^*$. There exist two pairs $(\phi_1, c_1) \in \text{epi } f_{\Phi}^*$ and $(\phi_2, c_2) \in \text{epi } (g \circ L)_{\Phi}^*$ such that $\phi = \phi_1 + \phi_2$ and $c_1 + c_2 = (f + g \circ L)_{\Phi}^*(\phi)$. Thanks to Theorem 2, for any $\phi \in \Phi$, we have

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi) \geq (f + g \circ L)_{\Phi}^*(\phi),$$

so we need to prove the reverse inequality. Consider

$$\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi) \leq (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi). \quad (4.34)$$

Assume (a). We can find $\psi_0 \in \Psi$ such that $\phi_2 \leq \psi_0 \circ L$ and $g_{\Psi}^*(\psi_0) \leq c_2$. Note that

$$\phi_1 = \phi - \phi_2 \geq \phi - \psi_0 \circ L.$$

Hence,

$$\begin{aligned}
\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi) &\leq (f + \psi_0 \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi_0) \\
&\leq \sup_{x \in X} \{\phi(x) - f(x) - \psi_0(Lx)\} + c_2 \\
&\leq \sup_{x \in X} \{\phi_1(x) - f(x)\} + c_2 \\
&= f_{\Phi}^*(\phi_1) + c_2 \\
&\leq c_1 + c_2 = (f + g \circ L)_{\Phi}^*(\phi). \tag{4.35}
\end{aligned}$$

Inequality (4.35) gives us

$$(f + g \circ L)_{\Phi}^*(\phi) \leq (f + \psi_0 \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi_0) \leq (f + g \circ L)_{\Phi}^*(\phi) < +\infty,$$

and the infimum in (ii) is attained at ψ_0 .

Now let (b) hold. There exists $\psi_0 \in \Psi$ such that $g \circ L \leq \psi_0 \circ L$ and $g_{\Psi}^*(\psi_0) \leq 0$. We have

$$\begin{aligned}
\inf_{\psi \in \Psi} (f + \psi \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi) &\leq (f + \psi_0 \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\psi_0) \\
&\leq \sup_{x \in X} \{\phi_1(x) - f(x)\} + \sup_{x \in X} \{\phi_2(x) - \psi_0(Lx)\} \\
&\leq f_{\Phi}^*(\phi_1) + \sup_{x \in X} \{\phi_2(x) - g(Lx)\} \\
&\leq f_{\Phi}^*(\phi_1) + (g \circ L)_{\Phi}^*(\phi_2) \\
&\leq c_1 + c_2 = (f + g \circ L)_{\Phi}^*(\phi), \tag{4.36}
\end{aligned}$$

where we have used $\phi = \phi_1 + \phi_2$ in the second inequality. The attainment of the infimum of (ii) at $\psi_0 \in \Psi$ follows in the same way as in the proof with condition (a).

(ii) $\Rightarrow$ (i): Let $(\phi_1, c_1) \in \text{epi } f_{\Phi}^*$, $(\phi_2, c_2) \in \text{epi } (g \circ L)_{\Phi}^*$ we have

$$c_1 + c_2 \geq f_{\Phi}^*(\phi_1) + (g \circ L)_{\Phi}^*(\phi_2) \geq (f + g \circ L)_{\Phi}^*(\phi_1 + \phi_2),$$

so $(\phi_1 + \phi_2, c_1 + c_2) \in \text{epi } (f + g \circ L)_\Phi^*$, which means

$$\text{epi } (f + g \circ L)_\Phi^* \supset \text{epi } f_\Phi^* + \text{epi } (g \circ L)_\Phi^*.$$

We want to prove the reverse inclusion. Take $(\phi, r) \in \text{epi } (f + g \circ L)_\Phi^*$. From (ii), there exists $\bar{\psi} \in \Psi$ which is a solution to the problem,

$$\begin{aligned} \inf_{\psi \in \Psi} (f + \psi \circ L)_\Phi^*(\phi) + g_\Psi^*(\psi) &= (f + \bar{\psi} \circ L)_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}) \\ &= (f + g \circ L)_\Phi^*(\phi) \leq r. \end{aligned} \quad (4.37)$$

We assume (c) holds. There exist $\phi_1, \phi_2 \in \Phi$ such that $\phi \geq \phi_1 + \bar{\psi} \circ L$, $\bar{\psi} \circ L \geq \phi_2$, so that $(g \circ L)_\Phi^*(\phi_2) \leq g_\Psi^*(\bar{\psi})$ and

$$(f + \bar{\psi} \circ L)_\Phi^*(\phi) = \sup_{x \in x} \phi(x) - \bar{\psi}(Lx) - f(x) \geq \sup_{x \in x} \phi_1(x) - f(x) = f_\Phi^*(\phi_1). \quad (4.38)$$

Combining this with (4.37) and (4.38), we obtain

$$f_\Phi^*(\phi_1) + (g \circ L)_\Phi^*(\phi_2) \leq (f + \bar{\psi} \circ L)_\Phi^*(\phi) + g_\Psi^*(\bar{\psi}) \leq r.$$

By taking $(\phi_2, (g \circ L)_\Phi^*(\phi_2)) \in \text{epi } (g \circ L)_\Phi^*$, we can make $(\phi_1, r - (g \circ L)_\Phi^*(\phi_2)) \in \text{epi } f_\Phi^*$. This means

$$(\phi_2, (g \circ L)_\Phi^*(\phi_2)) + (\phi_1, r - (g \circ L)_\Phi^*(\phi_2)) = (\phi, r).$$

Thus, we have

$$\text{epi } (f + g \circ L)_\Phi^* \subset \text{epi } f_\Phi^* + \text{epi } (g \circ L)_\Phi^*. \quad (4.39)$$

Now, let (d) hold. Let $\bar{\psi} \in \Psi$ be a solution to the infimum problem in (ii). For every $\varepsilon_1, \varepsilon_2 > 0$, there exist $x_{\varepsilon_2} \in X$, $\phi_{\varepsilon_2} \in \partial_{\varepsilon_2, \Phi}(g \circ L)(x_{\varepsilon_2})$ such

that

$$-\sup_{x \in X} \{\bar{\psi}(Lx) - \phi_{\varepsilon_2}(x)\} > -\bar{\psi}(Lx_{\varepsilon_2}) + \phi_{\varepsilon_2}(x_{\varepsilon_2}) - \varepsilon_1.$$

We have

$$\begin{aligned} (f + \bar{\psi} \circ L)_{\Phi}^*(\phi) &= \sup_{x \in X} \{\phi(x) - f(x) - \bar{\psi}(Lx)\} \\ &\geq \sup_{x \in X} \{\phi(x) - \phi_{\varepsilon_2}(x) - f(x)\} - \sup_{x \in X} \{\bar{\psi}(Lx) - \phi_{\varepsilon_2}(x)\}. \end{aligned} \quad (4.40)$$

As $\phi - \phi_{\varepsilon_2} := \varphi_{\varepsilon_2} \in \Phi$, we can write $\sup_{x \in X} \{\phi(x) - \phi_{\varepsilon_2}(x) - f(x)\} = f_{\Phi}^*(\varphi_{\varepsilon_2})$. Since $g_{\Psi}^*(\bar{\psi}) \geq \bar{\psi}(y) - g(y)$ for any $y \in Y$, we can set $y = Lx_{\varepsilon_2}$. Combining this with (4.40) gives us

$$\begin{aligned} (f + \bar{\psi} \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\bar{\psi}) &> f_{\Phi}^*(\varphi_{\varepsilon_2}) - \bar{\psi}(Lx_{\varepsilon_2}) + \phi_{\varepsilon_2}(x_{\varepsilon_2}) - \varepsilon_1 + \bar{\psi}(Lx_{\varepsilon_2}) - g(Lx_{\varepsilon_2}) \\ &> f_{\Phi}^*(\varphi_{\varepsilon_2}) + \phi_{\varepsilon_2}(x_{\varepsilon_2}) - g(Lx_{\varepsilon_2}) - \varepsilon_1 \\ &> f_{\Phi}^*(\varphi_{\varepsilon_2}) + (g \circ L)_{\Phi}^*(\phi_{\varepsilon_2}) - \varepsilon_1 - \varepsilon_2, \end{aligned}$$

as $\phi_{\varepsilon_2} \in \partial_{\varepsilon_2, \Phi} g(Lx_{\varepsilon_2})$. From (4.37), we have

$$\begin{aligned} r &> (f + \bar{\psi} \circ L)_{\Phi}^*(\phi) + g_{\Psi}^*(\bar{\psi}) \\ &> f_{\Phi}^*(\varphi_{\varepsilon_2}) + (g \circ L)_{\Phi}^*(\phi_{\varepsilon_2}) - \varepsilon_2 - \varepsilon_1. \end{aligned}$$

We obtain $(\varphi_{\varepsilon_2}, r + \varepsilon_1 + \varepsilon_2 - (g \circ L)_{\Phi}^*(\phi_{\varepsilon_2})) \in \text{epi } f_{\Phi}^*$ and $(\phi_{\varepsilon_2}, (g \circ L)_{\Phi}^*(\phi_{\varepsilon_2})) \in \text{epi } (g \circ L)_{\Phi}^*$. This means $\text{epi } (f + g \circ L)_{\Phi}^* \subset \text{epi } f_{\Phi}^* + \text{epi } (g \circ L)_{\Phi}^*$, so we have (4.39). $\square$

Remark 8.

- In Theorem 6 and Corollary 2, to obtain condition (ii), the assumptions mostly rely on g and the set of elementary functions Ψ . Thus, with appropriate choices of Ψ , we can guarantee the attainment of the infimum in (ii). The motivation comes from the fact that in the construction of

the conjugate dual, we perturb g but not f .

- If $0 \in \Phi$, then the condition $\Phi - \Phi \subset \Phi$ implies symmetry of the set Φ . Because we can take $0 - \phi \in \Phi$ for any $\phi \in \Phi$.
- It is true that $\text{epi } f^* = \text{supp } f$. To see this, let us take $(\phi, r) \in \text{epi } f^*$, then $\phi(x) - r \leq f(x)$ for all $x \in X$. As Φ is closed under addition of constant, we have $\phi - r \in \Phi$ and $\phi - r \in \text{supp } f$. Conversely, if $\phi \in \text{supp } f$ then $(\phi, 0) \in \text{epi } f^*$ (see [56]).
- Corollary 2-(i) is closed to the additivity of the support of the conjugate in [56, Theorem 5.1]

$$\text{supp}_\Phi(f + g \circ L) = \text{supp}_\Phi f + \text{supp}_\Phi(g \circ L),$$

while in general, [56, Proposition 2.2], we have

$$\text{supp}_\Phi(f + g \circ L) = \text{co}_\Phi(\text{supp}_\Phi f + \text{supp}_\Phi(g \circ L)),$$

where $\text{co}_\Phi A$ is the Φ -convex hull of $A \subset \Phi$ i.e.

$$\text{co}_\Phi A = \text{supp}_\Phi f_A, \quad \text{where } f_A(x) = \sup_{\phi \in A} \phi(x), \quad \forall x \in X.$$

- In the classical convex analysis, when X and Y are separated locally convex spaces. For lower semicontinuous proper convex functions f, g and a continuous linear mapping L , [16, Theorem 14.1], condition ensuring strong duality are expressed with the help of regularity condition $(RC)_5^\Phi$ as defined in [16, Chapter 4, Section 14] while in Theorem 6 and Corollary 2, we are using assumptions related to the additivity of epigraph of the conjugate. We refer the reader to Section 16 in [16] for the discussion of relationship between different regularity conditions ensuring strong duality.

We give an example of a convex problem satisfying the assumption (i) of Theorem 6.

Example 11. Let $X = \mathbb{R}^2$ and the functions $f(x, y) = x^2 + y^2, g(x) = 2x^2, L(x, y) = x + y$ and

$$\begin{aligned}\Phi_{isc}^{\geq} &= \{ \phi(x, y) = -a(x^2 + y^2) + b_1x + b_2y, a \geq 0, b_1, b_2 \in \mathbb{R} \}, \\ \Psi &= \{ \psi(x) = cx, c \in \mathbb{R} \}.\end{aligned}$$

Condition (4.31), $\phi + \psi \circ L \in \Phi$, always holds for all $\phi \in \Phi, \psi \in \Psi$ (see Example 8-3). Let us denote

$$\begin{aligned}\phi(x, y) &= -a(x^2 + y^2) + b_1x + b_2y, a \geq 0, \\ \varphi(x, y) &= -a_1(x^2 + y^2) + b_{11}x + b_{12}y, a_1 \geq 0, \\ \psi(x) &= cx.\end{aligned}$$

We want to calculate

$$\begin{aligned}(\phi, c_1) \in \text{epi } (f + g \circ L)_{\Phi}^* &\Leftrightarrow a \geq 0, \frac{4b_1b_2 - (3+a)(b_1^2 + b_2^2)}{4(4 - (3+a)^2)} \leq c_1, \\ (\varphi, c_2) \in \text{epi } f_{\Phi}^* &\Leftrightarrow a_1 \geq 0, \frac{b_{11}^2 + b_{12}^2}{4(a_1 + 1)} \leq c_2, \\ (\psi, c_3) \in \text{epi } g_{\Psi}^* &\Leftrightarrow \psi \in \Psi, \frac{c^2}{8} \leq c_3.\end{aligned}$$

Let us take $(\phi, c_1) \in \text{epi } (f + g \circ L)_{\Phi}^*$ then we want decompose $\phi = \varphi + \psi \circ L$ and $c_1 = c_2 + c_3$ where $(\varphi, c_2) \in \text{epi } f_{\Phi}^*$ and $(\psi, c_3) \in \text{epi } g_{\Psi}^*$. We have the following system

$$\begin{cases} a = a_1, \\ c + b_{11} = b_1, \\ c + b_{12} = b_2. \end{cases}$$

We can choose $c = c_3 = 0, a = a_1 = 0, b_1 = b_2$ so $b_1 = b_{11} = b_2 = b_{12}$

and $c_1 = c_2$. Taking $b_1 = 1 = c_1$ and assumption (i) of Theorem 6 is satisfied. Thus, we have strong duality for conjugate duality. Notice that if Φ is composed of affine functions only, we arrive at the same conclusion.

4.4 Lagrange Duality

4.4.1 Construction of Lagrangian Primal-Dual Problems

Of equal importance beside conjugate dual is Lagrange dual. Results from [12] give a connection between the Conjugate Duality and Lagrange Duality. For other construction of Lagrangian dual proposed recently, see e.g. [20] and the references therein.

In this Subsection, we point out the relationship between conjugate dual and Lagrange dual problems. We give new results for zero duality gap for Lagrange dual of composite problems. To construct Lagrangian dual, we introduce the Lagrangian function $\mathcal{L} : X \times \Psi \rightarrow (-\infty, +\infty]$ as follow

$$\mathcal{L}(x, \psi) := f(x) + \psi(Lx) - g_{\Psi}^*(\psi), \quad (4.41)$$

where Ψ is a set of elementary functions defined on Y . In the case Ψ is a convex set then $\mathcal{L}(x, \psi)$ is concave with respect to $\psi \in \Psi$ thanks to the property of conjugate function. The Lagrangian dual is

$$val(LD) := \sup_{\psi \in \Psi} \inf_{x \in X} \mathcal{L}(x, \psi) = \sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi), \quad (LD)$$

and the corresponding Lagrangian primal

$$val(LP) := \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = \inf_{x \in X} \sup_{\psi \in \Psi} f(x) + \psi(Lx) - g_{\Psi}^*(\psi). \quad (LP)$$

Notice that Lagrange dual (LD) is actually the conjugate dual (DCP) by expressing the conjugate of function f . Let us state the weak duality for

Lagrangian duality.

Theorem 7. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ and $L : X \rightarrow Y$ be a mapping. Let Φ, Ψ be sets of elementary functions defined on X and Y , respectively. It holds*

$$\sup_{\psi \in \Psi} \inf_{x \in X} \mathcal{L}(x, \psi) \leq \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi).$$

This result holds true for any functions not necessarily of the form (4.41). First we notice that Lagrangian primal (LP) is not equivalent to the composite problem (CP) as

$$\begin{aligned} \inf_{x \in X} \sup_{\psi \in \Psi} f(x) + \psi(Lx) - g_{\Psi}^*(\psi) &= \inf_{x \in X} f(x) + \sup_{\psi \in \Psi} \psi(Lx) - g_{\Psi}^*(\psi) \\ &= \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx) \\ &\leq \inf_{x \in X} f(x) + g(Lx). \end{aligned}$$

Even though the Lagrange dual and conjugate dual are the same, the primal problems are different. Since our main focus is (CP), we discuss the assumptions which make these two primal problems equivalent. Clearly, we can assume

$$\inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx) = \inf_{x \in X} f(x) + g(Lx). \quad (4.42)$$

Usually, in the classical convex approach, g is lsc and convex iff $g_{\Psi}^{**} = g$ so condition (4.42) holds. Generally, we obtain (4.42) whenever g is Ψ -convex [12, Proposition 5.1]. Conversely, we have the following.

Proposition 9. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ where X is a nonempty set and Y is a vector space. Let $L : X \rightarrow Y$ be a mapping. Let Φ, Ψ be the sets of elementary functions defined on X and Y , respectively. Assume (4.42) holds. If the primal problem is solvable at x_0 then g is Ψ -convex at Lx_0 .*

Proof. By assumption, We have

$$\inf_{x \in X} f(x) + g(Lx) \geq f(x_0) + g(Lx_0).$$

From (4.42) we have

$$f(x_0) + g(Lx_0) \leq \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx) \leq f(x_0) + g_{\Psi}^{**}(Lx_0),$$

which implies $g(Lx_0) \leq g_{\Psi}^{**}(Lx_0)$. From the definition of biconjugate function, $g_{\Psi}^{**} \leq g$, so we have $g(Lx_0) = g_{\Psi}^{**}(Lx_0)$. Thus, g is Ψ -convex at Lx_0 . $\square$

4.4.2 Lagrange Zero Duality Gap

In the present subsection, we discuss zero duality gap for Lagrange duality. We follow the result of [12] and exploit the intersection property to prove zero duality gap. In our context, Theorem 6.1 in [12] takes the following form.

Theorem 8. (Theorem 6.1 [12]) *Let $X = Y$ be a vector space, Φ be a convex set of elementary functions defined on X . Let $\mathcal{L}(x, \psi)$ be Lagrangian given by (4.41) with $L = Id$. Then the following are equivalent.*

- (i) *For every $\alpha < \inf_{x \in X} \sup_{\psi \in \Phi} \mathcal{L}(x, \psi)$, there exists $\psi_1, \psi_2 \in \Phi$ and $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1), \phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$ such that ϕ_1, ϕ_2 have the intersection property at level α ; i.e., for all $t \in [0, 1]$*

$$[t\phi_1 + (1-t)\phi_2 < \alpha] \cap [\phi_1 < \alpha] = \emptyset \text{ or } [t\phi_1 + (1-t)\phi_2 < \alpha] \cap [\phi_2 < \alpha] = \emptyset,$$

$$\text{where } [\phi_1 < \alpha] := \{x \in X : \phi_1(x) < \alpha\}.$$

- (ii) *$\text{val}(LP) = \text{val}(DCP)$ i.e. $\inf_{x \in X} \sup_{\psi \in \Phi} \mathcal{L}(x, \psi) = \sup_{\psi \in \Phi} \inf_{x \in X} \mathcal{L}(x, \psi)$.*

The proof can be found in [87].

Remark 9. In the case where for every $\alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi)$, we have $\psi_1 = \psi_2$, i.e. ϕ_1, ϕ_2 belong to the same support set $\text{supp } \mathcal{L}(\cdot, \psi_1)$ for some $\psi_1 \in \Psi$, the assumption of concavity of the Lagrangian can be removed. (cf. [86]).

For convenience, we state a result from [12].

Lemma 3. (*Lemma 6.2 [12]*) *Let X be a set, $\alpha \in \mathbb{R}$ and let $\phi_1, \phi_2 : X \rightarrow \mathbb{R}$. Two functions ϕ_1, ϕ_2 have the intersection property at level α if and only if there exists $t_0 \in [0, 1]$ such that $t_0 \phi_1(x) + (1 - t_0) \phi_2(x) \geq \alpha$ for all $x \in X$.*

Observe that in Theorem 6.3, the class of elementary functions Φ is arbitrary. Below, we relate the intersection property to the lower semi-continuity of the optimal value function, V , in the case where elementary functions satisfy additional conditions. For other results along this line, see e.g. [81, 54].

Let us define the optimal value function $V : Y \rightarrow [-\infty, +\infty)$ as follow

$$V(y) := \inf_{x \in X} \beta(x, y), \quad (4.43)$$

where the function β is defined at the beginning of Section 4.1. Similar to (4.42), we also define the following condition for any $y \in Y$

$$\inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx + y) = \inf_{x \in X} f(x) + g(Lx + y). \quad (4.44)$$

A function $V : Y \rightarrow [-\infty, +\infty)$ is lower semi-continuous at a point $y_0 \in Y$, if for every $\varepsilon > 0$, there exists a neighborhood $W(y_0)$ such that

$$V(y) > V(y_0) - \varepsilon, \quad \forall y \in W(y_0). \quad (4.45)$$

The following theorem below connects Theorem 3, the intersection property and the lower semi-continuity of the objective function at $y_0 = 0$.

Theorem 9. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$. Let $L : X \rightarrow Y$ be a mapping from X to Y with $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Let Φ, Ψ be sets of elementary functions defined on X and Y , respectively. Let $\mathcal{L}(x, \psi)$ be the Lagrangian defined in (4.41). Assume Ψ is convex, $0 \in \Phi$ and $\inf_{x \in X} f(x) + g(Lx) < +\infty$. Consider the following.*

- (i) *For every $\varepsilon > 0$, there exist $x_\varepsilon \in X$ and $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that $0 \in \partial_{\varepsilon, \Phi} (f + \psi_\varepsilon \circ L)(x_\varepsilon)$.*
- (ii) *For every $\alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi)$, there exists $\phi_1, \phi_2 \in \Phi$, $\psi_1, \psi_2 \in \Psi$, $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1)$, $\phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$ such that the intersection property holds for ϕ_1, ϕ_2 at level α .*
- (iii) *The function V is lower semi-continuous at 0.*

We have (i) $\Rightarrow$ (ii). If condition (4.42) holds, then (ii) $\Leftrightarrow$ (i). Moreover, if (4.42) holds and Ψ is composed of lower semi-continuous functions, then (ii) $\Rightarrow$ (iii). (iii) $\Rightarrow$ (ii) when Ψ is a collection of continuous functions and (4.44) holds.

Proof. (i) $\Rightarrow$ (ii): For every $\alpha \in \mathbb{R}$, let $\alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi)$. For every $\varepsilon > 0$, we can find $x_\varepsilon \in X$ and $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that $0 \in \partial_{\varepsilon, \Phi} (f + \psi_\varepsilon \circ L)(x_\varepsilon)$,

$$(\forall z \in X) \quad f(z) + \psi_\varepsilon(Lz) \geq f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) - \varepsilon. \quad (4.46)$$

We also have $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ so $g(Lx_\varepsilon) + g_\Psi^*(\psi_\varepsilon) \leq \psi_\varepsilon(Lx_\varepsilon) + \varepsilon$. Putting this into (4.46),

$$\begin{aligned} f(z) + \psi_\varepsilon(Lz) &\geq f(x_\varepsilon) + \psi_\varepsilon(Lx_\varepsilon) - \varepsilon \\ &\geq f(x_\varepsilon) + g(Lx_\varepsilon) + g_\Psi^*(\psi_\varepsilon) - 2\varepsilon, \end{aligned}$$

so

$$\mathcal{L}(z, \psi_\varepsilon) = f(z) + \psi_\varepsilon(Lz) - g_\Psi^*(\psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - 2\varepsilon. \quad (4.47)$$

As $\varepsilon > 0$ is arbitrary, we choose $\varepsilon = (f(x_\varepsilon) + g(Lx_\varepsilon) - \alpha)/2 > 0$ so that $\mathcal{L}(z, \psi_\varepsilon) > \alpha$. Pick $\phi_1 \in \Phi$ and $\psi_1 \in \Psi$ such that $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1)$. By choosing $\phi_2 = \alpha \in \Phi$, we have $\phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_\varepsilon)$. From Lemma 3 with $t = 1$ we have

$$\mathcal{L}(z, t\psi_\varepsilon + (1-t)\psi_1) = \mathcal{L}(z, \psi_\varepsilon) > t\alpha + (1-t)\phi_1(x) = \alpha,$$

so the intersection property holds for ϕ_1, ϕ_2 at level α . As α is arbitrary, we have (ii).

(ii) $\Rightarrow$ (i): As (4.42) holds, we have $\text{val}(\text{CP}) = \text{val}(\text{LP})$. From the primal problem (CP), for every $\varepsilon > 0$, there exists an $x_\varepsilon \in X$ such that

$$\inf_{x \in X} f(x) + g(Lx) > f(x_\varepsilon) + g(Lx_\varepsilon) - \varepsilon.$$

Denoting $\alpha := \text{val}(\text{CP}) - \varepsilon$, we have $\alpha < \text{val}(\text{CP})$. There exist $\phi_1, \phi_2 \in \Phi$ and $\psi_1, \psi_2 \in \Psi$ such that $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1)$, $\phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$ and the intersection property holds at level α . Lemma 3 gives us a $t_0 \in [0, 1]$ such that

$$t_0\phi_1(x) + (1-t_0)\phi_2(x) \geq \alpha, \quad \forall x \in X. \quad (4.48)$$

Because $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1)$, $\phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$, we have

$$\mathcal{L}(x, \psi_1) \geq \phi_1(x) \text{ and } \mathcal{L}(x, \psi_2) \geq \phi_2(x), \quad \forall x \in X.$$

As Ψ is a convex set, $t_0\psi_1 + (1-t_0)\psi_2 \in \Psi$. Moreover, $\mathcal{L}(x, \psi)$ is concave in

the second variable, i.e. for all $t \in [0, 1]$ and $\psi_1, \psi_2 \in \Psi$

$$\begin{aligned} \mathcal{L}(x, t\psi_1 + (1-t)\psi_2) &= f(x) + t\psi_1(Lx) + (1-t)\psi_2(Lx) - g_{\Psi}^*(t\psi_1 + (1-t)\psi_2) \\ &\geq t[f(x) + \psi_1(Lx) - g_{\Psi}^*(\psi_1)] + (1-t)[f(x) + \psi_2(Lx) - g_{\Psi}^*(\psi_2)] \\ &= t\mathcal{L}(x, \psi_1) + (1-t)\mathcal{L}(x, \psi_2), \end{aligned} \quad (4.49)$$

where the inequality holds due to the concavity of $-g_{\Psi}^*$. Combining inequality (4.49) with (4.48), we get

$$\begin{aligned} \mathcal{L}(x, t_0\psi_1 + (1-t_0)\psi_2) &\geq t_0\mathcal{L}(x, \psi_1) + (1-t_0)\mathcal{L}(x, \psi_1) \\ &\geq t_0\phi_1(x) + (1-t_0)\phi_2(x) \geq \alpha \\ &= \text{val}(\text{CP}) - \varepsilon \\ &= \inf_{x \in X} f(x) + g(Lx) - \varepsilon \\ &> f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - 2\varepsilon. \end{aligned}$$

We obtain

$$\mathcal{L}(x, t_0\psi_1 + (1-t_0)\psi_2) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - 2\varepsilon, \quad \forall x \in X$$

By denoting $\bar{\psi} = t_0\psi_1 + (1-t_0)\psi_2$, we get

$$\mathcal{L}(x, \bar{\psi}) = f(x) + \bar{\psi}(Lx) - g_{\Psi}^*(\bar{\psi}) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - 2\varepsilon. \quad (4.50)$$

Rearranging both sides, we obtain

$$f(x) - f(x_{\varepsilon}) > -\bar{\psi}(Lx) + g(Lx_{\varepsilon}) + g_{\Psi}^*(\bar{\psi}) - 2\varepsilon.$$

We have $g(Lx) + g_{\Psi}^*(\psi) \geq \psi(Lx)$ for all $x \in X$ and $\psi \in \Psi$. From the previous inequality,

$$f(x) - f(x_{\varepsilon}) > -\bar{\psi}(Lx) + \bar{\psi}(Lx_{\varepsilon}) - 2\varepsilon.$$

Because, in general, $-\bar{\psi} \circ L$ does not belong to Φ , we cannot claim that $-\bar{\psi} \circ L \in \partial_{2\varepsilon, \Phi} f(x_\varepsilon)$, but we have

$$f(x) + \bar{\psi}(Lx) - f(x_\varepsilon) - \bar{\psi}(Lx_\varepsilon) > -2\varepsilon,$$

which means $0 \in \partial_{2\varepsilon, \Phi}(f + \bar{\psi} \circ L)(x_\varepsilon)$. On the other hand, from (4.50), for any $x \in X$, we have

$$f(x) + \bar{\psi}(Lx) - g_\Psi^*(\bar{\psi}) > f(x_\varepsilon) + g(Lx_\varepsilon) - 2\varepsilon.$$

By choosing $x = x_\varepsilon$, we can write

$$\bar{\psi}(Lx_\varepsilon) - g_\Psi^*(\bar{\psi}) > g(Lx_\varepsilon) - 2\varepsilon,$$

or

$$\bar{\psi}(Lx_\varepsilon) + 2\varepsilon > g(Lx_\varepsilon) + g_\Psi^*(\bar{\psi}).$$

This gives $\bar{\psi} \in \partial_{2\varepsilon, \Psi} g(Lx_\varepsilon)$. Thus, (i) holds.

(ii) $\Rightarrow$ (iii): As the intersection property holds for every $\alpha \in \mathbb{R}$, for every $\varepsilon > 0$, we can take $\alpha_\varepsilon = \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) - \varepsilon/2 = V(0) - \varepsilon/2$. Now we can find $\phi_1, \phi_2 \in \Phi, \psi_1, \psi_2 \in \Psi$ such that $\phi_1(\cdot) \leq \mathcal{L}(\cdot, \psi_1), \phi_2 \leq \mathcal{L}(\cdot, \psi_2)$ for all $x \in X$, and ϕ_1, ϕ_2 satisfy the intersection property at level α_ε . Using Lemma 3, there exists $t_0 \in [0, 1]$ such that for all $x \in X$,

$$\begin{aligned} \alpha_\varepsilon &\leq t_0 \phi_1(x) + (1 - t_0) \phi_2(x) \leq t_0 \mathcal{L}(x, \psi_1) + (1 - t_0) \mathcal{L}(x, \psi_2) \\ &\leq \mathcal{L}(x, t_0 \psi_1 + (1 - t_0) \psi_2). \end{aligned} \tag{4.51}$$

On the other hand, we have

$$\begin{aligned}\mathcal{L}(x, \psi) &= f(x) + \psi(Lx) - g_{\Psi}^*(\psi) \\ &= \inf_{y \in Y} f(x) + g(y) + \psi(Lx) - \psi(y) \\ &\leq f(x) + g(Lx + y) + \psi(Lx) - \psi(Lx + y),\end{aligned}$$

where $x \in X, y \in Y$. Denoting $\psi_0 = t_0\psi_1 + (1 - t_0)\psi_2 \in \Psi$, as Ψ is a convex set, (4.51) gives us

$$\begin{aligned}\alpha_{\varepsilon} &\leq f(x) + g(Lx + y) + \psi_0(Lx) - \psi_0(Lx + y) \\ \alpha_{\varepsilon} + \psi_0(Lx + y) - \psi_0(Lx) &\leq f(x) + g(Lx + y),\end{aligned}$$

for any $x \in X, y \in Y$. Since $\psi_0 \in \Psi$ is lower semi-continuous, the function $\psi_x(y) := \psi_0(Lx + y)$ is also lower semi-continuous for all $y \in Y$. There exists a neighborhood $W(0) \subset Y$ such that

$$\psi_x(y) \geq \psi_x(0) - \varepsilon/2, \quad \forall y \in W(0).$$

Thus,

$$f(x) + g(Lx + y) \geq \alpha_{\varepsilon} + \psi_x(y) - \psi_x(0) \geq \alpha - \varepsilon/2,$$

for all $y \in W(0)$. The inequality holds for all $x \in X$, we can take the infimum with respect to $x \in X$ on both sides and obtain

$$V(y) = \inf_{x \in X} f(x) + g(Lx + y) \geq \alpha_{\varepsilon} - \varepsilon/2 = V(0) - \varepsilon,$$

for all $y \in W(0)$. Hence, $V(y)$ is lower semi-continuous at $y = 0$.

(iii) $\Rightarrow$ (ii): Conversely, for every $\alpha \in \mathbb{R}, \alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = V(0)$, we choose $3\varepsilon = V(0) - \alpha > 0$. There exists a neighborhood $W(0) \subset Y$ such that

$$V(0) - \varepsilon \leq V(y), \quad \forall y \in W(0).$$

Now considering

$$f(x) + g_{\Psi}^{**}(Lx + y) = \sup_{\psi \in \Psi} f(x) + \psi(Lx + y) - g_{\Psi}^*(\psi).$$

We can find $\psi_{\varepsilon} \in \Psi$ such that

$$\begin{aligned} f(x) + g_{\Psi}^{**}(Lx + y) &< f(x) + \psi_{\varepsilon}(Lx + y) - g_{\Psi}^*(\psi_{\varepsilon}) + \varepsilon \\ &\leq f(x) - g_{\Psi}^*(\psi_{\varepsilon}) + \psi_{\varepsilon}(Lx) + 2\varepsilon, \quad \forall y \in W_1(0). \end{aligned}$$

In the last inequality, we can find a neighborhood $W_1(0) \subset Y$ such that ψ_{ε} is continuous at 0. Hence, for all $y \in W_1(0) \cap W(0)$, we have

$$\begin{aligned} V(0) - \varepsilon &\leq V(y) = \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx + y) \\ &\leq \sup_{\psi \in \Psi} f(x) - g_{\Psi}^*(\psi) + \psi(Lx + y), \quad \forall x \in X \\ &< f(x) - g_{\Psi}^*(\psi_{\varepsilon}) + \psi_{\varepsilon}(Lx + y) + \varepsilon \\ &\leq f(x) - g_{\Psi}^*(\psi_{\varepsilon}) + \psi_{\varepsilon}(Lx) + 2\varepsilon, \quad \forall y \in W_1(0) \\ &= \mathcal{L}(x, \psi_{\varepsilon}) + 2\varepsilon. \end{aligned}$$

Finally, $\alpha = V(0) - 3\varepsilon \leq \mathcal{L}(x, \psi_{\varepsilon})$ for all $x \in X$, we can take $\phi = \alpha \in \Phi$ and $\phi \in \text{supp } \mathcal{L}(\cdot, \psi_{\varepsilon})$. Then ϕ has the intersection property at level α with any $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi)$, for $\psi \in \Psi$. $\square$

Now we can state zero duality gap for Lagrange dual problem.

Proposition 10. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$. Let $L : X \rightarrow Y$ be a mapping from X to Y with $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Let Φ, Ψ be sets of elementary functions defined on X and Y , respectively. Let $\mathcal{L}(x, \psi)$ be the Lagrangian defined by (4.41). Assume Ψ is convex, $0 \in \Phi$ and (4.42) holds. We further assume $\inf_{x \in X} f(x) + g(Lx) < +\infty$. The following are equivalent.*

(i) *For every $\alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi)$, there exists $\psi_1, \psi_2 \in \Psi$ and $\bar{\phi}_1 \in$*

$\text{supp } \mathcal{L}(\cdot, \psi_1), \bar{\phi}_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$ such that $\bar{\phi}_1, \bar{\phi}_2$ have the intersection property at level α .

$$(ii) \inf_{x \in X} f(x) + g(Lx) = \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = \sup_{\psi \in \Psi} \inf_{x \in X} \mathcal{L}(x, \psi) < +\infty.$$

Proof. By applying Theorem 9 and Theorem 3, we obtain the assertion. $\square$

Remark 10.

- In order to work with the intersection property, we need to find $\psi_1, \psi_2 \in \Psi$ and $\phi_1 \in \text{supp } \mathcal{L}(\cdot, \psi_1), \phi_2 \in \mathcal{L}(\cdot, \psi_2)$. Because $\alpha < \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi)$, so $\alpha \in \text{supp } \mathcal{L}(\cdot, \psi)$ for any $\psi \in \Psi$. We can set $\phi_1 = \alpha \in \Phi$, and the intersection property always holds for any $\phi_2 \in \text{supp } \mathcal{L}(\cdot, \psi_2)$, as $[\phi_1 < \alpha] = \emptyset$.
- Let Ψ be symmetric, and for every $\varepsilon > 0$, there exist $\bar{x} \in X, \psi_1 \circ L \in \Phi$ for $\psi_1 \in \Psi$, such that $\psi_1 \circ L \in \partial_{\varepsilon, \Phi} f(\bar{x})$. We have, for all $y \in X$

$$\begin{aligned} f(y) - f(\bar{x}) &\geq \psi_1(Ly) - \psi_1(L\bar{x}) - \varepsilon \\ f(y) - \psi_1(Ly) - g_{\Psi}^*(-\psi_1) &\geq f(\bar{x}) - \psi_1(L\bar{x}) - g_{\Psi}^*(-\psi_1) - \varepsilon \\ \mathcal{L}(y, -\psi_1) &\geq \mathcal{L}(\bar{x}, -\psi_1) - \varepsilon, \end{aligned}$$

so $\bar{x}$ is a ε -approximate solution to the problem $\inf_{x \in X} \mathcal{L}(x, -\psi_1)$.

We can replace assumption (4.42) with a stronger one by the following lemma.

Lemma 4. Assume $\inf_{x \in X} f(x) + g(Lx) < +\infty$ and let $\mathcal{L}(x, \psi)$ be the Lagrangian function given by (4.41). Then condition (4.42) i.e.,

$$\inf_{x \in X} f(x) + g(Lx) = \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx),$$

holds if and only if for every $\varepsilon \geq 0$, there exists $x_{\varepsilon} \in X$ such that

$$\inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - \varepsilon. \quad (4.52)$$

Proof. If condition (4.42) holds i.e.,

$$\inf_{x \in X} f(x) + g(Lx) = \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx),$$

then for every $\varepsilon > 0$, one can find an $x_{\varepsilon} \in X$ such that $\inf_{x \in X} f(x) + g(Lx) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - \varepsilon$. Hence,

$$\inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - \varepsilon.$$

Conversely, assume that

$$\inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - \varepsilon.$$

We have

$$\inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) > f(x_{\varepsilon}) + g(Lx_{\varepsilon}) - \varepsilon \geq \inf_{x \in X} f(x) + g(Lx) - \varepsilon.$$

Recall that,

$$\inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx).$$

Because $g(Lx) \geq g_{\Psi}^{**}(Lx)$ for all $x \in X$, we can write

$$\inf_{x \in X} f(x) + g(Lx) \geq \inf_{x \in X} f(x) + g_{\Psi}^{**}(Lx) \geq \inf_{x \in X} f(x) + g(Lx) - \varepsilon.$$

Both sides do not depend on ε , we can let $\varepsilon \rightarrow 0$ and get the equality. $\square$

In fact, condition (4.52) is stronger than intersection property, as we can see in the corollary below.

Corollary 3. *Let $f : X \rightarrow (-\infty, +\infty]$, $g : Y \rightarrow (-\infty, +\infty]$ and let $L : X \rightarrow Y$ be a mapping from X to Y with $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$. Let Φ, Ψ be sets of elementary functions defined on X and Y , respectively. Let the Lagrangian function be defined by (4.41). Assume Ψ is convex, $0 \in \Phi$,*

$\inf_{x \in X} f(x) + g(Lx) < +\infty$. The following are equivalent.

- (i) For every $\varepsilon > 0$, there exists $x_\varepsilon \in X, \psi_\varepsilon \in \Psi$ such that $\inf_{x \in X} \mathcal{L}(x, \psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - \varepsilon$.
- (ii) For every $\varepsilon > 0$, there exist $x_\varepsilon \in X$ and $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ such that $0 \in \partial_{\varepsilon, \Phi} (f + \psi_\varepsilon \circ L)(x_\varepsilon)$.
- (iii) $\inf_{x \in X} f(x) + g(Lx) = \inf_{x \in X} \sup_{\psi \in \Psi} \mathcal{L}(x, \psi) = \sup_{\psi \in \Psi} \inf_{x \in X} \mathcal{L}(x, \psi) < +\infty$.

Proof. Thanks to Theorem 3, Theorem 9 and Proposition 10, we have (ii) $\Leftrightarrow$

(iii). We only need to prove (i) $\Leftrightarrow$ (ii).

(i) $\Rightarrow$ (ii): For every $\varepsilon > 0$ and every $x \in X$, we have

$$\mathcal{L}(x, \psi_\varepsilon) \geq \inf_{x \in X} \mathcal{L}(x, \psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - \varepsilon.$$

While $\mathcal{L}(x, \psi_\varepsilon) = f(x) + \psi_\varepsilon(Lx) - g_\Psi^*(\psi_\varepsilon)$, we have

$$f(x) + \psi_\varepsilon(Lx) - g_\Psi^*(\psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - \varepsilon. \quad (4.53)$$

Let $x = x_\varepsilon$ and we obtain

$$\psi_\varepsilon(Lx_\varepsilon) - g_\Psi^*(\psi_\varepsilon) \geq g(Lx_\varepsilon) - \varepsilon,$$

so $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$. On the other hand, by using the property of Φ -conjugate,

$$(\forall x \in X) \quad g(Lx) + g_\Psi^*(\psi_\varepsilon) \geq \psi_\varepsilon(Lx).$$

By putting this in (4.53), we have $0 \in \partial_{\varepsilon, \Phi} (f + \psi_\varepsilon \circ L)(x_\varepsilon)$. Hence, we prove (ii).

(ii) $\Rightarrow$ (i): For every $\varepsilon > 0$, there exist $x_\varepsilon \in X, \psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$ and $0 \in \partial_{\varepsilon, \Phi} (f + \psi_\varepsilon \circ L)(x_\varepsilon)$, i.e. for all $z \in X$,

$$f(z) + \psi_\varepsilon(Lz) - f(x_\varepsilon) - \psi_\varepsilon(Lx_\varepsilon) \geq -\varepsilon.$$

Because $\psi_\varepsilon \in \partial_{\varepsilon, \Psi} g(Lx_\varepsilon)$, using inequality (3.9) we get,

$$\begin{aligned} f(z) + \psi_\varepsilon(Lz) - f(x_\varepsilon) &\geq \psi_\varepsilon(Lx_\varepsilon) - \varepsilon \\ &\geq g_\Psi^*(\psi_\varepsilon) + g(Lx_\varepsilon) - 2\varepsilon. \end{aligned}$$

Thus

$$\mathcal{L}(z, \psi_\varepsilon) = f(z) + \psi_\varepsilon(Lz) - g_\Psi^*(\psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - 2\varepsilon.$$

The right-hand side does not depend on $z \in X$, taking the infimum with respect to $z \in X$ gives us

$$\inf_{z \in X} \mathcal{L}(z, \psi_\varepsilon) \geq f(x_\varepsilon) + g(Lx_\varepsilon) - 2\varepsilon.$$

We have proved (i). □

Remark 11. Our idea in Corollary 3 is to make a connection between Lagrange duality and Conjugate duality. In statement (i.), we ensure that the Lagrange primal and the primal coincide, and the Lagrange primal has to be bounded below by the ε -solution to the primal problem. Thus x_ε and ψ_ε in (i.) become the same x_ε and ψ_ε in statement (ii.).

Having the class Ψ of elementary functions, we recall the notion of X -convexity in [72, Proposition 1.2.3], by using formula $\phi(x) = x(\phi)$ for $x \in X, \phi \in \Phi$. Together with the intersection property, we can prove Lagrange zero duality gap as follow.

Proposition 11. *Assume that for every $\alpha \in \mathbb{R}, \alpha \leq \inf_{x \in X} f(x) + \psi(Lx)$, there exist $\psi \in \text{supp } g, x_0 \in X$, such that $f(x_0) \leq \alpha \leq \psi(Lx_0)$. If $x_0 \in A = \{x \in X : Lx \in \text{supp } g_\Psi^*\} \neq \emptyset$, then we have $\text{val}(LP) = \text{val}(LD)$.*

Proof. Let $x_0 \in A, \psi \in \text{supp } g$ be such that

$$f(x_0) \leq \alpha \leq \psi(Lx_0).$$

Then $g_{\Psi}^*(\psi) \leq 0$. On the other hand, $x_0 \in A$ means

$$(\forall \psi \in \Psi) \psi(Lx_0) \leq g_{\Psi}^*(\psi) \Leftrightarrow \sup_{\psi \in \Psi} \psi(Lx_0) - g_{\Psi}^*(\psi) \leq 0.$$

Moreover, $\psi(Lx_0) \geq \alpha$ and

$$\begin{aligned} \alpha &\leq \inf_{x \in X} f(x) + \psi(Lx) \\ &\leq \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi), \quad \text{as } g_{\Psi}^*(\psi) \leq 0 \\ &\leq \sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi). \end{aligned}$$

On the other hand,

$$\begin{aligned} \alpha &\geq f(x_0) \geq \sup_{\psi \in \Psi} f(x_0) + \psi(Lx_0) - g_{\Psi}^*(\psi) \\ &\geq \inf_{x \in X} \sup_{\psi \in \Psi} f(x) + \psi(Lx) - g_{\Psi}^*(\psi). \end{aligned}$$

Combining the two above inequalities, we obtain

$$\sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi) \geq \inf_{x \in X} \sup_{\psi \in \Psi} f(x) + \psi(Lx) - g_{\Psi}^*(\psi).$$

□

Remark 12. In Proposition 11, we can replace the assumption $x_0 \in A$ with $g(Lx_0) \leq 0$, for $x_0 \in X$. Then zero duality gap holds between (CP) and (LD). Because, for $\psi_0 \in \text{supp } g$ and $\psi_0 \geq \alpha \geq f(x_0)$, we have

$$\begin{aligned} \sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(Lx) - g_{\Psi}^*(\psi) &\geq \inf_{x \in X} f(x) + \psi_0(Lx) - g_{\Psi}^*(\psi_0) \\ &\geq \inf_{x \in X} f(x) + \psi_0(Lx), \quad \text{as } g_{\Psi}^*(\psi_0) \leq 0 \\ &\geq \alpha \geq f(x_0) \geq f(x_0) + g(Lx_0) \\ &\geq \inf_{x \in X} f(x) + g(Lx). \end{aligned}$$

We complete this Section with two examples illustrating Proposition 11.

Example 12. Consider the functions $f(x) = 3x^2 - 3x - 10$, $g(x) = -2x^2 + x - 8$ and $L = Id$. Let

$$\Phi = \Psi = \{\psi(x) = -ax^2 + bx + c, a \geq 0, b, c \in \mathbb{R}\},$$

is our sets of elementary functions. We want to find $\text{supp } g$, i.e. the set of all $\psi \in \Psi$ such that

$$\psi(x) \leq g(x) \quad \forall x \in \mathbb{R}.$$

This gives us $\psi \in \Psi$ where $a > 2$, $c \leq -\frac{(1-b)^2}{4(a-2)} - 8$ or $\psi(x) = -2x^2 + x - 8$. Consider $a > 2$, for every $\alpha \in \mathbb{R}$ such that

$$\alpha \leq \inf_x f(x) + \psi(x) = \begin{cases} c - 10 & \text{if } a = 3, b = 3, c \leq -9, \\ -\frac{(b-3)^2}{4(3-a)} - 10 + c & \text{if } 2 < a < 3, c \leq -\frac{(1-b)^2}{4(a-2)} - 8, \\ -\infty & \text{if } a > 3. \end{cases}$$

We need to find x_0 such that $f(x_0) \leq \alpha \leq \psi(x_0)$, we can let $\alpha = \inf_x f(x) + \psi(x)$. There are two cases

- $\alpha = c - 10$ or $\alpha \leq -19$, we cannot find x_0 such that $f(x_0) \leq -19$ as $f_{\min} = -\frac{43}{4}$.
- $\alpha = -\frac{(b-3)^2}{4(3-a)} - 10 + c$, we have to solve $f(x_0) \leq \alpha \leq \psi(x_0)$ for x_0 . We arrive at the following system

$$\begin{aligned} -ax^2 + bx + c &\geq -\frac{(b-3)^2}{4(3-a)} - 10 + c, \\ 3x^2 - 3x - 10 &\leq -\frac{(b-3)^2}{4(3-a)} - 10 + c, \\ 2 < a < 3, c &\leq -\frac{(1-b)^2}{4(a-2)} - 8. \end{aligned}$$

By direct calculations, there are no solution a, b, c, x for the above system of inequalities.

For $\psi(x) = -2x^2 + x - 8$, and this turns into

$$\inf_{x \in \mathbb{R}} f(x) + g(x) = \alpha = \inf_{x \in \mathbb{R}} f(x) + \psi(x) \leq \sup_{\psi \in \Psi} \inf_{x \in \mathbb{R}} f(x) + \psi(x),$$

which is zero duality gap.

Example 13. Now consider the functions $f(x) = x^2 - 3x - 10$, $g(x) = 2x + 1$ with $L = Id$, and the set of elementary functions is

$$\Phi = \Psi = \{\psi(x) = ax + b, \text{ where } a, b \in \mathbb{R}\}.$$

Calculating $\psi \in \text{supp } g$ gives us $a = 2$ and $b \leq 1$. Now we find

$$\alpha = \inf_{x \in \mathbb{R}} f(x) + \psi(x) = \inf_{x \in \mathbb{R}} x^2 - x - 10 + b = -\frac{21}{2} + b.$$

We also calculate

$$g_{\Psi}^*(\psi) = \sup_{x \in X} \psi(x) - g(x) = b - 1,$$

for $b \leq 1$. Observe that

$$\begin{aligned} \sup_{\psi \in \text{supp } g} \inf_{x \in X} f(x) + \psi(x) - g_{\Psi}^*(\psi) &= \sup_{\psi \in \text{supp } g} \inf_{x \in X} x^2 - x - 10 + b - b + 1 \\ &= \sup_{\psi \in \text{supp } g} \inf_{x \in X} x^2 - x - 9 \\ &= \inf_{x \in X} x^2 - x - 9 = \inf_{x \in \mathbb{R}} f(x) + g(x), \end{aligned}$$

while

$$\sup_{\psi \in \text{supp } g} \inf_{x \in X} f(x) + \psi(x) - g_{\Psi}^*(\psi) \leq \sup_{\psi \in \Psi} \inf_{x \in X} f(x) + \psi(x) - g_{\Psi}^*(\psi).$$

Thus we have zero duality gap.

Part II

Algorithms for Composite Optimization Problems involving $\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions

Chapter 5

$\Phi_{lsc}^{\mathbb{R}}$ -Proximal Algorithms

In this chapter, we consider the class $\Phi_{lsc}^{\mathbb{R}}$ -convex function. Our aim is to design proximal-based algorithms using $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials and discuss convergence of these algorithms. The outline of of this chapter is as follow:

- Section 5.1: We construct $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator using $\Phi_{lsc}^{\mathbb{R}}$ duality map (Definition 9) and present its properties (Theorem 10 and 11).
- Section 5.2: First, we give a general auxiliary convergence result which can be applied throughout this chapter (Theorem 12). Then we propose $\Phi_{lsc}^{\mathbb{R}}$ -proximal point algorithm and provide convergence results (Theorem 13).
- Section 5.3: We propose $\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward algorithm and provide convergence analysis (Theorem 15).
- Section 5.4: We propose and provide convergence results for $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient algorithm (Proposition 16 and Theorem 16).
- Section 5.5: We illustrate $\Phi_{lsc}^{\mathbb{R}}$ -proximal point algorithm and $\Phi_{lsc}^{\mathbb{R}}$ -projected subgradient algorithm with some numerical examples.

5.1 $\Phi_{lsc}^{\mathbb{R}}$ -Proximal Operator

Let X be a Hilbert space. In this section, we introduce the proximal operator related to the class $\Phi_{lsc}^{\mathbb{R}}$ -convex functions. Observe that $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials lie in the set $\Phi_{lsc}^{\mathbb{R}}$ which is not a subset of X . Therefore, we need a mapping which can serve as a link between X and $\Phi_{lsc}^{\mathbb{R}}$. In analogy to the classical constructions we consider the function g_{γ} from Example 3.

Definition 9. Let $\gamma > 0$ and $g_{\gamma}(x) = \frac{1}{2\gamma}\|x\|^2$, we define $\Phi_{lsc}^{\mathbb{R}}$ -duality map $J_{\gamma} : X \rightrightarrows \Phi_{lsc}^{\mathbb{R}}$ as

$$J_{\gamma}(x) := \partial_{lsc}^{\mathbb{R}} g_{\gamma}(x) = \partial_{lsc}^{\mathbb{R}} \left(\frac{1}{2\gamma} \|x\|^2 \right).$$

Its inverse $J_{\gamma}^{-1} : \Phi_{lsc}^{\mathbb{R}} \rightrightarrows X$ is

$$J_{\gamma}^{-1}(\phi) := (\partial_{lsc}^{\mathbb{R}} g_{\gamma})^{-1}(\phi).$$

The explicit forms of J_{γ} and $J_{\gamma}^{-1}(\phi)$ are calculated in Example 3.

$$J_{\gamma}(x) = \left\{ \phi = (a, u) \in \Phi_{lsc}^{\mathbb{R}} : \phi = \left(a, \left(\frac{1}{\gamma} + 2a \right) x \right), 2\gamma a \geq -1 \right\}. \quad (5.1)$$

$$(J_{\gamma})^{-1}(\phi) = (J_{\gamma})^{-1}(a, u) = \left\{ x \in X : \begin{cases} \left\{ \frac{\gamma}{1+2\gamma a} u \right\} & \text{if } 2\gamma a > -1 \\ X & \text{if } a = -\frac{1}{2\gamma}, u = 0 \\ \emptyset & \text{otherwise} \end{cases} \right\}.$$

When $\gamma = 1$ and $\Phi = \Phi_{lsc}^0$ then $J_1 = \partial_{lsc}^0(\frac{1}{2}\|\cdot\|^2)$ which is the classical duality mapping known in [73, Example 2.26].

We consider the problem

$$\min_{x \in X} f(x), \quad (5.2)$$

where $f : X \rightarrow (-\infty, +\infty]$ which is proper $\Phi_{lsc}^{\mathbb{R}}$ -convex. Finding a global solution to (5.2) means that we need to find $x_0 \in \text{dom } f$ such that

$$(0, 0) \in \partial_{lsc}^{\mathbb{R}} f(x_0), \quad (5.3)$$

see [82, Proposition 7.13]. By using Definition 9, we define $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator, $\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}} : X \rightrightarrows X$ (set-valued map)

$$\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x) := (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1} J_{\gamma}(x). \quad (5.4)$$

The concept of $\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}$ is related to the concept of resolvent operator which is defined for classical convex subdifferentials as $(Id + \partial_{lsc}^0 f)^{-1}$. When f is convex, this is also known as proximity operator

$$\text{prox}_{\gamma f}(x_0) = \arg \min_{z \in X} f(z) + \frac{1}{2\gamma} \|z - x_0\|^2. \quad (5.5)$$

Remark 13. When reduced to convex analysis, the mapping J_{γ} is actually the gradient of the norm square function which is single-valued and this implies the equivalence between optimality condition (5.3) and the fixed point of the proximal operator [6, Proposition 12.29]. While in the weakly convex case, we have $\partial_{lsc}^a f(x)$ as the subdifferentials with fixed a . Then the fixed point of proximal operator is equivalent to a -critical point of f [10, Corollary 1].

For the class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions, we show that the global minimizers of f are related to fixed points of the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator of f .

Theorem 10. *Let X be a Hilbert space. Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function.*

- *If x_0 is a global minimizer of (5.2), then x_0 is a fixed point of $\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}$.*
- *If $x_0 \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0)$, then x_0 is a critical point in the sense of Definition 5. Additionally, if either*

(i) $(-1/2\gamma, 0) \in \partial_{lsc}^{\mathbb{R}}(f + \frac{1}{2\gamma}\|\cdot\|^2)(x_0)$, or

(ii) $\exists \phi_1, \phi_2 \in J_{\gamma}(x_0), \phi_1 - \phi_2 = (a_1 - a_2, u_1 - u_2) \in \partial_{lsc}^{\mathbb{R}}f(x_0), a_1 \leq a_2$,

then x_0 is a global minimizer of f .

Proof. Let x_0 be a global minimizer of f . Then x_0 satisfies (5.3), as $(0, 0) \in \partial_{lsc}^{\mathbb{R}}f(x_0)$, and

$$\phi \in \partial_{lsc}^{\mathbb{R}}f(x_0) + J_{\gamma}(x_0),$$

for any $\phi \in J_{\gamma}(x_0)$. This means that

$$x_0 \in (\partial_{lsc}^{\mathbb{R}}f + J_{\gamma})^{-1}(\phi) \subset (\partial_{lsc}^{\mathbb{R}}f + J_{\gamma})^{-1}J_{\gamma}(x_0). \quad (5.6)$$

Therefore, $x_0 \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0)$.

On the other hand, let us assume that x_0 is a fixed point of $\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}$. From (5.6), there exists $\phi_1 \in J_{\gamma}(x_0)$ such that

$$\phi_1 \in \partial_{lsc}^{\mathbb{R}}f(x_0) + J_{\gamma}(x_0). \quad (5.7)$$

If $\phi_1 = (-\frac{1}{2\gamma}, 0)$ then x_0 is a minimizer of f as

$$(-\frac{1}{2\gamma}, 0) \in \partial_{lsc}^{\mathbb{R}}f(x_0) + J_{\gamma}(x_0) \subseteq \partial_{lsc}^{\mathbb{R}}(f + \frac{1}{2\gamma}\|\cdot\|^2)(x_0).$$

If $\phi_1 = (a_1, (\frac{1}{\gamma} + 2a_1)x_0), a_1 > -1/2\gamma$, by (5.7), there exist $\phi_2 \in J_{\gamma}(x_0)$ and $\phi_3 \in \partial_{lsc}^{\mathbb{R}}f(x_0)$ such that $\phi_1 = \phi_2 + \phi_3$ or

$$\phi_1 - \phi_2 = \phi_3 \in \partial_{lsc}^{\mathbb{R}}f(x_0).$$

As $\phi_2 \in J_{\gamma}(x_0), \phi_2 = (a_2, (\frac{1}{\gamma} + 2a_2)x_0)$ for a_1, a_2 such that $2\gamma a_2 \geq -1$. By the

definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, for all $y \in X$,

$$\begin{aligned}
 f(y) - f(x_0) &\geq (\phi_1 - \phi_2)(y) - (\phi_1 - \phi_2)(x_0) \\
 &\geq -(a_1 - a_2) (\|y\|^2 - \|x_0\|^2) + \langle (\frac{1}{\gamma} + 2a_1)x_0 - (\frac{1}{\gamma} + 2a_2)x_0, y - x_0 \rangle \\
 &= -(a_1 - a_2) (\|y\|^2 - \|x_0\|^2) + 2(a_1 - a_2) \langle x_0, y - x_0 \rangle \\
 &= (a_2 - a_1) \|y - x_0\|^2,
 \end{aligned} \tag{5.8}$$

the right hand side is non-negative if $a_2 \geq a_1$, which implies that x_0 is a minimizer of f . From (5.8), in general, we have x_0 is $(a_2 - a_1)$ -critical point of f . $\square$

The next result characterizes the elements of $\Phi_{lsc}^{\mathbb{R}}$ -proximal map.

Theorem 11. *Let X be a Hilbert space and $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function. Let $x_0 \in \text{dom } f, \gamma > 0$. Then*

$$x \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0) \Leftrightarrow \exists a_0 \geq -\frac{1}{2\gamma} \text{ s.t. } x \in \arg \min_{z \in X} \left[f(z) + \left(\frac{1}{2\gamma} + a_0 \right) \|z - x_0\|^2 \right]. \tag{5.9}$$

Proof. Let $x \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0)$. By definition, there exists $\phi_0 \in J_{\gamma}(x_0)$ such that

$$x \in (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1}(\phi_0).$$

By (3.16), there exists $a_0 \geq -1/2\gamma$ such that $\phi_0 := \left(a_0, \left(\frac{1}{\gamma} + 2a_0 \right) x_0 \right) \in J_{\gamma}(x_0)$ which satisfies

$$\left(a_0, \left(\frac{1}{\gamma} + 2a_0 \right) x_0 \right) \in \partial_{lsc}^{\mathbb{R}} f(x) + J_{\gamma}(x) \subseteq \partial_{lsc}^{\mathbb{R}} (f + \frac{1}{2\gamma} \|\cdot\|^2)(x).$$

Hence, by the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferential, for all $y \in X$

$$f(y) + \frac{1}{2\gamma}\|y\|^2 - f(x) - \frac{1}{2\gamma}\|x\|^2 \geq -a_0(\|y\|^2 - \|x\|^2) + \left(\frac{1}{\gamma} + 2a_0\right)\langle x_0, y - x \rangle. \quad (5.10)$$

After simplifying all the quadratic terms in (5.10), we obtain

$$\begin{aligned} f(y) - f(x) &\geq -\left(\frac{1}{2\gamma} + a_0\right)\|y - x\|^2 - \left(\frac{1}{\gamma} + 2a_0\right)\langle x, y - x \rangle + \left(\frac{1}{\gamma} + 2a_0\right)\langle x_0, y - x \rangle \\ &\geq -\left(\frac{1}{2\gamma} + a_0\right)\|y - x\|^2 + \left(\frac{1}{\gamma} + 2a_0\right)\langle x_0 - x, y - x \rangle. \end{aligned} \quad (5.11)$$

By writing the inner product in terms of the norms, we have

$$f(y) - f(x) \geq \left(\frac{1}{2\gamma} + a_0\right) [\|x_0 - x\|^2 - \|y - x_0\|^2],$$

which results in

$$x \in \arg \min_{y \in X} f(y) + \left(\frac{1}{2\gamma} + a_0\right) \|y - x_0\|^2. \quad (5.12)$$

Conversely, let us assume (5.12). Then for all $y \in X$

$$\begin{aligned} f(y) - f(x) &\geq \left(\frac{1}{2\gamma} + a_0\right) [\|x_0 - x\|^2 - \|y - x_0\|^2] \\ &= -\left(\frac{1}{2\gamma} + a_0\right) (\|y\|^2 - \|x\|^2) + \left(\frac{1}{\gamma} + 2a_0\right) \langle x_0, y - x \rangle. \end{aligned} \quad (5.13)$$

This is equivalent to

$$\left(a_0, \left(\frac{1}{\gamma} + 2a_0\right)x_0\right) \in \partial_{lsc}^{\mathbb{R}} \left(f + \frac{1}{2\gamma}\|\cdot\|^2\right)(x).$$

To finish the proof, we need to prove

$$\partial_{lsc}^{\mathbb{R}} \left(f + \frac{1}{2\gamma} \|\cdot\|^2 \right) (x) = (\partial_{lsc}^{\mathbb{R}} f + J_{\gamma})(x).$$

We only have to prove that

$$\partial_{lsc}^{\mathbb{R}} \left(f + \frac{1}{2\gamma} \|\cdot\|^2 \right) (x) \subset (\partial_{lsc}^{\mathbb{R}} f + J_{\gamma})(x).$$

Let us take $(a, u) \in \partial_{lsc}^{\mathbb{R}} \left(f + \frac{1}{2\gamma} \|\cdot\|^2 \right) (x)$, so $(a + \frac{1}{2\gamma}, u) \in \partial_{lsc}^{\mathbb{R}} f(x)$ and $(-\frac{1}{2\gamma}, 0) \in J_{\gamma}(x)$ from Example 3. Hence, we conclude the proof. $\square$

Remark 14. When considering ρ -weakly convex function f , we can use proximal subdifferentials (Definition 7) instead of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials. Then $J_{\gamma}(x) = \partial_{\rho}(\frac{1}{2\gamma}\|x\|^2) = \frac{x}{\gamma}$ as the function is differentiable convex. Therefore,

$$x \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0) = (J_{\gamma} + \partial_{\rho} f)^{-1} J_{\gamma}(x_0) \Leftrightarrow \frac{x_0 - x}{\gamma} \in \partial_{\rho} f(x).$$

By [10, Proposition 5] with $\varepsilon = 0$, we have

$$x \in \arg \min_{z \in X} \left[f(z) + \frac{1}{2\gamma} \|z - x_0\|^2 \right] \Rightarrow \frac{x_0 - x}{\gamma} \in \partial_{\rho} f(x),$$

and the equivalence holds when $2\gamma\rho \leq 1$.

Assertion (5.9) shows that for every $\gamma > 0$, the proximal operator $\text{prox}_{\gamma f}$ from (5.5) is an element of $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator of f .

Example 14. We consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x| + x^2$ which is 2-strongly convex. We calculate the elements of $\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x)$ for any $x \in \mathbb{R}$. Firstly, we compute $\partial_{lsc}^{\mathbb{R}} f(x)$.

- When $x = 0$, we need to find $(a, u) \in \Phi_{lsc}^{\mathbb{R}}$ such that

$$(\forall y \in \mathbb{R}) \quad |y| + y^2 \geq -ay^2 + uy. \quad (5.14)$$

We divide into three cases of $y \in \mathbb{R}$. When $y = 0$, (5.14) holds for all $(a, u) \in \Phi_{lsc}^{\mathbb{R}}$. When $y > 0$, we have

$$(a + 1)y + 1 - u \geq 0,$$

for $a \geq -1$ and $u \leq 1$. When $y < 0$, we have

$$(a + 1)y - (1 + u) \leq 0,$$

for $a \geq -1$ and $u \geq -1$. Hence, $\partial_{lsc}^{\mathbb{R}} f(0) = \{(a, u) : a \geq -1, -1 \leq u \leq 1\}$.

- When $x > 0$, we have

$$(\forall y \in \mathbb{R}) \quad |y| + y^2 - x - x^2 \geq -a(y^2 - x^2) + u(y - x). \quad (5.15)$$

When $y \geq 0$, we simplify (5.15)

$$(a + 1)(y^2 - x^2) \geq (u - 1)(y - x).$$

If $a = -1$ then $u = 1$, otherwise, we have

$$(a + 1) \left[y - x - \frac{u - 1 - 2(a + 1)x}{2(a + 1)} \right]^2 \geq \frac{[u - 1 - 2(a + 1)x]^2}{4(a + 1)},$$

which implies $a > -1$ and $u = 1 + 2(a + 1)x$. When $y < 0$, we have

$$(a + 1)(y^2 - x^2) \geq (u + 1)(y - x).$$

Since $y - x \neq 0$, we can further have

$$(a + 1)(y + x) \leq (u + 1).$$

As $y < 0$, we need $a \geq -1$ and $u \geq -1 + (a + 1)x$. Hence, for $x > 0$,

$$\partial_{isc}^{\mathbb{R}} f(x) = \{(a, u) : a \geq -1, u = 1 + 2(a + 1)x\}.$$

- When $x < 0$, similar to the previous case, we obtain

$$\partial_{isc}^{\mathbb{R}} f(x) = \{(a, u) : a \geq -1, u = -1 + 2(a + 1)x\}.$$

In conclusion,

$$\partial_{isc}^{\mathbb{R}} f(x) = \left\{ (a, u) \in \Phi_{isc}^{\mathbb{R}} : \begin{cases} a \geq -1, u = -1 + 2(a + 1)x & \text{when } x < 0 \\ a \geq -1, u = 1 + 2(a + 1)x & \text{when } x > 0 \\ a \geq -1, -1 \leq u \leq 1 & \text{when } x = 0 \end{cases} \right\}. \quad (5.16)$$

Let us take $x_0 \in X$, for a given $(a_0, u_0) \in J_{\gamma}(x_0)$, $x \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0)$ means there exists $(a, u) \in J_{\gamma}(x)$ such that

$$(a_0 - a, u_0 - u) \in \partial_{isc}^{\mathbb{R}} f(x).$$

By (5.16), we consider the following cases

- If $x < 0$ then a has to satisfy $a_0 - a \geq -1$ and

$$u_0 - u = -1 + 2(a_0 - a + 1)x.$$

We substitute u_0 and u from Example 3, and arrive at

$$x = \frac{\left(\frac{1}{\gamma} + 2a_0\right)x_0 + 1}{\left(\frac{1}{\gamma} + 2a_0 + 1\right)}.$$

- If $x > 0$, we also have a satisfies $a_0 - a \geq -1$ and

$$x = \frac{\left(\frac{1}{\gamma} + 2a_0\right)x_0 - 1}{\left(\frac{1}{\gamma} + 2a_0 + 1\right)}.$$

- If $x = 0$ then $1 \leq u_0 - u \leq 1$ or

$$-\frac{\gamma}{1 + 2\gamma a_0} \leq x_0 \leq \frac{\gamma}{1 + 2\gamma a_0}.$$

Notice that when $2\gamma a_0 = -1$ then $u_0 = 0$ which also satisfies the above inequality.

Hence, we have

$$\text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_0) = \begin{cases} \frac{\left(\frac{1}{\gamma} + 2a_0\right)x_0 + 1}{\left(\frac{1}{\gamma} + 2a_0 + 1\right)} & \text{when } \left(\frac{1}{\gamma} + 2a_0\right)x_0 < -1, \ 2\gamma a_0 \geq -1, \\ \frac{\left(\frac{1}{\gamma} + 2a_0\right)x_0 - 1}{\left(\frac{1}{\gamma} + 2a_0 + 1\right)} & \text{when } \left(\frac{1}{\gamma} + 2a_0\right)x_0 > 1, \ 2\gamma a_0 \geq -1, \\ 0 & \text{otherwise.} \end{cases}$$

5.2 $\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point Algorithm

5.2.1 Auxiliary Convergence Results

This subsection presents the auxiliary convergence results which will be used in convergence proofs of $\Phi_{lsc}^{\mathbb{R}}$ -proximal point algorithm introduced in Section 5.2 and $\Phi_{lsc}^{\mathbb{R}}$ -forward-backward algorithm introduced in Section 5.3. These auxiliary results are based on Fejér and quasi-Fejér monotonicity of the iterate (cf. [6, Chapter 5]).

Lemma 5 (Lemma 3.1, [35]). *Let $\chi \in (0, 1]$, $(\alpha_n)_{n \in \mathbb{N}}, (\beta_n)_{n \in \mathbb{N}}, (\varepsilon_n)_{n \in \mathbb{N}}$ be non-negative sequences with $\sum_{n \in \mathbb{N}} \varepsilon_n < +\infty$ such that*

$$\alpha_{n+1} \leq \chi \alpha_n - \beta_n + \varepsilon_n. \quad (5.17)$$

Then

- (i) $(\alpha_n)_{n \in \mathbb{N}}$ is bounded and converges.
- (ii) $(\beta_n)_{n \in \mathbb{N}}$ is summable.
- (iii) If $\chi < 1$ then $(\alpha_n)_{n \in \mathbb{N}}$ is summable.

Theorem 12. Let $h : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\text{dom } h$, $(\alpha_n)_{n \in \mathbb{N}}, (\beta_n)_{n \in \mathbb{N}}$ be positive sequences on the real line. Assuming the following holds for all $n \in \mathbb{N}$

$$\alpha_{n+1} \|x^* - x_{n+1}\|^2 \leq \alpha_n \|x^* - x_n\|^2 - \beta_n \|x_n - x_{n+1}\|^2, \quad (5.18)$$

for some $x^* \in S = \text{argmin } h \neq \emptyset$. The following holds

- (i) $(\alpha_n \|x^* - x_n\|^2)_{n \in \mathbb{N}}$ converges.
- (ii) $(\alpha_n \text{dist}^2(x_n, S))_{n \in \mathbb{N}}$ is decreasing and converges.
- (iii) $\sum_{n \in \mathbb{N}} \beta_n \|x_n - x_{n+1}\|^2 < +\infty$. If $0 < \beta \leq \beta_n$ for all $n \in \mathbb{N}$ then $\sum_{n \in \mathbb{N}} \|x_n - x_{n+1}\|^2 < +\infty$.
- (iv) If $\sum_{n \in \mathbb{N}} 1/\sqrt{\beta_n} < +\infty$ then $\sum_{n \in \mathbb{N}} \|x_n - x_{n+1}\| < +\infty$.

If α_n is non-decreasing then

- (v) $(\|x^* - x_n\|^2)_{n \in \mathbb{N}}$ converges.
- (vi) $(\text{dist}^2(x_n, S))_{n \in \mathbb{N}}$ is decreasing and converges.

Proof. We can consider $\alpha_n \|x^* - x_n\|^2$ as $\tilde{\alpha}_n$ and $\beta_n \|x_n - x_{n+1}\|^2$ as $\tilde{\beta}_n$ and apply Lemma 5 with $\varepsilon_n = 0, \chi = 1$ to arrive at (i) and by Lemma 5-(ii)

$$\sum_{n \in \mathbb{N}} \beta_n \|x_n - x_{n+1}\|^2 < +\infty.$$

It follows from (5.18) for all $n \in \mathbb{N}$, we have

$$\alpha_{n+1}\|x^* - x_{n+1}\|^2 \leq \alpha_n\|x^* - x_n\|^2. \quad (5.19)$$

As α_n does not depend on $x^* \in S$ for all $n \in \mathbb{N}$, (ii) holds by taking the infimum with respect to $x^* \in S$ on both sides of (5.19).

From (5.18), we know that

$$\begin{aligned} \alpha_{n+1}\|x^* - x_{n+1}\|^2 + \beta\|x_n - x_{n+1}\|^2 &\leq \alpha_{n+1}\|x^* - x_{n+1}\|^2 + \beta\|x_n - x_{n+1}\|^2 \\ &\leq \alpha_n\|x^* - x_n\|^2, \end{aligned}$$

which is

$$\beta\|x_n - x_{n+1}\|^2 \leq \alpha_n\|x^* - x_n\|^2 - \alpha_{n+1}\|x^* - x_{n+1}\|^2.$$

Summing the above inequality from $n = 0$ to $N \in \mathbb{N}$ we get

$$\beta \sum_{n=0}^N \|x_n - x_{n+1}\|^2 \leq \alpha_0\|x^* - x_0\|^2 - \alpha_{N+1}\|x^* - x_{N+1}\|^2 \leq \alpha_0\|x^* - x_0\|^2,$$

and letting N go to infinity, we have (iii). To show (iv), we infer from (5.18),

$$\|x_n - x_{n+1}\| \leq \sqrt{\frac{\alpha_n}{\beta_n}}\|x^* - x_n\| \leq \sqrt{\frac{\alpha_{n-1}}{\beta_n}}\|x^* - x_{n-1}\| \leq \sqrt{\frac{\alpha_0}{\beta_n}}\|x^* - x_0\|. \quad (5.20)$$

Thanks to the assumption that $\sum_{n \in \mathbb{N}} \frac{1}{\sqrt{\beta_n}} < +\infty$, $\|x_n - x_{n+1}\|$ is summable.

Let us assume that α_n is non-decreasing, since $\alpha_n > 0$, we can divide both sides of (5.19) by α_{n+1} to obtain

$$\|x^* - x_{n+1}\|^2 \leq \frac{\alpha_n}{\alpha_{n+1}}\|x^* - x_n\|^2 \leq \|x^* - x_n\|^2. \quad (5.21)$$

This proves (v) and also (vi).

□

5.2.2 $\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point Algorithm

Let X be a Hilbert space, we assume that the function $f : X \rightarrow (\infty, +\infty]$ is proper $\Phi_{lsc}^{\mathbb{R}}$ -convex, for all $x \in \text{dom } f$. Let us further assume f has a global minimizer i.e. $S = \text{argmin}_{x \in X} f(x)$ is non-empty. We aim to solve

$$\min_{x \in X} f(x) \quad (5.22)$$

We introduce $\Phi_{lsc}^{\mathbb{R}}$ -proximal point Algorithm ($\Phi_{lsc}^{\mathbb{R}}$ -PPA), starting with $x_0 \in \text{dom } f$ and stepsize $\gamma > 0$, is defined as follows

$$x_{n+1} \in \text{prox}_{\gamma f}^{\text{lsc}, \mathbb{R}}(x_n) = (J_{\gamma} + \hat{\partial}_{lsc}^{\mathbb{R}} f)^{-1} J_{\gamma}(x_n). \quad (\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$$

According to (5.4), the following condition must be satisfied for the iteration scheme ($\Phi_{lsc}^{\mathbb{R}}$ -PPA) to be well-defined,

$$(\forall n \in \mathbb{N}) \ J_{\gamma}(x_n) \cap \text{ran}(J_{\gamma} + \hat{\partial}_{lsc}^{\mathbb{R}} f) \neq \emptyset. \quad (5.23)$$

Consider the situation where a is fixed and $2\gamma a \geq -1$. Then $\phi_n = (a, u_n) \in J_{\gamma}(x_n)$ and $\phi_{n+1} = (a, u_{n+1}) \in J_{\gamma}(x_{n+1})$. If for all $n \in \mathbb{N}$,

$$\phi_n - \phi_{n+1} = (0, u_n - u_{n+1}) = \left(0, \left(\frac{1}{\gamma} + 2a\right)x_n - \left(\frac{1}{\gamma} + 2a\right)x_{n+1}\right) \in \hat{\partial}_{lsc}^{\mathbb{R}} f(x_{n+1}), \quad (5.24)$$

(the second equality come from Example 3), then

$$\left(\frac{1}{\gamma} + 2a\right)(x_n - x_{n+1}) \in \hat{\partial}_{lsc}^0 f(x_{n+1}),$$

where $\hat{\partial}_{lsc}^0 f$ is a subgradient of f in the sense of convex analysis.

Remark 15. When f is ρ -weakly convex, we can use proximal subdifferentials (Definition 7) instead of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials and $J_{\gamma}(x) = \partial_{\rho}(\frac{1}{2\gamma}\|x\|^2) =$

$\frac{x}{\gamma}$ as it is differentiable convex. Then, for $n \in \mathbb{N}$,

$$x_{n+1} \in (J_{\gamma} + \partial_{\rho} f)^{-1} J_{\gamma}(x_n) \Leftrightarrow \frac{x_n - x_{n+1}}{\gamma} \in \partial_{\rho} f(x_{n+1}),$$

which is exactly the iteration of proximal point method for weakly function f [53, Section 5].

Taking into account the content of Theorem 11, we can propose a realizable version of $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$ in Algorithm 1 to solve problem (5.22).

Algorithm 1: $\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point Algorithm

(i) **Initialize:** $\gamma > 0$, $x_0 \in \text{dom } f$

(ii) **For** $n \geq 0$, **update**

- Pick $(a_n, (\frac{1}{\gamma} + 2a_n)x_n) \in J_{\gamma}(x_n)$ according to (5.1)
 - $x_{n+1} \in \underset{z \in X}{\operatorname{argmin}} f(z) + \left(\frac{1}{2\gamma} + a_n\right) \|z - x_n\|^2$
-

Remark 16. Finding x_{n+1} above reduces to finding solution of a separate minimization problem. The complexity of this problem depends upon the properties of function f . It is not our aim here to discuss particular computational aspects of Algorithm 1. Instead, we are investigating the convergence properties of Algorithm 1.

We start with the result related to monotonicity of the objective function.

Proposition 12. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper, $\Phi_{lsc}^{\mathbb{R}}$ -convex function, and (5.23) hold. Let $(x_n)_{n \in \mathbb{N}}$ and $(a_n)_{n \in \mathbb{N}}$ be sequences generated by*

$(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$ with $\gamma > 0$. Then it holds for all $y \in X$ that

$$\begin{aligned} f(y) - f(x_{n+1}) &\geq \left(\frac{1}{2\gamma} + a_n \right) (\|x_{n+1} - x_n\|^2 - \|y - x_n\|^2) \\ &\quad + \left(\frac{1}{2\gamma} + a_{n+1} \right) \|y - x_{n+1}\|^2. \end{aligned} \quad (5.25)$$

In particular, $(f(x_n))_{n \in \mathbb{N}}$ is decreasing.

Proof. By (5.23), the sequence $(x_n)_{n \in \mathbb{N}}$ is well-defined, i.e. there exist $\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right)x_n\right) \in J_\gamma(x_n)$ and $\left(a_{n+1}, \left(\frac{1}{\gamma} + 2a_{n+1}\right)x_{n+1}\right) \in J_\gamma(x_{n+1})$ such that

$$\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right)x_n\right) - \left(a_{n+1}, \left(\frac{1}{\gamma} + 2a_{n+1}\right)x_{n+1}\right) \in \partial_{lsc}^{\mathbb{R}} f(x_{n+1}), \quad (5.26)$$

with the conditions $2\gamma a_n \geq -1$ and $2\gamma a_{n+1} \geq -1$. By definition of $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f at x_{n+1} , we have

$$\begin{aligned} (\forall y \in X) \quad f(y) - f(x_{n+1}) &\geq -(a_n - a_{n+1}) (\|y\|^2 - \|x_{n+1}\|^2) \\ &\quad + \left\langle \frac{1}{\gamma} (x_n - x_{n+1}) + 2(a_n x_n - a_{n+1} x_{n+1}), y - x_{n+1} \right\rangle \\ &= \frac{1}{\gamma} \langle x_n - x_{n+1}, y - x_{n+1} \rangle + 2 \langle a_n x_n - a_{n+1} x_{n+1}, y - x_{n+1} \rangle \\ &\quad - (a_n - a_{n+1}) (\|y\|^2 - \|x_{n+1}\|^2). \end{aligned} \quad (5.27)$$

The last two terms on the right hand side can be simplified to

$$\begin{aligned} &2 \langle a_n x_n - a_{n+1} x_{n+1}, y - x_{n+1} \rangle - (a_n - a_{n+1}) (\|y\|^2 - \|x_{n+1}\|^2) \\ &= 2a_n \langle x_n - x_{n+1}, y - x_{n+1} \rangle + (a_{n+1} - a_n) \|y - x_{n+1}\|^2. \end{aligned} \quad (5.28)$$

Plugging (5.28) back into (5.27) we obtain

$$\begin{aligned}
f(y) - f(x_{n+1}) &\geq \left(\frac{1}{\gamma} + 2a_n\right) \langle x_n - x_{n+1}, y - x_{n+1} \rangle + (a_{n+1} - a_n) \|y - x_{n+1}\|^2 \\
&= \left(\frac{1}{2\gamma} + a_n + a_{n+1} - a_n\right) \|y - x_{n+1}\|^2 + \left(\frac{1}{2\gamma} + a_n\right) \|x_{n+1} - x_n\|^2 \\
&\quad - \left(\frac{1}{2\gamma} + a_n\right) \|y - x_n\|^2 \\
&= \left(\frac{1}{2\gamma} + a_{n+1}\right) \|y - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n\right) \|y - x_n\|^2 \\
&\quad + \left(\frac{1}{2\gamma} + a_n\right) \|x_{n+1} - x_n\|^2, \tag{5.29}
\end{aligned}$$

which proves (5.25). By taking $y = x_n$, we obtain

$$f(x_n) - f(x_{n+1}) \geq \left(\frac{1}{\gamma} + a_{n+1} + a_n\right) \|x_n - x_{n+1}\|^2 \geq 0,$$

as $\frac{1}{\gamma} + a_{n+1} + a_n \geq 0$ from the conditions of a_n and a_{n+1} . Therefore, $(f(x_n))_{n \in \mathbb{N}}$ is decreasing. $\square$

Proposition 12 provides a description of the behavior of the objective function at each iterate. In the following result, the role of coefficients a_n in the convergence results is investigated. Depending on the subdifferentials of f , the behavior of $(a_n)_{n \in \mathbb{N}}$ can be divided into two cases. Below, we present the main convergence result of $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$.

Theorem 13. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function with the set $S = \operatorname{argmin}_{x \in X} f(x)$ nonempty. Assuming (5.23) holds. Let $(x_n)_{n \in \mathbb{N}}, (a_n)_{n \in \mathbb{N}}$ be sequences generated by $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$ with $\gamma > 0$, then the following hold.*

- (i) *If there exists $n_0 \in \mathbb{N}$ such that $\frac{1}{2\gamma} + a_{n_0} = 0$, x_{n_0+1} is a global minimizer of f .*

(ii) If $\frac{1}{2\gamma} + a_n > 0$ for all $n \in \mathbb{N}$, we have $\lim_{n \rightarrow \infty} f(x_n) = f(x^*)$, where $x^* \in S$. The assertions (i)-(iv) of Theorem 12 hold with $\alpha_n = \beta_n = \frac{1}{2\gamma} + a_n$, with $n \in \mathbb{N}$. Moreover, if $(a_n)_{n \in \mathbb{N}}$ is non-decreasing, then we also have (v)-(vi) of Theorem 12 i.e.

(v) $(\|x^* - x_n\|^2)_{n \in \mathbb{N}}$ converges.

(vi) $(\text{dist}^2(x_n, S))_{n \in \mathbb{N}}$ is decreasing and converges.

Proof. Thanks to assumption (5.23), $(x_n)_{n \in \mathbb{N}}$ is well-defined. In case (i), there exists an $n_0 \in \mathbb{N}$ such that $\frac{1}{2\gamma} + a_{n_0} = 0$, then at iteration $n_0 + 1$, by Proposition 12,

$$(\forall y \in X) \quad f(y) - f(x_{n_0+1}) \geq \left(\frac{1}{2\gamma} + a_{n_0+1} \right) \|y - x_{n_0+1}\|^2 \geq 0. \quad (5.30)$$

This means x_{n_0+1} is the global minimizer, so we can stop the algorithm after $n_0 + 1$ steps no matter if $(\frac{1}{2\gamma} + a_{n_0+1})$ equals to zero or not.

In case (ii), from inequality (5.25) in Proposition 12, by taking $y = x^* \in S$, we obtain

$$\begin{aligned} 0 \geq f(x^*) - f(x_{n+1}) &\geq \left(\frac{1}{2\gamma} + a_{n+1} \right) \|x^* - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n \right) \|x^* - x_n\|^2 \\ &\quad + \left(\frac{1}{2\gamma} + a_n \right) \|x_{n+1} - x_n\|^2. \end{aligned} \quad (5.31)$$

which coincides with Theorem 12 inequality (5.18) for $\alpha_n = \beta_n = \frac{1}{2\gamma} + a_n > 0$ for all $n \in \mathbb{N}$. Notice that the monotonicity of α_n depends on the monotonicity of a_n . When $(a_n)_{n \in \mathbb{N}}$ is non-decreasing, Theorem 12-(v,vi) hold for the function $h = f$.

For the limit of the objective function, using the second inequality of

(5.31) and Theorem 12-(i), we can skip the last term to arrive at

$$f(x^*) - f(x_{n+1}) \geq \left(\frac{1}{2\gamma} + a_{n+1} \right) \|x^* - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n \right) \|x^* - x_n\|^2. \quad (5.32)$$

Taking the limit on both sides, by Theorem 12-(i), the right hand side converges to zero while the left hand side is non-positive, which implies that

$$\lim_{n \rightarrow \infty} f(x_{n+1}) = f(x^*).$$

□

Remark 17. The $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials of f influence the behavior of $(a_n)_{n \in \mathbb{N}}$ in $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$. If for any $(a, u) \in \partial_{lsc}^{\mathbb{R}} f(x)$, $a > 0$ for all $x \in \text{dom } f$, then by (5.26), at each iteration, we must have $a_n - a_{n+1} > 0$ so $(a_n)_{n \in \mathbb{N}}$ is decreasing and consequently α_n is decreasing. Then the assertions (v,vi,vii) of Theorem 13 will not be met. For example, the function $f(x) = -x^2$ has $\partial_{lsc}^{\mathbb{R}} f(x) = \{(a, u) \in \Phi_{lsc}^{\mathbb{R}} : a \geq 1, u = 2(a-1)x\}$ for all $x \in \mathbb{R}$.

If α_n is bounded from below by a positive constant in $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$, then Theorem 12-(iii) implies $\|x_n - x_{n+1}\|^2$ is summable and $(x_n)_{n \in \mathbb{N}}$ is bounded.

To see that $(x_n)_{n \in \mathbb{N}}$ is bounded, let us fix $x^* \in S$ and let $A \leq \alpha_n$ for all $n \in \mathbb{N}$ for some constant $A > 0$. By contradiction, assume that $(x_n)_{n \in \mathbb{N}}$ is unbounded i.e. there exists a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ such that

$$Ak^2 \leq \alpha_{n_k} \|x_{n_k} - x^*\|^2, \forall k \in \mathbb{N}.$$

By Theorem 12-(i), $(\alpha_n \|x_n - x^*\|^2)_{n \in \mathbb{N}}$ converges. Let $\delta = \lim_{n \rightarrow \infty} \alpha_n \|x_n - x^*\|^2$, then

$$|Ak^2 - \delta| \leq |\alpha_{n_k} \|x_{n_k} - x^*\|^2 - \delta|, \forall k \in \mathbb{N}.$$

The left hand side tends to infinity while the right hand side goes to zero which leads to a contradiction.

Since x_n is bounded, there exists a weakly convergent subsequence of $(x_n)_{n \in \mathbb{N}}$. However, we do not know if a weak limit point of $(x_n)_{n \in \mathbb{N}}$ lies in S . This is because we only know that the function f is lsc, thanks to $\Phi_{lsc}^{\mathbb{R}}$ -convexity which does not mean that f is weak lsc as, in general, it is not convex.

Remark 18. Theorem 13-(i) provides a stopping criterion for $(\Phi_{lsc}^{\mathbb{R}}\text{-PPA})$. As stated in Theorem 10, a minimizer is also a fixed point of the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator. However, as the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator is a set-valued operator, it can return a fixed point and not a solution (see Theorem 10).

Example 15. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a proper lsc ρ -weakly convex. We analyse Algorithm 1 with $a_n, n \in \mathbb{N}$ and stepsize $\gamma > 0$ chosen such that

$$\frac{1}{2\gamma} + a_n > \frac{\rho}{2}. \quad (5.33)$$

Let us observe that the only constraint on a_n is provided by the J_γ in (5.1) which is $2\gamma a_n \geq -1$. This is satisfied when we choose a_n according to (5.33). Moreover, (5.33) ensures us the uniqueness of x_{n+1} in the algorithm. By Theorem 12, by choosing $\beta_n = \frac{1}{2\gamma} + a_n$, formula (5.18) is satisfied in view of Proposition 12. Hence, by Theorem 12, $((1/2\gamma + a_n)\|x^* - x_n\|^2)_{n \in \mathbb{N}}$ converges for any $x^* \in S$. In consequence, $(x_n)_{n \in \mathbb{N}}$ is bounded for a suitable choice of a_n (for example, a_n is lower bounded). Hence, $(x_n)_{n \in \mathbb{N}}$ contains a convergent subsequence $(x_{n_k})_{k \in \mathbb{N}}$ with the limit $\bar{x}$. By the lower semi-continuity of f , we have $f(\bar{x}) = \inf f(x)$ which proves that for suitable choice of γ and a_n , the sequence $(x_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 contains a subsequence converging to a global solution.

5.3 $\Phi_{lsc}^{\mathbb{R}}$ -Forward Backward Algorithm

Inspired by the construction of $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator in (5.4), now we go a step further by considering the problem

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

which can be approached by finding a point $x_0 \in X$ such that

$$(0, 0) \in (\partial_{lsc}^{\mathbb{R}} f + \partial_{lsc}^{\mathbb{R}} g)(x_0) \subseteq \partial_{lsc}^{\mathbb{R}}(f + g)(x_0), \quad (5.34)$$

where the functions $f, g : X \rightarrow (-\infty, +\infty]$ are proper $\Phi_{lsc}^{\mathbb{R}}$ -convex and $\text{dom } f \cap \text{dom } g \neq \emptyset$. Moreover, let us assume that the set $S = \text{argmin}_{x \in X}(f + g)(x)$ is non-empty.

For any $x_0 \in \text{dom } f \cap \text{dom } g$, we define the following update

$$x_{n+1} \in (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1} (J_{\gamma} - \partial_{lsc}^{\mathbb{R}} g)(x_n). \quad (\Phi_{lsc}^{\mathbb{R}}\text{-FB})$$

To ensure that the iterate in $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ is well-defined, we assume the following condition

$$(\forall n \in \mathbb{N}) \quad (J_{\gamma} - \partial_{lsc}^{\mathbb{R}} g)(x_n) \cap \text{ran}(J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f) \neq \emptyset. \quad (5.35)$$

We refer to algorithm $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ as $\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward Algorithm.

Similar to Theorem 11, one can interpret $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ as follow.

Theorem 14. *Let X be a Hilbert space and $f, g : X \rightarrow (-\infty, +\infty]$ be proper $\Phi_{lsc}^{\mathbb{R}}$ -convex functions. Let $x_0 \in \text{dom } f, \gamma > 0$. If*

$$x \in (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1} (J_{\gamma} - \partial_{lsc}^{\mathbb{R}} g)(x_0), \quad (5.36)$$

then there exists $(a_0, (1/\gamma + 2a_0)x_0) \in J_\gamma(x_0)$ and $(a_0^g, u_0^g) \in \partial_{lsc}^{\mathbb{R}} g(x_0)$ such that

$$x \in \operatorname{argmin}_{y \in X} f(y) + \langle u_0^g - 2a_0^g x_0, y \rangle + \left(\frac{1}{2\gamma} + a_0 - a_0^g \right) \|y - x_0\|^2.$$

Proof. Let (5.36) hold, it means there exist

$$\left(a_0, \left(\frac{1}{\gamma} + 2a_0 \right) x_0 \right) \in J_\gamma(x_0), \quad \left(a, \left(\frac{1}{\gamma} + 2a \right) x \right) \in J_\gamma(x), \quad (a_0^g, u_0^g) \in \partial_{lsc}^{\mathbb{R}} g(x_0),$$

such that

$$\left(a_0, \left(\frac{1}{\gamma} + 2a_0 \right) x_0 \right) - \left(a, \left(\frac{1}{\gamma} + 2a \right) x \right) - (a_0^g, u_0^g) \in \partial_{lsc}^{\mathbb{R}} f(x).$$

$\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials gives us, for all $y \in X$,

$$\begin{aligned} & f(y) - f(x) \\ & \geq - (a_0 - a - a_0^g) (\|y\|^2 - \|x\|^2) + \left\langle \left(\frac{1}{\gamma} + 2a_0 \right) x_0 - \left(\frac{1}{\gamma} + 2a \right) x - u_0^g, y - x \right\rangle \\ & = - (a_0 - a - a_0^g) \|y - x\|^2 + \left\langle \left(\frac{1}{\gamma} + 2a_0 \right) (x_0 - x) - u_0^g + 2a_0^g x, y - x \right\rangle \\ & = \left(\frac{1}{2\gamma} + a + a_0^g \right) \|y - x\|^2 + \left(\frac{1}{2\gamma} + a_0 \right) [\|x_0 - x\|^2 - \|y - x_0\|^2] \\ & \quad + \langle 2a_0^g x_0 - u_0^g, y - x \rangle + 2a_0^g \langle x - x_0, y - x \rangle \\ & = \left(\frac{1}{2\gamma} + a \right) \|y - x\|^2 + \left(\frac{1}{2\gamma} + a_0 - a_0^g \right) [\|x_0 - x\|^2 - \|y - x_0\|^2] \\ & \quad - \langle u_0^g - 2a_0^g x_0, y - x \rangle. \end{aligned}$$

As $\left(a, \left(\frac{1}{\gamma} + 2a \right) x \right) \in J_\gamma(x)$, we have $2\gamma a \geq -1$ which infers

$$\begin{aligned} f(y) - f(x) & \geq - \left(\frac{1}{2\gamma} + a_0 - a_0^g \right) [\|y - x_0\|^2 + \|x_0 - x\|^2] \\ & \quad - \langle u_0^g - 2a_0^g x_0, y - x \rangle. \end{aligned}$$

or

$$x \in \operatorname{argmin}_{y \in X} f(y) + \langle u_0^g - 2a_0^g x_0, y \rangle + \left(\frac{1}{2\gamma} + a_0 - a_0^g \right) \|y - x_0\|^2.$$

□

From Theorem 14, we propose Algorithm 2 which is one way to implement $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$.

Algorithm 2: $\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward Algorithm

- (i) **Initialize:** $\gamma > 0$, $x_0 \in \operatorname{dom} f$ and
 $\left(a_0, \left(\frac{1}{\gamma} + 2a_0 \right) x_0 \right) \in J_\gamma(x_0), (a_0^g, u_0^g) \in \partial_{lsc}^{\mathbb{R}} g(x_0)$
 - (ii) **For** $n > 0$, **update**
 - Pick $\left(a_n, \left(\frac{1}{\gamma} + 2a_n \right) x_n \right) \in J_\gamma(x_n), (a_n^g, u_n^g) \in \partial_{lsc}^{\mathbb{R}} g(x_n)$
 - $x_{n+1} \in \operatorname{argmin}_{z \in X} f(z) + \langle u_n^g - 2a_n^g x_n, z \rangle + \left(\frac{1}{2\gamma} + a_n - a_n^g \right) \|z - x_n\|^2$
-

We start with the following technical fact.

Proposition 13. *Let $g : X \rightarrow (-\infty, +\infty]$ be proper $\Phi_{lsc}^{\mathbb{R}}$ -convex. If $(J_\gamma - \partial_{lsc}^{\mathbb{R}} g) \subseteq \operatorname{dom} J_\gamma^{-1}$, then*

$$\begin{aligned} (J_\gamma + \partial_{lsc}^{\mathbb{R}} f)^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} g) &\subset (J_\gamma + \partial_{lsc}^{\mathbb{R}} f)^{-1} J_\gamma J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} g) \\ &= \operatorname{prox}_{\gamma f}^{\operatorname{lsc}, \mathbb{R}} (J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} g)). \end{aligned} \quad (5.37)$$

Proof. Let $x \in \operatorname{dom} f \cap \operatorname{dom} g$ take $\phi \in (J_\gamma - \partial_{lsc}^{\mathbb{R}} g)(x)$. Then by assumption, $\phi \in \operatorname{dom} J_\gamma^{-1}$ which implies $\phi \in J_\gamma J_\gamma^{-1}(\phi)$ and

$$\phi \in J_\gamma J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} g)(x).$$

Then it is obvious that

$$(J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1} (J_{\gamma} - \partial_{lsc}^{\mathbb{R}} g) \subset (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} f)^{-1} J_{\gamma} J_{\gamma}^{-1} (J_{\gamma} - \partial_{lsc}^{\mathbb{R}} g).$$

□

Let us consider the function g to be Fréchet differentiable on the whole domain with Lipschitz continuous gradient with Lipschitz constant $L_g > 0$. We have an interesting relationship between $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials and gradient of g .

Lemma 6. *Let $g : X \rightarrow (-\infty, +\infty]$ be Fréchet-differentiable on X with Lipschitz continuous gradient with Lipschitz constant $L_g > 0$. Then $\partial_{lsc}^{\mathbb{R}} g(x) \neq \emptyset$ for all $x \in X$. Moreover, for $x \in X$, any $(a, u) \in \partial_{lsc}^{\mathbb{R}} g(x)$ satisfies $a + \frac{L_g}{2} \geq 0$ and $u - 2ax = \nabla g(x)$.*

Proof. By [88, Proposition 3.6] and [11, Proposition 2] which infers $\partial_{lsc}^{\geq} g(x) \neq \emptyset$ for all $x \in X$. By Proposition 2, $\Phi_{lsc}^{\geq} \subset \Phi_{lsc}^{\mathbb{R}}$, so $\partial_{lsc}^{\mathbb{R}} g(x) \neq \emptyset$ for all $x \in X$. For the second assertion, let $x \in X$. Since g is Fréchet-differentiable at $x \in X$ with Lipschitz continuous gradient, we have, for all $y \in X$

$$\frac{L_g}{2} \|y - x\|^2 + \langle \nabla g(x), y - x \rangle \geq g(y) - g(x), \quad (5.38)$$

which is the well-known descent Lemma [6, Lemma 2.64].

On the other hand, by taking $(a, u) \in \partial_{lsc}^{\mathbb{R}} g(x)$, we have

$$(\forall y \in X) \quad g(y) - g(x) \geq -a (\|y\|^2 - \|x\|^2) + \langle u, y - x \rangle. \quad (5.39)$$

Combining (5.39), (5.38) and separating x and y ,

$$\left(\frac{L_g}{2} + a \right) \|y\|^2 + \langle \nabla g(x) - L_g x - u, y \rangle \geq \left(\frac{L_g}{2} + a \right) \|x\|^2 + \langle \nabla g(x) - L_g x - u, x \rangle.$$

As this holds for all $y \in X$, x is the global minimizer of the function

$$h(z) = \left(\frac{L_g}{2} + a \right) \|z\|^2 + \langle \nabla g(x) - L_g x - u, z \rangle,$$

which is quadratic. This implies that $\frac{L_g}{2} + a \geq 0$ and

$$\nabla h(x) = 0 \Leftrightarrow u - 2ax = \nabla g(x).$$

Since $(a, u) \in \partial_{isc}^{\mathbb{R}} g(x)$ is taken arbitrarily, the above inequality holds for all the elements in $\partial_{isc}^{\mathbb{R}} g(x)$. $\square$

With the result obtained in Lemma 6, we can improve Algorithm 2 with the following Algorithm 3.

Algorithm 3: $\Phi_{isc}^{\mathbb{R}}$ -Forward-Backward Algorithm with g as in Lemma 6

- (i) **Initialize:** $\gamma > 0$, $x_0 \in \text{dom } f$ and $\left(a_0, \left(\frac{1}{\gamma} + 2a_0\right)x_0\right) \in J_{\gamma}(x_0)$, $(a_0^g, u_0^g) \in \partial_{isc}^{\mathbb{R}} g(x_0)$
 - (ii) **For** $n > 0$, **update**
 - Pick $\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right)x_n\right) \in J_{\gamma}(x_n)$, $(a_n^g, u_n^g) \in \partial_{isc}^{\mathbb{R}} g(x_n)$
 - $x_{n+1} \in \underset{z \in X}{\operatorname{argmin}} \ f(z) + \langle \nabla g(x_n), z \rangle + \left(\frac{1}{2\gamma} + a_n - a_n^g\right) \|z - x_n\|^2$
-

Below, we give the estimation for the behavior of the objective functions for $(\Phi_{isc}^{\mathbb{R}}\text{-FB})$.

Proposition 14. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{isc}^{\mathbb{R}}$ -convex function and $g : X \rightarrow (-\infty, +\infty]$ be a proper Fréchet differentiable on X with Lipschitz continuous gradient $L_g > 0$. Let (5.35) hold and $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by $(\Phi_{isc}^{\mathbb{R}}\text{-FB})$ with stepsize $\gamma > 0$. For all $y \in X$ and $n \in \mathbb{N}$, we*

have

$$\begin{aligned} (f+g)(y) - (f+g)(x_{n+1}) &\geq \left(\frac{1}{2\gamma} + a_{n+1}\right) \|y - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n\right) \|y - x_n\|^2 \\ &\quad + \left(\frac{1}{2\gamma} + a_n - a_n^g - \frac{L_g}{2}\right) \|x_{n+1} - x_n\|^2, \end{aligned} \quad (5.40)$$

where $\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right) x_n\right) \in J_\gamma(x_n)$, $\left(a_{n+1}, \left(\frac{1}{\gamma} + 2a_{n+1}\right) x_{n+1}\right) \in J_\gamma(x_{n+1})$, and $(a_n^g, u_n^g) \in \partial_{lsc}^{\mathbb{R}} g(x_n)$. Moreover, if $\frac{1}{\gamma} + a_n + a_{n+1} \geq a_n^g + \frac{L_g}{2}$, then $(f+g)(x_n) \geq (f+g)(x_{n+1})$.

Proof. By the definition of the updated in $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$, there exist $\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right) x_n\right) \in J_\gamma(x_n)$, $\left(a_{n+1}, \left(\frac{1}{\gamma} + 2a_{n+1}\right) x_{n+1}\right) \in J_\gamma(x_{n+1})$ and $(a_n^g, u_n^g) \in \partial_{lsc}^{\mathbb{R}} g(x_n)$ with $a_n^g + \frac{L_g}{2} \geq 0$ such that

$$\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right) x_n\right) - \left(a_{n+1}, \left(\frac{1}{\gamma} + 2a_{n+1}\right) x_{n+1}\right) - (a_n^g, u_n^g) \in \partial_{lsc}^{\mathbb{R}} f(x_{n+1}). \quad (5.41)$$

Using the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials of f , we have, for all $y \in X$

$$\begin{aligned} f(y) - f(x_{n+1}) &\geq -(a_n - a_n^g - a_{n+1}) (\|y\|^2 - \|x_{n+1}\|^2) \\ &\quad + \left\langle \left(\frac{1}{\gamma} + 2a_n\right) x_n - u_n^g - \left(\frac{1}{\gamma} + 2a_{n+1}\right) x_{n+1}, y - x_{n+1} \right\rangle. \end{aligned}$$

We proceed similarly as in Proposition 12 to obtain

$$\begin{aligned} f(y) - f(x_{n+1}) &\geq \left(\frac{1}{2\gamma} + a_{n+1}\right) \|y - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n\right) \|y - x_n\|^2 \\ &\quad + \left(\frac{1}{2\gamma} + a_n\right) \|x_{n+1} - x_n\|^2 \\ &\quad + a_n^g (\|y\|^2 - \|x_{n+1}\|^2) - \langle u_n^g, y - x_{n+1} \rangle. \end{aligned} \quad (5.42)$$

Let us focus on the last terms on the right hand side of (5.42)

$$\begin{aligned}
& -a_n^g (\|y\|^2 - \|x_{n+1}\|^2) + \langle u_n^g, y - x_{n+1} \rangle \\
& = -a_n^g (\|y\|^2 - \|x_n\|^2) + \langle u_n^g, y - x_n \rangle \\
& \quad + a_n^g (\|x_{n+1}\|^2 - \|x_n\|^2) + \langle u_n^g, x_n - x_{n+1} \rangle \\
& \leq g(y) - g(x_n) + a_n^g (\|x_{n+1}\|^2 - \|x_n\|^2) + \langle u_n^g, x_n - x_{n+1} \rangle \\
& = g(y) - g(x_n) + a_n^g \|x_{n+1} - x_n\|^2 \\
& \quad + \langle u_n^g - 2a_n^g x_n, x_n - x_{n+1} \rangle.
\end{aligned} \tag{5.43}$$

Plugging (5.43) back into (5.42) we get

$$\begin{aligned}
(f+g)(y) - f(x_{n+1}) - g(x_n) & \geq \left(\frac{1}{2\gamma} + a_{n+1} \right) \|y - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n \right) \|y - x_n\|^2 \\
& \quad + \left(\frac{1}{2\gamma} + a_n - a_n^g \right) \|x_{n+1} - x_n\|^2 \\
& \quad - \langle u_n^g - 2a_n^g x_n, x_n - x_{n+1} \rangle.
\end{aligned} \tag{5.44}$$

On the other hand, since g is Fréchet differentiable with Lipschitz continuous gradient, we apply Lemma 6

$$\begin{aligned}
\langle u_n^g - 2a_n^g x_n, x_n - x_{n+1} \rangle & = \langle \nabla g(x_n), x_n - x_{n+1} \rangle \\
& \leq g(x_n) - g(x_{n+1}) + \frac{L_g}{2} \|x_{n+1} - x_n\|^2.
\end{aligned} \tag{5.45}$$

Putting (5.45) back into (5.44) to obtain

$$\begin{aligned}
(f+g)(y) - (f+g)(x_{n+1}) & \geq \left(\frac{1}{2\gamma} + a_{n+1} \right) \|y - x_{n+1}\|^2 - \left(\frac{1}{2\gamma} + a_n \right) \|y - x_n\|^2 \\
& \quad + \left(\frac{1}{2\gamma} + a_n - a_n^g - \frac{L_g}{2} \right) \|x_{n+1} - x_n\|^2,
\end{aligned}$$

which is (5.40). On the other hand, letting $y = x_n$ in (5.40),

$$(f + g)(x_n) - (f + g)(x_{n+1}) \geq \left(\frac{1}{\gamma} + a_{n+1} + a_n - a_n^g - \frac{L_g}{2} \right) \|x_n - x_{n+1}\|^2. \quad (5.46)$$

The RHS of (5.46) is non-negative when $\frac{1}{\gamma} + a_n + a_{n+1} \geq a_n^g + \frac{L_g}{2}$. This conclude the proof. $\square$

The results obtained in Proposition 14 are similar to the one obtained for the Forward-Backward in [11, Corollary 3]. In analogy to Proposition 12, Proposition 14 gives us a crucial estimate of the $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ Algorithm which contributes to the convergence result below.

Theorem 15. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function and $g : X \rightarrow (-\infty, +\infty]$ be a proper Fréchet differentiable function on X with Lipschitz continuous gradient with Lipschitz constant $L_g > 0$. Let $(x_n)_{n \in \mathbb{N}}$, $(a_n)_{n \in \mathbb{N}}$, and $(a_n^g)_{n \in \mathbb{N}}$ be sequences generated by $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ with stepsize $\gamma > 0$. Assume that $\frac{1}{\gamma} + a_n + a_{n+1} \geq a_n^g + \frac{L_g}{2}$ for all $n \in \mathbb{N}$ where*

$$\left(a_n, \left(\frac{1}{\gamma} + 2a_n \right) x_n \right) \in J_\gamma(x_n), \quad (a_n^g, u_n^g) \in \partial_{lsc}^{\mathbb{R}} g(x_n).$$

We have the following

- (i) *If there exists $n_0 \in \mathbb{N}$ such that $\frac{1}{2\gamma} + a_{n_0} = a_{n_0}^g + \frac{L_g}{2} = 0$ then x_{n_0+1} is the global minimizer.*
- (ii) *If $\alpha_n = \frac{1}{2\gamma} + a_n > 0$ and $\beta_n = \frac{1}{2\gamma} + a_n - a_n^g - \frac{L_g}{2} > 0$ for all $n \in \mathbb{N}$, Theorem 12 holds and $\lim_{n \rightarrow \infty} (f + g)(x_n) = \inf_{x \in X} (f + g)(x)$.*

Proof. The proof follows in the same manner as in Theorem 12. $\square$

5.4 $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient

We saw in the previous section that the $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ Algorithm can be written with the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator. In this section, we investigate the case of the $\Phi_{lsc}^{\mathbb{R}}$ -proximity operator of the indicator function of a closed convex set. This is equivalent to solving the constrained problem

$$\min_{x \in C} f(x), \quad (5.47)$$

where $C \subset X$ is a closed convex set in Hilbert space and $f : X \rightarrow (-\infty, +\infty]$ is proper $\Phi_{lsc}^{\mathbb{R}}$ -convex. We can rewrite this problem in the form

$$\min_{x \in X} f(x) + \iota_C(x).$$

To solve (5.47), we propose $\Phi_{lsc}^{\mathbb{R}}$ -projected subgradient algorithm which is formally given in Algorithm 4.

Algorithm 4: $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient Algorithm

Initialize: $x_0 \in \text{dom } f$

Set: $(\gamma_n)_{n \in \mathbb{N}}$ positive

Compute: $\left(a_n, \left(\frac{1}{\gamma_n} + 2a_n\right)x_n\right) \in J_{\gamma_n}(x_n)$ and $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$
such that

$$2\gamma_n(a_n - a_n^f) > -1$$

Update: $x_{n+1} = \text{Proj}_C \left(\frac{(1 + 2\gamma_n a_n)x_n - \gamma_n u_n^f}{1 + 2\gamma_n(a_n - a_n^f)} \right)$

Return: x_{n+1}

Further details about the algorithm and the variable stepsize γ_n are given in (5.58) and in Theorem 16 below. In the following proposition, we show that the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator of indicator function of C coincides with the projection onto C without convexity assumption.

Proposition 15. *Let C be a closed subset of X with more than one element*

and let $x \in X, \gamma > 0$. If $(-\frac{1}{2\gamma}, 0) \notin J_{\gamma}(x) \cap \text{ran}(J_{\gamma} + \partial_{lsc}^{\mathbb{R}}\iota_C)$ then we have

$$(J_{\gamma} + \partial_{lsc}^{\mathbb{R}}\iota_C)^{-1} J_{\gamma}(x) = \text{Proj}_C(x).$$

Proof. Recall that

$$J_{\gamma}(x) \cap \text{ran}(J_{\gamma} + \partial_{lsc}^{\mathbb{R}}\iota_C) = \{\phi \in J_{\gamma}(x) : \exists x^+ \in X \text{ s.t. } \phi \in J_{\gamma}(x^+) + \partial_{lsc}^{\mathbb{R}}\iota_C(x^+)\}. \quad (5.48)$$

Let $x^+ \in (J_{\gamma} + \partial_{lsc}^{\mathbb{R}}\iota_C)^{-1} J_{\gamma}(x)$. By assumption, (5.48) implies that there exist $(a, (1/\gamma + 2a)x) \in J_{\gamma}(x)$ and $(a^+, (1/\gamma + 2a^+)x^+) \in J_{\gamma}(x^+)$ such that

$$\left(a - a^+, \left(\frac{1}{\gamma} + 2a\right)x - \left(\frac{1}{\gamma} + 2a^+\right)x^+\right) \in \partial_{lsc}^{\mathbb{R}}\iota_C(x^+), \quad (5.49)$$

with $2\gamma a > -1$, $2\gamma a^+ \geq -1$. If $x^+ \notin C$ then $\partial_{lsc}^{\mathbb{R}}\iota_C(x^+) = \emptyset$, which contradicts the above inclusion. Hence, it must be $x^+ \in C$. By the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, for any $y \in X$, we have

$$\begin{aligned} \iota_C(y) - \iota_C(x^+) &\geq -(a - a^+) \left(\|y\|^2 - \|x^+\|^2\right) \\ &\quad + \left\langle \left(\frac{1}{\gamma} + 2a\right)x - \left(\frac{1}{\gamma} + 2a^+\right)x^+, y - x^+ \right\rangle. \end{aligned} \quad (5.50)$$

By taking $y \in C$, (5.50) reads

$$(a - a^+) \left(\|y\|^2 - \|x^+\|^2\right) \geq \left\langle \left(\frac{1}{\gamma} + 2a\right)x - \left(\frac{1}{\gamma} + 2a^+\right)x^+, y - x^+ \right\rangle. \quad (5.51)$$

Further simplifying (5.51)

$$(a - a^+) \|y - x^+\|^2 \geq \left(\frac{1}{\gamma} + 2a\right) \langle x - x^+, y - x^+ \rangle. \quad (5.52)$$

From (5.52), we have

$$(a - a^+) \|y - x^+\|^2 \geq \left(\frac{1}{2\gamma} + a\right) \left[\|y - x^+\|^2 + \|x - x^+\|^2 - \|y - x\|^2 \right],$$

so that

$$\left(\frac{1}{2\gamma} + a\right) \|y - x\|^2 \geq \left(\frac{1}{2\gamma} + a^+\right) \|y - x^+\|^2 + \left(\frac{1}{2\gamma} + a\right) \|x - x^+\|^2.$$

As $2\gamma a > -1$, $2\gamma a^+ \geq -1$, all the coefficients are nonnegative, this infers

$$\left(\frac{1}{2\gamma} + a\right) \|y - x\|^2 \geq \left(\frac{1}{2\gamma} + a\right) \|x - x^+\|^2, \quad (5.53)$$

for all $y \in C$. By (5.49), $x^+ \in \text{Proj}_C(x)$.

On the other hand, let us assume that $x^+ \in \text{Proj}_C(x)$, we have

$$(\forall y \in C) \quad \|y - x\| \geq \|x^+ - x\|. \quad (5.54)$$

Taking square both sides of (5.54) and multiply with $\frac{1}{2\gamma} + a > 0$ for some $a \in \mathbb{R}$, we obtain

$$\left(\frac{1}{2\gamma} + a\right) \|y - x\|^2 \geq \left(\frac{1}{2\gamma} + a\right) \|x^+ - x\|^2.$$

This is equivalent to

$$0 \geq -\left(\frac{1}{2\gamma} + a\right) (\|y\|^2 - \|x^+\|^2) + \left\langle \left(\frac{1}{\gamma} + 2a\right) x, y - x^+ \right\rangle. \quad (5.55)$$

Because $x^+ \in C$, (5.55) implies that $\left(\frac{1}{2\gamma} + a, \left(\frac{1}{\gamma} + 2a\right)x\right) \in \partial_{lsc}^{\mathbb{R}} \iota_C(x^+)$. We

can write

$$\begin{aligned} \left(\frac{1}{2\gamma} + a, \left(\frac{1}{\gamma} + 2a \right) x \right) &= \left(a - \left(-\frac{1}{2\gamma} \right), \left(\frac{1}{\gamma} + 2a \right) x - \left(\frac{1}{\gamma} - \frac{1}{\gamma} \right) x^+ \right) \\ &\in J_{\gamma}(x) - J_{\gamma}(x^+). \end{aligned}$$

This infers $\left(a, \left(\frac{1}{\gamma} + 2a \right) x \right) \in (\partial_{lsc}^{\mathbb{R}} \iota_C + J_{\gamma})(x^+)$, so that $x^+ \in \text{prox}_{\gamma \iota_C}^{\text{lsc}, \mathbb{R}}(x)$. $\square$

Remark 19. The assumption $(-\frac{1}{2\gamma}, 0) \notin J_{\gamma}(x) \cap \text{ran}(J_{\gamma} + \partial_{lsc}^{\mathbb{R}} \iota_C)$ also implies the uniqueness of the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator of an indicator function. Assume that there are $x_1, x_2 \in (J_{\gamma} + \partial_{lsc}^{\mathbb{R}} \iota_C)^{-1} J_{\gamma}(x)$. Following (5.53) in the proof of Proposition 15, we have

$$\|x - x_2\|^2 \geq \|x - x_1\|^2 \geq \|x - x_2\|^2,$$

so $x_1 = x_2$.

Let us go back to problem (5.47). Since C is closed, the function ι_C is lsc and so it is $\Phi_{lsc}^{\mathbb{R}}$ -convex (by virtue of [82, Proposition 6.3]). Motivated by $(\Phi_{lsc}^{\mathbb{R}}\text{-FB})$ and Proposition 15, we propose $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient Algorithm,

$$x_{n+1} \in \text{Proj}_C \left(J_{\gamma}^{-1} \left(J_{\gamma} - \partial_{lsc}^{\mathbb{R}} f \right) (x_n) \right). \quad (\Phi_{lsc}^{\mathbb{R}}\text{-PSG})$$

Observe that if $\text{ran} J_{\gamma} \cap \text{ran}(J_{\gamma} - \partial_{lsc}^{\mathbb{R}} f) \neq \emptyset$, Algorithm $\Phi_{lsc}^{\mathbb{R}}$ -PSG is well-defined i.e. for every $n \in \mathbb{N}$, there exists x_{n+1} such that $(\Phi_{lsc}^{\mathbb{R}}\text{-PSG})$ holds.

In fact, $\text{ran} J_{\gamma} \cap \text{ran}(J_{\gamma} - \partial_{lsc}^{\mathbb{R}} f)$ is always nonempty. Let us prove this by taking $x \in \text{dom } \partial_{lsc}^{\mathbb{R}} f$, $(a_f, u_f) \in \partial_{lsc}^{\mathbb{R}} f(x)$ and $(a, (\frac{1}{\gamma} + 2a)x) \in J_{\gamma}(x)$ with $\gamma > 0$ and $2a\gamma \geq -1$. Then

$$J_{\gamma}^{-1} \left(a - a_f, \left(\frac{1}{\gamma} + 2a \right) x - u_f \right) \subseteq J_{\gamma}^{-1} \left(J_{\gamma} - \partial_{lsc}^{\mathbb{R}} f \right) (x). \quad (5.56)$$

By formula (3.17) in Example 3, the domain of J_{γ}^{-1} is nonempty when $a -$

$a_f \geq -1/(2\gamma)$. While $-1/(2\gamma) \leq a$, we can take a large enough so that $a - a_f > -1/(2\gamma)$. Again, by formula (3.17), J_γ^{-1} is a single valued map and we obtain

$$J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} f) (x) = \frac{\left(\frac{1}{\gamma} + 2a\right) x - u^f}{\frac{1}{\gamma} + 2(a - a^f)} = \frac{(1 + 2\gamma a)x - \gamma u^f}{1 + 2\gamma(a - a^f)}.$$

On the other hand, when

$$a - a_f = -\frac{1}{2\gamma}, \text{ and } \left(\frac{1}{\gamma} + 2a\right) x - u_f = 0, \quad (5.57)$$

then by (3.17)

$$J_\gamma^{-1} \left(a - a_f, \left(\frac{1}{\gamma} + 2a\right) x - u_f \right) = X.$$

Moreover, (5.57) implies that x is a $a + 1/(2\gamma)$ -critical point of f (see Definition 5).

With the well-defined inner operator of $(\Phi_{lsc}^{\mathbb{R}}\text{-PSG})$, we have

$$J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} f) (x_n) = \frac{\left(\frac{1}{\gamma} + 2a_n\right) x_n - u_n^f}{\frac{1}{\gamma} + 2(a_n - a_n^f)} = \frac{(1 + 2\gamma a_n)x_n - \gamma u_n^f}{1 + 2\gamma(a_n - a_n^f)},$$

with $\gamma > 0, 2\gamma(a_n - a_n^f) > -1$. $(\Phi_{lsc}^{\mathbb{R}}\text{-PSG})$ takes an explicit form

$$x_{n+1} \in \text{Proj}_C (J_\gamma^{-1} (J_\gamma - \partial_{lsc}^{\mathbb{R}} f) (x_n)) = \text{Proj}_C \left(\frac{(1 + 2\gamma a_n)x_n - \gamma u_n^f}{1 + 2\gamma(a_n - a_n^f)} \right). \quad (5.58)$$

Let us state an estimate related to the objective function of the $\Phi_{lsc}^{\mathbb{R}}$ -PSG Algorithm to problem (5.47) with the stepsize $\gamma > 0$ being kept constant.

Lemma 7. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function and C be a closed convex set. Let $(x_n)_{n \in \mathbb{N}}$ be the sequence generated by (5.58), then*

the following holds for any $x \in C$

$$\begin{aligned} (1 + 2\gamma(a_n - a_n^f))\|x - x_{n+1}\|^2 &\leq (1 + 2\gamma a_n)\|x - x_n\|^2 + \frac{\gamma^2 \|2a_n^f x_n - u_n^f\|^2}{1 + 2\gamma(a_n - a_n^f)} \\ &\quad + 2\gamma[f(x) - f(x_n)], \end{aligned} \quad (5.59)$$

where for any $n \in \mathbb{N}$

$$\left(a_n, \left(\frac{1}{\gamma} + 2a_n\right)x_n\right) \in J_\gamma(x_n), \quad (a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n), \quad 2\gamma(a_n - a_n^f) > -1.$$

Proof. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence from (5.58) and $x \in C$ so that $x = \text{Proj}_C(x)$. We consider

$$\|x - x_{n+1}\|^2 = \left\| \text{Proj}_C(x) - \text{Proj}_C\left(\frac{(1 + 2\gamma a_n)x_n - \gamma u_n^f}{1 + 2\gamma(a_n - a_n^f)}\right) \right\|^2.$$

Since C is closed and convex, the projection operator Proj_C is nonexpansive. Together with (5.56), we continue the above equality

$$\begin{aligned} \|x - x_{n+1}\|^2 &\leq \left\| x - \frac{(1 + 2\gamma a_n)x_n - \gamma u_n^f}{1 + 2\gamma(a_n - a_n^f)} \right\|^2 = \left\| x - x_n - \frac{2\gamma a_n^f x_n - \gamma u_n^f}{1 + 2\gamma(a_n - a_n^f)} \right\|^2 \\ &= \|x - x_n\|^2 + \frac{\gamma^2}{\left(1 + 2\gamma(a_n - a_n^f)\right)^2} \|2a_n^f x_n - u_n^f\|^2 \\ &\quad + \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} \langle x - x_n, u_n^f - 2a_n^f x_n \rangle. \end{aligned}$$

By using the definition of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, we obtain

$$\begin{aligned}
\|x - x_{n+1}\|^2 &\leq \|x - x_n\|^2 + \frac{\gamma^2}{\left(1 + 2\gamma(a_n - a_n^f)\right)^2} \|2a_n^f x_n - u_n^f\|^2 \\
&\quad + \frac{2\gamma[f(x) - f(x_n) + a_n^f(\|x\|^2 - \|x_n\|^2) - 2a_n^f\langle x - x_n, x_n \rangle]}{1 + 2\gamma(a_n - a_n^f)} \\
&= \left(1 + \frac{2\gamma a_n^f}{1 + 2\gamma(a_n - a_n^f)}\right) \|x - x_n\|^2 + \frac{\gamma^2 \|2a_n^f x_n - u_n^f\|^2}{\left(1 + 2\gamma(a_n - a_n^f)\right)^2} \\
&\quad + \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} [f(x) - f(x_n)]. \tag{5.60}
\end{aligned}$$

From the assumption, $1 + 2\gamma(a_n - a_n^f) > 0$, we can further simplify (5.60) to obtain (5.59). $\square$

From Lemma 7, we notice that $\left(1 + \frac{2\gamma a_n^f}{1 + 2\gamma(a_n - a_n^f)}\right) \geq 0$ as $2\gamma(a_n - a_n^f) > -1$ from the assumption. When $a_n^f \leq 0$, then $\frac{1 + 2\gamma a_n}{1 + 2\gamma(a_n - a_n^f)} \leq 1$, from (5.59), we obtain

$$\begin{aligned}
\|x - x_{n+1}\|^2 &\leq \frac{1 + 2\gamma a_n}{1 + 2\gamma(a_n - a_n^f)} \|x - x_n\|^2 + \frac{2\gamma[f(x) - f(x_n)]}{1 + 2\gamma(a_n - a_n^f)} \\
&\quad + \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)}\right)^2 \|2a_n^f x_n - u_n^f\|^2 \\
&\leq \|x - x_n\|^2 + \frac{2\gamma[f(x) - f(x_n)]}{1 + 2\gamma(a_n - a_n^f)} \\
&\quad + \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)}\right)^2 \|2a_n^f x_n - u_n^f\|^2. \tag{5.61}
\end{aligned}$$

The above expression is similar to the ones obtained for the class of subgra-

dient descent algorithms in the convex case [17]. Let us denote S is the set of minimiser of problem (5.47). To proceed with the convergence analysis, we assume the followings.

Assumption 1. For $n \in \mathbb{N}$, $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$. Let us consider the following:

- (i) There exists a constant $U > 0$ such that $\|2a_n^f x_n - u_n^f\| \leq U$ for all $n \in \mathbb{N}$.
- (ii) For $\gamma > 0$, $\sum_{n \in \mathbb{N}} \frac{1}{[1+2\gamma(a_n - a_n^f)]^2} < +\infty$ and $\sum_{n \in \mathbb{N}} \frac{1}{1+2\gamma(a_n - a_n^f)} = +\infty$ where
 $(a_n, (\frac{1}{\gamma} + 2a_n)x_n) \in J_{\gamma}(x_n)$.
- (iii) For $(\gamma_n)_{n \in \mathbb{N}}$ positive, $\sum_{n \in \mathbb{N}} \frac{\gamma_n^2}{1+2\gamma_n(a_n - a_n^f)} < +\infty$ and $\sum_{n \in \mathbb{N}} \gamma_n = +\infty$ where
 $(a_n, (\frac{1}{\gamma_n} + 2a_n)x_n) \in J_{\gamma_n}(x_n)$.

Assumption 1-(i) ensures that the quantity $\|2a_n^f x_n - u_n^f\|$ is uniformly bounded. Assumption 1-(ii) with fixed stepsize is related to the convex subgradient method with square summable but not summable stepsize [1]. This ensures the weak convergence of the sequence and $\lim_{n \rightarrow \infty} f(x_n) = f(x^*)$. However, in our general setting of $\Phi_{lsc}^{\mathbb{R}}$ -subdifferentials, we obtain convergence results with Assumption 1-(ii) only when $a_n^f \leq 0$ for all $n \in \mathbb{N}$, $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$.

Proposition 16. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex function and C be a closed convex subset of X . Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by (5.58) with stepsize $\gamma > 0$. Let Assumption 1-(i,ii) hold and $a_n^f \leq 0$ for all $n \in \mathbb{N}$, $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$. Consider a point $x^* \in S$, then the followings hold*

- (i) $\|x^* - x_n\|^2$ converges and is bounded.
- (ii) $\text{dist}^2(x_n, S)$ is decreasing and converges.
- (iii) $\lim_{n \rightarrow \infty} f(x_n) = f(x^*)$.

Proof. Since $a_n^f \leq 0$ for all $n \in \mathbb{N}$, estimate (5.61) holds. Taking $x^* \in S$, we have $f(x^*) \leq f(x_n)$ for all $n \in \mathbb{N}$, and from (5.61), we obtain

$$\begin{aligned} \|x^* - x_{n+1}\|^2 &\leq \|x^* - x_n\|^2 + \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)} \right)^2 \|2a_n^f x_n - u_n^f\|^2 \\ &\quad + \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} [f(x^*) - f(x_n)] \\ &\leq \|x^* - x_n\|^2 + \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)} \right)^2 U^2. \end{aligned} \quad (5.62)$$

By Assumption 1-(ii), the last term on RHS of (5.62) is summable. By Lemma 5, the assertions (i) and (ii) are proved.

For (iii), we consider (5.61) again and by taking the finite sum till $N \in \mathbb{N}$, we obtain

$$\begin{aligned} &\sum_{n=0}^N \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} [f(x_n) - f(x^*)] \\ &\leq \sum_{n=0}^N [\|x^* - x_n\|^2 - \|x^* - x_{n+1}\|^2] + \sum_{n=0}^N \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)} \right)^2 U^2 \\ &= [\|x^* - x_0\|^2 - \|x^* - x_{N+1}\|^2] + \sum_{n=0}^N \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)} \right)^2 U^2 \\ &\leq \|x^* - x_0\|^2 + \sum_{n=0}^N \left(\frac{\gamma}{1 + 2\gamma(a_n - a_n^f)} \right)^2 U^2, \end{aligned} \quad (5.63)$$

Letting $N \rightarrow \infty$ and using Assumption 1-(ii), we obtain

$$\sum_{n=0}^{\infty} \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} [f(x_n) - f(x^*)] < +\infty.$$

Since $\sum_{n=0}^{\infty} \frac{2\gamma}{1 + 2\gamma(a_n - a_n^f)} = +\infty$, and $f(x_n) \geq f(x^*)$ for all $n \in \mathbb{N}$, we infer

$$\lim_{n \rightarrow \infty} f(x_n) - f(x^*) = 0. \quad (5.64)$$

□

Observe that we require $a_n^f \leq 0$ in order to obtain the convergence results. This is different to the previous Sections, where a part of the information of the next iterate is known i.e. a_{n+1} from $J_\gamma(x_{n+1})$. The results obtained in Proposition 16 are also similar to [1, Proposition 1 and Lemma 1]. In Proposition 16, we keep the stepsize γ constant so the adaptivity is transferred from the stepsize to a_n in $J_\gamma(x_n)$.

In general, we do not know the sign of a_n^f for $n \in \mathbb{N}$. Hence, to obtain convergence results for $(\Phi_{lsc}^{\mathbb{R}}\text{-PSG})$, we need to restrict the next element a_{n+1} in $J_\gamma(x_{n+1})$. The condition on a_{n+1} can be relaxed by changing the stepsize γ_n instead of keeping it constant.

Theorem 16. *Let $f : X \rightarrow (-\infty, +\infty]$ be proper $\Phi_{lsc}^{\mathbb{R}}$ -convex and C be a nonempty closed convex set. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by (5.58) with positive stepsize $(\gamma_n)_{n \in \mathbb{N}}$. Assume that Assumption 1-(i) and (iii) hold with*

$$(\forall n \in \mathbb{N}) \quad 2\gamma_n(a_n - a_n^f) \geq 2\gamma_{n+1}a_{n+1}, \quad (5.65)$$

where $(a_n, (1/\gamma_n + 2a_n)x_n) \in J_{\gamma_n}(x_n)$, $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$ and $2\gamma_n(a_n - a_n^f) > -1$. Consider a point $x^* \in S$. Then

- Theorem 12-(i, ii) hold for $\alpha_n = 1 + 2\gamma_n a_n$, $\beta_n = 0$ and

$$\liminf_{n \rightarrow \infty} f(x_n) - f(x^*) = 0.$$

- If α_n is non-decreasing then we have Theorem 12-(v,vi).

Proof. Let us take estimate (5.59) in Lemma 7 with $x = x^* \in S$ with assumption (5.65) to obtain

$$\begin{aligned}
 (1 + 2\gamma_{n+1}a_{n+1})\|x^* - x_{n+1}\|^2 &\leq (1 + 2\gamma_n a_n) \|x^* - x_n\|^2 + 2\gamma_n [f(x^*) - f(x_n)] \\
 &\quad + \frac{\gamma_n^2}{1 + 2\gamma_n(a_n - a_n^f)} \|2a_n^f x_n - u_n^f\|^2 \\
 &\leq (1 + 2\gamma_n a_n) \|x^* - x_n\|^2 + \frac{\gamma_n^2 U^2}{1 + 2\gamma_n(a_n - a_n^f)}.
 \end{aligned} \tag{5.66}$$

Since the last term on the right hand side is summable, Lemma 5 gives us statement (i)-(ii). For (iii), the proof follows along the same steps as in Proposition 16 combined with Assumption 1-(iii). Additionally, when α_n is non-decreasing, we divide (5.66) by $1 + 2\gamma_{n+1}a_{n+1}$ and obtain

$$\begin{aligned}
 \|x^* - x_{n+1}\|^2 &\leq \frac{1 + 2\gamma_n a_n}{1 + 2\gamma_{n+1}a_{n+1}} \|x^* - x_n\|^2 + \frac{\gamma_n^2 U^2}{(1 + 2\gamma_{n+1}a_{n+1})(1 + 2\gamma_n(a_n - a_n^f))} \\
 &\leq \|x^* - x_n\|^2 + \frac{\gamma_n^2}{1 + 2\gamma_n(a_n - a_n^f)} \frac{U^2}{1 + 2\gamma_0 a_0}.
 \end{aligned}$$

Combining the above inequality, which is analogous to inequality (5.62) of Proposition 16, with Assumption 1-(iii), we infer (iv) and (v). Lastly, we have that

$$\liminf_{n \rightarrow \infty} f(x_n) - f(x^*) = 0,$$

by following the same argument as in [1, Proposition 2-(i)]. $\square$

Remark 20. When γ is kept constant, the restriction (5.65) becomes $a_n - a_n^f \geq a_{n+1}$. Combining this with $2\gamma(a_n - a_n^f) > -1$ can cause early stopping

when $a_n^f \geq -1/2\gamma$. Because

$$-\frac{1}{2\gamma} \leq a_{n+1} \leq a_n - a_n^f \leq a_0 - \sum_{i=0}^n a_i^f,$$

so we need a_0 big enough to ensure the algorithm converges.

Remark 21. Instead of taking $x^* \in S$ as the global minimizer, we can choose a point in the set

$$\bar{S} := \{x \in C : f(x) \leq f(x_n), \forall n \in \mathbb{N}\}. \quad (5.67)$$

Then Proposition 15 and Theorem 16 still hold for $x^* \in \bar{S}$.

5.5 Numerical Examples

In this section, we give some numerical examples for the different algorithm that are analyzed in this paper.

$\Phi_{lsc}^{\mathbb{R}}$ -Proximal Point Algorithm: We continue Example 14 to illustrate the performance of the $\Phi_{lsc}^{\mathbb{R}}$ -proximal point Algorithm works. We use the starting point $x_0 = -10$ $a_0 = 1$ and fix the number of iteration to $N = 101$. Thanks to the closed form of $\Phi_{lsc}^{\mathbb{R}}$ -subgradient and $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator of f , we can set the update rule for a_n to

$$a_{n+1} = a_n + 0.9,$$

which satisfies the condition $a_n - a_{n+1} \geq -1$ of $\partial_{lsc}^{\mathbb{R}} f(x)$. The function f has a minimizer at $x = 0$ with $f(0) = 0$. We test for multiple values of $\gamma = 0.01, 0.1, 1, 10$ and plot the absolute value of the function value, the distance between two iterates and the distance of the iterate to the minimizer in Figure 5.1).

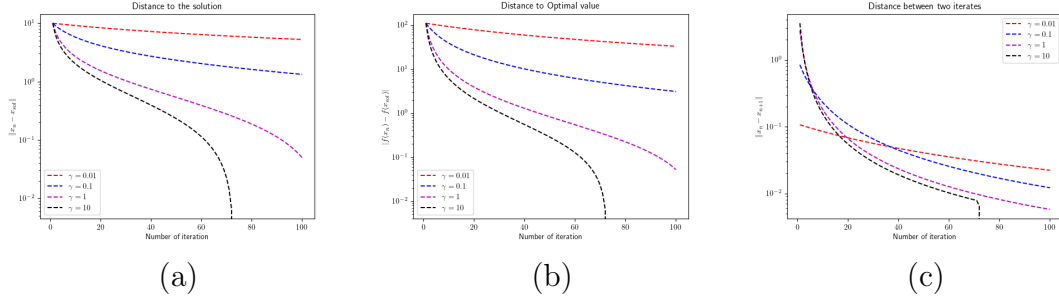


Figure 5.1: From left to right: distance between the current iteration and the minimizer; distance to the optimal value; distance between two successive iterates.

$\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient Algorithm: We illustrate Algorithm 4 by solving the following problem

$$\min_{x \in C} f(x) = \min_{x \in C} \langle x, Qx \rangle, \quad (5.68)$$

where $C = B(0, 1) \subset \mathbb{R}^n$ is the unit ball and $Q \in \mathbb{R}^{n \times n}$ is a full rank symmetric matrix. We can the problem as

$$\min_{x \in \mathbb{R}^n} \langle x, Qx \rangle + \iota_C(x),$$

In Examples 3 and 5, we derived expressions for J_γ , J_γ^{-1} and the $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f . We use (5.58) to state Algorithm ($\Phi_{lsc}^{\mathbb{R}}$ -PSG) in this case as

$$\begin{aligned} x_{n+1} &= \text{Proj}_C \left(\frac{\left[\left(\frac{1}{\gamma} + 2a_n - 2a_n^f \right) Id - 2Q \right] x_n}{\frac{1}{\gamma} + 2a_n - 2a_n^f} \right) \\ &= \text{Proj}_C \left(x_n - 2\gamma \frac{Qx_n}{1 + 2\gamma a_n - 2\gamma a_n^f} \right). \end{aligned}$$

Example 16. In the case where $C = B(0, 1)$, the problem (5.68) is equivalent

to

$$\min_{y \in B(0,1)} \langle y, Dy \rangle, \quad (5.69)$$

where $y = Px$ and D is a diagonal matrix of eigenvalues of Q by using matrix decomposition $Q = PDP^{-1}$. Thanks to Example 5, we have

$$\|y\| = \sqrt{\|y\|^2} = \sqrt{\langle Px, Px \rangle} = \|x\| \leq 1.$$

As D is a diagonal matrix, problem (5.69) is

$$\min_{y \in B(0,1)} \sum_{i=0}^N d_i y_i^2,$$

where d_i is the i -th element of diagonal matrix D . This allows us to check the distance between the current iterate and the solution of (5.69). We consider two scenarios of Assumption 1 where the stepsize is constant versus adaptive stepsize in order to show the early stopping in the former case. We give three different plots for each scenario: the optimal value of the objective function, the step length $|x_{n+1} - x_n|$ and the distance to the solution $|x_n - x^*|$.

(i) We consider the matrix

$$Q = \begin{bmatrix} -2 & 2 & 2 \\ 2 & 2 & -2 \\ 2 & -2 & 2 \end{bmatrix},$$

which has $(-4, 2, 4)$ as eigenvalues. Then for all the pair $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$, $a_n^f \geq 4$. We fix the maximum number of iteration $N = 101$, the starting point $x_0 = (-5, 5, -5)$.

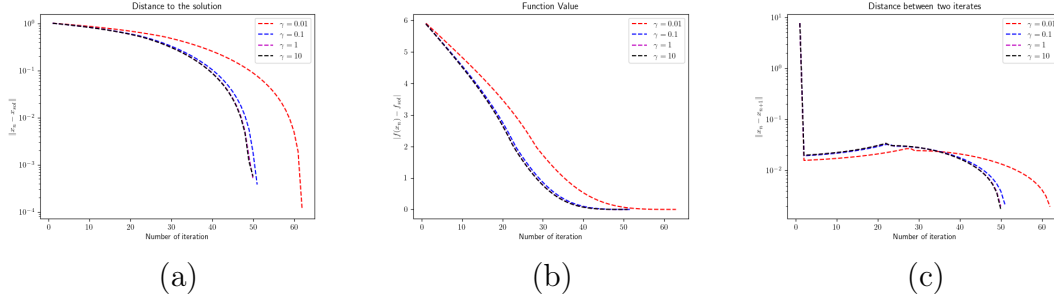


Figure 5.2: Example 16, case 1, constant stepsize. From left to right: distance between the current iteration and the solution; distance to the optimal value; distance between two successive iterates.

Constant Stepsize: We take $a_n^f = 4$ for all $n \in \mathbb{N}$. The initial value $a_0 = 200$ and $a_{n+1} = a_n - a_n^f$. The stopping criterion will be

$$a_{n+1} \leq a_{n+1}^f - \frac{1}{2\gamma} = 4 - \frac{1}{2\gamma}.$$

We test for multiple values of stepsize $\gamma = 0.01, 0.1, 1, 10$. Since a_n^f is positive for all $n \in \mathbb{N}$, the sequence a_n will be decreasing with the lower bound a_n^f . This is why we intentionally take a large initial a_0 , otherwise, if a_0 is closed to a_n^f then the algorithm may stop just after several iteration.

Figure 5.2 depicts the behaviors of the sequence $(x_n)_{n \in \mathbb{N}}$ and the function values. As we can see, for all values of γ , x_n and $f(x_n)$ tend to the optimal solution and optimal values, respectively. Observe that for $\gamma = 0.01$ the algorithm takes more time to get to the optimal solution and optimal function value compare to the larger values of γ . On the other hand, since we fixed the total number of steps $N = 101$, the algorithm stops before the number of iteration reach N . This behavior supports the observation mentioned in Remark 20.

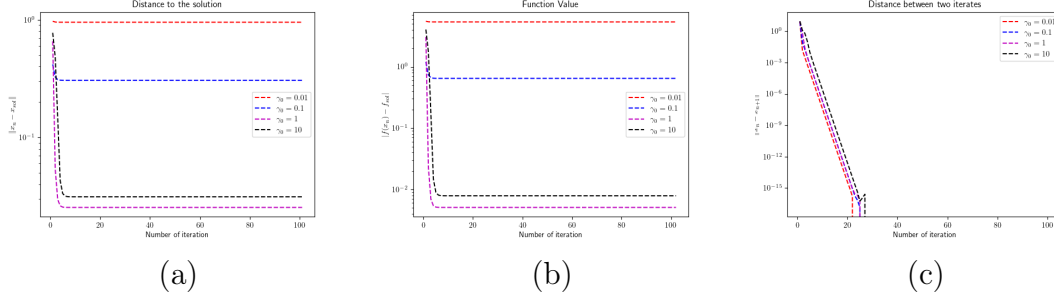


Figure 5.3: Example 16, case 1, adaptive stepsize. From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.

Adaptive Stepsize: With the same initial point and the same starting stepsize γ_0 as in the previous case, we fix $a_n^f = 4$ for all $n \in \mathbb{N}$, and set $N = 101$. We consider the following

$$(\forall n \in \mathbb{N}) \quad a_n = 5, \quad \gamma_{n+1} = \frac{\gamma_n}{a_{n+1}}(a_n - 4),$$

with the stopping criterion $2\gamma_n(a_n - a_n^f) \leq -1$. In this scenario, we can both fix a_n and a_n^f for all $n \in \mathbb{N}$. At each iteration, the value of the stepsize γ_n varies preventing early stopping of the algorithm. The plots are illustrated in Figure 5.3. The algorithm continues to the final iteration N and converges for all starting values of γ_0 . The number of iterations taken to arrive at the limit is less than the case with constant stepsize. However, notice that x_n converges to some point which is not the optimal solution of the problem. This happens as we keep a_n, a_n^f to be constants and the starting γ_0 is small. Hence γ_n has to go to zero, so x_n tends to stay at the same position. As we allow for bigger γ_0 , the sequence $(x_n)_{n \in \mathbb{N}}$ moves toward the solution.

(ii) Now we consider a matrix with two negative eigenvalues, namely

$$Q = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

which has $(-3, -1, 1, 2, 2)$ as eigenvalues. The lower bound for a_n^f where $(a_n^f, u_n^f) \in \partial_{lsc}^{\mathbb{R}} f(x_n)$, will be $a_n^f \geq 3$. We fix the maximum number of iterations to $N = 101$. We also consider constant stepsize and adaptive stepsize with the same starting γ_0 as in the first case. However, since there are two negative eigenvalues this time, we consider two different starting point $x_{01} = (-10, -10, -10, -10, -10)$ and $x_{02} = (-10, 10, -10, 10, -10)$.

Constant Stepsize: We keep the same setting as in the first case with $a_n^f = 3$ for all $n \in \mathbb{N}$. The initial value $a_0 = 200$ and $a_{n+1} = a_n - a_n^f$. The stopping criterion will be

$$a_{n+1} \leq a_{n+1}^f - \frac{1}{2\gamma} = 3 - \frac{1}{2\gamma}.$$

The plots for x_{01}, x_{02} are shown in Figure 5.4 and Figure 5.5, respectively. Notice that in this case, we have two different negative eigenvalues. The algorithm will converge to the direction of the closest eigenvector with respect to the negative eigenvalue. Here, the starting point x_{01} is actually closer to the eigenvector with respect to eigenvalue -1 while x_{02} is closer to the solution with respect to the eigenvalue -3 which is our true solution. This explains the different behaviors of the two graphs in Figure 5.4 and Figure 5.5.

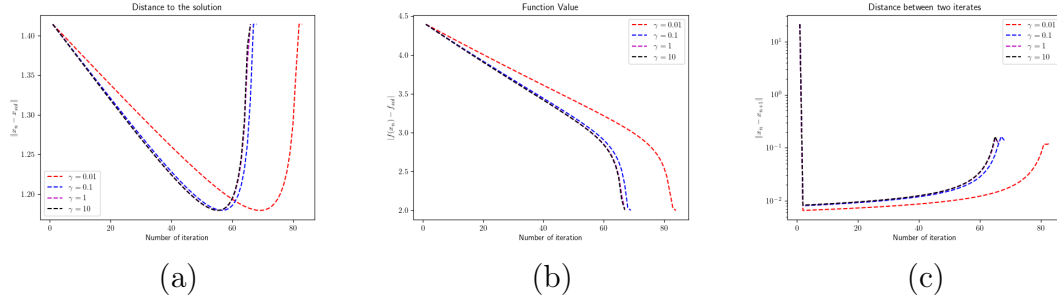


Figure 5.4: Example 16, case 2, constant stepsize, initialization x_{01} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.

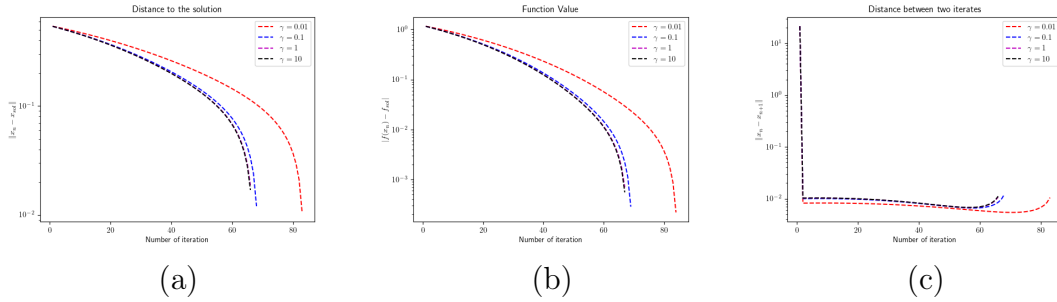


Figure 5.5: Example 16, case 2, constant stepsize, initialization x_{02} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.

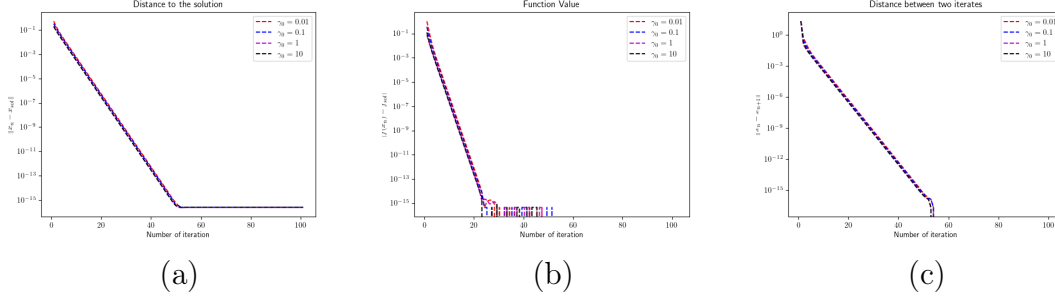


Figure 5.6: Example 16, case 2, adaptive stepsize, initialization x_{02} . From left to right: distance between the current iteration and the solution; function value at each iteration; distance between two successive iterates.

Adaptive Stepsize: We fix $a_n^f = 3$, maximum iteration $N = 101$. Compare to the previous case with adaptive stepsize, we use the following update with $\varepsilon = 1$

$$a_0 = 4, \quad a_n = -\frac{1}{2\gamma_n} + a_n^f + \varepsilon, \quad \gamma_{n+1} = \frac{\gamma_n(a_n - a_n^f) + 1}{a_n^f + \varepsilon}.$$

This rule ensures that $2\gamma_n(a_n - a_n^f) > -1$ for all $n \in \mathbb{N}$. We only run this case for starting point x_{02} , the plots are depicted in Figure 5.6. As both γ_n and a_n are changing, the sequence (x_n) converges to the solution for all cases of γ_0 comparing to Figure 5.3. It is obvious that as the scale of the problem increases, it takes more time for the sequence to converge to the solution.

Example 17. Instead of the quadratic function in (5.68), we consider a non-quadratic function which has one negative eigenvalue in its Hessian. Consider

$$f(x, y) = \frac{x^4}{12} + \frac{x^2}{2} - \frac{y^4}{12} - \frac{y^2}{2},$$

with the Hessian matrix

$$\nabla^2 f(x, y) = \begin{bmatrix} x^2 + 1 & 0 \\ 0 & -y^2 - 1 \end{bmatrix}.$$

It is difficult to calculate $\Phi_{lsc}^{\mathbb{R}}$ -subgradient of f , but we can use Lemma 6 as f has a Lipschitz continuous gradient. We compute $(a^f, u^f) \in \partial_{lsc}^{\mathbb{R}} f(x, y)$ coordinate-wise i.e.

$$\begin{aligned} a_x^f &= y^2 + 1 + \varepsilon, u_x^f = 2a_x^f x + \nabla_x f(x, y), \\ a_y^f &= y^2 + 1 + \varepsilon, u_y^f = 2a_y^f y + \nabla_y f(x, y). \end{aligned}$$

As the eigenvalues of the Hessian determines the convexity of f , we want to keep a_x^f, a_y^f above the smallest eigenvalues by some $\varepsilon > 0$. We use the same update rule for $a_{n+1} = a_n - a_n^f$ and the same stopping criterion. We test for constant stepsizes $\gamma = 0.01, 0.1, 1$.

We start the algorithm with initial $a_n = (200, 200), (x_0, y_0) = (-5, -1), \varepsilon = 0.1$ and max iteration $N = 1001$. Here we show only the function value and the distance to the solution in Figure 5.7.

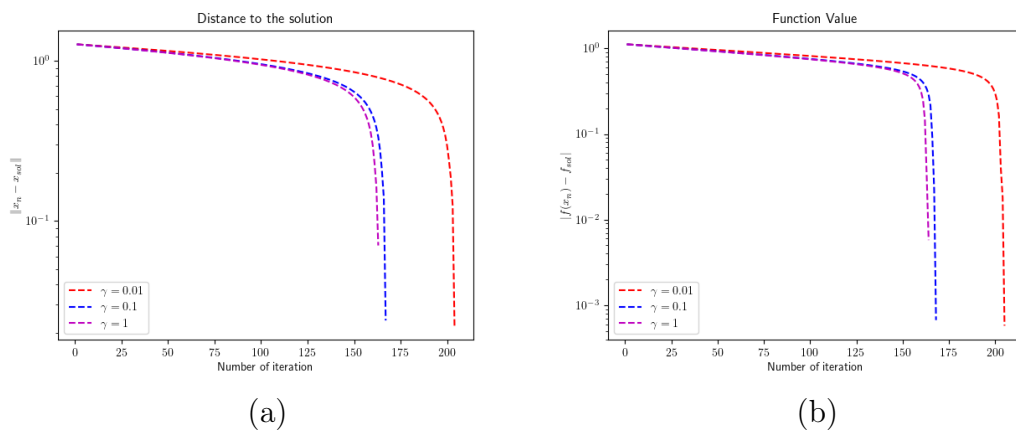


Figure 5.7: $\Phi_{lsc}^{\mathbb{R}}$ -(PSG) for Example 17. From left to right: distance between the current iterate and the solution; function value at each iteration.

Chapter 6

Primal-Dual algorithms for $\Phi_{lsc}^{\mathbb{R}}$ -Convex Functions

In this Chapter, we consider the following problem

$$\inf_{x \in X} f(x) + g(Lx), \quad (6.1)$$

where X, Y are Hilbert spaces, the functions $f : X \rightarrow (-\infty, +\infty]$ is proper lsc, $g : Y \rightarrow (-\infty, +\infty]$ is proper lsc and the linear operator $L : X \rightarrow Y$ is bounded. As we successfully constructed and investigated conjugate dual problem and Lagrange dual for (6.1) and studied duality in Chapter 4 using abstract convexity. Our next step would be to propose primal-dual algorithms which helps us solving problem (6.1) and its conjugate dual problem.

The outline of this chapter is the following:

- Section 6.1: The recollection of duality result for weakly convex functions and the concept of sharpness of the Lagrangian (Definition 11 and 12).
- Section 6.2: We propose Primal-Dual algorithms for composite minimization problem for weakly convex functions and investigate its convergence with the help of sharpness condition (Theorem 18 and 19).

- Section 6.3: We propose $\Phi_{lsc}^{\mathbb{R}}$ -primal-dual algorithm based on our result in the previous chapter and analyze the convergence (Theorem 20).
- Section 6.4: We conclude this chapter with some numerical examples of $\Phi_{lsc}^{\mathbb{R}}$ -primal-dual algorithm.

Our approach is based on the primal-dual algorithm developed by [25]. It was firstly proposed in [74, 46], by using splitting approach to find saddle point of the Lagrange primal dual problems. There has been numerous works and variants of primal-dual algorithm in the convex setting [64, 30, 24, 28, 27, 29, 97]. For examples, [51, 52] investigated the roles of stepsizes which affect the convergence of primal-dual algorithms. They also gave a convergence analysis with a more general choices of stepsizes for solving Total Variation. [26, 75] gave an improved analysis of primal-dual algorithm and studied its behavior when the computations are inexact. On the other hand, [66] combined primal-dual algorithm with linesearch approach to lessen computational effort compare to the original algorithm.

For nonconvex setting, there have been several attempt to investigate the convergence of primal-dual algorithm. [68] considered primal-dual algorithm for the sum of semiconvex (weakly convex) and strongly convex functions. [67] developed a block primal-dual splitting for nonconvex objective based on nonlinear primal-dual algorithm. They also derived parameters update and acceleration rules for their algorithm. [89] considered the same objective but in Banach space, they investigated the convergence of the algorithm using Bregman divergence. [33] applied primal-dual algorithm to the saddle point problem constructed from generalized conjugation.

6.1 Primal-Dual relationship of Composite Minimization Problem with weakly convex function

In this section, we consider the problem (6.1) in the case when the function f is weakly convex and g is convex. Let us recall that a function $f : X \rightarrow (-\infty, +\infty]$ is ρ -weakly convex if $f + \frac{\rho}{2} \|\cdot\|^2$ is convex (see Section 3.2). We would like to construct a conjugate dual problem for (6.1) from the results in Section 4.1. As shown in Section 3.2, a proper lsc ρ -weakly convex function is also abstract convex with respect to the class of elementary functions $\Phi_{lsc}^{\rho/2}$. Then it is natural to consider abstract conjugate with the respective class of elementary functions. For completeness, we give a definition of abstract conjugate with respect to the class $\Phi_{lsc}^{\rho/2}$ for weakly convex function. For a general version of abstract conjugate, we refer the reader to Definition 2.

Definition 10. Let Y be a Hilbert space, we define ρ_g -conjugate for ρ_g -weakly convex function $g : Y \rightarrow (-\infty, +\infty]$ as a function $g_{\rho_g}^* : Y \rightarrow \mathbb{R}$

$$g_{\rho_g}^*(u) = \sup_{y \in Y} \langle u, y \rangle - \frac{\rho_g}{2} \|y\|^2 - g(y). \quad (6.2)$$

The conjugate dual problem for f weakly convex and g convex has been mentioned in Example 8

$$\sup_{\psi \in \Psi_{conv}} -f_{\Phi_{lsc}^a}^*(-\psi \circ L) - g_{\Psi_{conv}}^*(\psi), \quad (6.3)$$

where Ψ_{conv} is the class of affine functions and Φ_{lsc}^a is the class of quadratic function. In Section 4.4, we show that the conjugate dual can be expressed as Lagrange dual with Lagrangian

$$\mathcal{L}(x, \psi) = f(x) + \psi(Lx) - g_{\Psi}^*(\psi).$$

For our specific setting, the Lagrangian takes the form

$$\mathcal{L}(x, \psi) = f(x) + \langle Lx, y \rangle - g^*(y).$$

Below we show that the fully weakly convex case (f, g weakly convex) can be reduced to our current setting.

Lemma 8. *Let X, Y be Hilbert spaces, consider the functions $f : X \rightarrow (-\infty, +\infty]$ be proper lsc ρ_f -weakly convex and $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc ρ_g -weakly convex, the problem (6.1)*

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

is equivalent to

$$\inf_{x \in X} F(x) + G(Lx), \quad (6.4)$$

where $F : X \rightarrow (-\infty, +\infty]$ is proper lsc weakly convex and $G : Y \rightarrow (-\infty, +\infty]$ is proper lsc convex. Moreover, the Lagrangian induced from the conjugate dual problems constructing from (6.1) and (6.4) have the same property which is and weakly convex in the primal variable and concave in the dual variable. The respective Lagrange primal problems are the same as the original problem (6.1) and (6.4).

Proof. Since g is ρ_g -weakly convex, we can turn (6.1) into

$$\inf_{x \in X} f(x) - \frac{\rho_g}{2} \|Lx\|^2 + \frac{\rho_g}{2} \|Lx\|^2 + g(Lx).$$

Denoting $F(x) = f(x) - \frac{\rho_g}{2} \|Lx\|^2$ and $G(y) = g(y) + \frac{\rho_g}{2} \|y\|^2$, we obtain (6.4) where F is $\rho_f + \rho_g \|L\|^2$ -weakly convex (see Proposition 8) and the function G is convex. Next, we construct the conjugate dual for (6.1) and (6.4) as in (4.26). The conjugate dual problem with respect to functions F and G is

$$\sup_{y \in Y} - \left(F - \frac{\rho_f + \rho_g \|L\|^2}{2} \|\cdot\|^2 \right)^*_{\rho_f + \rho_g \|L\|^2} (-L^*y) - G^*(y), \quad (6.5)$$

where the first term is the $\rho_f + \rho_g \|L\|^2$ -conjugate for weakly convex function $F - \frac{\rho_f + \rho_g \|L\|^2}{2} \|\cdot\|^2$ and G^* is the convex conjugate of G . Similarly, we have the conjugate dual problem for (6.1) is

$$\sup_{y \in Y} - \left(f - \frac{\rho_g}{2} \|L \cdot\|_Y^2 - \frac{\rho_f}{2} \|\cdot\|^2 \right)_{\rho_f}^* (-\langle L^* v, \cdot \rangle) - g_{\rho_g}^*(y). \quad (6.6)$$

We can write (6.5) by expressing $\rho_f + \rho_g \|L\|^2$ -conjugate in the first term to obtain

$$\sup_{y \in Y} \inf_{x \in X} F(x) + \langle Lx, y \rangle - G^*(y), \quad (6.7)$$

and (6.6) with ρ_f -conjugate

$$\sup_{y \in Y} \inf_{x \in X} f(x) + \langle Lx, y \rangle - \frac{\rho_g}{2} \|Lx\|^2 - g_{\rho_g}^*(y). \quad (6.8)$$

Notice that both (6.7) and (6.8) are Lagrange dual problems with respective Lagrangians

$$\mathcal{L}_{FG}(x, y) = F(x) + \langle Lx, y \rangle - G^*(y), \quad (6.9)$$

$$\mathcal{L}_{fg}(x, y) = f(x) + \langle Lx, y \rangle - \frac{\rho_g}{2} \|Lx\|^2 - g_{\rho_g}^*(y). \quad (6.10)$$

Observe that both Lagrangians are equal $\mathcal{L}_{FG}(x, y) = \mathcal{L}_{fg}(x, y)$ for each $x \in X$ and $y \in Y$. Now let us consider Lagrange primal problem with respect to (6.9),

$$\inf_{x \in X} \sup_{y \in Y} \mathcal{L}_{FG}(x, y) = \inf_{x \in X} \sup_{y \in Y} F(x) + \langle Lx, y \rangle - G^*(y) = \inf_{x \in X} F(x) + G^{**}(Lx).$$

Because G is lsc convex, $G^{**} = G$ ([6, Theorem 13.37]) and we return to (6.4). With the same argument for Lagrangian in (6.10), we have

$$\inf_{x \in X} \sup_{y \in Y} \mathcal{L}_{fg}(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x) + \langle Lx, y \rangle - \frac{\rho_g}{2} \|Lx\|^2 - g_{\rho_g}^*(y) = \inf_{x \in X} f(x) + g_{\rho_g}^{**}(Lx),$$

where

$$g_{\rho_g}^{**}(Lx) = \sup_{y \in Y} \langle Lx, y \rangle - \frac{\rho_g}{2} \|Lx\|^2 - g_{\rho_g}^*(y),$$

is the ρ_g -biconjugate of g from the definition of ρ_g -conjugate. Moreover, we know that g is lsc weakly convex, it is also abstract convex with the class $\Phi_{lsc}^{\rho_g/2}$ by Theorem 1. Hence $g_{\rho_g}^{**} = g$ so the Lagrange primal is the primal problem (6.1). $\square$

6.1.1 Sharpness

By the property of conjugate function, the Lagrangian is always convex in the dual variable. On the other hand, it is nonconvex (weakly convex) in the primal variable. Our main interest lies in the respective Lagrange dual problem

$$\sup_{y \in Y} \inf_{x \in X} \mathcal{L}(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x) + \langle Lx, y \rangle - g^*(y). \quad (6.11)$$

The Lagrange primal problem

$$\inf_{x \in X} \sup_{y \in Y} \mathcal{L}(x, y). \quad (6.12)$$

We recall that Lagrange primal (6.12), in general, is not equivalent to the primal problem (6.1) unless function g is Ψ -convex [12, Proposition 5.1]. As we consider g convex in the current setting, problem (6.1) is equivalent to (6.12) for $\Psi = \Psi_{conv}$. By Theorem 7, the duality gap $\mathcal{G}$ is non-negative,

$$\mathcal{G} = \inf_{x \in X} \sup_{\hat{y} \in Y} \mathcal{L}(x, \hat{y}) - \sup_{y \in Y} \inf_{\hat{x} \in X} \mathcal{L}(\hat{x}, y) \geq 0. \quad (6.13)$$

By definition, (x^*, y^*) is a saddle point of Lagrangian given by (6.11) which satisfy

$$(\forall (x, y) \in X \times Y) \quad \sup_{y \in Y} \mathcal{L}(x^*, y) \geq \mathcal{L}(x^*, y^*) \geq \inf_{x \in X} \mathcal{L}(x, y^*). \quad (6.14)$$

Furthermore, for any saddle point (x^*, y^*) of (6.14), x^* is a minimizer of (6.1). From (6.14), we define the gap function $\mathcal{G} : X \times Y \rightarrow [0, \infty)$,

$$\begin{aligned} \mathcal{G}(x, y) &:= f(x) + g^*(y) + \sup_{(\hat{x}, \hat{y}) \in X \times Y} \langle Lx, \hat{y} \rangle - \langle L\hat{x}, y \rangle - f(\hat{x}) - g^*(\hat{y}) \\ &= f(x) + g^*(y) + g^{**}(Lx) - f^*(-L^*y). \end{aligned} \quad (6.15)$$

Hence $\mathcal{G}(x, y)$ can take $+\infty$ when $f(\hat{x}) = -\infty$ or $g^*(y) = +\infty$. Moreover, any point (x^*, y^*) satisfying $\mathcal{G}(x^*, y^*) = 0$ is a saddle point. To see this, let us take $(x^*, y^*) \in X \times Y$ such that $\mathcal{G}(x^*, y^*) = 0$ i.e.

$$\sup_{y \in Y} \inf_{x \in X} \mathcal{L}(x, y) \geq \inf_{x \in X} \mathcal{L}(x, y^*) = \sup_{y \in Y} \mathcal{L}(x^*, y) \geq \inf_{x \in X} \sup_{y \in Y} \mathcal{L}(x, y).$$

This infers equality in all the relations above. Hence,

$$\mathcal{L}(x, y^*) \geq \mathcal{L}(x^*, y^*) \geq \mathcal{L}(x^*, y),$$

for all $(x, y) \in X \times Y$ which is the definition of saddle point. By assumption of this section, the gap function $\mathcal{G}(x, y)$,

$$\mathcal{G}(x, y) = f(x) + g^*(y) + \sup_{(\hat{x}, \hat{y}) \in X \times Y} \langle Lx, \hat{y} \rangle - \langle L\hat{x}, y \rangle - f(\hat{x}) - g^*(\hat{y}), \quad (6.16)$$

is weakly convex with respect to x , and convex with respect to y . It is even jointly weakly convex with respect to x and y thanks to its structure, i.e. there exists a constant $\rho_{\mathcal{G}} > 0$ such that for any $(x, y) \in X \times Y$, $\mathcal{G}(x, y) + \frac{\rho_{\mathcal{G}}}{2} \|(x, y)\|^2$ is a convex function. Indeed, let $\lambda \in [0, 1]$, $(x_1, y_1), (x_2, y_2) \in$

$X \times Y$, we have

$$\begin{aligned}
& \mathcal{G}(x_1 + (1 - \lambda)x_2, y_1 + (1 - \lambda)y_2) \\
&= f(x_1 + (1 - \lambda)x_2) + g^*(y_1 + (1 - \lambda)y_2) \\
&+ \sup_{(\hat{x}, \hat{y}) \in X \times Y} \langle L(x_1 + (1 - \lambda)x_2), \hat{y} \rangle - \langle L\hat{x}, y_1 + (1 - \lambda)y_2 \rangle - f(\hat{x}) - g^*(\hat{y}) \\
&\leq f(x_1) + g^*(y_1) + (1 - \lambda)(f(x_2) + g^*(y_2)) \\
&+ \sup_{(\hat{x}, \hat{y}) \in X \times Y} \langle Lx_1, \hat{y} \rangle - \langle L\hat{x}, y_1 \rangle - f(\hat{x}) - g^*(\hat{y}) \\
&+ (1 - \lambda) \sup_{(\hat{x}, \hat{y}) \in X \times Y} \langle Lx_2, \hat{y} \rangle - \langle L\hat{x}, y_2 \rangle - f(\hat{x}) - g^*(\hat{y}) + \lambda(1 - \lambda) \frac{\rho_f}{2} \|x_1 - x_2\|^2 \\
&\leq \mathcal{G}(x_1, y_1) + (1 - \lambda)\mathcal{G}(x_2, y_2) + \lambda(1 - \lambda) \frac{\rho_f}{2} \|(x_1, y_1) - (x_2, y_2)\|_{X \times Y}^2,
\end{aligned}$$

where we utilize weak convexity of f with modulus ρ_f and convexity of g^* and property of supremum. Hence $\mathcal{G}$ is jointly weakly convex with modulus ρ_f .

Let us introduce sharpness condition for the Lagrangian which measures how far we are from the saddle point in term of difference of Lagrangian value.

Definition 11 (Sharpness). Let us denote $z = (x, y) \in X \times Y = Z$ and let $S \subset Z$ be the nonempty set of saddle point of the Lagrange primal-dual problems. We say that the Lagrangian $\mathcal{L}$ is sharp if there exists a positive constant $\mu > 0$ such that

$$(\forall z \in Z) \quad \mathcal{G}(z) \geq \mu \text{dist}(z, S), \quad (6.17)$$

where the distance function is taken as

$$\text{dist}(z, S) = \inf_{\bar{z} \in S} \|z - \bar{z}\| = \inf_{(\bar{x}, \bar{y}) \in S} \sqrt{\|x - \bar{x}\|_X^2 + \|y - \bar{y}\|_Y^2}. \quad (6.18)$$

Let us observe that Definition 11 coincides with sharpness definition of a

proper function $F : X \rightarrow (-\infty, +\infty]$ with respect to its minimizer

$$F(x) - \inf_{\bar{x} \in X} F(\bar{x}) \geq \mu \text{dist}(x, \arg \min F).$$

Sharpness condition is commonly known as error bound condition and is often used to speed up convergence rate of algorithms [80, 34, 47]. In nonconvex setting, sharpness is used to investigate convergence of algorithm [39, 62, 31]. For necessary and sufficient conditions to obtain error bound condition, the reader is refer to the works of [42, 41, 36].

For primal-dual algorithms, the concept of Lagrangian sharpness is not widely studied. Recently, [3] exploits sharpness condition for Lagrangian to achieve linear convergence rate in linear programming. At the moment, this is the only work in the literature that we found which used a concept analogous to our sharpness definition.

While the properties of the gap function $\mathcal{G}(x, y)$ are related to the properties of weak duality of Lagrange dual problem, it is not often used in algorithms to determine saddle point, especially for unbounded problems. Moreover, the sharpness of the gap function is often used to improve convergence rate rather than determine the convergence of algorithms [93, 47, 34, 3].

To investigate convergence of our primal-dual algorithm, we introduce another definition of sharpness which will be crucial in the convergence analysis, see e.g. Theorem 18 and Theorem 19.

Let us denote $z = (x, y) \in X \times Y = Z$ and let $\mathcal{L} : X \times Y \rightarrow (-\infty, +\infty]$ be any function. We will refer to this function as a Lagrangian.

Definition 12 (Inf-Sharpness). Let $S \subset Z$ be the nonempty set of saddle points of the Lagrangian $\mathcal{L}$. We say that the Lagrangian $\mathcal{L}$ is *inf-sharp* if there exists a positive constant $\mu > 0$ such that for all $(x, y) \in X \times Y$,

$$\mathcal{H}(x, y) := \inf_{(\bar{x}, \bar{y}) \in S} \{\mathcal{L}(x, \bar{y}) - \mathcal{L}(\bar{x}, y)\} \geq \mu \text{dist}((x, y), S). \quad (6.19)$$

Observe that $\mathcal{H}(x, y) \geq 0$ for any $(x, y) \in X \times Y$ and if $(x, y) \in S$ then $\mathcal{H}(x, y) = 0$. The converse is not true, see Example 23 below. Moreover, we have

$$\mathcal{G}(x, y) \geq \sup_{(\bar{x}, \bar{y}) \in S} \{\mathcal{L}(x, \bar{y}) - \mathcal{L}(\bar{x}, y)\} \geq \mathcal{H}(x, y) \geq \inf_{(\bar{x}, \bar{y}) \in X \times Y} \{\mathcal{L}(x, \bar{y}) - \mathcal{L}(\bar{x}, y)\},$$

which means if (6.19) is satisfied, then we also have Lagrangian sharpness in the sense of Definition 11 with the same constant $\mu > 0$. On the other hand, sharpness condition (6.19) gives us a necessary and sufficient condition for the point $\mathcal{H}(x, y) = 0$ becomes a saddle point i.e.

$$\mathcal{H}(x, y) = 0 \Leftrightarrow (x, y) \in S.$$

Example 18. Let us give a simple example of Lagrangian that satisfies Definition 12. Consider $\mathcal{L} : \mathbb{R} \times \mathbb{R} \rightarrow (-\infty, +\infty]$ of the following form,

$$\mathcal{L}(x, y) = |x| - |y|, \tag{6.20}$$

which is convex concave. It has a unique saddle point at $x = y = 0$. We have

$$\mathcal{H}(x, y) = \mathcal{L}(x, 0) - \mathcal{L}(0, y) = |x| + |y| \geq \sqrt{x^2 + y^2},$$

so the Lagrangian is inf-sharp with constant 1.

Example 19. Even though the Lagrangian from the previous example is inf-sharp, it does not have the bilinear term as it is in (6.11). Let us consider another example inspired by (6.20),

$$\mathcal{L}(x, y) = |x| + xy - \frac{y^2}{8}, \tag{6.21}$$

which has a unique saddle point at $x = y = 0$ (see figure 6.1). Moreover, the

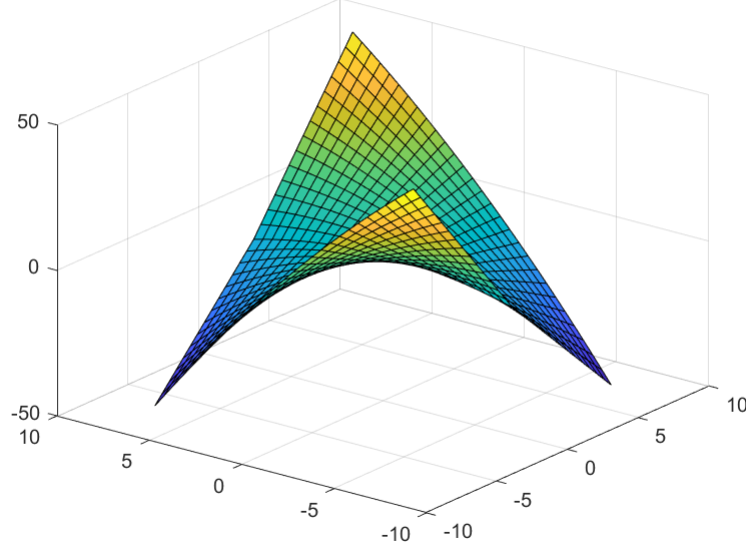


Figure 6.1: Lagrange function from Example 19

primal and dual problems corresponding to (6.21) are

$$\begin{aligned}\inf_{x \in \mathbb{R}} \sup_{y \in \mathbb{R}} \mathcal{L}(x, y) &= \inf_{x \in \mathbb{R}} |x| + 2x^2, \\ \sup_{y \in \mathbb{R}} \inf_{x \in \mathbb{R}} \mathcal{L}(x, y) &= \sup_{y \in \mathbb{R}} -\frac{y^2}{8}.\end{aligned}$$

In order to check for inf-sharp condition, we calculate

$$\mathcal{H}(x, y) = \mathcal{L}(x, 0) - \mathcal{L}(0, y) = |x| + \frac{y^2}{8},$$

and check for the value of $\mu > 0$ such that for all $(x, y) \in X \times Y$,

$$\mathcal{H}(x, y) - \mu \sqrt{x^2 + y^2} > 0. \quad (6.22)$$

We test $\mu = 1, 0.5, 0.1$. To see the area of interest, we plot the contour

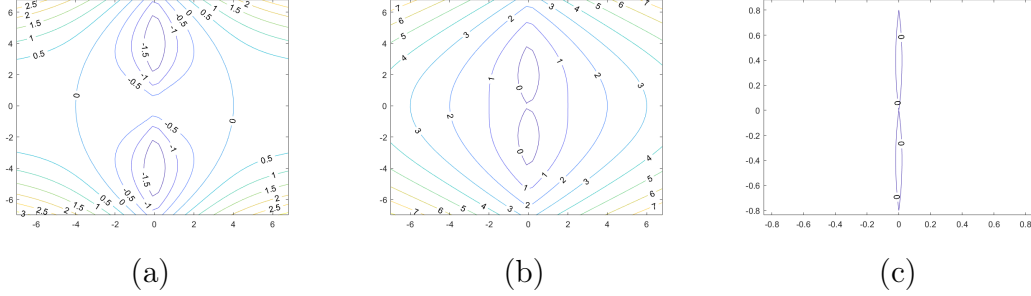


Figure 6.2: From left to right: contour plots of $\mathcal{H}(x, y) - \mu\sqrt{x^2 + y^2}$ at $\mu = 1, 0.5, 0.1$.

of (6.22) in Figure 6.2. As we can see, Lagrangian is not inf-sharp around $(0, 0)$ with $\mu = 1$. As we decrease μ , we gain more area for which $\mathcal{H}(x, y) - \mu\sqrt{x^2 + y^2} > 0$ around $(0, 0)$ but never a full neighborhood of $(0, 0)$ (even with $\mu = 0.1$). Hence, the Lagrangian in this example is not inf-sharp.

Example 20. Let us consider Example 18 with the bilinear term as

$$\mathcal{L}(x, y) = |x| + xy - |y|,$$

which has a saddle point at $(0, 0)$. The respective primal and dual problems for this Lagrangian are

$$\inf_{-1 \leq x \leq 1} |x|, \quad \sup_{-1 \leq y \leq 1} -|y|.$$

We calculate the function $\mathcal{H}(x, y)$

$$\mathcal{H}(x, y) = \mathcal{L}(x, 0) - \mathcal{L}(0, y) = |x| + |y| \geq \sqrt{x^2 + y^2} = \text{dist}((x, y), (0, 0)).$$

Therefore, the Lagrangian in this case is inf-sharp with constant 1. We also illustrate the difference $\mathcal{H}(x, y) - \mu \text{dist}((x, y), (0, 0))$ for $\mu = 0.5, 0.9, 1$ in Figure 6.3.

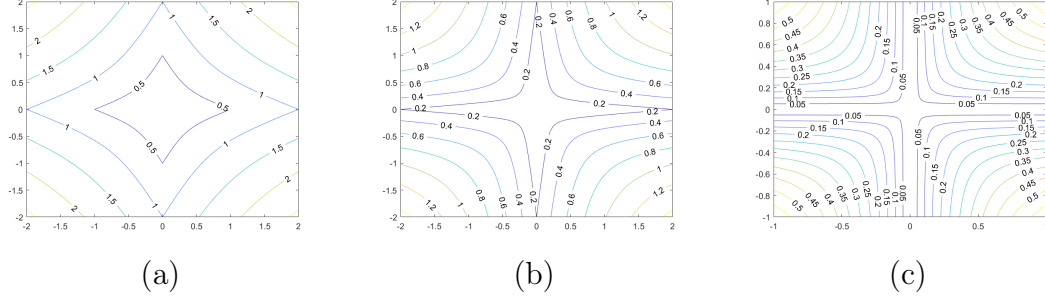


Figure 6.3: From left to right: contour plots of $\mathcal{H}(x, y) - \mu \text{dist}((x, y), (0, 0))$ at $\mu = 0.5, 0.9, 1$.

Example 21. We give one more example for a more complex function. Consider the Lagrangian

$$\mathcal{L}(x, y) = |x| + |x^2 - 1| + xy - |y| - |y^2 - 1|,$$

which has a saddle point at $(0, 0)$ (see Figure 6.4). The Lagrangian $\mathcal{L}(x, y)$ is weakly convex in x and weakly concave in y .

We calculate the function $\mathcal{H}(x, y)$

$$\begin{aligned} \mathcal{H}(x, y) &= \mathcal{L}(x, 0) - \mathcal{L}(0, y) = |x| + |x^2 - 1| + |y| + |y^2 - 1| - 2 \\ &\geq \sqrt{x^2 + y^2} - 2 = \text{dist}((x, y), (0, 0)) - 2. \end{aligned}$$

This implies that Lagrangian is not inf-sharp with constant 1. In fact, it is locally inf-sharp with constant $\mu < 1$. We illustrate the difference $\mathcal{H}(x, y) - \mu \text{dist}((x, y), (0, 0))$ with $\mu = 0.5, 0.9, 1$ in Figure 6.5.

Notice that there are others points other than the saddle point that $\mathcal{H}(x, y) = 0$ e.g. $\mathcal{H}(0, 1) = \mathcal{H}(1, 1) = \mathcal{H}(1, 0) = 0$.

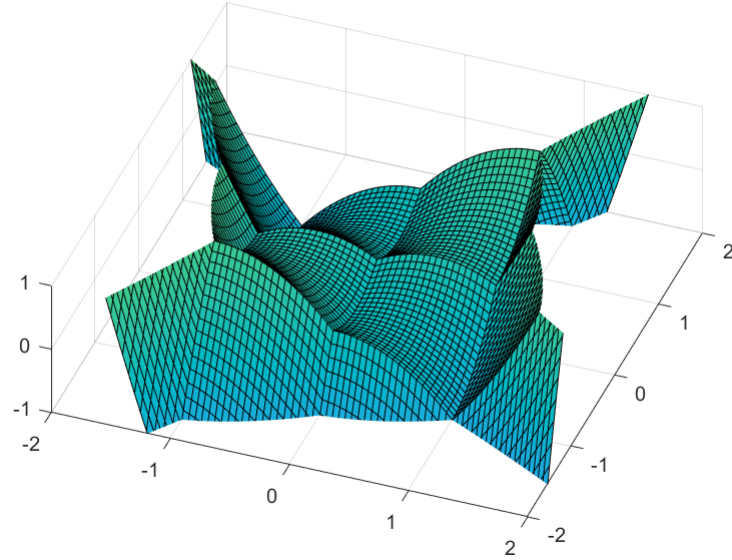
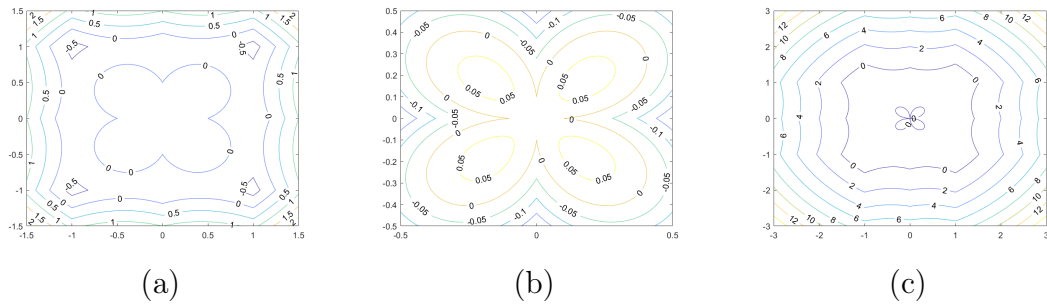


Figure 6.4: Lagrange function from Example 21.

Figure 6.5: From left to right: contour plots of $\mathcal{H}(x, y) - \mu \text{dist}((x, y), (0, 0))$ at $\mu = 0.5, 0.9, 1$.

6.2 Convergence Analysis of the Primal-Dual Algorithm for Weakly Convex Functions

In this section, we investigate the sequences generated by the primal-dual Algorithm 5 for problem (6.1)

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

where f is proper lsc weakly convex function and g is proper lsc convex defined on Hilbert spaces X and Y , respectively, and $L : X \rightarrow Y$ is a bounded linear operator with the adjoint operator $L^* : Y \rightarrow X$.

Our standing assumption in this section is that $\text{dom } f \cap \text{dom } g \circ L \neq \emptyset$ and the set of saddle points S is nonempty.

For convex functions f and g , the analogous of Algorithm 5 has been studied in [25]. In [51], the authors consider Algorithm 5 for convex case with no relaxation ($\theta = 0$). They emphasize that the algorithm may diverge even if the Lagrangian is convex-concave.

In our setting, under inf-sharpness condition defined in Definition 11, we investigate the convergence of Algorithm 5 where function f is weakly convex while g is convex (see Theorem 18). For weakly convex function f , we use proximal subdifferentials. It has been proved that for this special class of functions, the proximal subdifferentials can be globalized (see [59, Proposition 3.1]).

Definition 13. Let $F : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function and $x \in X$. The proximal subgradient of F is an element $u \in X$ satisfies the following inequality

$$(\forall z \in X) \ F(z) - F(x) \geq \langle u, z - x \rangle - \frac{\rho}{2} \|z - x\|^2.$$

We denote the collection of all proximal subgradients of the function F at the point x as $\partial_\rho F(x)$.

6.2.1 Primal-Dual algorithm with Dual Update

We investigate the convergence of primal-dual Algorithm 5 by exploiting the conjugate dual under the setting that the function f is ρ -weakly convex while g is convex on their respective domains (see (6.3)). Recall that under our standing assumption, we have $\text{dom } f \cap \text{dom } g \circ L \neq \emptyset$ and the set of saddle points S is nonempty.

Algorithm 5: Primal-Dual Algorithm

Initialize: $x_0 \in \text{dom } f \cap \text{dom } g \circ L$ and $\bar{y}_0 = y_0$

Set: $\tau, \sigma > 0, \sigma\rho < 1, \theta \in [0, 1], \sqrt{\sigma\tau}\|L\| < 1$

Update:

$$y_{n+1} = \operatorname{argmin}_{y \in Y} g^*(y) + \frac{1}{2\tau} \|y - y_n - \tau Lx_n\|^2 \quad (\text{dual update})$$

$$\bar{y}_{n+1} = y_{n+1} + \theta(y_{n+1} - y_n)$$

$$x_{n+1} = \operatorname{argmin}_{x \in X} f(x) + \frac{1}{2\sigma} \|x - x_n + \sigma L^* \bar{y}_{n+1}\|^2 \quad (\text{primal update})$$

We remind that we consider the Lagrangian of the form $\mathcal{L}(x, y) = f(x) + \langle Lx, y \rangle - g^*(y)$ as in problem (6.11). In Algorithm 5, we have to solve two minimization problems where the objectives are convex thanks to the appropriate choice of stepsize $\sigma\rho < 1$,

$$\begin{aligned} 0 &\in \partial \left(g^* + \frac{1}{2\tau} \|\cdot - y_n - \tau Lx_n\|^2 \right) (y_{n+1}), \\ 0 &\in \partial \left(f + \frac{1}{2\sigma} \|\cdot - x_n + \sigma L^* \bar{y}_{n+1}\|^2 \right) (x_{n+1}), \end{aligned}$$

where ∂ is the subdifferential in the sense of convex analysis. We can apply convex subdifferentials sum rule to the dual update as both functions are convex [6, Corollary 16.48].

On the other hand, since f is weakly convex, we utilize the following sum

rule for proximal subdifferentials for weakly convex function.

Theorem 17. *[10, Theorem 2] Let X be a Hilbert space. For $i = 0, 1$, let function $f_i : X \rightarrow (-\infty, +\infty]$ be proper lower semicontinuous and ρ_i -weakly convex on X with $\rho_i \geq 0$. Then, for all $x \in \text{dom } f_0 \cap \text{dom } f_1$ and for all $\varepsilon_0, \varepsilon_1 \geq 0$ we have*

$$\partial_{\varepsilon_0, \rho_0} f_0(x) + \partial_{\varepsilon_1, \rho_1} f_1(x) \subseteq \partial_{\varepsilon, \rho}(f_0 + f_1)(x) \quad (6.23)$$

for all $\varepsilon \geq \varepsilon_0 + \varepsilon_1$ and for all $\rho \geq \rho_0 + \rho_1$. The equality

$$\partial_{\varepsilon, \rho_0 + \rho_1}(f_0 + f_1)(x) = \bigcup_{\varepsilon_0, \varepsilon_1 \mid \varepsilon_0 + \varepsilon_1 \leq \varepsilon} \partial_{\varepsilon_0, \rho_0} f_0(x) + \partial_{\varepsilon_1, \rho_1} f_1(x) \quad (6.24)$$

holds when $\text{dom } f_0 \cap \text{dom } f_1$ contains a point at which either $f_0 + \rho_0/2 \|\cdot\|^2$ or $f_1(\cdot) + \rho_1/2 \|\cdot\|^2$ is continuous.

Using the above theorem with $\varepsilon = 0$, we obtain

$$\begin{aligned} \frac{y_n - y_{n+1}}{\tau} + Lx_n &\in \partial g^*(y_{n+1}), \\ \frac{x_n - x_{n+1}}{\sigma} - L^* \bar{y}_{n+1} &\in \partial_\rho f(x_{n+1}). \end{aligned} \quad (6.25)$$

Let us give an estimate of the Lagrangian values for the iterates generated by Algorithm 5.

Proposition 17. *Let X, Y be Hilbert spaces, $f : X \rightarrow (-\infty, +\infty]$ be proper lsc ρ -weakly convex function and $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex defined on Hilbert spaces X and Y , respectively, and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be sequences generated by Algorithm 5.*

Then for any $(x, y) \in X \times Y$, we have

$$\begin{aligned}
& \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
& \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{n+1}\|^2 - \frac{1}{2\tau} \|y - y_n\|^2 + \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
& \quad + \frac{1 - \sigma\rho - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x - x_{n+1}\|^2 - \frac{1}{2\sigma} \|x - x_n\|^2 + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n+1}\|^2.
\end{aligned} \tag{6.26}$$

Proof. From the update of Algorithm 5 and (6.25), by definitions of subdifferentials, we have

$$\begin{aligned}
(\forall x \in X) \quad f(x) - f(x_{n+1}) & \geq \frac{1}{\sigma} \langle x_n - x_{n+1}, x - x_{n+1} \rangle - \langle L^* \bar{y}_{n+1}, x - x_{n+1} \rangle \\
& \quad - \frac{\rho}{2} \|x - x_{n+1}\|^2, \\
(\forall y \in Y) \quad g^*(y) - g^*(y_{n+1}) & \geq \frac{1}{\tau} \langle y_n - y_{n+1}, y - y_{n+1} \rangle + \langle Lx_n, y - y_{n+1} \rangle,
\end{aligned}$$

Summing the above inequalities and taking into account the Lagrangian given in (6.11), we have

$$\begin{aligned}
& \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
& \geq \left\langle \frac{y_n - y_{n+1}}{\tau}, y - y_{n+1} \right\rangle + \langle Lx_n, y - y_{n+1} \rangle + \left\langle \frac{x_n - x_{n+1}}{\sigma}, x - x_{n+1} \right\rangle \\
& \quad - \langle L^* \bar{y}_{n+1}, x - x_{n+1} \rangle - \frac{\rho}{2} \|x - x_{n+1}\|^2 + \langle Lx, y_{n+1} \rangle - \langle Lx_{n+1}, y \rangle \\
& = \left\langle \frac{y_n - y_{n+1}}{\tau}, y - y_{n+1} \right\rangle + \left\langle \frac{x_n - x_{n+1}}{\sigma}, x - x_{n+1} \right\rangle - \frac{\rho}{2} \|x - x_{n+1}\|^2 \\
& \quad + \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle + \langle L(x - x_{n+1}), y_{n+1} - \bar{y}_{n+1} \rangle.
\end{aligned} \tag{6.27}$$

For the last line of (6.27), we have

$$\begin{aligned}
& \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle + \langle L(x - x_{n+1}), y_{n+1} - \bar{y}_{n+1} \rangle \\
&= \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle + \theta \langle L(x - x_{n+1}), y_n - y_{n+1} \rangle \\
&\geq -\sqrt{\sigma\tau} \|L\| \left(\frac{\|x_n - x_{n+1}\|^2}{2\sigma} + \frac{\|y - y_{n+1}\|^2}{2\tau} + \theta \frac{\|x - x_{n+1}\|^2}{2\sigma} + \theta \frac{\|y_n - y_{n+1}\|^2}{2\tau} \right).
\end{aligned}$$

Inequality (6.27) becomes

$$\begin{aligned}
& \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
&\geq \frac{1}{2\tau} (\|y - y_{n+1}\|^2 - \|y - y_n\|^2 + \|y_n - y_{n+1}\|^2) - \frac{\rho}{2} \|x - x_{n+1}\|^2 \\
&\quad + \frac{1}{2\sigma} (\|x - x_{n+1}\|^2 - \|x - x_n\|^2 + \|x_n - x_{n+1}\|^2) \\
&\quad - \sqrt{\sigma\tau} \|L\| \left(\frac{\|x_n - x_{n+1}\|^2}{2\sigma} + \frac{\|y - y_{n+1}\|^2}{2\tau} + \theta \frac{\|x - x_{n+1}\|^2}{2\sigma} + \theta \frac{\|y_n - y_{n+1}\|^2}{2\tau} \right) \\
&= \frac{1 - \sqrt{\sigma\tau} \|L\|}{2\tau} \|y - y_{n+1}\|^2 - \frac{1}{2\tau} \|y - y_n\|^2 + \frac{1 - \theta \sqrt{\sigma\tau} \|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
&\quad + \frac{1 - \sigma\rho - \theta \sqrt{\sigma\tau} \|L\|}{2\sigma} \|x - x_{n+1}\|^2 - \frac{1}{2\sigma} \|x - x_n\|^2 + \frac{1 - \sqrt{\sigma\tau} \|L\|}{2\sigma} \|x_n - x_{n+1}\|^2,
\end{aligned}$$

which is (6.26). $\square$

Estimation (6.26) is not enough to obtain convergence of Algorithm 5. This is where we utilize inf-sharpness assumption in the sense of Definition 12 to analyze the convergence of Algorithm 5.

We show that, if (x_{n+1}, y_{n+1}) are closed enough to the solution set S then we have decreasing property for the distance function of (x_n, y_n) to the set S .

Proposition 18. *Let X and Y be Hilbert, $f : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function, $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 5. Let us assume that the stepsizes τ, σ satisfy the*

inequality $\sigma\rho + \theta\sqrt{\sigma\tau}\|L\| < 1$, where $\theta \in [0, 1]$, and the Lagrangian is inf-sharp in the sense of Definition 12 with respect to the set of saddle point $S \neq \emptyset$. At iteration $n \in \mathbb{N}$, if

$$\text{dist}((x_n, y_n), S) \leq \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A}, \quad (6.28)$$

then

$$\text{dist}((x_{n+1}, y_{n+1}), S) \leq \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A},$$

where $A := \min\left\{\frac{1-\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\theta\sqrt{\sigma\tau}\|L\|}{2\sigma}\right\}$. Moreover,

$$\text{dist}^2((x_n, y_n), S) \geq \text{dist}^2((x_{n+1}, y_{n+1}), S). \quad (6.29)$$

Proof. Let us prove (6.29) first. By (6.26) from Proposition 17, for any $(x, y) \in X \times Y$, we have

$$\begin{aligned} & \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) + \frac{1}{2\sigma} \|x - x_n\|^2 + \frac{1}{2\tau} \|y - y_n\|^2 \\ & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{n+1}\|^2 + \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\ & \quad + \frac{1 - \sigma\rho - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x - x_{n+1}\|^2 + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n+1}\|^2 \\ & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{n+1}\|^2 + \frac{1 - \sigma\rho - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x - x_{n+1}\|^2 \\ & \geq A \inf_{(x,y) \in S} \|x - x_{n+1}\|^2 + \|y - y_{n+1}\|^2 = A \text{dist}^2((x_{n+1}, y_{n+1}), S), \end{aligned} \quad (6.30)$$

where $A := \min\left\{\frac{1-\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\theta\sqrt{\sigma\tau}\|L\|}{2\sigma}\right\} > 0$ thanks to the assumption on the stepsizes. The last inequality in (6.30) holds as we can take any $(x, y) \in S$. Let us take infimum on both sides of (6.30) with respect to $(x, y) \in S$, combine with the distance function (6.18)

$$\inf_{(x,y) \in S} \|x - x_n\|^2 + \|y - y_n\|^2 = \text{dist}^2((x_n, y_n), S).$$

We obtain

$$\begin{aligned} & \max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \text{dist}^2((x_n, y_n), S) \\ & \geq A \text{dist}^2((x_{n+1}, y_{n+1}), S) + \inf_{(x, y) \in S} \{ \mathcal{L}(x_{n+1}, y) - \mathcal{L}(x, y_{n+1}) \}. \end{aligned} \quad (6.31)$$

We apply sharpness assumption to (6.31) and obtain

$$\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \text{dist}^2((x_n, y_n), S) \geq A \text{dist}^2((x_{n+1}, y_{n+1}), S) + \mu \text{dist}((x_{n+1}, y_{n+1}), S). \quad (6.32)$$

To obtain (6.29), we need

$$\begin{aligned} & A \text{dist}^2((x_{n+1}, y_{n+1}), S) + \mu \text{dist}((x_{n+1}, y_{n+1}), S) \\ & \geq \max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \text{dist}^2((x_{n+1}, y_{n+1}), S), \end{aligned}$$

which holds when

$$\text{dist}((x_{n+1}, y_{n+1}), S) \leq \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A}. \quad (6.33)$$

The quantity on RHS of (6.33) is well-defined as

$$\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \geq \min \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} > \min \left\{ \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1 - \sigma\rho - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \right\},$$

thanks to the assumptions on the stepsizes τ, σ . Next we will prove that for any $n \in \mathbb{N}$

$$\text{dist}((x_n, y_n), S) \leq \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A} \Rightarrow \text{dist}((x_{n+1}, y_{n+1}), S) \leq \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A}.$$

Let us start from (6.32)

$$\begin{aligned}
& \max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \frac{\mu^2}{\left(\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A \right)^2} \\
& \geq \max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} \text{dist}^2((x_n, y_n), S) \\
& \geq A \text{dist}^2((x_{n+1}, y_{n+1}), S) + \mu \text{dist}((x_{n+1}, y_{n+1}), S). \tag{6.34}
\end{aligned}$$

Inequality (6.34) becomes a quadratic inequality with respect to $\text{dist}((x_{n+1}, y_{n+1}), S)$. The quadratic equation taken from (6.34) has two distinct solutions. This implies that

$$0 \leq \text{dist}((x_{n+1}, y_{n+1}), S) \leq \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A}.$$

□

When inequality in (6.28) is strict, we obtain convergence to zero of the distance function.

Corollary 4. *Let X and Y be Hilbert, $f : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function, $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 5. Let us assume that $\sigma\rho + \theta\sqrt{\sigma\tau}\|L\| < 1$ and the Lagrangian is inf-sharp in the sense of Definition 12 with respect to the set of saddle points $S \neq \emptyset$. If the starting point (x_0, y_0) satisfies*

$$\text{dist}((x_0, y_0), S) < \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A},$$

where $A = \min \left\{ \frac{1-\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \right\}$. Then $\text{dist}((x_n, y_n), S)$ tends to zero as $n \rightarrow \infty$.

Proof. By assumption and Proposition 18, $\text{dist}((x_n, y_n), S)$ is non-increasing

and

$$\text{dist}((x_n, y_n), S) < \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A},$$

for all $n \in \mathbb{N}$. Moreover, we have

$$\begin{aligned} & \left(A + \frac{\mu}{\text{dist}((x_0, y_0), S)} \right) \text{dist}^2((x_{n+1}, y_{n+1}), S) \\ & \leq \left(A + \frac{\mu}{\text{dist}((x_{n+1}, y_{n+1}), S)} \right) \text{dist}^2((x_{n+1}, y_{n+1}), S) \\ & \leq \max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} \text{dist}^2((x_n, y_n), S). \end{aligned} \quad (6.35)$$

Inequality (6.35) implies that

$$\text{dist}^2((x_{n+1}, y_{n+1}), S) \leq (B)^{n-1} \text{dist}^2((x_0, y_0), S), \quad (6.36)$$

where

$$B := \frac{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\}}{A + \frac{\mu}{\text{dist}((x_0, y_0), S)}}.$$

By assumption on $\text{dist}((x_0, y_0), S)$, we have $0 < B < 1$ since

$$A + \frac{\mu}{\text{dist}((x_0, y_0), S)} > \max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\}.$$

As (6.36) holds for any $n \in \mathbb{N}$, we let $n \rightarrow \infty$ and we infer that

$$\lim_{n \rightarrow \infty} \text{dist}((x_n, y_n), S) = 0.$$

□

Next, we prove the convergence of $(x_n, y_n)_{n \in \mathbb{N}}$ toward the set of saddle points in S .

Theorem 18. *Let X and Y be Hilbert, $f : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function, $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex and*

$L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 5. Let us assume that the stepsizes τ, σ satisfy $\sigma\rho + \theta\sqrt{\sigma\tau}\|L\| < 1$, and the Lagrangian is inf-sharp in the sense of Definition 12 with respect to the set of saddle point $S \neq \emptyset$. If the starting point (x_0, y_0) satisfies

$$\text{dist}((x_0, y_0), S) < \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A},$$

where $A = \min\left\{\frac{1-\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\theta\sqrt{\sigma\tau}\|L\|}{2\sigma}\right\}$. Then (x_n, y_n) converges to a saddle point.

Proof. By (6.26) from Proposition 17 and the inf-sharpness of Lagrangian, we have

$$\begin{aligned} & \max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} \text{dist}^2((x_n, y_n), S) \\ & \geq A \text{dist}^2((x_{n+1}, y_{n+1}), S) + \mu \text{dist}((x_{n+1}, y_{n+1}), S) \\ & \quad + \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n+1}\|^2 \\ & \geq \min\left\{\frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma}\right\} \|(x_n, y_n) - (x_{n+1}, y_{n+1})\|_{X \times Y}^2. \end{aligned} \quad (6.37)$$

Taking square root on both sides of (6.37) and summing up till $N \in \mathbb{N}$, combining with the result from Corollary 4, we obtain

$$C \sum_{i=0}^N B^{\frac{i-1}{2}} \geq \sum_{i=0}^N \|(x_i, y_i) - (x_{i+1}, y_{i+1})\|_{X \times Y}, \quad (6.38)$$

where

$$B = \frac{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\}}{A + \frac{\mu}{\text{dist}((x_0, y_0), S)}}, \quad C := \sqrt{\frac{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\}}{\min\left\{\frac{1-\theta\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sqrt{\sigma\tau}\|L\|}{2\sigma}\right\}}} \text{dist}((x_0, y_0), S).$$

We have $0 < B < 1$ as follow from the proof of Corollary 4. Letting $N \rightarrow \infty$,

the LHS of (6.38) is summable which implies that

$$\sum_i \|(x_i, y_i) - (x_{i+1}, y_{i+1})\|_{X \times Y} < +\infty.$$

As X, Y are Hilbert, so is $X \times Y$. By [14, Theorem 1], $(x_n, y_n)_{n \in \mathbb{N}}$ is Cauchy, so it converges. Combining this with the result obtained in Corollary 4 which is $\text{dist}((x_n, y_n), S) \rightarrow 0$ as $n \rightarrow \infty$, $(x_n, y_n)_{n \in \mathbb{N}}$ must converge to a saddle point in S . $\square$

6.2.2 Primal-Dual algorithm with primal update first

In this subsection, we consider Algorithm 6 which differs from Algorithm 5 by updating the primal variable x_n first instead of y_n . We show the convergence results similar to Proposition 18 and Theorem 18 of the previous subsection with some differences concerning the assumption on the stepsizes $\tau, \sigma > 0$.

Algorithm 6: Primal-Dual Algorithm with primal update first

Initialize: $x_0 \in \text{dom } f \cap \text{dom } g \circ L$ and $\bar{x}_0 = x_0$

Set: $\tau, \sigma > 0, \sigma\rho < 1, \theta \in [0, 1], \sqrt{\sigma\tau}\|L\| < 1$

Update:

$$x_{n+1} = \operatorname{argmin}_{x \in X} f(x) + \frac{1}{2\sigma} \|x - x_n + \tau L^* y_n\|^2$$

$$\bar{x}_{n+1} = x_{n+1} + \theta(x_{n+1} - x_n)$$

$$y_{n+1} = \operatorname{argmin}_{y \in Y} g^*(y) + \frac{1}{2\tau} \|y - y_n - \tau L \bar{x}_{n+1}\|^2$$

Analogously to Proposition 17, we start by analyzing the behavior of the Lagrangian values given as in problem (6.11).

Proposition 19. *Let X, Y be Hilbert spaces, $f : X \rightarrow (-\infty, +\infty]$ be proper lsc ρ -weakly convex function and $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex*

defined on Hilbert spaces X and Y , respectively, and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be sequences generated by Algorithm 6. Then for any $(x, y) \in X \times Y$, we have

$$\begin{aligned} & \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\ & \geq \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{n+1}\|^2 - \frac{1}{2\tau} \|y - y_n\|^2 + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\ & \quad + \frac{1 - \sigma\rho - \sqrt{\sigma\tau}\|L\|}{2\sigma} \|x - x_{n+1}\|^2 - \frac{1}{2\sigma} \|x - x_n\|^2 + \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n+1}\|^2. \end{aligned} \quad (6.39)$$

Proof. We proceed similar to the proof of Proposition 17 and obtain the following

$$\begin{aligned} & \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\ & \geq \left\langle \frac{y_n - y_{n+1}}{\tau}, y - y_{n+1} \right\rangle + \langle L\bar{x}_{n+1}, y - y_{n+1} \rangle + \left\langle \frac{x_n - x_{n+1}}{\sigma}, x - x_{n+1} \right\rangle \\ & \quad - \langle L^* y_n, x - x_{n+1} \rangle - \frac{\rho}{2} \|x - x_{n+1}\|^2 + \langle Lx, y_{n+1} \rangle - \langle Lx_{n+1}, y \rangle \\ & = \left\langle \frac{y_n - y_{n+1}}{\tau}, y - y_{n+1} \right\rangle + \left\langle \frac{x_n - x_{n+1}}{\sigma}, x - x_{n+1} \right\rangle - \frac{\rho}{2} \|x - x_{n+1}\|^2 \\ & \quad + \langle L(\bar{x}_{n+1} - x_{n+1}), y - y_{n+1} \rangle + \langle L(x - x_{n+1}), y_{n+1} - y_n \rangle. \end{aligned} \quad (6.40)$$

For the last terms of (6.40), we have

$$\begin{aligned} & \langle L(\bar{x}_{n+1} - x_{n+1}), y - y_{n+1} \rangle + \langle L(x - x_{n+1}), y_{n+1} - y_n \rangle \\ & = \theta \langle L(x_{n+1} - x_n), y - y_{n+1} \rangle + \langle L(x - x_{n+1}), y_{n+1} - y_n \rangle \\ & \geq -\sqrt{\sigma\tau}\|L\| \left(\theta \frac{\|x_n - x_{n+1}\|^2}{2\sigma} + \theta \frac{\|y - y_{n+1}\|^2}{2\tau} + \frac{\|x - x_{n+1}\|^2}{2\sigma} + \frac{\|y_n - y_{n+1}\|^2}{2\tau} \right). \end{aligned}$$

Inequality (6.40) becomes

$$\begin{aligned}
& \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
& \geq \frac{1}{2\tau} (\|y - y_{n+1}\|^2 - \|y - y_n\|^2 + \|y_n - y_{n+1}\|^2) - \frac{\rho}{2} \|x - x_{n+1}\|^2 \\
& \quad + \frac{1}{2\sigma} (\|x - x_{n+1}\|^2 - \|x - x_n\|^2 + \|x_n - x_{n+1}\|^2) \\
& \quad - \sqrt{\sigma\tau} \|L\| \left(\theta \frac{\|x_n - x_{n+1}\|^2}{2\sigma} + \theta \frac{\|y - y_{n+1}\|^2}{2\tau} + \frac{\|x - x_{n+1}\|^2}{2\sigma} + \frac{\|y_n - y_{n+1}\|^2}{2\tau} \right) \\
& = \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{n+1}\|^2 - \frac{1}{2\tau} \|y - y_n\|^2 + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
& \quad + \frac{1 - \sigma\rho - \sqrt{\sigma\tau}\|L\|}{2\sigma} \|x - x_{n+1}\|^2 - \frac{1}{2\sigma} \|x - x_n\|^2 + \frac{1 - \theta\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n+1}\|^2,
\end{aligned}$$

which is (6.39). $\square$

As we can see, the difference between Proposition 17 and Proposition 19 is position of θ which appears in the term $\|y - y_{n+1}\|^2$ instead of $\|x - x_{n+1}\|^2$. This will cause a small change in the bound of $\text{dist}((x_n, y_n), S)$ as we see in the next result.

Proposition 20. *Let X and Y be Hilbert, $f : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function, $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 6. Let us assume that the stepsizes τ, σ satisfy $\sigma\rho + \sqrt{\sigma\tau}\|L\| < 1$ and the Lagrangian is inf-sharp in the sense of Definition 12 with respect to the set of saddle points $S \neq \emptyset$. At iteration $n \in \mathbb{N}$, if*

$$\text{dist}((x_n, y_n), S) \leq \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A_1}, \quad (6.41)$$

then

$$\text{dist}((x_{n+1}, y_{n+1}), S) \leq \frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A_1},$$

where $A_1 := \min \left\{ \frac{1-\theta\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\sqrt{\sigma\tau}\|L\|}{2\sigma} \right\}$. Moreover,

$$\text{dist}^2((x_n, y_n), S) \geq \text{dist}^2((x_{n+1}, y_{n+1}), S). \quad (6.42)$$

As a consequence, if the initials (x_0, y_0) satisfy

$$\text{dist}((x_0, y_0), S) < \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A_1}.$$

Then $\text{dist}((x_n, y_n), S)$ tends to zero as $n \rightarrow \infty$.

The proof of Proposition 20 follows the same lines as the proof of Proposition 18 and Corollary 4 so we will not give it here. As $(\text{dist}((x_n, y_n), S))_{n \in \mathbb{N}}$ with $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ generated by Algorithm 6 behaves in the same way as in the case of Algorithm 5, so is the convergence of $(x_n, y_n)_{n \in \mathbb{N}}$.

Theorem 19. *Let X and Y be Hilbert, $f : X \rightarrow (-\infty, +\infty]$ be a proper lsc ρ -weakly convex function, $g : Y \rightarrow (-\infty, +\infty]$ be proper lsc convex and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 6. Let us assume that the stepsizes τ, σ satisfy $\sigma\rho + \sqrt{\sigma\tau}\|L\| < 1$ and the Lagrangian is inf-sharp in the sense of Definition 12 with respect to the set of saddle point $S \neq \emptyset$. If the starting point (x_0, y_0) satisfies*

$$\text{dist}((x_0, y_0), S) < \frac{\mu}{\max \left\{ \frac{1}{2\sigma}, \frac{1}{2\tau} \right\} - A_1}, \quad (6.43)$$

where $A_1 = \min \left\{ \frac{1-\theta\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\sqrt{\sigma\tau}\|L\|}{2\sigma} \right\}$. Then (x_n, y_n) converges to a saddle point.

The condition (6.43) gives us the radius of an open ball around the set S such that Algorithm 6 converges to a saddle point provided we start inside the ball. When starting outside the ball, the sequences can converges to other points but the saddle point. We will observe this behavior in the numerical example below.

6.3 Primal-Dual algorithm using $\Phi_{lsc}^{\mathbb{R}}$ -prox

In this section, we apply our newly defined $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator (5.4) to propose a primal-dual algorithm and investigate its convergence. We consider the primal problem (6.1) and its Lagrange dual (6.11), where $g : Y \rightarrow (-\infty, +\infty]$ is proper lsc convex, $f : X \rightarrow (-\infty, +\infty]$ is $\Phi_{lsc}^{\mathbb{R}}$ -convex and $L : X \rightarrow Y$ is a bounded linear operator. Recall that $(\hat{x}, \hat{y}) \in X \times Y$ is a saddle point of Lagrangian in (6.11) if the following are satisfied

$$(\forall (x, y) \in X \times Y) \quad \mathcal{L}(x, \hat{y}) \geq \mathcal{L}(\hat{x}, \hat{y}) \geq \mathcal{L}(\hat{x}, y).$$

Applying this definition to the Lagrangian in (6.11), we have

$$\begin{aligned} (\forall x \in X) \quad f(x) - f(\hat{x}) &\geq -\langle \hat{y}, L(x - \hat{x}) \rangle, \\ (\forall y \in Y) \quad g^*(y) - g^*(\hat{y}) &\geq \langle L\hat{x}, L(y - \hat{y}) \rangle. \end{aligned}$$

We can rewrite the above inequalities in the form of subdifferentials

$$\begin{aligned} L\hat{x} &\in \partial g^*(\hat{y}), \\ (0, -L^*\hat{y}) &\in \partial_{lsc}^{\mathbb{R}} f(\hat{x}). \end{aligned} \tag{6.44}$$

This is known as KKT condition for the Lagrangian primal-dual problems (6.11).

In Algorithm 7, we use the update inspired by Chambolle-Pock Algorithm [25, Algorithm 1]. For this case, we take $\theta = 1$. By replacing the primal update of [25, Algorithm 1] with $\Phi_{lsc}^{\mathbb{R}}$ -proximal update below, we call this new algorithm $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock Algorithm and show that we can obtain

convergence without relying on sharpness condition.

Algorithm 7: $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock Algorithm

Initialize: Choose $\tau, \sigma > 0$, $(x_0, y_0) \in X \times Y$ and $\bar{x}_0 = x_0$

Update: For $n \in \mathbb{N}$,

- $y_{n+1} = \operatorname{argmin}_{y \in Y} g^*(y) + \frac{1}{2\tau} \|y - y_n - \tau L \bar{x}_n\|^2$
 - We update the primal step as follow:
 - Pick $(a_n, (\frac{1}{\gamma} + 2a_n)x_n) \in J_\gamma(x_n)$ according to (5.1)
 - Pick $x_{n+1} \in \operatorname{argmin}_{z \in X} f(z) + \langle L^* y_{n+1}, z \rangle + \left(\frac{1}{2\gamma} + a_n\right) \|z - x_n\|^2$
 - $\bar{x}_{n+1} = 2x_{n+1} - x_n$
-

According formula (5.4), the $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator is defined as

$$\operatorname{prox}_{\gamma f}^{\operatorname{lsc}, \mathbb{R}}(x) := (J_\gamma + \partial_{lsc}^{\mathbb{R}} f)^{-1} J_\gamma(x),$$

which we use to update our primal variable $(x_n)_{n \in \mathbb{N}}$. We present our main results in the next theorem.

Theorem 20. *Let $f : X \rightarrow (-\infty, +\infty]$ be a proper $\Phi_{lsc}^{\mathbb{R}}$ -convex, $g : Y \rightarrow (-\infty, +\infty]$ be a proper lsc convex function and $L : X \rightarrow Y$ be a bounded linear operator. Let $(x_n)_{n \in \mathbb{N}}$, $(y_n)_{n \in \mathbb{N}}$, $(\bar{x}_n)_{n \in \mathbb{N}}$ and $(a_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 7. Assume that $\tau\sigma\|L\|^2 < 1$ and $1 + 2\sigma a_n > \sqrt{\sigma\tau}\|L\|$ for all $n \in \mathbb{N}$, then we have*

(i) For any $(x, y) \in X \times Y$,

$$\begin{aligned} & \sum_{n=0}^N \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\ & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{N+1}\|^2 - \frac{1}{2\tau} \|y - y_0\|^2 \\ & + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x - x_0\|^2, \end{aligned} \quad (6.45)$$

where $\mathcal{L}(x, y) = f(x) + \langle Lx, y \rangle - g^*(y)$ is the Lagrangian.

(ii) The sequences $(x_n, y_n)_{n \in \mathbb{N}}$ are bounded.

(iii) If X, Y are finite dimensional and $(a_n)_{n \in \mathbb{N}}$ converges, then $(x_n, y_n)_{n \in \mathbb{N}}$ converges to a saddle point in the sense of (6.44).

Proof. From the update of Algorithm 7, we have

$$\begin{aligned} \frac{y_n - y_{n+1}}{\tau} + L\bar{x}_n & \in \partial g^*(y_{n+1}), \\ \phi_n - \phi_{n+1} & \in \partial_{lsc}^{\mathbb{R}}(f + \langle L^* y_{n+1}, \cdot \rangle)(x_{n+1}). \end{aligned} \quad (6.46)$$

In the same way as in Proposition 17, we obtain an estimation analogous to (6.27),

$$\begin{aligned} & \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\ & \geq \frac{1}{2\tau} [\|y - y_{n+1}\|^2 + \|y_n - y_{n+1}\|^2 - \|y - y_n\|^2] \\ & + \left(\frac{1}{2\sigma} + a_{n+1} \right) \|x - x_{n+1}\|^2 + \left(\frac{1}{2\sigma} + a_n \right) (\|x_n - x_{n+1}\|^2 - \|x - x_n\|^2) \\ & + \langle L\bar{x}_n, y - y_{n+1} \rangle - \langle y_{n+1}, L(x - x_{n+1}) \rangle + \langle y_{n+1}, Lx \rangle - \langle y, Lx_{n+1} \rangle. \end{aligned} \quad (6.47)$$

By using $\bar{x}_n = 2x_n - x_{n-1}$ in the last line of (6.47),

$$\begin{aligned}
& \langle L\bar{x}_n, y - y_{n+1} \rangle - \langle y_{n+1}, L(x - x_{n+1}) \rangle + \langle y_{n+1}, Lx \rangle - \langle y, Lx_{n+1} \rangle \\
&= \langle L(2x_n - x_{n-1} - x_{n+1}), y - y_{n+1} \rangle \\
&= \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle - \langle L(x_{n-1} - x_n), y - y_n \rangle + \langle L(x_n - x_{n-1}), y_n - y_{n+1} \rangle \\
&\leq \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle - \langle L(x_{n-1} - x_n), y - y_n \rangle \\
&\quad + \frac{\sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 + \frac{\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n-1}\|^2.
\end{aligned} \tag{6.48}$$

Plugging (6.48) back into (6.47),

$$\begin{aligned}
& \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
&\geq \frac{1}{2\tau} [\|y - y_{n+1}\|^2 - \|y - y_n\|^2] + \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_{n+1} \right) \|x - x_{n+1}\|^2 - \left(\frac{1}{2\sigma} + a_n \right) \|x - x_n\|^2 \\
&\quad + \langle L(x_n - x_{n+1}), y - y_{n+1} \rangle - \langle L(x_{n-1} - x_n), y - y_n \rangle \\
&\quad + \left(\frac{1}{2\sigma} + a_n \right) \|x_n - x_{n+1}\|^2 - \frac{\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_n - x_{n-1}\|^2.
\end{aligned} \tag{6.49}$$

Summing (6.49) to $N \in \mathbb{N}$, and we obtain

$$\begin{aligned}
& \sum_{n=0}^N \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\
&\geq \frac{1}{2\tau} [\|y - y_{N+1}\|^2 - \|y - y_0\|^2] + \sum_{n=0}^N \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x - x_0\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_N \right) \|x_N - x_{N+1}\|^2 + \sum_{n=0}^{N-1} \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_{n-1} \right) \|x_n - x_{n-1}\|^2 \\
&\quad + \langle L(x_N - x_{N+1}), y - y_{N+1} \rangle.
\end{aligned} \tag{6.50}$$

We separate the inner product term on RHS of (6.50) using Young's inequality

$$\langle L(x_N - x_{N+1}), y - y_{N+1} \rangle \leq \frac{\sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{N+1}\|^2 + \frac{\sqrt{\sigma\tau}\|L\|}{2\sigma} \|x_N - x_{N+1}\|^2.$$

Hence, we have

$$\begin{aligned} & \sum_{n=0}^N \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) \\ & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{N+1}\|^2 - \frac{1}{2\tau} \|y - y_0\|^2 + \sum_{n=0}^N \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\ & + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x - x_0\|^2 \\ & + \sum_{n=0}^N \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_n \right) \|x_n - x_{n+1}\|^2. \end{aligned} \quad (6.51)$$

By assumption, $\sqrt{\sigma\tau}\|L\| < 1$ and $1 + 2\sigma a_n > \sqrt{\sigma\tau}\|L\|$ for all $n \in \mathbb{N}$, we have

$$\begin{aligned} \sum_{n=0}^N \mathcal{L}(x, y_{n+1}) - \mathcal{L}(x_{n+1}, y) & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y - y_{N+1}\|^2 - \frac{1}{2\tau} \|y - y_0\|^2 \\ & + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x - x_0\|^2, \end{aligned}$$

which is (6.45). Since (6.45) holds for all $(x, y) \in X \times Y$, we can take $(x, y) = (x^*, y^*)$ as the saddle point. By (6.13), we have

$$\begin{aligned} 0 & \geq \sum_{n=0}^N \mathcal{L}(x^*, y_{n+1}) - \mathcal{L}(x_{n+1}, y^*) \\ & \geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y^* - y_{N+1}\|^2 - \frac{1}{2\tau} \|y^* - y_0\|^2 \\ & + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x^* - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x^* - x_0\|^2. \end{aligned}$$

This implies that $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ are bounded.

Let X, Y be finite dimensional. In view of the above, let $(x_{n_k})_{k \in \mathbb{N}}, (y_{n_k})_{k \in \mathbb{N}}$ be the subsequences of $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ which converge to $\bar{x}$ and $\bar{y}$, respectively. By (6.51), with $(x, y) = (x^*, y^*)$ as a saddle point, we have

$$\begin{aligned}
0 &\geq \sum_{n=0}^N \mathcal{L}(x^*, y_{n+1}) - \mathcal{L}(x_{n+1}, y^*) \\
&\geq \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y^* - y_{N+1}\|^2 - \frac{1}{2\tau} \|y^* - y_0\|^2 + \sum_{n=0}^N \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_{N+1} \right) \|x^* - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x^* - x_0\|^2 \\
&\quad + \sum_{n=0}^N \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_n \right) \|x_n - x_{n+1}\|^2 \\
&\geq \sum_{n=0}^N \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 + \sum_{n=0}^N \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_n \right) \|x_n - x_{n+1}\|^2 \\
&\quad - \frac{1}{2\tau} \|y^* - y_0\|^2 - \left(\frac{1}{2\sigma} + a_0 \right) \|x^* - x_0\|^2.
\end{aligned}$$

Letting $N \rightarrow \infty$, we have the two sums above are finite. This means

$$\|y_n - y_{n+1}\|^2 \rightarrow 0, \quad \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_n \right) \|x_n - x_{n+1}\|^2 \rightarrow 0.$$

By assumption, $1 + 2\sigma a_n > \sqrt{\sigma\tau}\|L\|$ for all $n \in \mathbb{N}$ and $(a_n)_{n \in \mathbb{N}}$ converges, we infer

$$\lim_{n \rightarrow \infty} \|x_n - x_{n+1}\| = \lim_{n \rightarrow \infty} \|y_n - y_{n+1}\| = 0.$$

Then $x_{n_k+1} \rightarrow \bar{x}$ and $y_{n_k+1} \rightarrow \bar{y}$ as well. Now we want to prove that $(\bar{x}, \bar{y})$ is a saddle point. By (6.46), we have, for every $y \in Y$,

$$g^*(y) - g^*(y_{n_k+1}) \geq \left\langle \frac{y_{n_k} - y_{n_k+1}}{\tau} + L\bar{x}_{n_k}, y - y_{n_k+1} \right\rangle.$$

Since g^* is lsc convex [6, Proposition 13.13], letting $k \rightarrow \infty$, we have, for

every $y \in Y$,

$$\begin{aligned}
g^*(y) - g^*(\bar{y}) &\geq \liminf_{k \rightarrow \infty} g^*(y) - g^*(y_{n_k+1}) \\
&\geq \liminf_{k \rightarrow \infty} \left\langle \frac{y_{n_k} - y_{n_k+1}}{\tau} + L\bar{x}_{n_k} \cdot y - y_{n_k+1} \right\rangle \\
&= \langle L\bar{x}, y - \bar{y} \rangle.
\end{aligned} \tag{6.52}$$

Similarly for lsc f , by (6.51), for every $x \in X$,

$$\begin{aligned}
f(x) - f(x_{n_k+1}) &\geq (\phi_{n_k} - \phi_{n_k+1})(x) - (\phi_{n_k} - \phi_{n_k+1})(x_{n_k+1}) \\
&\quad - \langle L^*y_{n_k+1}, x - x_{n_k+1} \rangle \\
&= \left(\frac{1}{2\sigma} + a_{n_k+1} \right) \|x - x_{n_k+1}\|^2 - \left(\frac{1}{2\sigma} + a_{n_k} \right) \|x - x_{n_k}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_{n_k} \right) \|x_{n_k+1} - x_{n_k}\|^2 - \langle L^*y_{n_k+1}, x - x_{n_k+1} \rangle,
\end{aligned}$$

where $\phi_n = (a_n, (1/\sigma + 2a_n)x_n) \in \Phi_{lsc}^{\mathbb{R}}$, the second equality can be followed as in the proof of Proposition 12 for (5.25).

Letting $k \rightarrow \infty$ and note that $(a_n)_{n \in \mathbb{N}}$ converges, we obtain

$$\begin{aligned}
f(x) - f(\bar{x}) &\geq \liminf_{k \rightarrow \infty} f(x) - f(x_{n_k+1}) \\
&\geq \liminf_{k \rightarrow \infty} \left[\left(\frac{1}{2\sigma} + a_{n_k+1} \right) \|x - x_{n_k+1}\|^2 - \left(\frac{1}{2\sigma} + a_{n_k} \right) \|x - x_{n_k}\|^2 \right. \\
&\quad \left. + \left(\frac{1}{2\sigma} + a_{n_k} \right) \|x_{n_k+1} - x_{n_k}\|^2 - \langle L^*y_{n_k+1}, x - x_{n_k+1} \rangle \right] \\
&\geq -\langle L^*\bar{y}, x - \bar{x} \rangle.
\end{aligned} \tag{6.53}$$

As (6.53) and (6.52) hold for all $(x, y) \in X \times Y$, this means that $(\bar{x}, \bar{y})$ satisfies the KKT condition (6.44) and so it is a saddle point.

Next, we prove that the sequence $(x_n, y_n)_{n \in \mathbb{N}}$ converges to $(\bar{x}, \bar{y})$. Let us

consider the sum of (6.49) from n_k to $N > n_k$ with $(x, y) = (\bar{x}, \bar{y})$ and obtain

$$\begin{aligned}
0 &\geq \sum_{n=n_k}^N \mathcal{L}(\bar{x}, y_{n+1}) - \mathcal{L}(x_{n+1}, \bar{y}) \\
&\geq \frac{1}{2\tau} [\|\bar{y} - y_{N+1}\|^2 - \|\bar{y} - y_{n_k}\|^2] + \sum_{n=n_k}^N \frac{1 - \sqrt{\sigma\tau}\|L\|}{2\tau} \|y_n - y_{n+1}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_{N+1}\right) \|\bar{x} - x_{N+1}\|^2 - \left(\frac{1}{2\sigma} + a_{n_k}\right) \|\bar{x} - x_{n_k}\|^2 \\
&\quad + \left(\frac{1}{2\sigma} + a_N\right) \|x_N - x_{N+1}\|^2 + \sum_{n=n_k}^{N-1} \left(\frac{1 - \sqrt{\sigma\tau}\|L\|}{2\sigma} + a_{n-1}\right) \|x_n - x_{n+1}\|^2 \\
&\quad + \langle L(x_N - x_{N+1}), \bar{y} - y_{N+1} \rangle - \langle L(x_{n_k-1} - x_{n_k}), \bar{y} - y_{n_k} \rangle.
\end{aligned}$$

As $1 + 2\sigma a_n > 0$ for all $n \in \mathbb{N}$, we have

$$\begin{aligned}
&\frac{1}{2\tau} \|\bar{y} - y_{n_k}\|^2 + \left(\frac{1}{2\sigma} + a_{n_k}\right) \|\bar{x} - x_{n_k}\|^2 \\
&\geq \frac{1}{2\tau} \|\bar{y} - y_{N+1}\|^2 + \left(\frac{1}{2\sigma} + a_{N+1}\right) \|\bar{x} - x_{N+1}\|^2 \\
&\quad + \langle L(x_N - x_{N+1}), \bar{y} - y_{N+1} \rangle - \langle L(x_{n_k-1} - x_{n_k}), \bar{y} - y_{n_k} \rangle. \tag{6.54}
\end{aligned}$$

Letting $k \rightarrow \infty, N \rightarrow \infty$, the LHS and inner product terms of (6.54) go to zero, which implies $x_{N+1} \rightarrow \bar{x}$ and $y_{N+1} \rightarrow \bar{y}$. Hence, the whole sequences $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ converge to $\bar{x}$ and $\bar{y}$, respectively. $\square$

Remark 22. While we only obtain convergence in finite dimensional case, the results in Theorem 20 (iii) are obtained for a large class of functions which are nonconvex. For instance, in contrast to Algorithm 5 and Algorithm 6, we do not assume any sharpness condition to derive convergence as in Theorem 18 and Theorem 19. One important reason of this fact is that we have calculated $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator (5.4), while requiring $(a_n)_{n \in \mathbb{N}}$ converges in order to obtain convergence of the constructed sequences to a saddle point. Nonetheless, we do not need to rely on a stronger assumption such as inf-

sharpness like in the previous section.

6.4 Numerical Examples

We demonstrate the performance of Algorithm 5 with several examples.

Example 22. We recall Lagrangian in Example 20

$$\mathcal{L}(x, y) = |x| + xy - |y|.$$

The respective primal and dual problems for this Lagrangian are

$$\inf_{-1 \leq x \leq 1} |x|, \quad \sup_{-1 \leq y \leq 1} -|y|.$$

Setting: We set $\tau = 0.25, \sigma = 0.75, \theta = 1$, maximum number of iteration $N = 2001$. We randomized the starting point in the interval $[-10, 10]$. For each step, we calculate the proximal operator of the absolute function $f(x) = |x|$ which has an explicit form,

$$\text{prox}_{\gamma f}(x_0) = \begin{cases} x_0 - \gamma & x_0 \geq \gamma, \\ x_0 + \gamma & x_0 \leq -\gamma, \\ 0 & -\gamma < x_0 < \gamma. \end{cases}$$

We plot the distance at each iteration to the saddle point and the primal, dual iterate in Figure 6.6. This is a special case because by the nature of proximal operator for f , which return zero whenever the iterate is closed enough. That is why just after a few iterations, both primal and dual iterate take zero value.

Example 23. We consider the Lagrangian

$$\mathcal{L}(x, y) = |x| + |x^2 - 2| + xy - |y| - |y^2 - 2|,$$

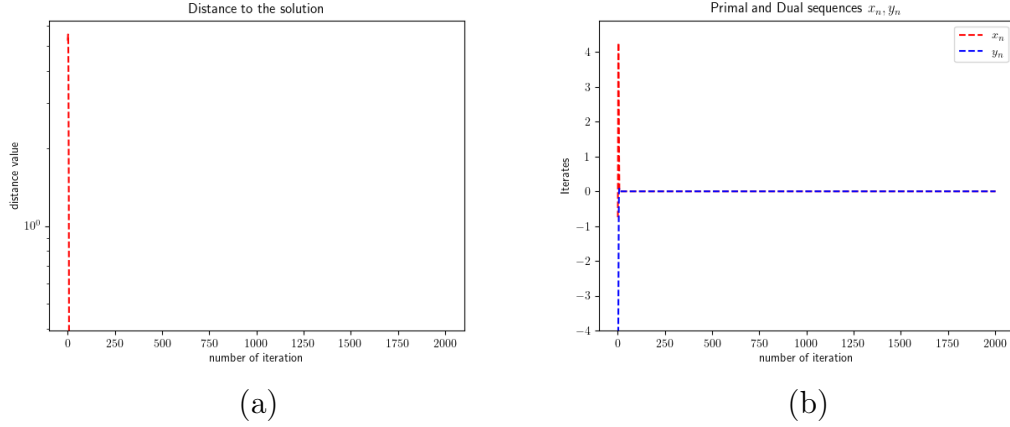


Figure 6.6: Example 22. From left to right: distance to the solution; primal-dual iterates.

which has a similar form as in Example 21. It also has one saddle point at $(0,0)$. We run Algorithm 5 for this Lagrangian in two cases: inside and outside the region mentioned in Theorem 18. To compute the primal and dual update of Algorithm 5, we use Scipy package in Python to approximate the proximal operator in each step.

Setting: We set $\tau = 0.25, \sigma = 0.35, \theta = 1$ and we take sharpness constant $\mu = 0.9$. The weakly convex modulus $\rho = 2$ so the quantity

$$\frac{\mu}{\max\left\{\frac{1}{2\sigma}, \frac{1}{2\tau}\right\} - A} \approx 0.3199, \quad (6.55)$$

where $A = \min\left\{\frac{1-\sqrt{\sigma\tau}\|L\|}{2\tau}, \frac{1-\sigma\rho-\theta\sqrt{\sigma\tau}\|L\|}{2\sigma}\right\} \approx 0.00599$. For the first case, we randomize the starting points $(x_0, y_0) \in [-0.3199, 0.3199]^2$ and for the second case, $[-10, 10]^2$. We plot the distance of the iterate to the saddle point and primal/ dual iterate in Figure 6.7 for the first case and Figure 6.8 for the second case. Notice the differences in two figures as the sequences converge to different points from the saddle point in the second case.

For $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock Algorithm 7, we give two illustrative examples.

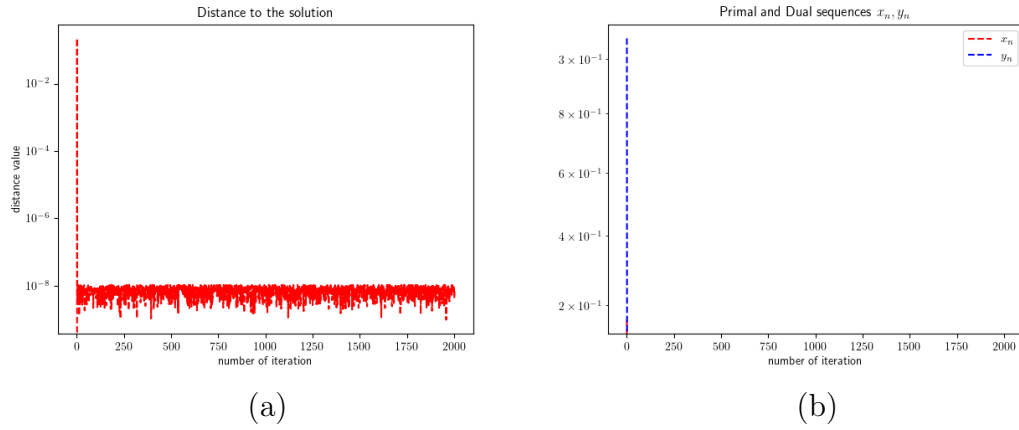


Figure 6.7: Example 23 first case. From left to right: distance to the solution; primal-dual iterates.

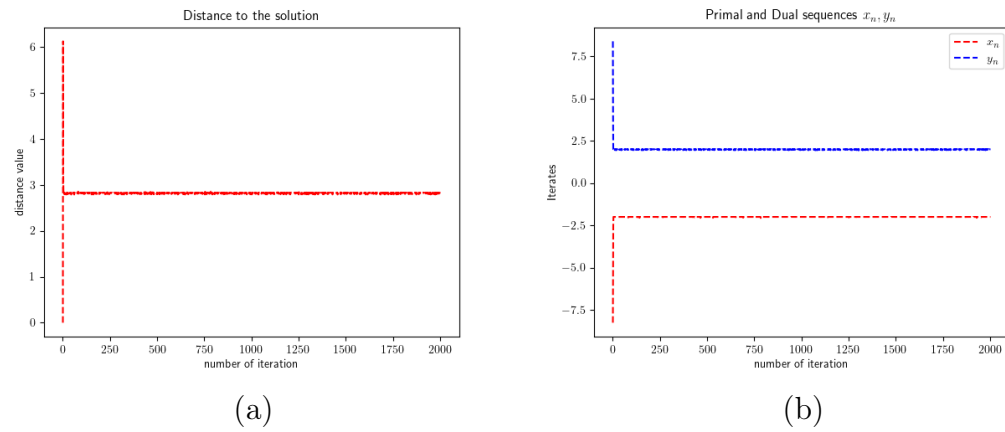


Figure 6.8: Example 23 second case. From left to right: distance to the solution; primal-dual iterates.

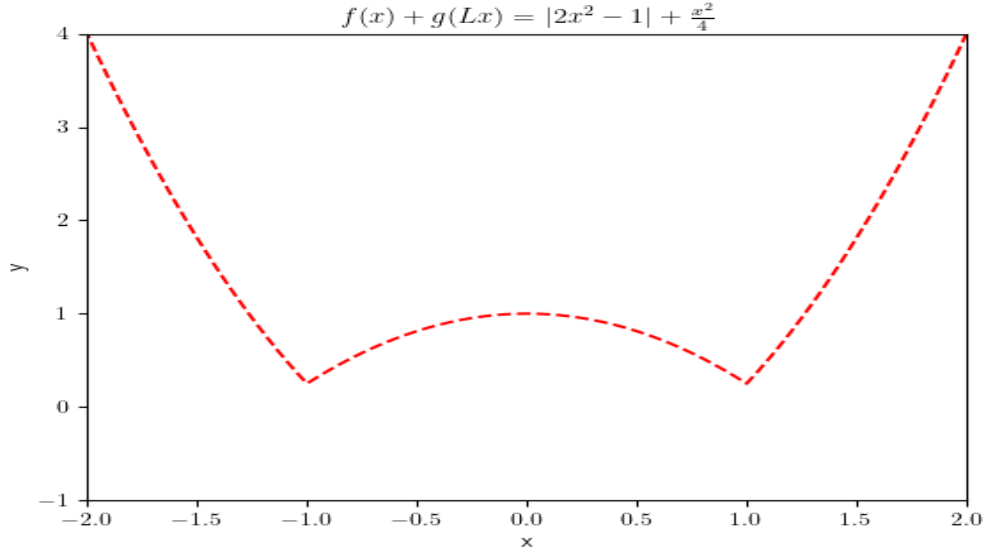


Figure 6.9: Objective function in Example 24.

Example 24. Consider the following problem in the real line

$$\min_{x \in \mathbb{R}} 2|x^2 - 1| + \frac{x^2}{4}. \quad (6.56)$$

We can treat this problem as a composite minimization (6.1) with $f(x) = 2|x^2 - 1|$, $g(y) = y^2$ and $L = 1/2$ is the scaling factor. The objective function is illustrated in figure 6.9. It is easy to identify the minimizers which is $x = \pm 1$. Thanks to the structure of function g , we can calculate g^* which is

$$g^*(y) = \frac{y^2}{4}.$$

Hence, the updates in Algorithm 7 looks like this

$$\begin{aligned} y_{n+1} &= \operatorname{argmin}_{y \in Y} \frac{y^2}{4} + \frac{1}{2\tau} (y - y_n - \tau L \bar{x}_n)^2, \\ x_{n+1} &\in \operatorname{prox}_{\sigma(f + \langle L^* y_{n+1}, \cdot \rangle)}^{\operatorname{lsc}, \mathbb{R}}(x_n), \\ \bar{x}_{n+1} &= 2x_{n+1} - x_n. \end{aligned}$$

The stepsizes $\tau, \sigma > 0$ are chosen such that $\tau\sigma\|L\|^2 < 1$. We can write the dual update explicitly as

$$y_{n+1} = \frac{2(y_n + \tau L \bar{x}_n)}{\tau + 2}.$$

While for the primal update, from Theorem 11,

$$x_{n+1} \in \operatorname{argmin}_{x \in \mathbb{R}} 2|x^2 - 1| + y_{n+1}Lx + \left(\frac{1}{2\sigma} + a_n\right)(x - x_n)^2,$$

where $a_n \geq -\frac{1}{2\sigma}$ will be chosen later.

Setting: We take $\tau = 0.25, \sigma = 0.75$ and fix the number of iteration $N = 2001$. The parameter $a_0 = 0.1$ and updating rule will be

$$a_{n+1} = a_n + 0.1.$$

For starting points, (x_0, y_0) will be randomized uniformly in $[-10, 10]$. To find x_{n+1} , we use approximation method BFGS which is available in Scipy package (Python). For illustration, we plot the following: the objective function value, the distance of the primal iterate to the solution and the behavior of the dual iterate (see figure 6.10).

Example 25 (Binary tomography). Binary Tomography (BT) is a special case of Discrete Tomography (DT) which aims at reconstructing binary images starting from a limited number of their projections [84, 61]. The image

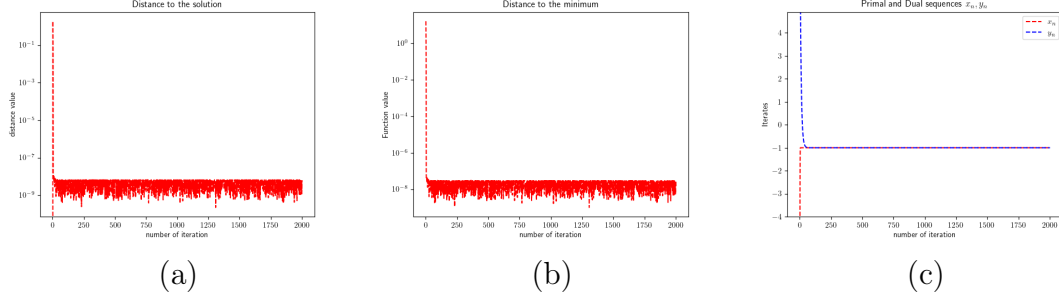


Figure 6.10: From left to right: distance from the iterate to the minimizer; distance to the optimal value; primal and dual iterates.

$\bar{x} \in \mathbb{R}^n$ is represented as a grid of $n = n_1 \times n_2$ pixels taking values $x_j \in \{-1, 1\}$ for $j = 1, \dots, n$. The projections y_i for $i = 1, \dots, m$ are linear combinations of the pixels along m directions. The linear transformation from image to projections is modelled as

$$y = \mathbf{A}\bar{x} + \omega, \quad (6.57)$$

where x_j denotes the value of the image in the j -th cell, y_i is the weighted sum of the image along the i -th ray, the element $\mathbf{A}_{i,j}$ of matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is proportional to the length of the i -th ray in the j -th cell and finally $\omega \in \mathbb{R}^m$ is an additive noise with Gaussian distribution and standard deviation $\sigma > 0$. In general, the projection matrix $\mathbf{A}$ has a low rank, meaning that $\text{rank}(\mathbf{A}) < \min\{n, m\}$ [61]. We model the problem as

$$\underset{x \in \mathbb{R}^n, x_i \in \{-1, 1\}}{\text{minimize}} \quad \frac{\|\mathbf{A}x - y\|_2^2}{2}. \quad (6.58)$$

Instead of considering the binary variable x , we can turn problem (6.58) into a continuous one by adding the function $F : \mathbb{R}^n \rightarrow \mathbb{R}$ defined as

$$F(x) = \sum_{i=1}^n |x_i^2 - 1|, \quad (6.59)$$

which is separable as the sum of functions $f_i(x) = |x_i^2 - 1|$. Observe that F

is a weakly convex function with the set of minimisers

$$S_F = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \in \{-1, 1\} \text{ for } i = 1, \dots, n\} = \{-1, 1\}^n,$$

and can be seen as a non-smooth counterpart to the function $x \mapsto \sum_{i=1}^n x_i^2(x_i - 1)^2$, $x \in \mathbb{R}^n$, proposed in [94] to promote binary integer solutions (with values in $\{0, 1\}^n$) in Markov Random Field approaches. To the best of our knowledge, this is the first work where function F is used in such a context.

Problem (6.58) transforms into

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{\|\mathbf{A}x - y\|_2^2}{2} + F(x). \quad (6.60)$$

We can write the objective function of (6.60) in the form of composite functions $F(x) + G(Ax)$ and implement $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock Algorithm 7 to solve this problem.

The function $G : \mathbb{R}^m \rightarrow \mathbb{R}$ is $G(z) = \frac{\|z - y\|^2}{2}$ which has the conjugate

$$G^*(v) = \frac{\|v\|^2}{2} + \langle v, y \rangle,$$

and we can write the dual update similar to the previous examples, which is

$$v_{n+1} = \frac{v_n + \tau(A\bar{x}_n - y)}{\tau + 1}.$$

For the primal step, it can be calculated as a proximal step as we mention in Algorithm 1 from Chapter 5. Specifically, since the function F is weakly convex, we can obtain an explicit form of convex proximal operator for each component of F i.e. the function $\mathcal{F}(x) = |x^2 - 1|$ (see [11, Remark 8]). For

$\alpha > 0, y \in \mathbb{R}$, we have

$$(\forall y \in \mathbb{R}) \quad \text{prox}_{\alpha\mathcal{F}}(y) = \begin{cases} \frac{y}{1+2\alpha} & \text{if } |y| > 1 + 2\alpha, \\ \frac{y}{1-2\alpha} & \text{if } |y| < 1 - 2\alpha, \\ \frac{y}{|y|} & \text{otherwise.} \end{cases} \quad (6.61)$$

Remark 23. The model in (6.58) (for appropriate choices of matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, vector $y \in \mathbb{R}^m$ and possible additional convex constraints) could be applied to other Binary Quadratic Programs (a special class of QCQP problems having the equality constraint $x_i^2 = 1$ for every $i \in \{1, \dots, n\}$) arising in Computer Vision, such as Graph Bisection, Graph Matching and Image Co-Segmentation (see [91, Table 2] and the references therein).

Setting: For $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock Algorithm, we fix the number of iteration to 500, stepsizes $\tau = \sigma = \frac{0.9}{\|A\|}$ to ensure that $\tau\sigma\|A\|^2 < 1$. For the parameter a_n , we start with $a_0 = 100$ and updating $a_{n+1} = a_n - \frac{1}{n^2}$.

Numerical tests For our simulations we used the MATLAB codes and data from [61, 60]. We reproduced the same synthetic setting and test from simple to complex images: Consider multiples phantoms (*Apple, Bird, Lizard, Octopus, Spring, Horse, Pocket Watch* and *Blood Vessels*), corresponding to binary images of size 64 x 64 pixels. Operator $\mathbf{A}$ models an X-ray tomographic scan with 64 detectors and a parallel beam acquisition geometry with four angles ($0^\circ, 50^\circ, 100^\circ, 150^\circ$). We set $\sigma = 0.01$ for the additive noise. Figure 6.11 illustrates, from left to right, the original image and the reconstructions obtained with the Least Squares QR method (LSQR), with the Truncated Least Squares QR method (TLSQR), with the DUAL method proposed in [61] and finally our proposed Algorithm 7 applied to (6.60), which we refer to as $\Phi_{lsc}^{\mathbb{R}}$ -Chambolle-Pock method ($\Phi_{lsc}^{\mathbb{R}}$ -CP). All methods are initialised with a vector of zeros.

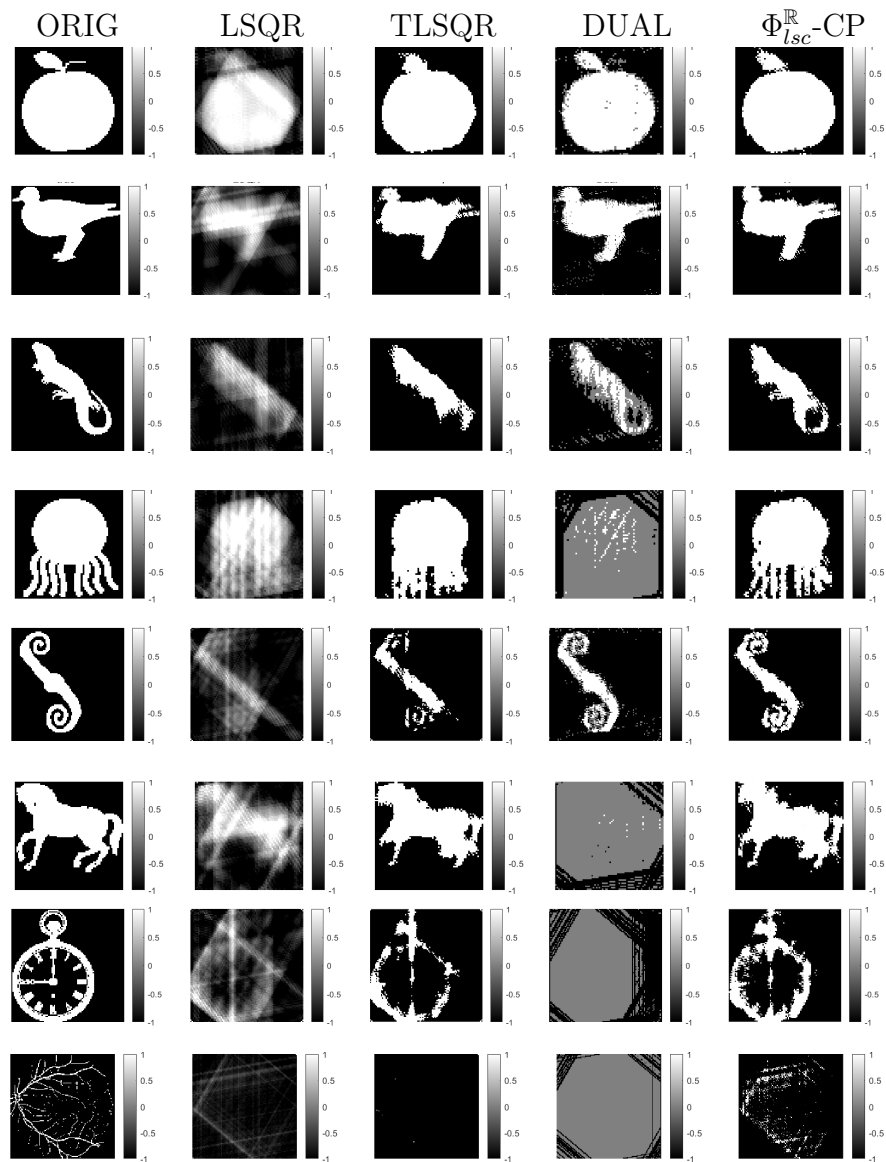


Figure 6.11: Comparisons between different solutions. From left to right: original image, least squares solution, thresholded least squares solution, solution obtained with the dual method proposed in [61], solution obtained with our model.

Chapter 7

Conclusion and Future Work

The thesis investigates the class of abstract convex functions which are minorized by elementary functions with specific structure Φ . We point out that a proper lsc convex function is also abstract convex with respect to the class of affine function. More generally, the function is also abstract convex when the class of elementary functions is quadratic ($\Phi_{lsc}^{\mathbb{R}}$). The class of $\Phi_{lsc}^{\mathbb{R}}$ -convex functions also consists of strongly convex and weakly convex functions.

Together with abstract convexity, one can define abstract conjugate and abstract subdifferentials which take on similar forms as in classical convex conjugate and convex subdifferentials. With these powerful tools, we successfully construct an abstract conjugate dual problem to the general composite minimization problem where the objective functions are nonconvex. By using abstract subdifferentials, we prove zero duality where the primal and dual problems have the same value. We move one step further by establishing conditions which strong duality is obtained.

On the other hand, conjugate dual problem can be written as Lagrange dual problem. We then determine Lagrangian and define Lagrange primal problem and state the connection to the composite minimisation problem. By making use of "intersection property", we manage to prove Lagrange zero duality gap and also conjugate zero duality gap from Lagrange duality.

The second part of this work is about algorithms. The final goal is to have a primal-dual algorithm for solving the composite minimisation problem through abstract conjugate dual. Instead of considering a general abstract function, we limit ourselves to the class of $\Phi_{lsc}^{\mathbb{R}}$ as elementary functions. This helps us define $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator and propose $\Phi_{lsc}^{\mathbb{R}}$ -proximal based algorithms such as $\Phi_{lsc}^{\mathbb{R}}$ -proximal point, $\Phi_{lsc}^{\mathbb{R}}$ -Forward-Backward or $\Phi_{lsc}^{\mathbb{R}}$ -Projected Subgradient for several types of minimisation problems. We also give a detailed convergence analysis of these algorithms and remarks the similarities to the one in convex analysis.

For primal-dual algorithm, we manage to apply the new $\Phi_{lsc}^{\mathbb{R}}$ -proximal operator to conjugate primal-dual problems where one function is convex and other is $\Phi_{lsc}^{\mathbb{R}}$ -convex. Since $\Phi_{lsc}^{\mathbb{R}}$ -convex functions are related to weakly convex, we turn our attention to conjugated primal-dual problems where one function is convex and other is weakly convex. We propose different types of primal-dual updates and investigate their convergence. To obtain convergence, we assume that the Lagrangian is sharp with respect to the set of saddle points. Otherwise, if the primal objective function is sharp to its solutions, then primal-dual algorithm converges under certain assumptions.

During and after this work, there are several questions which I cannot answer or have not gone into it. These questions can serve as starting points for future discussions.

- (i) **Subdifferentials sum rule:** We want to find conditions which abstract subdifferentials of sum of functions can be expressed as the sum of subdifferentials of each function. From the observation of zero duality gap i.e. Theorem 4 and Corollary 2, there is a connection between conjugate duality and subdifferentials sum rule. It is a difficult problem if we consider two different functions with two different classes of elementary functions.
- (ii) **Abstract convex set:** Even though we investigate conjugate duality and Lagrange duality using abstract convexity, we only use abstract

conjugate and abstract subdifferentials for functions. We have not touch the concept of abstract convex set and topological space in the presence of abstract convexity. This will help us develop a complete understanding of abstract convexity and the differences to the classical convexity.

- (iii) **Gradient/ subgradient algorithms for abstract convex function:** With the definition of abstract subdifferentials, one can think of subgradient methods to solve minimization problem. However, abstract subgradient is considered as a function, how can we make a subgradient update? We know that an abstract subgradient is a minorized function, then how can we define a direction toward minimizer using abstract subgradient?
- (iv) **$\Phi_{lsc}^{\mathbb{R}}$ -proximal algorithms:** As we see in Example 16, $\Phi_{lsc}^{\mathbb{R}}$ -Projected subgradient algorithm can terminate early. Similar situations can occur when the function is not convex. Hence, we want to find conditions which help us overcome this problem and make the algorithm run indefinitely.
- (v) **General abstract proximal operator:** We successfully define an abstract proximal operator and design proximal based algorithms for abstract convex functions. However, this is for only a specific class of elementary functions $\Phi_{lsc}^{\mathbb{R}}$. The next step would be to define an abstract proximal operator which can work on arbitrary class Φ and study its differences to the classical one in convex analysis.

Bibliography

- [1] Ya I Alber, Alfredo N Iusem, and Mikhail V Solodov. “On the projected subgradient method for nonsmooth convex optimization in a Hilbert space”. In: *Mathematical Programming* 81 (1998), pp. 23–35.
- [2] Mikhail Andramonov. “A survey of methods of abstract convex programming”. In: *Journal of Statistics and Management Systems* 5.1-3 (2002), pp. 21–37.
- [3] David Applegate et al. “Faster first-order primal-dual methods for linear programming using restarts and sharpness”. In: *Mathematical Programming* 201.1 (2023), pp. 133–184.
- [4] Felipe Atenas et al. “A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods”. In: *SIAM Journal on Optimization* 33.1 (2023), pp. 89–115.
- [5] Hedy Attouch and Dominique Aze. “Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method”. In: *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*. Vol. 10. 3. Elsevier. 1993, pp. 289–312.
- [6] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2nd. Springer Publishing Company, Incorporated, 2017. ISBN: 978-3-319-48310-8.

- [7] Amir Beck, Petre Stoica, and Jian Li. “Exact and approximate solutions of source localization problems”. In: *IEEE Transactions on signal processing* 56.5 (2008), pp. 1770–1778.
- [8] Amir Beck and Marc Teboulle. “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems”. In: *IEEE transactions on image processing* 18.11 (2009), pp. 2419–2434.
- [9] Ewa Bednarczuk and The Hung Tran. “Duality for composite optimization problem within the framework of abstract convexity”. In: *Optimization* 72.1 (2023), pp. 37–80.
- [10] Ewa Bednarczuk et al. “Calculus rules for proximal ε -subdifferentials and inexact proximity operators for weakly convex functions”. In: *2023 European Control Conference (ECC)*. IEEE. 2023, pp. 1–8.
- [11] Ewa Bednarczuk et al. “Convergence analysis of an inexact Forward-Backward algorithm for problems involving weakly convex functions”. In: *arXiv preprint arXiv:2303.14021* (2023).
- [12] Ewa M Bednarczuk and Monika Syga. “On duality for nonconvex minimization problems within the framework of abstract convexity”. In: *Optimization* 71.4 (2022), pp. 949–971.
- [13] Frédéric Bernard and Lionel Thibault. “Prox-regular functions in Hilbert spaces”. In: *Journal of Mathematical Analysis and Applications* 303.1 (2005), pp. 1–14.
- [14] Jerome Bolte, Shoham Sabach, and Marc Teboulle. “Proximal alternating linearized minimization for nonconvex and nonsmooth problems”. In: *Mathematical Programming* 146.1-2 (2014), pp. 459–494. DOI: 10.1007/s10107-013-0701-9. URL: <https://hal.inria.fr/hal-00916090>.
- [15] J Frédéric Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

- [16] Radu Ioan Bot. *Conjugate duality in convex optimization*. Vol. 637. Springer Science & Business Media, 2009.
- [17] Stephen Boyd, Lin Xiao, and Almir Mutapcic. “Subgradient methods”. In: *lecture notes of EE392o, Stanford University, Autumn Quarter 2004* (2003), pp. 2004–2005.
- [18] Kristian Bredies, Dirk Lorenz, et al. *Mathematical image processing*. Springer, 2018.
- [19] Hoa T Bui et al. “Zero duality gap conditions via abstract convexity”. In: *Optimization* 71.4 (2022), pp. 811–847.
- [20] Regina S Burachik. “On asymptotic Lagrangian duality for nonsmooth optimization”. In: *ANZIAM journal* 58 (2016), pp. C93–C123.
- [21] Regina Sandra Burachik and Alex Rubinov. “Abstract convexity and augmented Lagrangians”. In: *SIAM Journal on Optimization* 18.2 (2007), pp. 413–436.
- [22] Regina Sandra Burachik and Alex Rubinov. “On abstract convexity and set valued analysis”. In: *Journal of Nonlinear and Convex Analysis* 9.1 (2008), p. 105.
- [23] Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. Vol. 58. Springer Science & Business Media, 2004.
- [24] Antonin Chambolle and Juan Pablo Contreras. “Accelerated Bregman primal-dual methods applied to optimal transport and Wasserstein Barycenter problems”. In: *SIAM Journal on Mathematics of Data Science* 4.4 (2022), pp. 1369–1395.
- [25] Antonin Chambolle and Thomas Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. In: *Journal of mathematical imaging and vision* 40 (2011), pp. 120–145.

- [26] Antonin Chambolle and Thomas Pock. “On the ergodic convergence rates of a first-order primal–dual algorithm”. In: *Mathematical Programming* 159.1 (2016), pp. 253–287.
- [27] Antonin Chambolle et al. “Stochastic Primal–Dual Hybrid Gradient Algorithm with Adaptive Step Sizes”. In: *Journal of Mathematical Imaging and Vision* (2024), pp. 1–20.
- [28] Xiao-Kai Chang, Junfeng Yang, and Hongchao Zhang. “Golden ratio primal-dual algorithm with linesearch”. In: *SIAM Journal on Optimization* 32.3 (2022), pp. 1584–1613.
- [29] Xiaokai Chang and Junfeng Yang. “A Golden Ratio Primal–Dual Algorithm for Structured Convex Optimization”. In: *Journal of Scientific Computing* 87 (2021), pp. 1–26.
- [30] Xiaokai Chang and Junfeng Yang. “GRPDA Revisited: Relaxed Condition and Connection to Chambolle–Pock’s Primal–Dual Algorithm”. In: *Journal of Scientific Computing* 93.3 (2022), p. 70.
- [31] Shixiang Chen, Alfredo Garcia, and Shahin Shahrampour. “On distributed nonconvex optimization: Projected subgradient method for weakly convex problems in networks”. In: *IEEE Transactions on Automatic Control* 67.2 (2021), pp. 662–675.
- [32] Francis H Clarke et al. *Nonsmooth analysis and control theory*. Vol. 178. Springer Science & Business Media, 2008.
- [33] Christian Clason, Stanislav Mazurenko, and Tuomo Valkonen. “Primal–dual proximal splitting and generalized conjugation in non-smooth non-convex optimization”. In: *Applied Mathematics & Optimization* 84.2 (2021), pp. 1239–1284.
- [34] Matthew J Colbrook. “WARPd: A linearly convergent first-order primal–dual algorithm for inverse problems with approximate sharpness conditions”. In: *SIAM Journal on Imaging Sciences* 15.3 (2022), pp. 1539–1575.

-
- [35] Patrick L Combettes. “Quasi-Fejérian analysis of some optimization algorithms”. In: *Studies in Computational Mathematics*. Vol. 8. Elsevier, 2001, pp. 115–152.
 - [36] Nguyen Duy Cuong and Alexander Y Kruger. “Error bounds revisited”. In: *Optimization* 71.4 (2022), pp. 1021–1053.
 - [37] Aris Daniilidis and Pando Georgiev. “Approximate convexity and submonotonicity”. In: *Journal of mathematical analysis and applications* 291.1 (2004), pp. 292–301.
 - [38] Damek Davis and Dmitriy Drusvyatskiy. “Stochastic model-based minimization of weakly convex functions”. In: *SIAM Journal on Optimization* 29.1 (2019), pp. 207–239.
 - [39] Damek Davis et al. “Subgradient methods for sharp weakly convex functions”. In: *Journal of Optimization Theory and Applications* 179 (2018), pp. 962–982.
 - [40] Maxim Vladimirovich Dolgopolik. “Abstract convex approximations of nonsmooth functions”. In: *Optimization* 64.7 (2015), pp. 1439–1469.
 - [41] Dmitriy Drusvyatskiy, Alexander D Ioffe, and Adrian S Lewis. “Nonsmooth optimization using Taylor-like models: error bounds, convergence, and termination criteria”. In: *Mathematical Programming* 185 (2021), pp. 357–383.
 - [42] Dmitriy Drusvyatskiy and Adrian S Lewis. “Error bounds, quadratic growth, and linear convergence of proximal methods”. In: *Mathematics of Operations Research* 43.3 (2018), pp. 919–948.
 - [43] J Dutta, JE Martinez-Legaz, and AM Rubinov. “Monotonic analysis over cones: I”. In: *Optimization* 53.2 (2004), pp. 129–146.
 - [44] AC Eberhard and Hossein Mohebi. “Maximal abstract monotonicity and generalized Fenchel’s conjugation formulas”. In: *Set-Valued and Variational Analysis* 18.1 (2010), pp. 79–108.

- [45] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. SIAM, 1999.
- [46] Ernie Esser, Xiaoqun Zhang, and Tony F Chan. “A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science”. In: *SIAM Journal on Imaging Sciences* 3.4 (2010), pp. 1015–1046.
- [47] Olivier Fercoq. “Quadratic error bound of the smoothed gap and the restarted averaged primal-dual hybrid gradient”. In: *Open Journal of Mathematical Optimization* 4 (2023), pp. 1–34.
- [48] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.
- [49] VV Gorokhovich and AS Tykoun. “Abstract Convexity of Functions with Respect to the Set of Lipschitz (Concave) Functions”. In: *Proceedings of the Steklov Institute of Mathematics* 309.1 (2020), S36–S46.
- [50] VV Gorokhovich and AS Tykoun. “Support points of lower semicontinuous functions with respect to the set of Lipschitz concave functions”. In: *Doklady of the National Academy of Sciences of Belarus* 63.6 (2020), pp. 647–653.
- [51] Bingsheng He, Yanfei You, and Xiaoming Yuan. “On the convergence of primal-dual hybrid gradient algorithm”. In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2526–2537.
- [52] Bingsheng He and Xiaoming Yuan. “Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective”. In: *SIAM Journal on Imaging Sciences* 5.1 (2012), pp. 119–149.
- [53] Tim Hoheisel, Maxime Laborde, and Adam Oberman. “On proximal point-type algorithms for weakly convex functions and their connection to the backward euler method”. In: ().

- [54] XX Huang and XQ Yang. “Further study on augmented Lagrangian duality theory”. In: *Journal of Global Optimization* 31.2 (2005), pp. 193–210.
- [55] D. H. Hyers and S. M. Ulam. “Approximately convex functions”. In: *Proceedings of the American Mathematical Society* 3 (1952), pp. 821–828.
- [56] Vaithilingam Jeyakumar, AM Rubinov, and ZY Wu. “Generalized Fenchel’s conjugation formulas and duality for abstract convex functions”. In: *Journal of optimization theory and applications* 132.3 (2007), pp. 441–458.
- [57] Zhichao Jia and Benjamin Grimmer. “First-order methods for nonsmooth nonconvex functional constrained optimization with or without slater points”. In: *arXiv preprint arXiv:2212.00927* (2022).
- [58] He Jiang, Weihua Zheng, and Yao Dong. “Sparse and robust estimation with ridge minimax concave penalty”. In: *Information Sciences* 571 (2021), pp. 154–174.
- [59] Abderrahim Jourani. “Subdifferentiability and subdifferential monotonicity of 1-paraconvex functions”. In: *Control and Cybernetics* 25.4 (1996).
- [60] A. Kadu. *BinaryTomo Toolbox*. <https://github.com/ajinkyakadu/BinaryTomo>.
- [61] Ajinkya Kadu and Tristan van Leeuwen. “A convex formulation for binary tomography”. In: *IEEE Transactions on Computational Imaging* 6 (2019), pp. 1–11.
- [62] Xiao Li et al. “Incremental methods for weakly convex optimization”. In: *arXiv preprint arXiv:1907.11687* (2019).

- [63] Dirk Lorenz, Ewa Bednarczuk, and The Hung Tran. “Proximal Algorithms for a class of abstract convex functions”. In: *arXiv preprint arXiv:2402.17072* (2024).
- [64] Dirk A Lorenz and Felix Schneppe. “Chambolle–Pock’s Primal-Dual Method with Mismatched Adjoint”. In: *Applied Mathematics & Optimization* 87.2 (2023), p. 22.
- [65] D Russell Luke, Shoham Sabach, and Marc Teboulle. “Optimization on spheres: models and proximal algorithms with computational performance comparisons”. In: *SIAM Journal on Mathematics of Data Science* 1.3 (2019), pp. 408–445.
- [66] Yura Malitsky and Thomas Pock. “A first-order primal-dual algorithm with linesearch”. In: *SIAM Journal on Optimization* 28.1 (2018), pp. 411–432.
- [67] Stanislav Mazurenko, Jyrki Jauhainen, and Tuomo Valkonen. “Primal-dual block-proximal splitting for a class of non-convex problems”. In: *Electron. Trans. Numer. Anal.* 52 (2020), pp. 509–552. DOI: 10.1553/etna_vol52s509.
- [68] Thomas Mollenhoff et al. “The primal-dual hybrid gradient method for semiconvex splittings”. In: *SIAM Journal on Imaging Sciences* 8.2 (2015), pp. 827–857.
- [69] Jean Jacques Moreau. “Inf-convolution, sous-additivité, convexité des fonctions numériques”. In: *Journal de Mathématiques Pures et Appliquées* (1970).
- [70] Werner Oettli and Dirk Schläger. “Conjugate functions for convex and nonconvex duality”. In: *Journal of Global Optimization* 13.4 (1998), pp. 337–347.
- [71] Zsolt Páles. “On approximately convex functions”. In: *Proceedings of the American Mathematical Society* 131.1 (2003), pp. 243–252.

-
- [72] Diethard Ernst Pallaschke and Stefan Rolewicz. *Foundations of mathematical optimization: convex analysis without linearity*. Vol. 388. Springer Science & Business Media, 2013.
 - [73] Robert R Phelps. *Convex functions, monotone operators and differentiability*. Springer, 2009.
 - [74] Thomas Pock et al. “An algorithm for minimizing the Mumford-Shah functional”. In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE. 2009, pp. 1133–1140.
 - [75] Julian Rasch and Antonin Chambolle. “Inexact first-order primal–dual algorithms”. In: *Computational Optimization and Applications* 76.2 (2020), pp. 381–430.
 - [76] R Tyrrell Rockafellar. “Favorable classes of Lipschitz continuous functions in subgradient optimization”. In: (1981).
 - [77] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*. Vol. 317. Springer Science & Business Media, 2009.
 - [78] S Rolewicz. “On paraconvex multifunctions”. In: *Oper. Research Verf.(Methods of Oper Res)* 31 (1979), pp. 540–546.
 - [79] Stefan Rolewicz. “On uniformly approximate convex and strongly alpha (.)-paraconvex functions”. In: *Control and Cybernetics* 30.3 (2001), pp. 323–330.
 - [80] Vincent Roulet and Alexandre d’Aspremont. “Sharpness, restart and acceleration”. In: *Advances in Neural Information Processing Systems* 30 (2017).
 - [81] Alex M Rubinov, XX Huang, and XQ Yang. “The zero duality gap property and lower semicontinuity of the perturbation function”. In: *Mathematics of Operations Research* 27.4 (2002), pp. 775–791.
 - [82] Alexander M Rubinov. *Abstract convexity and global optimization*. Vol. 44. Springer Science & Business Media, 2013.

-
- [83] Alexander M Rubinov and M Yu Andramonov. “Minimizing increasing star-shaped functions based on abstract convexity”. In: *Journal of Global Optimization* 15.1 (1999), pp. 19–39.
 - [84] Thomas Schüle et al. “Discrete tomography by convex–concave regularization and DC programming”. In: *Discrete Applied Mathematics* 151.1-3 (2005), pp. 229–243.
 - [85] Ivan Singer. *Abstract convex analysis*. Vol. 25. John Wiley & Sons, 1997.
 - [86] Monika Syga. “Minimax theorems for ϕ -convex functions: sufficient and necessary conditions”. In: *Optimization* 65.3 (2016), pp. 635–649.
 - [87] Monika Syga. “Minimax theorems for extended real-valued abstract convex–concave functions”. In: *Journal of Optimization Theory and Applications* 176.2 (2018), pp. 306–318.
 - [88] Monika Syga. “On global properties of lower semicontinuous quadratically minorized functions”. In: *arXiv preprint arXiv:1912.04644* (2019).
 - [89] Tuomo Valkonen. “First-Order Primal–Dual Methods for Nonsmooth Non-convex Optimization”. In: *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*. Springer, 2022, pp. 1–42.
 - [90] Jean-Philippe Vial. “Strong and weak convexity of sets and functions”. In: *Mathematics of Operations Research* 8.2 (1983), pp. 231–259.
 - [91] Peng Wang et al. “Large-Scale Binary Quadratic Optimization Using Semidefinite Relaxation and Applications”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.3 (2017), pp. 470–485. DOI: 10.1109/TPAMI.2016.2541146.
 - [92] Fei Wen et al. “A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning”. In: *IEEE Access* 6 (2018), pp. 69883–69906.

-
- [93] Zikai Xiong and Robert Michael Freund. “Computational Guarantees for Restarted PDHG for LP based on" Limiting Error Ratios" and LP Sharpness”. In: *arXiv preprint arXiv:2312.14774* (2023).
 - [94] Xu Yang and Zhi-Yong Liu. “A Doubly Graduated Method for Inference in Markov Random Field”. In: *SIAM Journal on Imaging Sciences* 14.3 (2021), pp. 1354–1373.
 - [95] Constantin Zalinescu. *Convex analysis in general vector spaces*. World scientific, 2002.
 - [96] Jinshan Zeng and Wotao Yin. “On nonconvex decentralized gradient descent”. In: *IEEE Transactions on signal processing* 66.11 (2018), pp. 2834–2848.
 - [97] Yuzixuan Zhu, Deyi Liu, and Quoc Tran-Dinh. “New primal-dual algorithms for a class of nonsmooth and nonlinear convex-concave minimax problems”. In: *SIAM Journal on Optimization* 32.4 (2022), pp. 2580–2611.