ITERATIVE METHODS FOR SOLVING OPTIMIZATION PROBLEMS

Shoham Sabach

ITERATIVE METHODS FOR SOLVING OPTIMIZATION PROBLEMS

Research Thesis

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Shoham Sabach

Submitted to the Senate of the
Technion - Israel Institute of Technology

Iyar 5772 Haifa May 2012

The research thesis was written under the supervision of

Prof. Simeon Reich in the Department of Mathematics

Publications:

(i) Reich, S. and Sabach, S.: Three strong convergence theorems for iterative methods for solving equilibrium problems in reflexive Banach spaces, *Optimization Theory and Related Topics*, Contemporary Mathematics, vol. 568, Amer. Math. Soc., Providence, RI, 2012, 225–240.

(ii) Martín-Márquez, V., Reich, S. and Sabach, S.: Iterative methods for approximating fixed points of Bregman nonexpansive operators, *Discrete and Continuous Dynamical Systems*, accepted for publication. **Impact Factor: 0.986**.

(iii) Sabach, S.: Products of finitely many resolvents of maximal monotone mappings in reflexive Banach spaces, *SIAM Journal on Optimization* **21** (2011), 1289–1308. **Impact Factor: 2.091**.

(iv) Kassay, G., Reich, S. and Sabach, S.: Iterative methods for solving systems of variational inequalities in reflexive Banach spaces, *SIAM J. Optim.* **21** (2011), 1319–1344. **Impact Factor: 2.091**.

(v) Censor, Y., Gibali, A., Reich S. and Sabach, S.: The common variational inequality point problem, *Set-Valued and Variational Analysis* **20** (2012), 229–247. **Impact Factor: 0.333**.

(vi) Borwein, J. M., Reich, S. and Sabach, S.: Characterization of Bregman firmly nonexpansive operators using a new type of monotonicity, *J. Nonlinear Convex Anal.* **12** (2011), 161–183. **Impact Factor: 0.738**.

(vii) Reich, S. and Sabach, S.: A projection method for Bregman firmly nonexpansive operators in reflexive Banach spaces, *J. Fixed Point Theory Appl.* **9** (2011), 101–116. **Impact Factor: 0.545**.

(viii) Reich, S. and Sabach, S.: Two strong convergence theorems for Bregman strongly nonexpansive operators in reflexive Banach spaces, *Nonlinear Analysis* **73** (2010), 122–135. **Impact Factor: 1.279**.

(ix) Butnariu, D., Reich, S. and Sabach, S.: A strong convergence theorem for resolvents of monotone operators, *J. Convex Anal.* **17** (2010), 991–1006. **Impact Factor: 0.908**.

(x) Reich, S. and Sabach, S.: Existence and approximation of fixed points of Bregman firmly nonexpansive mappings in reflexive Banach spaces, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, 2010, 299–314.

(xi) Reich, S. and Sabach, S.: Two strong convergence theorems for a proximal method in reflexive Banach spaces, *Numer. Funct. Anal. Optim.* **31** (2010), 22–44. **Impact Factor: 0.687**.

(xii) Reich, S. and Sabach, S.: A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces, *J. Nonlinear Convex Anal.* **10** (2009), 471–485. **Impact Factor: 0.738**.

THE GENEROUS FINANCIAL SUPPORT OF THE

TECHNION IS GRATEFULLY ACKNOWLEDGED

Acknowledgments

First and foremost I want to express my sincere gratitude to my advisor Professor Simeon Reich for the continuous support of my Ph.D study and research, for his patience, motivation and immense knowledge. His guidance helped me in all the time of research and writing this thesis. The joy and enthusiasm he has for his research were contagious and motivational for me, even during tough times in my Ph.D. pursuit. Because of his vast knowledge and understanding of mathematics, my view of Optimization Theory broadened immensely. It has been a great privilege to learn from him.

I finished my master studies at the University of Haifa in 2008 under the supervision of Professor Dan Butnariu who passed away two months after I graduated. Simeon then gave me the option to study for my doctoral degree under his supervision at the Technion. I am grateful to the Department of Mathematics for their help and support.

Thanks to Simeon's advice, I had the opportunity to study and work with several amazing mathematicians during my doctoral studies. I am sure that this experience improved my thesis novelty and my research abilities. Among these collaborators, I would like to express my deepest appreciation to Professor Amir Beck form the Technion for his enthusiasm, encouragements and able guidance. It was an amazing stimulating environment to study and conduct a fruitful research.

My sincere thanks also go to Professors Heinz H. Bauschke, Jonathan M. Borwein, Regina S. Burachik, Yair Censor, Andreas H. Hamel, Gabor Kassay, Elena Resmerita and Kaya C. Yalçın for discussing problems and making observations that improved much the novelty of my thesis.

Many many thanks to my friends Dr. Aviv Gibali and Dr. Victoria Martín-Márquez who have always been supportive, encouraging and amazing partners for research and fun.

Most of all, I wish to thank my parents Eyal and Yael, my wife Orli and my son Jonathan for sharing these unforgettable years with me.

Table of Contents

Table of Contents - Continued

List of Tables

List of Figures

List of Algorithms

Abstract

This dissertation concerns iterative methods for solving diverse optimization problems in infinite-dimensional and Euclidean spaces. It contains six chapters. My contributions to the fields of Optimization Theory, Nonlinear Analysis and Numerical Methods are interpreted on the broad spectrum between practical methods for solving real-world problems to iterative algorithms for approximating solutions of optimization problems in infinite-dimensional spaces.

The first five chapters of this dissertation focus on my research in the infinite-dimensional case. The iterative methods proposed in the third chapter are based on several results in Fixed Point Theory and Convex Analysis which were obtained in the first two chapters. We first studied new properties of Bregman distances with respect to two classes of convex functions: Legendre functions and totally convex functions. These developments lead to many results regarding fixed points of nonexpansive operators which are defined with respect to Bregman distances instead of the norm. We deal with a wide variety of optimization problems such as fixed point problems, equilibrium problems, minimization problems, variational inequalities and the convex feasibility problem. The fourth chapter is devoted to a long and detailed study of the problem of finding zeroes of monotone mappings. In this area we wish to develop iterative methods which generate approximation sequences which converge strongly to a zero.

My research in finite-dimensional spaces appears in the last but not the least chapter and deals with developing an algorithm for solving optimization problems which arise from real-world applications. We developed a new numerical method for finding minimum norm solutions of convex optimization problems. This algorithm is the first attempt to solve such problems directly and not by solving a "sequence" of subproblems.

List of Symbols

X	A real reflexive Banach space
X^*	The dual space of X
X^{**}	The bidual space of X
$\|\cdot\|$	The norm in X
$\|\cdot\|_*$	The norm in X^*
$\langle \xi, x \rangle$	The value of the functional $\xi \in X^*$ at $x \in X$
$\operatorname{int} K$	The interior of the set K
$\operatorname{cl} K$	The closure of the set K
$\operatorname{bdr} K$	The boundary of the set K
f	A function from X to $(-\infty, +\infty]$
f^*	The Fenchel conjugate of f
f^{**}	The biconjugate of f
$\operatorname{dom} f$	The set $\{x \in X : f(x) < +\infty\}$
$\operatorname{epi} f$	The set $\{(x, t) \in X \times \mathbb{R} : f(x) \leqslant t\}$
$\operatorname{graph} f$	The set $\{(x, f(x)) \in X \times \mathbb{R} : x \in \operatorname{dom} f\}$
$\varphi_f(y, x; t)$	The function $(f(x + ty) - f(x))/t$
$f^\circ(x, y)$	The limit of $\varphi_f(y, x; t)$ at $t \searrow 0$
$\mathcal{BS}$	The Boltzmann-Shannon entropy in $\mathbb{R}_{++}$
$\mathcal{BS}_n$	The Boltzmann-Shannon entropy in $\mathbb{R}^n_{++}$
$\mathcal{FD}$	The Fermi-Dirac entropy in $(0, 1)$
f_p	The p-norm function, $\|\cdot\|^p$, $0 < p < 1$
D_f	The Bregman distance with respect to the function f
$\operatorname{lev}^f_\alpha$	The sub-level set of f on level α
ι_K	The indicator function of the set K
$\mathbb{I}_S$	The set-valued indicator operator of a subset S of X
N_K	The normal cone of the set K
$\operatorname{Fix}(T)$	The fixed point set of the operator T
$\widehat{\operatorname{Fix}}(T)$	The asymptotic fixed point set of the operator T
$\operatorname{dom} A$	The set $\{x \in X : Ax \neq \varnothing\}$
$\operatorname{ran} A$	The set $\{\xi \in Ax : x \in \operatorname{dom} A\}$
$\operatorname{graph} A$	The set $\{(x, \xi) \in X \times X^* : x \in \operatorname{dom} A, \xi \in Ax\}$
∇f	The gradient of the function f
∂f	The set-valued subdifferential mapping of the function f
J_ϕ	The duality mapping with respect to a gauge function ϕ
J_X	The normalized duality mapping of the space X
J_p	The duality mapping of the p-norm function f_p

Res_A^f	The f-resolvent of the mapping A
GRes_A^f	The generalized resolvent of the mapping A with respect to the function f
Res_g^f	The resolvent of the bifunction g with respect to the function f
A^f	The anti-resolvent of the mapping A with respect to the function f
R_A	The classical resolvent of the mapping A
S_T	The single-valued mapping defined by $\nabla f - \nabla f \circ T$
A_T	The set-valued mapping defined by $\nabla f \circ T^{-1} - \nabla f$
proj_K^f	The Bregman projection onto K with respect to the function f
P_K	The metric projection onto K
W^f	$W^f(\xi, x) = f(x) - \langle \xi, x \rangle + f^*(\xi)$
T_M	The proj-grad mapping of a function f with constant $M > 0$
G_M	The gradient mapping of a function f with constant $M > 0$
T_B	The block operator
W	The Lambert W function
ν_f	The modulus of total convexity of the function f
δ_f	The modulus of uniform convexity of the function f
δ_X	The modulus of uniform convexity of the space X
μ_X	The modulus of local uniform convexity of the space X
$A^{-1}(0^*)$	The zero set of the mapping A
EP	The solution set of the corresponding equilibrium problem
VI	The solution set of the corresponding variational inequality problem

Introduction

Many problems arising in different areas of Mathematics, such as Convex Analysis, Variational Analysis, Optimization, Monotone Mapping Theory and Differential Equations, can be modeled by the problem

$$x = Tx,$$

where T is a nonlinear operator defined on a metric space. Solutions to this equation are called fixed points of T. If T is a *strict contraction* defined on a complete metric space X, Banach's contraction principle establishes that T has a unique fixed point and for any $x \in X$, the sequence of *Picard iterates* $\{T^n x\}_{n\in\mathbb{N}}$ strongly converges to the unique fixed point of T. However, if the operators T is a *nonexpansive* operator, that is,

$$d(Tx, Ty) \leqslant d(x, y), \ \forall x, y \in X,$$

then we must assume additional conditions on T and/or on the underlying space to ensure the existence of fixed points. Since the sixties, the study of the class of nonexpansive operators is one of the major and most active research areas of nonlinear analysis. This is due to the connection with the geometry of Banach spaces along with the relevance of these operators to the theory of monotone and accretive mappings. The concepts of monotonicity and accretivity have turned out to be very powerful in diverse fields such as Operator Theory, Numerical Analysis, Differentiability of Convex Functions and Partial Differential Equations; see, for instance, [47, 56, 77, 113]. In particular, one of the reasons is that the class of monotone mappings is broad enough to cover subdifferentials of proper, convex and lower semicontinuous functions, which are mappings of increasing importance in Optimization Theory.

The relationship in Hilbert spaces between the Theory of Monotone Mappings and the Theory of Nonexpansive Operators is basically determined by two facts: (1) if T is a nonexpansive operator, then the complementary operator $I - T$ is monotone and (2) the *classical resolvent*, $(I + A)^{-1}$, of a monotone mapping A is nonexpansive. Moreover, in both cases the fixed point set of the nonexpansive operator coincides with the zero set of the monotone mapping. See [11] for a detailed study of these two concepts.

In this dissertation you will find results concerning these connections in general reflexive Banach spaces. The first chapter is a collection of several notions, definitions and basic results needed in the whole work. This chapter mainly concentrates on functions, operators and mappings. The rest of the work includes five more chapters. The second chapter is devoted to results on Bregman distances and on "nonexpansive" operators with respect to Bregman distances. The next chapter uses the tools and the new developments to propose and study several iterative methods for approximating fixed points of such operators which are based on the Picard iteration. The fourth chapter deals with the very related area of the Theory of Monotone Mappings. We propose and study iterative methods for approximating zeroes of monotone mappings. Taking into account the iterative methods proposed in the third and the fourth chapters, we modified them for solving diverse optimization problems, such as variational inequalities, equilibrium problems and convex feasibility problems. All these four chapters are formulated in the general context of infinite-dimensional reflexive Banach spaces. The last, but not the least, chapter is the most practical aspect of this work. We took one of the algorithms proposed in the previous chapters and modified it to solve the problem of finding minimal norm solutions of convex optimization problems.

Let me briefly describe these chapters.

Chapter 2 - Fixed Point Properties of Bregman Nonexpansive Operators

In 2003, Bauschke, Borwein and Combettes [8] first introduced the class of Bregman firmly nonexpansive (BFNE) operators which is a generalization of the classical firmly nonexpansive operators (FNE). A few years before Reich studied the class of Bregman strongly nonexpansive (BSNE) operators in the case of common fixed points. Many other researchers studied several other classes of operators of Bregman nonexpansive type. Very recently, in several projects we studied in depth these operators from the aspects of Fixed Point Theory. The results obtained in this chapter bring out new tools and techniques which are used by us and many other researchers, to develop iterative methods for approximating fixed points of operators of Bregman nonexpansive type.

In [91] we studied the existence and approximation of fixed points of Bregman firmly nonexpansive operators in reflexive Banach spaces. In this paper we first obtained necessary and sufficient conditions for BFNE operators to have a (common) fixed point. We also found under which conditions the asymptotic fixed point set of BFNE operators coincides with the fixed point set (demi-closedness principle). The concept of asymptotic fixed points was first introduced in [88] and plays a key role in analyzing iterative methods. In practice, it is much easier to prove convergence to an asymptotic fixed point than to a fixed point and therefore we were motivated to determine when and under what conditions these two sets coincide.

Motivated by this work and during my visit to CARMA, we collaborated with Jonathan M. Borwein and found a characterization of BFNE operators in general reflexive Banach spaces. This characterization allows one to construct many Bregman firmly nonexpansive operators explicitly. We also provided several examples of such operators with respect to two important Bregman functions: the Boltzmann-Shannon entropy and the Fermi-Dirac entropy in Euclidean spaces. We have studied these entropies in detail because of their importance in applications. These two entropies form a large part of the basis for classical Information Theory.

Then after a visit to the University of Seville we collaborated with Victoria Martín-Márquez and continued our research on certain Bregman nonexpansive classes of operators. We mainly put forward a clear picture of the existence and approximation of their (asymptotic) fixed points. In particular, the asymptotic behavior of Picard and Mann type iterations are discussed for quasi-Bregman nonexpansive (QBNE) operators. We also presented parallel algorithms for approximating common fixed points of a finite family of Bregman strongly nonexpansive operators by means of a block operator which preserves the Bregman strong nonexpansivity.

Chapter 3 - Iterative Methods for Approximating Fixed Points

Fixed point iterative methods started with the celebrated Picard method and were further developed for computing and constructing fixed points of various types of nonexpansive operators and in various types of topologies and spaces (for instance, weak/strong, Hilbert/Banach and metric/normed). We have been mostly interested in such methods in general reflexive Banach spaces for operators of Bregman nonexpansive type. In addition, several of our algorithms are new even in the framework of Hilbert spaces and of Euclidean spaces. Our main motivation was, and still is, to develop iterative methods which generate a strongly convergent sequence. The importance of strong convergence over weak convergence plays a key role in applications and as a result, many researchers are motivated to develop strongly convergent iterative methods, even under stronger conditions.

In this chapter we propose several variants of the classical Picard method for operators of Bregman nonexpansive type. In 2010 we studied the convergence of two iterative algorithms for finding fixed points of Bregman strongly nonexpansive operators in reflexive Banach spaces. Both algorithms take into account possible computational errors and we established two strong convergence results. In these methods we calculate the value of the operator at the current point and, in contrast with the Picard iterative method, the next iteration is the Bregman projection onto the intersection of two half-spaces which contain the solution set. These two algorithms are more complicated than the Picard method because of the additional projection step, but on the other hand, they generate a sequence which converges strongly to a certain fixed point (the operator may have many fixed points). Another advantage of these algorithms is the nature of the limit point, which is not only a fixed point, but the one which is closest to the initial starting point of the algorithm with respect to the Bregman distance.

These algorithms are proposed for finding common fixed points of finitely many operators. In this case one seeks to find the Bregman projection onto the intersection of $N+1$ half-spaces (N is the number of operators). The motivation for studying such common fixed-point problems with $N > 1$

stems from the simple observation that this is a generalization of the well-known Convex Feasibility Problem (CFP).

A few months later we proposed in [93] a projection method for solving the common fixed point problem of N Bregman firmly nonexpansive operators. In this paper there is an algorithm which is based on the shrinking projection method. The advantage of this method is that the number of subsets onto which we project in every step is N (not $N+1$ as in the previous algorithms). But these subsets are not necessarily half-spaces as in the previous case.

Chapter 4 - Iterative Methods for Approximating Zeroes

In this chapter we are concerned with the problem of finding zeroes of mappings $A : X \to 2^{X^*}$, that is, finding $x \in \operatorname{dom} A$ such that

$$0^* \in Ax. \tag{0.0.1}$$

Many problems have reformulations which require us to find zeroes, for instance, differential equations, evolution equations, complementarity problems, mini-max problems, variational inequalities and optimization problems. It is well known that minimizing a convex function f can be reduced to finding zeroes of the corresponding subdifferential mapping $A = \partial f$.

One of the most important techniques for solving the inclusion (0.0.1) goes back to the work of Browder [29] in the sixties. One of the basic ideas in the case of a Hilbert space $\mathcal{H}$ is reducing (0.0.1) to a fixed point problem for the operator $R_A : \mathcal{H} \to 2^{\mathcal{H}}$ defined by

$$R_A = (I + A)^{-1},$$

which we call in what follows the *classical resolvent* of A. When $\mathcal{H}$ is a Hilbert space and A satisfies some monotonicity conditions, the classical resolvent of A is with full domain and *nonexpansive*, that is,

$$\|R_A x - R_A y\| \leqslant \|x - y\| \;\; \forall x, y \in \mathcal{H},$$

and even *firmly nonexpansive*, that is,

$$\|R_A x - R_A y\|^2 \leqslant \langle R_A x - R_A y, x - y \rangle \;\; \forall x, y \in \mathcal{H}.$$

These properties of the resolvent ensure that its Picard iterates $x_{n+1} = R_A x_n$ converge weakly, and sometimes even strongly, to a fixed point of R_A, which is necessarily a zero of A. Rockafellar introduced this iteration method and called it the Proximal Point Algorithm (see [100, 101]).

Methods for finding zeroes of monotone mappings in Hilbert space are based on the good properties of the resolvent R_A such as nonexpansivity, but when we try to extend these methods to Banach spaces we encounter several difficulties (see, for example, [41]).

One way to overcome these difficulties is to use, instead of the classical resolvent, a new type of resolvent: the *f-resolvent* first introduced by Teboulle [108] in 1992 for the subdifferential mapping case and one year later by Eckstein [51] for a general monotone mapping (see also [46, 88, 8]). If $f : X \to (-\infty, +\infty]$ is a Legendre function, then the operator $\mathrm{Res}_A^f : X \to 2^X$ given by

$$\mathrm{Res}_A^f = (\nabla f + A)^{-1} \circ \nabla f \tag{0.0.2}$$

is well defined when A is maximal monotone and $\operatorname{int}\operatorname{dom} f \bigcap \operatorname{dom} A \neq \varnothing$. Moreover, similarly to the classical resolvent, a fixed point of Res_A^f is a solution of (0.0.1). This leads to the question whether, and under what conditions on A and f, the iterates of Res_A^f approximate a fixed point of Res_A^f.

In order to modify the proximal point algorithm for the new resolvent and prove the convergence of the iterates of Res_A^f, we need the nonexpansivity properties of this resolvent (the theory of which we develop in Chapter 2) as in the case of the classical resolvent.

We propose several modifications of the classical proximal point algorithm. We first modify this algorithm to obtain algorithms which generate sequences which converge strongly to zeroes. In this dissertation we also solve the problems of finding common zeroes of finitely many maximal monotone mappings. In addition, we allow in these algorithms several kinds of computational errors.

In a recent single-authored paper [103], which has already appeared in the prestigious SIAM Journal on Optimization, we focused on the common zeroes problem. The algorithms proposed in this paper are based on the concept of finite products of resolvents. This concept has led me to an algorithm which solves several problems, but in the main step of the algorithm one should project onto the intersection of just two half-spaces. Therefore the number of operators does not influence the number of the half-spaces that should be constructed at each step.

Chapter 5 - Applications - Equilibrium, Variational and Convex Feasibility Problems

This chapter contains three main applications of the iterative methods proposed in the previous chapters. We start by studying equilibrium problems in the context of reflexive Banach spaces. The second application is devoted to iterative methods for solving variational inequalities. Connections between these two problems to fixed point problems are given. At the end we apply our algorithms to solving the well-known Convex Feasibility Problem in the framework of reflexive Banach spaces.

Chapter 6 - Minimal Norm Solutions of Convex Optimization Problems

We were interested in the following problem. Suppose one has two objective functions f and ω. The problem of minimizing the function f over a constrains set S may have multiple solutions. Among these solutions of the "core" problem we wish to find a solution which minimizes the "outer" objective function ω.

A particular case of this problem was studied from the seventies by many leading researchers. This particular case is when the "other" function is taken as the norm, that is, $\omega = \|\cdot\|^2$. In this case we wish to find a minimal norm solution of the "core" problem. Since we consider a more general function ω than the usual norm, we call these solutions minimal norm-like solutions of optimization problems.

A well-known approach to finding the minimal norm solutions of convex optimization problem is via the celebrated Tikhonov regularization. This approach involves solving the original problem by solving many emerging subproblems. These subproblems are simpler than the original problem, but still need a different approach in order to be solved. In addition, this approach suffers a few drawbacks which makes it, from a practical point of view, problematic for implementation.

Recently, I jointly presented with Amir Beck [17] a direct first order method for finding minimal norm-like solutions of convex optimization problems. We proposed the minimal norm gradient method which is aimed at solving the problem directly and not "stage by stage" or via solving a sequence of related problems.

At each iteration of the minimal norm gradient method, the required computations are (i) a gradient evaluation of the objective function f, (ii) an orthogonal projection onto the feasible set of the "core" problem and (iii) a solution of a problem consisting of minimizing the objective function ω subject to the intersection of two half-spaces. The convergence of the sequence generated by the method is established along with an $O\left(1/\sqrt{k}\right)$ convergence of the sequence of function values (k being the iteration index). We support our analysis with a numerical example of a portfolio optimization problem.

Chapter 1

Preliminaries

This chapter contains three sections that deal with several basic notions concerning functions, operators and mappings which will be needed in our later discussion (see [97]). Let X be a reflexive infinite-dimensional Banach space. The space X is equipped with the norm $\|\cdot\|$ and X^* represents the (topological) dual of X whose norm is denoted by $\|\cdot\|_*$. We denote the value of the functional $\xi \in X^*$ at $x \in X$ by $\langle \xi, x \rangle$.

From now on we will use the following notations for functions, bifunctions, operators and single/set-valued mappings.

- A *function* f which maps X into $(-\infty, +\infty]$ will be denoted by $f : X \to (-\infty, +\infty]$.
- A *bifunction* g which maps $X \times X$ into $\mathbb{R}$ will be denoted by $g : X \to (-\infty, +\infty]$.
- An *operator* T which maps X into X will be denoted by $T : X \to X$.
- A *single-valued mapping* A which maps $\operatorname{dom} A \subset X$ into X^* will be denoted by $A : X \to X^*$.
- A *set-valued mapping* A which maps X into 2^{X^*} will be denoted by $A : X \to 2^{X^*}$.

We will also use the following notations.

- The set of all nonnegative integers is denoted by $\mathbb{N}$.
- The set of all real numbers is denoted by $\mathbb{R}$.
- The set of all nonnegative real numbers is denoted by $\mathbb{R}_+$.
- The set of all positive real numbers is denoted by $\mathbb{R}_{++}$.
- The *closure* of a set K will be denoted by $\operatorname{cl} K$.
- The *interior* of a set K will be denoted by $\operatorname{int} K$.
- The *boundary* of a set K will be denoted by $\operatorname{bdr} K$.
- The *unit sphere* of X is denoted by $S_X = \{x \in X : \|x\| = 1\}$.

- Given a sequence $\{x_n\}_{n\in\mathbb{N}}$ and $x \in X$, the strong convergence (weak convergence) of $\{x_n\}_{n\in\mathbb{N}}$ to x is denoted by $x_n \to x$ ($x_n \rightharpoonup x$) as $n \to \infty$ or $\lim_{n\to\infty} x_n = x$ ($w - \lim_{n\to\infty} x_n = x$).
- We will denote by $\mathcal{H}$ a Hilbert space.

1.1 Functions

This section is devoted to the notions related to functions that are needed in our results. We present four subsections with basic notions and results about continuity of functions, subdifferentiability and differentiability properties and conjugate functions. At the end of this section we present another subsection on the geometry of Banach spaces.

1.1.1 Lower Semicontinuity, Continuity and Lipschitz Continuity

We will start with the basic notions needed for a discussion of functions.

Definition 1.1.1 (Basic notions for functions)**.** *Let $f : X \to (-\infty, +\infty]$ be a function.*

(i) *The domain of f is the following set:*

$$\operatorname{dom} f := \{x \in X : f(x) < +\infty\}.$$

(ii) *The epigraph of f is the following set:*

$$\operatorname{epi} f := \{(x, t) \in X \times \mathbb{R} : f(x) \leqslant t\}.$$

(iii) *The function f is called proper if* $\operatorname{dom} f \neq \varnothing$.

Definition 1.1.2 (Convexity)**.** *The function $f : X \to (-\infty, +\infty]$ is called convex if it satisfies*

$$f(\lambda x + (1-\lambda) y) \leqslant \lambda f(x) + (1-\lambda) f(y) \tag{1.1.1}$$

for any two points $x, y \in \operatorname{dom} f$ and for any real number $\lambda \in [0, 1]$.

Remark 1.1.3 (Convexity properties)**.** (i) *If the inequality in* (1.1.1) *is strict, then the function f is called strictly convex.*

(ii) *It easy to check that f is convex if and only if* $\operatorname{epi} f$ *is a convex set in $X \times \mathbb{R}$.*

(iii) *If f is a convex function, then* $\operatorname{dom} f$ *is a convex set.* ◇

Definition 1.1.4 (Lower and upper semicontinuity)**.** *Let $f : X \to (-\infty, +\infty]$ be a function.*

(i) *The function* f *is called lower semicontinuous if, for each real number* r*, the set* $\{x \in X : f(x) \leqslant r\}$ *is closed.*

(ii) *The function* f *is called weakly lower semicontinuous if, for any real number* r*, the set* $\{x \in X : f(x) \leqslant r\}$ *is weakly closed in* X*.*

(iii) *The function* $f : X \to (-\infty, +\infty]$ *is called upper semicontinuous at a point* $x \in \operatorname{dom} f$ *if, for each open set* V *in* $(-\infty, +\infty]$ *containing* $f(x)$*, there is a neighborhood* U *of* x *such that* $f(U) \subset V$*.*

Remark 1.1.5 (Sufficient condition for lower semicontinuity)**.** *If* $f : X \to (-\infty, +\infty]$ *is convex and continuous on* $\operatorname{dom} f$ *which is a closed set, then* f *is lower semicontinuous.* ◇

The following two propositions show connections among various continuity properties of convex functions (see, for instance, [35, Propostion 1.1.5, page 6] and [82, Proposition 2.3, page 22], respectively).

Proposition 1.1.6 (Continuity properties of convex functions)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a convex function with* $\operatorname{int}\operatorname{dom} f \neq \varnothing$*. The following statements are equivalent.*

(i) *The function* f *is locally bounded from above on* $\operatorname{int}\operatorname{dom} f$*.*

(ii) *The function* f *is locally bounded on* $\operatorname{int}\operatorname{dom} f$*.*

(iii) *The function* f *is locally Lipschitzian on* $\operatorname{int}\operatorname{dom} f$*.*

(iv) *The function* f *is continuous on* $\operatorname{int}\operatorname{dom} f$*.*

Moreover, if f *is lower semicontinuous, then all these statements hold.*

Corollary 1.1.7 (Continuity property)**.** *Every proper, convex and lower semicontinuous function* $f : X \to (-\infty, +\infty]$ *is continuous on* $\operatorname{int}\operatorname{dom} f$*.*

Proposition 1.1.8 (Lower semicontinuity properties)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a convex function. The following statements are equivalent.*

(i) *The function* f *is lower semicontinuous on* X*.*

(ii) *For each* $\bar{x} \in X$ *and for each net* $\{x_\beta\}_{\beta \in I} \subset X$ *converging to* $\bar{x} \in X$*, we have*

$$f(\bar{x}) \leqslant \liminf_{\beta \in I} f(x_\beta).$$

(iii) *The set* $\operatorname{epi} f$ *is closed in* $X \times \mathbb{R}$*.*

(iv) *The function* f *is weakly lower semicontinuous on* X*.*

1.1.2 Subdifferentiability

Definition 1.1.9 (Directional derivative)**.** *Let $f : X \to (-\infty, +\infty]$ be a function. Given $x \in \operatorname{int} \operatorname{dom} f$ and $y \in X$, define $\varphi_f(y, x; \cdot) : \mathbb{R} \backslash \{0\} \to (-\infty, +\infty]$ by*

$$\varphi_f(y, x; t) := \frac{1}{t}(f(x + ty) - f(x)). \tag{1.1.2}$$

The directional derivative of f at x in the direction y is given by

$$f^{\circ}(x, y) := \lim_{t \searrow 0} \varphi_f(y, x; t). \tag{1.1.3}$$

The following result presents several properties of the directional derivative of convex functions (see, for example, [35, Propostion 1.1.2, page 2]).

Proposition 1.1.10 (Properties of directional derivatives)**.** *Let $f : X \to (-\infty, +\infty]$ be a proper and convex function. If $x \in \operatorname{int} \operatorname{dom} f$, then the following statements are true.*

(i) *The function $\varphi_f(y, x; \cdot)$ is increasing on each of the intervals $(0, +\infty)$ and $(-\infty, 0)$ for any $y \in X$.*

(ii) *If f is also strictly convex, then $\varphi_f(y, x; \cdot)$ is strictly increasing on each of the intervals $(0, +\infty)$ and $(-\infty, 0)$ for any $y \in X$.*

(iii) *The directional derivative $f^{\circ}(x, y)$ exists for any $y \in X$, and we have*

$$f^{\circ}(x, y) \leqslant f(x + y) - f(x). \tag{1.1.4}$$

If, in addition, f is strictly convex, then the inequality in (1.1.4) is strict.

(iv) *We have*

$$-f^{\circ}(x, -y) \leqslant f^{\circ}(x, y) \quad \forall y \in X.$$

(v) *The limit $f^{\circ}(x, y)$ is finite for any $y \in X$ if and only if $x \in \operatorname{int} \operatorname{dom} f$. In this case, the function $y \to f^{\circ}(x, y)$ is sublinear.*

(vi) *If, in addition, f is lower semicontinuous, then the function $f^{\circ}(x, \cdot)$ is Lipschitzian on X.*

Here is a consequence of Proposition 1.1.10 which will be used below (*cf.* [35, Propostion 1.1.4, page 4]).

Proposition 1.1.11 (Characterization of strict convexity)**.** *Let $f : X \to (-\infty, +\infty]$ be a convex function such that $\operatorname{int} \operatorname{dom} f \neq \varnothing$. The following statements are equivalent.*

(i) *The function f is strictly convex on $\operatorname{int} \operatorname{dom} f$.*

(ii) *For any* $x, y \in \operatorname{int}\operatorname{dom} f$ *such that* $x \neq y$, *we have*

$$f^{\circ}(x, y-x) + f^{\circ}(y, x-y) < 0.$$

Definition 1.1.12 (Subgradient and subdifferential). *Let* $f : X \to (-\infty, +\infty]$ *be a function.*

(i) *A vector* $\xi \in X^*$ *is called a subgradient of* f *at a point* $x \in \operatorname{dom} f$ *if*

$$f(y) - f(x) \geqslant \langle \xi, y-x \rangle \quad \forall y \in \operatorname{dom} f. \tag{1.1.5}$$

(ii) *If there exists a subgradient* ξ *of* f *at* x, *we say that* f *is subdifferentiable at* x.

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

The next result shows a strong connection between the notions of subdifferentiability and convexity.

Proposition 1.1.13 (Sufficient condition for convexity). *Let* $f : X \to (-\infty, +\infty]$ *be a function and let* K *be a convex subset of* X. *If* f *is subdifferentiable at each point of* K, *then* f *is convex on* K.

The following result brings out several properties of the subdifferential mapping (*cf.* [35, Propostion 1.1.7, page 8]).

Proposition 1.1.14 (Properties of the subdifferential mapping). *Let* $f : X \to (-\infty, +\infty]$ *be a convex function such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$. *The following statements are true.*

(i) *For any* $x \in \operatorname{int}\operatorname{dom} f$, *the set* $\partial f(x)$ *is a convex and weak** *closed subset of* X^*.

(ii) *If, in addition,* f *is continuous on* $\operatorname{int}\operatorname{dom} f$, *then, for each* $x \in \operatorname{int}\operatorname{dom} f$, *the set* $\partial f(x)$ *is nonempty and weak** *compact. In this case, for each* $y \in X$, *we have*

$$f^{\circ}(x, y) = \max\left\{\langle \xi, y \rangle : \xi \in \partial f(x)\right\}.$$

(iii) *If* $x \in X$, *then*

$$\partial f(x) = \left\{\xi \in X^* : -f^{\circ}(x, -y) \leqslant \langle \xi, y \rangle \leqslant f^{\circ}(x, y) \quad \forall y \in X\right\}.$$

(iv) *If, in addition,* f *is continuous on* $\operatorname{int}\operatorname{dom} f$, *then the set-valued mapping* $x \to \partial f(x)$ *is norm-to-weak** *upper semicontinuous on* $\operatorname{int}\operatorname{dom} f$.

Proposition 1.1.14(iv) shows that the set-valued subdifferential mapping ∂f associated to a convex function $f : X \to (-\infty, +\infty]$ is norm-to-weak* upper semicontinuous. A question which occurs in optimization is whether the set-valued mapping ∂f is bounded on bounded subsets. The answer to this question is given by the following technical result (*cf.* [35, Propostion 1.1.11, page 16]).

Proposition 1.1.15 (Characterization of boundedness on bounded subsets). *If $f : X \to \mathbb{R}$ is a continuous and convex function, then the set-valued mapping $\partial f : X \to 2^{X^*}$ is bounded on bounded subsets of* $\operatorname{dom} \partial f$ *if and only if the function f itself is bounded on bounded subsets X.*

1.1.3 Gâteaux and Fréchet Differentiability

Definition 1.1.16 (Gâteaux differentiability). *Let $f : X \to (-\infty, +\infty]$ be a proper and convex function. The function f is called Gâteaux differentiable at $x \in \operatorname{int} \operatorname{dom} f$ if the limit*

$$f'(x)(y) := \lim_{t \to 0} \varphi_f(y, x; t) = \lim_{t \to 0} \frac{1}{t} \left(f(x + ty) - f(x) \right) \tag{1.1.6}$$

exists.

Remark 1.1.17 (Characterization of Gâteaux differentiable functions). *Since in our case $f : X \to (-\infty, +\infty]$ is always assumed to be proper and convex, it follows from Proposition 1.1.10(v) that the function $y \to f^\circ(x, y)$ is sublinear. Hence, in our setting, f is Gâteaux differentiable at $x \in \operatorname{int} \operatorname{dom} f$ if and only if $-f^\circ(x, -y) = f^\circ(x, y)$ for any $y \in X$. If, in addition, f is lower semicontinuous, then $f'(x)(\cdot) = f^\circ(x, \cdot)$ is continuous and belongs to X^* (see Proposition 1.1.10(vi)).* $\diamond$

Definition 1.1.18 (Gradient). *Let $f : X \to (-\infty, +\infty]$ be a function. The gradient of f, ∇f, is the linear function $x \to f'(x)(\cdot)$ when it exists (see [83, Definition 1.3, page 3]).*

The next result establishes characteristic continuity properties of Gâteaux derivatives of convex and lower semicontinuous functions (*cf.* [35, Propostion 1.1.10(i), page 13]).

Proposition 1.1.19 (The subdifferential is a gradient). *Let $f : X \to (-\infty, +\infty]$ be a convex and lower semicontinuous function with $\operatorname{int} \operatorname{dom} f \neq \varnothing$. The function f is Gâteaux differentiable at a point $x \in \operatorname{int} \operatorname{dom} f$ if and only if $\partial f(x)$ consists of a single element. In this case, $\partial f(x) = \{\nabla f(x)\}$.*

Definition 1.1.20 (Fréchet differentiability). *Let $f : X \to (-\infty, +\infty]$ be a function. We say that f is:*

(i) *Fréchet differentiable if it is Gâteaux differentiable and the limit* (1.1.6) *is attained uniformly for every $y \in S_X$ (see [83]).*

(ii) *Uniformly Fréchet differentiable on bounded subsets of* X *if for any bounded subset* E *of* X *the limit* (1.1.6) *is attained uniformly for any* $x \in E$ *and every* $y \in S_X$.

The following result brings out a connection between differentiability properties of f and continuity properties of ∂f (*cf.* [83, Proposition 2.8, page 19]).

Proposition 1.1.21 (Continuity of the subdifferential mapping)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a convex and continuous function with* $\operatorname{int}\operatorname{dom} f \neq \varnothing$. *Then* f *is Gâteaux (respectively Fréchet) differentiable at* $x \in \operatorname{int}\operatorname{dom} f$ *if and only if there is a selection* φ *for the subdifferential mapping* ∂f *which is norm-to-weak** *(norm-to-norm) continuous at* x.

In this connection we will also have the following result (*cf.* [89, Proposition 2.1, page 474] and [3, Theorem 1.8, page 13]).

Proposition 1.1.22 (Properties of uniformly Fréchet differentiable functions)**.** *Let* $f : X \to \mathbb{R}$ *be a convex function which is both bounded and uniformly Fréchet differentiable on bounded subsets of* X. *The following statements hold.*

(i) *The function* f *is uniformly continuous on bounded subsets of* X.

(ii) *The gradient* ∇f *is uniformly continuous on bounded subsets of* X *from the strong topology of* X *to the strong topology of* X^*.

Proof. (i) See [3, Theorem 1.8, page 13].

(ii) If this result were not true, there would be two bounded sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$, and a positive number ε such that $\|x_n - y_n\| \to 0$ as $n \to \infty$ and

$$\langle \nabla f(x_n) - \nabla f(y_n), w_n \rangle \geqslant 2\varepsilon,$$

where $\{w_n\}_{n\in\mathbb{N}}$ is a sequence in S_X, that is, $\|w_n\| = 1$ for each $n \in \mathbb{N}$. Since f is uniformly Fréchet differentiable (see Definition 1.1.20(ii)), there is a positive number δ such that

$$f(y_n + tw_n) - f(y_n) - t\langle \nabla f(y_n), w_n \rangle \leqslant \varepsilon t$$

for all $0 < t < \delta$ and each $n \in \mathbb{N}$. From the subdifferentiability inequality (see (1.1.5)) we have

$$\langle \nabla f(x_n), (y_n + tw_n) - x_n \rangle \leqslant f(y_n + tw_n) - f(x_n) \quad \forall n \in \mathbb{N}.$$

In other words,

$$t\langle \nabla f(x_n), w_n \rangle \leqslant f(y_n + tw_n) - f(y_n) + \langle \nabla f(x_n), x_n - y_n \rangle + f(y_n) - f(x_n).$$

Hence

$$\begin{aligned}2\varepsilon t &\leqslant t\left\langle \nabla f\left(x_n\right)-\nabla f\left(y_n\right),w_n\right\rangle \leqslant \left[f\left(y_n+tw_n\right)-f\left(y_n\right)-t\left\langle \nabla f\left(y_n\right),w_n\right\rangle\right]\\&\quad+\left\langle \nabla f\left(x_n\right),x_n-y_n\right\rangle+f\left(y_n\right)-f\left(x_n\right)\\&\leqslant \varepsilon t+\left\langle \nabla f\left(x_n\right),x_n-y_n\right\rangle+f\left(y_n\right)-f\left(x_n\right).\end{aligned}$$

Since ∇f is bounded on bounded subsets of X (see Proposition 1.1.15), it follows that $\left\langle \nabla f\left(x_n\right),x_n-y_n\right\rangle$ converges to zero as $n\to\infty$, while $\left[f\left(y_n\right)-f\left(x_n\right)\right]\to 0$ as $n\to\infty$ since f is uniformly continuous on bounded subsets which follows from assertion (i). But this would yield that $2\varepsilon t\leqslant \varepsilon t$, a contradiction. □

Definition 1.1.23 (Positively homogeneous)**.** *Let $f:X\to(-\infty,+\infty]$ be a function. We say that f is positively homogeneous of degree $\alpha\in\mathbb{R}$ if $f(tx)=t^{\alpha}f(x)$ for all $x\in X$ and any $t>0$.*

The following result, which seems to be well-known, appears here for the sake of completeness (*cf.* [91, Proposition 15.2, page 302]).

Proposition 1.1.24 (Property of positively homogeneous functions)**.** *If $f:X\to\mathbb{R}$ is a positively homogeneous function of degree $\alpha\in\mathbb{R}$, then ∇f is a positively homogeneous mapping of degree $\alpha-1$.*

Proof. Let $y\in X$. From the definition of the gradient (see Definition 1.1.18) we have

$$\begin{aligned}\nabla f(tx)(y)&=\lim_{h\to 0}\frac{1}{h}\left(f(tx+hy)-f(tx)\right)=\lim_{h\to 0}\frac{1}{th}\left(f(tx+thy)-f(tx)\right)\\&=\frac{t^{\alpha}}{t}\lim_{h\to 0}\frac{1}{h}\left(f(x+hy)-f(x)\right)=t^{\alpha-1}\nabla f(x)(y)\end{aligned}$$

for any $x\in X$ and all $t>0$. □

1.1.4 The Fenchel Conjugate

Definition 1.1.25 ((Fenchel) conjugate function)**.** *Let $f:X\to(-\infty,+\infty]$ be a function. The function $f^{*}:X^{*}\to[-\infty,+\infty]$ defined by*

$$f^{*}(\xi):=\sup_{x\in X}\left\{\langle\xi,x\rangle-f(x)\right\}$$

is called the (Fenchel) conjugate function of f.

Remark 1.1.26 (Basic properties of conjugate functions)**.** *Let $f:X\to(-\infty,+\infty]$ be a proper, convex and lower semicontinuous function.*

(i) *The conjugate function f^* is convex and lower semicontinuous since it is the supremum of a family of convex and continuous functions and, therefore, convex and lower semicontinuous functions.*

(ii) *The conjugate function f^* is also proper since f is proper.* $\diamond$

Definition 1.1.27 (Biconjugate function)**.** *Let $f : X \to (-\infty, +\infty]$ be a function. The biconjugate function $f^{**} = \left(f^*\right)^*$ is defined by*

$$f^{**}(x) := \sup_{\xi \in X^*} \left\{ \langle \xi, x \rangle - f^*(\xi) \right\}.$$

Obviously, for any $x \in X$ and every $\xi \in X^*$, we have

$$f(x) \geqslant \langle \xi, x \rangle - f^*(\xi). \tag{1.1.7}$$

It is known as the *Young-Fenchel inequality.* This inequality implies that $f(x) \geqslant f^{**}(x)$, for all $x \in X$.

Example 1.1.28 (Conjugate of the norm-p function)**.** *The conjugate function of $f_p(x) = (1/p)\|x\|^p$, $p \in (0, +\infty)$, is $f_p^*(\xi) = (1/q)\|\xi\|_*^q$ where $1/p + 1/q = 1$.*

Several basic properties of Fenchel conjugate functions are summarized in the next result (see [102, Proposition 1.4.1, page 18]).

Proposition 1.1.29 (Properties of Fenchel conjugate functions)**.** *Let $f : X \to (-\infty, +\infty]$ and $g : X \to (-\infty, +\infty]$ be two proper functions. The following statements are true.*

(i) *If $f \leqslant g$ then $g^* \leqslant f^*$.*

(ii) *For any $\lambda \geqslant 0$, we have $(\lambda f)^*(\xi) = \lambda f^*(\xi/\lambda)$.*

(iii) *For any $\lambda \in \mathbb{R}$, we have $(f + \lambda)^* = f^* - \lambda$.*

(iv) *For any $y \in X$, we have $(f(\cdot - y))^*(\xi) = f^*(\xi) + \langle \xi, y \rangle$.*

(v) *For any $x \in X$, we have*

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

(vi) *If $\partial f(x) \neq \varnothing$ then $f(x)$ and $f^*(\xi)$ are finite.*

The next result brings out several connections among Fenchel conjugates and subdifferentiability (*cf.* [22, Propostion 2.118, page 82]).

Proposition 1.1.30 (Fenchel conjugany and subdifferentiability)**.** *Assume that $f : X \to (-\infty, +\infty]$ is a function. The following statements are true.*

(i) *If for some $x \in X$ the value $f^*(x)$ is finite, then,*

$$\partial f^{**}(x) = \operatorname{argmax}_{\xi \in X^*} \left\{ \langle \xi, x \rangle - f^*(\xi) \right\}.$$

(ii) *If f is subdifferentiable at $x \in X$, then $f^{**}(x) = f(x)$.*

(iii) *If $f^{**}(x) = f(x)$ and is finite, then $\partial f(x) = \partial f^{**}(x)$.*

The Young-Fenchel inequality (see (1.1.7)) implies that $f \geqslant f^*$. The next theorem shows that $f = f^{**}$ when f is a proper, convex and lower semicontinuous function (*cf.* [112, Theorem 2.33, page 77]).

Proposition 1.1.31 (Biconjugate)**.** *Let $f : X \to (-\infty, +\infty]$ be a proper, convex and lower semicontinuous function. Then $f = f^{**}$.*

The subdifferential mapping of conjugate functions is given by the inverse of the subdifferential mapping of the function (see [102, Proposition 1.4.4, page 20]).

Proposition 1.1.32 (Subdifferential of conjugate functions)**.** *Let $f : X \to (-\infty, +\infty]$ be a proper, convex and lower semicontinuous function on X. Then $\partial f^* = (\partial f)^{-1}$, where the inverse mapping $(\partial f)^{-1} : X^* \to 2^X$ is defined by*

$$(\partial f)^{-1}(\xi) = \{x \in X : \xi \in \partial f(x)\}.$$

1.1.5 Geometry of Banach Spaces

This subsection gathers some basic definitions and geometrical properties of Banach spaces which can be found in the book [47] (see also [87]).

Definition 1.1.33 (Types of Banach spaces)**.** *A Banach space X*

(i) *is smooth or has a Gâteaux differentiable norm if the limit*

$$\lim_{t \to 0} \frac{\|x + ty\| - \|x\|}{t} \tag{1.1.8}$$

exists for each $x, y \in S_X$;

(ii) *has a uniformly Gâteaux differentiable norm if for each $y \in S_X$ the limit (1.1.8) is attained uniformly for any $x \in S_X$;*

(iii) *is uniformly smooth if the limit (1.1.8) is attained uniformly for any $x, y \in S_X$;*

(iv) *is uniformly convex if the modulus of uniform convexity of the space X, that is, the function $\delta_X : [0,2] \to [0,1]$ defined by*

$$\delta_X(t) := \begin{cases} \inf\left\{1 - \left(\frac{1}{2}\right)\|x+y\| : \|x\| = 1 = \|y\|, \|x-y\| \geqslant t\right\} & , \quad t > 0 \\ 0 & , \quad t = 0 \end{cases}$$

satisfies $\delta_X(t) > 0$ for all $t > 0$.

A very related concept to the geometry of Banach spaces is the theory of duality mappings. Recall that a *gauge* is a continuous and strictly increasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\phi(0) = 0$ and $\lim_{t\to\infty} \phi(t) = \infty$. Associated with a gauge function ϕ, the *duality mapping* is the mapping $J_\phi : X \to 2^{X^*}$ given by

$$J_\phi(x) := \left\{ j_\phi(x) \in X^* : \langle j_\phi(x), x\rangle = \|j_\phi(x)\| \|x\|, \ \phi(\|x\|) = \|j_\phi(x)\| \right\}. \tag{1.1.9}$$

Remark 1.1.34 (Full domain of duality mappings). *According to the Hahn-Banach Theorem, $J_\phi(x)$ is a nonempty subset of X^* for every $x \in X$. Hence,* $\operatorname{dom} J_\phi = X$*.* ◇

Remark 1.1.35 (Properties of duality mappings). *It follows from* (1.1.9) *that J_ϕ is an odd mapping (i.e., $J_\phi(-x) = -J_\phi(x)$) and positively homogeneous (i.e., $\lambda J_\phi(x) = J_\phi(\lambda x)$ for any $\lambda > 0$).* ◇

If a gauge function ϕ is given by $\phi(t) = t$ for all $t \in \mathbb{R}^+$, then the corresponding duality mapping J_ϕ is called the *normalized duality mapping*, and is denoted by J_X. It follows from (1.1.9) that the normalized duality mapping J_X is defined by

$$J_X(x) := \left\{ \xi \in X^* : \langle \xi, x\rangle = \|x\|^2 = \|\xi\|_*^2 \right\}. \tag{1.1.10}$$

We can use another way to describe duality mappings. Given a gauge function ϕ, we define

$$\Phi(t) := \int_0^t \phi(s)\, ds.$$

Then it can be proved that Φ is convex and, for any $x \in X$, we have

$$J_\phi(x) = \partial\Phi(\|x\|).$$

Thus we have from the subdifferential inequality (see (1.1.5)) that for any $x, y \in X$,

$$\Phi(\|x+y\|) \leqslant \Phi(\|x\|) + \langle y, j_\phi(x+y)\rangle, \quad j_\phi(x+y) \in J_\phi(x+y).$$

For the normalized duality mapping J_X, the subdifferential inequality (see (1.1.5)) turns into

$$\|x+y\|^2 \leqslant \|x\|^2 + 2\langle y, j(x+y)\rangle, \quad j(x+y) \in J_X(x+y).$$

The relation between the normalized duality mapping J_X and a general duality mapping J_ϕ is easily given by the following identity:

$$J_\phi(x) = \frac{\phi(\|x\|)}{\|x\|} J_X(x), \quad x \neq 0, \quad x \in X.$$

The following result gathers relations between geometric properties of three classes of Banach spaces and features of duality mappings.

Proposition 1.1.36 (Characterization of Banach spaces)**.** *Let X be a Banach space. Given any gauge function ϕ, the following statements are true.*

(i) *The space X is smooth if and only if the duality mapping J_ϕ is single-valued (cf. [47, Theorem 1.10, page 46]).*

(ii) *The space X is uniformly smooth if and only if the duality mapping J_ϕ is single-valued and norm-to-norm uniformly continuous on bounded subsets of X (cf. [47, Theorem 2.16, page 54]).*

(iii) *If the space X has a uniformly Gâteaux differentiable norm then, J_ϕ is norm-to-weak* uniformly continuous on bounded subsets of X (cf. [47, Corollar 1.5, page 43] and [86]).*

Remark 1.1.37 (Duality mapping of the p-norm function)**.** *Take $\phi_p(t) := pt^{p-1}$. Then $\Phi_p(t) = t^p$. We denote the duality mapping with respect to ϕ_p by $J_p := \partial\Phi_p(\|x\|)$. In this case the function ϕ_p is invertible and*

$$\psi_p(t) = \phi_p^{-1}(t) = \left(\frac{t}{p}\right)^{1/(p-1)}$$

is again a gauge function. Define

$$\Psi_p = \int_0^t \psi_p(s)\, ds = (p-1)\, p^{p/(1-p)} t^{p/(p-1)}.$$

The duality mapping with respect to ψ_p is the mapping from X^* to 2^X given by

$$J_p^* = \partial\Psi_p\left(\|\xi\|_*\right) = p^{1/(1-p)} \|\xi\|_*^{1/(p-1)} \left(\|\cdot\|_*\right)'(\xi). \qquad \diamond$$

1.2 Bregman Distances

From now on we will use admissible functions.

Definition 1.2.1 (Admissible function)**.** *A function* $f : X \to (-\infty, +\infty]$ *is called admissible if it is proper, convex and lower semicontinuous, and Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f$*.*

In 1967 (*cf.* [27]) Bregman defined the following bifunction D_f.

Definition 1.2.2 (Bregman distance)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Gâteaux differentiable function. Then*

$$D_f(y, x) = f(y) - f(x) - \langle \nabla f(x), y - x \rangle . \tag{1.2.1}$$

A few years later, Censor and Lent [45] called it the *Bregman distance* with respect to the function f. The Bregman distance is at the core of this dissertation mainly because of its importance in optimization as a substitute for the usual distance or, more exactly, for the square of the norm of X. During the last 30 years, Bregman distances, have been studied in this connection by many researchers; for example, Bauschke, Borwein, Burachik, Butnariu, Censor, Combettes, Iusem, Reich and Resmerita (see, among many others, [7, 8, 33, 35, 41] and the references therein). Over the last 10 years the usage of this concept has been extended to many fields like Clustering, Image Reconstruction, Information Theory and Machine Learning. Because of all these reasons we are motivated to develop more tools for working with Bregman distances.

Remark 1.2.3 (Bregman distance is not a usual distance)**.** *It should be noted that* D_f *(see* (1.2.1)*) is not a distance in the usual sense of the term. In general,* D_f *is not symmetric and it does not satisfy the triangle inequality. Clearly,* $D_f(x, x) = 0$*, but* $D_f(y, x) = 0$ *may not imply* $x = y$ *as it happens, for instance, when* f *is a linear function on* X*.* ◇

In general we have the following result (*cf.* [7, Theorem 7.3(vi), page 642]).

Proposition 1.2.4 (Property of Bregman distances)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre function (see Definition 1.2.7 below). Then* $D_f(y, x) = 0$ *if and only if* $y = x$*.*

If f is a Gâteaux differentiable function, then Bregman distances have the following two important properties.

- The *three point identity*: for any $x \in \operatorname{dom} f$ and $y, z \in \operatorname{int}\operatorname{dom} f$, we have (see [50])

$$D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle \nabla f(z) - \nabla f(y), x - y \rangle . \tag{1.2.2}$$

- The *four point identity*: for any $y, w \in \operatorname{dom} f$ and $x, z \in \operatorname{int}\operatorname{dom} f$, we have

$$D_f(y, x) - D_f(y, z) - D_f(w, x) + D_f(w, z) = \langle \nabla f(z) - \nabla f(x), y - w \rangle . \tag{1.2.3}$$

Bregman distances are very interesting also because of the following property. In the following remark we emphasize we prove that Bregman distances can be considered as a generalization of the usual metric distance.

Remark 1.2.5 (Generalization of the metric distance)**.** *It is easy to check that when the Banach space X is a Hilbert space and $f(\cdot) = \|\cdot\|^2$, then $D_f(y,x) = \|y - x\|^2$, that is, the metric distance squared.* ◇

Directly from the definition of the Bregman distance (see (1.2.1)) we see that any property of D_f is inherited from a property of the function f. First, one may wonder if the Bregman distance can be defined for non-differentiable functions.

Remark 1.2.6 (Bregman distance of non-differentiable functions)**.** *For a non-differentiable function $f : X \to (-\infty, +\infty]$, there is a generalization of the Bregman distance (see* (1.2.1)*):*

$$D_f(y,x) := f(y) - f(x) - f^{\circ}(x, y - x). \qquad \diamond$$

In order to obtain more interesting and important properties of the Bregman distance (see (1.2.1)) we will study more deeply two classes of functions, the Legendre functions and the totally convex functions.

1.2.1 Legendre Functions

The notion of Legendre functions in a general Banach space X was introduced first by Bauschke, Borwein and Combettes in [7, Definition 5.2, page 634]. In our setting the Banach space X is reflexive, thus, from [7, Theorems 5.4 and 5.6, page 634] we can equivalently define the notion of Legendre functions as follows.

Definition 1.2.7 (Legendre)**.** *A function $f : X \to (-\infty, +\infty]$ is called Legendre if it satisfies the following two conditions.*

(L1) $\operatorname{int}\operatorname{dom} f \neq \varnothing$ *and the subdifferential ∂f is single-valued on its domain.*

(L2) $\operatorname{int}\operatorname{dom} f^* \neq \varnothing$ *and ∂f^* is single-valued on its domain.*

Since X is reflexive, we also have that $\nabla f = \left(\nabla f^*\right)^{-1}$ (see [7, Theorem 5.10, page 636]). This fact, combined with Conditions (L1) and (L2), implies the following equalities which will be very useful in the sequel:

(i)

$$\operatorname{ran}\nabla f = \operatorname{dom}\nabla f^* = \operatorname{int}\operatorname{dom} f^*. \tag{1.2.4}$$

(ii)

$$\operatorname{ran}\nabla f^* = \operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f. \tag{1.2.5}$$

Conditions (L1) and (L2), in conjunction with [7, Theorem 5.4, page 634], imply that both functions f and f^* are strictly convex and Gâteaux differentiable in the interior of their respective domains.

Several interesting examples of Legendre functions are presented in [6, 7]. Among them are the functions $(1/p)\|\cdot\|^p$ with $p \in (1, \infty)$, where the Banach space X is smooth and strictly convex, in particular, a Hilbert space.

1.2.2 Totally Convex Functions

The notion of totally convex functions was first introduced by Butnariu, Censor and Reich [33] in the context of the Euclidean space $\mathbb{R}^n$ because of its usefulness for establishing convergence of a Bregman projection method for finding common points of infinite families of closed and convex sets. In this finite dimensional environment total convexity hardly differs from strict convexity. In fact, a function with a closed domain in a finite dimensional Banach space is totally convex if and only if it is strictly convex. The relevancy of total convexity as a strengthened form of strict convexity becomes apparent when the Banach space on which the function is defined is of infinite dimension. In this case, total convexity is a property stronger than strict convexity but weaker than locally uniform convexity. Total convexity in the infinite dimensional case is studied intensively by Butnariu and Iusem and summarized in the book [35].

Total convexity is a property of the modulus of total convexity of the function which ensures that some sequential convergence properties which are true in the uniformity-like structure defined on the space via Bregman distances with respect to totally convex functions are inherited by the norm topology of the space. Therefore, in order to establish convergence and/or "good behavior" of some algorithms in infinite-dimensional settings, it is enough to do so with respect to the uniformity-like structure determined by the Bregman distance associated to totally convex functions.

This naturally leads to the question of whether totally convex functions with predesignate properties exist on a given Banach space. It is shown in [35] that totally convex functions can be found on any separable as well as on any reflexive Banach space.

In this subsection we present several properties of the modulus of total convexity associated to a convex function f. The interest in the modulus of total convexity and totally convex functions comes from the usefulness of these concepts when dealing with a class of recursive procedures for computing common fixed points for large families of operators and, in particular, solutions to optimization, convex feasibility, variational inequality and equilibrium problems as shown in the following chapters of this thesis.

Definition 1.2.8 (Total convexity at a point). *A function $f : X \to (-\infty, +\infty]$ is called totally convex at a point $x \in \operatorname{int}\operatorname{dom} f$ if its modulus of total convexity at x, that is, the function $\nu_f(x, \cdot) : [0, +\infty) \to [0, +\infty]$ defined by*

$$\nu_f(x, t) = \inf \{D_f(y, x) : y \in \operatorname{dom} f, \quad \|y - x\| = t\} \tag{1.2.6}$$

is positive whenever $t > 0$.

Remark 1.2.9 (Totally convex function). *A function $f : X \to (-\infty, +\infty]$ is called totally convex when it is totally convex at any point of* $\operatorname{int}\operatorname{dom} f$. ◇

Definition 1.2.10 (Total convexity on bounded subsets). *A function $f : X \to (-\infty, +\infty]$ is called totally convex on bounded subsets if $\nu_f(E, t)$ is positive for any nonempty and bounded subset E of X and for any $t > 0$, where the modulus of total convexity of the* *function f on the set E defined by*

$$\nu_f(E,t) := \inf\left\{\nu_f(x,t) : x \in E \bigcap \operatorname{int}\operatorname{dom} f\right\}.$$

The following proposition summarizes several properties of the modulus of total convexity (*cf.* [33, Proposition 2.4, page 26] and [35, Propostion 1.2.2, page 18]).

Proposition 1.2.11 (Properties of the modulus of total convexity)**.** *Let $f : X \to (-\infty, +\infty]$ be an admissible function. If $x \in \operatorname{int}\operatorname{dom} f$, then the following assertion hold.*

(i) *The domain of $\nu_f(x,\cdot)$ is an interval $[0, \tau_f(x))$ or $[0, \tau_f(x)]$ with $\tau_f(x) \in (0, +\infty]$.*

(ii) *If $c \in [1, +\infty)$ and $t \geqslant 0$, then $\nu_f(x, ct) \geqslant c\nu_f(x,t)$.*

(iii) *The function $\nu_f(x,\cdot)$ is superadditive, that is, for any $s, t \in [0, +\infty)$, we have*

$$\nu_f(x, s+t) \geqslant \nu_f(x,s) + \nu_f(x,t).$$

(iv) *The function $\nu_f(x,\cdot)$ is nondecreasing; it is strictly increasing if and only if f is totally convex at x.*

Moreover, if $X = \mathbb{R}^n$ and $f : C \to \mathbb{R}$, where C is an open, convex and unbounded subset of $\mathbb{R}^n$, then the following statements also hold.

(v) *The modulus of total convexity $\upsilon_f(x,\cdot)$ is continuous from the right on $(0, +\infty)$.*

(vi) *If $\bar{f} : \overline{C} \to \mathbb{R}$ is a convex and continuous extension of f to $\overline{C}$ and if $\upsilon_f(x,\cdot)$ is continuous, then, for each $t \in [0, +\infty)$, we have*

$$\upsilon_f(x,t) = \inf\left\{D_{\bar{f}}(y,x) : y \in \overline{C},\ \|y-x\| = t\right\}.$$

Definition 1.2.12 (Cofinite)**.** *A function $f : X \to (-\infty, +\infty]$ is called cofinite if* $\operatorname{dom} f^* = X^*$*.*

The following proposition follows from [37, Proposition 2.3, page 39] and [112, Theorem 3.5.10, page 164].

Proposition 1.2.13 (Sufficient condition for cofiniteness)**.** *If $f : X \to (-\infty, +\infty]$ is Fréchet differentiable and totally convex, then f is cofinite.*

Uniformly Convex Functions

The applications of totally convex functions discussed in this work requires us to know whether on a given Banach space totally convex functions exist and, eventually, how rich the class of totally convex functions on a given Banach space X. For that we compare the notion of the modulus of total convexity with the modulus of uniform convexity. Totally convex functions are strictly convex, but there exists a strictly convex function which is not totally convex (see [34]). In addition, a strongly related concept to total convexity is the concept of uniform convexity which was first introduced and studied in [110, 112] (see also [23, 26]).

Definition 1.2.14 (Uniform convexity (function)). *A function $f : X \to (-\infty, +\infty]$ is called uniformly convex if the function $\delta_f : [0, +\infty) \to [0, +\infty]$, defined by*

$$\delta_f(t) := \inf\left\{\frac{1}{2}f(x) + \frac{1}{2}f(y) - f\left(\frac{x+y}{2}\right) : \|y - x\| = t, \quad x, y \in \operatorname{dom} f\right\}, \tag{1.2.7}$$

is positive whenever $t > 0$. The function $\delta_f(\cdot)$ is called the modulus of uniform convexity of f.

Remark 1.2.15 (Uniform convexity implies total convexity). *According to [35, Proposition 1.2.5, page 25], if $x \in \operatorname{int}\operatorname{dom} f$ and $t \in [0, +\infty)$, then $\nu_f(x, t) \geqslant \delta_f(t)$ and, therefore, if f is uniformly convex, then it also is totally convex.* ◇

The converse implication is not generally valid, that is, a function f may be totally convex without being uniformly convex (for such an example see [35, Section 1.3, page 30]). Even for Gâteaux differentiable functions, the notions of uniformly and totally convex are not equivalent. However, if f is Fréchet differentiable, then we have the following result (*cf.* [37, Proposition 2.3, page 40]).

Proposition 1.2.16 (Total and uniform convexity coincide). *Suppose that $f : X \to (-\infty, +\infty]$ is a lower semicontinuous function. If $x \in \operatorname{int}\operatorname{dom} f$ and f is Fréchet differentiable at x, then f is totally convex at x if and only if f is uniformly convex at x.*

The notions of uniformly and totally convex on bounded subsets are equivalent under less restrictive conditions on the function, which can be seen in the following result (*cf.* [37, Proposition 4.2, page 53]).

Proposition 1.2.17 (Total and uniform convexity on bounded subsets). *Suppose that $f : X \to (-\infty, +\infty]$ is an admissible function. The function f is totally convex on bounded subsets if and only if f is uniformly convex on bounded subsets.*

Examples of Totally Convex Functions

In Banach spaces of infinite dimension finding totally convex functions is a challenging problem. This happens because, in an infinite dimensional context, we need to find totally convex functions designed in such a way that specific methods like the proximal point algorithm with Bregman distance and/or the projection type algorithms are effectively and efficiently computable. For instance, the function $f_p := \|\cdot\|^p$, $p \in (1, +\infty)$, is totally convex in any uniformly convex Banach space X (*cf.* [36, Theorem 1, page 322]).

Proposition 1.2.18 (Total convexity of p-norm in uniformly convex Banach spaces)**.** *If X is a uniformly convex Banach space, then, for each $p \in (1, +\infty)$, the function $\|\cdot\|^p$ is totally convex.*

Remark 1.2.19 (Total convexity of p-norm in locally uniformly convex Banach spaces)**.** *The function $\|\cdot\|^p$, $p \in (1, +\infty)$, is totally convex even if X is only locally uniformly convex, that is, if for each x in the unit ball of the space X, the function $\mu_X(x, \cdot) : [0, 2] \to [0, 1]$ defined by*

$$\mu_X(x,t) = \begin{cases} \inf\left\{1 - \left(\frac{1}{2}\right)\|x+y\| : \|y\| = 1, \|x-y\| \geqslant t\right\}, & t > 0 \\ 0, & t = 0 \end{cases}$$

is positive whenever $t > 0$. The function $\mu_X(x, \cdot)$ is called the modulus of locally uniform convexity of the space X. ◇

It is proved in [112] that $\|\cdot\|^p$, $p \in (1, +\infty)$, are uniformly convex at any point, and, hence are totally convex at any point. In [96] Resmerita proves that a Banach space on which the functions $\|\cdot\|^p$, $p \in (1, +\infty)$, are totally convex is necessarily strictly convex and has the *Kadeč-Klee property*, that is,

$$\left(\text{w-}\lim_{n\to\infty} x_n = x \quad \text{and} \quad \lim_{n\to\infty}\|x_n\| = \|x\|\right) \Longrightarrow \lim_{n\to\infty} x_n = x.$$

More precisely, she proves that those spaces are exactly the E-spaces (*cf.* [96, Theorem 3.2, page 8]).

Definition 1.2.20 (Sequentially consistent)**.** *A function $f : X \to (-\infty, +\infty]$ is called sequentially consistent (see [41]) if for any two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ in* dom f *and* int dom f*, respectively, such that the first one is bounded,*

$$\lim_{n\to\infty} D_f(y_n, x_n) = 0 \Rightarrow \lim_{n\to\infty}\|y_n - x_n\| = 0.$$

In the following result we see that in uniformly convex Banach spaces the function $\|\cdot\|^p$, $p \in (1, +\infty)$, is sequentially consistent (*cf.* [36, Corollary 1, page 326]).

Proposition 1.2.21 (Sequentially consistency of the p-norm function in uniformly convex Banach spaces)**.** *If X is a uniformly convex Banach space, then, for each $p \in (1, +\infty)$, the function $\|\cdot\|^p$ is sequentially consistent.*

The following result gives us a tool to generate more examples of totally convex functions (*cf.* [35, Proposition 1.2.7, page 28]).

Proposition 1.2.22 (Arithmetic of totally convex functions)**.** *The following statements hold.*

(i) *Let* $f_i : X \to (-\infty, +\infty]$*,* $1 \leqslant i \leqslant N$*, be totally convex functions with domains* $D_1, D_2, \ldots, D_N$*, respectively. Assume that*

$$\bigcap_{i=1}^{N} \operatorname{int} D_i \neq \varnothing.$$

Then, for any N *nonnegative real numbers* $c_1, c_2, \ldots, c_N$ *such that* $\sum_{i=1}^{N} c_i > 0$*, the function* $h := \sum_{i=1}^{N} c_i f_i$ *is totally convex and, for any* $x \in \bigcap_{i=1}^{N} \operatorname{int} D_i$ *and for all* $t \in [0, +\infty)$*, we have*

$$\nu_h(x,t) \geqslant \sum_{i=1}^{N} c_i \nu_{f_i}(x,t).$$

(ii) *If* $f : X \to (-\infty, +\infty]$ *is totally convex and lower semicontinuous with open domain* D*, and if* ϕ *is a real convex function defined, differentiable and strictly increasing on an open interval which contains* $f(D)$*, then the function* $g : X \to (-\infty, +\infty]$ *defined by* $g(x) = \phi(f(x))$*, if* $x \in D$*, and* $g(x) = +\infty$ *otherwise, is totally convex and we have*

$$\nu_g(x,t) \geqslant \phi'(f(x))\, \nu_f(x,t)$$

for all $x \in D$ *and any* $t \geqslant 0$*.*

Convergence analysis of many iterative algorithms for solving convex optimization problems in Banach spaces show that the produced sequences are bounded and that any weak accumulation points of which are optimal solutions of the problems these algorithms are supposed to solve. Obviously the identification of a convergent subsequence of a given sequence is difficult, if not impossible. Thus such algorithms can be used to compute approximate solutions of the given problem only to the extent to which either the objective function of the optimization problem is strictly convex because in such case the sequences those algorithms generate converge weakly to the necessarily unique optimal solution of the problem or one can regularize the problem by replacing the objective function with a strictly convex approximation of it in such a way that the optimal solution of the regularized problem exists and is close enough to the optimal solution set of the original problem.

Keeping the optimal solution of the regularized problem close to the optimal solution set of the original problem usually demands that the strictly convex approximation of the objective function should be uniform on bounded subsets. Also, the regularization process often requires the use of functions satisfying, among other conditions, a stronger form of strict convexity, namely, total convexity. Because of this, for a large number of optimization algorithms regularization of the optimization problem using totally convex and sufficiently uniform approximations of the objective function, which preserve some if not all of its continuity properties, is an implicit guarantee of a better convergence behavior of the computational procedure. Thus the abundance of totally convex functions and the possibility of using them as good approximations of given convex functions are crucial in numerous optimization algorithms.

Butnariu, Reich and Zaslavski [40] prove that whenever there exists a function which is totally convex at each point of K, and Lipschitz continuous on any bounded subset of K, then the set of totally convex functions on K, the set of lower semicontinuous and totally convex functions on K, the set of continuous and totally convex functions on K, as well as the set of Lipschitz continuous and totally convex functions on K are large in the sense that they contain countable intersections of open (in the weak topology) and everywhere dense (in the strong topology) subsets. This result is meaningful because it implies the existence of large pools of totally convex functions.

At the same time it guarantees that given a convex function f with some continuity features, one can find uniform on bounded subsets totally convex approximations of it which not only preserve the continuity features of f, but also have corresponding Bregman distances which are uniformly close on bounded subsets to the Bregman distance corresponding to f itself.

Examples of Totally Convex Functions in Euclidian Spaces

Let $X = \mathbb{R}$. In this section we study in detail the total convexity of the *Boltzmann-Shannon entropy*

$$\mathcal{BS}(x) := x \log(x) - x, \quad 0 < x < +\infty \tag{1.2.8}$$

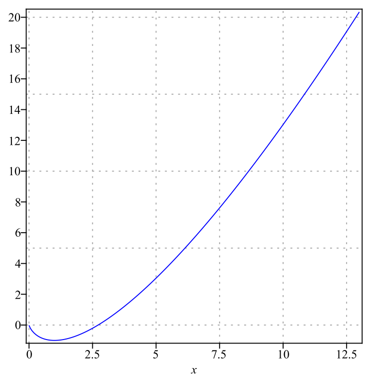

Figure 1.1: The Boltzmann-Shannon entropy

and the *Fermi-Dirac entropy*

$$\mathcal{FD}(x) := x \log(x) + (1 - x) \log(1 - x), \quad 0 < x < 1. \tag{1.2.9}$$

Each of these functions can be defined to be zero, by its limits, at the endpoints of their domains.

We study the entropies $\mathcal{BS}$ and $\mathcal{FD}$ in detail because of their importance in applications. These two functions, which form a large part of the basis for the classical information theory,

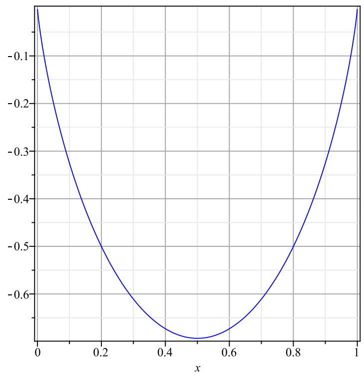

Figure 1.2: The Fermi-Dirac entropy

arguably provide the only consistent measures of the average uncertainty in predicting outcomes of a random experiment (see [68]).

Moreover, both $D_{\mathcal{BS}}$ and $D_{\mathcal{FD}}$ are jointly convex [24, 26], an uncommon property which they share with $(x, y) \mapsto \|x - y\|^2$. The utility of both the Boltzmann-Shannon and the Fermi-Dirac entropies is enhanced because they are totally convex. In the following two results (Propositions 1.2.23 and 1.2.24) we calculate the modulus of total convexity of the $\mathcal{BS}$ entropy and show that $\mathcal{BS}$ is totally convex (see [33, 35]). Propositions 1.2.25 and 1.2.26 are analogous results concerning the $\mathcal{FD}$ entropy.

We start with formula of the modulus of total convexity of $\mathcal{BS}$ (*cf.* [25, Propsoition 4.2, page 170]).

Proposition 1.2.23 (Modulus of total convexity of $\mathcal{BS}$)**.** *The modulus of total convexity of $\mathcal{BS}$ on $(0, +\infty)$ is given by*

$$\nu_{\mathcal{BS}}(x, t) = x\left[\left(1 + \frac{t}{x}\right)\log\left(1 + \frac{t}{x}\right) - \frac{t}{x}\right], \quad x \in (0, +\infty), \quad t \geqslant 0. \tag{1.2.10}$$

Proof. Let $x_0 \in (0, +\infty)$ and $0 < t < x_0$. It is clear from the definition of the modulus of total convexity (see (1.2.6)) that

$$\begin{aligned}\nu_{\mathcal{BS}}(x_0, t) &= \min\left\{D_{\mathcal{BS}}(x_0 + t, x_0), D_{\mathcal{BS}}(x_0 - t, x_0)\right\} \\ &= \min\left\{(x_0 + t)\log\left(\frac{x_0 + t}{x_0}\right) - t, (x_0 - t)\log\left(\frac{x_0 - t}{x_0}\right) + t\right\} \\ &= \min\left\{x_0\left[\left(1 + \frac{t}{x_0}\right)\log\left(1 + \frac{t}{x_0}\right) - \frac{t}{x_0}\right],\right. \\ &\left. x_0\left[\left(1 - \frac{t}{x_0}\right)\log\left(1 - \frac{t}{x_0}\right) + \frac{t}{x_0}\right]\right\}.\end{aligned}$$

In order to find this minimum we define a function $\varphi : [0, x_0) \to \mathbb{R}$ by

$$\varphi(t) := x_0\left[\left(1 - \frac{t}{x_0}\right)\log\left(1 - \frac{t}{x_0}\right) + \frac{t}{x_0} - \left(1 + \frac{t}{x_0}\right)\log\left(1 + \frac{t}{x_0}\right) + \frac{t}{x_0}\right].$$

It is clear that $\varphi(0) = 0$ and

$$\varphi'(t) = -\log\left(1 - \left(\frac{t}{x_0}\right)^2\right),$$

and so φ is increasing for all $t < x_0$. Thus $\varphi(t) > 0$ for any $t < x_0$, which means that

$$\nu_{\mathcal{BS}}(x_0, t) = x_0\left[\left(1 + \frac{t}{x_0}\right)\log\left(1 + \frac{t}{x_0}\right) - \frac{t}{x_0}\right]$$

for any $t < x_0$. If $t \geqslant x_0$ then the point $t - x_0$ does not belong to the domain of $\mathcal{BS}$ and therefore

$$\nu_{\mathcal{BS}}(x_0, t) = D_{\mathcal{BS}}(x_0 + t, x_0) = x_0\left[\left(1 + \frac{t}{x_0}\right)\log\left(1 + \frac{t}{x_0}\right) - \frac{t}{x_0}\right].$$

Hence (1.2.10) holds for any $t \geqslant 0$. □

Now we will prove that $\mathcal{BS}$ is totally convex on $(0, +\infty)$ but not uniformly convex (*cf.* [25, Propsoition 4.3, page 171]).

Proposition 1.2.24 (Total convexity of $\mathcal{BS}$ on $(0, +\infty)$)**.** *The function $\mathcal{BS}$ is totally convex, but not uniformly convex on* $(0, +\infty)$.

Proof. We need to show that $\nu_{\mathcal{BS}}(x_0, t) > 0$ for any $t > 0$. We know that $\nu_{\mathcal{BS}}(x_0, 0) = 0$ and from Proposition 1.2.23 we obtain that

$$\frac{\partial}{\partial t}\left(\nu_{\mathcal{BS}}(x_0, t)\right) = \log\left(1 + \frac{t}{x_0}\right) > 0, \quad t > 0.$$

This means that $\nu_{\mathcal{BS}}(x_0, t)$ is a strictly increasing function for all $t > 0$. Thence, $\nu_{\mathcal{BS}}(x_0, t) > 0$ for any $t > 0$ and so $\mathcal{BS}$ is totally convex on $(0, +\infty)$, as asserted.

From Remark 1.2.15 we have, for any $t > 0$, that

$$0 \leqslant \delta_{\mathcal{BS}}(t) \leqslant \lim_{x \to \infty} \nu_{\mathcal{BS}}(x, t) = 0.$$

It follows that $\delta_{\mathcal{BS}}(t) = 0$ and thus $\mathcal{BS}$ is not uniformly convex. □

In [33] it is mentioned that the modulus of total convexity of the function $f(x) = x\log(x)$ is also given by (1.2.10) and that f is totally convex. Note that $D_f = D_{\mathcal{BS}}$.

The following results show that $\mathcal{FD}$ is both totally convex and uniformly convex (*cf.* [25, Propsoitions 4.5 and 4.6, pages 171–172]).

Proposition 1.2.25 (Modulus of total convexity of $\mathcal{FD}$)**.** *The modulus of total convexity of $\mathcal{FD}$ is given by*

$$\nu_{\mathcal{FD}}(x,t) = x\left[\left(1+\frac{t}{x}\right)\log\left(1+\frac{t}{x}\right)+\left(\frac{1-t}{x}-1\right)\log\left(1-\frac{t}{1-x}\right)\right], \tag{1.2.11}$$

when $0 < x \leqslant 1/2$ *and* $0 < t < 1-x$, *and by*

$$\nu_{\mathcal{FD}}(x,t) = x\left[\left(1-\frac{t}{x}\right)\log\left(1-\frac{t}{x}\right)+\left(\frac{1+t}{x}-1\right)\log\left(1+\frac{t}{1-x}\right)\right] \tag{1.2.12}$$

when $1/2 \leqslant x < 1$ *and* $0 < t < x$.

Proof. Let $x_0 \in (0,1)$. Denote $M = \max\{x_0, 1-x_0\}$ and $m = \min\{x_0, 1-x_0\}$. If $0 < t < m$, then it is clear from the definition of the modulus of total convexity (see (1.2.6)) that

$$\begin{aligned}\nu_{\mathcal{FD}}(x_0,t) &= \min\left\{D_{\mathcal{FD}}(x_0+t,x_0), D_{\mathcal{FD}}(x_0-t,x_0)\right\}\\ &= \min\left\{x_0\left[\left(1+\frac{t}{x_0}\right)\log\left(1+\frac{t}{x_0}\right)+\left(\frac{1-t}{x_0}-1\right)\log\left(1-\frac{t}{1-x_0}\right)\right],\right.\\ &\left. x_0\left[\left(1-\frac{t}{x_0}\right)\log\left(1-\frac{t}{x_0}\right)+\left(\frac{1+t}{x_0}-1\right)\log\left(1+\frac{t}{1-x_0}\right)\right]\right\}.\end{aligned}$$

In order to find this minimum we define a function $\psi : [0,m) \to \mathbb{R}$ by

$$\begin{aligned}\psi(t) := x_0&\left[\left(1-\frac{t}{x_0}\right)\log\left(1-\frac{t}{x_0}\right)+\left(\frac{1+t}{x_0}-1\right)\log\left(1+\frac{t}{1-x_0}\right)\right]\\ -x_0&\left[\left(1+\frac{t}{x_0}\right)\log\left(1+\frac{t}{x_0}\right)+\left(\frac{1-t}{x_0}-1\right)\log\left(1-\frac{t}{1-x_0}\right)\right].\end{aligned}$$

It is clear that $\psi(0) = 0$ and

$$\psi'(t) = \log\left(1-\left(\frac{t}{1-x_0}\right)^2\right)-\log\left(1-\left(\frac{t}{x_0}\right)^2\right).$$

Therefore, for any $0 < t < m$, the function ψ is increasing when $0 < x \leqslant 1/2$ and decreasing when $1/2 \leqslant x < 1$. Hence, for any $0 < t < m$, the function $\psi(t) > 0$ when $0 < x \leqslant 1/2$ and $\psi(t) < 0$ when $1/2 \leqslant x < 1$. If $m \leqslant t < M$, then one of the points $x_0 - t$ or $x_0 + t$ belongs to the domain of $\mathcal{FD}$ and the second does not. Therefore the modulus of total convexity of $\mathcal{FD}$ is given by (1.2.11) and (1.2.12) in all cases. □

Proposition 1.2.26 (Total convexity of $\mathcal{FD}$ on $(0,1)$)**.** *The function $\mathcal{FD}$ is totally convex on* $(0,1)$.

Proof. We need to show that $\nu_{\mathcal{FD}}(x_0,t) > 0$ for any $t > 0$. We know that $\nu_{\mathcal{FD}}(x_0,0) = 0$

and from Proposition 1.2.25 we obtain that

$$\frac{\partial}{\partial t}\left(\nu_{\mathcal{FD}}\left(x_0,t\right)\right)=\begin{cases}\log\left(1+\frac{t}{x(1-x-t)}\right), & 0<x\leqslant 1/2,\quad 0<t<1-x,\\ \log\left(1+\frac{t}{(1-x)(x-t)}\right), & 1/2\leqslant x<1,\qquad 0<t<x.\end{cases}$$

This means that $\nu_{\mathcal{FD}}\left(x_0,t\right)$ is a strictly increasing function for all $t>0$. Thence, we get that $\nu_{\mathcal{FD}}\left(x_0,t\right)>0$ for any $t>0$ and so $\mathcal{FD}$ is totally convex on $(0,1)$, as asserted. □

The following technical result will be useful in order to prove that $\mathcal{FD}$ is uniformly convex on $(0,1)$ (*cf.* [25, Lemma 4.7, page 172]).

Lemma 1.2.27 (Uniform convexity of a one variable function)**.** *Let* $f:(a,b)\to\mathbb{R}$ *be twice differentiable function. If* $f''(x)\geqslant m>0$ *on* (a,b)*, then* f *is uniformly convex there.*

Proof. Let $x,y\in(a,b)$ with $\|y-x\|=t>0$. Then

$$f(x)=f\left(\frac{x+y}{2}\right)+f'\left(\frac{x+y}{2}\right)\left(\frac{x-y}{2}\right)+\frac{f''(\alpha)}{2}\left(\frac{x-y}{2}\right)^2$$

and

$$f(y)=f\left(\frac{x+y}{2}\right)+f'\left(\frac{x+y}{2}\right)\left(\frac{y-x}{2}\right)+\frac{f''(\beta)}{2}\left(\frac{y-x}{2}\right)^2$$

for some $\alpha,\beta\in(a,b)$. Therefore

$$\frac{f(x)}{2}+\frac{f(y)}{2}-f\left(\frac{x+y}{2}\right)=\frac{f''(\alpha)}{4}\left(\frac{x-y}{2}\right)^2+\frac{f''(\beta)}{4}\left(\frac{y-x}{2}\right)^2\geqslant\frac{mt^2}{8}>0,$$

as asserted. □

The following result follows immediately from the previous lemma (*cf.* [25, Proposition 4.8, page 172]).

Proposition 1.2.28 (Uniform convexity of $\mathcal{FD}$ on $(0,1)$)**.** *The function* $\mathcal{FD}$ *is uniformly convex on* $(0,1)$.

Proof. This result follows immediately from Lemma 1.2.27 because $\mathcal{FD}''(x)\geqslant 4$ for all $x\in(0,1)$. □

Now, assume that $X=\mathbb{R}^n$. In this case the *Boltzmann-Shannon entropy* is the function $\mathcal{BS}_n:\mathbb{R}^n_{++}\to\mathbb{R}$ defined by

$$\mathcal{BS}_n(x):=\sum_{i=1}^{n}x_i\log\left(x_i\right)-x_i,\quad x\in\mathbb{R}^n_{++}.$$

The following result shows that $\mathcal{BS}_n$ is a totally convex function on $\mathbb{R}^n_{++}$ (*cf.* [25, Proposition 4.19, page 178]).

Proposition 1.2.29 (Total convexity of $\mathcal{BS}_n$ on $\mathbb{R}^n_{++}$)**.** *The function $\mathcal{BS}_n$ is totally convex on $\mathbb{R}^n_{++}$ and its modulus of total convexity satisfies*

$$\nu_{\mathcal{BS}_n}(x,t) \geqslant \min_{1\leqslant i\leqslant n}\left\{x_i\left[\left(1+\frac{t}{x_i\sqrt{n}}\right)\log\left(1+\frac{t}{x_i\sqrt{n}}\right)-\frac{t}{x_i\sqrt{n}}\right]\right\}.$$

Proof. Let $\overline{\mathcal{BS}}:[0,+\infty)\to\mathbb{R}$ be the convex and continuous function defined by

$$\overline{\mathcal{BS}}(x) := \begin{cases} x\log(x)-x, & x>0 \\ 0, & x=0. \end{cases}$$

It is clear that the restriction of $\overline{\mathcal{BS}}$ to $(0,+\infty)$ is exactly $\mathcal{BS}$. The function $\overline{\mathcal{BS}_n}:\mathbb{R}^n_{++}\to\mathbb{R}$ defined by

$$\overline{\mathcal{BS}_n}(x) := \sum_{i=1}^{n}\overline{\mathcal{BS}}(x_i)$$

is convex, continuous and its restriction to $\mathbb{R}^n_{++}$ is exactly $\mathcal{BS}_n$. Let

$$\nu_{\overline{\mathcal{BS}_n}}(x,t) = \inf\left\{D_{\overline{\mathcal{BS}_n}}(y,x) : y\in\mathbb{R}^n_+,\ \|y-x\|=t\right\}.$$

Since the set $\left\{y\in\mathbb{R}^n_+ : \|y-x\|=t\right\}$ is compact in $\mathbb{R}^n$ and $D_{\overline{\mathcal{BS}_n}}(\cdot,x)$ is continuous on this set, there exists $\bar{y}\in\mathbb{R}^n_+$ such that $\|x-\bar{y}\|=t$ and

$$\nu_{\mathcal{BS}_n}(x,t) \geqslant \nu_{\overline{\mathcal{BS}_n}}(x,t) = D_{\overline{\mathcal{BS}_n}}(\bar{y},x) = \sum_{i=1}^{n}D_{\overline{\mathcal{BS}}}(\bar{y}_i,x_i).$$

The modulus of total convexity of $\mathcal{BS}$ is given by (1.2.10) and is continuous in t. Therefore we can apply Proposition 1.2.11(vi) and obtain that, for each $1\leqslant i\leqslant n$,

$$D_{\overline{\mathcal{BS}}}(\bar{y}_i,x_i) \geqslant \nu_{\mathcal{BS}}(x_i,|x_i-\bar{y}_i|).$$

Hence

$$\nu_{\mathcal{BS}_n}(x,t) \geqslant \sum_{i=1}^{n}\nu_{\mathcal{BS}}(x_i,|x_i-\bar{y}_i|). \tag{1.2.13}$$

When $t>0$, we have that $|x_i-\bar{y}_i|>0$ for at least one index i. As noted in Proposition 1.2.23, the function $\mathcal{BS}$ is totally convex. Consequently, $\nu_{\mathcal{BS}}(x_i,|x_i-\bar{y}_i|)>0$ for at least one index i. This and (1.2.13) show that, if $t>0$, then $\nu_{\mathcal{BS}_n}(x,t)>0$, *i.e.*, $\mathcal{BS}_n$ is totally convex on $\mathbb{R}^n_{++}$.

Since for at least one index i_0 we have $|x_i - \bar{y}_i| \geqslant t/\sqrt{n}$, we deduce from (1.2.13) that

$$\begin{aligned}\nu_{\mathcal{BS}_n}(x,t) &\geqslant \sum_{i=1}^{n} \nu_{\mathcal{BS}}(x_i, |x_i - \bar{y}_i|) \geqslant \nu_{\mathcal{BS}}(x_{i_0}, |x_{i_0} - \bar{y}_{i_0}|) \\ &\geqslant \nu_{\mathcal{BS}}\left(x_{i_0}, t/\sqrt{n}\right) \geqslant \min_{1 \leqslant i \leqslant n}\left\{\nu_{\mathcal{BS}}\left(x_i, t/\sqrt{n}\right)\right\}.\end{aligned}$$

When combined with (1.2.10), this inequality completes the proof. □

1.2.3 The Bregman Projection

If f is strictly convex on $\operatorname{int}\operatorname{dom} f$, then so is $D_f(\cdot, x)$. Therefore, if f is strictly convex on $\operatorname{int}\operatorname{dom} f$ and if K is a subset of X such that $\operatorname{int}\operatorname{dom} f \bigcap K \neq \varnothing$, then there exists at most one point $\operatorname{proj}_K^f(x) \in \operatorname{int}\operatorname{dom} f \bigcap K$ such that

$$D_f\left(\operatorname{proj}_K^f(x), x\right) = \inf\left\{D_f(y,x) : y \in \operatorname{int}\operatorname{dom} f \bigcap K\right\}. \tag{1.2.14}$$

This point (if any) is called the *Bregman projection* of $x \in \operatorname{int}\operatorname{dom} f$ onto K (*cf.* [45]). A question of essential interest in the sequel is whether, and under which conditions concerning the function f, the Bregman projection $\operatorname{proj}_K^f(x)$ exists and is unique for each $x \in \operatorname{int}\operatorname{dom} f$. The following three conditions are sufficient for ensuring the existence and uniqueness of the Bregman projection $\operatorname{proj}_K^f(x)$ for each $x \in \operatorname{int}\operatorname{dom} f$.

(A1) The set $\operatorname{dom} f$ is closed with nonempty interior and f is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f$.

(A2) The function f is strictly convex and continuous on $\operatorname{dom} f$.

(A3) For each $x \in \operatorname{int}\operatorname{dom} f$ and for any real number α, the *sub-level set* $\operatorname{lev}_\alpha^{D_f}(x)$ defined by

$$\operatorname{lev}_\alpha^{D_f}(x) = \{y \in \operatorname{dom} f : D_f(y,x) \leqslant \alpha\}$$

is bounded.

Applying Proposition 1.1.10(v) and (vi), we deduce the following result (*cf.* [35, Corollary 1.1.6, page 8]).

Proposition 1.2.30 (Continuity of the Bregman distance). *Let* $f : X \to (-\infty, +\infty]$ *be a convex function such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$. *If* f *is continuous on* $\operatorname{int}\operatorname{dom} f$, *then, for any* $x \in \operatorname{int}\operatorname{dom} f$, *the function* $f^\circ(x, \cdot)$ *is finite and Lipschitzian on* X. *Also, the function* $D_f(\cdot, x)$ *is locally Lipschitzian on* $\operatorname{int}\operatorname{dom} f$. *In particular, these statements hold when* f *is lower semicontinuous.*

Now we prove that Conditions (A1)–(A3) guarantee that there exists a unique Bregman projection for any $x \in \operatorname{int}\operatorname{dom} f$ (*cf.* [36, Lemma 2.2, page 273]).

Proposition 1.2.31 (Existence and uniqueness of the Bregman projection). *Let* $f : X \to (-\infty, +\infty]$ *be a function which satisfies Conditions (A1)–(A3) and let* K *be a nonempty, closed and convex subset of* X *such that* $\operatorname{int}\operatorname{dom} f \bigcap K \neq \varnothing$. *Then the Bregman projection* $\operatorname{proj}_K^f(x)$ *exists and is unique for any* $x \in \operatorname{int}\operatorname{dom} f$.

Proof. Let $z \in \operatorname{int}\operatorname{dom} f \bigcap K$ and define $r = D_f(z, x)$. Denote by C the following intersection: $\operatorname{int}\operatorname{dom} f \bigcap K \bigcap \operatorname{lev}_r^{D_f}(x)$. The set C is nonempty, bounded, closed and convex and, thus, it is weakly compact. The function $D_f(\cdot, x)$ is convex and continuous (see Proposition 1.2.30) and, therefore, weakly lower semicontinuous on the weakly compact set C. Consequently, $D_f(\cdot, x)$ achieves its minimal value at a point of C. This point obviously satisfies

$$D_f\left(\operatorname{proj}_K^f(x), x\right) = \inf\left\{D_f(y, x) : y \in \operatorname{int}\operatorname{dom} f \bigcap K\right\}$$

and it is the unique point with this property since $D_f(\cdot, x)$ is strictly convex. □

Proposition 1.2.31 shows that Conditions (A1)–(A3) are sufficient to ensure that the Bregman projection operator $x \to \operatorname{proj}_K^f(x) : \operatorname{int}\operatorname{dom} f \to K$ is well defined for any nonempty, closed and convex subset K of X such that $\operatorname{int}\operatorname{dom} f \bigcap K \neq \varnothing$. Verification of these conditions, and especially of Condition (A3), in particular, may be, sometimes, difficult. In the pervious section we proved several properties of totally convex functions. This class of functions is also important here since any totally convex function which satisfies Condition (A1) also satisfies Condition (A3).

The following result shows that in finite-dimensional spaces, functions which satisfy Conditions (A1) and (A2) also satisfy Condition (A3). If the space is of infinite-dimension, then totally convex functions which satisfies Condition (A1) also satisfy Condition (A3) (*cf.* [36, Lemma 2.5, page 274]).

Proposition 1.2.32 (Property of totally convex functions). *Let* $f : X \to (-\infty, +\infty]$ *be a totally convex function which satisfies Condition (A1). Then* f *satisfies Condition (A3). Moreover, if the Banach space* X *has finite dimension and, in addition to Condition (A1), the function* f *also satisfies Condition (A2), then* f *is totally convex and the modulus of total convexity* $\nu_f(x, \cdot)$ *is continuous from the left on* $(0, +\infty)$ *for any* $x \in X$.

Proof. If the result does not hold, then for some $\alpha \in \mathbb{R}$, there exists an unbounded sequence $\{u_n\}_{n\in\mathbb{N}}$ in $\operatorname{lev}_\alpha^{D_f}(x)$. Hence, for each nonnegative integer n, we have from Proposition 1.2.11(ii) that

$$\alpha \geqslant D_f(u_n, x) = \nu_f(x, \|x - u_n\|) \geqslant \|x - u_n\| \, \nu_f(x, 1).$$

Since $\nu_f(x, 1) > 0$ and $\lim_{n\to\infty} \|x - u_n\| = \infty$, it results that α cannot be finite and this is a contradiction.

Now suppose that X is of finite dimension and f is continuous and strictly convex on the closed set $\operatorname{dom} f$. Fix $x \in \operatorname{int}\operatorname{dom} f$ and $t \in (0, +\infty)$. Note that the set

$$\{y \in \operatorname{dom} f : \|x - y\| = t\}$$

is compact in X since it is bounded and closed. We also know that $D_f(\cdot, x)$ is continuous on $\operatorname{dom} f$ (see Proposition 1.2.30). Consequently, there exists a point $y \in \operatorname{dom} f$ with $\|x - y\| = t$ such that $\nu_f(x, t) = D_f(y, x)$. Since f is strictly convex on $\operatorname{dom} f$, we have from Proposition 1.1.10(iii) that

$$f^{\circ}(x, y - x) < f(y) - f(x).$$

This implies that $D_f(y, x) > 0$ (see Remark 1.2.6). Hence $\nu_f(x, t) = D_f(y, x) > 0$ and this proves that f is totally convex.

Now, suppose that $\bar{t} \in (0, +\infty)$ and let $\{t_n\}_{n \in \mathbb{N}} \subset (0, +\infty)$ be an increasingly sequence that converges to $\bar{t}$. For each nonnegative integer n, let $y_n \in \operatorname{dom} f$ be a point such that $\|x - y_n\| = t_n$ and $\nu_f(x, t_n) = D_f(y_n, x)$. The sequence $\{y_n\}_{n \in \mathbb{N}}$ is bounded since $\{t_n\}_{n \in \mathbb{N}}$ converges. Hence, there exists a convergent subsequence $\{y_{n_k}\}_{k \in \mathbb{N}}$ of $\{y_n\}_{n \in \mathbb{N}}$. Let $\bar{y} = \lim_{k \to \infty} y_{n_k}$. Then $\bar{y} \in \operatorname{dom} f$ and $\|x - \bar{y}\| = \bar{t}$. Since, for any $k \in \mathbb{N}$, $\nu_f(x, \bar{t}) \geqslant \nu_f(x, t_n)$, and the sequence $\{\nu_f(x, t_n)\}_{n \in \mathbb{N}}$ is increasing, we get

$$\nu_f(x, \bar{t}) \geqslant \lim_{n \to \infty} \nu_f(x, t_n) = \lim_{k \to \infty} \nu_f(x, t_{n_k}) = \lim_{k \to \infty} D_f(y_{n_k}, x) = D_f(\bar{y}, x) \geqslant \nu_f(x, \bar{t}).$$

Consequently, $\nu_f(x, \bar{t}) = \lim_{n \to \infty} \nu_f(x, t_n)$, *i.e.*, $\nu_f(x, \cdot)$ is continuous from the left at $\bar{t}$. □

Proposition 1.2.31 shows that if f satisfies Conditions (A1)–(A3), then the Bregman projection exists and unique. The next result shows that the Bregman projection exists and is unique also under two different conditions (*cf.* [41, Proposition 4.1, page 21]). One of the conditions is coercivity.

Definition 1.2.33 (Coercivity). *A function $f : X \to (-\infty, +\infty]$ is called*

(i) *coercive if* $\lim_{\|x\| \to \infty} f(x) = +\infty$*;*

(ii) *super-coercive if* $\lim_{\|x\| \to \infty} (f(x) / \|x\|) = +\infty$*.*

Proposition 1.2.34 (Another existence and uniqueness result). *Let $f : X \to (-\infty, +\infty]$ be a strictly convex function on* $\operatorname{dom} f$*. Let $x \in \operatorname{int}\operatorname{dom} f$ and let $K \subset \operatorname{int}\operatorname{dom} f$ be a nonempty, closed and convex set. If the sub-level sets $\operatorname{lev}_{\alpha}^{D_f}(x)$ are bounded for any $\alpha \in [0, +\infty)$, then there exists a unique Bregman projection of x onto K with respect to f. In particular, this happens if any of the following conditions hold:*

(i) f *is totally convex at each point of* $\operatorname{dom} f$*;*

(ii) f *is super-coercive.*

Proof. Since f is strictly convex on $\operatorname{int}\operatorname{dom} f$, so is $D_f(\cdot, x)$. This guarantees that there is no more than one vector y satisfying (1.2.14). Since the function $D_f(\cdot, x)$ is also convex, lower semicontinuous and $\operatorname{lev}_\alpha^{D_f}(x)$ are bounded for any $\alpha \in [0, +\infty)$, it results that $D_f(\cdot, x)$ has at least one minimizer in the convex set K, that is, the Bregman projection of x onto K exists and is unique.

(i) Suppose that f is totally convex and that, for some $\alpha \geqslant 0$, the set $\operatorname{lev}_\alpha^{D_f}(x)$ is unbounded. Then there exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ contained in $\operatorname{lev}_\alpha^{D_f}(x)$ such that $\lim_{n\to\infty}\|y_n\| = +\infty$. From the definition of the modulus of total convexity (see (1.2.6)), for any $x \in \operatorname{int}\operatorname{dom} f$ and for any $n \in \mathbb{N}$ such that $\|y_n - x\| \geqslant 1$, one has from Proposition 1.2.11(ii) that

$$\alpha \geqslant D_f(y_n, x) \geqslant \nu_f(x, \|y_n - x\|) = \|y_n - x\|\,\nu_f(x, 1)\,. \tag{1.2.15}$$

Since f is totally convex at x, it results that $\nu_f(x, 1) > 0$. Therefore, by letting $n \to \infty$ in (1.2.15) one gets a contradiction. Hence the set $\operatorname{lev}_\alpha^{D_f}(x)$ is bounded for all $\alpha \in [0, +\infty)$.

(ii) Now suppose that f is super-coercive and that, for some $\alpha \geqslant 0$, there exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ contained in $\operatorname{lev}_\alpha^{D_f}(x)$ such that $\lim_{n\to\infty}\|y_n\| = +\infty$. Then from Proposition 1.1.29(v) we get

$$\begin{aligned}\alpha \geqslant D_f(y_n, x) &= f(y_n) - f(x) - \langle \nabla f(x), y_n - x\rangle \\ &= f(y_n) + f^*(\nabla f(x)) - \langle \nabla f(x), y_n\rangle\,.\end{aligned}$$

Now from the Cauchy-Schwarz inequality we get

$$\begin{aligned}\alpha &\geqslant f^*(\nabla f(x)) + f(y_n) - \|\nabla f(x)\|_* \|y_n\| \\ &= f^*(\nabla f(x)) + \|y_n\|\left(\frac{f(y_n)}{\|y_n\|} - \|\nabla f(x)\|_*\right).\end{aligned} \tag{1.2.16}$$

Letting $n \to \infty$ in (1.2.16) one gets a contradiction. Hence the set $\operatorname{lev}_\alpha^{D_f}(x)$ is bounded for all $\alpha \in [0, +\infty)$. □

Similarly to the metric projection in Hilbert spaces, the Bregman projections have a variational characterization. These properties extend to the Bregman projection with respect to totally convex and admissible functions (*cf.* [41, Corollary 4.4, page 23]).

Proposition 1.2.35 (Characterizations of the Bregman projection)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function which is totally convex. Let* $x \in \operatorname{int}\operatorname{dom} f$ *and let* $K \subset \operatorname{int}\operatorname{dom} f$ *be a nonempty, closed and convex set. If* $\hat{x} \in K$*, then the following statements are equivalent.*

(i) *The vector* $\hat{x}$ *is the Bregman projection of* x *onto* K*.*

(ii) *The vector* $\hat{x}$ *is the unique solution of the variational inequality*

$$\langle \nabla f(x) - \nabla f(z), z - y \rangle \geqslant 0, \qquad \forall y \in K. \tag{1.2.17}$$

(iii) *The vector* $\hat{x}$ *is the unique solution of the inequality*

$$D_f(y, z) + D_f(z, x) \leqslant D_f(y, x), \qquad \forall y \in K.$$

Proof. (i) $\Leftrightarrow$ (ii) Suppose that (i) holds. Then, for any $u \in K$ one has $D_f(\hat{x}, x) \leqslant D_f(u, x)$. In particular, this holds for $u = (1-\tau)\hat{x} + \tau y$ for all $y \in K$ and for all $\tau \in [0, 1]$. Since f is strictly convex and continuous function (see Corollary 1.1.7) so it is also true for $D_f(\cdot, y)$ (see Proposition 1.2.30), one obtains from the subdifferential inequality (see (1.1.5)) that

$$0 \geqslant D_f(\hat{x}, y) - D_f((1-\tau)\hat{x} + \tau y, y) \geqslant \left\langle [D_f(\cdot, y)]'((1-\tau)\hat{x} + \tau y), \tau(\hat{x} - y) \right\rangle,$$

where $[D_f(\cdot, y)]' = \nabla f - \nabla f(y)$. Therefore, for any $\tau \in (0, 1]$, one has

$$0 \geqslant \langle \nabla f((1-\tau)\hat{x} + \tau y) - \nabla f(y), \hat{x} - y \rangle$$

and, letting here $\tau \to 0^+$, one obtains (1.2.17) because the function

$$\langle \nabla f(\cdot) - \nabla f(y), \hat{x} - y \rangle$$

is continuous due to the norm-to-weak* continuity of the gradient ∇f (see Proposition 1.1.21). Now, suppose that $\hat{x} \in K$ satisfies (1.2.17). Then, for any $y \in K$, one has again from the subdifferentiability inequality (see (1.1.5)) that

$$D_f(y, x) - D_f(\hat{x}, x) \geqslant \left\langle [D_f(\cdot, x)]'(\hat{x}), y - \hat{x} \right\rangle = \langle \nabla f(\hat{x}) - \nabla f(x), y - \hat{x} \rangle \geqslant 0,$$

showing that $\hat{x}$ minimizes $D_f(y, \cdot)$ over K, that is, $\hat{x} = \operatorname{proj}_K^f(x)$.

(ii) $\Leftrightarrow$ (iii) It is sufficient to observe from the three point identity (see (1.2.2)) that

$$D_f(\hat{x}, x) + D_f(y, \hat{x}) - D_f(y, x) = \langle \nabla f(\hat{x}) - \nabla f(x), \hat{x} - y \rangle$$

for any $y \in K$. □

Computing Bregman projections may not be an easy task. In the special case where $f = \|\cdot\|^p$, $p \in (1, +\infty)$, and X is a uniformly convex and smooth Banach space (see Definition 1.1.33(i) and (iv)), Butnariu, Iusem and Resmerita found an explicit formula for the Bregman projection onto hyperplane or half-space (*cf.* [36, Theorem 2, page 326]).

Proposition 1.2.36 (Bregman projection onto hyperplane)**.** *Let X be uniformly convex and smooth Banach space and let $f_p = \|\cdot\|^p$, $p \in (1, +\infty)$. Denote*

$$K = \{z \in X : \langle \xi, z \rangle = \alpha\},$$

where $\xi \in X \backslash \{0^*\}$ and $\alpha \in \mathbb{R}$. The following statements are true.

(i) *For any $x \in X$ the equation*

$$\left\langle \xi, J_p^* \left(\beta\xi + J_p(x)\right)\right\rangle = \alpha \tag{1.2.18}$$

has solutions β such that $\operatorname{sign}\beta = \operatorname{sign}(\alpha - \langle \xi, x \rangle)$.

(ii) *The Bregman projection* $\operatorname{proj}_K^f(x)$ *is given by*

$$\operatorname{proj}_K^f(x) = J_p^*\left(\beta\xi + J_p(x)\right) \tag{1.2.19}$$

with $\beta \in \mathbb{R}$ being a solution of the equation (1.2.18).

(iii) *Formula* (1.2.19) *remains true when K is the half-space $\{z \in X : \langle \xi, z \rangle \geqslant \alpha\}$ and $\beta \in \mathbb{R}$ is a nonnegative solution of* (1.2.18).

Remark 1.2.37. *As we already noted, when the Banach space X is a Hilbert space $\mathcal{H}$ and $f = \|\cdot\|^2$, then the Bregman distance is the metric distance squared (see Remark 1.2.5). Therefore in this setting the Bregman projection is exactly the metric projection. Here, for each $x \in \mathcal{H}$ and each nonempty, closed and convex subset K of $\mathcal{H}$, the metric projection $P_K(x)$ is defined as the unique point which satisfies*

$$\|x - P_K(x)\| = \inf\{\|x - y\| : y \in K\}.$$ ◇

The metric projection is characterized in the following way (*cf.* [11, Theorem 3.14, page 46]).

Proposition 1.2.38 (Characterization of the metric projection)**.** *Let K be a nonempty, closed and convex subset of $\mathcal{H}$. Given $x \in \mathcal{H}$ and $z \in K$, then $z = P_K(x)$ if and only if*

$$\langle x - z, y - z \rangle \leqslant 0, \quad \forall y \in K.$$

As a consequence we have the following properties of the metric projection (*cf.* [11, Proposition 4.8, page 61]).

Corollary 1.2.39 (Properties of the metric projection)**.** *Let K be a nonempty, closed and convex subset of $\mathcal{H}$. The following statements are true.*

(i) *For any $x, y \in \mathcal{H}$ we have*

$$\|P_K(x) - P_K(y)\|^2 \leqslant \langle x - y, P_K(x) - P_K(y)\rangle .$$

(ii) *For all $x \in \mathcal{H}$ and $y \in K$ we have*

$$\|x - P_K(x)\|^2 \leqslant \|x - y\|^2 - \|y - P_K(x)\|^2 .$$

(iii) *If K is a closed subspace, then P_K coincides with the orthogonal projection from $\mathcal{H}$ onto K, that is, for any $x \in \mathcal{H}$, the vector $x - P_K(x)$ is orthogonal to K (i.e., $\langle x - P_K(x), y\rangle = 0$ for each $y \in K$).*

Remark 1.2.40 (Special cases of the metric projection)**.** *Let K be a nonempty, closed and convex subset with a particulary simple structure. Then the projection P_K has a closed form expression as described below.*

(i) *If $K = \{x \in \mathcal{H} : \|x - u\| \leqslant r\}$ is a closed ball centered at $u \in \mathcal{H}$ with radius $r > 0$, then*

$$P_K(x) = \begin{cases} u + r\frac{(x-u)}{\|x-u\|}, & x \notin K \\ x, & x \in K. \end{cases} \tag{1.2.20}$$

(ii) *If $K = [\mathbf{a}, \mathbf{b}]$ is a closed rectangle in $\mathbb{R}^n$, where $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$, then, for any $1 \leqslant i \leqslant n$, $P_K(x)$ has the i^{th} coordinate given by*

$$(P_K(x))_i = \begin{cases} a_i, & x_i < a_i, \\ x_i, & x_i \in [a_i, b_i], \\ b_i, & x_i > b_i. \end{cases} \tag{1.2.21}$$

(iii) *If $K = \{y \in \mathcal{H} : \langle a, y\rangle = \alpha\}$ is a hyperplane, with $a \neq 0$ and $\alpha \in \mathbb{R}$, then*

$$P_K(x) = x - \frac{\langle a, x\rangle - \alpha}{\|a\|^2} a. \tag{1.2.22}$$

(iv) *If* $K = \{y \in \mathcal{H} : \langle a, y \rangle \leqslant \alpha\}$ *is a closed half-space, with* $a \neq 0$ *and* $\alpha \in \mathbb{R}$*, then*

$$P_K(x) = \begin{cases} x - \frac{\langle a,x \rangle - \alpha}{\|a\|^2} a, & \langle a, x \rangle > \alpha \\ x, & \langle a, x \rangle \leqslant \alpha. \end{cases} \tag{1.2.23}$$

(v) *If* K *is the range of an* $m \times n$ *matrix* A *with full column rank, then* $P_K(x) = A\left(A^* A\right)^{-1} A^* x$ *where* A^* *is the adjoint of* A. ◇

The following example appears in [12, Definition 3.1, page 66].

Example 1.2.41 (Metric projection - intersection of two half-spaces)**.** *In the Hilbert space setting, the orthogonal projection onto the intersection of two half-spaces*

$$T = \{x \in \mathcal{H} : \langle a_1, x \rangle \leqslant b_1, \langle a_2, x \rangle \leqslant b_2\} \quad (a_1, a_2 \in \mathcal{H}, b_1, b_2 \in \mathbb{R})$$

is given by the following explicit formula:

$$P_T(x) = \begin{cases} x, & \alpha \leqslant 0 \text{ and } \beta \leqslant 0, \\ x - (\beta/\nu)\, a_2, & \alpha \leqslant \pi(\beta/\nu) \text{ and } \beta > 0, \\ x - (\alpha/\mu)\, a_1, & \beta \leqslant \pi(\alpha/\mu) \text{ and } \alpha > 0, \\ x + (\alpha/\rho)(\pi a_2 - \nu a_1) + (\beta/\rho)(\pi a_1 - \mu a_2), & \text{otherwise}, \end{cases}$$

where here

$$\pi = \langle a_1, a_2 \rangle, \ \mu = \|a_1\|^2, \ \nu = \|a_2\|^2, \ \rho = \mu\nu - \pi^2, \ \alpha = \langle a_1, x \rangle - b_1 \text{ and } \beta = \langle a_2, x \rangle - b_2.$$

1.2.4 Properties of Bregman Distances

With an admissible function $f : X \to (-\infty, +\infty]$ (see Definition 1.2.1), we associate the bifunction $W^f : \operatorname{dom} f^* \times \operatorname{dom} f \to [0, +\infty]$ defined by

$$W^f(\xi, x) = f(x) - \langle \xi, x \rangle + f^*(\xi). \tag{1.2.24}$$

Now we list several properties of the bifunction W^f (*cf.* [74, Proposition 1, page 5]).

Proposition 1.2.42 (Properties of W^f)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function. The following assertions are true.*

(i) *The function* $W^f(\cdot, x)$ *is convex for any* $x \in \operatorname{dom} f$.

(ii) $W^f(\nabla f(x), y) = D_f(y, x)$ *for any* $x \in \operatorname{int}\operatorname{dom} f$ *and* $y \in \operatorname{dom} f$.

(iii) *For any* $\xi, \eta \in \operatorname{dom} f^*$ *and* $x \in \operatorname{dom} f$, *we have*

$$W^f(\xi, x) + \left\langle \eta, \left(\nabla f^*\right)(\xi) - x \right\rangle \leqslant W^f(\xi + \eta, x).$$

Proof. (i) This is clear since f^* is convex (see Remark 1.1.26(i)).

(ii) Let $x \in \operatorname{int} \operatorname{dom} f$ and let $y \in \operatorname{dom} f$. It is known from Proposition 1.1.29(v) that

$$f(x) + f^*(\nabla f(x)) = \langle \nabla f(x), x \rangle.$$

Therefore

$$\begin{aligned} W^f(\nabla f(x), y) &= f(y) - \langle \nabla f(x), y \rangle + f^*(\nabla f(x)) \\ &= f(y) - \langle \nabla f(x), y \rangle + [\langle \nabla f(x), x \rangle - f(x)] \\ &= f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\ &= D_f(y, x). \end{aligned}$$

(iii) Let $x \in \operatorname{dom} f$ be given. Define a function $g : X^* \to (-\infty, +\infty]$ by $g(\xi) = W^f(\xi, x)$. Then

$$\nabla g(\xi) = \nabla \left(f^* - \langle \cdot, x \rangle\right)(\xi) = \nabla f^*(\xi) - x.$$

Hence from the subdifferentiability inequality (see (1.1.5)) we get

$$g(\xi + \eta) - g(\xi) \geqslant \left\langle \eta, \nabla f^*(\xi) - x \right\rangle,$$

that is,

$$W^f(\xi, x) + \left\langle \eta, \nabla f^*(\xi) - x \right\rangle \leqslant W^f(\xi + \eta, x)$$

for all $\xi, \eta \in \operatorname{dom} f^*$. □

In order to prove several properties of Bregman distances we first prove simple observation of strictly convex functions which is essential for our later study (*cf.* [74, Lemma 6.1, page 14]).

Lemma 1.2.43 (Property of strictly convex functions)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a strictly convex function and let* $\{t_i\}_{i=1}^N \subset (0, 1)$ *which satisfy* $\sum_{i=1}^N t_i = 1$. *Let* $\{x_i\}_{i=1}^N$ *be a subset of* $\operatorname{int} \operatorname{dom} f$ *and assume that*

$$f\left(\sum_{i=1}^N t_i x_i\right) = \sum_{i=1}^N t_i f(x_i). \tag{1.2.25}$$

Then $x_1 = x_2 = \ldots = x_N$.

Proof. Assume, by way of contradiction, that $x_k \neq x_l$ for some $k, l \in \{1, 2, \ldots, N\}$. Then from the strict convexity of f we get

$$f\left(\frac{t_k}{t_k + t_l}x_k + \frac{t_l}{t_k + t_l}x_l\right) < \frac{t_k}{t_k + t_l}f(x_k) + \frac{t_l}{t_k + t_l}f(x_l).$$

Using this inequality, we obtain

$$\begin{aligned} f\left(\sum_{i=1}^{N} t_i x_i\right) &= f\left((t_k + t_l)\left(\frac{t_k}{t_k + t_l}x_k + \frac{t_l}{t_k + t_l}x_l\right) + \sum_{i \neq k,l} t_i x_i\right) \\ &\leqslant (t_k + t_l) f\left(\frac{t_k}{t_k + t_l}x_k + \frac{t_l}{t_k + t_l}x_l\right) + \sum_{i \neq k,l} t_i f(x_i) \\ &< (t_k + t_l)\left(\frac{t_k}{t_k + t_l}f(x_k) + \frac{t_l}{t_k + t_l}f(x_l)\right) + \sum_{i \neq k,l} t_i f(x_i) \\ &= \sum_{i=1}^{N} t_i f(x_i). \end{aligned}$$

This contradicts the assumption (1.2.25). □

Using the previous technical result we now prove the following lemma which concerns the Bregman distance (*cf.* [74, Lemma 6.2, page 15]).

Lemma 1.2.44 (Basic property of Bregman distances)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function and let $\{t_i\}_{i=1}^{N} \subset (0,1)$ which satisfy $\sum_{i=1}^{N} t_i = 1$. Let $z \in X$ and let $\{x_i\}_{i=1}^{N}$ be a finite subset in* $\operatorname{int}\operatorname{dom} f$ *such that*

$$D_f\left(z, \nabla f^*\left(\sum_{i=1}^{N} t_i \nabla f(x_i)\right)\right) = \sum_{i=1}^{N} t_i D_f(z, x_i). \tag{1.2.26}$$

Then $x_1 = x_2 = \ldots = x_N$.

Proof. Equality (1.2.26) can be reformulated as follows (see Proposition 1.2.42(ii))

$$D_f\left(z, \nabla f^*\left(\sum_{i=1}^{N} t_i \nabla f(x_i)\right)\right) = W^f\left(\sum_{i=1}^{N} t_i \nabla f(x_i), z\right) = \sum_{i=1}^{N} t_i D_f(z, x_i). \tag{1.2.27}$$

Now from the definition of W^f (see (1.2.24)) and the definition of the Bregman distance (see (1.2.1)) we get that the second equality in (1.2.27) can be written as

$$f(z) + f^*\left(\sum_{i=1}^{N} t_i \nabla f(x_i)\right) - \left\langle \sum_{i=1}^{N} t_i \nabla f(x_i), z \right\rangle = \sum_{i=1}^{N} t_i \left(f(z) - f(x_i) - \langle \nabla f(x_i), z - x_i \rangle\right).$$

Thus

$$f^*\left(\sum_{i=1}^{N} t_i \nabla f(x_i)\right) = \sum_{i=1}^{N} t_i \left(\langle \nabla f(x_i), x_i\rangle - f(x_i)\right).$$

Since $f(x_i) + f^*(\nabla f(x_i)) = \langle \nabla f(x_i), x_i\rangle$ for any $1 \leqslant i \leqslant N$ (see Proposition 1.1.29(v)), we obtain

$$f^*\left(\sum_{i=1}^{N} t_i \nabla f(x_i)\right) = \sum_{i=1}^{N} t_i f^*(\nabla f(x_i)).$$

Since f Legendre, f^* is strictly convex on $\operatorname{int}\operatorname{dom} f^*$ and from Lemma 1.2.43 it follows that $\nabla f(x_1) = \nabla f(x_2) = \ldots = \nabla f(x_N)$ and therefore $x_1 = x_2 = \ldots = x_N$, as claimed. □

The following proposition will be very useful for proving our main results. This result shows an important property of totally convex functions (*cf.* [96, Proposition 2.2, page 3]).

Proposition 1.2.45 (Convergence in the Bregman distance)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a convex function and take* $x \in \operatorname{dom} f$*. Then* f *is totally convex at* x *if and only if* $\lim_{n\to\infty} D_f(y_n, x) = 0$ *implies that* $\lim_{n\to\infty} \|y_n - x\| = 0$ *for any sequence* $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$*.*

Proof. Suppose that f is totally convex at x (see Definition 1.2.6). Take $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$ such that

$$\lim_{n\to\infty} D_f(y_n, x) = 0.$$

Since, by definition, $\nu_f(x, \|y_n - x\|) \leqslant D_f(y_n, x)$ for all $n \in \mathbb{N}$, it follows that

$$\lim_{n\to\infty} \nu_f(x, \|y_n - x\|) = 0. \tag{1.2.28}$$

Suppose, by way of contradiction, that there exist a positive number ε and a subsequence $\{y_{n_k}\}_{k\in\mathbb{N}}$ of $\{y_n\}_{n\in\mathbb{N}}$ such that $\|y_{n_k} - x\| \geqslant \varepsilon$ for all $k \in \mathbb{N}$. It was shown in Proposition 1.2.11(iv) that the function $\nu_f(x, \cdot)$ is strictly increasing whenever $x \in \operatorname{int}\operatorname{dom} f$. It is easy to see that this result is still valid when $x \in \operatorname{dom} f$. Consequently, we get that

$$\lim_{k\to\infty} \nu_f(x, \|y_{n_k} - x\|) > \nu_f(x, \varepsilon) > \nu_f(x, 0) = 0,$$

contradicting (1.2.28). Conversely, suppose that there exists $t_0 > 0$ such that $\nu_f(x, t_0) = 0$, that is, there exists a sequence $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$ such that $\|y_n - x\| = t_0$ and in addition $\lim_{n\to\infty} D_f(y_n, x) = 0$. Then $\lim_{n\to\infty} \|y_n - x\| = 0$ yields $t_0 = 0$, a contradiction. Therefore, the function f is totally convex at $x \in \operatorname{dom} f$. □

The following result (*cf.* [35, Lemma 2.1.2, page 67]) shows a strong connection between the two concepts of sequential consistency (see Definition 1.2.20) and of total convexity on bounded subsets (see Definition 1.2.10).

Proposition 1.2.46 (Characterization of sequential consistency)**.** *A function* $f : X \to (-\infty, +\infty]$ *is totally convex on bounded subsets if and only if it is sequentially consistent.*

Proof. Suppose that f is totally convex on bounded subsets (see Definition 1.2.10) and suppose, by way of contradiction, that there exists two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ in $\operatorname{dom} f$ and $\operatorname{int}\operatorname{dom} f$, respectively, such that $\{x_n\}_{n\in\mathbb{N}}$ is bounded and $\lim_{n\to\infty} D_f(y_n, x_n) = 0$ but $\{\|y_n - x_n\|\}_{n\in\mathbb{N}}$ does not converge to zero. Then, there exists a positive number α and subsequences $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_{n_k}\}_{k\in\mathbb{N}}$ of $\{y_n\}_{n\in\mathbb{N}}$, such that $\alpha \leqslant \|y_{n_k} - x_{n_k}\|$ for all $n \in \mathbb{N}$. The set E of all x_k's is bounded since $\{x_n\}_{n\in\mathbb{N}}$ is bounded. Therefore, for all $n \in \mathbb{N}$, we have from Proposition 1.2.11(iv) that

$$D_f(y_{n_k}, x_{n_k}) \geqslant \nu_f(x_{n_k}, \|y_{n_k} - x_{n_k}\|) \geqslant \nu_f(x_{n_k}, \alpha) \geqslant \inf_{x\in E} \nu_f(x, \alpha),$$

which implies that $\inf_{x\in E} \nu_f(x, \alpha) = 0$ and, thus, contradicts our assumption.

Conversely, suppose, by way of contradiction, that there exists a nonempty and bounded subset E of $\operatorname{dom} f$ such that $\inf_{x\in E} \nu_f(x, t) = 0$ for some positive real number t. Then there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ contained in E such that, for each positive integer n, we have

$$\frac{1}{n} > \nu_f(x_n, t) = \inf\{D_f(y, x_n) : \|y - x_n\| = t\}.$$

Therefore, there exists a sequence $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$ such that, for each integer $n \geqslant 1$, one has $\|y_n - x_n\| = t$ and $D_f(y_n, x_n) < 1/n$. The sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded because it is contained in E. Also, we have that $\lim_{n\to\infty} D_f(y_n, x_n) = 0$. Hence,

$$0 < t = \lim_{n\to\infty} \|y_n - x_n\| = 0$$

and this is a contradiction. □

Now we prove several technical results which will be very useful in the proofs of our main results (*cf.* [90, Lemma 3.1, page 31] and [74, Proposition 10, page 10]).

Proposition 1.2.47 (Boundedness property - left variable)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre and totally convex function. Let* $x \in \operatorname{int}\operatorname{dom} f$ *and* $\{x_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$*. If* $\{D_f(x_n, x)\}_{n\in\mathbb{N}}$ *is bounded, then the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *is bounded too.*

Proof. Since the sequence $\{D_f(x_n, x)\}_{n\in\mathbb{N}}$ is bounded then there exists $M > 0$ such that $D_f(x_n, x) < M$ for any $n \in \mathbb{N}$. Therefore, from (1.2.6), the sequence $\{\nu_f(x, \|x_n - x\|)\}_{n\in\mathbb{N}}$ is bounded by M since

$$\nu_f(x, \|x_n - x\|) \leqslant D_f(x_n, x) \leqslant M. \tag{1.2.29}$$

The function f is totally convex (see Definition 1.2.8), therefore from Proposition 1.2.11(iv) the function $\nu_f(x, \cdot)$ is strictly increasing and positive on $(0, \infty)$. This implies, in particular, that $\nu_f(x, 1) > 0$ for all $x \in X$. Now suppose, by way of contradiction, that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is not bounded. Then it contains a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \|x_{n_k}\| = +\infty.$$

Consequently, $\lim_{k\to\infty} \|x_{n_k} - x\| = +\infty$. This shows that $\{\nu_f(x, \|x_n - x\|)\}_{n\in\mathbb{N}}$ is not bounded. Indeed, there exists some $k_0 > 0$ such that $\|x_{n_k} - x\| > 1$ for all $k > k_0$ and then, from Proposition 1.2.11(ii), we see

$$\lim_{k\to\infty} \nu_f(x, \|x_{n_k} - x\|) \geqslant \lim_{k\to\infty} \|x_{n_k} - x\| \, \nu_f(x, 1) = +\infty,$$

because, as noted above, $\nu_f(x, 1) > 0$. This contradicts (1.2.29). Hence the sequence $\{x_n\}_{n\in\mathbb{N}}$ is indeed bounded, as claimed. □

Proposition 1.2.48 (Boundedness property - right variable)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function such that* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let* $x \in \operatorname{dom} f$ *and* $\{x_n\}_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$*. If* $\{D_f(x, x_n)\}_{n\in\mathbb{N}}$ *is bounded, so is the sequence* $\{x_n\}_{n\in\mathbb{N}}$*.*

Proof. Let β be an upper bound of the sequence $\{D_f(x, x_n)\}_{n\in\mathbb{N}}$. Then from the definition of W^f (see (1.2.24)) and Proposition 1.2.42(ii) we obtain

$$f(x) - \langle \nabla f(x_n), x\rangle + f^*(\nabla f(x_n)) = W^f(\nabla f(x_n), x) = D_f(x, x_n) \leqslant \beta.$$

This implies that $\{\nabla f(x_n)\}_{n\in\mathbb{N}}$ is contained in the sub-level set, $\operatorname{lev}^{\psi}_{\leqslant}(\beta - f(x))$, of the function $\psi := f^* - \langle \cdot, x\rangle$. Since the function f^* is proper and lower semicontinuous (see Remark 1.1.26), an application of the Moreau-Rockafellar Theorem (see [7, Fact 3.1, page 623]) shows that ψ is super-coercive (see Definition 1.2.33(ii)). Consequently, all sub-level sets of ψ are bounded. Indeed, if this is not the case then there is a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ in $\operatorname{lev}^{\psi}_{\leqslant}(\alpha)$ such that $\|\xi_n\| \to \infty$ as $n \to \infty$. Then we have that

$$\frac{\psi(\xi_n)}{\|\xi_n\|} \leqslant \frac{\alpha}{\|\xi_n\|}.$$

This, since ψ is super-coercive, implies that the left-hand side converges to ∞ as $n \to \infty$, which is a contradiction since the right-hand side converges to zero.

Hence the sequence $\{\nabla f(x_n)\}_{n\in\mathbb{N}}$ is bounded. By hypothesis, ∇f^* is bounded on bounded subsets of $\operatorname{int}\operatorname{dom} f^*$. Therefore the sequence $\left\{x_n = \nabla f^*(\nabla f(x_n))\right\}_{n\in\mathbb{N}}$ is bounded too, as claimed. □

Remark 1.2.49. *The previous result can be also proved by combining known results. More precisely, according to [7, Theorem 3.3, page 624], f is super-coercive (see Definition 1.2.33(ii)) because* $\operatorname{dom} \nabla f^* = X^*$ *and* ∇f^* *is bounded on bounded subsets of* X^**. From [7, Lemma 7.3(viii), page 642] it follows that* $D_f(x, \cdot)$ *is coercive (see Definition 1.2.33(i)). If the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *were unbounded, then there would exist a subsequence* $\{x_{n_k}\}_{k\in\mathbb{N}}$ *with* $\|x_{n_k}\| \to \infty$ *as* $k \to \infty$*. This, since* $D_f(x, \cdot)$ *is coercive, implies that* $D_f(x, x_{n_k}) \to \infty$ *as* $k \to \infty$*, which is a contradiction. Thus* $\{x_n\}_{n\in\mathbb{N}}$ *is indeed bounded, as claimed.* ◇

Proposition 1.2.50 (Property of Bregman distances)**.** *Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X*. Suppose that a sequence* $\{x_n\}_{n\in\mathbb{N}}$ *is bounded and* $\{e_n\}_{n\in\mathbb{N}}$ *is a sequence which satisfies* $\lim_{n\to\infty} \|e_n\| = 0$*. If*

$$\lim_{n\to\infty} D_f(x_{n+1}, x_n) = 0 \tag{1.2.30}$$

then

$$\lim_{n\to\infty} D_f(x_{n+1}, x_n + e_n) = 0.$$

Proof. From Proposition 1.2.46, (1.2.30) and the fact that $\{x_n\}_{n\in\mathbb{N}}$ is bounded, we have

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \tag{1.2.31}$$

It follows from the definition of the Bregman distance (see (1.2.1)) that

$$\begin{aligned} D_f(x_n, x_n + e_n) &= f(x_n) - f(x_n + e_n) - \langle \nabla f(x_n + e_n), x_n - (x_n + e_n) \rangle \\ &= f(x_n) - f(x_n + e_n) + \langle \nabla f(x_n + e_n), e_n \rangle. \end{aligned}$$

The function f is bounded on bounded subsets of X and therefore ∇f is also bounded on bounded subsets of X (see Proposition 1.1.15). In addition, f is uniformly Fréchet differentiable on bounded subsets of X and therefore f is uniformly continuous on bounded subsets of X (see Proposition 1.1.22(i)). Hence, since $\lim_{n\to\infty} \|e_n\| = 0$, we have

$$\lim_{n\to\infty} D_f(x_n, x_n + e_n) = 0. \tag{1.2.32}$$

The three point identity (see (1.2.2)) implies that

$$\begin{aligned} D_f(x_{n+1}, x_n + e_n) = {} & D_f(x_{n+1}, x_n) + D_f(x_n, x_n + e_n) \\ & + \langle \nabla f(x_n) - \nabla f(x_n + e_n), x_{n+1} - x_n \rangle. \end{aligned}$$

Since ∇f is bounded on bounded subsets of X, we get from (1.2.30), (1.2.31) and (1.2.32)

$$\lim_{n\to\infty} D_f\left(x_{n+1}, x_n + e_n\right) = 0,$$

as required. □

The following result is frequently used in this thesis because of the usage of the total convexity property to prove strong convergence when weak convergence is already known (*cf.* [90, Lemma 3.2, page 31]).

Proposition 1.2.51 (Strong converges result)**.** *Let $f : X \to (-\infty, +\infty]$ be an admissible and totally convex function, $x_0 \in X$ and let K be a nonempty, closed and convex subset of* $\operatorname{dom} f$*. Suppose that a sequence $\{x_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} f$ is bounded and any weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to K. If $D_f\left(x_n, x_0\right) \leqslant D_f\left(\operatorname{proj}_K^f\left(x_0\right), x_0\right)$ for any $n \in \mathbb{N}$, then $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\operatorname{proj}_K^f\left(x_0\right)$*.*

Proof. Denote $\operatorname{proj}_K^f\left(x_0\right) = \tilde{u}$. The three point identity (see (1.2.2)) and the assumption that $D_f\left(x_n, x_0\right) \leqslant D_f\left(\tilde{u}, x_0\right)$ yields

$$\begin{aligned} D_f\left(x_n, \tilde{u}\right) &= D_f\left(x_n, x_0\right) + D_f\left(x_0, \tilde{u}\right) - \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), x_n - x_0 \right\rangle \\ &\leqslant D_f\left(\tilde{u}, x_0\right) + D_f\left(x_0, \tilde{u}\right) - \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), x_n - x_0 \right\rangle \\ &= \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), \tilde{u} - x_0 \right\rangle - \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), x_n - x_0 \right\rangle \\ &= \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), \tilde{u} - x_n \right\rangle. \end{aligned} \tag{1.2.33}$$

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded, it has weakly convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and denote its weak limit by v. We know that $v \in K$. It follows from (1.2.33) and Proposition 1.2.35(ii) that

$$\limsup_{k\to\infty} D_f\left(x_{n_i}, \tilde{u}\right) \leqslant \limsup_{k\to\infty} \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), \tilde{u} - x_{n_k} \right\rangle = \left\langle \nabla f\left(\tilde{u}\right) - \nabla f\left(x_0\right), \tilde{u} - v \right\rangle \leqslant 0.$$

Hence

$$\lim_{k\to\infty} D_f\left(x_{n_k}, \tilde{u}\right) = 0.$$

Since f is totally convex (see Remark 1.2.9), Proposition 1.2.45 now implies that $x_{n_k} \to \tilde{u}$ as $k \to \infty$. It follows that the whole sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\tilde{u} = \operatorname{proj}_K^f\left(x_0\right)$, as claimed. □

Definition 1.2.52 (Weakly sequentially continuous)**.** *A mapping $A : X \to X^*$ is called weakly sequentially continuous if $x_n \rightharpoonup x$ implies that $Ax_n \rightharpoonup Ax$.*

Using this definition for the gradient ∇f leads to the following result (*cf.* [74, Proposition 9, page 10]).

Proposition 1.2.53 (Weak convergence result)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function such that ∇f is weakly sequentially continuous. Suppose that a sequence $\{x_n\}_{n\in\mathbb{N}}$ in* $\operatorname{int}\operatorname{dom} f$ *is bounded and*

$$\lim_{n\to\infty} D_f(u, x_n) \tag{1.2.34}$$

exists for any weak subsequential limit u of $\{x_n\}_{n\in\mathbb{N}}$. Then $\{x_n\}_{n\in\mathbb{N}}$ converges weakly.

Proof. It is suffices to show that there is exactly one weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$. Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded and X is reflexive, there is at least one weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$. Assume that u and v are two weak subsequential limits of $\{x_n\}_{n\in\mathbb{N}}$. From (1.2.34) we have that the limit

$$\lim_{n\to\infty} \left(D_f(u, x_n) - D_f(v, x_n)\right)$$

exists. From the definition of the Bregman distance (see (1.2.1)) we get that

$$\begin{aligned} D_f(u, x_n) - D_f(v, x_n) &= [f(u) - f(x_n) - \langle \nabla f(x_n), u - x_n\rangle] \\ &\quad - [f(v) - f(x_n) - \langle \nabla f(x_n), v - x_n\rangle] \\ &= f(u) - f(v) + \langle \nabla f(x_n), v - u\rangle \end{aligned}$$

and therefore

$$\lim_{n\to\infty} \langle \nabla f(x_n), v - u\rangle$$

exists. Since u and v are weak subsequential limits of $\{x_n\}_{n\in\mathbb{N}}$, there are subsequences $\{x_{n_k}\}_{k\in\mathbb{N}}$ and $\{x_{n_m}\}_{m\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup u$ and $x_{n_m} \rightharpoonup v$ as $k \to \infty$ and $m \to \infty$, respectively. Since ∇f is weakly sequentially continuous, we have that $\nabla f(x_{n_k}) \rightharpoonup \nabla f(u)$ and $\nabla f(x_{n_m}) \rightharpoonup \nabla f(v)$ as $k \to \infty$ and $m \to \infty$, respectively. Then we have

$$\begin{aligned} \langle \nabla f(u), v - u\rangle &= \lim_{k\to\infty} \langle \nabla f(x_{n_k}), v - u\rangle = \lim_{n\to\infty} \langle \nabla f(x_n), v - u\rangle \\ &= \lim_{m\to\infty} \langle \nabla f(x_{n_m}), v - u\rangle = \langle \nabla f(v), v - u\rangle. \end{aligned}$$

Thus $\langle \nabla f(v) - \nabla f(u), v - u\rangle = 0$ implies that $u = v$ since f is strictly convex because f is Legendre. □

The following result will play a key tool in the proof of several results in this thesis.

Proposition 1.2.54 (Closed and convex half-space)**.** *Let $f : X \to \mathbb{R}$ be an admissible function. Let $u, v \in X$. Then the set*

$$K = \{z \in X : D_f(z, u) \leqslant D_f(z, v)\}$$

is a closed and convex half-space.

Proof. If K is empty then the result is obvious. Now assume that K is nonempty. Directly from the definition of the Bregman distance (see (1.2.1)) we can write the set K in the following way:

$$K = \{z \in X : \langle \nabla f(v) - \nabla f(u), z \rangle \leqslant \langle \nabla f(u), u \rangle - \langle \nabla f(v), v \rangle + f(u) - f(v)\}.$$

This of course proves that K is a half-space. We first show that K is closed. Let $\{z_n\}_{n \in \mathbb{N}}$ be a sequence in K which converges strongly to z. From the definition of the Bregman distance (see (1.2.1)), for any $n \in \mathbb{N}$, we have that

$$f(z_n) - f(u) - \langle \nabla f(u), z_n - u \rangle = D_f(z_n, u) \leqslant D_f(z_n, v) = f(z_n) - f(v) - \langle \nabla f(v), z_n - v \rangle,$$

that is,

$$f(v) - f(u) \leqslant \langle \nabla f(u), z_n - u \rangle - \langle \nabla f(v), z_n - v \rangle.$$

Letting $n \to \infty$, we get that

$$f(v) - f(u) \leqslant \langle \nabla f(u), z - u \rangle - \langle \nabla f(v), z - v \rangle,$$

that is,

$$f(z) - f(u) - \langle \nabla f(u), z - u \rangle = D_f(z, u) \leqslant D_f(z, v) = f(z) - f(v) - \langle \nabla f(v), z - v \rangle,$$

which means that $z \in K$, this proves that K is closed. Now we show that K is convex. Let $z_1, z_2 \in K$ and $t \in (0, 1)$. Denote $z_t = tz_1 + (1 - t) z_2$. Then

$$f(v) - f(u) \leqslant \langle \nabla f(u), z_1 - u \rangle - \langle \nabla f(v), z_1 - v \rangle$$

and

$$f(v) - f(u) \leqslant \langle \nabla f(u), z_2 - u \rangle - \langle \nabla f(v), z_2 - v \rangle.$$

If we multiply the first inequality by t and the second by $(1 - t)$ and summing up, then we get

$$f(v) - f(u) \leqslant \langle \nabla f(u), z_t - u \rangle - \langle \nabla f(v), z_t - v \rangle,$$

that is,

$$f(z_t) - f(u) - \langle \nabla f(u), z_t - u \rangle = D_f(z_t, u) \leqslant D_f(z_t, v) = f(z_t) - f(v) - \langle \nabla f(v), z_t - v \rangle$$

which means that $z_t \in K$. This proves that K is convex. $\square$

1.3 Operators

In this section we introduce and study several classes of nonexpansive operators. The theory of nonexpansive operators in Banach spaces is a recent branch of nonlinear functional analysis. It has flourished during the last hundred years with many papers, results, and still many unsolved problems. The simplest and perhaps the most useful result in Fixed Point Theory is the Banach fixed point theorem from 1922. The theorem holds for any complete metric space, in particular for Banach spaces.

Let K be a nonempty and convex subset of a Banach space X. An operator $T : K \to K$ is said to be *nonexpansive* (or 1-Lipschitz) if

$$\|Tx - Ty\| \leqslant \|x - y\| \tag{1.3.1}$$

for all $x, y \in K$. The operator T is called a *strict contraction* if its Lipschitz constant smaller than 1. The *Banach fixed point theorem* is the following result (*cf.* [56, Theorem 1.1, page 2]).

Theorem 1.3.1 (Banach's fixed point theorem)**.** *Let K be a nonempty, closed, and convex subset of a Banach space X. If $T : K \to K$ is a strict contraction, then it has a unique fixed point p and $\lim_{n\to\infty} T^n x = p$ for all $x \in K$.*

Remark 1.3.2 (Nonexpansive operator without fixed point)**.** *Theorem 1.3.1 requires the Lipschitz constant, L, of T to satisfy $L < 1$. If $L = 1$, i.e., T is nonexpansive (see* (1.3.1)*), then T need not have a fixed point as the example $T(x) = x + 1$, $x \in \mathbb{R}$, shows.* ◇

Definition 1.3.3 (Fixed point property)**.** *We say that a closed and convex subset K of X has the fixed point property for nonexpansive operators if every nonexpansive $T : K \to K$ has a fixed point.*

Browder [28] proved in 1965 that if X is uniformly convex Banach space (see Definition 1.1.33(iv)) and K is closed, convex and bounded, then K has the fixed point property. Notice that uniqueness may not hold as the example $T(x) = x$, $x \in K = [0, 1]$, shows.

It turns out that nonexpansive fixed point theory in Hilbert spaces can be applied to the solution of diverse problems such as finding zeroes of monotone mappings and solutions to certain evolution equations, as well as solving convex feasibility (CFP), variational inequality (VIP) and equilibrium problems (EP) (these problems will be studied in full detail in the following chapters of this dissertation). In some cases it is enough to assume that an operator $T : K \to K$ is *quasi-nonexpansive*, that is,

$$\|p - Tx\| \leqslant \|p - x\| \tag{1.3.2}$$

for all $p \in \operatorname{Fix}(T)$ and $x \in K$, where $\operatorname{Fix}(T)$ stands for the (nonempty) *fixed point set* of T.

There are many papers that deal with methods for finding fixed points of nonexpansive and quasi-nonexpansive operators in Hilbert spaces. Another class of operators which is very useful in Fixed Point Theory is the class of firmly nonexpansive operators. Recall that an operator $T : K \to K$ is called *firmly nonexpansive* if

$$\|Tx - Ty\|^2 \leqslant \langle Tx - Ty, x - y\rangle \tag{1.3.3}$$

for all $x, y \in K$.

Remark 1.3.4. *It is clear that the following implications hold:*

firmly nonexpansive $\Longrightarrow$ *nonexpansive* $\Longrightarrow$ *quasi-nonexpansive,*

where the second implication is true only if $\operatorname{Fix}(T) \neq \varnothing$. ◇

When we try to extend this theory to Banach spaces we encounter some difficulties because many of the useful examples of nonexpansive operators in Hilbert space are no longer firmly nonexpansive or even nonexpansive in Banach spaces. For example, the classical resolvent $R_A = (I + A)^{-1}$ of a maximal monotone mapping $A : \mathcal{H} \to 2^{\mathcal{H}}$ and the metric projection P_K onto a nonempty, closed and convex subset K of $\mathcal{H}$ (for more details see the relevant chapters). There are several ways to overcome these difficulties. The way we choose in this thesis is to use Bregman distances (see (1.2.1)) with respect to convex functions instead of with respect to the norm. Then the definitions of nonexpansive, quasi-nonexpansive and firmly nonexpansive will be defined with respect to the Bregman distance instead of with respect to the norm.

These definitions are useful in the setting of Banach spaces since we have several examples of operators which satisfy them, for example, the Bregman projection and the f-resolvent (see (1.2.14) and (0.0.2), respectively). In addition, if we go back to Hilbert space and take these new definitions with respect to the function $f = (1/2)\|\cdot\|^2$, then they coincide with the usual definitions.

A naive way to define nonexpansive operator with respect to the Bregman distance is by the following inequality

$$D_f(Tx, Ty) \leqslant D_f(x, y)$$

for any $x, y \in K \subset \operatorname{int}\operatorname{dom} f$.

But it turns out that this notion of nonexpansive operators with respect to Bregman distances encounters several difficulties. This generalization does not satisfy any of the properties that the classical nonexpansive operators do (for instance, the Bregman projection is not necessarily Bregman nonexpansive). Therefore it seems that the well-defined notions with respect to the Bregman distance are firmly, strongly and quasi-nonexpansive.

1.3.1 Bregman Nonexpansive Operators

We fix an admissible function f (see Definition 1.2.1) and let K and S be two nonempty subsets of $\operatorname{int}\operatorname{dom} f$. We next list significant types of nonexpansivity with respect to the Bregman distance (see 1.2.1).

Definition 1.3.5 (Bregman nonexpansivity). *Let K and S be two nonempty subsets of* $\operatorname{int}\operatorname{dom} f$. *We say that an operator $T : K \subset \operatorname{int}\operatorname{dom} f \to \operatorname{int}\operatorname{dom} f$ is:*

(i) *Bregman firmly nonexpansive (BFNE) if*

$$\langle \nabla f(Tx) - \nabla f(Ty), Tx - Ty \rangle \leqslant \langle \nabla f(x) - \nabla f(y), Tx - Ty \rangle \tag{1.3.4}$$

for any $x, y \in K$, or equivalently,

$$D_f(Tx, Ty) + D_f(Ty, Tx) + D_f(Tx, x) + D_f(Ty, y) \leqslant D_f(Tx, y) + D_f(Ty, x). \tag{1.3.5}$$

(ii) *Quasi-Bregman firmly nonexpansive (QBFNE) with respect to S if*

$$0 \leqslant \langle \nabla f(x) - \nabla f(Tx), Tx - p \rangle \; \forall\, x \in K,\, p \in S, \tag{1.3.6}$$

or equivalently,

$$D_f(p, Tx) + D_f(Tx, x) \leqslant D_f(p, x). \tag{1.3.7}$$

(iii) *Quasi-Bregman nonexpansive (QBNE) with respect to S if*

$$D_f(p, Tx) \leqslant D_f(p, x), \; \forall\, x \in K,\, p \in S. \tag{1.3.8}$$

The class of Bregman firmly nonexpansive operators was introduced first by Bauschke, Borwein and Combettes in [8] (they call those operators D_f-firmly nonexpansive).

The natural option for the set S in Definition 1.3.5 is the fixed point set of the operator. Another option that seems to be important in applications is the asymptotic fixed point set defined first by Reich in [88].

Definition 1.3.6 (Asymptotic fixed point). *A point $u \in K$ is said to be an asymptotic fixed point of $T : K \to K$ if there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in K such that $x_n \rightharpoonup u$ and $\|x_n - Tx_n\| \to 0$ as $n \to \infty$. We denote the asymptotic fixed point set of T by $\widehat{\operatorname{Fix}}(T)$.*

Remark 1.3.7 (Types of quasi-Bregman nonexpansivity). *We will use the following particular cases.*

(i) *An operator which satisfies* (1.3.7) *(or* (1.3.8)*) with respect to $S := \operatorname{Fix}(T)$ is called properly QBFNE (or properly QBNE).*

(ii) *An operator which satisfies* (1.3.7) *(or* (1.3.8)*) with respect to $S := \widehat{\operatorname{Fix}}(T)$ is called strictly QBFNE (or strictly QBNE).*

(iii) *An operator which satisfies* (1.3.7) *(or* (1.3.8)*) with respect to $S := \operatorname{Fix}(T) = \widehat{\operatorname{Fix}}(T)$ is called QBFNE (or QBNE).*

◇

Another class of operators which was introduced in [46, 88] is the class of Bregman strongly nonexpansive operators.

Definition 1.3.8 (Bregman strongly nonexpansive)**.** *We say that an operator* $T : K \subset \operatorname{int}\operatorname{dom} f \to \operatorname{int}\operatorname{dom} f$ *is Bregman strongly nonexpansive (BSNE) with respect to* $S \subset \operatorname{dom} f$ *if*

$$D_f(p, Tx) \leqslant D_f(p, x) \tag{1.3.9}$$

for all $p \in S$ *and* $x \in K$*, and if whenever* $\{x_n\}_{n\in\mathbb{N}} \subset K$ *is bounded,* $p \in S$*, and*

$$\lim_{n\to\infty} \left(D_f(p, x_n) - D_f(p, Tx_n)\right) = 0, \tag{1.3.10}$$

it follows that

$$\lim_{n\to\infty} D_f(Tx_n, x_n) = 0. \tag{1.3.11}$$

Remark 1.3.9 (Types of Bregman strong nonexpansivity)**.** *We will use the following particular cases.*

(i) *An operator which satisfies* (1.3.9)–(1.3.11) *with respect to* $S := \operatorname{Fix}(T)$ *is called properly BSNE.*

(ii) *An operator which satisfies* (1.3.9)–(1.3.11) *with respect to* $S := \widehat{\operatorname{Fix}}(T)$ *is called strictly BSNE (this class of operators was first defined in [88]).*

(iii) *An operator which satisfies* (1.3.9)–(1.3.11) *with respect to* $S := \operatorname{Fix}(T) = \widehat{\operatorname{Fix}}(T)$ *is called BSNE.* ◇

The relations among all these classes of Bregman nonexpansive operators are summarized in the following scheme.

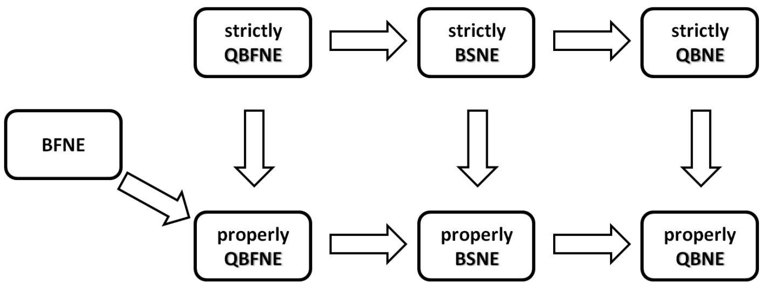

Figure 1.3: Implications between types of Bregman nonexpansivity

Remark 1.3.10 (Particular cases - nonexpansivity)**.** *Assume that* $f = \|\cdot\|^2$ *and the space* X *is a Hilbert space* $\mathcal{H}$*. In this case we have that* $\nabla f = 2I$ *(where* I *is the identity* *operator) and* $D_f(y,x) = \|x-y\|^2$ *(see Remark 1.2.5). Thence, Definition 1.3.5(i)-(iii) with* $S = \mathrm{Fix}(T)$ *implies the known classes of nonexpansive operators.* ◇

Definition 1.3.11 (Nonexpansivity)**.** *Let* K *be a subset of* $\mathcal{H}$*. We say that an operator* $T : K \to \mathcal{H}$ *is:*

(i') firmly nonexpansive (FNE) if

$$\|Tx - Ty\|^2 \leqslant \langle x - y, Tx - Ty \rangle, \; \forall \, x, y \in \mathcal{H}; \tag{1.3.12}$$

(ii') quasi-firmly nonexpansive (QFNE) if

$$\|Tx - p\|^2 + \|Tx - x\|^2 \leqslant \|x - p\|^2, \tag{1.3.13}$$

for any $x \in \mathcal{H}$ *and* $p \in \mathrm{Fix}(T)$*, or equivalently,*

$$0 \leqslant \langle x - Tx, Tx - p \rangle; \tag{1.3.14}$$

(iii') quasi-nonexpansive (QNE) if

$$\|Tx - p\| \leqslant \|x - p\|, \; \forall \, x \in \mathcal{H}, \, p \in \mathrm{Fix}(T). \tag{1.3.15}$$

The analog of Definition 1.3.8 for the particular case when $f = \|\cdot\|^2$ and the space X is a Hilbert space $\mathcal{H}$ is presented in the following definition. This latter class of operators was first studied in [32].

Definition 1.3.12 (Strong nonexpansivity)**.** *Let* K *be a subset of* $\mathcal{H}$*. We say that an operator* $T : K \to \mathcal{H}$ *is strongly nonexpansive (SNE) if it nonexpansive and for any two bounded sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{y_n\}_{n\in\mathbb{N}}$ *satisfying*

$$\lim_{n\to\infty} \left(\|x_n - y_n\| - \|Tx_n - Ty_n\|\right) = 0, \tag{1.3.16}$$

it follows that

$$\lim_{n\to\infty} \left((x_n - y_n) - (Tx_n - Ty_n)\right) = 0. \tag{1.3.17}$$

Since the norm variant does not follow from the Bregman case as do the other classes we emphasize the connection between the two classes of Bregman strongly nonexpansive and strongly nonexpansive.

Remark 1.3.13 (Connection between BSNE and SNE operators)**.** *Let* K *be a subset of* $\mathcal{H}$*. When* $f = \|\cdot\|^2$ *and* $S = \mathrm{Fix}(T)$*, Definition 1.3.12 means that* $T : K \to \mathcal{H}$ *is SNE with respect to* Fix (T) *if* T *is QNE (see* (1.3.15)*) and if for any bounded sequence* $\{x_n\}_{n\in\mathbb{N}}$

satisfying

$$\lim_{n\to\infty}\left(\|x_n - p\|^2 - \|Tx_n - p\|^2\right) = 0 \tag{1.3.18}$$

for all $p \in \operatorname{Fix}(T)$*, it follows that*

$$\lim_{n\to\infty}\|x_n - Tx_n\| = 0. \tag{1.3.19}$$

One is able to show that, in this case, strong nonexpansivity implies properly Bregman strong nonexpansivity. Indeed, if T *is SNE, the quasi-nonexpansivity is guaranteed by definition. Now, given a bounded sequence* $\{x_n\}_{n\in\mathbb{N}}$ *satisfying* (1.3.18) *for some* $p \in \operatorname{Fix}(T)$*, we have*

$$\lim_{n\to\infty}\left(\|x_n - p\| - \|Tx_n - p\|\right) = 0. \tag{1.3.20}$$

By taking in Definition 1.3.12 the sequence $\{y_n\}_{n\in\mathbb{N}}$ *to be the constant sequence defined by* $y_n = p$ *for all* $n \in \mathbb{N}$*, we see that* (1.3.19) *follows from* (1.3.17)*, so* T *is properly BSNE, as claimed. The converse does not hold in general, mainly because nonexpansivity is required.*

Note that if $S = \widehat{\operatorname{Fix}}(T)$*, the previous implication is no longer true. However, in the finite dimensional case,* $\mathcal{H} = \mathbb{R}^n$*, if* T *is continuous, then* $\operatorname{Fix}(T) = \widehat{\operatorname{Fix}}(T)$*. This happens, in particular, when* T *is SNE. Therefore, in finite dimensions, any SNE mapping (called paracontraction in [46]) is also strictly BSNE.*

To sum up, we can say that Bregman strong nonexpansivity turns out to be a generalization of strong nonexpansivity. ◇

Definition 1.3.14 (Asymptotically regular)**.** *An operator* $T : X \to X$ *is called asymptotically regular if for any* $x \in X$ *we have*

$$\lim_{n\to\infty}\left\|T^{n+1}x - T^n x\right\| = 0. \tag{1.3.21}$$

1.4 Monotone Mappings

Definition 1.4.1 (Notions of mappings)**.** *Let* $A : X \to 2^{X^*}$ *be a mapping.*

(i) *The domain of* A *is the set*

$$\operatorname{dom} A = \{x \in X : Ax \neq \varnothing\}.$$

(ii) *The range of* A *is the set*

$$\operatorname{ran} A = \{\xi \in Ax : x \in \operatorname{dom} A\}.$$

(iii) *The graph of A is the subset of $X \times X^*$ defined by*

$$\operatorname{graph} A = \left\{ (x, \xi) \in X \times X^* : \xi \in Ax \right\}.$$

Definition 1.4.2 (Monotone mapping)**.** *Let $A : X \to 2^{X^*}$ be a mapping.*

(i) *The mapping A is said to be monotone if for any $x, y \in \operatorname{dom} A$, we have*

$$\xi \in Ax \quad \text{and} \quad \eta \in Ay \quad \Longrightarrow \quad \langle \xi - \eta, x - y \rangle \geqslant 0. \tag{1.4.1}$$

(ii) *The mapping A is called strictly monotone if the inequality in* (1.4.1) *is strict whenever $x \neq y$.*

Example 1.4.3 (Monotonicity of the subdifferential mapping)**.** *Let $f : X \to (-\infty, +\infty]$ be a proper and convex function. The subdifferential mapping $\partial f : X \to 2^{X^*}$ (see Definition 1.1.12) is monotone since for any $x, y \in \operatorname{dom} \partial f$, any $\xi \in \partial f(x)$ and for any $\eta \in \partial f(y)$, we have from the subdifferential inequality (see* (1.1.5)*) that*

$$f(y) - f(x) \geqslant \langle \xi, y - x \rangle \quad \text{and} \quad f(x) - f(y) \geqslant \langle \eta, x - y \rangle.$$

Summing up these two inequalities we get that $\langle \xi - \eta, x - y \rangle \geqslant 0$ for any $x, y \in \operatorname{dom} \partial f$, that is, ∂f is a monotone mapping (see (1.4.1)*). If f is a strictly convex function then ∂f is strictly monotone. Indeed, if $\xi \in \partial f(x)$ then again from* (1.1.5) *we obtain*

$$f(x + t(y - x)) - f(x) \geqslant \langle \xi, x + t(y - x) - x \rangle \quad \forall t > 0, \quad y \in X.$$

Hence

$$f^{\circ}(x, y - x) = \lim_{t \searrow 0} \frac{f(x + t(y - x)) - f(x)}{t} \geqslant \langle \xi, y - x \rangle, \quad y \in X.$$

From Proposition 1.1.10(ii) we get that

$$\langle \xi, y - x \rangle \leqslant f^{\circ}(x, y - x) < f(y) - f(x).$$

In the same way, if $\eta \in \partial f(y)$, then $\langle \eta, x - y \rangle \leqslant f^{\circ}(y, x - y) < f(x) - f(y)$. Adding these two inequalities and we get that

$$\langle \xi - \eta, x - y \rangle > 0, \quad \forall x, y \in \operatorname{dom} \partial f.$$

Hence ∂f is a strictly monotone mapping (see Definition 1.4.2(ii)).

Example 1.4.4 (Monotonicity of an increasing one variable function). *Increasing function* $f : \mathbb{R} \to \mathbb{R}$ *determines a single-valued monotone mapping which is defined by* $Ax = \{f(x)\}$.

Definition 1.4.5 (Inverse mapping). *Let* $A : X \to 2^{X^*}$ *be mapping. The inverse mapping* $A^{-1} : X^* \to 2^X$ *is defined by*

$$A^{-1}\xi = \{x \in X : \xi \in Ax\}.$$

Remark 1.4.6 (Monotonicity of inverse mapping). *A mapping* $A : X \to 2^{X^*}$ *is monotone if and only if the inverse mapping* A^{-1} *is monotone.* ◇

Definition 1.4.7 (Demi-closed). *A mapping* $A : X \to 2^{X^*}$ *is called demi-closed at* $x \in \operatorname{dom} A$ *if*

$$\left.\begin{array}{c} x_n \longrightarrow x \\ \xi_n \in Ax_n \\ \xi_n \rightharpoonup \xi \end{array}\right\} \implies \xi \in Ax. \tag{1.4.2}$$

Definition 1.4.8 (Hemicontinuous). *The mapping* $A : X \to X^*$ *is called hemicontinuous if for any* $x \in \operatorname{dom} A$ *we have*

$$\left.\begin{array}{c} x + t_n y \in \operatorname{dom} A,\ y \in X \\ \lim_{n\to\infty} t_n = 0^+ \end{array}\right\} \implies A(x + t_n y) \rightharpoonup Ax. \tag{1.4.3}$$

Definition 1.4.9 (Maximal monotone). *A mapping* $A : X \to 2^{X^*}$ *is called maximal monotone if it is monotone and there does not exist a monotone mapping* $B : X \to 2^{X^*}$ *such that* $\operatorname{graph} A \subsetneqq \operatorname{graph} B$.

Remark 1.4.10 (Maximal monotonicity of the inverse mapping). *Note that a mapping* $A : X \to 2^{X^*}$ *is maximal monotone if and only if* A^{-1} *is maximal monotone mapping.* ◇

Let A be a monotone mapping. A maximal monotone mapping $\bar{A}$ such that $\operatorname{graph} A \subset \operatorname{graph} \bar{A}$ is called a *maximal monotone extension* of A.

Proposition 1.4.11 (Maximal monotone extension). *If* $A : X \to 2^{X^*}$ *is a monotone mapping, then there exists at least one maximal monotone extension* $\bar{A} : X \to 2^{X^*}$ *of* A.

In the case of demi-closed and monotone mappings which are closed- and convex-valued (that is, Ax is closed and convex for any $x \in \operatorname{dom} A$), any two maximal monotone extensions differ on the boundary of their domain only. This means that if the domain of a demi-closed and monotone mapping A which is closed- and convex-valued, is an open set, then it has a single maximal monotone extension (*cf.* [1, Lemma 2.2, page 7]).

Proposition 1.4.12 (Uniqueness of maximal monotone extension). *Let* $A : X \to 2^{X^*}$ *be a monotone and demi-closed mapping. If* $x \in \operatorname{int}\operatorname{dom} A$ *and if* Ax *is closed and convex, then any maximal monotone extension* $\bar{A}$ *of* A *satisfies* $\bar{A}x = Ax$.

The following result provides a characterization of maximal monotone mappings.

Proposition 1.4.13 (Characterization of maximal monotonicity). *A mapping $A : X \to 2^{X^*}$ is maximal monotone if and only if*

$$\left.\begin{array}{r}\forall\,(y,\eta) \in \operatorname{graph} A \\ \langle \xi - \eta, x - y\rangle \geqslant 0\end{array}\right\} \quad \Longrightarrow \quad \xi \in Ax.$$

Corollary 1.4.14 (Maximal monotonicity implies demi-closedness). *Any maximal monotone mapping $A : X \to 2^{X^*}$ is demi-closed.*

Maximal monotone mappings are with closed and convex values as shown in the following result (*cf.* [82, page 105]).

Proposition 1.4.15 (Closed and convex values). *If a mapping $A : X \to 2^{X^*}$ is maximal monotone then, for any $x \in \operatorname{dom} A$, the set Ax is closed and convex in X^*.*

Definition 1.4.16 (Surjectivity). *A mapping $A : X \to 2^{X^*}$ is called surjective if for each element $\xi \in X^*$ there exists an element $x \in \operatorname{dom} A$ such that $\xi \in Ax$, i.e.,* $\operatorname{ran} A = X^*$.

The following result gives a characterization of maximal monotone mappings by means of surjectivity (*cf.* [47, Theorem 3.11, page 166]).

Proposition 1.4.17 (Surjectivity result). *Let X be a strictly convex and smooth Banach space and let $A : X \to 2^{X^*}$ be a monotone mapping. Then A is a maximal monotone mapping if and only if $A + J_X$ is surjective.*

The following is a generalization of this result (*cf.* [15, Corollary 2.3, page 59]).

Proposition 1.4.18 (General surjectivity result). *Let $A : X \to 2^{X^*}$ be a monotone mapping. Assume that $f : X \to \mathbb{R}$ is a Gâteaux differentiable, strictly convex, and cofinite function. Then A is maximal monotone if and only if* $\operatorname{ran}(A + \nabla f) = X^*$.

Among the most important examples of maximal monotone mappings are the subdifferential of proper, convex and lower semicontinuous functions. Maximal monotonicity of such subdifferentials was shown in [98] (see also [82, Theorem 2.13, page 124]).

Proposition 1.4.19 (Maximal monotonicity of the subdifferential mapping). *Let f be a proper, convex and lower semicontinuous function. Then the subdifferential mapping $\partial f : X \to 2^{X^*}$ is maximal monotone.*

Definition 1.4.20 (Sum of monotone mappings). *The sum of two mappings $A_1 : X \to 2^{X^*}$ and $A_2 : X \to 2^{X^*}$ is defined by*

$$(A_1 + A_2)\,x := \begin{cases} \varnothing & if x \notin (\operatorname{dom} A_1) \bigcap (\operatorname{dom} A_2) \\ A_1x + A_2x & if x \in (\operatorname{dom} A_1) \bigcap (\operatorname{dom} A_2) \end{cases},$$

where the addition defined by

$$A_1x + A_2x = \{\xi + \eta : \xi \in A_1x, \eta \in A_2x\}.$$

Remark 1.4.21. *Note that the set of all monotone mappings is closed under addition.* ◇

In the spirit of the previous remark, the problem of under which conditions, the sum of two maximal monotone mappings is again a maximal monotone mapping is essential and is of interest for many researchers. For instance, we present the following result in this direction (*cf.* [82, Theorem 3.6, page 142]).

Proposition 1.4.22 (Maximality of the sum of two mappings)**.** *Let* $A : X \to 2^{X^*}$ *and* $B : X \to 2^{X^*}$ *be two maximal monotone mappings. If*

$$\operatorname{int} \operatorname{dom} A \bigcap \operatorname{dom} B \neq \varnothing,$$

then the sum $A + B$ *is maximal monotone too.*

Corollary 1.4.23 (Maximality of the sum of two subdifferential mappings)**.** *Suppose that* $f : X \to (-\infty, +\infty]$ *and* $g : X \to (-\infty, +\infty]$ *are two proper, convex and lower semicontinuous functions, such that the domain of one of them intersects the interior of the domain of the other. Then*

$$\partial (f + g)(x) = \partial f(x) + \partial g(x) \quad \forall x \in \operatorname{dom}(f + g).$$

The following concept of monotonicity generalizes the classical notion.

Definition 1.4.24 (T-monotonicity)**.** *Let* $A : X \to 2^{X^*}$ *be a mapping,* $K \subset \operatorname{dom} A$ *and let* $T : K \to X$ *be an operator. We say that the mapping* A *is monotone with respect to the operator* T*, or* T*-monotone, if*

$$0 \leqslant \langle \xi - \eta, Tx - Ty \rangle \tag{1.4.4}$$

for any $x, y \in K$*, where* $\xi \in Ax$ *and* $\eta \in Ay$*.*

Clearly, when $T = I$ the classes of monotone and I-monotone operators coincide.

Definition 1.4.25 (Set-valued indicator)**.** *The set-valued indicator of a subset* S *of* X *is defined by*

$$\mathbb{I}_S : x \mapsto \begin{cases} \{0\}, & x \in S; \\ \varnothing, & \textit{otherwise}. \end{cases}$$

The concept of T-monotonicity can also be defined by using this set-valued indicator (as kindly suggested by Heinz H. Bauschke).

Remark 1.4.26 (T-monotonicity via set-valued indicator)**.** *A mapping* $A : X \to 2^{X^*}$ *is* T*-monotone if and only if* $T \circ \left(A^{-1} + \mathbb{I}_{A(K)}\right)$ *is monotone.* ◇

Remark 1.4.27 (Other T-monotonicity concept)**.** *An unrelated concept of a* T*-monotone operator can be found in several papers of Calvert (see, for example, [43]).* ◇

Remark 1.4.28 (d-accretive)**.** *Let* $F : X \to X$ *be an operator which satisfies*

$$0 \leqslant \langle J_X(x) - J_X(y), Fx - Fy\rangle \tag{1.4.5}$$

for any $x, y \in \operatorname{dom} f$*. An operator* F *which satisfies inequality* (1.4.5) *is called d-accretive (see [2]). Clearly in our terms* J_X *is* F*-monotone whenever* F *is d-accretive.* ◇

1.4.1 Bregman Inverse Strongly Monotone Mappings

This class of mappings was introduced by Butnariu and Kassay (see [38]). We assume that the Legendre function f (see Definition 1.2.7) satisfies the following range condition:

$$\operatorname{ran}(\nabla f - A) \subseteq \operatorname{ran} \nabla f. \tag{1.4.6}$$

Definition 1.4.29 (Bregman inverse strongly monotone)**.** *Let* Y *be a subset of* X*. A mapping* $A : X \to 2^{X^*}$ *is called Bregman inverse strongly monotone (BISM for short) on the set* Y *if*

$$Y \bigcap (\operatorname{dom} A) \bigcap (idomf) \neq \varnothing \tag{1.4.7}$$

and for any $x, y \in Y \bigcap (\operatorname{int}\operatorname{dom} f)$ *and* $\xi \in Ax$*,* $\eta \in Ay$*, we have*

$$\left\langle \xi - \eta, \nabla f^*\left(\nabla f(x) - \xi\right) - \nabla f^*\left(\nabla f(y) - \eta\right)\right\rangle \geqslant 0. \tag{1.4.8}$$

From the definition of the bifunction W^f (see (1.2.24)) it is easy to check that (1.4.8) is equivalent to

$$\begin{aligned} W^f\left(\xi, \nabla f^*\left(\nabla f(x) - \xi\right)\right) + W^f\left(\eta, \nabla f^*\left(\nabla f(y) - \eta\right)\right) \leqslant W^f\left(\xi, \nabla f^*\left(\nabla f(y) - \eta\right)\right) \\ + W^f\left(\eta, \nabla f^*\left(\nabla f(x) - \xi\right)\right). \end{aligned}$$

Remark 1.4.30 (Particular cases of BISM)**.** *The BISM class of mappings is a generalization of the class of firmly nonexpansive operators in Hilbert spaces (see* (1.3.12)*). Indeed, if* $f = (1/2)\|\cdot\|^2$*, then* $\nabla f = \nabla f^* = I$*, where* I *is the identity operator, and* (1.4.8) *becomes*

$$\langle \xi - \eta, x - \xi - (y - \eta)\rangle \geqslant 0, \tag{1.4.9}$$

that is,

$$\|\xi - \eta\|^2 \leqslant \langle x - y, \xi - \eta \rangle . \tag{1.4.10}$$

In other words, A is a (single-valued) firmly nonexpansive operator.

It is interesting to note that if, instead of the function $f = (1/2)\|\cdot\|^2$, we take the Hilbert space $\mathcal{H}$ with the Legendre function $(1/(2\alpha))\|\cdot\|^2$ for some positive real number α, then the inequality in (1.4.8) *becomes the usual α-inverse strongly monotone operator, that is, operator which satisfies*

$$\alpha \|Tx - Ty\|^2 \leqslant \langle Tx - Ty, x - y \rangle$$

for all $x, y \in K$. ◇

The following example shows that a BISM mapping might not be maximal monotone (see [66, Example 1, page 1324]).

Example 1.4.31 (BISM mapping which is not maximal monotone)**.** *Let K be any proper, closed and convex subset of X. Let $A : X \to 2^{X^*}$ be any BISM mapping with* $\operatorname{dom} A = K$ *such that Ax is a bounded set for any $x \in X$. Then A is not maximal monotone. Indeed,* $\operatorname{cl} K = K \neq X$*, which means that* $\operatorname{bdr} K = \operatorname{cl} K \backslash \operatorname{int} K \neq \varnothing$*. Now for any $x \in \operatorname{bdr} K$ we know that Ax is a nonempty and bounded set. On the other hand, Ax is unbounded whenever A is maximal monotone, since we know that the image of a point on the boundary of the domain of a maximal monotone mapping, if non-empty, is unbounded because it contains a half-line.*

A very simple particular case is the following one: X is a Hilbert space, $f = (1/2)\|\cdot\|^2$ (in this case BISM reduces to firm nonexpansivity (see Remark 1.4.30), K is a nonempty, closed, convex and bounded subset of X (e.g., a closed ball) and A is any single-valued BISM operator on K (e.g., the identity) and $\varnothing$ otherwise.

Problem 1. *Since a BISM mapping need not be maximal monotone, it is of interest to determine if it must be a monotone mapping.*

Remark 1.4.32 (BISM is not necessarily FNE)**.** *It is important to note that a mapping A (even in a Hilbert space provided that f is not $(1/2)\|\cdot\|^2$) does not have to be firmly nonexpansive (see* (1.3.12)*) in order to be BISM on Y (see, for example, [38, pages 2108-2109]).* ◇

Chapter 2

Fixed Point Properties of Bregman Nonexpansive Operators

In this chapter we present properties of Bregman nonexpansive operators from the point of view of their fixed points. We will present properties of the fixed point set of Bregman nonexpansive operators. In addition, existence results (sufficient and necessary conditions) are presented too. A characterization of BFNE operators is presented. It leads us to finding many examples of BFNE operators in Euclidean spaces and in Hilbert spaces.

2.1 Properties of Bregman Nonexpansive Operators

We will start with the following simple property (*cf.* [91, Lemma 15.5, page 305]) of the fixed point set of properly QBNE operators (see Definition 1.3.5 and Remark 1.3.7).

Proposition 2.1.1 (Fixed point set is closed and convex)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre function. Let* K *be a nonempty, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$*, and let* $T : K \to K$ *be a properly QBNE operator. Then* $\operatorname{Fix}(T)$ *is closed and convex.*

Proof. If $\operatorname{Fix}(T)$ is empty then the result follows immediately. Otherwise we assume that $\operatorname{Fix}(T)$ is nonempty. We first show that $\operatorname{Fix}(T)$ is closed. To this end, let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\operatorname{Fix}(T)$ such that $x_n \to \bar{x}$ as $n \to \infty$. From the definition of strictly QBNE operator (see (1.3.8)) it follows that

$$D_f(x_n, T\bar{x}) \leqslant D_f(x_n, \bar{x}) \tag{2.1.1}$$

for any $n \in \mathbb{N}$. Since f is continuous at $\bar{x} \in K \subset \operatorname{int}\operatorname{dom} f$ (see Corollary 1.1.7) and $x_n \to \bar{x}$

as $n \to \infty$, it follows that

$$\begin{aligned}\lim_{n\to\infty} D_f(x_n, T\bar{x}) &= \lim_{n\to\infty}\left[f(x_n) - f(T\bar{x}) - \langle \nabla f(T\bar{x}), x_n - T\bar{x}\rangle\right]\\ &= \left[f(\bar{x}) - f(T\bar{x}) - \langle \nabla f(T\bar{x}), \bar{x} - T\bar{x}\rangle\right] = D_f(\bar{x}, T\bar{x}).\end{aligned}$$

On the other hand, replacing $T\bar{x}$ with $\bar{x}$, one gets

$$\lim_{n\to\infty} D_f(x_n, \bar{x}) = D_f(\bar{x}, \bar{x}) = 0.$$

Thus (2.1.1) implies that $D_f(\bar{x}, T\bar{x}) = 0$ and therefore it follows from Proposition 1.2.4 that $\bar{x} = T\bar{x}$. Hence $\bar{x} \in \operatorname{Fix}(T)$ and this means that $\operatorname{Fix}(T)$ is closed, as claimed.

Next we show that $\operatorname{Fix}(T)$ is convex. For any $x, y \in \operatorname{Fix}(T)$ and $t \in (0,1)$, put $z = tx + (1-t)y$. We have to show that $Tz = z$. Indeed, from the definition of the Bregman distance (see (1.2.1)) and the definition of strictly QBNE operator (see (1.3.8)) it follows that

$$\begin{aligned}D_f(z, Tz) &= f(z) - f(Tz) - \langle \nabla f(Tz), z - Tz\rangle\\ &= f(z) - f(Tz) - \langle \nabla f(Tz), tx + (1-t)y - Tz\rangle\\ &= f(z) + tD_f(x, Tz) + (1-t)D_f(y, Tz) - tf(x) - (1-t)f(y)\\ &\leqslant f(z) + tD_f(x, z) + (1-t)D_f(y, z) - tf(x) - (1-t)f(y)\\ &= \langle \nabla f(z), z - tx - (1-t)y\rangle = 0.\end{aligned}$$

Again from Proposition 1.2.4 it follows that $Tz = z$. Therefore $\operatorname{Fix}(T)$ is also convex, as asserted. □

Next we show that if f is an admissible function (see Definition 1.2.1) which is bounded and uniformly Fréchet differentiable on bounded subsets of X (see Definition 1.1.20(ii)), and T is a BFNE operator (see Definition 1.3.5(i)), then the fixed point set of T coincides with the set of its asymptotic fixed points (*cf.* [91, Lemma 15.6, page 306]).

Proposition 2.1.2 (Sufficient condition for $\widehat{\operatorname{Fix}}(T) = \operatorname{Fix}(T)$)**.** *Let $f : X \to \mathbb{R}$ be an admissible function which is uniformly Fréchet differentiable and bounded on bounded subsets of X. Let K be a nonempty, closed and convex subset of X and let $T : K \to \operatorname{int}\operatorname{dom} f$ be a BFNE operator. Then $\widehat{\operatorname{Fix}}(T) = \operatorname{Fix}(T)$.*

Proof. The inclusion $\operatorname{Fix}(T) \subset \widehat{\operatorname{Fix}}(T)$ is obvious. To show that $\operatorname{Fix}(T) \supset \widehat{\operatorname{Fix}}(T)$, let $u \in \widehat{\operatorname{Fix}}(T)$ be given. Then, from Definition 1.3.6 we get a sequence $\{x_n\}_{n\in\mathbb{N}}$ in K such that both $x_n \rightharpoonup u$ and $\|x_n - Tx_n\| \to 0$ as $n \to \infty$. Since f is bounded and uniformly Fréchet differentiable on bounded subsets of X, ∇f is uniformly continuous on bounded subsets of

X (see Proposition 1.1.22(ii)). Hence $\left\|\nabla f\left(Tx_n\right)-\nabla f\left(x_n\right)\right\|_* \to 0$ as $n \to \infty$ and therefore

$$\lim_{n\to\infty}\left\langle \nabla f\left(Tx_n\right)-\nabla f\left(x_n\right), y\right\rangle = 0 \tag{2.1.2}$$

for any $y \in X$, and

$$\lim_{n\to\infty}\left\langle \nabla f\left(Tx_n\right)-\nabla f\left(x_n\right), x_n\right\rangle = 0 \tag{2.1.3}$$

because $\left\{x_n\right\}_{n\in\mathbb{N}}$ is bounded as a weakly convergent sequence. On the other hand, since T is a BFNE operator (see (1.3.4)), we have

$$0 \leqslant D_f\left(Tx_n, u\right) - D_f\left(Tx_n, Tu\right) + D_f\left(Tu, x_n\right) - D_f\left(Tu, Tx_n\right). \tag{2.1.4}$$

From the three point identity (see (1.2.2)) and (2.1.4) we now obtain

$$\begin{aligned}
D_f\left(u, Tu\right) &= D_f\left(Tx_n, Tu\right) - D_f\left(Tx_n, u\right) - \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&\leqslant D_f\left(Tu, x_n\right) - D_f\left(Tu, Tx_n\right) - \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&= \left[f\left(Tu\right) - f\left(x_n\right) - \left\langle \nabla f\left(x_n\right), Tu - x_n\right\rangle\right] - \\
&\left[f\left(Tu\right) - f\left(Tx_n\right) - \langle \nabla f(Tx_n), Tu - Tx_n\rangle\right] \\
&- \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&= f\left(Tx_n\right) - f\left(x_n\right) - \left\langle \nabla f\left(x_n\right), Tu - x_n\right\rangle + \left\langle \nabla f\left(Tx_n\right), Tu - Tx_n\right\rangle \\
&- \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&= -\left[f\left(x_n\right) - f\left(Tx_n\right) - \left\langle \nabla f\left(Tx_n\right), x_n - Tx_n\right\rangle\right] - \left\langle \nabla f\left(Tx_n\right), x_n - Tx_n\right\rangle \\
&- \left\langle \nabla f\left(x_n\right), Tu - x_n\right\rangle + \left\langle \nabla f\left(Tx_n\right), Tu - Tx_n\right\rangle \\
&- \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&= -D_f\left(x_n, Tx_n\right) - \left\langle \nabla f\left(Tx_n\right), x_n - Tx_n\right\rangle - \left\langle \nabla f\left(x_n\right), Tu - x_n\right\rangle \\
&+ \left\langle \nabla f\left(Tx_n\right), Tu - Tx_n\right\rangle - \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&\leqslant -\left\langle \nabla f\left(Tx_n\right), x_n - Tx_n\right\rangle - \left\langle \nabla f\left(x_n\right), Tu - x_n\right\rangle \\
&+ \left\langle \nabla f\left(Tx_n\right), Tu - Tx_n\right\rangle - \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - u\right\rangle \\
&= \left\langle \nabla f\left(x_n\right) - \nabla f\left(Tx_n\right), x_n - Tu\right\rangle - \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), Tx_n - x_n\right\rangle \\
&- \left\langle \nabla f\left(u\right) - \nabla f\left(Tu\right), x_n - u\right\rangle.
\end{aligned}$$

From (2.1.2), (2.1.3), and the hypotheses that both $x_n \rightharpoonup u$ and $\left\|x_n - Tx_n\right\| \to 0$ as $n \to \infty$, we get that $D_f\left(u, Tu\right) \leqslant 0$. Consequently $D_f\left(u, Tu\right) = 0$ and from Proposition 1.2.4 it follows that $Tu = u$. That is, $u \in \operatorname{Fix}\left(T\right)$, as required. □

Remark 2.1.3 (BFNE is BSNE)**.** *From Proposition 2.1.2 it follows that if an admissible* *function $f : X \to \mathbb{R}$ is uniformly Fréchet differentiable and bounded on bounded subsets of X, then any BFNE operator is also a BSNE operator (see Figure 1.3).* $\diamond$

Now we obtain necessary and sufficient conditions for BFNE operators to have a (common) fixed point in general reflexive Banach spaces. We begin with a theorem for a single strictly QBNE operator; hence it also holds for a BFNE operator.

Proposition 2.1.4 (Necessary condition for $\widehat{\operatorname{Fix}}(T)$ to be nonempty)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of* $\operatorname{int}\operatorname{dom} f$ *and let $T : K \to K$ be a strictly QBNE operator. If* $\widehat{\operatorname{Fix}}(T)$ *is nonempty, then $\{T^n y\}_{n\in\mathbb{N}}$ is bounded for each $y \in K$.*

Proof. We know from the definition of strictly QBNE operators (see (1.3.8)) that

$$D_f(p, Ty) \leqslant D_f(p, y)$$

for any $p \in \widehat{\operatorname{Fix}}(T)$ and $y \in K$. Therefore

$$D_f(p, T^n y) \leqslant D_f(p, y)$$

for any $p \in \widehat{\operatorname{Fix}}(T)$ and $y \in K$. This inequality shows that the nonnegative sequence $\{D_f(p, T^n y)\}_{n\in\mathbb{N}}$ is bounded. Now Proposition 1.2.48 implies that the sequence $\{T^n y\}_{n\in\mathbb{N}}$ is bounded too, as claimed. $\square$

A result in this spirit for properly QBNE operators was first proved in [91, Theorem 15.7, page 307].

Corollary 2.1.5 (Necessary condition for $\operatorname{Fix}(T)$ to be nonempty)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of* $\operatorname{int}\operatorname{dom} f$ *and let $T : K \to K$ be a properly QBNE operator. If* $\operatorname{Fix}(T)$ *is nonempty, then $\{T^n y\}_{n\in\mathbb{N}}$ is bounded for each $y \in K$.*

Proof. Follow the arguments in the proof of Proposition 2.1.4 and replace $p \in \widehat{\operatorname{Fix}}(T)$ with $p \in \operatorname{Fix}(T)$. $\square$

For an operator $T : K \to K$, let $S_n(z) := (1/n)\sum_{k=1}^{n} T^k z$ for all $z \in K$. The next result give a sufficient condition for BFNE operators to have a fixed point (*cf.* [91, Theorem 15.8, page 310]).

Proposition 2.1.6 (Sufficient condition for $\operatorname{Fix}(T)$ to be nonempty)**.** *Let $f : X \to (-\infty, +\infty]$ be an admissible function. Let K be a nonempty, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$ *and let $T : K \to K$ be a BFNE operator. If there exists $y \in K$ such that $\|S_n(y)\| \nrightarrow \infty$ as $n \to \infty$, then* $\operatorname{Fix}(T)$ *is nonempty.*

Proof. Suppose that there exists $y \in K$ such that $\|S_n(y)\| \nrightarrow \infty$ as $n \to \infty$. Let $x \in K$, $k \in \mathbb{N}$ and $n \in \mathbb{N}$ be given. Since T is BFNE (see (1.3.4)), we have

$$D_f\left(T^{k+1}y, Tx\right) + D_f\left(Tx, T^{k+1}y\right) \leqslant D_f\left(Tx, T^k y\right) + D_f\left(T^{k+1}y, x\right). \tag{2.1.5}$$

From the three point identity (see (1.2.2)) we get that

$$\begin{aligned} D_f\left(T^{k+1}y, Tx\right) + D_f\left(Tx, T^{k+1}y\right) \leqslant{} & D_f\left(Tx, T^k y\right) + D_f\left(T^{k+1}y, Tx\right) + D_f\left(Tx, x\right) \\ & + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), T^{k+1}y - Tx \right\rangle. \end{aligned}$$

This implies that

$$0 \leqslant D_f\left(Tx, x\right) + D_f\left(Tx, T^k y\right) - D_f\left(Tx, T^{k+1}y\right) + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), T^{k+1}y - Tx \right\rangle.$$

Summing up these inequalities with respect to $k = 0, 1, \ldots, n-1$, we now obtain

$$0 \leqslant nD_f\left(Tx, x\right) + D_f\left(Tx, y\right) - D_f\left(Tx, T^n y\right) + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), \sum_{k=0}^{n-1} T^{k+1}y - nTx \right\rangle$$

where $T^0 = I$ is the identity operator. Dividing this inequality by n, we have

$$0 \leqslant D_f\left(Tx, x\right) + \frac{1}{n}\left[D_f\left(Tx, y\right) - D_f\left(Tx, T^n y\right)\right] + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), S_n\left(y\right) - Tx \right\rangle$$

and

$$0 \leqslant D_f\left(Tx, x\right) + \frac{1}{n}D_f\left(Tx, y\right) + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), S_n\left(y\right) - Tx \right\rangle. \tag{2.1.6}$$

Since $\|S_n(y)\| \nrightarrow \infty$ as $n \to \infty$ by assumption, there exists a subsequence $\{S_{n_k}(y)\}_{k \in \mathbb{N}}$ of $\{S_n(y)\}_{n \in \mathbb{N}}$ such that $S_{n_k}(y) \rightharpoonup u \in K$ as $k \to \infty$. Letting $k \to \infty$ in (2.1.6), we obtain

$$0 \leqslant D_f\left(Tx, x\right) + \left\langle \nabla f\left(Tx\right) - \nabla f\left(x\right), u - Tx \right\rangle. \tag{2.1.7}$$

Setting $x = u$ in (2.1.7), we get from the four point identity (see (1.2.3)) that

$$\begin{aligned} 0 &\leqslant D_f\left(Tu, u\right) + \left\langle \nabla f\left(Tu\right) - \nabla f\left(u\right), u - Tu \right\rangle \\ &= D_f\left(Tu, u\right) + D_f\left(u, u\right) - D_f\left(u, Tu\right) - D_f\left(Tu, u\right) + D_f\left(Tu, Tu\right) \\ &= -D_f\left(u, Tu\right). \end{aligned}$$

Hence $D_f(u, Tu) \leqslant 0$ and so $D_f(u, Tu) = 0$. It now follows from Proposition 1.2.4 that $Tu = u$. That is, $u \in \operatorname{Fix}(T)$. □

Remark 2.1.7 (Non-spreading)**.** *As can be seen from the proof of Proposition 2.1.6, the* *result remains true for those operators which only satisfy* (2.1.5). *In the special case where* $f = (1/2)\left\|\cdot\right\|^2$, *such operators are called non-spreading. For more information see [69].* $\diamond$

We remark in passing that we still do not know if the analog of Proposition 2.1.6 for nonexpansive operators holds outside Hilbert space (*cf.* [84, Remark 2, page 275]).

The following corollary brings out conditions for the fixed point property of BFNE operators (*cf.* [91, Corollary 15.11, page 309]).

Corollary 2.1.8 (Fixed point property of BFNE operators)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function. Every nonempty, bounded, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$ *has the fixed point property for BFNE self-operators.*

In order to prove a common fixed point theorem, we need the following lemma.

Lemma 2.1.9 (Common fixed point - finite family)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function. Let* K *be a nonempty, bounded, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$. *Let* $\{T_1, T_2, \ldots, T_N\}$ *be a commutative finite family of* N *BFNE operators from* K *into itself. Then* $\{T_1, T_2, \ldots, T_N\}$ *has a common fixed point.*

Proof. The proof is by way of induction over N. We first show the result for the case $N = 2$. From Proposition 2.1.1 and Corollary 2.1.8, $\operatorname{Fix}(T_1)$ is nonempty, bounded, closed and convex. It follows from $T_1 \circ T_2 = T_2 \circ T_1$ that if $u \in \operatorname{Fix}(T_1)$, then we have $T_1 \circ T_2 u = T_2 \circ T_1 u = T_2 u$. Thus $T_2 u \in \operatorname{Fix}(T_1)$. Hence the restriction of T_2 to $\operatorname{Fix}(T_1)$ is a BFNE self-operator. From Corollary 2.1.8, T_2 has a fixed point in $\operatorname{Fix}(T_1)$, that is, we have $v \in \operatorname{Fix}(T_1)$ such that $T_2 v = v$. Consequently, $v \in \operatorname{Fix}(T_1) \bigcap \operatorname{Fix}(T_2)$.

Suppose that for some $N \geqslant 2$, $F = \bigcap_{i=1}^{N} \operatorname{Fix}(T_i)$ is nonempty. Then F is a nonempty, bounded, closed and convex subset of K and the restriction of T_{N+1} to F is BFNE self-operator. From Corollary 2.1.8, T_{N+1} has a fixed point in F. This shows that $F \bigcap \operatorname{Fix}(T_{N+1})$ is nonempty. This complete the proof. $\square$

Using Lemma 2.1.9, we finally prove the following common fixed point theorem for a commutative family of BFNE operators (*cf.* [91, Theorem 15.12, page 309]).

Theorem 2.1.10 (Common fixed point - infinite family)**.** *Let* $f : X \to (-\infty, +\infty]$ *be an admissible function. Let* K *be a nonempty, bounded, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$. *Let* $\{T_\alpha\}_{\alpha \in A}$ *be a commutative family of BFNE operators from* K *into itself. Then the family* $\{T_\alpha\}_{\alpha \in A}$ *has a common fixed point.*

Proof. From Proposition 2.1.1 we know that each $\operatorname{Fix}(T_\alpha)$, $\alpha \in A$, is closed and convex subset of K. Since X is reflexive and K is bounded, closed and convex, K is weakly compact. Thus, to show that $\bigcap_{\alpha \in A} \operatorname{Fix}(T_\alpha)$ is nonempty, it is sufficient to show that $\{\operatorname{Fix}(T_\alpha)\}_{\alpha \in A}$ has the finite intersection property. From Lemma 2.1.9 we know that $\{\operatorname{Fix}(T_\alpha)\}_{\alpha \in A}$ has this property. Thus the proof is complete. $\square$

Now we present two important properties of strictly BSNE operators (see Definition 1.3.8) which were proved in [88, Lemma 1, page 314] and [88, Lemma 2, page 314]. Both results deal with the composition of N strictly BSNE operators. We start with the following result.

Proposition 2.1.11 (Asymptotic fixed points of strictly BSNE operators). *Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of X. If $\{T_i : 1 \leqslant i \leqslant N\}$ are N strictly BSNE operators from K into itself, and the set*

$$\widehat{F} = \bigcap \left\{ \widehat{\operatorname{Fix}}\,(T_i) : 1 \leqslant i \leqslant N \right\}$$

is not empty, then $\widehat{\operatorname{Fix}}\,(T_N \circ \cdots \circ T_1) \subset \widehat{F}$.

Proof. Let $u \in \widehat{F}$. Given $x \in \widehat{\operatorname{Fix}}\,(T)$, Definition 1.3.6 implies that there exists a sequence $\{x_n\}_{n \in \mathbb{N}} \subset K$ converging weakly to x such that

$$\lim_{n \to \infty} \|x_n - Tx_n\| = 0. \tag{2.1.8}$$

we first note that since the function f is bounded on bounded subsets of X, the gradient ∇f is also bounded on bounded subsets of X (see Proposition 1.1.15). Thus the sequences $\{x_n\}_{n \in \mathbb{N}}$ and $\{\nabla f\,(Tx_n)\}_{n \in \mathbb{N}}$ are bounded. Since f is uniformly Fréchet differentiable on bounded subsets of X, it is also uniformly continuous on bounded subsets of X (see Proposition 1.1.22(i)) and therefore

$$\lim_{n \to \infty} \left(f\,(Tx_n) - f\,(x_n)\right) = 0. \tag{2.1.9}$$

In addition, from Proposition 1.1.22(ii) we obtain that ∇f is also uniformly continuous on bounded subsets of X and thus

$$\lim_{n \to \infty} \left\|\nabla f\,(Tx_n) - \nabla f\,(x_n)\right\|_* = 0. \tag{2.1.10}$$

From the definition of the Bregman distance (see (1.2.1)) we obtain that

$$\begin{aligned} D_f\,(u, x_n) - D_f\,(u, Tx_n) &= \left[f\,(u) - f\,(x_n) - \langle \nabla f\,(x_n)\,, u - x_n\rangle\right] \\ &\quad - \left[f\,(u) - f\,(Tx_n) - \langle \nabla f\,(Tx_n)\,, u - Tx_n\rangle\right] \\ &= f\,(Tx_n) - f\,(x_n) - \langle \nabla f\,(x_n)\,, u - x_n\rangle + \langle \nabla f\,(Tx_n)\,, u - Tx_n\rangle \\ &= f\,(Tx_n) - f\,(x_n) - \langle \nabla f\,(x_n) - \nabla f\,(Tx_n)\,, u - x_n\rangle \\ &\quad + \langle \nabla f\,(Tx_n)\,, x_n - Tx_n\rangle. \end{aligned}$$

Hence from the fact that both the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{\nabla f(Tx_n)\}_{n\in\mathbb{N}}$ are bounded, along with (2.1.8), (2.1.9) and (2.1.10) we obtain that

$$\lim_{n\to\infty}\left(D_f(u,x_n)-D_f(u,Tx_n)\right)=0. \tag{2.1.11}$$

Set $y_n = T_{N-1}\circ\cdots\circ T_1 x_n$ so that $T_N y_n = Tx_n$. From the first part of the definition of strictly QBNE operator (see (1.3.9)) we get

$$D_f(u,Tx_n)=D_f(u,T_N y_n)\leqslant D_f(u,y_n)\leqslant D_f(u,x_n). \tag{2.1.12}$$

Hence from (2.1.11) we get that

$$\lim_{n\to\infty}\left(D_f(u,y_n)-D_f(u,T_N y_n)\right)=\lim_{n\to\infty}\left(D_f(u,x_n)-D_f(u,Tx_n)\right)=0. \tag{2.1.13}$$

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded and both f and ∇f are bounded on bounded subsets of X, we have that $\{D_f(u,x_n)\}_{n\in\mathbb{N}}$ is also bounded. Therefore if follows from (2.1.12) that $\{D_f(u,y_n)\}_{n\in\mathbb{N}}$ is bounded too. Since ∇f^* is bounded on bounded subsets of $\operatorname{int}\operatorname{dom} f^*$ it follows from Proposition 1.2.48 that $\{y_n\}_{n\in\mathbb{N}}$ is bounded. This together with (2.1.13) implies that

$$\lim_{n\to\infty} D_f(T_N y_n, y_n)=0,$$

because T_N is strictly BSNE (see Definition 1.3.8). Since $\{y_n\}_{n\in\mathbb{N}}$ is bounded, Proposition 1.2.46 now implies that $\lim_{n\to\infty}\|y_n - T_N y_n\| = 0$. Consequently,

$$\lim_{n\to\infty}\|x_n - T_{N-1}\circ\cdots\circ T_1 x_n\| = \lim_{n\to\infty}\|x_n-y_n\|\leqslant\lim_{n\to\infty}\left(\|x_n-Tx_n\|+\|y_n-T_N y_n\|\right)=0.$$

This implies, on one hand, that the sequence $\{y_n\}_{n\in\mathbb{N}}$ also converges weakly to x and thus $x\in\widehat{\operatorname{Fix}}(T_N)$, and on the other hand, that $x\in\widehat{\operatorname{Fix}}(T_{N-1}\circ\cdots\circ T_1)$. Repeating the same argument we obtain that $x\in\widehat{\operatorname{Fix}}(T_i)$ for any $i=1,2,\ldots,N-1$, thence $x\in\widehat{F}$, as asserted. □

The next result shows that the composition of N strictly BSNE operators is also strictly BSNE operator.

Proposition 2.1.12 (Composition of strictly BSNE operators)**.** *Let $f: X\to\mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of X. Let $\{T_i : 1\leqslant i\leqslant N\}$ be N strictly BSNE operators from K into itself and let $T = T_N\circ T_{N-1}\circ\cdots\circ T_1$. If the sets*

$$\widehat{F}=\bigcap\left\{\widehat{\operatorname{Fix}}(T_i):1\leqslant i\leqslant N\right\}$$

and $\widehat{\mathrm{Fix}}(T)$ *are not empty, then* T *is also strictly BSNE.*

Proof. Let$u \in \widehat{\mathrm{Fix}}(T)$ and $x \in K$, then the first part of the definition of strictly BSNE operator (see (1.3.9)) is satisfied because $u \in \hat{F}$ by Proposition 2.1.11 and since any strictly BSNE operator is strictly QBNE (see Figure 1.3). Assume that $u \in \widehat{\mathrm{Fix}}(T)$ and $\{x_n\}_{n\in\mathbb{N}}$ is a bounded sequence such that

$$\lim_{n\to\infty} \left(D_f\left(u, x_n\right) - D_f\left(u, Tx_n\right)\right) = 0.$$

In order to prove the second part of the definition of strictly BSNE operator (see (1.3.11)), note that for any $i = 2, 3, \ldots, N$, we have from (1.3.9) that

$$0 \leqslant D_f\left(u, T_{i-1} \circ \cdots \circ T_1 x_n\right) - D_f\left(u, T_i \circ T_{i-1} \circ \cdots \circ T_1 x_n\right) \leqslant D_f\left(u, x_n\right) - D_f\left(u, Tx_n\right),$$

and using the same arguments as in the proof of Proposition 2.1.11, we get

$$\lim_{n\to\infty} D_f\left(T_i \circ T_{i-1} \circ \cdots \circ T_1 x_n, T_{i-1} \circ \cdots \circ T_1 x_n\right) = 0,$$

where the sequence $\{T_{i-1} \circ \cdots \circ T_1 x_n\}_{n\in\mathbb{N}}$ is bounded. Now Proposition 1.2.46 implies that

$$\lim_{n\to\infty} \left\|T_i \circ T_{i-1} \circ \cdots \circ T_1 x_n - T_{i-1} \circ \cdots \circ T_1 x_n\right\| = 0$$

for each $i = 2, 3, \ldots, N$. Since

$$\|x_n - Tx_n\| \leqslant \|x_n - T_1 x_n\| + \|T_1 x_n - T_2 \circ T_1 x_n\| + \cdots + \|T_{N-1} \circ T_{N-2} \circ \cdots \circ T_1 x_n - Tx_n\|$$

we get that

$$\lim_{n\to\infty} \|x_n - Tx_n\| = 0.$$

The function f is bounded on bounded subsets of X and therefore ∇f is also bounded on bounded subsets of X (see Proposition 1.1.15). Thus both the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{\nabla f(Tx_n)\}_{n\in\mathbb{N}}$ are bounded. Since f is also uniformly continuous on bounded subsets of X (see Proposition 1.1.22(i)), we have that

$$\lim_{n\to\infty} \left(f\left(Tx_n\right) - f\left(x_n\right)\right) = 0.$$

So from the definition of the Bregman distance (see (1.2.1)) we obtain that

$$\lim_{n\to\infty} D_f\left(Tx_n, x_n\right) = 0.$$

Hence T is strictly L-BSNE, as asserted. □

In applications it seems that the assumption $\widehat{\text{Fix}}(T) = \text{Fix}(T)$ imposed on the operator T is essential for the convergence of iterative methods. In Proposition 2.1.2 we gave sufficient condition for BFNE operators to satisfy this condition (see also Remark 2.1.3). In the following remark we show that this condition holds for the composition of N strictly BSNE operators where each operator satisfy this condition.

Remark 2.1.13 (Property of the composition)**.** *Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of X. Let $\{T_i : 1 \leqslant i \leqslant N\}$ be N strictly BSNE operators and let $T = T_N \circ T_{N-1} \circ \cdots \circ T_1$. If $F = \bigcap \{\text{Fix}(T_i) : 1 \leqslant i \leqslant N\}$ and $\text{Fix}(T)$ are nonempty, then T is also strictly BSNE with $\widehat{\text{Fix}}(T) = \text{Fix}(T)$. Indeed, from Proposition 2.1.11 we get that*

$$\text{Fix}(T) \subset \widehat{\text{Fix}}(T) \subset \bigcap \left\{\widehat{\text{Fix}}(T_i) : 1 \leqslant i \leqslant N\right\} = \bigcap \{\text{Fix}(T_i) : 1 \leqslant i \leqslant N\} \subset \text{Fix}(T),$$

which implies that $\widehat{\text{Fix}}(T) = \text{Fix}(T)$, as claimed.

In the following result we prove that any BSNE operator is asymptotically regular (*cf.* [74, Proposition 11, page 11]).

Proposition 2.1.14 (BSNE operators are asymptotically regular)**.** *Assume that $f : X \to (-\infty, +\infty]$ is a Legendre function which is totally convex on bounded subsets of* $\operatorname{int}\operatorname{dom} f$ *and assume that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty subset of* $\operatorname{int}\operatorname{dom} f$*. Let T be a strictly* (*properly*) *BSNE operator from K into itself such that $\widehat{\text{Fix}}(T) \neq \varnothing$* ($\text{Fix}(T) \neq \varnothing$)*. Then T is asymptotically regular.*

Proof. Assume that T is strictly BSNE. Let $u \in \widehat{\text{Fix}}(T)$ and let $x \in K$. From (1.3.9) we get that

$$D_f\left(u, T^{n+1}x\right) \leqslant D_f\left(u, T^n x\right) \leqslant \ldots \leqslant D_f\left(u, Tx\right).$$

Thus $\lim_{n\to\infty} D_f(u, T^n x)$ exists and the sequence $\{D_f(u, T^n x)\}_{n\in\mathbb{N}}$ is bounded. Now Proposition 1.2.48 implies that $\{T^n x\}_{n\in\mathbb{N}}$ is also bounded for any $x \in K$. Since the limit $\lim_{n\to\infty} D_f(u, T^n x)$ exists, we have

$$\lim_{n\to\infty}\left(D_f\left(u, T^n x\right) - D_f\left(u, T^{n+1}x\right)\right) = 0.$$

From the definition of strictly BSNE operator (see (1.3.10) and (1.3.11)) we get

$$\lim_{n\to\infty} D_f\left(T^{n+1}x, T^n x\right) = 0.$$

Since $\{T^n x\}_{n\in\mathbb{N}}$ is bounded, we now obtain from Proposition 1.2.46 that

$$\lim_{n\to\infty} \left\|T^{n+1}x - T^n x\right\| = 0.$$

In other words, T is asymptotically regular (see Definition 1.3.14). The proof when T is properly BSNE is identical when we take $u \in \operatorname{Fix}(T)$. □

Definition 2.1.15 (Block operator)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function. Let $\{T_i : 1 \leqslant i \leqslant N\}$ be N operators from X to X and let $\{w_i\}_{i=1}^N \subset (0,1)$ satisfy $\sum_{i=1}^N w_i = 1$. Then the block operator corresponding to $\{T_i : 1 \leqslant i \leqslant N\}$ and $\{w_i : 1 \leqslant i \leqslant N\}$ is defined by*

$$T_B := \nabla f^* \left(\sum_{i=1}^N w_i \nabla f\left(T_i\right)\right). \tag{2.1.14}$$

The following inequality will be essential in our next results. From Proposition 1.2.42(i) and (ii) we have

$$\begin{aligned} D_f\left(p, T_B x\right) &= D_f\left(p, \nabla f^*\left(\sum_{i=1}^N w_i \nabla f\left(T_i x\right)\right)\right) = W^f\left(\sum_{i=1}^N w_i \nabla f\left(T_i x\right), p\right) \\ &\leqslant \sum_{i=1}^N w_i W^f\left(\nabla f\left(T_i x\right), p\right) = \sum_{i=1}^N w_i D_f\left(p, T_i x\right) \end{aligned} \tag{2.1.15}$$

In our next result we prove that the block operator defined by (2.1.14) is properly QBNE when each T_i, $1 \leqslant i \leqslant N$, is properly QBNE (*cf.* [74, Proposition 12, page 16]).

Proposition 2.1.16 (Block operator of properly QBNE operators)**.** *Assume that $f : X \to (-\infty, +\infty]$ is a Legendre function and let $\{T_i : 1 \leqslant i \leqslant N\}$ be N properly QBNE operators from X into X such that $F = \bigcap\{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\} \neq \varnothing$. Let $\{w_i\}_{i=1}^N \subset (0,1)$ which satisfy $\sum_{i=1}^N w_i = 1$. Then T_B is properly QBNE with respect to $F = \operatorname{Fix}(T_B)$.*

Proof. Let $p \in F$. Since each T_i, $i = 1, 2, \ldots, N$, is properly QBNE (see (1.3.8)), we obtain from (2.1.15) that

$$D_f\left(p, T_B x\right) \leqslant \sum_{i=1}^N w_i D_f\left(p, T_i x\right) \leqslant \sum_{i=1}^N w_i D_f\left(p, x\right) = D_f\left(p, x\right) \tag{2.1.16}$$

for all $x \in X$. Thus T_B is a properly QBNE operator with respect to F. Next we show that $\operatorname{Fix}(T_B) = F$.

The inclusion $F \subset \operatorname{Fix}(T_B)$ is obvious, so it is enough to show that $\operatorname{Fix}(T_B) \subset F$. To this end, let $u \in \operatorname{Fix}(T_B)$ and take $k \in \{1, 2, \ldots, N\}$. For all $p \in F$, such that $p \neq u$, we obtain from (2.1.15) that

$$D_f(p,u) = D_f(p,T_Bu) \leqslant \sum_{i=1}^{N} w_i D_f(p,T_iu) \leqslant \sum_{i\neq k} w_i D_f(p,u) + w_k D_f(p,T_ku).$$

Therefore

$$w_k D_f(p,u) = \left(1 - \sum_{i\neq k} w_i\right) D_f(p,u) \leqslant w_k D_f(p,T_ku),$$

that is,

$$w_k D_f(p,u) \leqslant w_k D_f(p,T_ku).$$

Since $w_k > 0$, it follows that $D_f(p,u) \leqslant D_f(p,T_ku)$. On the other hand, since T_k is properly QBNE and $p \in F \subset \operatorname{Fix}(T_k)$, we have that $D_f(p,T_ku) \leqslant D_f(p,u)$. Thus $D_f(p,u) = D_f(p,T_ku)$ for all $k \in \{1,2,\ldots,N\}$. Hence

$$D_f\left(p, \nabla f^*\left(\sum_{i=1}^{N} w_i \nabla f(T_iu)\right)\right) = D_f(p,T_Bu) = D_f(p,u) = \sum_{i=1}^{N} w_i D_f(p,T_iu). \quad (2.1.17)$$

Now Lemma 1.2.44 implies that $T_1u = T_2u = \ldots = T_nu$. Therefore $u \in F$. □

In the following result we prove that the asymptotic fixed point set of the block operator is a subset of the intersection of the asymptotic fixed point sets of the strictly BSNE operators generating the block operator (*cf.* [74, Proposition 13, page 17]).

Proposition 2.1.17 (Asymptotic fixed point of the block operator)**.** *Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Let K be a nonempty subset of X. If each T_i, $i = 1,2,\ldots,N$, is a strictly BSNE operator from X into itself, and the set*

$$\widehat{F} := \bigcap \left\{\widehat{\operatorname{Fix}}(T_i) : 1 \leqslant i \leqslant N\right\}$$

is not empty, then $\widehat{\operatorname{Fix}}(T_B) \subset \widehat{F}$.

Proof. Let $u \in \widehat{F}$ and let $x \in \widehat{\operatorname{Fix}}(T_B)$. Then, from the definition of asymptotic fixed point (see Definition 1.3.6), there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ which converges weakly to x such that $\lim_{n\to\infty} \|x_n - T_Bx_n\| = 0$. Since the function f is bounded on bounded subsets of X, ∇f is also bounded on bounded subsets of X (see Proposition 1.1.15). So the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{\nabla f(T_Bx_n)\}_{n\in\mathbb{N}}$ are bounded. Since f is also uniformly Fréchet differentiable on bounded subsets of X, it is uniformly continuous on bounded subsets of

X (see Proposition 1.1.22(i)), and therefore

$$\lim_{n\to\infty}\left(f\left(T_B x_n\right)-f\left(x_n\right)\right)=0. \tag{2.1.18}$$

In addition, from Proposition 1.1.22(ii) we obtain that ∇f is also uniformly continuous on bounded subsets of X and thus

$$\lim_{n\to\infty}\left\|\nabla f\left(T_B x_n\right)-\nabla f\left(x_n\right)\right\|_{*}=0. \tag{2.1.19}$$

From the definition of the Bregman distance (see (1.2.1)) we obtain that

$$\begin{aligned} D_f\left(u,x_n\right)-D_f\left(u,T_B x_n\right) &= \left[f\left(u\right)-f\left(x_n\right)-\left\langle \nabla f\left(x_n\right),u-x_n\right\rangle\right] \\ &\quad -\left[f\left(u\right)-f\left(T_B x_n\right)-\left\langle \nabla f\left(T_B x_n\right),u-T_B x_n\right\rangle\right] \\ &= f\left(T_B x_n\right)-f\left(x_n\right)-\left\langle \nabla f\left(x_n\right),u-x_n\right\rangle \\ &\quad +\left\langle \nabla f\left(T_B x_n\right),u-T_B x_n\right\rangle \\ &= f\left(T_B x_n\right)-f\left(x_n\right)-\left\langle \nabla f\left(x_n\right)-\nabla f\left(T_B x_n\right),u-x_n\right\rangle \\ &\quad +\left\langle \nabla f\left(T_B x_n\right),x_n-T_B x_n\right\rangle. \end{aligned}$$

Since $\lim_{n\to\infty}\left\|x_n-T_B x_n\right\|=0$, the sequences $\left\{x_n\right\}_{n\in\mathbb{N}}$ and $\left\{\nabla f\left(T_B(x_n)\right)\right\}_{n\in\mathbb{N}}$ are bounded, (2.1.18) and (2.1.19), we obtain that

$$\lim_{n\to\infty}\left(D_f\left(u,x_n\right)-D_f\left(u,T_B x_n\right)\right)=0. \tag{2.1.20}$$

Since each operator T_i, $i=1,2,\ldots,N$, is strictly BSNE, we deduce from (1.3.9) and (2.1.15) that for any $k=1,2,\ldots,N$,

$$\begin{aligned} D_f\left(u,T_B x_n\right) &\leqslant \sum_{i=1}^{N} w_i D_f\left(u,T_i x_n\right)=w_k D_f\left(u,T_k x_n\right)+\sum_{i\neq k} w_i D_f\left(u,T_i x_n\right) \\ &\leqslant w_k D_f\left(u,T_k x_n\right)+\sum_{i\neq k} w_i D_f\left(u,x_n\right) \\ &= w_k D_f\left(u,T_k x_n\right)+\left(1-w_k\right)D_f\left(u,x_n\right) \\ &= w_k\left(D_f\left(u,T_k x_n\right)-D_f\left(u,x_n\right)\right)+D_f\left(u,x_n\right). \end{aligned}$$

Hence, for any $k\in\{1,2,\ldots,N\}$, we have from (2.1.20) that

$$\lim_{n\to\infty} w_k\left(D_f\left(u,x_n\right)-D_f\left(u,T_k x_n\right)\right)\leqslant \lim_{n\to\infty}\left(D_f\left(u,x_n\right)-D_f\left(u,T_B x_n\right)\right)=0.$$

Thence

$$\lim_{n\to\infty}\left(D_f\left(u,x_n\right)-D_f\left(u,T_kx_n\right)\right)=0$$

for any $k\in\{1,2,\ldots,N\}$. Since each operator T_i, $i=1,2,\ldots,N$, is strictly BSNE and $\{x_n\}_{n\in\mathbb{N}}$ is bounded, we get from (1.3.10) and (1.3.11) that

$$\lim_{n\to\infty}D_f\left(T_ix_n,x_n\right)=0.$$

Since f is totally convex (see Definition 1.2.8) and $\{x_n\}_{n\in\mathbb{N}}$ is bounded, it follows from Proposition 1.2.46 that

$$\lim_{n\to\infty}\left\|T_ix_n-x_n\right\|=0.$$

This means that x belongs to $\widehat{\operatorname{Fix}}\left(T_i\right)$ because we also know that $x_n\rightharpoonup x$ as $n\to\infty$. Therefore $x\in\widehat{F}$, which proves that $\widehat{\operatorname{Fix}}\left(T_B\right)\subset\widehat{F}$, as claimed. □

Now we prove that the block operator of strictly BSNE operators also is a strictly BSNE operator (*cf.* [74, Proposition 14, page 18]).

Proposition 2.1.18 (Block operator of strictly BSNE operators)**.** *Let $f:X\to\mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Assume that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom}f^*$. *If each T_i, $i=1,2,\ldots,N$, is a strictly BSNE operator from X into itself, and the sets*

$$\widehat{F}:=\bigcap\left\{\ \widehat{\operatorname{Fix}}\left(T_i\right):1\leqslant i\leqslant N\right\}$$

and $\widehat{\operatorname{Fix}}\left(T_B\right)$ *are not empty, then T_B is also strictly BSNE.*

Proof. If $u\in\widehat{\operatorname{Fix}}\left(T_B\right)$, then $u\in\widehat{F}$ by Proposition 2.1.17. Therefore the fact that each T_i, $i=1,2,\ldots,N$, is strictly BSNE, with respect to $\widehat{\operatorname{Fix}}\left(T_i\right)$, implies that (1.3.9) holds for T_B and any $x\in X$.

Now we assume that there exists a bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ in X such that

$$\lim_{n\to\infty}\left(D_f\left(u,x_n\right)-D_f\left(u,T_Bx_n\right)\right)=0$$

and therefore, as we proved in Proposition 2.1.17, we get

$$\lim_{n\to\infty}\left(D_f\left(u,x_n\right)-D_f\left(u,T_ix_n\right)\right)=0$$

for any $i\in\{1,2,\ldots,N\}$. Since each T_i, $i=1,2,\ldots,N$, is strictly BSNE and $u\in\widehat{\operatorname{Fix}}\left(T_B\right)\subset$

$\widehat{\mathrm{Fix}}\,(T_i)$, it follows from (1.3.10) and (1.3.11) that

$$\lim_{n\to\infty} D_f\left(T_i x_n, x_n\right) = 0.$$

Since f is totally convex and $\{x_n\}_{n\in\mathbb{N}}$ is bounded, it follows from Proposition 1.2.46 that

$$\lim_{n\to\infty} \left\| T_i x_n - x_n \right\| = 0.$$

Since f is bounded and uniformly Fréchet differentiable on bounded subsets of X, it follows from Proposition 1.1.22(ii) that ∇f is uniformly continuous on bounded subsets of X and thus

$$\lim_{n\to\infty} \left\| \nabla f\left(T_i x_n\right) - \nabla f\left(x_n\right) \right\|_* = 0.$$

By the definition of the block operator (see (2.1.14)), we have

$$\nabla f\left(T_B x_n\right) - \nabla f\left(x_n\right) = \sum_{i=1}^{N} w_i \left(\nabla f\left(T_i x_n\right) - \nabla f\left(x_n\right)\right)$$

and therefore

$$\lim_{n\to\infty} \left\| \nabla f\left(T_B x_n\right) - \nabla f\left(x_n\right) \right\|_* = 0. \tag{2.1.21}$$

On the other hand, from the definition of the Bregman distance (see (1.2.1)) we obtain that

$$D_f\left(T_B x_n, x_n\right) + D_f\left(x_n, T_B x_n\right) = \left\langle \nabla f\left(T_B x_n\right) - \nabla f\left(x_n\right), T_B x_n - x_n \right\rangle. \tag{2.1.22}$$

Note that each sequence $\{T_i x_n\}_{n\in\mathbb{N}}$, $i = 1, 2, \ldots, N$, is bounded because so is the sequence $\{x_n\}_{n\in\mathbb{N}}$ and $\lim_{n\to\infty} \|T_i x_n - x_n\| = 0$. Since ∇f and ∇f^* are bounded on bounded subsets of X and $\mathrm{int}\,\mathrm{dom}\, f^*$, respectively, it follows that $\{T_B x_n\}_{n\in\mathbb{N}}$ is bounded too. Whence, combining (2.1.21) and (2.1.22), we deduce that

$$\lim_{n\to\infty} \left(D_f\left(T_B x_n, x_n\right) + D_f\left(x_n, T_B x_n\right)\right) = 0.$$

Therefore

$$\lim_{n\to\infty} D_f\left(T_B x_n, x_n\right) = 0.$$

This means that (1.3.10) implies (1.3.11) for T_B and this proves that T_B is strictly BSNE, as required. □

When we generate a block operator from properly BSNE operators, we have that its fixed point set is the intersection of the fixed point sets of the operators generating the block operator (*cf.* [74, Proposition 15, page 19]).

Proposition 2.1.19 (Block operator of properly BSNE operators). *Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Assume that ∇f^* is bounded on bounded subsets of* $\operatorname{int} \operatorname{dom} f^*$. *If each T_i, $i = 1, 2, \ldots, N$, is a properly BSNE operator from X into itself, and the set*

$$F := \bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$$

is not empty, then T_B is also properly BSNE and $F = \operatorname{Fix}(T_B)$.

Proof. On the one hand, since each T_i, $i = 1, 2, \ldots, N$, is properly BSNE, it is also properly QBNE (see Figure 1.3). Then the fact that $F \neq \varnothing$ makes it possible to apply Proposition 2.1.16 so that $F = \operatorname{Fix}(T_B)$ and T_B is properly QBNE, that is, it satisfies inequality (1.3.9) for any $p \in \operatorname{Fix}(T_B)$.

On the other hand, given a bounded sequence $\{x_n\}_{n \in \mathbb{N}}$ such that, for any $u \in \operatorname{Fix}(T_B)$, we have

$$\lim_{n \to \infty} \left(D_f(u, x_n) - D_f(u, T_B x_n)\right) = 0,$$

analogously to the argument used in Proposition 2.1.18, one is able to deduce that

$$\lim_{n \to \infty} D_f(T_B x_n, x_n) = 0.$$

Thus T_B is indeed properly BSNE, as asserted. □

2.1.1 Characterization of BFNE Operators

In this section we establish a characterization of BFNE operators. This characterization emphasizes the strong connection between the nonexpansivity of T and the monotonicity of S_T, where

$$S_T := \nabla f - (\nabla f) \circ T. \tag{2.1.23}$$

Results in this direction have been known for a long time. We cite the one of Rockafellar [100] from 1976 and the one of Bauschke, Wang and Yao [15] from 2008.

Proposition 2.1.20 (Characterization of firmly nonexpansive operators). *Let K be a subset of a Hilbert space $\mathcal{H}$ and let $T : K \to \mathcal{H}$ be an operator. Then T is firmly nonexpansive if and only if $I - T$ is T-monotone.*

Proposition 2.1.21 (Property of BFNE operators). *Let K be a subset of X and let $T : K \to X$ be an operator. Fix an admissible function $f : X \to \mathbb{R}$ and set*

$$A_T := \nabla f \circ T^{-1} - \nabla f.$$

If T is BFNE, then A_T is monotone (this operator is not necessarily single-valued).

Motivated by these results, we offer the following characterization (*cf.* [25, Theorem 3.3, page 167]).

Theorem 2.1.22 (Characterization of BFNE operators)**.** *Let $K \subset \operatorname{int}\operatorname{dom} f$ and suppose that $T : K \to \operatorname{int}\operatorname{dom} f$ for an admissible function $f : X \to (-\infty, +\infty]$. Then T is BFNE if and only if $S_T = \nabla f - (\nabla f) \circ T$ is T-monotone.*

Proof. Suppose that T is BFNE (see Definition 1.3.5(i)). Take x, y in K and denote $\xi = S_T(x)$ and $\eta = S_T(y)$. Then by the definition of S_T (see (2.1.23)) we obtain

$$\nabla f(Tx) = \nabla f(x) - \xi \quad \text{and} \quad \nabla f(Ty) = \nabla f(y) - \eta. \tag{2.1.24}$$

Since T is BFNE, we have from (1.3.4) that

$$\langle \nabla f(Tx) - \nabla f(Ty), Tx - Ty \rangle \leqslant \langle \nabla f(x) - \nabla f(y), Tx - Ty \rangle. \tag{2.1.25}$$

Now, substituting (2.1.24) on the left-hand side of (2.1.25), we obtain

$$\langle (\nabla f(x) - \xi) - (\nabla f(y) - \eta), Tx - Ty \rangle \leqslant \langle \nabla f(x) - \nabla f(y), Tx - Ty \rangle,$$

which means that

$$0 \leqslant \langle S_T(x) - S_T(y), Tx - Ty \rangle.$$

Thus S_T is T-monotone (see Definition 1.4.24). Conversely, if S_T is T-monotone, then

$$0 \leqslant \langle S_T(x) - S_T(y), Tx - Ty \rangle$$

for any $x, y \in K$ and therefore from (2.1.23) we have

$$0 \leqslant \langle (\nabla f(x) - \nabla f(Tx)) - (\nabla f(y) - \nabla f(Ty)), Tx - Ty \rangle,$$

which means that

$$\langle \nabla f(Tx) - \nabla f(Ty), Tx - Ty \rangle \leqslant \langle \nabla f(x) - \nabla f(y), Tx - Ty \rangle$$

for any $x, y \in K$. In other words, T is indeed a BFNE operator. □

Remark 2.1.23 (Theorem 2.1.22 implies Proposition 2.1.20)**.** *It is clear that when X is a Hilbert space and $f = \|\cdot\|^2$, BFNE operators are firmly nonexpansive operators (see Remark 1.3.10) and in this case $S_T = I - T$. Therefore Proposition 2.1.20 is an immediate consequence of Theorem 2.1.22.* ◇

Remark 2.1.24 (Theorem 2.1.22 implies Proposition 2.1.21)**.** *If T is a BFNE operator, then S_T is T-monotone by Theorem 2.1.22. Take $\xi \in A_T(x)$ and $\eta \in A_T(y)$. From the definition of A_T (see* (2.1.21)*) we get $\xi = \nabla f(z) - \nabla f(x)$, where $Tz = x$, and $\eta = \nabla f(w) - \nabla f(y)$, where $Tw = y$. Hence*

$$\begin{aligned}\langle \xi - \eta, x - y\rangle &= \langle (\nabla f(z) - \nabla f(Tz)) - (\nabla f(w) - \nabla f(Tw)), Tz - Tw\rangle \\ &= \langle S_T(z) - S_T(w), Tz - Tw\rangle \\ &\geqslant 0\end{aligned}$$

for all $x, y \in \operatorname{dom} A_T$, and so A_T is monotone (see Definition 1.4.2(i)). Hence Proposition 2.1.21 follows from Theorem 2.1.22. ◇

Motivated by our characterization (see Theorem 2.1.22), we now show that the converse implication of Proposition 2.1.21 is also true (*cf.* [25, Proposition 3.6, page 168]).

Proposition 2.1.25 (Another characterization of BFNE operators)**.** *Let $K \subset \operatorname{int}\operatorname{dom} f$ and suppose that $T : K \to \operatorname{int}\operatorname{dom} f$ for an admissible function $f : X \to \mathbb{R}$. The mapping A_T is monotone if and only if T is BFNE.*

Proof. If T is BFNE then from Proposition 2.1.25 we get that A_T is monotone (see Definition 1.4.2(i)). Conversely, suppose that A_T is monotone. Then for any $x, y \in \operatorname{dom} A_T$, we have

$$0 \leqslant \langle \xi - \eta, x - y\rangle$$

for any $\xi \in A_T(x)$ and $\eta \in A_T(y)$. Let $w, z \in K$. Set $\xi = \nabla f(z) - \nabla f(x)$, where $Tz = x$, and $\eta = \nabla f(w) - \nabla f(y)$, where $Tw = y$. We have from the monotonicity of A_T (see (1.4.1)) that

$$0 \leqslant \langle (\nabla f(z) - \nabla f(x)) - (\nabla f(w) - \nabla f(y)), x - y\rangle,$$

which means that

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \leqslant \langle \nabla f(z) - \nabla f(w), x - y\rangle.$$

Thus

$$\langle \nabla f(Tz) - \nabla f(Tw), Tz - Tw\rangle \leqslant \langle \nabla f(z) - \nabla f(w), Tz - Tw\rangle$$

and so T is a BFNE operator (see Definition 1.3.5(i)), as asserted. □

Remark 2.1.26 (Comparison between these two characterizations)**.** *Our characterization of BFNE operators is based on a new type of monotonicity, the T-monotonicity (see Definition 1.4.24), which seems to be harder to check than the classical monotonicity (see*

Definition 1.4.2(i)). On the other hand, our mapping S_T is defined without any inverse operation, and hence is easier to compute. In the case of the mapping A_T, similar computations seem to be much harder because of the presence of the inverse operator T^{-1}. ◇

2.1.2 Examples of BFNE Operators in Euclidean Spaces

In this section we use Theorem 2.1.22 to present various examples of BFNE operators in Euclidean spaces. Indeed, we have already seen that BFNE operators can be generated from T-monotone mappings. Moreover, the notion of T-monotonicity can be simplified in the case of the real line.

Remark 2.1.27 (The real line case)**.** *If $X = \mathbb{R}$ and both T and S_T are increasing (decreasing), then S_T is T-monotone.* ◇

The next remark allows us to explicitly produce BFNE operators.

Remark 2.1.28 (Characterization of BFNE operators on the real line)**.** *Let $f : \mathbb{R} \to (-\infty, +\infty]$ be an admissible function and let K be a nonempty subset of $\operatorname{int}\operatorname{dom} f$. From Theorem 2.1.22 we know that an increasing (decreasing) operator T is BFNE if S_T is also increasing (decreasing). If, in addition, T is differentiable on $\operatorname{int} K$, then $S_T' = f'' - f''(T)\,T'$.*

We conclude that a differentiable operator $T : K \to \operatorname{int}\operatorname{dom} f$ is BFNE on K with respect to an admissible twice-differentiable function f as soon as

$$0 \leqslant T'(x) \leqslant \frac{f''(x)}{f''(T(x))}$$

for all $x \in \operatorname{int} K$. ◇

The following result gives sufficient conditions for an operator T to be BFNE with respect to the Boltzmann-Shannon entropy $\mathcal{BS}$ (see (1.2.8)). We use the term $\mathcal{BS}$-BFNE for operators $T : K \to (0, +\infty)$ which are BFNE with respect to $\mathcal{BS}$ (*cf.* [25, Proposition 4.12, page 174]).

Proposition 2.1.29 (Conditions for $\mathcal{BS}$-BFNE)**.** *Let K be a nonempty subset of $(0, +\infty)$ and let $T : K \to (0, +\infty)$ be an operator. Assume that one of the following conditions holds.*

(i) *T is increasing and $T(x)/x$ is decreasing for every $x \in \operatorname{int} K$.*

(ii) *T is differentiable on $\operatorname{int} K$ and its derivative T' satisfies*

$$0 \leqslant T'(x) \leqslant \frac{T(x)}{x} \tag{2.1.26}$$

for every $x \in \operatorname{int} K$.

(iii) T *is decreasing and* $T(x)/x$ *is increasing for every* $x \in \operatorname{int} K$.

(iv) T *is differentiable on* $\operatorname{int} K$ *and its derivative* T' *satisfies*

$$\frac{T(x)}{x} \leqslant T'(x) \leqslant 0$$

for every $x \in \operatorname{int} K$.

Then T *is an* $\mathcal{BS}$*-BFNE operator on* K.

Proof. This result follows immediately from Theorem 2.1.22 and Remark 2.1.28. □

Remark 2.1.30. *The only solution of the differential equation*

$$T'(x) = \frac{T(x)}{x}$$

is $T(x) = \alpha x$ *for any* $\alpha \in \mathbb{R}$, *but in our case* $\alpha \in (0, +\infty)$ *since* $T(x) \in (0, +\infty)$ *for any* $x \in K \subset (0, +\infty)$. ◇

Using the conditions provided in Proposition 2.1.29, we give examples of $\mathcal{BS}$-BFNE operators (*cf.* [25, Example 4.14, page 174]).

$T(x)$	Domain
$\alpha x + \beta \quad \alpha, \beta \in (0, +\infty)$	$(0, +\infty)$
$x^p \quad p \in (0, 1]$	$(0, +\infty)$
$\alpha x - x^p \quad p \in [1, +\infty), \alpha \in (0, +\infty)$	$\left(0, (\alpha/p)^{1/(p-1)} \infty\right)$
$\alpha \log x \quad \alpha \in [0, +\infty)$	$[e, +\infty)$
$\sin x$	$(0, \pi/2]$
$\alpha e^x \quad \alpha \in (0, +\infty)$	$(0, 1]$

Table 2.1: Examples of $\mathcal{BS}$-BFNE operators

The following result gives sufficient conditions for an operator T to be BFNE with respect to the Fermi-Dirac entropy $\mathcal{FD}$ (see (1.2.9)). We use the term $\mathcal{FD}$-BFNE for operators $T : K \to (0, 1)$ which are BFNE with respect to $\mathcal{FD}$ (*cf.* [25, Proposition 4.16, page 176]).

Proposition 2.1.31 (Conditions for $\mathcal{FD}$-BFNE)**.** *Let* K *be a nonempty subset of* $(0, 1)$ *and let* $T : K \to (0, 1)$ *be an operator. Assume that one of the following conditions holds.*

(i) *T is increasing and*

$$\frac{T(x)(1-x)}{x(1-T(x))}$$

is decreasing for every $x \in \operatorname{int} K$.

(ii) *T is differentiable and its derivative T' satisfies*

$$0 \leqslant T'(x) \leqslant \frac{T(x)(1-T(x))}{x(1-x)}$$

for every $x \in \operatorname{int} K$.

(iii) *T is decreasing and*

$$\frac{T(x)(1-x)}{x(1-T(x))}$$

is increasing for every $x \in \operatorname{int} K$.

(iv) *T is differentiable on $\operatorname{int} K$ and its derivative T' satisfies*

$$\frac{T(x)(1-T(x))}{x(1-x)} \leqslant T'(x) \leqslant 0$$

for every $x \in \operatorname{int} K$.

Then T is an $\mathcal{FD}$-BFNE operator on K.

Proof. This result follows immediately from Theorem 2.1.22 and Remark 2.1.28. □

Remark 2.1.32. *The only solution of the differential equation*

$$T'(x) = \frac{T(x)(1-T(x))}{x(1-x)}$$

is

$$T(x) = \frac{\alpha x}{(1-x+\alpha x)}$$

for any $\alpha \in \mathbb{R}$, but in our case $\alpha \in (0,+\infty)$ since $T(x) \in (0,1)$ for any $x \in K \subset (0,1)$. ◇

Using Proposition 2.1.31, we now give examples of $\mathcal{FD}$-BFNE operators (*cf.* [25, Example 4.18, page 177]).

In the following table we summarize sufficient conditions on the operator T to be BFNE with respect to various choices of functions f.

Remark 2.1.33 (Product constructions)**.** *For each $i = 1, 2, \ldots, n$, let $f_i : \mathbb{R} \to \mathbb{R}$ be an admissible function, and define the function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}$ by*

$$\mathbf{F}(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i).$$

$T(x)$	Domain
$\alpha \quad \alpha \in (0,1)$	$(0,1)$
$\alpha x \quad \alpha \in (0,1)$	$(0,1)$
$x^p \quad p \in (0,1]$	$(0,1)$
$\sin x$	$(0,1]$

Table 2.2: Examples of $\mathcal{FD}$-BFNE operators

$f(x)$	Domain	Condition
$\mathcal{BS}(x)$	$(0,+\infty)$	$0 \leqslant T'(x) \leqslant \frac{T(x)}{x}$
$\mathcal{FD}(x)$	$-(0,1)$	$0 \leqslant T'(x) \leqslant \frac{T(x)(1-T(x))}{x(1-x)}$
$\cosh x$	$\mathbb{R}$	$0 \leqslant T'(x) \leqslant \frac{\cosh(x)}{\cosh(T(x))}$
$x^2/2$	$\mathbb{R}$	$0 \leqslant T'(x) \leqslant 1$
$x^4/4$	$\mathbb{R}$	$0 \leqslant T'(x) \leqslant \frac{x^2}{(T(x))^2}$
e^x	$\mathbb{R}$	$0 \leqslant T'(x) \leqslant \frac{e^x}{e^{T(x)}}$
$-\log(x)$	$(0,+\infty)$	$0 \leqslant T'(x) \leqslant \frac{(T(x))^2}{x^2}$

Table 2.3: Conditions for T to be a BFNE operator

For each $i = 1, 2, \ldots, n$, *let* K_i *be a nonempty subset of* $\operatorname{int}\operatorname{dom} f_i$. *Let* $\mathbf{T} : \times_{i=1}^n K_i \to \times_{i=1}^n \operatorname{int}\operatorname{dom} f_i$ *be an operator which is defined by* $\mathbf{T} = (T_1, \ldots, T_n)$, *where* $T_i : K_i \to \operatorname{int}\operatorname{dom} f_i$ *for each* $1 \leqslant i \leqslant n$. *If each* T_i, $i = 1, \ldots, n$, *satisfies the hypotheses of Theorem 2.1.22, then the operator* T *is BFNE with respect to* F *on* $\times_{i=1}^n K_i$. $\diamond$

Chapter 3

Iterative Methods for Approximating Fixed Points

In this section the function f is always assumed to be admissible (see Definition 1.2.1). Let K be a nonempty, closed and convex subset of a Banach space X and let $T : K \to K$ be an operator. Iterative methods are often used to solve the fixed point equation $Tx = x$. The most well-known method is perhaps the *Picard successive iteration* method when T is a strict contraction (see (1.3.1)). Picard's iterative method generates a sequence $\{x_n\}_{n \in \mathbb{N}}$ successively by the following algorithm.

Picard Iterative Method
Initialization: $x_0 \in K$.
General Step ($n = 1, 2, \ldots$):

$$x_{n+1} = Tx_n. \tag{3.0.1}$$

A sequence generated by the Picard iterative method converges in norm to the unique fixed point of T. However, if T is not a strict contraction (for instance, if T is nonexpansive (see (1.3.1)) even with a unique fixed point), then Picard's successive iteration fails, in general, to converge. It suffices, for example, to take for T a rotation of the unit disk in the plane around the origin of coordinates (see, for example, [72]). Krasnoselski [70], however, has shown that in this example, one can obtain a convergent sequence of successive approximations if instead of T one takes the auxiliary nonexpansive operator $(1/2)(I + T)$, where I denotes the identity operator of X, *i.e.*, a sequence of successive approximations which is defined by the following algorithm.

Krasnoselski Iterative Method
Initialization: $x_0 \in K$.
General Step ($n = 1, 2, \ldots$):

$$x_{n+1} = \frac{1}{2}(I + T)x_n. \tag{3.0.2}$$

It is easy to see that the operators T and $(1/2)(I+T)$ have the same fixed point set, so that the limit of a convergent sequence defined by Algorithm (3.0.2) is necessarily a fixed point of T.

However, a more general iterative scheme is the following (see [72]).

Mann Iterative Method
Input: $\{\alpha_n\}_{n\in\mathbb{N}} \subset (0,1)$.
Initialization: $x_0 \in K$.
General Step $(n = 1, 2, \ldots)$**:**

$$x_{n+1} = \alpha_n x_n + (1-\alpha_n) T x_n. \tag{3.0.3}$$

In an infinite-dimensional Hilbert space, the Mann iterative method has only weak convergence, in general, even for nonexpansive operators (see [14, 55]). Therefore, many authors have tried to modify Mann's iteration process in order to obtain strong convergence for nonexpansive operators (see also [58]). One way to get strong convergence in Hilbert spaces is to use the method proposed by Haugazeau in [60].

Haugazeau Iterative Method
Initialization: $x_0 \in \mathcal{H}$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n = T x_n, \\ H_n = \{z \in \mathcal{H} : \langle x_n - y_n, y_n - z\rangle \geqslant 0\}, \\ Q_n = \{z \in \mathcal{H} : \langle x_n - z, x_0 - x_n\rangle \geqslant 0\}, \\ x_{n+1} = P_{H_n \cap Q_n}(x_0). \end{cases} \tag{3.0.4}$$

Haugazeau proved that a sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (3.0.4) converges strongly to a fixed point of T. Later many authors studied and developed this method (in the context of Hilbert spaces see, for example, [9, 105], and in Banach spaces see, for example, [10, 48, 54]).

In the next sections we present several methods for finding fixed points of operators in reflexive Banach space which generalize previously mentioned results. We focus our study on explicit methods (which we call iterative methods or algorithms) except for one result about approximation of fixed point for BFNE operators by an implicit method (see Theorem 3.4.1). Our algorithms allow for computational errors in some cases and find common fixed points of finitely many operators.

3.1 Picard's Iteration for Bregman Nonexpansive Operators

The main result in this section is the following one (*cf.* [74, Theorem 4.1, page 12]).

Theorem 3.1.1 (Picard iteration)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function such that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty, closed and* *convex subset of* $\operatorname{int}\operatorname{dom} f$ *and let* $T : K \to K$ *be a strictly QBNE operator. Then the following assertions hold.*

(i) *If* $\widehat{\operatorname{Fix}}(T)$ *is nonempty, then* $\{T^n x\}_{n\in\mathbb{N}}$ *is bounded for each* $x \in K$.

(ii) *If, furthermore,* T *is asymptotically regular, then, for each* $x \in K$*, any weak subsequential limit of* $\{T^n x\}_{n\in\mathbb{N}}$ *belongs to* $\widehat{\operatorname{Fix}}(T)$.

(iii) *If, furthermore,* ∇f *is weakly sequentially continuous, then* $\{T^n x\}_{n\in\mathbb{N}}$ *converges weakly to an element in* $\widehat{\operatorname{Fix}}(T)$ *for each* $x \in K$.

Proof. (i) This result follows directly from Proposition 2.1.4.

(ii) Since $\{T^n x\}_{n\in\mathbb{N}}$ is bounded (by assertion (i)), there is a subsequence $\{T^{n_k} x\}_{k\in\mathbb{N}}$ which converges weakly to some u. Define $x_n = T^n x$ for any $n \in \mathbb{N}$. Since T is asymptotically regular, it follows from Definition 1.3.14 that $\|x_n - Tx_n\| \to 0$ as $n \to \infty$. Therefore we have both $x_{n_k} \rightharpoonup u$ and $\|x_{n_k} - Tx_{n_k}\| \to 0$ as $k \to \infty$, which means that $u \in \widehat{\operatorname{Fix}}(T)$.

(iii) From assertion (ii) and since T is strictly QBNE, we already know (part of the proof of Proposition 2.1.4) that the limit $\lim_{n\to\infty} D_f(u, T^n x)$ exists for any weak subsequential limit u of the sequence $\{T^n x\}_{n\in\mathbb{N}}$. The result now follows immediately from Proposition 1.2.53. □

Corollary 3.1.2 (Picard iteration for BSNE operators). *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre function which is totally convex on bounded subsets of* X*. Suppose that* ∇f *is weakly sequentially continuous and* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let* K *be a nonempty, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$*. Let* $T : K \to K$ *be a BSNE operator such that* $\operatorname{Fix}(T) = \widehat{\operatorname{Fix}}(T) \neq \varnothing$*. Then* $\{T^n x\}_{n\in\mathbb{N}}$ *converges weakly to an element in* $\operatorname{Fix}(T)$ *for each* $x \in K$.

Proof. The result follows immediately from Theorem 3.1.1 and Proposition 2.1.14. □

Remark 3.1.3 (The case $\operatorname{Fix}(T) \neq \widehat{\operatorname{Fix}}(T)$). *If* $\operatorname{Fix}(T) \neq \widehat{\operatorname{Fix}}(T)$*, but* $\widehat{\operatorname{Fix}}(T) \neq \varnothing$*, then we only know that, for a strictly BSNE operator* T*,* $\{T^n x\}_{n\in\mathbb{N}}$ *converges weakly to an element in* $\widehat{\operatorname{Fix}}(T)$ *for each* $x \in K$*. This result was previously proved in [88, Lemma 4, page 315] under somewhat different assumptions.* ◇

Remark 3.1.4 (Picard iteration for BFNE operators). *Let* $f : X \to \mathbb{R}$ *be a function which is uniformly Fréchet differentiable and bounded on bounded subsets of* X*. From Proposition 2.1.2 and Corollary 3.1.2 we get that Theorem 3.1.1 holds for BFNE operators. It is well known that in Hilbert spaces, the Picard iteration of firmly nonexpansive operators (see* (1.3.12)*) converges weakly to a fixed point of the operator (see, for instance, [56]).* ◇

Remark 3.1.5 (Common fixed point - composition case)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Let K be a nonempty, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$.

Let $\{T_i : 1 \leqslant i \leqslant N\}$ be N BSNE operators such that $\widehat{\operatorname{Fix}}(T_i) = \operatorname{Fix}(T_i) \neq \varnothing$ for each $1 \leqslant i \leqslant N$ and let $T = T_N \circ T_{N-1} \circ \cdots \circ T_1$. From Proposition 2.1.12 and Remark 2.1.13 we obtain that if $\bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\} \neq \varnothing$, then T is also BSNE such that $\widehat{\operatorname{Fix}}(T) = \operatorname{Fix}(T) = \bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$.

From Theorem 3.1.1 we now get that $\{T^n x\}_{n\in\mathbb{N}}$ converges weakly to a common fixed point of the family of BSNE operators. Similarly, if we just assume that each T_i is strictly BSNE, with $\widehat{\operatorname{Fix}}(T_i) \neq \varnothing$, $1 \leqslant i \leqslant N$, then we get weak convergence of the sequence $\{T^n x\}_{n\in\mathbb{N}}$ to a common asymptotic fixed point. ◇

As a consequence of the previous result, we now see that the Picard iteration provides a method for approximating common fixed points of a finite family of BSNE operators. We can also use the block operator (see Definition 2.1.15) for finding common fixed point.

Remark 3.1.6 (Common fixed point - block operator case)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$.

Let $\{T_i : 1 \leqslant i \leqslant N\}$ be N BSNE operators such that $\widehat{\operatorname{Fix}}(T_i) = \operatorname{Fix}(T_i) \neq \varnothing$ and let T_B be the block operator defined by (2.1.14). *If $F := \bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$ and* $\operatorname{Fix}(T_B)$ *are nonempty, then from Proposition 2.1.17 we know that T_B is BSNE. Furthermore, from Proposition 2.1.16 we get that*

$$\operatorname{Fix}(T_B) \subset \widehat{\operatorname{Fix}}(T_B) \subset F \subset \operatorname{Fix}(T_B),$$

which implies that $\widehat{\operatorname{Fix}}(T_B) = \operatorname{Fix}(T_B) \neq \varnothing$.

Therefore, Theorem 3.1.1 applies to guarantee that $\{T_B^n x\}_{n\in\mathbb{N}}$ converges weakly to an element in F under appropriate conditions. ◇

3.2 Mann's Iteration for Bregman Nonexpansive Operators

In this section we study a modification of the Mann iterative method (see Algorithm (3.0.3)), which is defined by using convex combinations with respect to a convex function f, a concept which was first introduced in the case of Euclidean spaces in [46].

f-Mann Iterative Method
Input: $f : X \to \mathbb{R}$ and $\{\alpha_n\}_{n\in\mathbb{N}} \subset (0,1)$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f\left(x_n\right) + (1-\alpha_n)\nabla f\left(Tx_n\right)\right). \tag{3.2.1}$$

Remark 3.2.1 (Particular case). *When the Banach space X is a Hilbert space and $f = (1/2)\left\|\cdot\right\|^2$ then $\nabla f = \nabla f^* = I$ and the f-Mann iterative method is exactly the Mann iterative method (see Algorithm* (3.0.3)*).* ◇

In the following result we prove weak convergence of the sequence generated by the f-Mann iterative method (*cf.* [74, Theorem 5.1, page 13]).

Theorem 3.2.2 (f-Mann iteration). *Let $T : X \to X$ be a strictly BSNE operator with $\widehat{\mathrm{Fix}}(T) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence generated by the f-Mann iterative method (see Algorithm* (3.2.1)*) where $\{\alpha_n\}_{n\in\mathbb{N}} \subset [0,1]$ satisfies $\limsup_{n\to\infty}\alpha_n < 1$. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in $\widehat{\mathrm{Fix}}(T)$.*

Proof. We divide the proof into 3 steps.

Step 1. *The sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded.*

Let $p \in \widehat{\mathrm{Fix}}(T)$. From Proposition 1.2.42(i) and (ii), and the first part of the definition of strictly BSNE operator (see (1.3.9)) we have for all $n \in \mathbb{N}$,

$$\begin{aligned}
D_f\left(p, x_{n+1}\right) &= D_f\left(p, \nabla f^*\left(\alpha_n \nabla f\left(x_n\right) + \left(1-\alpha_n\right)\nabla f\left(Tx_n\right)\right)\right) \\
&= W^f\left(\alpha_n \nabla f\left(x_n\right) + \left(1-\alpha_n\right)\nabla f\left(Tx_n\right), p\right) \\
&\leqslant \alpha_n W^f\left(\nabla f\left(x_n\right), p\right) + \left(1-\alpha_n\right) W^f\left(\nabla f\left(Tx_n\right), p\right) \\
&= \alpha_n D_f\left(p, x_n\right) + \left(1-\alpha_n\right) D_f\left(p, Tx_n\right) \\
&\leqslant \alpha_n D_f\left(p, x_n\right) + \left(1-\alpha_n\right) D_f\left(p, x_n\right) \\
&= D_f\left(p, x_n\right).
\end{aligned} \tag{3.2.2}$$

This shows that the nonnegative sequence $\{D_f\left(p, x_n\right)\}_{n\in\mathbb{N}}$ is decreasing, thus bounded, and $\lim_{n\to\infty} D_f\left(p, x_n\right)$ exists. From Proposition 1.2.48 we obtain that $\{x_n\}_{n\in\mathbb{N}}$ is bounded, as claimed.

Step 2. *Every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to $\widehat{\mathrm{Fix}}(T)$.*

For any $p \in \widehat{\mathrm{Fix}}(T)$ we see, from the first inequality of (3.2.2), that

$$D_f(p, x_{n+1}) \leqslant D_f(p, x_n) + (1 - \alpha_n)(D_f(p, Tx_n) - D_f(p, x_n)).$$

Hence

$$(1 - \alpha_n)(D_f(p, x_n) - D_f(p, Tx_n)) \leqslant D_f(p, x_n) - D_f(p, x_{n+1}) \tag{3.2.3}$$

for all $n \in \mathbb{N}$. We already know that $\lim_{n\to\infty} D_f(p, x_n)$ exists. Since $\limsup_{n\to\infty} \alpha_n < 1$, it follows that

$$\lim_{n\to\infty} (D_f(p, x_n) - D_f(p, Tx_n)) = 0.$$

Now, since T is strictly BSNE and $p \in \widehat{\mathrm{Fix}}(T)$, we obtain

$$\lim_{n\to\infty} D_f(Tx_n, x_n) = 0.$$

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded (see Step 1), Proposition 1.2.46 implies that

$$\lim_{n\to\infty} \|Tx_n - x_n\| = 0.$$

Therefore, if there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ which converges weakly to some $v \in X$ as $k \to \infty$, then $v \in \widehat{\mathrm{Fix}}(T)$.

Step 3. *The sequence $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in $\widehat{\mathrm{Fix}}(T)$.*

Since ∇f is weakly sequentially continuous (see Definition 1.2.52), the result follows immediately from Proposition 1.2.53 since $\lim_{n\to\infty} D_f(u, x_n)$ exists for any weak subsequential limit u of the sequence $\{x_n\}_{n\in\mathbb{N}}$ by Step 2. □

Corollary 3.2.3 (f-Mann iteration for BSNE operators)**.** *Let $T : X \to X$ be a BSNE operator such that $\mathrm{Fix}(T) = \widehat{\mathrm{Fix}}(T) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\mathrm{int}\,\mathrm{dom}\, f^*$. *Let $\{x_n\}_{n\in\mathbb{N}}$ be the sequence generated by Algorithm* (3.2.1)*, where $\{\alpha_n\}_{n\in\mathbb{N}} \subset [0, 1]$ satisfies $\limsup_{n\to\infty} \alpha_n < 1$. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in* $\mathrm{Fix}(T)$.

Remark 3.2.4 (Particular case of f-Mann iteration)**.** *If $f = (1/2)\|\cdot\|^2$ and X is a Hilbert space, then both ∇f and ∇f^* are the identity operator, and Algorithm 3.2.1 coincides with the Mann iteration (see Algorithm* (3.0.3)*). In this case the weak convergence of which for nonexpansive operators is well known, even in more general Banach spaces, under the assumption that $\sum_{n\in\mathbb{N}} \alpha_n(1 - \alpha_n) = \infty$ (see [84]).* ◇

Remark 3.2.5 (Common fixed point - composition case)**.** *Let $f : X \to \mathbb{R}$ be a Legendre* *function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X*. Suppose that* ∇f *is weakly sequentially continuous and* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*.*

Let $\{T_i : 1 \leqslant i \leqslant N\}$ *be* N *BSNE operators such that* $\widehat{\operatorname{Fix}}(T_i) = \operatorname{Fix}(T_i) \neq \varnothing$ *for each* $1 \leqslant i \leqslant N$ *and let* $T = T_N \circ T_{N-1} \circ \cdots \circ T_1$*. Then from Proposition 2.1.12 and Remark 2.1.13 we obtain that, if* $\bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\} \neq \varnothing$*, then* T *is also BSNE such that* $\widehat{\operatorname{Fix}}(T) = \operatorname{Fix}(T) = \bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$*.*

Now from Theorem 3.2.2 we get that a sequence $\{x_n\}_{n\in\mathbb{N}}$ *generated by Algorithm* (3.2.1) *for* $T = T_N \circ T_{N-1} \circ \cdots \circ T_1$ *converges weakly to an element in* $\bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$ *for each* $x_0 \in X$*.*

In the case where each T_i*,* $i = 1, \ldots, N$*, is strictly BSNE with* $\widehat{\operatorname{Fix}}(T_i) \neq \varnothing$*, a sequence* $\{x_n\}_{n\in\mathbb{N}}$ *generated by Algorithm* (3.2.1) *for* $T = T_N \circ T_{N-1} \circ \cdots \circ T_1$ *weakly converges to a common asymptotic fixed point of the family* $\{T_i : 1 \leqslant i \leqslant N\}$ *whenever such a point exists.* ◇

Remark 3.2.6 (Common fixed point - block operator case)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X*. Suppose that* ∇f *is weakly sequentially continuous and* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*.*

Let $\{T_i : 1 \leqslant i \leqslant N\}$ *be* N *BSNE operators such that* $\widehat{\operatorname{Fix}}(T_i) = \operatorname{Fix}(T_i) \neq \varnothing$ *and let* T_B *be the block operator defined by* (2.1.14)*. If* $F := \bigcap \{\operatorname{Fix}(T_i) : 1 \leqslant i \leqslant N\}$ *and* $\operatorname{Fix}(T_B)$ *are nonempty, then from Propositions 2.1.17 we know that* T_B *is BSNE. Furthermore, from Proposition 2.1.16 we get that*

$$\operatorname{Fix}(T_B) \subset \widehat{\operatorname{Fix}}(T_B) \subset F \subset \operatorname{Fix}(T_B),$$

which implies that $\widehat{\operatorname{Fix}}(T_B) = \operatorname{Fix}(T_B) \neq \varnothing$*.*

Therefore, Theorem 3.2.2 applies to guarantee the weak convergence of a sequence $\{x_n\}_{n\in\mathbb{N}}$ *generated by Algorithm* (3.2.1) *for* $T = T_B$ *to an element in* F*.* ◇

3.3 Haugazeau's Iteration for Bregman Nonexpansive Operators

Let $T : X \to X$ be an operator such that $\operatorname{Fix}(T) \neq \varnothing$. A first modification of Algorithm (3.0.4) to general reflexive Banach spaces has been proposed by Bauschke and Combettes [10]. More precisely, they have introduced the following algorithm (for a single operator).

f-Haugazeau Iterative Method
Input: $f : X \to \mathbb{R}$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} y_n = Tx_n, \\ H_n = \{z \in X : \langle \nabla f(x_n) - \nabla f(y_n), z - y_n \rangle \leqslant 0\}, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leqslant 0\}, \\ x_{n+1} = \operatorname{proj}_{H_n \cap Q_n}(x_0). \end{cases} \tag{3.3.1}$$

Remark 3.3.1 (Particular case)**.** *Where the Banach space is a Hilbert space and* $f = (1/2)\|\cdot\|^2$, *Algorithms* (3.0.4) *and* (3.3.1) *coincide.* ◇

Now we present our modification of Algorithm (3.3.1) for finding common fixed points of finitely many QBFNE operators (see (1.3.6)). Our algorithm allows for computational errors. More precisely, let $T_i : X \to X$, $i = 1, 2, \ldots, N$, be QBFNE operators and denote $F := \bigcap_{i=1}^{N} \operatorname{Fix}(T_i) \neq \varnothing$. We study the following algorithm.

Minimal Norm-Like Picard Iterative Method
Input: $f : X \to (-\infty, +\infty]$ and $\{e_n^i\}_{n \in \mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} y_n^i = T_i(x_n + e_n^i), \\ H_n^i = \{z \in X : \langle \nabla f(x_n + e_n^i) - \nabla f(y_n^i), z - y_n^i \rangle \leqslant 0\}, \\ H_n := \bigcap_{i=1}^{N} H_n^i, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leqslant 0\}, \\ x_{n+1} = \operatorname{proj}_{H_n \cap Q_n}(x_0). \end{cases} \tag{3.3.2}$$

Let $T : X \to X$ be an operator such that $\operatorname{Fix}(T) \neq \varnothing$. Another modification of Algorithm (3.0.4) in Hilbert spaces has been proposed by Bauschke and Combettes [9]. They introduce, for example, the following algorithm (see [9, Theorem 5.3(ii), page 257] for a single operator and $\lambda_n = 1/2$).

Bauschke-Combettes Iterative Method
Initialization: $x_0 \in \mathcal{H}$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} y_n = Tx_n, \\ C_n = \{z \in \mathcal{H} : \|y_n - z\| \leqslant \|x_n - z\|\}, \\ Q_n = \{z \in \mathcal{H} : \langle x_0 - x_n, z - x_n \rangle \leqslant 0\}, \\ x_{n+1} = P_{C_n \cap Q_n}(x_0). \end{cases} \tag{3.3.3}$$

We introduce the following modification of the Bauschke-Combettes iterative method.

f-Bauschke-Combettes Iterative Method
Input: $f : X \to (-\infty, +\infty]$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n = Tx_n, \\ C_n = \{z \in X : D_f(z, y_n) \leqslant D_f(z, x_n)\}, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leqslant 0\}, \\ x_{n+1} = \mathrm{proj}_{C_n \cap Q_n}(x_0). \end{cases} \tag{3.3.4}$$

Remark 3.3.2. *When the Banach space is a Hilbert space and $f = (1/2)\|\cdot\|^2$, Algorithms* (3.3.3) *and* (3.3.4) *coincide.* ◇

Now we present a modification of Algorithm (3.3.4) for finding common fixed points of finitely many QBNE operators. Our algorithm allows for computational errors. More precisely, let $T_i : X \to X$, $i = 1, 2, \ldots, N$, be N QBNE operators such that $F := \bigcap_{i=1}^{N} \mathrm{Fix}(T_i) \neq \varnothing$. We study the following algorithm.

Minimal Norm-Like Bauschke-Combettes Iterative Method
Input: $f : X \to \mathbb{R}$ and $\{e_n^i\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n^i = T_i(x_n + e_n^i), \\ C_n^i = \{z \in X : D_f(z, y_n^i) \leqslant D_f(z, x_n + e_n^i)\}, \\ C_n := \bigcap_{i=1}^{N} C_n^i, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n \rangle \leqslant 0\}, \\ x_{n+1} = \mathrm{proj}_{C_n \cap Q_n}(x_0). \end{cases} \tag{3.3.5}$$

3.3.1 Convergence Analysis

Since the proofs that these algorithms generate sequences which converge strongly to a common fixed point are somewhat similar, we first prove several lemmata which are common to all the proofs and then present the statements and the proofs of our main results.

In order to prove our lemmata, we consider two more general versions of these algorithms. More precisely, we consider the following two algorithms.

Input: $f : X \to \mathbb{R}$ and $\{e_n^i\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n^i = S_n^i\left(x_n + e_n^i\right), \\ H_n^i = \left\{z \in X : \left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), z - y_n^i \right\rangle \leqslant 0\right\}, \\ H_n := \bigcap_{i=1}^N C_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}_{H_n \cap Q_n}\left(x_0\right). \end{cases} \tag{3.3.6}$$

Input: $f : X \to \mathbb{R}$ and $\{e_n^i\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n^i = S_n^i\left(x_n + e_n^i\right), \\ C_n^i = \left\{z \in X : D_f\left(z, y_n^i\right) \leqslant D_f\left(z, x_n + e_n^i\right)\right\}, \\ C_n := \bigcap_{i=1}^N C_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}_{C_n \cap Q_n}\left(x_0\right). \end{cases} \tag{3.3.7}$$

Here $S_n^i : X \to X$ are given operators for each $i = 1, 2, \ldots, N$. All our lemmata are proved under several assumptions, which we summarize in the following condition.

Condition 1. *Let* $S_n^i : X \to X$*,* $i = 1, \ldots, N$ *and* $n \in \mathbb{N}$*, be QBNE operators such that* $\Omega := \bigcap_{n\in\mathbb{N}} \bigcap_{i=1}^N \operatorname{Fix}\left(S_n^i\right) \neq \varnothing$ *. Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X*. Suppose that* ∇f^* *is bounded on bounded subsets of* $\operatorname{int} \operatorname{dom} f^*$*. Assume that, for each* $i = 1, \ldots, N$*, the sequence of errors* $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ *satisfies* $\lim_{n\to\infty} \|e_n^i\| = 0$*.*

Now we prove a sequence of lemmata. We start by proving that both algorithms are well defined.

Lemma 3.3.3. *Assume, in addition to Condition 1, that each* $S_n^i : X \to X$*,* $i = 1, 2, \ldots, N$ *and* $n \in \mathbb{N}$*, is a QBFNE operator. Then Algorithm* (3.3.6) *is well defined.*

Proof. The point y_n^i is well defined for each $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$. Hence we only have to show that $\{x_n\}_{n\in\mathbb{N}}$ is well defined. To this end, we will prove that the Bregman projection onto $H_n \bigcap Q_n$ is well defined (see (1.2.14)), that is, we need to show that $H_n \bigcap Q_n$ is a nonempty, closed and convex subset of X for each $n \in \mathbb{N}$ (see Proposition 1.2.34). Let $n \in \mathbb{N}$. It is not difficult to check that H_n^i are closed half-spaces for any $i = 1, 2, \ldots, N$. Hence their intersection H_n is a closed polyhedral set. It is also obvious that Q_n is a closed half-space. Let $u \in \Omega$. For any $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$, we obtain from the definition of QBFNE operator (see (1.3.6)) that

$$\left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), u - y_n^i \right\rangle \leqslant 0,$$

which implies that $u \in H_n^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in H_n$. Thus $\Omega \subset H_n$ for any $n \in \mathbb{N}$. On the other hand, it is obvious that $\Omega \subset Q_0 = X$. Thus $\Omega \subset H_0 \bigcap Q_0$, and therefore $x_1 = \mathrm{proj}^f_{H_0 \cap Q_0}(x_0)$ is well defined. Now suppose that $\Omega \subset H_{n-1} \bigcap Q_{n-1}$ for some $n \geqslant 1$. Then $x_n = \mathrm{proj}^f_{H_{n-1} \cap Q_{n-1}}(x_0)$ is well defined because $H_{n-1} \bigcap Q_{n-1}$ is a nonempty, closed and convex subset of X. So from Proposition 1.2.35(ii) we have

$$\langle \nabla f(x_0) - \nabla f(x_n), y - x_n \rangle \leqslant 0$$

for any $y \in H_{n-1} \bigcap Q_{n-1}$. Hence we obtain that $\Omega \subset Q_n$. Therefore $\Omega \subset H_n \bigcap Q_n$ and so $H_n \bigcap Q_n$ is nonempty. Hence $x_{n+1} = \mathrm{proj}^f_{H_n \cap Q_n}(x_0)$ is well defined. Consequently, we see that $\Omega \subset H_n \bigcap Q_n$ for any $n \in \mathbb{N}$. Thus the sequence we constructed is indeed well defined and satisfies Algorithm (3.3.6), as claimed. □

Lemma 3.3.4. *Algorithm* (3.3.7) *is well defined.*

Proof. The point y_n^i is well defined for each $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$. Hence we only have to show that $\{x_n\}_{n \in \mathbb{N}}$ is well defined. To this end, we will prove that the Bregman projection onto $C_n \bigcap Q_n$ is well defined (see (1.2.14)), that is, we need to show that $C_n \bigcap Q_n$ is a nonempty, closed and convex subset of X for each $n \in \mathbb{N}$ (see Proposition 1.2.34). Let $n \in \mathbb{N}$. It follows from Proposition 1.2.54 that C_n^i are closed half-spaces for any $i = 1, 2, \ldots, N$. Hence their intersection C_n is a closed polyhedral set. It is also obvious that Q_n is a closed half-space. Let $u \in \Omega$. For any $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$, we obtain from the definition of QBNE operator (see (1.3.8)) that

$$D_f\left(u, y_n^i\right) = D_f\left(u, S_n^i\left(x_n + e_n^i\right)\right) \leqslant D_f\left(u, x_n + e_n^i\right),$$

which implies that $u \in C_n^i$. Since this holds for any $i = 1, 2, \ldots, N$, it follows that $u \in C_n$. Thus $\Omega \subset C_n$ for any $n \in \mathbb{N}$. The rest of the proof is identical to the proof of Lemma 3.3.3 by replacing H_n with C_n. □

From now on we fix an arbitrary sequence $\{x_n\}_{n \in \mathbb{N}}$ which is generated by Algorithm (3.3.6) or by Algorithm (3.3.7).

Lemma 3.3.5. *The sequences* $\{D_f(x_n, x_0)\}_{n \in \mathbb{N}}$, $\{x_n\}_{n \in \mathbb{N}}$ *and* $\{y_n^i\}_{n \in \mathbb{N}}$, $i = 1, 2, \ldots, N$, *are bounded.*

Proof. Denote by Δ_n the intersection $H_n \cap Q_n$ in the case of Algorithm (3.3.6) and the intersection $C_n \cap Q_n$ in the case of Algorithm (3.3.7). It follows from the definition of Q_n and from Proposition 1.2.35(ii) that $\mathrm{proj}^f_{Q_n}(x_0) = x_n$. Furthermore, from Proposition 1.2.35(iii), for each $u \in \Omega$, we have

$$D_f(x_n, x_0) = D_f\left(\mathrm{proj}^f_{Q_n}(x_0), x_0\right) \leqslant D_f(u, x_0) - D_f\left(u, \mathrm{proj}^f_{Q_n}(x_0)\right) \leqslant D_f(u, x_0).$$

Hence the sequence $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is bounded by $D_f(u, x_0)$ for any $u \in \Omega$. Therefore by Proposition 1.2.47 the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded too, as claimed.

Now we will prove that each sequence $\{y^i_n\}_{n\in\mathbb{N}}$, $i = 1, 2, \ldots, N$, is bounded. Let $u \in \Omega$. From the three point identity (see (1.2.2)) we get

$$\begin{aligned} D_f(u, x_n + e_n) &= D_f(u, x_n) - D_f(x_n + e_n, x_n) + \langle \nabla f(x_n + e_n) - \nabla f(x_n), u - (x_n + e_n)\rangle \\ &\leqslant D_f(u, x_n) + \langle \nabla f(x_n + e_n) - \nabla f(x_n), u - (x_n + e_n)\rangle. \end{aligned} \tag{3.3.8}$$

We also have

$$D_f(u, x_n) = D_f\left(u, \mathrm{proj}^f_{\Delta_{n-1}}(x_0)\right) \leqslant D_f(u, x_0)$$

because of Proposition 1.2.35(iii) and since $\Omega \subset \Delta_{n-1}$. On the other hand, since f is uniformly Fréchet differentiable and bounded on bounded subsets of X, we obtain from Proposition 1.1.22(ii) that

$$\lim_{n\to\infty} \|\nabla f(x_n + e_n) - \nabla f(x_n)\|_* = 0$$

because $\lim_{n\to\infty} \|e_n\| = 0$. This means that if we take into account that $\{x_n\}_{n\in\mathbb{N}}$ is bounded, then we get

$$\lim_{n\to\infty} \langle \nabla f(x_n) - \nabla f(x_n + e_n), u - (x_n + e_n)\rangle = 0. \tag{3.3.9}$$

Combining these facts, we obtain from (3.3.8) that $\{D_f(u, x_n + e_n)\}_{n\in\mathbb{N}}$ is bounded. Using the inequality

$$D_f\left(u, y^i_n\right) \leqslant D_f(u, x_n + e_n),$$

we see that $\{D_f(u, y^i_n)\}_{n\in\mathbb{N}}$ is bounded too. The boundedness of the sequence $\{y^i_n\}_{n\in\mathbb{N}}$ now follows from Proposition 1.2.48. □

Lemma 3.3.6. *For any $i = 1, 2, \ldots, N$, we have the following facts.*

(i)

$$\lim_{n\to\infty} \left\|y^i_n - \left(x_n + e^i_n\right)\right\| = 0. \tag{3.3.10}$$

(ii)

$$\lim_{n\to\infty} \left\|\nabla f\left(y_n^i\right) - \nabla f\left(x_n + e_n^i\right)\right\|_* = 0. \tag{3.3.11}$$

(iii)

$$\lim_{n\to\infty} \left(f\left(y_n^i\right) - f\left(x_n + e_n^i\right)\right) = 0. \tag{3.3.12}$$

Proof. Since $x_{n+1} \in Q_n$ and $\mathrm{proj}_{Q_n}^f(x_0) = x_n$, it follows from Proposition 1.2.35(iii) that

$$D_f\left(x_{n+1}, \mathrm{proj}_{Q_n}^f(x_0)\right) + D_f\left(\mathrm{proj}_{Q_n}^f(x_0), x_0\right) \leqslant D_f(x_{n+1}, x_0)$$

and hence

$$D_f(x_{n+1}, x_n) + D_f(x_n, x_0) \leqslant D_f(x_{n+1}, x_0). \tag{3.3.13}$$

Therefore the sequence $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is increasing and since it is also bounded (see Lemma 3.3.5), $\lim_{n\to\infty} D_f(x_n, x_0)$ exists. Thus from (3.3.13) it follows that

$$\lim_{n\to\infty} D_f(x_{n+1}, x_n) = 0. \tag{3.3.14}$$

Proposition 1.2.50 now implies that

$$\lim_{n\to\infty} D_f\left(x_{n+1}, x_n + e_n^i\right) = 0. \tag{3.3.15}$$

Now we split our proof into two parts according to the differences between Algorithms (3.3.6) and (3.3.7). In both cases we will prove that $\lim_{n\to\infty} D_f\left(x_{n+1}, y_n^i\right) = 0$.

(i) For any $i = 1, 2, \ldots, N$, it follows from the inclusion $x_{n+1} \in H_n^i$ that

$$\left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), x_{n+1} - y_n^i \right\rangle \leqslant 0. \tag{3.3.16}$$

The three point identity (see (1.2.2)) now implies that

$$\begin{aligned} D_f\left(x_{n+1}, y_n^i\right) &= D_f\left(x_{n+1}, x_n + e_n^i\right) - D_f\left(y_n^i, x_n + e_n^i\right) \\ &\quad + \left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), x_{n+1} - y_n^i \right\rangle \\ &\leqslant D_f\left(x_{n+1}, x_n + e_n^i\right) + \left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), x_{n+1} - y_n^i \right\rangle. \end{aligned}$$

From (3.3.16) we get that

$$D_f\left(x_{n+1}, y_n^i\right) \leqslant D_f\left(x_{n+1}, x_n + e_n^i\right),$$

hence (3.3.15) leads to $\lim_{n\to\infty} D_f\left(x_{n+1}, y_n^i\right) = 0$.

(ii) For any $i = 1, 2, \ldots, N$, it follows from the inclusion $x_{n+1} \in C_n^i$ that

$$D_f\left(x_{n+1}, y_n^i\right) \leqslant D_f\left(x_{n+1}, x_n + e_n^i\right).$$

Hence from (3.3.15) it follows that $\lim_{n\to\infty} D_f\left(x_{n+1}, y_n^i\right) = 0$.

Since $\{y_n^i\}_{n\in\mathbb{N}}$ is bounded (see Lemma 3.3.5), Proposition 1.2.46 now implies that

$$\lim_{n\to\infty} \left\|y_n^i - x_{n+1}\right\| = 0.$$

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded (see Lemma 3.3.5), it follows from Proposition 1.2.46 and (3.3.14) that

$$\lim_{n\to\infty} \left\|y_n^i - x_{n+1}\right\| = 0.$$

Therefore, for any $i = 1, 2, \ldots, N$, we have

$$\lim_{n\to\infty} \left\|y_n^i - x_n\right\| \leqslant \lim_{n\to\infty} \left(\left\|y_n^i - x_{n+1}\right\| + \left\|x_{n+1} - x_n\right\|\right) = 0.$$

Since $\lim_{n\to\infty} \left\|e_n^i\right\| = 0$, it also follows that

$$\lim_{n\to\infty} \left\|y_n^i - \left(x_n + e_n^i\right)\right\| = 0.$$

The function f is uniformly Fréchet differentiable and bounded on bounded subsets of X. Hence from Proposition 1.1.22(ii) we get

$$\lim_{n\to\infty} \left\|\nabla f\left(y_n^i\right) - \nabla f\left(x_n + e_n^i\right)\right\|_* = 0$$

for any $i = 1, 2, \ldots, N$. Finally, since f is uniformly Fréchet differentiable on bounded subsets of X, it is also uniformly continuous on bounded subsets of X (see Proposition 1.1.22(i)) and therefore

$$\lim_{n\to\infty} \left(f\left(y_n^i\right) - f\left(x_n + e_n^i\right)\right) = 0$$

for any $i = 1, 2, \ldots, N$. □

Lemma 3.3.7. *If any weak subsequential limit of* $\{x_n\}_{n\in\mathbb{N}}$ *belongs to* Ω*, then the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_\Omega^f(x_0)$.

Proof. From Proposition 2.1.1 it follows that $\operatorname{Fix}\left(S_n^i\right)$ is closed and convex for each $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$. Therefore Ω is nonempty, closed and convex, and the Bregman projection $\operatorname{proj}_\Omega^f$ is well defined. Since $x_{n+1} = \operatorname{proj}_{C_n\cap Q_n}^f(x_0)$ and Ω is contained in K in both cases (see Lemmas 3.3.3 and 3.3.4), we have $D_f(x_{n+1}, x_0) \leqslant D_f(\tilde{u}, x_0)$. Therefore

Proposition 1.2.51 implies that $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\tilde{u} = \mathrm{proj}_{\Omega}^{f}(x_0)$, as claimed. □

Now we prove the convergence of the Minimal Norm-like Picard Iterative Method (see Algorithm (3.3.2)).

Theorem 3.3.8 (Convergence of Algorithm (3.3.2))**.** *Let $T_i : X \to X$, $i = 1, 2, \ldots, N$, be N QBFNE operators such that $F := \bigcap_{i=1}^{N} \mathrm{Fix}(T_i) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Then, for each $x_0 \in X$ there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (3.3.2)*. If, for each $i = 1, 2, \ldots, N$, the sequence of errors $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to\infty} \|e_n^i\| = 0$, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\mathrm{proj}_F^f(x_0)$ *as $n \to \infty$.*

Proof. We denote $S_n^i := T_i$ for any $i = 1, 2, \ldots, N$ and all $n \in \mathbb{N}$. Therefore $\Omega = F$. We see that Condition 1 holds and therefore we can apply our lemmata.

From Lemmata 3.3.3 and 3.3.5, any sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (3.3.2) is well defined and bounded. From now on we let $\{x_n\}_{n\in\mathbb{N}}$ be an arbitrary sequence which is generated by Algorithm (3.3.2).

We claim that every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to F. From Lemma 3.3.6 we have

$$\begin{aligned}\lim_{n\to\infty} \left\|y_n^i - \left(x_n + e_n^i\right)\right\| &= \lim_{n\to\infty} \left\|S_n^i\left(x_n + e_n^i\right) - \left(x_n + e_n^i\right)\right\| \\ &= \lim_{n\to\infty} \left\|T_i\left(x_n + e_n^i\right) - \left(x_n + e_n^i\right)\right\| = 0 \end{aligned} \tag{3.3.17}$$

for any $i = 1, 2, \ldots, N$. Now let $\{x_{n_k}\}_{k\in\mathbb{N}}$ be a weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and denote its weak limit by v. Let $z_n^i = x_n + e_n^i$. Since $x_{n_k} \rightharpoonup v$ and $e_{n_k}^i \to 0$ as $k \to \infty$, it is obvious that for any $i = 1, \ldots, N$, the sequence $\left\{z_{n_k}^i\right\}_{k\in\mathbb{N}}$ also converges weakly to v. We also have $\lim_{k\to\infty} \left\|T_i z_{n_k}^i - z_{n_k}^i\right\| = 0$ by (3.3.17). This means that $v \in \widehat{\mathrm{Fix}}(T_i)$ for any $i = 1, 2, \ldots, N$. Since each T_i is a QBFNE operator, it follows that $v \in \mathrm{Fix}(T_i)$ for any $i = 1, 2, \ldots, N$. Therefore $v \in F$, as claimed.

Now Theorem 3.3.8 is seen to follow from Lemma 3.3.7. □

Now we prove the convergence of the Minimal Norm-Like Bauschke-Combettes iterative method (see Algorithm (3.3.5)). The analysis of this algorithm was first done in [92, Theorem 1, page 126].

Theorem 3.3.9 (Convergence of Algorithm (3.3.5))**.** *Let $T_i : X \to X$, $i = 1, 2, \ldots, N$, be N QBNE operators such that $F := \bigcap_{i=1}^{N} \mathrm{Fix}(T_i) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded* *subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Then, for each $x_0 \in X$ there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (3.3.5)*. If, for each $i = 1, 2, \ldots, N$, the sequence of errors $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ satisfies* $\lim_{n\to\infty} \|e_n^i\| = 0$*, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\operatorname{proj}_F^f(x_0)$ *as $n \to \infty$.*

Proof. We denote $S_n^i := T_i$ for any $i = 1, 2, \ldots, N$ and all $n \in \mathbb{N}$. Therefore $\Omega = F$. We see that Condition 1 holds and therefore we can apply our lemmata.

By Lemmata 3.3.4 and 3.3.5, any sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (3.3.5) is well defined and bounded. From now on we let $\{x_n\}_{n\in\mathbb{N}}$ be an arbitrary sequence which is generated by Algorithm (3.3.5).

The rest of the proof is identical to the proof of 3.3.8. □

3.4 An Implicit Method for Approximating Fixed Points

In this section we prove a strong convergence theorem of Browder's type for BFNE operators (see Definition 1.3.5) with respect to a well chosen function (*cf.* [91, Theorem 15.13, page 310]).

Theorem 3.4.1 (Implicit method for approximating fixed points)**.** *Let $f : X \to \mathbb{R}$ be a Legendre, totally convex function which is positively homogeneous of degree $\alpha > 1$, uniformly Fréchet differentiable and bounded on bounded subsets of X. Let K be a nonempty, bounded, closed and convex subset of X with $0 \in K$, and let T be a BFNE self-operator. Then the following two assertions hold.*

(i) *For each $t \in (0, 1)$, there exists a unique $u_t \in K$ satisfying $u_t = tTu_t$.*

(ii) *The net $\{u_t\}_{t\in(0,1)}$ converges strongly to* $\operatorname{proj}_{\operatorname{Fix}(T)}^f\left(\nabla f^*(0)\right)$ *as $t \to 1^-$.*

Proof. (i) Fix $t \in (0, 1)$ and let S_t be the operator defined by $S_t = tT$. Since $0 \in K$ and K is convex, S_t is an operator from K into itself. We next show that S_t is a BFNE operator (see (1.3.4)). Indeed, if $x, y \in K$, then, since T is BFNE (see (1.3.4)), it follows from Proposition 1.1.24 that

$$\begin{aligned}\langle \nabla f(S_t x) - \nabla f(S_t y), S_t x - S_t y\rangle &= t^\alpha \langle \nabla f(Tx) - \nabla f(Ty), Tx - Ty\rangle \\ &\leqslant t^\alpha \langle \nabla f(x) - \nabla f(y), Tx - Ty\rangle \\ &= t^{\alpha-1} \langle \nabla f(x) - \nabla f(y), S_t x - S_t y\rangle \\ &\leqslant \langle \nabla f(x) - \nabla f(y), S_t x - S_t y\rangle .\end{aligned} \tag{3.4.1}$$

Thus S_t is also BFNE (see (1.3.4)). Since K is bounded, it follows from Corollary 2.1.8 that S_t has a fixed point. We next show that $\operatorname{Fix}(S_t)$ consists of exactly one point. If $u, u' \in \operatorname{Fix}(S_t)$, then it follows from (3.4.1) that

$$\begin{aligned}\langle \nabla f(u) - \nabla f(u'), u - u' \rangle &= \langle \nabla f(S_t u) - \nabla f(S_t u'), S_t u - S_t u' \rangle \\ &\leqslant t^{\alpha-1} \langle \nabla f(u) - \nabla f(u'), S_t u - S_t u' \rangle \\ &= t^{\alpha-1} \langle \nabla f(u) - \nabla f(u'), u - u' \rangle .\end{aligned} \tag{3.4.2}$$

From (3.4.2) and the monotonicity of ∇f (see Example 1.4.3), we have

$$\langle \nabla f(u) - \nabla f(u'), u - u' \rangle = 0.$$

Since f is Legendre (see Definition 1.2.7), then f is strictly convex and hence ∇f is strictly monotone (see again Example 1.4.3) and therefore $u = u'$. Thus there exists a unique $u_t \in K$ such that $u_t = S_t u_t$.

(ii) Let $\{t_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,1)$ such that $t_n \to 1^-$ as $n \to \infty$. Put $x_n = u_{t_n}$ for all $n \in \mathbb{N}$. From Propositions 2.1.1 and 2.1.6, $\operatorname{Fix}(T)$ is nonempty, closed and convex. Thus the Bregman projection $\operatorname{proj}^f_{\operatorname{Fix}(T)}$ is well defined. In order to show that $u_t \to \operatorname{proj}^f_{\operatorname{Fix}(T)}\left(\nabla f^*(0)\right)$ as $t \to 1^-$, it is sufficient to show that $x_n \to \operatorname{proj}^f_{\operatorname{Fix}(T)}\left(\nabla f^*(0)\right)$ as $n \to \infty$. Since K is bounded, there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup v$ as $k \to \infty$. By the definition of x_n, we have $\|x_n - Tx_n\| = (1 - t_n)\|Tx_n\|$ for all $n \in \mathbb{N}$. So, we have that $\|x_n - Tx_n\| \to 0$ as $n \to \infty$ and hence $v \in \widehat{\operatorname{Fix}}(T)$. Proposition 2.1.2 now implies that $v \in \operatorname{Fix}(T)$. We next show that $x_{n_k} \to v$ as $k \to \infty$. Let $y \in \operatorname{Fix}(T)$ be given and fix $n \in \mathbb{N}$. Then, since T is BFNE, we have from (1.3.4) that

$$\langle \nabla f(Tx_n) - \nabla f(Ty), Tx_n - Ty \rangle \leqslant \langle \nabla f(x_n) - \nabla f(y), Tx_n - Ty \rangle .$$

That is

$$0 \leqslant \langle \nabla f(x_n) - \nabla f(Tx_n), Tx_n - y \rangle .$$

Since

$$\begin{aligned}\nabla f(x_n) - \nabla f(Tx_n) &= \nabla f(t_n Tx_n) - \nabla f(Tx_n) \\ &= t_n^{\alpha-1} \nabla f(Tx_n) - \nabla f(Tx_n) = \left(t_n^{\alpha-1} - 1\right) \nabla f(Tx_n),\end{aligned}$$

we have

$$0 \leqslant \left\langle \left(t_n^{\alpha-1} - 1\right) \nabla f(Tx_n), Tx_n - y \right\rangle .$$

This yields

$$0 \leqslant \langle -\nabla f(Tx_n), Tx_n - y \rangle \tag{3.4.3}$$

and

$$\langle \nabla f(y) - \nabla f(Tx_n), y - Tx_n \rangle \leqslant \langle \nabla f(y), y - Tx_n \rangle . \tag{3.4.4}$$

Since $x_{n_k} \rightharpoonup v$ and $x_{n_k} - Tx_{n_k} \to 0$ as $k \to \infty$, it follows that $Tx_{n_k} \rightharpoonup v$ as $k \to \infty$. Hence from (3.4.4) we obtain that

$$\begin{aligned} \limsup_{k\to\infty} \langle \nabla f(y) - \nabla f(Tx_{n_k}), y - Tx_{n_k} \rangle &\leqslant \limsup_{k\to\infty} \langle \nabla f(y), y - Tx_{n_k} \rangle \\ &= \langle \nabla f(y), y - v \rangle . \end{aligned} \tag{3.4.5}$$

Substituting $y = v$ in (3.4.5), we get

$$0 \leqslant \limsup_{k\to\infty} \langle \nabla f(v) - \nabla f(Tx_{n_k}), v - Tx_{n_k} \rangle \leqslant 0.$$

Thus

$$\lim_{k\to\infty} \langle \nabla f(v) - \nabla f(Tx_{n_k}), v - Tx_{n_k} \rangle = 0.$$

Since

$$D_f(v, Tx_{n_k}) + D_f(Tx_{n_k}, v) = \langle \nabla f(v) - \nabla f(Tx_{n_k}), v - Tx_{n_k} \rangle ,$$

it follows that

$$\lim_{k\to\infty} D_f(v, Tx_{n_k}) = \lim_{k\to\infty} D_f(Tx_{n_k}, v) = 0.$$

Since f is totally convex (see Definition 1.2.8), Proposition 1.2.45 now implies that $Tx_{n_k} \to v$ as $k \to \infty$. Finally, we claim that $v = \mathrm{proj}^f_{\mathrm{Fix}(T)}\left(\nabla f^*(0)\right)$. Since ∇f is norm-to-weak* continuous on bounded subsets (see Proposition 1.1.21), it follows that $\nabla f(Tx_{n_k}) \rightharpoonup \nabla f(v)$ as $k \to \infty$. Letting $k \to \infty$ in (3.4.3), we obtain

$$0 \leqslant \langle -\nabla f(v), v - y \rangle$$

for any $y \in \mathrm{Fix}(T)$. Hence

$$0 \leqslant \left\langle \nabla f\left(\nabla f^*(0)\right) - \nabla f(v), v - y \right\rangle$$

for any $y \in \mathrm{Fix}(T)$. Thus Proposition 1.2.35(ii) implies that $v = \mathrm{proj}^f_{\mathrm{Fix}(T)}\left(\nabla f^*(0)\right)$. Consequently, the whole net $\{u_t\}_{t\in(0,1)}$ converges strongly to $\mathrm{proj}^f_{\mathrm{Fix}(T)}\left(\nabla f^*(0)\right)$ as $t \to 1^-$. This completes the proof. □

Remark 3.4.2 (Browder's type result for nonexpansive operators)**.** *Early analogs of Theorem 3.4.1 for nonexpansive mappings in Hilbert and Banach spaces may be found in [30, 58, 85].*

We specialize Theorem 3.4.1 to the case where $f = \|\cdot\|^2$ and X is a uniformly smooth and uniformly convex Banach space (see Definition 1.1.33). In this case the function $f = \|\cdot\|^2$ is Legendre (*cf.* [7, Lemma 6.2, page 639]) is bounded and uniformly Fréchet differentiable on bounded subsets of X. According to Proposition 1.2.18, since X is uniformly convex, f is totally convex. Thus we obtain the following corollary. As a matter of fact, this corollary is known to hold even when X is only a smooth and uniformly convex Banach space (see [69]).

Corollary 3.4.3 (Particular case)**.** *Let X be a uniformly smooth and uniformly convex Banach space. Let K be a nonempty, bounded, closed and convex subset of X with $0 \in K$, and let T be a BFNE self-operator with respect to $\|\cdot\|^2$. Then the following two assertions hold.*

(i) *For each $t \in (0, 1)$, there exists a unique $u_t \in K$ satisfying $u_t = tTu_t$.*

(ii) *The net $\{u_t\}_{t\in(0,1)}$ converges strongly to $\operatorname{proj}^{\|\cdot\|^2}_{\operatorname{Fix}(T)}(0)$ as $t \to 1^-$.*

Chapter 4

Iterative Methods for Approximating Zeroes

A problem of interest in Optimization Theory is that of finding zeroes of mappings $A : X \to 2^{X^*}$. Formally, the problem can be written as follows:

$$\text{Find } x \in X \text{ such that } 0^* \in Ax. \tag{4.0.1}$$

This problem occurs in practice in various forms. For instance, minimizing a convex and lower semicontinuous function $f : X \to (-\infty, +\infty]$, a basic problem of optimization, amounts to finding a zero of the mapping $A = \partial f$, where $\partial f(x)$ stands for the subdifferential (see Definition 1.1.12(iii)) of f at the point $x \in X$. Finding solutions of some classes of differential equations can also be reduced to finding zeroes of certain mappings $A : X \to 2^{X^*}$.

One of the most important methods for solving (4.0.1) consists of replacing (4.0.1) in the case of a Hilbert space, $\mathcal{H}$, with the fixed point problem for the operator $R_A : \mathcal{H} \to 2^{\mathcal{H}}$ defined by

$$R_A := (I + A)^{-1}.$$

When $\mathcal{H}$ is a Hilbert space, and provided that A satisfies some monotonicity conditions, the *classical resolvent* R_A of A is single-valued, nonexpansive and even firmly nonexpansive (see (1.3.1) and (1.3.12), respectively) which ensure that its iterates $x_{n+1} = R_A x_n$, based on Picard's method (see Algorithm (3.0.1) and Remark 3.1.4), converge weakly, and sometimes even strongly, to fixed points of the resolvent R_A which are necessarily zeroes of A (see [100]) as we will explain later on in this chapter. As in the case of fixed point problems, when X is not a Hilbert space, or if A fails to be monotone, the convergence of the iterates of R_A to a fixed point of R_A and, thus, to a solution of (4.0.1), is more difficult to ensure (see [41]).

One way to overcome this difficulty is to use, instead of the classical resolvent, a new type of resolvent which first introduced by Teboulle [108] in 1992 for the subdifferential mapping case and one year later by Eckstein [51] for a general monotone mapping (see also [46, 88, 8]). If $f : X \to (-\infty, +\infty]$ is an admissible function (see Definition 1.2.1), then the *f-resolvent* is the operator $\mathrm{Res}_A^f : X \to 2^X$ given by

$$\mathrm{Res}_A^f := (\nabla f + A)^{-1} \circ \nabla f. \tag{4.0.2}$$

It is well defined when A is monotone and $\operatorname{int}\operatorname{dom} f \bigcap \operatorname{dom} A \neq \varnothing$. Moreover, similarly to the classical resolvent, the fixed points of Res_A^f are solutions of (4.0.1). This leads to the question whether, and under which conditions concerning A and f, the iterates of Res_A^f approximate fixed points of Res_A^f. Some partial results in this direction are already known (see [41]).

In this section we present several methods for finding zeroes of maximal monotone mappings which improve and generalize previous results. The literature contains several other methods for finding zeroes of monotone mappings. See, for example, [4, 13, 14, 37, 41, 42, 38, 32, 39, 51, 81, 76, 78, 100, 101] and the references therein. Many of them are fixed point methods which calculate fixed points of the resolvent.

In the next sections we are motivated by the methods proposed in Chapter 3 for approximating fixed points. We describe various methods for finding zeroes of monotone mappings and prove convergence theorems for these methods. In the following result we see that any monotone mapping with bounded effective domain has zeroes (*cf.* [89, Lemma 4.1, page 480]).

Proposition 4.0.4 (Zeroes of mappings with bounded domains)**.** *If $A : X \to 2^{X^*}$ is a maximal monotone mapping with a bounded effective domain, then $A^{-1}\left(0^*\right) \neq \varnothing$.*

Proof. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers which converges to zero. The mapping $A + \varepsilon_n J_X$ is surjective for any $n \in \mathbb{N}$ because A is a maximal monotone operator (see Proposition 1.4.17). Therefore, for any $n \in \mathbb{N}$, there exists $x_n \in \operatorname{dom} A$ such that $0^* \in (A + \varepsilon_n J_X) x_n$. Consequently, for any $n \in \mathbb{N}$, there are $\xi_n \in A x_n$ and $\eta_n \in J_X(x_n)$ such that $\xi_n + \varepsilon_n \eta_n = 0^*$. Therefore from the definition of the normalized duality mapping (see (1.1.10)) we have

$$\lim_{n\to\infty} \|\xi_n\|_* = \lim_{n\to\infty} \varepsilon_n \|\eta_n\|_* = \lim_{n\to\infty} \varepsilon_n \|x_n\| \to 0$$

because $\{x_n\}_{n\in\mathbb{N}}$ is a bounded sequence. Hence there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ which converges weakly to some $x_0 \in X$. Since A is monotone (see (1.4.1)), we have

$$\langle \zeta - \xi_{n_k}, v - x_{n_k} \rangle \geqslant 0, \quad \forall k \in \mathbb{N} \tag{4.0.3}$$

for any $(v, \zeta) \in \operatorname{graph} A$. Letting $k \to \infty$ in (4.0.3), we obtain $\langle \zeta, v - x_0 \rangle \geqslant 0$ for all $(v, \zeta) \in \operatorname{graph} A$ and from the maximality of A it follows that $x_0 \in A^{-1}\left(0^*\right)$ (see Proposition 1.4.13). Hence $A^{-1}\left(0^*\right) \neq \varnothing$, as claimed □

4.1 Properties of f-Resolvents

Let $A : X \to 2^{X^*}$ be a mapping such that

$$\operatorname{int}\operatorname{dom} f \bigcap \operatorname{dom} A \neq \varnothing. \tag{4.1.1}$$

Remark 4.1.1 (Particular case of f-resolvents). *If K is a nonempty, closed and convex subset of X, then the indicator function ι_K of K, that is, the function*

$$\iota_K(x) := \begin{cases} 0 & \text{if } x \in K \\ +\infty & \text{if } x \notin K, \end{cases}$$

is proper, convex and lower semicontinuous, and therefore $\partial\iota_K$ exists and is a maximal monotone mapping with domain K (see Proposition 1.4.19). The operator $\operatorname{Res}^f_{\partial\iota_K}$ is exactly the Bregman projection onto K with respect to f which we already defined in (1.2.14). *As we already noted there, this operator is denoted by* proj^f_K.

Now, we present several properties of f-resolvents which will be used later (*cf.* [8, Proposition 3.8, page 604]).

Proposition 4.1.2 (Properties of f-resolvents). *Let $f : X \to (-\infty, +\infty]$ be an admissible function and let $A : X \to 2^{X^*}$ be a mapping. The following statements hold.*

(i) $\operatorname{dom}\operatorname{Res}^f_A \subset \operatorname{int}\operatorname{dom} f$.

(ii) $\operatorname{ran}\operatorname{Res}^f_A \subset \operatorname{int}\operatorname{dom} f$.

(iii) $\operatorname{Fix}\left(\operatorname{Res}^f_A\right) = \operatorname{int}\operatorname{dom} f \bigcap A^{-1}(0^*)$.

(iv) *Suppose, in addition, that A is a monotone mapping and $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex (and, in particular, if f is Legendre). Then the following hold.*

(a) *The operator* Res^f_A *is single-valued on its domain.*

(b) *The operator* Res^f_A *is BFNE.*

(c) *Suppose that*

$$\operatorname{ran}\nabla f \subseteq \operatorname{ran}(\nabla f + A). \tag{4.1.2}$$

Then $\operatorname{dom}\operatorname{Res}^f_A = \operatorname{int}\operatorname{dom} f$ *and* $\operatorname{Fix}\left(\operatorname{Res}^f_A\right)$ *is convex set.*

Proof. (i) It is clear from (1.2.5) that

$$\operatorname{dom}\operatorname{Res}^f_A = \operatorname{dom}(\nabla f + A)^{-1} \circ \nabla f \subseteq \operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f.$$

(ii) Again we have from (1.2.5) that

$$\begin{aligned}\operatorname{ran}\operatorname{Res}_A^f &\subset \operatorname{ran}(\nabla f + A)^{-1} = \operatorname{dom}(\nabla f + A)\\ &= \operatorname{dom}\nabla f \bigcap \operatorname{dom} A \subset \operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f.\end{aligned}$$

(iii) From assertion (i) we have $\operatorname{Fix}\left(\operatorname{Res}_A^f\right) \subset \operatorname{int}\operatorname{dom} f$ and we know that $0^* \in Ax$ if and only if $x \in \operatorname{Res}_A^f x$ for any $x \in \operatorname{int}\operatorname{dom} f$, indeed,

$$\begin{aligned}0^* \in Ax &\Leftrightarrow \nabla f(x) \in Ax + \nabla f(x) = (\nabla f + A)(x)\\ &\Leftrightarrow x \in (\nabla f + A)^{-1}(\nabla f(x)) = \operatorname{Res}_A^f x.\end{aligned}$$

Hence, we have

$$\operatorname{int}\operatorname{dom} f \bigcap A^{-1}\left(0^*\right) = \operatorname{int}\operatorname{dom} f \bigcap \operatorname{Fix}\left(\operatorname{Res}_A^f\right) = \operatorname{Fix}\left(\operatorname{Res}_A^f\right).$$

(iv) Suppose that A is a monotone mapping and $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex.

(a) Fix $x \in \operatorname{dom}\operatorname{Res}_A^f$ and $\{u, v\} \subset \operatorname{Res}_A^f x$. Then (1.3.4) implies that

$$\langle \nabla f(u) - \nabla f(v), u - v\rangle \leqslant 0.$$

But the converse inequality is also true since ∇f is monotone (see Example 1.4.3). The function f is strictly convex on $\operatorname{int}\operatorname{dom} f$. Thus ∇f is strictly monotone on $\operatorname{int}\operatorname{dom} f$ (see again Example 1.4.3). Since $\{u, v\} \in \operatorname{int}\operatorname{dom} f$, we obtain that $u = v$.

(b) In view of assertions (i) and (ii), we have to show that (1.3.4) is satisfied for any $x, y \in \operatorname{dom}\operatorname{Res}_A^f$. Then from the definition of the f-resolvent (see (4.0.2)) we have that $\nabla f(x) - \nabla f\left(\operatorname{Res}_A^f x\right) \in A\left(\operatorname{Res}_A^f x\right)$ and $\nabla f(y) - \nabla f\left(\operatorname{Res}_A^f y\right) \in A\left(\operatorname{Res}_A^f y\right)$. Indeed, if $x \in \operatorname{dom}\operatorname{Res}_A^f$ then we have that

$$\begin{aligned}\operatorname{Res}_A^f x = (\nabla f + A)^{-1} \circ \nabla f(x) &\Leftrightarrow (\nabla f + A)\left(\operatorname{Res}_A^f x\right) = \nabla f(x)\\ &\Leftrightarrow \nabla f(x) - \nabla f\left(\operatorname{Res}_A^f x\right) = A\left(\operatorname{Res}_A^f x\right),\end{aligned}$$

and the same for $y \in \operatorname{dom}\operatorname{Res}_A^f$. Consequently, since A is monotone (see (1.4.1)), we have $\langle \xi - \eta, x - y\rangle \geqslant 0$ for any $\xi \in Ax$ and for any $\eta \in Ay$. Therefore, we get that

$$\left\langle \nabla f(x) - \nabla f\left(\operatorname{Res}_A^f x\right) - \left(\nabla f(y) - \nabla f\left(\operatorname{Res}_A^f y\right)\right), \operatorname{Res}_A^f x - \operatorname{Res}_A^f y\right\rangle \geqslant 0;$$

thus,

$$\left\langle \nabla f\left(\operatorname{Res}_A^f x\right) - \nabla f\left(\operatorname{Res}_A^f y\right), \operatorname{Res}_A^f x - \operatorname{Res}_A^f y\right\rangle \leqslant \left\langle \nabla f(x) - \nabla f(y), \operatorname{Res}_A^f x - \operatorname{Res}_A^f y\right\rangle.$$

Hence the operator Res_A^f is BFNE.

(c) Suppose that (4.1.2) holds. Then we have that $\operatorname{ran} \nabla f \subset \operatorname{dom}(\nabla f + A)^{-1}$ and therefore from (1.2.5) we get

$$\operatorname{dom} \operatorname{Res}_A^f = \operatorname{dom} \nabla f = \operatorname{int} \operatorname{dom} f.$$

Indeed, from assertion (i) we have that $\operatorname{dom} \operatorname{Res}_A^f \subset \operatorname{int} \operatorname{dom} f$. It remains to show that $\operatorname{int} \operatorname{dom} f \subset \operatorname{dom} \operatorname{Res}_A^f$. For any $x \in \operatorname{int} \operatorname{dom} f$ we have from (4.1.2) that $\nabla f(x) \in \operatorname{dom}(\nabla f + A)^{-1}$ and, therefore, $(\nabla f + A)^{-1} \circ \nabla f(x) \neq \varnothing$ that is $\operatorname{Res}_A^f(x) \neq \varnothing$, it means that $\operatorname{dom} \operatorname{Res}_A^f \supset \operatorname{dom} \nabla f = \operatorname{int} \operatorname{dom} f$. In view of assertion (b) of (iv) above and Proposition 2.1.1 (see also Figure 1.3), $\operatorname{Fix}\left(\operatorname{Res}_A^f\right)$ is closed and convex. □

The following result gives two other conditions which guarantee that an f-resolvent of a maximal monotone mapping satisfies $\operatorname{dom} \operatorname{Res}_A^f = \operatorname{int} \operatorname{dom} f$ (*cf.* [8, Corollary 3.14, page 606]).

Proposition 4.1.3 (Sufficient conditions for $\operatorname{dom} \operatorname{Res}_A^f = \operatorname{int} \operatorname{dom} f$)**.** *Let $f : X \to (-\infty, +\infty]$ be an admissible function and let $\lambda > 0$. Suppose that $A : X \to 2^{X^*}$ is a maximal monotone mapping such that $A^{-1}\left(0^*\right) \neq \varnothing$. If one of the following conditions holds:*

(i) $\operatorname{ran} \nabla f$ *is open and* $\operatorname{dom} A \subset \operatorname{int} \operatorname{dom} f$*;*

(ii) f *is Legendre and* $\operatorname{dom} A \subset \operatorname{int} \operatorname{dom} f$*,*

then $\operatorname{dom} \operatorname{Res}_{\lambda A}^f = \operatorname{int} \operatorname{dom} f$*.*

Corollary 4.1.4 (f-resolvent with full domain - case I)**.** *Let $f : X \to \mathbb{R}$ be a Legendre function and let $A : X \to 2^{X^*}$ be a maximal monotone mapping such that $A^{-1}\left(0^*\right) \neq \varnothing$. Then* $\operatorname{dom} \operatorname{Res}_{\lambda A}^f = X$*. If f is also cofinite, then* $\operatorname{dom} \operatorname{Res}_{\lambda A}^f = X$ *implies that A is maximal monotone.*

The following result presents another property of maximal monotone mappings with possibly empty zeroes set (*cf.* [8, Theorem 3.13(iv), page 606]).

Proposition 4.1.5 (f-resolvent with full domain - case II)**.** *Let $f : X \to (-\infty, +\infty]$ be an admissible function and let $\lambda > 0$. Suppose that $A : X \to 2^{X^*}$ is a maximal monotone* *mapping such that* $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$ *and satisfies (4.1.1). If* f *is Legendre and cofinite then* $\operatorname{dom}\operatorname{Res}^f_{\lambda A} = X$.

The next result shows the strong connection between resolvents of monotone mappings and BFNE operators (*cf.* [15, Proposition 5.1, page 7]). In this connection see Section 2.1.1 for another characterization.

Proposition 4.1.6 (Characterization of BFNE operators)**.** *Let* $f : X \to \mathbb{R}$ *be a cofinite and Legendre function. Let* K *be a subset of* X, $T : K \to X$ *be an operator and set* $A_T := \nabla f \circ T^{-1} - \nabla f$. *Suppose that* $A : X \to 2^{X^*}$ *is a maximal monotone mapping. The following assertions hold.*

(i) *The* f*-resolvent* $\operatorname{Res}^f_{A_T}$ *is exactly* T *and* $A^{-1}\left(0^*\right) = A_T^{-1}\left(0^*\right)$.

(ii) *If* T *is BFNE, then* A_T *is monotone.*

(iii) *If* T *is BFNE, then* $K = X$ *if and only if* A_T *is maximal monotone.*

Proof. (i) From the definition of A_T we get that

$$\begin{aligned} A_T = \nabla f \circ T^{-1} - \nabla f &\Leftrightarrow A_T + \nabla f = \nabla f \circ T^{-1} \Leftrightarrow \left(A_T + \nabla f\right)^{-1} = \left(\nabla f \circ T^{-1}\right)^{-1} \\ &\Leftrightarrow \left(A_T + \nabla f\right)^{-1} = T \circ \nabla f^{-1} \Leftrightarrow \left(A_T + \nabla f\right)^{-1} \circ \nabla f = T \circ \nabla f^{-1} \circ \nabla f = T \\ &\Leftrightarrow \operatorname{Res}^f_{A_T} = T. \end{aligned}$$

It is easy to check that $A^{-1}\left(0^*\right) = A_T^{-1}\left(0^*\right)$.

(ii) Take (u, ξ) and (v, η) in $\operatorname{graph} A_T$. Then

$$\begin{aligned} \xi \in A_T u = \nabla f \circ T^{-1} u - \nabla f(u) &\Leftrightarrow \xi + \nabla f(u) \in \nabla f \circ T^{-1} u \\ \Leftrightarrow u \in \left(\nabla f \circ T^{-1}\right)^{-1}\left(\xi + \nabla f(u)\right) &\Leftrightarrow u = T \circ \nabla f^{-1}\left(\xi + \nabla f(u)\right), \end{aligned}$$

and analogously $v = T \circ \nabla f^{-1}\left(\eta + \nabla f(v)\right)$. Since T is BFNE, we know from (1.3.4) that

$$\begin{aligned} &\left\langle \nabla f(u) - \nabla f(v), u - v \right\rangle \\ &= \left\langle \nabla f\left(T \circ \nabla f^{-1}\left(\xi + \nabla f(u)\right)\right) - \nabla f\left(T \circ \nabla f^{-1}\left(\eta + \nabla f(v)\right)\right), u - v \right\rangle \\ &\leqslant \left\langle \nabla f\left(\nabla f^{-1}\left(\xi + \nabla f(u)\right)\right) - \nabla f\left(\nabla f^{-1}\left(\eta + \nabla f(v)\right)\right), u - v \right\rangle \\ &= \left\langle \left(\xi + \nabla f(u)\right) - \left(\eta + \nabla f(v)\right), u - v \right\rangle, \end{aligned}$$

that is, $\langle \xi - \eta, u - v \rangle \geqslant 0$ which means that T is monotone (see (1.4.1)).

(iii) Suppose that T is BFNE. By assertion (ii), A_T is monotone. Using assertion (i) and Corollary 4.1.4, we obtain that A_T is maximal monotone if and only if $\operatorname{dom} \operatorname{Res}_{A_T}^f = \operatorname{dom} T = K = X$. □

4.2 Examples of f-Resolvents

As we explained, any BFNE operator is an f-resolvent (see (4.0.2)) of a monotone mapping (see Proposition 4.1.6(ii)). Since f-resolvents play an important role in the analysis of optimization problems, in the following subsection we provide several examples of f-resolvents with respect to different choices of admissible functions f, for example, the Boltzmann-Shannon entropy (see (1.2.8)) and the Fermi-Dirac entropy (see (1.2.9)).

4.2.1 Examples of $\mathcal{BS}$-Resolvents

Let $A : (0, +\infty) \to \mathbb{R}$ be a monotone mapping. Then the $\mathcal{BS}$-resolvent of A is

$$\operatorname{Res}_A^{\mathcal{BS}} := (\log + A)^{-1} \circ \log .$$

Remark 4.2.1 (Another formulation of the $\mathcal{BS}$-resolvent)**.** *We can also write the $\mathcal{BS}$-resolvent as follows:*

$$\operatorname{Res}_A^{\mathcal{BS}} = \left(\left((\log + A)^{-1} \circ \log\right)^{-1}\right)^{-1} = \left((\log)^{-1} \circ (\log + A)\right)^{-1} = \left(e^{(\log + A)}\right)^{-1},$$

where $\left(e^{(\log + A)}\right)(x) = xe^{A(x)}$*. This naturally leads us to the Lambert W function.* ◇

Recall [24, 26] that the *Lambert* W *function*, W, is defined to be the inverse of $x \mapsto xe^x$ and is implemented in both Maple and Mathematica. Its principal branch on the real axis is shown in Figure 4.1. Like log, it is concave increasing, and its domain is $(-1/e, +\infty)$.

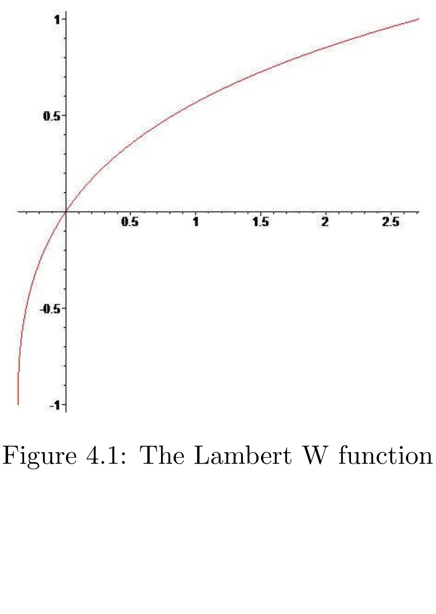

Figure 4.1: The Lambert W function

We now give several examples of $\mathcal{BS}$-resolvents.

Example 4.2.2 ($\mathcal{BS}$-resolvents in the real line)**.** (i) *If* $A(x) = \alpha$, $\alpha \in \mathbb{R}$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = e^{-\alpha}x$ *for all* $x \in \mathbb{R}_{++}$.

In particular, if $\alpha = 0$ *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = x$, $x \in \mathbb{R}_{++}$.

(ii) *If* $A(x) = \alpha x + \beta$, $\alpha, \beta \in \mathbb{R}$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = (1/\alpha)\,\mathrm{W}\left(\alpha e^{-\beta}x\right)$ *for all* $x \in \mathbb{R}_{++}$.

Hence, if $\alpha = 1$ *and* $\beta = 0$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = \mathrm{W}(x)$, $x \in \mathbb{R}_{++}$.

(iii) *If* $A(x) = \alpha \log(x)$, $\alpha \in \mathbb{R}$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = x^{1/(1+\alpha)}$ *for all* $x \in \mathbb{R}_{++}$.

Therefore, if $\alpha = 1$ *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = \sqrt{x}$, $x \in \mathbb{R}_{++}$.

(iv) *If* $A(x) = x^p/p$, $p > 1$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = \left(\mathrm{W}(x^p)\right)^{1/p}$ *for all* $x \in \mathbb{R}_{++}$.

Thus, if $p = 2$ *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = \sqrt{\mathrm{W}(x^2)}$, $x \in \mathbb{R}_{++}$.

(v) *If* $A(x) = \mathrm{W}(\alpha x^p)$, $\alpha \in \mathbb{R}$ *and* $p \geqslant 1$, *then*

$$\mathrm{Res}_A^{\mathcal{BS}}(x) = \left(\frac{x}{\alpha(p+1)}\right)^{\frac{1}{p+1}} \left(\mathrm{W}\left(\alpha(p+1)x^p\right)\right)^{\frac{1}{p+1}}$$

for all $x \in \mathbb{R}_{++}$.
Therefore, if $\alpha = 2$ *and* $p = 1$, *then* $\mathrm{Res}_A^{\mathcal{BS}}(x) = \sqrt{\frac{x}{4}}\sqrt{\mathrm{W}(4x)}$, $x \in \mathbb{R}_{++}$. ◇

We now present an example of a $\mathcal{BS}$-resolvent in $\mathbb{R}^2$.

Example 4.2.3 ($\mathcal{BS}$-resolvent in $\mathbb{R}^2$)**.** *Let* $\mathcal{BS}_2(x, y) := x\log(x) + y\log(y) - x - y$. *Thus* $\nabla\mathcal{BS}_2(x, y) = (\log(x), \log(y))$. *Let* $\theta \in [0, \pi/2]$ *and consider the rotation mapping* $A_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ *defined by*

$$A_\theta(x, y) := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

In particular, the $\mathcal{BS}$*-resolvent of the rotation mapping* $A_{\pi/2}$ *is the operator*

$$\mathrm{Res}_{A_{\pi/2}}^{\mathcal{BS}_2} := \left(\nabla\mathcal{BS}_2 + A_{\pi/2}\right)^{-1} \circ (\nabla\mathcal{BS}_2).$$

We claim that the inverse of $\nabla\mathcal{BS}_2 + A_{\pi/2}$ *uniquely exists. To see this, note that for any* $x, y \in (0, +\infty)$, *we have*

$$\left(\nabla\mathcal{BS}_2 + A_{\pi/2}\right)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \log(x) - y \\ \log(y) + x \end{pmatrix}.$$

Thus we have to show that for any $(z, w) \in \mathbb{R}^2$*, there exist unique* $x, y \in (0, +\infty)$ *such that* $z = \log(x) - y$ *and* $w = \log(y) + x$*. These two equations can be written as*

$$x = e^{y+z} \quad \textit{and} \quad y = e^{w-x}.$$

Therefore, $x = e^{e^{w-x}+z}$*. This equation has indeed a unique solution in* $(0, +\infty)$*. To check this, define a function* $f : [0, +\infty) \to \mathbb{R}$ *by* $f(x) = x - e^{e^{w-x}+z}$*. Then it is easy to see that* $f(0) = -e^{e^{w}+z} < 0$ *and* $\lim_{x\to+\infty} f(x) = +\infty$*. Since the function* f *is continuous, it has at least one root. On the other hand,*

$$f'(x) = 1 - e^{e^{w-x}+z}\left(-e^{w-x}\right) = 1 + e^{e^{w-x}+w-x+z} > 0.$$

This means that f *has exactly one root, which is the unique solution of the equation* $x = e^{e^{w-x}+z}$*. The general case is similar but less explicit.*

4.2.2 Examples of $\mathcal{FD}$-Resolvents

Let $A : (0, 1) \to \mathbb{R}$ be a monotone mapping. Then the $\mathcal{FD}$-resolvent of A is

$$\mathrm{Res}_A^{\mathcal{FD}} := (\mathcal{FD}' + A)^{-1} \circ \mathcal{FD}',$$

where in this case $\mathcal{FD}'(x) = \log(x/(1-x))$ and therefore $(\mathcal{FD}')^{-1}(x) = e^x/(1+e^x)$.

Remark 4.2.4 (Another formulation of the $\mathcal{FD}$-resolvent)**.** *We can also write the resolvent in the following way:*

$$\begin{aligned}\mathrm{Res}_A^{\mathcal{FD}} &= \left(\left((\mathcal{FD}' + A)^{-1} \circ \mathcal{FD}'\right)^{-1}\right)^{-1} = \left((\mathcal{FD}')^{-1} \circ (\mathcal{FD}' + A)\right)^{-1} \\ &= \left(\frac{e^{(\mathcal{FD}'+A)}}{1 + e^{(\mathcal{FD}'+A)}}\right)^{-1},\end{aligned}$$

where

$$\left(\frac{e^{(\mathcal{FD}'+A)}}{1 + e^{(\mathcal{FD}'+A)}}\right)(x) = \frac{xe^{A(x)}}{1 - x + xe^{A(x)}}.$$

◇

Several examples of $\mathcal{FD}$-resolvents follow.

Example 4.2.5 ($\mathcal{FD}$-resolvents in the real line)**.** (i) *If* $A(x) = \alpha$*,* $\alpha \in (0, +\infty)$*, then*

$$\mathrm{Res}_A^{\mathcal{FD}}(x) = \frac{x}{x + e^{\alpha}(1-x)}, \quad x \in (0, 1).$$

If $\alpha = 0$*, then* $\mathrm{Res}_A^{\mathcal{FD}}(x) = x$*,* $x \in (0, 1)$*.*

(ii) *If* $A(x) = \log(x)$, *then*

$$\mathrm{Res}_A^{\mathcal{FD}}(x) = \frac{x - \sqrt{4x - 3x^2}}{2(x-1)}$$

for all $x \in (0, 1)$.

(iii) *If* $A(x) = \log(1 - x)$, *then*

$$\mathrm{Res}_A^{\mathcal{FD}}(x) = \frac{x}{1 - x}$$

for all $x \in (0, 1)$.

(vi) If $A(x) = 2\log(1 - x)$, *then*

$$\mathrm{Res}_A^{\mathcal{FD}}(x) = \frac{1 - x - \sqrt{5x^2 - 6x + 1}}{2(x-1)}$$

for all $x \in (0, 1/5]$.

Finally, the next table lists f-resolvents with respect to various choices of functions f. Here, for simplicity, we denote $\mathrm{Res}_A^f = g^{-1}$.

$f(x)$	Domain	$g(x)$
$\mathcal{BS}(x)$	$(0, +\infty)$	$xe^{A(x)}$
$\mathcal{FD}(x)$	$(0, 1)$	$\frac{xe^{A(x)}}{1-x+xe^{A(x)}}$
$x^2/2$	$\mathbb{R}$	$x + A(x)$
$x^4/4$	$\mathbb{R}$	$(x^3 + A(x))^{1/3}$
e^x	$\mathbb{R}$	$\log(e^x + A(x))$
$-\log(x)$	$(0, +\infty)$	$\frac{x}{1-xA(x)}$

Table 4.1: Examples of f-Resolvents

4.2.3 Examples of f-Resolvents in Hilbert Spaces

Following [15, Example 9.6, page 71], we consider the function $f_p : \mathcal{H} \to \mathbb{R}$ defined by $f_p(x) = \frac{1}{p}\|x\|^p$, where $\mathcal{H}$ is a Hilbert space and $p \in (1, +\infty)$. So the conjugate of f_p is the function $f_p^*(y) = \frac{1}{q}\|y\|^q$, where q is the conjugate exponent of p, that is, $1/p + 1/q = 1$. Then, for any $y \neq 0$, we have that $\nabla f_p^*(y) = \|y\|^{q-2} y$. Consider $A = I$, the identity mapping, and denote the f-resolvent of A by

$$T_p = \mathrm{Res}_A^{f_p} := (\nabla f_p + I)^{-1} \circ \nabla f_p.$$

Then

$$T_p = \begin{cases} 0 & \text{if } x = 0 \\ k_p(x)\, x & \text{if } x \neq 0, \end{cases}$$

where $k_p(x) \in (0,1)$ is the unique solution to the equation

$$k^{p-1} + k \|x\|^{2-p} = 1.$$

4.3 Iterative Methods Based on Haugazeau's Algorithm

A well-known method for finding zeroes of monotone mappings in Hilbert space is the celebrated Proximal Point Algorithm.

Proximal Point Algorithm
Input: $\{\lambda_n\}_{n \in \mathbb{N}} \subset (0, +\infty)$.
Initialization: $x_0 \in \mathcal{H}$.
General Step ($n = 1, 2, \ldots$):

$$x_{n+1} = R_{\lambda_n A}(x_n) = (I + \lambda_n A)^{-1} x_n. \tag{4.3.1}$$

This algorithm was first introduced by Martinet [75] and further developed by Rockafellar [100], who proves that the sequence generated by Algorithm (4.3.1) converges weakly to an element of $A^{-1}(0)$ when $A^{-1}(0)$ is nonempty and $\liminf_{n\to\infty} \lambda_n > 0$. Furthermore, Rockafellar [100] asks if the sequence generated by Algorithm (4.3.1) converges strongly. For general monotone mappings a negative answer to this question follows from [55]; see also [14]. In the case of the subdifferential this question was answered in the negative by Güler [57], who presented an example of a subdifferential for which the sequence generated by Algorithm (4.3.1) converges weakly but not strongly; see [14] for a more recent and simpler example. There are several ways to generalize the classical proximal point algorithm (see Algorithm (4.3.1)) so that strong convergence is guaranteed.

In Chapter 3 we have studied several algorithms for approximating fixed points of Bregman nonexpansive operators. In the following sections we modify these methods in order to find zeroes of monotone mappings.

4.3.1 The Solodov-Svaiter Iterative Method

Solodov and Svaiter [105] modified the classical proximal point algorithm (see Algorithm (4.3.1)) in order to generate a strongly convergent sequence (in this sense see also Algorithm (3.0.4)). They introduced the following algorithm.

Solodov-Svaiter Proximal Point Algorithm
Input: $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$.
Initialization: $x_0 \in \mathcal{H}$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} 0 = v_n + \frac{1}{\lambda_n}(y_n - x_n), & v_n \in Ay_n, \\ H_n = \{z \in \mathcal{H} : \langle v_n, z - y_n\rangle \leqslant 0\}, \\ Q_n = \{z \in \mathcal{H} : \langle x_0 - x_n, z - x_n\rangle \leqslant 0\}, \\ x_{n+1} = P_{H_n \cap Q_n}(x_0). \end{cases} \tag{4.3.2}$$

They prove that if $A^{-1}(0)$ is nonempty and $\liminf_{n\to\infty} \lambda_n > 0$, then the sequence generated by Algorithm (4.3.2) converges strongly to $P_{A^{-1}(0)}$. Kamimura and Takahashi [64] generalized this result to those Banach space X which are both uniformly convex and uniformly smooth (see Definition 1.1.33(iii) and (iv)). They introduced the following algorithm.

Kamimura-Takahashi Proximal Point Algorithm
Input: $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} 0^* = \xi_n + \frac{1}{\lambda_n}(J_X(y_n) - J_X(x_n)), & \xi_n \in Ay_n, \\ H_n = \{z \in X : \langle v_n, z - y_n\rangle \leqslant 0\}, \\ Q_n = \{z \in X : \langle J_X(x_0) - J_X(x_n), z - x_n\rangle \leqslant 0\}, \\ x_{n+1} = \operatorname{proj}^{\|\cdot\|^2}_{H_n \cap Q_n}(x_0). \end{cases} \tag{4.3.3}$$

They prove that if $A^{-1}(0^*)$ is nonempty and $\liminf_{n\to\infty} \lambda_n > 0$, then the sequence generated by Algorithm (4.3.3) converges strongly to $\operatorname{proj}^{\|\cdot\|^2}_{A^{-1}(0^*)}(x_0)$.

We study an extension of Algorithms (4.3.2) and (4.3.3) in all reflexive Banach spaces using a well-chosen convex function f. More precisely, we consider the following algorithm introduced by Bauschke and Combettes [10] (see also Gárciga Otero and Svaiter [54]).

Bauschke-Combettes Proximal Point Algorithm I
Input: $f : X \to \mathbb{R}$ and $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$.
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} 0^* = \xi_n + \frac{1}{\lambda_n}(\nabla f(y_n) - \nabla f(x_n)), & \xi_n \in Ay_n, \\ H_n = \{z \in X : \langle \xi_n, z - y_n\rangle \leqslant 0\}, \\ Q_n = \{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leqslant 0\}, \\ x_{n+1} = \operatorname{proj}^f_{Q_n \cap W_n}(x_0). \end{cases} \tag{4.3.4}$$

Algorithm (4.3.4) is more flexible than Algorithm (4.3.3) because it leaves us the freedom of fitting the function f to the nature of the mapping A (especially when A is the subdifferential of some function) and of the space X in ways which make the application of Algorithm (4.3.4) simpler than that of Algorithm (4.3.3). It should be observed that if X is a Hilbert space $\mathcal{H}$, then using in Algorithm (4.3.4) the function $f = (1/2)\|\cdot\|^2$, one obtains exactly the classical proximal point algorithm (see Algorithm (4.3.2)). If X is not a Hilbert space, but still a uniformly convex and uniformly smooth Banach space X (see Definition 1.1.33(iii) and (iv)), then setting $f = (1/2)\|\cdot\|^2$ in Algorithm (4.3.4), one obtains exactly Algorithm (4.3.3). We also note that the choice $f = (1/2)\|\cdot\|^2$ in some Banach spaces may make the computations in Algorithm (4.3.3) quite difficult. These computations can be simplified by an appropriate choice of f. For instance, if $X = \ell^p$ or $X = L^p$ with $p \in (1, +\infty)$, and $f_p = (1/p)\|\cdot\|^p$ in Algorithm (4.3.4), then the computations become simpler than those required in Algorithm (4.3.3), which corresponds to $f = (1/2)\|\cdot\|^2$.

We study the following algorithm when $Z := \bigcap_{i=1}^{N} A_i^{-1}(0^*) \neq \varnothing$.

Minimal Norm-Like Proximal Point Algorithm I
Input: $f : X \to \mathbb{R}$, $\{\lambda_n^i\}_{n\in\mathbb{N}} \subset (0, +\infty)$ and $\{\eta_n^i\}_{n\in\mathbb{N}}$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$)**:**

$$\begin{cases} \eta_n^i = \xi_n^i + \frac{1}{\lambda_n^i}\left(\nabla f(y_n^i) - \nabla f(x_n)\right), & \xi_n^i \in A_i y_n^i, \\ H_n = \left\{z \in X : \langle \xi_n^i, z - y_n^i\rangle \leqslant 0\right\}, \\ H_n := \bigcap_{i=1}^{N} H_n^i, \\ Q_n = \left\{z \in X : \langle \nabla f(x_0) - \nabla f(x_n), z - x_n\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}_{H_n \cap Q_n}^f(x_0). \end{cases} \tag{4.3.5}$$

Note that if $\eta_n^i = 0^*$, then

$$y_n^i = \operatorname{Res}_{\lambda_n^i A_i}^f(x_n).$$

Now we will prove a convergence result for Algorithm (4.3.5) (*cf.* [89, Theorem 3.1, page 477]).

Theorem 4.3.1 (Convergence result for Algorithm (4.3.5))**.** *Let $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$, be N maximal monotone operators such that $Z := \bigcap_{i=1}^{N} A_i^{-1}(0^*) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Then, for each $x_0 \in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (4.3.5)*. If, for each $i = 1, 2, \ldots, N$, $\liminf_{n\to\infty} \lambda_n^i > 0$, and the sequence of errors $\{\eta_n^i\}_{n\in\mathbb{N}} \subset X^*$ satisfies $\lim_{n\to\infty} \lambda_n^i \|\eta_n^i\|_* = 0$ and $\limsup_{n\to\infty} \langle \eta_n^i, y_n^i\rangle \leqslant 0$, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z^f(x_0)$ as $n \to \infty$.*

Proof. Note that $\operatorname{dom}\nabla f = X$ because $\operatorname{dom} f = X$ and f is Legendre (see Definition 1.2.7). Hence it follows from Proposition 4.1.4 that $\operatorname{dom}\operatorname{Res}_{\lambda A}^f = X$. Denote $S_n^i := \operatorname{Res}_{\lambda_n^i A_i}^f$. Therefore from Proposition 4.1.2(iv)(b) and Figure 1.3 we have that each S_n^i is BFNE and thus QBFNE. We also have that $\Omega = Z$ and that Condition 1 holds.

We split our proof into three steps.

Claim 1: *There are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy Algorithm* (4.3.5).

Proof. As a matter of fact, we will prove that, for each $x_0 \in X$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (4.3.5) with $\eta_n^i = 0^*$ for all $i = 1, 2, \ldots, N$ and $n \in \mathbb{N}$. In this case $y_n^i = S_n^i(x_n)$. Therefore our claim follows directly from Lemma 3.3.3.

From now on we fix an arbitrary sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying Algorithm (4.3.5). It is clear from the proof of Claim 1 that $Z \subset H_n \bigcap Q_n$ for each $n \in \mathbb{N}$.

Claim 2: *The sequences* $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ *and* $\{x_n\}_{n\in\mathbb{N}}$ *are bounded.*

Proof. It is easy to check that the proof of the facts that $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ and $\{x_n\}_{n\in\mathbb{N}}$ are bounded proceeds exactly as in the proof of Lemma 3.3.5.

Claim 3: *Every weak subsequential limit of* $\{x_n\}_{n\in\mathbb{N}}$ *belongs to* Z.

Proof. It follows from the definition of Q_n and Proposition 1.2.35(ii) that $\mathrm{proj}_{Q_n}^f(x_0) = x_n$. Since $x_{n+1} \in Q_n$, it follows from Proposition 1.2.35(iii) that

$$D_f\left(x_{n+1}, \mathrm{proj}_{Q_n}^f(x_0)\right) + D_f\left(\mathrm{proj}_{Q_n}^f(x_0), x_0\right) \leqslant D_f(x_{n+1}, x_0)$$

and hence

$$D_f(x_{n+1}, x_n) + D_f(x_n, x_0) \leqslant D_f(x_{n+1}, x_0). \tag{4.3.6}$$

Therefore the sequence $\{D_f(x_n, x_0)\}_{n\in\mathbb{N}}$ is increasing and since it is also bounded (see Claim 2), $\lim_{n\to\infty} D_f(x_n, x_0)$ exists. Thus from (4.3.6) it follows that

$$\lim_{n\to\infty} D_f(x_{n+1}, x_n) = 0. \tag{4.3.7}$$

Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded (see Claim 2), Proposition 1.2.46 now implies that

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0.$$

For any $i = 1, 2, \ldots, N$, it follows from the three point identity (see (1.2.2)) that

$$\begin{aligned}
&D_f(x_{n+1}, x_n) - D_f\left(y_n^i, x_n\right) \\
&= D_f\left(x_{n+1}, y_n^i\right) + \left\langle \nabla f(x_n) - \nabla f\left(y_n^i\right), y_n^i - x_{n+1}\right\rangle \\
&\geqslant \left\langle \nabla f(x_n) - \nabla f\left(y_n^i\right), y_n^i - x_{n+1}\right\rangle = \left\langle \lambda_n^i\left(\xi_n^i - \eta_n^i\right), y_n^i - x_{n+1}\right\rangle \\
&= \lambda_n^i\left\langle \xi_n^i, y_n^i - x_{n+1}\right\rangle - \lambda_n^i\left\langle \eta_n^i, y_n^i - x_{n+1}\right\rangle \geqslant -\lambda_n^i\left\langle \eta_n^i, y_n^i - x_{n+1}\right\rangle
\end{aligned}$$

because $x_{n+1} \in H_n^i$. We now have

$$\begin{aligned} D_f\left(y_n^i, x_n\right) &\leqslant D_f\left(x_{n+1}, x_n\right) + \left\langle \lambda_n^i \eta_n^i, y_n^i - x_{n+1}\right\rangle \\ &= D_f\left(x_{n+1}, x_n\right) + \lambda_n^i \left\langle \eta_n^i, y_n^i\right\rangle - \left\langle \lambda_n^i \eta_n^i, x_{n+1}\right\rangle \\ &\leqslant D_f\left(x_{n+1}, x_n\right) + \lambda_n^i \left\langle \eta_n^i, y_n^i\right\rangle + \left\|\lambda_n^i \eta_n^i\right\|_* \left\|x_{n+1}\right\|. \end{aligned}$$

Hence

$$\begin{aligned} \limsup_{n\to\infty} D_f\left(y_n^i, x_n\right) &\leqslant \limsup_{n\to\infty} D_f\left(x_{n+1}, x_n\right) \\ &\quad + \limsup_{n\to\infty} \lambda_n^i \left\langle \eta_n^i, y_n^i\right\rangle + \limsup_{n\to\infty} \left\|\lambda_n^i \eta_n^i\right\|_* \left\|x_{n+1}\right\|. \end{aligned}$$

Since $\lim_{n\to\infty} \lambda_n^i \left\|\eta_n^i\right\|_* = 0$, $\limsup_{n\to\infty} \langle \eta_n^i, y_n^i\rangle \leqslant 0$ and (4.3.7), we see that

$$\limsup_{n\to\infty} D_f\left(y_n^i, x_n\right) \leqslant 0.$$

Hence $\lim_{n\to\infty} D_f\left(y_n^i, x_n\right) = 0$. Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded (see Claim 2), Proposition 1.2.46 again implies that $\lim_{n\to\infty} \left\|y_n^i - x_n\right\| = 0$. Since the function f is bounded and uniformly Fréchet differentiable on bounded subsets of X we get from Proposition 1.1.22(ii) that

$$\lim_{n\to\infty} \left\|\nabla f\left(x_n\right) - \nabla f\left(y_n^i\right)\right\|_* = 0.$$

Since $\liminf_{n\to\infty} \lambda_n^i > 0$ and $\lim_{n\to\infty} \left\|\eta_n^i\right\|_* = 0$, it follows that

$$\begin{aligned} \lim_{n\to\infty} \left\|\xi_n^i\right\|_* &= \lim_{n\to\infty} \frac{1}{\lambda_n^i} \left\|\nabla f\left(x_n\right) - \nabla f\left(y_n^i\right) + \eta_n^i\right\|_* \\ &\leqslant \lim_{n\to\infty} \frac{1}{\lambda_n^i} \left(\left\|\nabla f\left(x_n\right) - \nabla f\left(y_n^i\right)\right\|_* + \left\|\eta_n^i\right\|_*\right) = 0^*, \end{aligned} \tag{4.3.8}$$

for any $i = 1, 2, \ldots, N$.

Now let $\{x_{n_k}\}_{k\in\mathbb{N}}$ be a weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and denote its weak limit by v. Then $\left\{y_{n_k}^i\right\}_{k\in\mathbb{N}}$ also converges weakly to v for any $i = 1, 2, \ldots, N$. Since $\xi_n^i \in Ay_n^i$ and A_i is monotone (see (1.4.1)), it follows that

$$\left\langle \eta - \xi_n^i, z - y_n^i\right\rangle \geqslant 0$$

for all $(z, \eta) \in \operatorname{graph} A_i$. This, in turn, implies that $\langle \eta, z - v\rangle \geqslant 0$ for all $(z, \eta) \in \operatorname{graph} A_i$. Therefore, using the maximal monotonicity of A_i (see Proposition 1.4.13), we now obtain that $v \in A_i^{-1}\left(0^*\right)$ for each $i = 1, 2, \ldots, N$. Thus $v \in Z$ and this proves Claim 3.

Now Theorem 4.3.1 is seen to follow from Lemma 3.3.7. □

Suppose now that the mappings A_i, $i = 1, 2, \ldots, N$, have no common zero. If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence satisfying Algorithm (4.3.5), then $\lim_{n\to\infty} \|x_n\| = +\infty$. This is because if $\{x_n\}_{n\in\mathbb{N}}$ were to have a bounded subsequence, then it would follow from Claim 3 in the proof of Theorem 4.3.1 that the mappings A_i, $i = 1, 2, \ldots, N$, did share a common zero. In the case of a single zero free mapping A, we can prove that such a sequence always exists (*cf.* [89, Theorem 4.2, page 481]).

Theorem 4.3.2 (Algorithm (4.3.5) is well-defined - zero free case)**.** *Let $A : X \to 2^{X^*}$ be a maximal monotone mapping. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Then, for each $x_0 \in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (4.3.5) *with $N = 1$. If $\liminf_{n\to\infty} \lambda_n > 0$, and the sequence of errors $\{\eta_n\}_{n\in\mathbb{N}} \subset X^*$ satisfies $\lim_{n\to\infty} \lambda_n \|\eta_n\|_* = 0$ and $\limsup_{n\to\infty} \langle \eta_n, y_n\rangle \leqslant 0$, then either $A^{-1}\left(0^*\right) \neq \varnothing$ and each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}^f_{A^{-1}(0^*)}(x_0)$ as $n \to \infty$, or $A^{-1}\left(0^*\right) = \varnothing$ and each such sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfies $\lim_{n\to\infty} \|x_n\| = +\infty$.*

Proof. In view of Theorem 4.3.1, we only need to consider the case where $A^{-1}\left(0^*\right) = \varnothing$. First of all we prove that in this case, for each $x_0 \in X$, there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ which satisfies Algorithm (4.3.5) with $\eta_n = 0^*$ for all $n \in \mathbb{N}$.

We prove this by induction. We first check that the initial step ($n = 0$) is well defined. The proximal subproblem

$$0^* \in Ax + \frac{1}{\lambda_0}\left(\nabla f(x) - \nabla f(x_0)\right)$$

always has a solution (y_0, ξ_0) because it is equivalent to the problem $x = \operatorname{Res}^f_{\lambda_0 A}(x_0)$ and this problem does have a solution since $\operatorname{dom}\operatorname{Res}^f_{\lambda A} = X$ (see Propositions 1.2.13 and 4.1.5). Now note that $Q_0 = X$. Since H_0 cannot be empty, the next iterate x_1 can be generated; it is the Bregman projection of x_0 onto $H_0 = Q_0 \bigcap H_0$.

Note that whenever x_n is generated, y_n and ξ_n can further be obtained because the proximal subproblems always have solutions. Suppose now that x_n and (y_n, ξ_n) have already been defined for $n = 0, 1, \ldots, \hat{n}$. We have to prove that $x_{\hat{n}+1}$ is also well defined. To this end, take any $z_0 \in \operatorname{dom} A$ and define

$$\rho := \max\left\{\|y_n - z_0\| : n = 0, 1, \ldots, \hat{n}\right\}$$

and

$$h(x) := \begin{cases} 0, & \|x - z_0\| \leqslant \rho + 1 \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $h : X \to (-\infty, +\infty]$ is a proper, convex and lower semicontinuous function, its subdifferential ∂h is maximal monotone (see Proposition 1.4.19). Define $A' = A + \partial h$, which is also maximal monotone (see Proposition 1.4.22). Furthermore,

$$A'(z) = A(z) \quad \text{for all} \quad \|z - z_0\| < \rho + 1.$$

Therefore $\xi_n \in A'y_n$ for $n = 0, 1, \ldots, \hat{n}$. We conclude that x_n and (y_n, ξ_n) also satisfy the conditions of Theorem 4.3.1 applied to the problem $0^* \in A'(x)$. Since A' has a bounded effective domain, this problem has a solution by Proposition 4.0.4. Thus it follows from Claim 1 in the proof of Theorem 4.3.1 that $x_{\hat{n}+1}$ is well defined. Hence the whole sequence $\{x_n\}_{n\in\mathbb{N}}$ is well defined, as asserted.

If $\{x_n\}_{n\in\mathbb{N}}$ were to have a bounded subsequence, then it would follow from Claim 3 in the proof of Theorem 4.3.1 that A had a zero. Therefore if $A^{-1}\left(0^*\right) = \varnothing$, then $\lim_{n\to\infty} \|x_n\| = +\infty$, as asserted. □

Algorithm (4.3.4) is a special case of Algorithm (4.3.5) when $N = 1$ and $\eta_n = 0^*$ for all $n \in \mathbb{N}$. Hence as a direct consequence of Theorem 4.3.1 we obtain the following result (*cf.* [54] and [89, Theorem 5.1, page 482]).

Corollary 4.3.3 (Convergence result for Algorithm (4.3.4))**.** *Let $A : X \to 2^{X^*}$ be a maximal monotone mapping. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X, and suppose that* $\liminf_{n\to\infty} \lambda_n > 0$*. Then for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm* (4.3.4) *is well defined, and either $A^{-1}\left(0^*\right) \neq \varnothing$ and $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\operatorname{proj}^f_{A^{-1}(0^*)}(x_0)$ *as $n \to \infty$, or $A^{-1}\left(0^*\right) = \varnothing$ and* $\lim_{n\to\infty} \|x_n\| = +\infty$.

Notable corollaries of Theorems 4.3.1 and 4.3.2 occur when the space X is both uniformly smooth and uniformly convex (see Definition 1.1.33(ii) and (iv)). In this case the function $f = \|\cdot\|^2$ is Legendre (*cf.* [7, Lemma 6.2, page 24]) and both bounded and uniformly Fréchet differentiable on bounded subsets of X. According to Proposition 1.2.21, f is sequentially consistent since X is uniformly convex and hence f is totally convex on bounded subsets of X. Therefore Theorems 4.3.1 and 4.3.2 hold in this context and lead us to the following two results (*cf.* [89, Theorem 5.2, page 482] and [89, Theorem 5.3, page 483]) which, in some sense, complement in [64, Theorem 8] (see also [105, Theorem 1, page 199]).

Corollary 4.3.4 (Convergence result for Algorithm (4.3.3))**.** *Let X be a uniformly smooth and uniformly convex Banach space and let $A : X \to 2^{X^*}$ be a maximal monotone mapping. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm* (4.3.3) *is well defined. If* $\liminf_{n\to\infty} \lambda_n > 0$*, then either $A^{-1}\left(0^*\right) \neq \varnothing$ and $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\operatorname{proj}^{\|\cdot\|^2}_{A^{-1}(0^*)}(x_0)$ *as $n \to \infty$, or $A^{-1}\left(0^*\right) = \varnothing$ and* $\lim_{n\to\infty} \|x_n\| = +\infty$.

Corollary 4.3.5 (Convergence result for Algorithm (4.3.2))**.** *Let $\mathcal{H}$ be a Hilbert space and let $A : X \to 2^X$ be a maximal monotone mapping. Then, for each $x_0 \in \mathcal{H}$, the sequence*

$\{x_n\}_{n\in\mathbb{N}}$ *generated by Algorithm* (4.3.2) *is well defined. If* $\liminf_{n\to\infty}\lambda_n > 0$*, then either* $A^{-1}(0) \neq \varnothing$ *and* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $P_{A^{-1}(0)}(x_0)$ *as* $n\to\infty$*, or* $A^{-1}(0) = \varnothing$ *and* $\lim_{n\to\infty}\|x_n\| = +\infty$*.*

These corollaries also hold, of course, in the presence of computational errors as in Theorems 4.3.1 and 4.3.2.

Let $g : X \to (-\infty, +\infty]$ be a proper, convex and lower semicontinuous function. Using Theorems 4.3.1 and 4.3.2 for the subdifferential of g, we obtain an algorithm for finding a minimizer of g (*cf.* [89, Proposition 6.1, page 483]).

Corollary 4.3.6 (Application of Algorithm (4.3.5) - finding minimizers)**.** *Let* $g : X \to (-\infty, +\infty]$ *be a proper, convex and lower semicontinuous function which attains its minimum over* X*. If* $f : X \to \mathbb{R}$ *is a Legendre function which is bounded, uniformly Fréchet differentiable, and totally convex on bounded subsets of* X*, and* $\{\lambda_n\}_{n\in\mathbb{N}}$ *is a positive sequence with* $\liminf_{n\to\infty}\lambda_n > 0$*, then, for each* $x_0 \in X$*, the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *generated by Algorithm* (4.3.5) *with* $A = \partial g$ *converges strongly to a minimizer of* g *as* $n \to \infty$*.*

If g *does not attain its minimum over* X*, then* $\lim_{n\to\infty}\|x_n\| = +\infty$*.*

Proof. The subdifferential ∂g of g is a maximal monotone mapping because g is a proper, convex and lower semicontinuous function (see Proposition 1.4.19). Since the zero set of ∂g coincides with the set of minimizers of g, the result follows immediately from Theorems 4.3.1 and 4.3.2. □

Next we prove a result similar to Theorem 4.3.1, but with a different type of errors than those in Algorithm (4.3.5).

Minimal Norm-Like Proximal Point Algorithm II
Input: $f : X \to \mathbb{R}$ and $\{e_n^i\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$)**:**

$$\begin{cases} y_n^i = \operatorname{Res}^f_{\lambda_n^i A_i}\left(x_n + e_n^i\right), \\ H_n = \left\{z \in X : \left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), z - y_n^i \right\rangle \leqslant 0\right\}, \\ H_n := \bigcap_{i=1}^N H_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}^f_{H_n\cap Q_n}\left(x_0\right). \end{cases} \tag{4.3.9}$$

Theorem 4.3.7 (Convergence result for Algorithm (4.3.9))**.** *Let* $A_i : X \to 2^{X^*}$*,* $i = 1, 2, \ldots, N$*, be* N *maximal monotone mappings such that* $Z := \bigcap_{i=1}^N A_i^{-1}\left(0^*\right) \neq \varnothing$*. Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X*. Suppose that* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Then, for each* $x_0 \in X$*, there are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy*

Algorithm (4.3.9). *If, for each* $i = 1, 2, \ldots, N$, $\liminf_{n\to\infty} \lambda_n^i > 0$, *and the sequence of errors* $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ *satisfies* $\lim_{n\to\infty} \|e_n^i\| = 0$, *then each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\mathrm{proj}_Z^f(x_0)$ *as* $n \to \infty$.

Proof. Note that $\mathrm{dom}\,\nabla f = X$ because $\mathrm{dom}\, f = X$ and f is Legendre (see Definition 1.2.7). Hence it follows from Proposition 4.1.4 that $\mathrm{dom}\,\mathrm{Res}_{\lambda A}^f = X$. Denote $S_n^i := \mathrm{Res}_{\lambda_n^i A_i}^f$. Therefore from Proposition 4.1.2(iv)(b) and Figure 1.3 we have that each S_n^i is BFNE and hence QBFNE. We also have $\Omega = Z$ and we see that Condition 1 holds so that we can apply our lemmata.

From Lemmata 3.3.3 and 3.3.5, any sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (4.3.9) is well defined and bounded. From now on we let $\{x_n\}_{n\in\mathbb{N}}$ be an arbitrary sequence which is generated by Algorithm (4.3.9).

We claim that every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to Z. From Lemma 3.3.6 we have

$$\lim_{n\to\infty} \left\|\nabla f\left(y_n^i\right) - \nabla f\left(x_n + e_n^i\right)\right\|_* = 0,$$

for any $i = 1, 2, \ldots, N$. The rest of the proof follows as the proof of Theorem 4.3.7. $\square$

Following the same arguments as in the proof of Theorem 4.3.2, we can prove the following result.

Theorem 4.3.8 (Algorithm (4.3.9) is well-defined - zero free case)**.** *Let* $A : X \to 2^{X^*}$ *be a maximal monotone mapping. Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X. *Suppose that* ∇f^* *is bounded on bounded subsets of* $\mathrm{int\,dom}\, f^*$. *Then, for each* $x_0 \in X$, *there are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy Algorithm* (4.3.9) *with* $N = 1$. *If* $\liminf_{n\to\infty} \lambda_n > 0$, *and the sequence of errors* $\{e_n\}_{n\in\mathbb{N}} \subset X$ *satisfies* $\lim_{n\to\infty} e_n = 0$, *then either* $A^{-1}\left(0^*\right) \neq \varnothing$ *and each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\mathrm{proj}_{A^{-1}(0^*)}^f(x_0)$ *as* $n \to \infty$, *or* $A^{-1}\left(0^*\right) = \varnothing$ *and each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *satisfies* $\lim_{n\to\infty} \|x_n\| = +\infty$.

The following algorithm allows for computational errors of the kind of Algorithm (4.3.9) but in a different way and with a weaker condition. The following algorithm combines the proximal point algorithm and the Mann methods (see Algorithm (3.0.3)). More precisely, we study the following algorithm when $Z := \bigcap_{i=1}^N A_i^{-1}\left(0^*\right) \neq \varnothing$.

Minimal Norm-Like Proximal Point Algorithm III
Input: $f : X \to \mathbb{R}$, $\{\lambda_n^i\} n \in \mathbb{N} \subset (0, +\infty)$ and $\{\alpha_n^i\} n \in \mathbb{N} \subset [0,1]$, $i = 1, 2, \ldots, N$ and $\{e_n\}_{n\in\mathbb{N}} \subset X$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} z_n = \nabla f^* \left(\alpha_n \nabla f\left(x_n\right) + \left(1-\alpha_n\right) \nabla f\left(e_n\right)\right), \\ 0^* = \xi_n^i + \frac{1}{\lambda_n^i}\left(\nabla f\left(y_n^i\right) - \nabla f\left(x_n\right)\right), \qquad \xi_n^i \in A_i y_n^i, \\ H_n = \left\{z \in X : \left\langle \xi_n^i, z - y_n^i \right\rangle \leqslant 0\right\}, \\ H_n := \bigcap_{i=1}^N H_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}_{H_n \cap Q_n}^f \left(x_0\right). \end{cases} \tag{4.3.10}$$

Theorem 4.3.9 (Convergence result for Algorithm (4.3.10))**.** *Let $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$, be N maximal monotone mappings such that $Z := \bigcap_{i=1}^N A_i^{-1}\left(0^*\right) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Assume that $\{\alpha_n\}_{n\in\mathbb{N}} \subset [0,1]$ satisfies $\lim_{n\to\infty} \alpha_n = 1$ and $\{e_n\}_{n\in\mathbb{N}}$ is the sequence of errors which satisfies $\|e_n\| \leqslant M$ (M is a positive constant) and* $\liminf_{n\to\infty} \lambda_n^i > 0$*, $i = 1, 2, \ldots, N$, then, for each $x_0 \in X$, each such sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by Algorithm* (4.3.10) *converges strongly to* $\operatorname{proj}_Z^f(x_0)$ *as $n \to \infty$.*

Proof. Note that $\operatorname{dom} \nabla f = X$ because $\operatorname{dom} f = X$ and f is Legendre (see Definition 1.2.7). Hence it follows from Proposition 4.1.4 that $\operatorname{dom} \operatorname{Res}_{\lambda A}^f = X$. It is easy to check that $y_n^i = \operatorname{Res}_{\lambda_n^i A_i}^f(z_n)$. Following the arguments in the proof of Theorem 4.3.1 we get that there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm (4.3.10) and $\{x_n\}_{n\in\mathbb{N}}$ is bounded.

Now we will prove that every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to Z. Again following the same arguments as in the proof of Theorem 4.3.1 we get that

$$\lim_{n\to\infty} D_f\left(x_{n+1}, x_n\right) = 0.$$

For any $i = 1, 2, \ldots, N$, it follows from the three point identity (see (1.2.2)) that

$$\begin{aligned}
& D_f\left(\operatorname{proj}^f_{H^i_n}(x_n), x_n\right) - D_f\left(y^i_n, x_n\right) \\
&= D_f\left(\operatorname{proj}^f_{H^i_n}(x_n), y^i_n\right) + \left\langle \nabla f(x_n) - \nabla f\left(y^i_n\right), y^i_n - \operatorname{proj}^f_{H^i_n}(x_n)\right\rangle \\
&\geqslant \left\langle \nabla f(x_n) - \nabla f\left(y^i_n\right), y^i_n - \operatorname{proj}^f_{H^i_n}(x_n)\right\rangle \\
&= \left\langle \lambda^i_n \xi^i_n, y^i_n - \operatorname{proj}^f_{H^i_n}(x_n)\right\rangle \geqslant 0
\end{aligned}$$

because $\operatorname{proj}^f_{H^i_n}(x_n) \in H^i_n$. Since, in addition, $x_{n+1} \in H^i_n$, we also have

$$D_f(x_{n+1}, x_n) \geqslant D_f\left(\operatorname{proj}^f_{H^i_n}(x_n), x_n\right) \geqslant D_f\left(y^i_n, x_n\right).$$

Hence $\lim_{n\to\infty} D_f\left(y^i_n, x_n\right) = 0$. Since $\{x_n\}_{n\in\mathbb{N}}$ is bounded, Proposition 1.2.46 now implies that $\lim_{n\to\infty} \left\|y^i_n - x_n\right\| = 0$. In addition, we have from the definition of W^f (see (1.2.24)) and Proposition 1.2.42(ii) that

$$\begin{aligned}
D_f(x_{n+1}, z_n) &= D_f\left(x_{n+1}, \nabla f^*\left(\alpha_n \nabla f(x_n) + (1-\alpha_n)\nabla f(e_n)\right)\right) \qquad (4.3.11) \\
&= W^f\left(\alpha_n \nabla f(x_n) + (1-\alpha_n)\nabla f(e_n), x_{n+1}\right) \\
&\leqslant \alpha_n W^f\left(\nabla f(x_n), x_{n+1}\right) + (1-\alpha_n) W^f\left(\nabla f(e_n), x_{n+1}\right) \\
&= \alpha_n D_f(x_{n+1}, x_n) + (1-\alpha_n) D_f(x_{n+1}, e_n).
\end{aligned}$$

The sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{e_n\}_{n\in\mathbb{N}}$ are bounded and since f and consequently ∇f are bounded on bounded subsets of X (see Proposition 1.1.15), it follows that $\{D_f(x_{n+1}, e_n)\}_{n\in\mathbb{N}}$ is also bounded. Since $\lim_{n\to\infty} \alpha_n = 1$, it follows that

$$\lim_{n\to\infty} D_f(x_{n+1}, z_n) = 0.$$

As we have already noted, the sequences $\{\nabla f(x_n)\}_{n\in\mathbb{N}}$ and $\{\nabla f(e_n)\}_{n\in\mathbb{N}}$ are bounded and since ∇f^* is bounded on bounded subsets of $\operatorname{int}\operatorname{dom} f^*$, it follows that $\{z_n\}_{n\in\mathbb{N}}$ is also bounded. Proposition 1.2.46 now implies that $\lim_{n\to\infty} \|x_{n+1} - z_n\| = 0$. Then it follows that

$$\lim_{n\to\infty} \|x_n - z_n\| = 0$$

because

$$\|x_n - z_n\| \leqslant \|x_n - x_{n+1}\| + \|x_{n+1} - z_n\|.$$

Therefore for any $i = 1, 2, \ldots, N$, it follows that $\lim_{n\to\infty} \left\|y^i_n - z_n\right\| = 0$.

Since f is uniformly Fréchet differentiable and bounded on bounded subsets of X, it follows from Proposition 1.1.22(ii) that

$$\lim_{n\to\infty}\left\|\nabla f\left(y_n^i\right)-\nabla f\left(x_n+e_n^i\right)\right\|_*=0 \tag{4.3.12}$$

for any $i=1,2,\ldots,N$. The rest of the proof follows as the proof of Theorem 4.3.7. □

Remark 4.3.10 (Convergence under different assumptions). *In Theorems 4.3.1, 4.3.7 and 4.3.9 we can replace the assumptions that* $\liminf_{n\to\infty}\lambda_n>0$ *and* f *is bounded and uniformly Fréchet differentiable on bounded subsets of* X *with the assumption that* $\lim_{n\to\infty}\lambda_n=+\infty$. ◇

In the case of Algorithm (4.3.10) we can also prove the following result, based on Theorem 4.3.9.

Theorem 4.3.11 (Algorithm (4.3.10) is well-defined - zero free case). *Let* $A:X\to 2^{X^*}$ *be maximal monotone mappings such that* $Z:=\bigcap_{i=1}^N A_i^{-1}\left(0^*\right)\neq\varnothing$. *Let* $f:X\to\mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X. *Suppose that* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom}f^*$. *Let* $\{\alpha_n\}_{n\in\mathbb{N}}\subset[0,1]$ *and* $\{e_n\}_{n\in\mathbb{N}}$ *be a sequence of errors. Then, for each* $x_0\in X$, *there are sequences* $\{x_n\}_{n\in\mathbb{N}}$ *which satisfy Algorithm* (4.3.10) *with* $N=1$.

If $\lim_{n\to\infty}\alpha_n=1$, $\|e_n\|\leqslant M$ *(*M *is a positive constant) and* $\liminf_{n\to\infty}\lambda_n>0$, *then either* $A^{-1}\left(0^*\right)\neq\varnothing$ *and each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}^f_{A^{-1}(0^*)}(x_0)$ *as* $n\to\infty$, *or* $A^{-1}\left(0^*\right)=\varnothing$ *and each such sequence* $\{x_n\}_{n\in\mathbb{N}}$ *satisfies* $\lim_{n\to\infty}\|x_n\|=+\infty$.

4.4 Iterative Methods Based on the Bauschke-Combettes Algorithm

Another modification of the classical proximal point algorithm (see Algorithm (4.3.1)) has been proposed by Bauschke and Combettes [9], who also have modified the proximal point algorithm in order to generate a strongly convergent sequence. They introduced, for example, the following algorithm (see [9, Corollary 6.1(ii), page 258] for a single operator and $\lambda_n=1/2$).

Bauschke-Combettes Proximal Point Algorithm II
Input: $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$
Initialization: $x_0 \in \mathcal{H}$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n = R_{\lambda_n A}\left(x_n\right), \\ C_n = \left\{z \in \mathcal{H} : \|y_n - z\| \leqslant \|x_n - z\|\right\}, \\ Q_n = \left\{z \in \mathcal{H} : \langle x_0 - x_n, z - x_n\rangle \leqslant 0\right\}, \\ x_{n+1} = P_{C_n \cap Q_n}\left(x_0\right). \end{cases} \tag{4.4.1}$$

They prove that if $A^{-1}(0)$ is nonempty and $\liminf_{n\to\infty} \lambda_n > 0$, then the sequence generated by Algorithm (4.4.1) converges strongly to $P_{A^{-1}(0)}$. Wei and Thou [111] generalized this result to those Banach spaces X which are both uniformly convex and uniformly smooth (see Definition 1.1.33(iii) and (iv)). They introduced the following algorithm.

Wei-Zhou Proximal Point Algorithm
Input: $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n = \operatorname{Res}^{(1/2)\|\cdot\|^2}_{\lambda_n A}\left(x_n\right), \\ C_n = \left\{z \in X : D_{(1/2)\|\cdot\|^2}\left(z, y_n\right) \leqslant D_{(1/2)\|\cdot\|^2}\left(z, x_n\right)\right\}, \\ Q_n = \left\{z \in X : \langle J_X x_0 - J_X x_n, z - x_n\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}^{(1/2)\|\cdot\|^2}_{C_n \cap Q_n}\left(x_0\right). \end{cases} \tag{4.4.2}$$

They prove that if $A^{-1}\left(0^*\right)$ is nonempty and $\liminf_{n\to\infty} \lambda_n > 0$, then the sequence generated by Algorithm (4.4.2) converges strongly to $\operatorname{proj}^{(1/2)\|\cdot\|^2}_{A^{-1}\left(0^*\right)}$.

We extend Algorithms (4.4.1) and (4.4.2) to general reflexive Banach spaces using a well chosen convex function f. More precisely, we introduce the following algorithm.

***f*-Bauschke-Combettes Proximal Point Algorithm I**
Input: $f : X \to \mathbb{R}$ and $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,+\infty)$
Initialization: $x_0 \in X$.
General Step $(n = 1, 2, \ldots)$**:**

$$\begin{cases} y_n = \operatorname{Res}^f_{\lambda_n A}\left(x_n\right), \\ C_n = \left\{z \in X : D_f\left(z, y_n\right) \leqslant D_f\left(z, x_n\right)\right\}, \\ Q_n = \left\{z \in X : \langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}^f_{C_n \cap Q_n}\left(x_0\right). \end{cases} \tag{4.4.3}$$

As we have already noted in Section 4.3, Algorithm (4.4.3) is more flexible than Algorithm (4.4.2) because it leaves us the freedom of fitting the function f to the nature of the mapping

A (especially when A is the subdifferential of some function) and of the space X in ways which make the application of Algorithm (4.4.3) simpler than that of Algorithm (4.4.2). It should be observed that if X is a Hilbert space $\mathcal{H}$, then using in Algorithm (4.4.3) the function $f = (1/2)\left\|\cdot\right\|^2$, one obtains exactly Algorithm (4.4.1). If X is not a Hilbert space, but still a uniformly convex and uniformly smooth Banach space X (see Definition 1.1.33(iii) and (iv)), then setting $f = (1/2)\left\|\cdot\right\|^2$ in Algorithm (4.4.3), one obtains exactly Algorithm (4.4.2).

We study the following algorithm when $Z := \bigcap_{i=1}^{N} A_i^{-1}\left(0^*\right) \neq \varnothing$.

f-Bauschke-Combettes Proximal Point Algorithm II
Input: $f : X \to \mathbb{R}$ and $\{\lambda_n^i\}\, n \in \mathbb{N} \subset (0, +\infty)$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$)**:**

$$\begin{cases} y_n^i = \operatorname{Res}_{\lambda_n^i A_i}^f \left(x_n + e_n^i\right), \\ C_n^i = \left\{z \in X : D_f\left(z, y_n^i\right) \leqslant D_f\left(z, x_n + e_n^i\right)\right\}, \\ C_n := \bigcap_{i=1}^{N} C_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \operatorname{proj}_{C_n \cap Q_n}^f \left(x_0\right). \end{cases} \tag{4.4.4}$$

In the following result we prove that Algorithm (4.4.4) generates a sequence which converges strongly to a common zero of the finite family of maximal monotone mappings $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$ (*cf.* [90, Theorem 4.2, page 35]).

Theorem 4.4.1 (Convergence result for Algorithm (4.4.4))**.** *Let $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$, be N maximal monotone mappings such that $Z := \bigcap_{i=1}^{N} A_i^{-1}\left(0^*\right) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Then, for each $x_0 \in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (4.4.4)*. If for each $i = 1, 2, \ldots, N$, $\liminf_{n\to\infty} \lambda_n^i > 0$, and the sequence of errors $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to\infty} \|e_n^i\| = 0$, then each such converges strongly to* $\operatorname{proj}_Z^f(x_0)$ *as $n \to \infty$.*

Proof. Note that $\operatorname{dom} \nabla f = X$ because $\operatorname{dom} f = X$ and f is Legendre (see Definition 1.2.7). Hence it follows from Proposition 4.1.4 that $\operatorname{dom} \operatorname{Res}_{\lambda A}^f = X$. Denote $S_n^i := \operatorname{Res}_{\lambda_n^i A_i}^f$. Therefore from Proposition 4.1.2(iv)(b) and Figure 1.3 we have that each S_n^i is BFNE and therefore QBNE. We also have $\Omega = Z$ and we see that Condition 1 holds and we can apply our lemmata.

By Lemmata 3.3.3 and 3.3.5, any sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by Algorithm (4.4.4) is well defined and bounded. From now on we let $\{x_n\}_{n\in\mathbb{N}}$ be an arbitrary sequence which is generated by Algorithm (4.4.4).

We claim that every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to Z. From Lemma 3.3.6 we have

$$\lim_{n\to\infty}\left\|\nabla f\left(y_n^i\right)-\nabla f\left(x_n+e_n^i\right)\right\|_*=0, \tag{4.4.5}$$

for any $i=1,2,\ldots,N$. The rest of the proof follows as the proof of Theorem 4.3.7. □

Now we propose two algorithms for finding common zeroes of finitely many maximal monotone mappings. Both algorithms are based on products of f-resolvents. For earlier results based on this method see, for example, [13, 32, 88, 95].

Algorithm (4.4.4) finds common zeroes of finitely many maximal monotone mappings. In this algorithm we build, at each step, N copies of the half-space C_n with respect to each mapping. Then the next iteration is the Bregman projection onto the intersection of $N+1$ half-spaces (N copies of C_n and Q_n). Now we propose a new variant of Algorithm (4.4.4) which also finds common zeroes of finitely many maximal monotone mappings. In the new algorithm we use the concept of products of resolvents and therefore we build, at each step, only one copy of the half-space C_n. Then the next iteration is the Bregman projection onto the intersection of two half-spaces (C_n and Q_n).

Sabach Proximal Point Algorithm I
Input: $f:X\to\mathbb{R}$ and $\{\lambda_n^i\}\,n\in\mathbb{N}\subset(0,+\infty)$, $i=1,2,\ldots,N$.
Initialization: $x_0\in X$.
General Step ($n=1,2,\ldots$):

$$\begin{cases} y_n=\mathrm{Res}^f_{\lambda_n^N A_N}\circ\cdots\circ\mathrm{Res}^f_{\lambda_n^1 A_1}\left(x_n+e_n\right), \\ C_n=\left\{z\in X: D_f\left(z,y_n\right)\leqslant D_f\left(z,x_n+e_n\right)\right\}, \\ Q_n=\left\{z\in X:\left\langle\nabla f\left(x_0\right)-\nabla f\left(x_n\right),z-x_n\right\rangle\leqslant 0\right\}, \\ x_{n+1}=\mathrm{proj}^f_{C_n\cap Q_n}\left(x_0\right). \end{cases} \tag{4.4.6}$$

The following algorithm is a modification of Algorithm (4.4.6), where at any step we calculate the Bregman projection onto only one set which is not necessarily a half-space. Even if we only project onto one set, the computation of the projection is harder since this set is a general convex set. We present and analyze this algorithm. Its proof is very similar to the one of Algorithm (4.4.6). More precisely, we introduce the following. algorithm

Sabach Proximal Point Algorithm II
Input: $f:X\to\mathbb{R}$ and $\{\lambda_n^i\}\,n\in\mathbb{N}\subset(0,+\infty)$, $i=1,2,\ldots,N$.
Initialization: $x_0\in X$.
General Step ($n=1,2,\ldots$):

$$\begin{cases} y_n=\mathrm{Res}^f_{\lambda_n^N A_N}\circ\cdots\circ\mathrm{Res}^f_{\lambda_n^1 A_1}\left(x_n+e_n\right), \\ C_{n+1}=\left\{z\in C_n: D_f\left(z,y_n^i\right)\leqslant D_f\left(z,x_n+e_n\right)\right\}, \\ x_{n+1}=\mathrm{proj}^f_{C_{n+1}}\left(x_0\right). \end{cases} \tag{4.4.7}$$

We have the following theorem (*cf.* [103, Theorem 3.1, page 1297]).

Theorem 4.4.2 (Convergence results for Algorithms (4.4.6) and (4.4.7))**.** *Let* $A_i : X \to 2^{X^*}$, $i = 1, 2, \ldots, N$, *be* N *maximal monotone mappings with* $Z := \bigcap_{i=1}^{N} A_i^{-1}\left(0^*\right) \neq \varnothing$. *Let* $f : X \to \mathbb{R}$ *be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of* X. *Suppose that* ∇f^* *is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Then, for each* $x_0 \in X$, *the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *which is generated either by Algorithm* (4.4.6) *or by Algorithm* (4.4.7) *is well defined. If the sequence of errors* $\{e_n\}_{n\in\mathbb{N}} \subset X$ *satisfies* $\lim_{n\to\infty} \|e_n\| = 0$ *and for each* $i = 1, 2, \ldots, N$, $\liminf_{n\to\infty} \lambda_n^i > 0$, *then the sequence* $\{x_n\}_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z^f(x_0)$ *as* $n \to \infty$.

Proof. Note that $\operatorname{dom} \nabla f = X$ because $\operatorname{dom} f = X$ and f is Legendre (see Definition 1.2.7). Hence it follows from Proposition 4.1.4 that $\operatorname{dom}\operatorname{Res}_{\lambda A}^f = X$. We denote by T_n^i the f-resolvent $\operatorname{Res}_{\lambda_n^i A_i}^f$ and by S_n^i the composition $T_n^i \circ \cdots \circ T_n^1$ for any $i = 1, 2, \ldots, N$ and for each $n \in \mathbb{N}$. Therefore $y_n = T_n^N \circ \cdots \circ T_n^1 (x_n + e_n) = S_n^N (x_n + e_n)$. We also assume that $S_n^0 = I$, where I is the identity operator.

From Proposition 4.1.2(iv)(b), Proposition 2.1.2 and Figure 1.3 we have that each T_n^i, $i = 1, 2, \ldots, N$, is BSNE. Therefore Proposition 2.1.12 now implies that also S_n^i is BSNE and therefore QBNE. From Remark 2.1.13 we have that $\operatorname{Fix}(S_n^i) = \bigcap_{i=1}^{n} \operatorname{Fix}(T_n^i)$.

Each f-resolvent $\operatorname{Res}_{\lambda_n^i A_i}^f$ is a QBNE operator and therefore S_n^N, a composition of QBNE operators, is also QBNE. Hence we get from (1.3.8) that

$$\begin{aligned} D_f(u, y_n) &= D_f\left(u, \operatorname{Res}_{\lambda_n^N A_N}^f \circ \cdots \circ \operatorname{Res}_{\lambda_n^1 A_1}^f (x_n + e_n)\right) = D_f\left(u, S_n^N (x_n + e_n)\right) \\ &\leqslant D_f\left(u, S_n^i (x_n + e_n)\right) \leqslant D_f(u, x_n + e_n) \end{aligned} \tag{4.4.8}$$

for any $i = 1, 2, \ldots, N-1$.

We have $\Omega = Z$ and therefore Condition 1 holds. Hence we can apply our lemmata.

From Lemmata 3.3.3 and 3.3.5, any sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by either Algorithm (4.4.6) or by Algorithm (4.4.7) is well defined and bounded. From now on we let $\{x_n\}_{n\in\mathbb{N}}$ be an arbitrary sequence which is generated by Algorithm (4.4.6) or by Algorithm (4.4.7).

We claim that every weak subsequential limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to Z. From Lemma 3.3.6 we have

$$\lim_{n\to\infty} \|y_n - (x_n + e_n)\| = 0, \quad \lim_{n\to\infty} \left(f(y_n) - f(x_n + e_n)\right) = 0$$

and

$$\lim_{n\to\infty} \left\|\nabla f(x_n + e_n) - \nabla f(y_n)\right\|_* = 0. \tag{4.4.9}$$

Hence, from the definition of the Bregman distance (see (1.2.1)), we get that

$$\lim_{n\to\infty} D_f(y_n, x_n + e_n) = \lim_{n\to\infty} \left[f(y_n) - f(x_n + e_n) - \langle \nabla f(x_n + e_n), y_n - (x_n + e_n)\rangle\right] = 0. \tag{4.4.10}$$

Let $u \in Z$. From the three point identity (see (1.2.2)) we obtain that

$$D_f(u, x_n + e_n) - D_f(u, y_n) = D_f(y_n, x_n + e_n) + \langle \nabla f(x_n + e_n) - \nabla f(y_n), y_n - u\rangle.$$

Since the sequence $\{y_n\}_{n\in\mathbb{N}}$ is bounded (see Lemma 3.3.5) we obtain from (4.4.9) and (4.4.10) that

$$\lim_{n\to\infty} \left(D_f(u, x_n + e_n) - D_f(u, y_n)\right) = 0. \tag{4.4.11}$$

Thence from (4.4.11) we get that

$$\lim_{n\to\infty} \left(D_f(u, x_n + e_n) - D_f\left(u, S_n^N(x_n + e_n)\right)\right) = 0$$

for any $u \in Z$. From (1.3.6), (1.3.8), (4.4.8) we get that

$$\begin{aligned} D_f\left(S_n^i(x_n + e_n), S_n^{i-1}(x_n + e_n)\right) &= D_f\left(T_n^i\left(S_n^{i-1}(x_n + e_n)\right), S_n^{i-1}(x_n + e_n)\right) \\ &\leqslant D_f\left(u, S_n^{i-1}(x_n + e_n)\right) - D_f\left(u, S_n^i(x_n + e_n)\right) \\ &\leqslant D_f(u, x_n + e_n) - D_f(u, y_n). \end{aligned}$$

Hence from (4.4.11) we get that

$$\lim_{n\to\infty} D_f\left(S_n^i(x_n + e_n), S_n^{i-1}(x_n + e_n)\right) = 0 \tag{4.4.12}$$

for any $i = 1, 2, \ldots, N$. Therefore from Proposition 1.2.46 and the fact that $\{S_n^i(x_n + e_n)\}_{n\in\mathbb{N}}$ is bounded (using similar arguments to those in the proof of Lemma 3.3.5), we obtain that

$$\lim_{n\to\infty} \left(S_n^i(x_n + e_n) - S_n^{i-1}(x_n + e_n)\right) = 0 \tag{4.4.13}$$

for any $i = 1, 2, \ldots, N$. From the three point identity (see (1.2.2)) we get that

$$\begin{aligned} &D_f\left(S_n^i(x_n + e_n), x_n + e_n\right) - D_f\left(S_n^{i-1}(x_n + e_n), x_n + e_n\right) \\ &= D_f\left(S_n^i(x_n + e_n), S_n^{i-1}(x_n + e_n)\right) \\ &+ \left\langle \nabla f(x_n + e_n) - \nabla f\left(S_n^{i-1}(x_n + e_n)\right), S_n^{i-1}(x_n + e_n) - S_n^i(x_n + e_n)\right\rangle. \end{aligned}$$

The sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{S_n^i(x_n + e_n)\}_{n\in\mathbb{N}}$ are bounded (see Lemma 3.3.5). Hence, from

(4.4.12) and (4.4.13) we get that

$$\lim_{n\to\infty}\left(D_f\left(S_n^i\left(x_n+e_n\right),x_n+e_n\right)-D_f\left(S_n^{i-1}\left(x_n+e_n\right),x_n+e_n\right)\right)=0. \tag{4.4.14}$$

Since

$$\lim_{n\to\infty}D_f\left(y_n,x_n+e_n\right)=\lim_{n\to\infty}D_f\left(S_n^N\left(x_n+e_n\right),x_n+e_n\right)=0,$$

we obtain from (4.4.14) that

$$\lim_{n\to\infty}D_f\left(S_n^i\left(x_n+e_n\right),x_n+e_n\right)=0$$

for any $i=1,2,\ldots,N$. Proposition 1.2.46 and the fact that $\{x_n+e_n\}_{n\in\mathbb{N}}$ is bounded (see Lemma 3.3.5), now imply that

$$\lim_{n\to\infty}\left\|S_n^i\left(x_n+e_n\right)-\left(x_n+e_n\right)\right\|=0 \tag{4.4.15}$$

for any $i=1,2,\ldots,N$, that is,

$$\lim_{n\to\infty}\left\|\mathrm{Res}^f_{\lambda_n^iA_i}\left(S_n^{i-1}\left(x_n+e_n\right)\right)-\left(x_n+e_n\right)\right\|=0$$

for any $i=1,2,\ldots,N$. From the definition of the f-resolvent (see (4.0.2)), it follows that

$$\nabla f\left(S_n^{i-1}\left(x_n+e_n\right)\right)\in\left(\nabla f+\lambda_n^iA_i\right)\left(S_n^i\left(x_n+e_n\right)\right).$$

Hence

$$\xi_n^i:=\frac{1}{\lambda_n^i}\left(\nabla f\left(S_n^{i-1}\left(x_n+e_n\right)\right)-\nabla f\left(S_n^i\left(x_n+e_n\right)\right)\right)\in A_i\left(S_n^i\left(x_n+e_n\right)\right) \tag{4.4.16}$$

for any $i=1,2,\ldots,N$. Applying Proposition 1.1.22(ii) to (4.4.13) we get that

$$\lim_{n\to\infty}\left\|\nabla f\left(S_n^{i-1}\left(x_n+e_n\right)\right)-\nabla f\left(S_n^i\left(x_n+e_n\right)\right)\right\|_*=0.$$

Now let $\{x_{n_k}\}_{k\in\mathbb{N}}$ be a weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and denote its weak limit by v. Then from (4.4.15) it follows that $\left\{S_{n_k}^i\left(x_{n_k}+e_{n_k}\right)\right\}_{k\in\mathbb{N}}$, $i=1,2,\ldots,N$, also converges weakly to v. Since $\liminf_{n\to\infty}\lambda_n^i>0$, it follows from (4.4.16) that

$$\lim_{n\to\infty}\left\|\xi_n^i\right\|_*=0^*$$

for any $i=1,2,\ldots,N$. From the monotonicity of A_i it follows that

$$\left\langle\eta-\xi_n^i,z-S_{n_k}^i\left(x_{n_k}+e_{n_k}\right)\right\rangle\geqslant 0$$

for all $(z, \eta) \in \operatorname{graph} A_i$ and for all $i = 1, 2, \ldots, N$. This, in turn, implies that $\langle \eta, z - v \rangle \geqslant 0$ for all $(z, \eta) \in \operatorname{graph} A_i$ for any $i = 1, 2, \ldots, N$. Therefore, using the maximal monotonicity of A_i (see Proposition 1.4.13), we now obtain that $v \in A_i^{-1}(0^*)$ for each $i = 1, 2, \ldots, N$. Thus $v \in Z$ and this proves the result.

Now Theorem 4.4.2 is seen to follow from Lemma 3.3.7. $\square$

Chapter 5

Applications - Equilibrium, Variational and Convex Feasibility Problems

In this chapter we modify the iterative methods proposed in Chapters 3 and 4 in order to solve diverse optimization problems. We focus our study on the following three problems.

(i) Equilibrium Problem (EP). Given a subset K of a Banach space X, and a bifunction $g : K \times K \to \mathbb{R}$, the equilibrium problem corresponding to g is to find $\bar{x} \in K$ such that

$$g(\bar{x}, y) \geqslant 0 \quad \forall y \in K. \tag{5.0.1}$$

(ii) Variational Inequality Problem (VIP). Given a subset K of a Banach space X, and a single-valued mapping $A : X \to 2^{X^*}$, the corresponding variational inequality is to find $\bar{x} \in K$ such that there exists $\xi \in A\bar{x}$ with

$$\langle \xi, y - \bar{x} \rangle \geqslant 0 \quad \forall y \in K. \tag{5.0.2}$$

(iii) Convex Feasibility Problem (CFP). Given N nonempty, closed and convex subsets K_i, $i = 1, 2, \ldots, N$, of a Banach space X, the convex feasibility problem is to find an element in the assumed nonempty intersection $\bigcap_{i=1}^{N} K_i$.

Thence this chapter is divided into three sections concerning each problem.

5.1 Equilibrium Problems

The equilibrium problem contains as special cases many optimization, fixed point and variational inequality problems (see [21, 49] for more details).

It is well known that many interesting and complicated problems in nonlinear analysis, such as complementarity, fixed point, Nash equilibrium, optimization, saddle point and variational inequality problems, can all be formulated as equilibrium problems as in (5.0.1) (see, *e.g.*, [21]). There are several papers available in the literature which are devoted to this problem. Most of the work on this issue deals with conditions for the existence of solutions (see, for example, [61, 63]). However, there are only a few papers that deal with iterative procedures for solving equilibrium problems in finite as well as infinite-dimensional spaces (see, for instance, [49, 62, 92, 93, 94, 106, 107]).

As in the case of finding zeroes of monotone mappings (see Chapter 4), the key tool for solving equilibrium problems is to define a resolvent (see [49] for the case of Hilbert spaces), this time with respect to a bifunction g instead of with respect to a mapping A (see (4.0.2)).

Definition 5.1.1 (Resolvent of bifunctions)**.** *The resolvent of a bifunction* $g : K \times K \to \mathbb{R}$ *is the operator* $\mathrm{Res}_g^f : X \to 2^K$*, defined by*

$$\mathrm{Res}_g^f(x) = \{z \in K : g(z, y) + \langle \nabla f(z) - \nabla f(x), y - z \rangle \geqslant 0 \quad \forall y \in K\}. \tag{5.1.1}$$

Actually there is a strong connection between the resolvent Res_g^f and the f-resolvent Res_A^f. We will show this in the next section.

5.1.1 Properties of Resolvents of Bifunctions

It is well known that for studying equilibrium problems, it is assumed that the corresponding bifunction g satisfies the following four assumptions (see, for example, [21]).

Assumption 1 (Basic assumptions on bifunctions)**.** *Given a subset* K *of a Banach space* X *and a bifunction* $g : K \times K \to \mathbb{R}$*, we make the following assumptions.*

(C1) $g(x, x) = 0$ *for all* $x \in K$.

(C2) g *is monotone, i.e.,* $g(x, y) + g(y, x) \leqslant 0$ *for all* $x, y \in K$.

(C3) *For all* $x, y, z \in K$, *we have*

$$\limsup_{t \downarrow 0} g\left(tz + (1-t)\, x, y\right) \leqslant g\left(x, y\right).$$

(C4) *For each* $x \in K$, $g\left(x, \cdot\right)$ *is convex and lower semicontinuous.*

In the following two lemmata we obtain several properties of these resolvents. We first show that $\operatorname{dom} \operatorname{Res}_g^f$ is the whole space X when f is a super-coercive (see Definition 1.2.33(ii)) and Gâteaux differentiable function (*cf.* [92, Lemma 1, page 130]).

Proposition 5.1.2 (Sufficient condition for $\operatorname{dom} \operatorname{Res}_g^f = X$)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a super-coercive and Gâteaux differentiable function. Let* K *be a closed and convex subset of* X. *If the bifunction* $g : K \times K \to \mathbb{R}$ *satisfies Conditions (C1)–(C4), then* $\operatorname{dom} \operatorname{Res}_g^f = X$.

Proof. First we show that for any $\xi \in X^*$, there exists $\bar{x} \in K$ such that

$$g\left(\bar{x}, y\right) + f\left(y\right) - f\left(\bar{x}\right) - \left\langle \xi, y - \bar{x} \right\rangle \geqslant 0 \tag{5.1.2}$$

for any $y \in K$. Since f is a super-coercive function, a function $h : X \times X \to (-\infty, +\infty]$, defined by

$$h\left(x, y\right) := f\left(y\right) - f\left(x\right) - \left\langle \xi, y - x \right\rangle,$$

satisfies

$$\lim_{\left\| x - y \right\| \to \infty} \frac{h\left(x, y\right)}{\left\| x - y \right\|} = -\infty$$

for each fixed $y \in K$. Therefore it follows from [21, Theorem 1, page 127] that (5.1.2) holds. Now we prove that (5.1.2) implies that

$$g\left(\bar{x}, y\right) + \left\langle \nabla f\left(\bar{x}\right), y - \bar{x} \right\rangle - \left\langle \xi, y - \bar{x} \right\rangle \geqslant 0$$

for any $y \in K$. We know that (5.1.2) holds for $y = t\bar{x} + (1-t)\,\bar{y}$, where $\bar{y} \in K$ and $t \in (0, 1)$. Hence

$$g\left(\bar{x}, t\bar{x} + (1-t)\,\bar{y}\right) + f\left(t\bar{x} + (1-t)\,\bar{y}\right) - f\left(\bar{x}\right) - \left\langle \xi, t\bar{x} + (1-t)\,\bar{y} - \bar{x} \right\rangle \geqslant 0 \tag{5.1.3}$$

for all $\bar{y} \in K$. Since

$$f\left(t\bar{x} + (1-t)\,\bar{y}\right) - f\left(\bar{x}\right) \leqslant \left\langle \nabla f\left(t\bar{x} + (1-t)\,\bar{y}\right), t\bar{x} + (1-t)\,\bar{y} - \bar{x} \right\rangle,$$

we get from (5.1.3) and Condition (C4) that

$$tg(\bar{x},\bar{x}) + (1-t)\,g(\bar{x},\bar{y}) + \langle \nabla f(t\bar{x} + (1-t)\,\bar{y}), t\bar{x} + (1-t)\,\bar{y} - \bar{x}\rangle$$
$$- \langle \xi, t\bar{x} + (1-t)\,\bar{y} - \bar{x}\rangle \geqslant 0$$

for all $\bar{y} \in K$. From Condition (C1) we know that $g(\bar{x},\bar{x}) = 0$. So, we have

$$(1-t)\,g(\bar{x},\bar{y}) + \langle \nabla f(t\bar{x} + (1-t)\,\bar{y}), (1-t)(\bar{y} - \bar{x})\rangle - \langle \xi, (1-t)(\bar{y} - \bar{x})\rangle \geqslant 0$$

and

$$(1-t)\left[g(\bar{x},\bar{y}) + \langle \nabla f(t\bar{x} + (1-t)\,\bar{y}), \bar{y} - \bar{x}\rangle - \langle \xi, \bar{y} - \bar{x}\rangle\right] \geqslant 0$$

for all $\bar{y} \in K$. Therefore

$$g(\bar{x},\bar{y}) + \langle \nabla f(t\bar{x} + (1-t)\,\bar{y}), \bar{y} - \bar{x}\rangle - \langle \xi, \bar{y} - \bar{x}\rangle \geqslant 0$$

for all $\bar{y} \in K$. Since f is a Gâteaux differentiable function, it follows that ∇f is norm-to-weak* continuous (see Proposition 1.1.21). Therefore, letting here $t \to 1^-$, we obtain that

$$g(\bar{x},\bar{y}) + \langle \nabla f(\bar{x}), \bar{y} - \bar{x}\rangle - \langle \xi, \bar{y} - \bar{x}\rangle \geqslant 0$$

for all $\bar{y} \in K$. Hence, for any $x \in X$, taking $\xi = \nabla f(x)$, we obtain $\bar{x} \in K$ such that

$$g(\bar{x},\bar{y}) + \langle \nabla f(\bar{x}) - \nabla f(x), \bar{y} - \bar{x}\rangle \geqslant 0$$

for all $\bar{y} \in K$, that is, $\bar{x} \in \operatorname{Res}_g^f(x)$. Hence $\operatorname{dom}\operatorname{Res}_g^f = X$. □

In the next lemma we list more properties of resolvents of bifunctions (*cf.* [92, Lemma 2, page 131]).

Proposition 5.1.3 (Properties of resolvents of bifunctions)**.** *Let $f : X \to (-\infty, +\infty]$ be a Legendre function. Let K be a closed and convex subset of X. If a bifunction $g : K \times K \to \mathbb{R}$ satisfies Conditions (C1)–(C4), then the following assertions hold.*

(i) *The resolvent Res_g^f is single-valued.*

(ii) *The resolvent Res_g^f is an BFNE operator.*

(iii) *The fixed point set of Res_g^f is the solutions set of the corresponding equilibrium problem, i.e.,* $\operatorname{Fix}\left(\operatorname{Res}_g^f\right) = \operatorname{EP}(g)$.

(iv) *The set $\operatorname{EP}(g)$ is closed and convex.*

Proof. (i) Let $z_1, z_2 \in \mathrm{Res}_g^f(x)$. From the definition of the resolvent Res_g^f (see (5.1.1)) we obtain

$$g(z_1, z_2) + \langle \nabla f(z_1) - \nabla f(x), z_2 - z_1 \rangle \geqslant 0$$

and

$$g(z_2, z_1) + \langle \nabla f(z_2) - \nabla f(x), z_1 - z_2 \rangle \geqslant 0.$$

Summing up these two inequalities, we get

$$g(z_1, z_2) + g(z_2, z_1) + \langle \nabla f(z_2) - \nabla f(z_1), z_1 - z_2 \rangle \geqslant 0.$$

From Condition (C2) it follows that

$$\langle \nabla f(z_2) - \nabla f(z_1), z_1 - z_2 \rangle \geqslant 0.$$

The function f is Legendre (see Definition 1.2.7) and therefore it is strictly convex. Hence ∇f is strictly monotone (see Example 1.4.3) and therefore $z_1 = z_2$.

(ii) For any $x, y \in K$, we have

$$g\left(\mathrm{Res}_g^f(x), \mathrm{Res}_g^f(y)\right) + \left\langle \nabla f\left(\mathrm{Res}_g^f(x)\right) - \nabla f(x), \mathrm{Res}_g^f(y) - \mathrm{Res}_g^f(x) \right\rangle \geqslant 0$$

and

$$g\left(\mathrm{Res}_g^f(y), \mathrm{Res}_g^f(x)\right) + \left\langle \nabla f\left(\mathrm{Res}_g^f(y)\right) - \nabla f(y), \mathrm{Res}_g^f(x) - \mathrm{Res}_g^f(y) \right\rangle \geqslant 0.$$

Summing up these two inequalities, we obtain that

$$\begin{aligned} & g\left(\mathrm{Res}_g^f(x), \mathrm{Res}_g^f(y)\right) + g\left(\mathrm{Res}_g^f(y), \mathrm{Res}_g^f(x)\right) \\ & + \left\langle \nabla f\left(\mathrm{Res}_g^f(x)\right) - \nabla f(x) + \nabla f(y) - \nabla f\left(\mathrm{Res}_g^f(y)\right), \mathrm{Res}_g^f(y) - \mathrm{Res}_g^f(x) \right\rangle \geqslant 0. \end{aligned}$$

From Condition (C2) it follows that

$$\left\langle \nabla f\left(\mathrm{Res}_g^f(x)\right) - \nabla f(x) + \nabla f(y) - \nabla f\left(\mathrm{Res}_g^f(y)\right), \mathrm{Res}_g^f(y) - \mathrm{Res}_g^f(x) \right\rangle \geqslant 0.$$

Hence

$$\begin{aligned} & \left\langle \nabla f\left(\mathrm{Res}_g^f(x)\right) - \nabla f\left(\mathrm{Res}_g^f(y)\right), \mathrm{Res}_g^f(x) - \mathrm{Res}_g^f(y) \right\rangle \\ & \leqslant \left\langle \nabla f(x) - \nabla f(y), \mathrm{Res}_g^f(x) - \mathrm{Res}_g^f(y) \right\rangle. \end{aligned}$$

This means that Res_g^f is an BFNE operator (see (1.3.4)), as claimed.

(iii) Indeed,

$$x \in \operatorname{Fix}\left(\operatorname{Res}_g^f\right) \iff x = \operatorname{Res}_g^f(x) \iff 0 \leqslant g(x,y) + \langle \nabla f(x) - \nabla f(x), y - x\rangle \quad \forall y \in K,$$

therefore

$$x \in \operatorname{Fix}\left(\operatorname{Res}_g^f\right) \iff 0 \leqslant g(x,y) \quad \forall y \in K \iff x \in \operatorname{EP}(g).$$

Therefore $\operatorname{Fix}\left(\operatorname{Res}_g^f\right) = \operatorname{EP}(g)$.

(iv) Since Res_g^f is a BFNE operator, the result follows immediately from Proposition 2.1.1 because of point (iii). □

As we have already noted, there is a strong connection between zeroes of maximal monotone mappings and solutions of equilibrium problems of bifunctions. Let $g : K \times K \to \mathbb{R}$ be a bifunction and define the mapping $A_g : X \to 2^{X^*}$ in the following way:

$$A_g(x) := \begin{cases} \{\xi \in X^* : g(x,y) \geqslant \langle \xi, y - x\rangle \quad \forall y \in K\} &, \quad x \in K \\ \varnothing &, \quad x \notin K. \end{cases} \tag{5.1.4}$$

In the following result we show that under suitable assumptions on the function f, the mapping A_g generated from a bifunction g is maximal monotone (see Definition 1.4.9).

Proposition 5.1.4 (Properties of A_g). *Let $f : X \to (-\infty, +\infty]$ be a super-coercive, Legendre, Fréchet differentiable and totally convex function. Let K be a closed and convex subset of X and assume that a bifunction $g : K \times K \to \mathbb{R}$ satisfies Conditions (C1)–(C4), then the following assertions hold.*

(i) $\operatorname{EP}(g) = A_g^{-1}(0^*)$.

(ii) *The mapping $A_g : X \to 2^{X^*}$ is maximal monotone.*

(iii) $\operatorname{Res}_g^f = \operatorname{Res}_{A_g}^f$.

Proof. (i) If $x \in K$ then from the definition of the mapping A_g (see (5.1.4)) we get that

$$x \in A_g^{-1}\left(0^*\right) \Leftrightarrow g(x,y) \geqslant 0 \quad \forall y \in K \Leftrightarrow x \in \operatorname{EP}(g).$$

(ii) We first prove that A_g is monotone mapping (see Definition 1.4.2(i)). Let (x, ξ) and (y, η) belong to the graph of A_g. From the definition of the mapping A_g (see (5.1.4)) we get that

$$g(x,z) \geqslant \langle \xi, z - x\rangle \text{ and } g(y,z) \geqslant \langle \eta, z - y\rangle$$

for any $z \in K$. In particular we have that

$$g(x,y) \geqslant \langle \xi, y-x \rangle \text{ and } g(y,x) \geqslant \langle \eta, x-y \rangle .$$

From Condition (C2) we obtain that

$$0 \geqslant g(x,y) + g(y,x) \geqslant \langle \xi - \eta, y - x \rangle$$

that is $\langle \xi - \eta, x - y \rangle \geqslant 0$ which means that A_g is monotone mapping (see (1.4.1)). In order to show that A_g is maximal monotone mapping it is enough to show that $\operatorname{ran}(A_g + \nabla f) = X^*$ (see Proposition 1.4.18). Let $\xi \in X^*$, from Proposition 1.2.13 we get that under the assumption here, f is cofinite, that is, $\operatorname{dom} f^* = X^*$ and therefore $\operatorname{ran} \nabla f = \operatorname{int} \operatorname{dom} f^* = X^*$ (see (1.2.4)) which means that ∇f is surjective. Then there exists $x \in X$ such that $\nabla f(x) = \xi$. From Proposition 5.1.2 we know that the resolvent Res_g^f of g has full domain and therefore from the definition of Res_g^f (see (5.1.1)) we get that

$$g\left(\operatorname{Res}_g^f(x), y\right) + \left\langle \nabla f\left(\operatorname{Res}_g^f(x)\right) - \nabla f(x), y - \operatorname{Res}_g^f(x)\right\rangle \geqslant 0$$

for any $y \in K$, that is,

$$g\left(\operatorname{Res}_g^f(x), y\right) \geqslant \left\langle \nabla f(x) - \nabla f\left(\operatorname{Res}_g^f(x)\right), y - \operatorname{Res}_g^f(x)\right\rangle$$

for any $y \in K$. This shows that $\nabla f(x) - \nabla f\left(\operatorname{Res}_g^f(x)\right) \in A_g\left(\operatorname{Res}_g^f(x)\right)$ (see (5.1.4)). Therefore

$$\xi = \nabla f(x) \in (\nabla f + A_g)\left(\operatorname{Res}_g^f(x)\right) \tag{5.1.5}$$

which means that $\xi \in \operatorname{ran}(A_g + \nabla f)$. This completes the proof.

(iii) From Proposition 4.1.2(iv)(a) and assertion (ii) we have that the resolvent, $\operatorname{Res}_{A_g}^f$, of a maximal monotone mapping A_g is single-valued. From Proposition 5.1.3(ii) the resolvent Res_g^f is single-valued too. Now we obtain from (5.1.5) that

$$\operatorname{Res}_{A_g}^f = (A_g + \nabla f)^{-1} \circ \nabla f = \operatorname{Res}_g^f$$

as asserted. □

As we have seen in Propositions 5.1.2 and 5.1.3(ii), the operator $A = \operatorname{Res}_g^f$ is BFNE and with full domain. Therefore, from Proposition 4.1.6(iii) the mapping $B = \nabla f \circ A^{-1} - \nabla f$ is maximal monotone. This fact also follows from Proposition 5.1.4(ii) where we proved that A_g is a maximal monotone mapping. Therefore $B = A_g$. Indeed, from Proposition

5.1.4(iii)

$$B = \nabla f \circ \left(\mathrm{Res}_g^f\right)^{-1} - \nabla f = \nabla f \circ \left(\mathrm{Res}_{A_g}^f\right)^{-1} - \nabla f = A_g.$$

Now we will show the converse connection holds. Let $A : X \to 2^{X^*}$ be a maximal monotone mapping and define the bifunction g_A in the following way:

$$g_A(x, y) := \sup \{\langle \xi, y - x \rangle : \xi \in Ax\}. \tag{5.1.6}$$

In the following result we show that under appropriate assumptions on the function f, the bifunction g_A satisfies Conditions (C1)–(C4).

Proposition 5.1.5 (Properties of g_A). *Let $f : X \to (-\infty, +\infty]$ be a super-coercive, Legendre, Fréchet differentiable and totally convex function. Let $A : X \to 2^{X^*}$ be a maximal monotone mapping with nonempty, closed and convex domain $K = \operatorname{dom} A$. Then the following assertions hold.*

(i) *The bifunction g_A satisfies Conditions (C1)–(C4).*

(ii) $A_{g_A} = A$.

(iii) $\mathrm{EP}(g_A) = A^{-1}(0^*)$

(iv) $\mathrm{Res}_{g_A}^f = \mathrm{Res}_A^f$.

Proof. (i) We first prove that Condition (C1) holds. Let $x \in K$. It is clear that

$$g_A(x, x) = \sup \{\langle \xi, x - x \rangle : \xi \in Ax\} = 0.$$

Now we prove that Condition (C2) holds. Let (x, ξ) and (y, η) belong to the graph of A. Since A is monotone mapping (see (1.4.1)), we have $-\langle \eta, x - y \rangle \geqslant \langle \xi, y - x \rangle$, which implies that

$$\inf \{-\langle \eta, x - y \rangle : \eta \in Ay\} \geqslant \sup \{\langle \xi, y - x \rangle : \xi \in Ax\} = g_A(x, y).$$

On the other hand, we have

$$\inf \{-\langle \eta, x - y \rangle : \eta \in Ay\} = -\sup \{\langle \eta, x - y \rangle : \eta \in Ay\} = -g_A(y, x)$$

and therefore $g_A(x, y) + g_A(y, x) \leqslant 0$ for any $x, y \in K$. In order to prove that Condition

(C3) holds, we use the following fact:

$$\begin{aligned} g_A(tz+(1-t)x,y) &= \sup\{\langle \xi, y-tz-(1-t)x\rangle : \xi \in Ax\} \\ &= \sup\{\langle \xi, y-x\rangle : \xi \in Ax\} - t\sup\{\langle \xi, z-x\rangle : \xi \in Ax\} \\ &= g_A(x,y) - tg_A(x,z). \end{aligned}$$

Therefore

$$\limsup_{t\downarrow 0} g(tz+(1-t)x,y) = \limsup_{t\downarrow 0}(g_A(x,y) - tg_A(x,z)) = g_A(x,y).$$

From the definition of g_A (see (5.1.6)), it is easy to check that Condition (C4) holds.

(ii) Let $x \in X$. If $A_{g_A}(x)$ is empty then it is contained in Ax. Otherwise $A_{g_A}(x)$ is nonempty and there exists $\xi \in A_{g_A}(x)$. The monotonicity of g_A which is Condition (C2) (proved in assertion (i)) implies that

$$\langle \xi, y-x\rangle \leqslant g_A(x,y) \leqslant -g_A(y,x) \leqslant -\langle \eta, x-y\rangle = \langle \eta, y-x\rangle$$

for any $\eta \in Ay$. Therefore $\langle \xi-\eta, x-y\rangle \geqslant 0$ for any $\eta \in Ay$. Since A is a maximal monotone mapping, we get from Proposition 1.4.13 that $\xi \in Ax$. Hence $A_{g_A}(x) \subset Ax$ for any $x \in X$. From Proposition 5.1.4(ii) and item (i) we have that A_{g_A} is a maximal monotone mapping. But A is also a maximal monotone mapping and therefore $A_{g_A} = A$.

(iii) From Proposition 5.1.4(i) we have

$$A^{-1}\left(0^*\right) = A_{g_A}^{-1}\left(0^*\right) = \mathrm{EP}(g_A),$$

as asserted.

(iv) Again from Proposition 5.1.4(iii) we have

$$\mathrm{Res}_A^f = \mathrm{Res}_{A_{g_A}}^f = \mathrm{Res}_{g_A}^f.$$ □

5.1.2 Iterative Methods for Solving Equilibrium Problems

Using the properties of resolvents of bifunctions and the connection between their fixed points and the solutions of the corresponding equilibrium problems, we can implement the iterative methods proposed in Chapter 3. There also are connections between solutions of equilibrium problems and zeroes of the corresponding monotone mappings. Therefore we can modify the iterative methods proposed in Chapter 4 in order to solve equilibrium problems. We present two of the possible modifications.

We begin by providing the modification of the Picard iterative method.

Picard Iterative Method for Solving Equilibrium Problems
Initialization: $x_0 \in K$.
General Step ($n = 1, 2, \ldots$):

$$x_{n+1} = \mathrm{Res}_g^f(x_n). \tag{5.1.7}$$

The convergence result for the Picard iterative method for solving equilibrium problems is formulated as follows.

Proposition 5.1.6 (Picard iteration for solving equilibrium problems)**.** *Let $f : X \to (-\infty, +\infty]$ be a super-coercive and Legendre function which is totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$*. Let K be a nonempty, closed and convex subset of* $\operatorname{int}\operatorname{dom} f$ *and assume that a bifunction $g : K \times K \to \mathbb{R}$ satisfies Conditions (C1)–(C4) such that* $\mathrm{EP}(g) \neq \varnothing$*. Then $\left\{\left(\mathrm{Res}_g^f\right)^n x\right\}_{n\in\mathbb{N}}$ converges weakly to an element in* $\mathrm{EP}(g)$ *for each $x \in K$.*

Proof. From Proposition 5.1.3(ii) we have that Res_g^f is BFNE and therefore BSNE (see Figure 1.3). In addition, from Propositions 2.1.2 and 5.1.3(iii) we have that $\widehat{\mathrm{Fix}}\left(\mathrm{Res}_g^f\right) = \mathrm{Fix}\left(\mathrm{Res}_g^f\right) = \mathrm{EP}(g) \neq \varnothing$. Now the result follows immediately from Corollary 3.1.2. □

Now we present a modification of Algorithm (4.4.7) which is based on the concept of products of resolvents.

Sabach Iterative Method for Solving Equilibrium Problems
Input: $f : X \to \mathbb{R}$, $\{\lambda_n^i\}\, n \in \mathbb{N} \subset (0, +\infty)$, $i = 1, 2, \ldots, N$, and $\{e_n\}_{n\in\mathbb{N}} \subset X$
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} y_n = \mathrm{Res}_{\lambda_n^N g_N}^f \circ \cdots \circ \mathrm{Res}_{\lambda_n^1 g_1}^f (x_n + e_n), \\ C_{n+1} = \left\{z \in C_n : D_f\left(z, y_n^i\right) \leqslant D_f\left(z, x_n + e_n\right)\right\}, \\ C_n := \bigcap_{i=1}^N C_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f(x_0) - \nabla f(x_n), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \mathrm{proj}_{C_n \cap Q_n}^f (x_0). \end{cases} \tag{5.1.8}$$

In this case we have the following result.

Proposition 5.1.7 (Convergence results for Algorithm (5.1.8)). *Let K_i, $i = 1, 2, \ldots, N$, be a nonempty, closed and convex subset of X. Let $g_i : K_i \times K_i \to \mathbb{R}$, $i = 1, 2, \ldots, N$, be N bifunctions which satisfy Conditions (C1)–(C4) such that $E := \bigcap_{i=1}^{N} \mathrm{EP}(g_i) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ which is generated by* (5.1.8) *is well defined. If the sequence of errors $\{e_n\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to\infty} \|e_n\| = 0$ and for each $i = 1, 2, \ldots, N$, $\liminf_{n\to\infty} \lambda_n^i > 0$, then the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to* $\mathrm{proj}_E^f(x_0)$ *as $n \to \infty$.*

Proof. The result follows immediately from Theorem 4.4.2 and Proposition 5.1.4. □

5.2 Variational Inequalities

Variational inequalities have turned out to be very useful in studying optimization problems, differential equations, minimax theorems and in certain applications to mechanics and economic theory. Important practical situations motivate the study of systems of variational inequalities (see [66] and the references therein). For instance, the flow of fluid through a fissured porous medium and certain models of plasticity lead to such problems (see, for instance, [104]). The variational inequality problem (VIP), was first introduced (with a single-valued mapping) by Hartman and Stampacchia in 1966 (see [59]).

Because of their importance, variational inequalities have been extensively analyzed in the literature (see, for example, [52, 67, 113] and the references therein). Usually, either the monotonicity or a generalized monotonicity property of the mapping A play a crucial role in these investigations.

The importance of VIPs stems from the fact that several fundamental problems in Optimization Theory can be formulated as VIPs, as the following few examples show.

Example 5.2.1 (Constrained minimization). *Let $K \subseteq X$ be a nonempty, closed and convex subset and let $g : X \to (-\infty, +\infty]$ be a Gâteaux differentiable function which is convex on K. Then x^* is a minimizer of g over K if and only if x^* solves the following VIP:*

$$\left\langle \nabla g\left(x^*\right), x - x^* \right\rangle \geqslant 0 \text{ for all } x \in K.$$

When g is not differentiable, we get the VIP

$$\left\langle u^*, x - x^* \right\rangle \geqslant 0 \text{ for all } x \in K,$$

where $u^* \in \partial g\left(x^*\right)$.

Example 5.2.2 (Nonlinear complementarity problem)**.** *When $X = \mathbb{R}^n$ and $K = \mathbb{R}^n_+$, then the VIP is exactly the nonlinear complementarity problem, that is, find a point $x^* \in \mathbb{R}^n_+$ and a point $u^* \in Ax^*$ such that $u^* \in \mathbb{R}^n_+$ and $\left\langle u^*, x^* \right\rangle = 0$.*

Indeed, if x^* solves (5.0.2) *and $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, then there exists $x^* \in \mathbb{R}^n_+$ such that $u^* \in Ax^*$ which satisfies*

$$\left\langle u^*, x - x^* \right\rangle \geqslant 0 \text{ for all } x \in \mathbb{R}^n_+.$$

So, in particular, if we take $x = 0$ we obtain $\left\langle u^*, x^* \right\rangle \leqslant 0$ and if we take $x = 2x^*$ we obtain $\left\langle u^*, x^* \right\rangle \geqslant 0$. Combining the above two inequalities, we see that $\left\langle u^*, x^* \right\rangle = 0$. As a consequence, this yields

$$\left\langle u^*, x \right\rangle \geqslant 0 \text{ for all } x \in \mathbb{R}^n_+$$

and hence $u^* \in \mathbb{R}^n_+$. Conversely, if x^* solves the nonlinear complementarity problem, then $\left\langle u^*, x - x^* \right\rangle = \left\langle u^*, x \right\rangle \geqslant 0$ for all $x \in \mathbb{R}^n_+$ (since $u^* \in \mathbb{R}^n_+$), which means that x^* solves (5.0.2) *with $N = 1$.*

Example 5.2.3 (Finding zeroes)**.** *When the set K is the whole space X, then the VIP obtained from* (5.0.2) *is equivalent to the problem of finding zeroes of a mapping $A : X \to 2^{X^*}$, i.e., to find an element $x^* \in X$ such that $0 \in A\left(x^*\right)$.*

Example 5.2.4 (Saddle-point problem)**.** *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces, and let K_1 and K_2 be two convex subsets of X_1 and X_2, respectively. Given a bifunction $g : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R}$, the saddle-point problem is to find a point $\left(u_1^*, u_2^*\right) \in K_1 \times K_2$ such that*

$$g\left(u_1^*, u_2\right) \leqslant g\left(u_1^*, u_2^*\right) \leqslant g\left(u_1, u_2^*\right) \text{ for all } (u_1, u_2) \in K_1 \times K_2.$$

This problem can be written as the VIP of finding $\left(u_1^*, u_2^*\right) \in K_1 \times K_2$ such that

$$\left\langle \begin{pmatrix} \nabla g_{u_1}\left(u_1^*, u_2^*\right) \\ -\nabla g_{u_2}\left(u_1^*, u_2^*\right) \end{pmatrix}, \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - \begin{pmatrix} u_1^* \\ u_2^* \end{pmatrix} \right\rangle \geqslant 0 \text{ for all } (u_1, u_2) \in K_1 \times K_2. \tag{5.2.1}$$

As in the case of finding zeroes of monotone mappings (see Chapter 4), the key tool for solving variational inequalities is to define a resolvent with respect to a mapping A as in the case of the f-resolvents (see (4.0.2)). In the case of variational inequalities we discuss two kinds of resolvents: the anti-resolvent and the generalized resolvent.

Definition 5.2.5 (Anti-resolvent)**.** *The anti-resolvent* $A^f : X \to 2^X$ *of a mapping* $A : X \to 2^{X^*}$ *is defined by*

$$A^f := \nabla f^* \circ (\nabla f - A) . \tag{5.2.2}$$

Definition 5.2.6 (Generalized resolvent)**.** *The generalized resolvent* $\mathrm{GRes}_A^f : X \to 2^X$ *of a mapping* $A : X \to 2^{X^*}$ *is defined by*

$$\mathrm{GRes}_A^f (x) := \{z \in K : \langle Az, y - z\rangle + \langle \nabla f (z) - \nabla f (x) , y - z\rangle \geqslant 0 \ \forall y \in K\} . \tag{5.2.3}$$

5.2.1 Properties of Anti-Resolvents

We begin by providing several basic properties of anti-resolvents (see Definition 5.2.5) which were proved in [38, Lemma 3.5, page 2109] (see also [66, Proposition 11, page 1326]).

Proposition 5.2.7 (Properties of anti-resolvents)**.** *Let* $f : X \to (-\infty, +\infty]$ *be a Legendre function which satisfies the range condition* (1.4.6). *Let* $A : X \to 2^{X^*}$ *be a mapping. The following statements are true.*

(i) $\mathrm{dom}\, A^f \subseteq \mathrm{dom}\, A \bigcap \mathrm{int}\,\mathrm{dom}\, f$.

(ii) $\mathrm{ran}\, A^f \subseteq \mathrm{int}\,\mathrm{dom}\, f$.

(iii) *The mapping A is BISM on its domain if and only if its anti-resolvent A^f is BFNE operator.*

(iv) $A^{-1}\left(0^*\right) = \mathrm{Fix}\left(A^f\right)$.

Proof. (i) Clear from Definition 5.2.5.

(ii) Clear from Definition 5.2.5.

(iii) Let $x, y \in \mathrm{dom}\, A^f$ and take $\xi \in Ax$, $\eta \in Ay$, $u \in A^f x$ and $v \in A^f y$. From the definition of BFNE operators (see (1.3.4)) the anti-resolvent A^f (see Definition 5.2.5) is BFNE if and only if

$$\langle \nabla f (u) - \nabla f (v) , u - v\rangle \leqslant \langle \nabla f (x) - \nabla f (y) , u - v\rangle , \tag{5.2.4}$$

which is equivalent to

$$\langle (\nabla f (x) - \xi) - (\nabla f (y) - \eta) , u - v\rangle \leqslant \langle \nabla f (x) - \nabla f (y) , u - v\rangle ,$$

that is, $\langle \xi - \eta, u - v \rangle \geqslant 0$. Since $u \in A^f x$ and $\xi \in Ax$, we get that $u = \nabla f^* (\nabla f(x) - \xi)$. The same holds for v, that is, $v = \nabla f^* (\nabla f(y) - \eta)$. Therefore (5.2.4) is equivalent to

$$\left\langle \xi - \eta, \nabla f^* (\nabla f(x) - \xi) - \nabla f^* (\nabla f(y) - \eta) \right\rangle \geqslant 0,$$

which means that A is a BISM mapping (see Definition 1.4.29).

(iv) From the definition of the anti-resolvent (see (5.2.5)) we get that

$$0^* \in Ax \Leftrightarrow \nabla f(x) \in \nabla f(x) - Ax = (\nabla f - A)(x) \Leftrightarrow x \in \nabla f^* \circ (\nabla f - A)(x) = A^f x.$$

□

Let K be a nonempty, closed and convex subset of X and let $A : X \to X^*$ be a mapping. The variational inequality corresponding to such a mapping A is to find $\bar{x} \in K$ such that

$$\langle A\bar{x}, y - \bar{x} \rangle \geqslant 0 \ \forall y \in K. \tag{5.2.5}$$

The solution set of (5.2.5) is denoted by $\mathrm{VI}(K, A)$.

In the following result we bring out the connections between the fixed point set of $\mathrm{proj}_K^f \circ A^f$ and the solution set of the variational inequality corresponding to a single-valued mapping $A : X \to X^*$ (*cf.* [66, Proposition 12, page 1327]).

Proposition 5.2.8 (Characterization of $\mathrm{VI}(K, A)$ as a fixed point set). *Let $A : X \to X^*$ be a mapping. Let $f : X \to (-\infty, +\infty]$ be a Legendre and totally convex function which satisfies the range condition* (1.4.6). *If K is a nonempty, closed and convex subset of X, then* $\mathrm{VI}(K, A) = \mathrm{Fix}\left(\mathrm{proj}_K^f \circ A^f\right)$.

Proof. From Proposition 1.2.35(ii) we obtain that $x = \mathrm{proj}_K^f \left(A^f x\right)$ if and only if

$$\left\langle \nabla f\left(A^f x\right) - \nabla f(x), x - y \right\rangle \geqslant 0$$

for all $y \in K$. This is equivalent to $\langle (\nabla f - A) x - \nabla f(x), x - y \rangle \geqslant 0$ for any $y \in K$, that is, $\langle -Ax, x - y \rangle \geqslant 0$ for each $y \in K$, which is obviously equivalent to $x \in \mathrm{VI}(K, A)$, as claimed. □

It is obvious that any zero of a mapping A which belongs to K is a solution of the variational inequality corresponding to A on the set K, that is, $A^{-1}(0^*) \cap K \subset \mathrm{VI}(K, A)$. In the following result we show that the converse implication holds for single-valued BISM mappings (*cf.* [66, Proposition 13, page 1327]).

Proposition 5.2.9 ($A^{-1}(0^*) \cap K = \mathrm{VI}(K, A)$ for BISM mappings). *Let $f : X \to (-\infty, +\infty]$ be a Legendre and totally convex function which satisfies the range condition*

(1.4.6). *Let K be a nonempty, closed and convex subset of* $\mathrm{dom}\, A \bigcap \mathrm{int}\,\mathrm{dom}\, f$. *If a BISM mapping $A : X \to X^*$ satisfies $Z := A^{-1}(0^*) \cap K \neq \varnothing$, then* $\mathrm{VI}(K, A) = Z$.

Proof. Let $x \in \mathrm{VI}(K, A)$. From Proposition 5.2.8 we know that $x = \mathrm{proj}_K^f\left(A^f x\right)$. From Proposition 1.2.35(iii) we now obtain that

$$D_f\left(u, \mathrm{proj}_K^f\left(A^f x\right)\right) + D_f\left(\mathrm{proj}_K^f\left(A^f x\right), A^f x\right) \leqslant D_f\left(u, A^f x\right)$$

for any $u \in K$. From Proposition 5.2.7(iii) we have that A^f is BFNE and therefore QBNE (see Figure 1.3). Hence

$$\begin{aligned} D_f(u, x) + D_f\left(x, A^f x\right) &= D_f\left(u, \mathrm{proj}_K^f\left(A^f x\right)\right) + D_f\left(\mathrm{proj}_K^f\left(A^f x\right), A^f x\right) \\ &\leqslant D_f\left(u, A^f x\right) \leqslant D_f(u, x) \end{aligned}$$

for any $u \in Z$. This implies that $D_f\left(x, A^f x\right) = 0$. It now follows from Proposition 1.2.4 that $x = A^f x$, that is, $x \in \mathrm{Fix}\left(A^f\right)$, and from Proposition 5.2.7(iv) we get that $x \in A^{-1}\left(0^*\right)$. Since $x = \mathrm{proj}_K^f\left(A^f x\right)$, it is clear that $x \in K$ and therefore $x \in Z$. Conversely, let $x \in Z$. Then $x \in K$ and $Ax = 0^*$, so it is obvious that (5.2.5) is satisfied. In other words, $x \in \mathrm{VI}(K, A)$. □

The following example shows that the assumption $Z \neq \varnothing$ in Proposition 5.2.9 is essential (*cf.* [66, Example 2, page 1328]).

Example 5.2.10 (Assumption $Z \neq \varnothing$ is essential)**.** *Let $X = \mathbb{R}$, $f = (1/2)\|\cdot\|^2$, $K = [1, +\infty)$ and let $A : \mathbb{R} \to \mathbb{R}$ be given by $Ax = x$ (the identity mapping). This is obviously a BISM mapping (which in our case means that it is firmly nonexpansive (see Remark 1.4.30), and all the assumptions of Proposition 5.2.9 hold, except $Z \neq \varnothing$. Indeed, we have $A^{-1}(0) = \{0\}$ and $0 \notin K$. However, $V = \{1\}$ since the only solution of the variational inequality $x(y - x) \geqslant 0$ for all $y \geqslant 1$ is $x = 1$ and therefore $Z = \varnothing$ is a proper subset of V.*

To sum up, the anti-resolvent A^f of a mapping A seems to be a more "complicated" operator than the other resolvents we mentioned since its nonexpansivity property holds only if the mapping A is assumed to be BISM. On the other hand, as we proved in Proposition 5.2.9, for BISM mappings, finding zeroes in K is exactly equivalent to solving a variational inequality over K. Therefore, solving variational inequalities for BISM mappings using anti-resolvents leads to a particular case of finding zeroes. We refer the interested reader to the paper [66] for a careful study of iterative methods for solving variational inequalities for BISM mappings.

Because of these drawbacks of the anti-resolvents we will study more carefully the generalized resolvent (see Definition 5.2.6).

5.2.2 Properties of Generalized Resolvents

We begin this section by proving that the bifunction $g(x,y)$ defined by $\langle Ax, y-x\rangle$ satisfies the basic conditions mentioned in Assumption 1 (*cf.* [66, Proposition 16, page 1338]).

Proposition 5.2.11 (Monotone mappings and bifunctions)**.** *Let $A : X \to X^*$ be a monotone mapping such that $K := \operatorname{dom} A$ is closed and convex. Assume that A is bounded on bounded subsets and semicontinuous on K. Then the bifunction $g(x,y) = \langle Ax, y-x\rangle$ satisfies Conditions (C1)–(C4).*

Proof. It is clear that $g(x,x) = \langle Ax, x-x\rangle = 0$ for any $x \in K$. From the monotonicity of the mapping A (see (1.4.1)) we obtain that

$$g(x,y) + g(y,x) = \langle Ax, y-x\rangle + \langle Ay, x-y\rangle = \langle Ax - Ay, y-x\rangle \leqslant 0$$

for any $x, y \in K$. To prove Condition (C3), fix $y \in X$ and choose a sequence $\{t_n\}_{n\in\mathbb{N}}$, converging to zero, such that

$$\limsup_{t\downarrow 0} g(tz + (1-t)x, y) = \lim_{n\to\infty} g(t_n z + (1-t_n)x, y).$$

Such a sequence exists by the definition of the limsup. Denote $u_n = t_n z + (1-t_n)x$. Then $\lim_{n\to\infty} u_n = x$ and $\{Au_n\}_{n\in\mathbb{N}}$ is bounded. Let $\{Au_{n_k}\}_{k\in\mathbb{N}}$ be a weakly convergent subsequence. Then its limit is Ax because A is hemicontinuous (see Definition 1.4.8) and we get

$$\begin{aligned}
\limsup_{t\downarrow 0} g(tz + (1-t)x, y) &= \lim_{k\to\infty} g(t_{n_k} z + (1-t_{n_k})x, y) = \\
&= \lim_{k\to\infty} \langle A(t_{n_k} z + (1-t_{n_k})x), y - t_{n_k} z - (1-t_{n_k})x\rangle \\
&= \lim_{k\to\infty} \langle A(u_{n_k}), y - u_{n_k}\rangle = \langle Ax, y-x\rangle = g(x,y)
\end{aligned}$$

for all $x, y, z \in K$, as required. Condition (C4) also holds because

$$\begin{aligned}
g(x, ty_1 + (1-t)y_2) &= \langle Ax, x - (ty_1 + (1-t)y_2)\rangle = t\langle Ax, x-y_1\rangle + (1-t)\langle Ax, x-y_2\rangle \\
&= tg(x,y_1) + (1-t)g(x,y_2);
\end{aligned}$$

thus the function $g(x,\cdot)$ is clearly convex and lower semicontinuous as it is (in particular) affine and continuous for any $x \in K$.

Therefore g indeed satisfies Conditions (C1)–(C4). □

Now we summarize several properties of generalized resolvents (see Definition 5.2.6).

Proposition 5.2.12 (Properties of generalized resolvents)**.** *Let $f : X \to (-\infty, +\infty]$ be a super-coercive and Legendre function. Let $A : X \to X^*$ be a monotone mapping such that $K := \operatorname{dom} A$ is a closed and convex subset of X. Assume that A is bounded on bounded subsets and hemicontinuous on K. Then the generalized resolvent of A has the following properties.*

(i) $\operatorname{dom} \mathrm{GRes}_A^f = X$.

(ii) GRes_A^f *is single-valued.*

(iii) GRes_A^f *is an BFNE operator.*

(iv) $\operatorname{Fix}\left(\mathrm{GRes}_A^f\right) = \mathrm{VI}\,(K, A)$.

(v) $\mathrm{VI}\,(K, A)$ *is a closed and convex subset of K.*

Proof. The result follows by combining Propositions 5.1.2, 5.1.3 and 5.2.11. □

A connection between f-resolvents, Res_A^f, and generalized resolvents, GRes_A^f, is brought out by the following remark.

Remark 5.2.13 (Connection between f-resolvents and generalized resolvents)**.** *If the domain of a mapping $A : X \to X^*$ is the whole space, then $\mathrm{VI}\,(X, A)$ is exactly the zero set of A. Therefore we obtain, for any $z \in \mathrm{GRes}_A^f(x)$, that $\langle Az, y - z\rangle + \langle \nabla f(z) - \nabla f(x), y - z\rangle \geqslant 0$ for any $y \in X$. This is equivalent to $\langle Az + \nabla f(z) - \nabla f(x), y - z\rangle \geqslant 0$ for any $y \in X$, and this, in turn, is the same as $\langle Az + \nabla f(z) - \nabla f(x), w\rangle \geqslant 0$ for any $w \in X$. But then we obtain that $\langle Az + \nabla f(z) - \nabla f(x), w\rangle = 0$ for any $w \in X$. This happens only if $Az + \nabla f(z) - \nabla f(x) = 0^*$, which means that $z = (\nabla f + A)^{-1} \circ \nabla f(x)$. This proves that the generalized resolvent GRes_A^f is a generalization of the resolvent Res_A^f.* ◇

5.2.3 Iterative Methods for Solving Variational Inequalities

Using the properties of generalized resolvents and the connection between their fixed points and the solutions of variational inequalities, we can implement the iterative methods proposed in Chapter 3.

We begin with a modification of the Mann iterative method (see Algorithm (3.0.3)), which is defined by using convex combinations with respect to a convex function f, for solving variational inequalities.

f-Mann Iterative Method for Solving Variational Inequalities
Input: $f : X \to \mathbb{R}$ and $\{\alpha_n\}_{n\in\mathbb{N}} \subset (0,1)$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$)**:**

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f(x_n) + (1-\alpha_n)\nabla f\left(\mathrm{GRes}_A^f(x_n)\right)\right). \tag{5.2.6}$$

In the following result we prove weak convergence of the sequence generated by Algorithm (5.2.6).

Proposition 5.2.14 (Convergence result for Algorithm (5.2.6))**.** *Let $A : X \to X^*$ be a monotone mapping such that $K := \mathrm{dom}\, A$ is a closed and convex subset of X. Assume that A is bounded on bounded subsets and hemicontinuous on K such that $\mathrm{VI}(K,A) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a super-coercive and Legendre function which is totally convex on bounded subsets of X. Suppose that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of* $\mathrm{int}\,\mathrm{dom}\, f^*$*. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence generated by Algorithm* (5.2.6) *where $\{\alpha_n\}_{n\in\mathbb{N}} \subset [0,1]$ satisfies $\limsup_{n\to\infty} \alpha_n < 1$. Then, for each $x_0 \in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in* $\mathrm{VI}(K,A)$.

Proof. From Proposition 5.2.12(iii) we have that GRes_A^f is BFNE and therefore BSNE (see Figure 1.3). In addition, from Propositions 2.1.2 and 5.2.12(iv) we have that $\widehat{\mathrm{Fix}}\left(\mathrm{GRes}_A^f\right) = \mathrm{Fix}\left(\mathrm{Res}_A^f\right) = \mathrm{VI}(K,A) \neq \varnothing$. Now the result follows immediately from Corollary 3.2.3. □

Now we present another algorithm for finding solutions of a system of a finite number of variational inequalities.

Minimal Norm-Like Iterative Method for Solving Variational Inequalities I
Input: $f : X \to \mathbb{R}$ and $\{e_n^i\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$):

$$\begin{cases} y_n^i = \mathrm{GRes}_{A_i}^f \left(x_n + e_n^i\right), \\ C_n^i = \left\{z \in X : D_f\left(z, y_n^i\right) \leqslant D_f\left(z, x_n + e_n^i\right)\right\}, \\ C_n := \bigcap_{i=1}^N C_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n \right\rangle \leqslant 0\right\}, \\ x_{n+1} = \mathrm{proj}_{C_n \cap Q_n}\left(x_0\right). \end{cases} \tag{5.2.7}$$

In this case Algorithm (5.2.7) generates a sequence which converges strongly to a solution of the system.

Proposition 5.2.15 (Convergence of Algorithm (5.2.7))**.** *Let $A_i : X \to X^*$, $i = 1, 2, \ldots, N$, be a monotone mapping such that $K_i := \operatorname{dom} A_i$ is a closed and convex subset of X. Assume that each A_i, $i = 1, 2, \ldots, N$, is bounded on bounded subsets and hemicontinuous on K_i such that $V := \bigcap_{i=1}^N \mathrm{VI}\left(K_i, A_i\right) \neq \varnothing$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that ∇f^* is bounded on bounded subsets of* $\operatorname{int}\operatorname{dom} f^*$. *Then, for each $x_0 \in X$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (5.2.7)*. If, for each $i = 1, 2, \ldots, N$, the sequence of errors $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to\infty} \|e_n^i\| = 0$, then each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly to $\mathrm{proj}_V^f\left(x_0\right)$ as $n \to \infty$.*

Proof. From Proposition 5.2.12(iii) we have that each $\mathrm{GRes}_{A_i}^f$, $i = 1, 2, \ldots, N$, is BFNE and therefore QBNE (see Figure 1.3). In addition, from Propositions 2.1.2 and 5.2.12(iv) we have that $\bigcap_{i=1}^N \mathrm{Fix}\left(\mathrm{Res}_{A_i}^f\right) = \bigcap_{i=1}^N \mathrm{VI}\left(K_i, A_i\right) \neq \varnothing$. Now the result follows immediately from Theorem 3.3.9. □

Now we present another approach for solving systems of variational inequalities corresponding to hemicontinuous mappings (see Definition 1.4.8). We use the following notation.

Definition 5.2.16 (Normal cone)**.** *Consider the normal cone N_K corresponding to $K \subset X$, which is defined by*

$$N_K\left(x\right) := \left\{\xi \in X^* : \left\langle \xi, x - y \right\rangle \geqslant 0, \ \forall y \in K\right\}, \ x \in K.$$

Now we have the following connection between the problems of solving variational inequalities and finding zeroes of maximal monotone mappings (*cf.* [99, Theorem 3, page 77]).

Proposition 5.2.17 (A maximal monotone mapping for solving VIP)**.** *Let K be a nonempty, closed and convex subset of X, and let $A : K \to X^*$ be a monotone and hemicontinuous mapping. Let $B : X \to 2^{X^*}$ be the mapping which is defined by*

$$Bx := \begin{cases} (A + N_K)\, x, & x \in K \\ \varnothing, & x \notin K. \end{cases} \tag{5.2.8}$$

Then B is maximal monotone and $B^{-1}\left(0^*\right) = \mathrm{VI}\,(K, A)$.

For each $i = 1, 2, \ldots, N$, let the mapping B_i, defined as in (5.2.8), correspond to the mapping A_i and the set K_i, and let $\{\lambda_n^i\}_{n\in\mathbb{N}}$, $i = 1, 2, \ldots, N$, be N sequences of positive real numbers. Using Proposition 5.2.17 we can modify the iterative methods proposed in Chapter 4 in order to solve variational inequalities. We present one of these modifications.

Minimal Norm-Like Iterative Method for Solving Variational Inequalities II
Input: $f : X \to \mathbb{R}$ and $\{e^i{}_n\}_{n\in\mathbb{N}} \subset X$, $i = 1, 2, \ldots, N$.
Initialization: $x_0 \in X$.
General Step ($n = 1, 2, \ldots$)**:**

$$\begin{cases} y_n^i = \mathrm{GRes}^f_{\lambda_n^i B_i}\left(x_n + e_n^i\right), \\ H_n = \left\{z \in X : \left\langle \nabla f\left(x_n + e_n^i\right) - \nabla f\left(y_n^i\right), z - y_n^i\right\rangle \leqslant 0\right\}, \\ H_n := \bigcap_{i=1}^N H_n^i, \\ Q_n = \left\{z \in X : \left\langle \nabla f\left(x_0\right) - \nabla f\left(x_n\right), z - x_n\right\rangle \leqslant 0\right\}, \\ x_{n+1} = \mathrm{proj}^f_{H_n \cap Q_n}\left(x_0\right). \end{cases} \tag{5.2.9}$$

Theorem 4.3.7 yields a method for solving systems of variational inequalities corresponding to hemicontinuous mappings.

Proposition 5.2.18 (Convergence result for Algorithm (5.2.9))**.** *Let K_i, $i = 1, 2, \ldots, N$, be N nonempty, closed and convex subsets of X such that $K := \bigcap_{i=1}^N K_i$. Let $A_i : K_i \to X^*$, $i = 1, 2, \ldots, N$, be N monotone and hemicontinuous mappings with $V := \bigcap_{i=1}^N VI\,(K_i, A_i) \neq \varnothing$. Let $\{\lambda_n^i\}_{n\in\mathbb{N}}$, $i = 1, 2, \ldots, N$, be N sequences of positive real numbers that satisfy $\liminf_{n\to\infty} \lambda_n^i > 0$. Let $f : X \to \mathbb{R}$ be a Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of X. Suppose that*

∇f^* is bounded on bounded subsets of $\operatorname{int}\operatorname{dom} f^*$. *If, for each $i = 1, 2, \ldots, N$, the sequence of errors $\{e_n^i\}_{n\in\mathbb{N}} \subset X$ satisfies $\lim_{n\to\infty} \|e_n^i\| = 0$, then for each $x_0 \in K$, there are sequences $\{x_n\}_{n\in\mathbb{N}}$ which satisfy Algorithm* (5.2.9)*, where each B_i is defined as in* (5.2.8). *Each such sequence $\{x_n\}_{n\in\mathbb{N}}$ converges strongly as $n \to \infty$ to* $\operatorname{proj}_V^f(x_0)$.

Proof. For each $i = 1, 2, \ldots, N$, we define the mapping B_i as in (5.2.8). Proposition 5.2.17 now implies that each B_i, $i = 1, 2, \ldots, N$, is a maximal monotone mapping and $V = \bigcap_{i=1}^N VI(K_i, A_i) = \bigcap_{i=1}^N B_i^{-1}(0^*) \neq \varnothing$.

Our result now follows immediately from Theorem 4.3.7 with $Z = V$. □

5.3 Convex Feasibility Problems

Let K_i, $i = 1, 2, \ldots, N$, be N nonempty, closed and convex subsets of X. The convex feasibility problem (CFP) is to find an element in the assumed nonempty intersection $\bigcap_{i=1}^N K_i$ (see [5]). It is clear that $\operatorname{Fix}\left(\operatorname{proj}_{K_i}^f\right) = K_i$ for any $i = 1, 2, \ldots, N$. If the Legendre function f is uniformly Fréchet differentiable and bounded on bounded subsets of X, then if follows from Proposition 4.1.2(iv)(b) that the Bregman projection $\operatorname{proj}_{K_i}^f$ is BFNE. In addition, from Propositions 2.1.2 and 4.1.2(iii) we have that $\widehat{\operatorname{Fix}}\left(\operatorname{proj}_{K_i}^f\right) = \operatorname{Fix}\left(\operatorname{proj}_{K_i}^f\right) = K_i$. Hence we can implement the iterative methods proposed in Chapter 3. We present the following modification of the Picard iterative method (see Algorithm (3.0.1)) for solving convex feasibility problems in reflexive Banach spaces. Define the block operator (see (2.1.14)) in the following way

$$T_B := \nabla f^* \left(\sum_{i=1}^N w_i \nabla f \left(\operatorname{proj}_{K_i}^f\right)\right).$$

It follows from Propositions 2.1.17 and 2.1.18 that T_B is a BSNE operator such that $\operatorname{Fix}(T_B) = \bigcap_{i=1}^N \operatorname{Fix}\left(\operatorname{proj}_{K_i}^f\right) = \bigcap_{i=1}^N K_i$. Now, if $f : X \to (-\infty, +\infty]$ is a Legendre function which is totally convex on bounded subsets of X such that ∇f is weakly sequentially continuous and ∇f^* is bounded on bounded subsets of $\operatorname{int}\operatorname{dom} f^*$, then it follows that the Picard iterative method of T_B generates a sequence which converges weakly to an element in $\bigcap_{i=1}^N K_i$, that is, a solution of the convex feasibility problem.

5.3.1 A Numerical Example

In this subsection we present a simple low-dimensional example (*cf.* [44, Subsection 4.6]).

We consider a two-disk convex feasibility problem in $\mathbb{R}^2$ and provide an explicit formulation of Algorithm (3.3.5) as well as some numerical results. More explicitly, let

$$K_1 = \left\{(x, y) \in \mathbb{R}^2 : (x - a_1)^2 + (y - b_1)^2 \leqslant r_1^2\right\}$$

and

$$K_2 = \left\{(x, y) \in \mathbb{R}^2 : (x - a_2)^2 + (y - b_2)^2 \leqslant r_2^2\right\}$$

with $K_1 \cap K_2 \neq \varnothing$.

Consider the problem of finding a point $\left(x^*, y^*\right) \in \mathbb{R}^2$ such that $\left(x^*, y^*\right) \in K_1 \cap K_2$. Observe that in this case $T_1 = P_{K_1}$ and $T_2 = P_{K_2}$. For simplicity we take $f = (1/2)\left\|\cdot\right\|^2$. Given the current iterate $x_n = (u, v)$, the explicit formulation of the iterative step of our algorithm becomes (see Remark 1.2.40(i)):

$$\begin{cases} y_n^1 = P_{K_1}\left(x_n\right) = \left(a_1 + \frac{r_1(u-a_1)}{\|(u-a_1,v-b_1)\|}, b_1 + \frac{r_1(v-b_1)}{\|(u-a_1,v-b_1)\|}\right), \\ y_n^2 = P_{K_1}\left(x_n\right) = \left(a_2 + \frac{r_2(u-a_2)}{\|(u-a_2,v-b_2)\|}, b_2 + \frac{r_2(v-b_2)}{\|(u-a_2,v-b_2)\|}\right), \\ C_n^1 = \left\{z \in \mathbb{R}^2 : \left\|z - y_n^1\right\| \leqslant \left\|z - x_n\right\|\right\}, \\ C_n^2 = \left\{z \in \mathbb{R}^2 : \left\|z - y_n^2\right\| \leqslant \left\|z - x_n\right\|\right\}, \\ Q_n = \left\{z \in \mathbb{R}^2 : \langle x_0 - x_n, z - x_n\rangle \leqslant 0\right\}, \\ x_{n+1} = P_{C_n^1 \cap C_n^2 \cap Q_n}\left(x_0\right). \end{cases} \tag{5.3.1}$$

In order to evaluate x^{n+1}, we solve the following constrained minimization problem:

$$\begin{cases} \min & \left\|x^0 - z\right\|^2 \\ \text{s.t.} & z \in C_1^n \cap C_2^n \cap Q^n. \end{cases} \tag{5.3.2}$$

Following the same technique as in Example 1.2.41, it is possible to obtain a solution to the problem (5.3.2) even for more than three half-spaces, but there are many subcases in the explicit formula (two to the power of the number of half-spaces).

Now we present some numerical results for the particular case where

$$K_1 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant 1\right\}$$

and

$$K_2 = \left\{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + y^2 \leqslant 1\right\}.$$

We choose two starting points $(-1/2, 3)$ and $(3, 3)$, and for each starting point we present a table with the (x, y) coordinates for the first 10 iterations of Algorithm (5.3.1). In addition, Figures 5.1 and 5.3 illustrate the geometry in each iterative step, *i.e.*, the disks and the three half-spaces C_1^n, C_2^n and Q^n.

Iteration Number	x-value	y-value
1	−0.500000000	3.0000000000
2	0.0263507717	1.9471923798
3	0.2898391508	1.4209450920
4	0.4211545167	1.1576070220
5	0.4687763141	1.0169184232
6	0.4862238741	0.9429308114
7	0.4935428246	0.9048859275
8	0.4968764116	0.8855650270
9	0.4984644573	0.8758239778
10	0.4992386397	0.8709324060

Table 5.1: The first 10 iterations of Algorithm (5.3.1) with $x_0 = (-1/2, 3)$

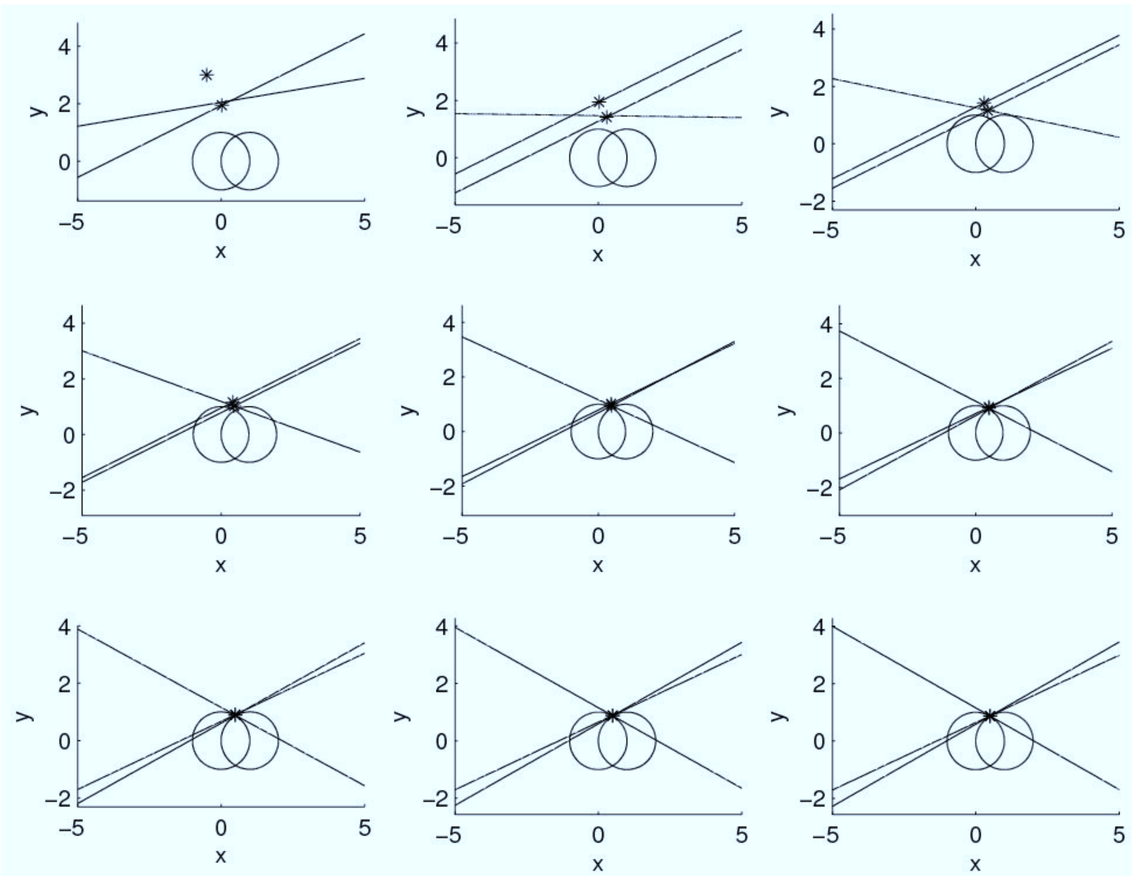

Figure 5.1: Geometric illustration of Algorithm (5.3.1) with $x_0 = (-1/2, 3)$

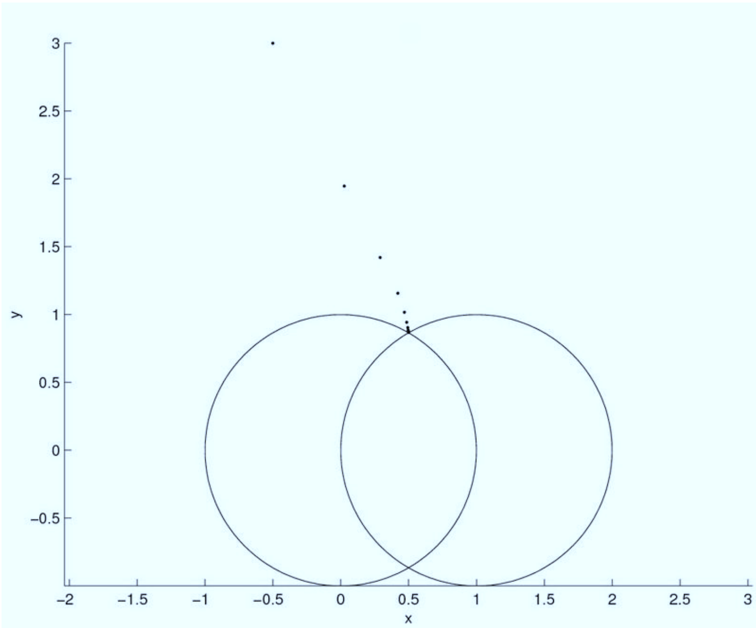

Figure 5.2: Plot of the first 10 iterations of Algorithm (5.3.1) with $x_0 = (-1/2, 3)$

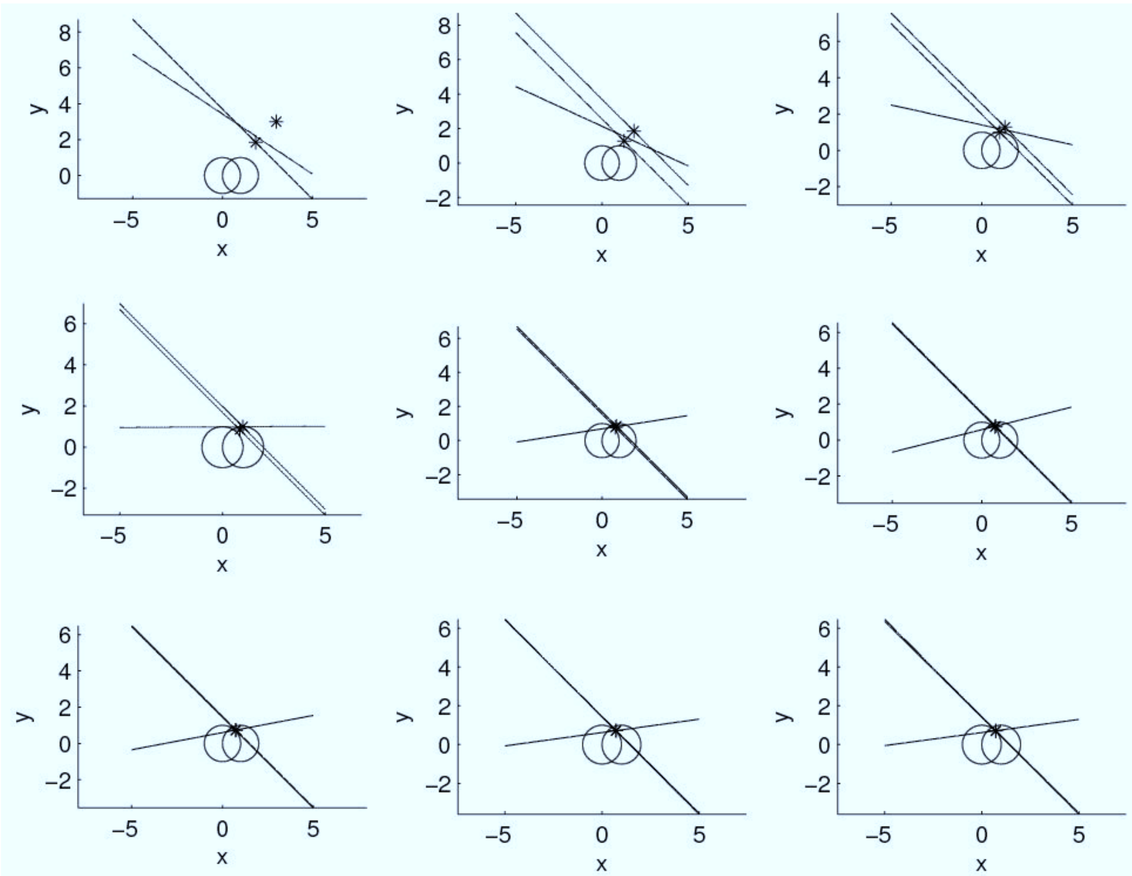

Figure 5.3: Geometric illustration of Algorithm (5.3.1) with $x_0 = (3, 3)$

Iteration Number	x-value	y-value
1	3.0000000000	3.0000000000
2	1.8536075595	1.8534992168
3	1.2802790276	1.2803811470
4	0.9937807510	0.9936561265
5	0.8503033752	0.8505218683
6	0.7789970157	0.7785224690
7	0.7423971596	0.7434698006
8	0.7264747366	0.7235683325
9	0.7115677773	0.7205826742
10	0.7260458319	0.6973591138

Table 5.2: The first 10 iterations of Algorithm (5.3.1) with $x_0 = (3, 3)$

Chapter 6

Minimal Norm-Like Solutions of Convex Optimizations Problems

Motivated by the algorithms proposed in Chapters 3, 4 and 5 for solving diverse problems such as fixed point problems, finding zeroes of monotone mappings, equilibrium, variational inequalities and convex feasibility problems in the setting of infinite-dimensional Banach spaces, we present on this paper a full analysis of a modification of Algorithm (3.3.2) in the setting of Euclidean spaces for solving the well-known problem of finding minimal norm solutions of convex optimization problems. This problem has very practical aspects and therefore we prove a rate of convergence result and show implementation to real-world problems. This chapter is based on a joint work with Professor Amir Beck.

More precisely, in this chapter we consider a general class of convex optimization problems in which one seeks to minimize a strongly convex function over a closed and convex set which is by itself an optimal set of another convex problem. We introduce a gradient-based method, called the minimal norm gradient method, for solving this class of problems, and establish the convergence of the sequence generated by the algorithm as well as a rate of convergence of the sequence of function values. A portfolio optimization example is given in order to illustrate our results.

6.1 Problem Formulation

Consider a general convex constrained optimization problem given by

$$\text{(P):} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in S, \end{array}$$

where the following assumptions are made throughout the chapter.

(i) S is a nonempty, closed and convex subset of $\mathbb{R}^n$.

(ii) The objective function f is convex and continuously differentiable over $\mathbb{R}^n$, and its gradient is Lipschitz with constant L:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant L \|\mathbf{x} - \mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{6.1.1}$$

(iii) The optimal set of (P), denoted by S^*, is nonempty. The optimal value is denoted by f^*.

Problem (P) might have multiple optimal solutions, and in this case it is natural to consider the *minimal norm solution problem* in which one seeks to find the optimal solution of (P) with a minimal Euclidean norm[1]:

$$\text{(Q):} \quad \min \left\{ \frac{1}{2} \|\mathbf{x}\|^2 : \mathbf{x} \in S^* \right\}.$$

We will denote the optimal solution of (Q) by $\mathbf{x}_{\mathrm{Q}}^*$. A well-known approach to tackling problem (Q) is via the celebrated Tikhonov regularization. More precisely, for a given $\varepsilon > 0$, consider the convex problem defined by

$$(\mathrm{Q}_\varepsilon): \quad \min \left\{ f(\mathbf{x}) + \frac{\varepsilon}{2} \|\mathbf{x}\|^2 : \mathbf{x} \in S \right\}.$$

The above problem is the so-called *Tikhonov regularized problem* [109]. Let us denote the unique optimal solution of (Q_ε) by $\mathbf{x}^\varepsilon$. In [109], Tikhonov showed in the linear case – that is, when f is a linear function and S is an intersection of half-spaces – that $\mathbf{x}^\varepsilon \to \mathbf{x}_{\mathrm{Q}}^*$ as $\varepsilon \to 0^+$. Therefore, for a small enough $\varepsilon > 0$, the vector $\mathbf{x}^\varepsilon$ can be considered as an approximation of the minimal norm solution $\mathbf{x}_{\mathrm{Q}}^*$. A stronger result in the linear case showing that for a small enough ε, $\mathbf{x}^\varepsilon$ is in fact *exactly the same* as $\mathbf{x}_{\mathrm{Q}}^*$ was established in [71] and was later on generalized to the more general convex case in [53].

[1] We use here the obvious property that the problems of minimizing the norm and of minimizing half of the squared norm are equivalent in the sense that they have the same unique optimal solution.

From a practical point of view, the connection just alluded to between the minimal norm solution and the solutions of the Tikhonov regularized problems, does not yield an explicit algorithm for solving (Q). It is not clear how to choose an appropriate sequence of regularization parameters $\varepsilon_k \to 0^+$, and how to solve the emerging subproblems. A different approach for solving (Q) in the linear case was developed in [65] where it was suggested to invoke a Newton-type method for solving a reformulation of (Q) as an unconstrained smooth minimization problem.

The main contribution in this work is the construction and analysis of a new first-order method for solving a generalization of problem (Q), which we call *the minimum norm-like solution problem* (MNP). Problem (MNP) consists of finding the optimal solution of problem (P) which minimizes a given strongly convex function ω. More precisely,

$$\text{(MNP):} \quad \min \{\omega(\mathbf{x}) : \mathbf{x} \in S^*\}.$$

The function ω is assumed to satisfy the following conditions.

(i) ω is a strongly convex function over $\mathbb{R}^n$ with parameter $\sigma > 0$.

(ii) ω is a continuously differentiable function.

From the strong convexity of ω, problem (MNP) has a unique solution which will be denoted by $\mathbf{x}^*_{\text{mn}}$.

For simplicity, problem (P) will be called the *core problem*, problem (MNP) will be called *the outer problem* and correspondingly, ω will be called the *outer objective function*. It is obvious that problem (Q) is a special case of problem (MNP) with the choice $\omega(\mathbf{x}) \equiv \frac{1}{2}\|\mathbf{x}\|^2$. The so-called *prox center* of ω is given by

$$\mathbf{a} := \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \omega(\mathbf{x}).$$

We assume without loss of generality that $\omega(\mathbf{a}) = 0$. Under this setting we also have

$$\omega(\mathbf{x}) \geqslant \frac{\sigma}{2}\|\mathbf{x} - \mathbf{a}\|^2 \ \text{ for all} \mathbf{x} \in \mathbb{R}^n. \tag{6.1.2}$$

6.1.1 Stage by Stage Solution

It is important to note that the minimal norm-like solution optimization problem (MNP) can also be formally cast as the following convex optimization problem.

$$\begin{aligned} \min \quad & \omega(\mathbf{x}) \\ \text{s.t.} \quad & f(\mathbf{x}) \leqslant f^*, \\ & \mathbf{x} \in S. \end{aligned} \tag{6.1.3}$$

Of course, the optimal value of the core problem f^* is not known in advance, which suggests a solution method that consists of two stages: first find the optimal value of the core problem, and then solve problem (6.1.3). This two-stage solution technique has two main drawbacks. First, the optimal value f^* is often not found *exactly* but rather up to some tolerance, which causes the feasible set of the outer problem to be incorrect or even infeasible. Second, even if it had been possible to compute f^* exactly, problem (6.1.3) inherently does not satisfy Slater's condition, which means that this two-stage approach will usually run into numerical problems. We also note that the lack of regularity condition for Problem (6.1.3) implies that known optimality conditions such as Karush-Kuhn-Tucker are not valid; see for example the work [18] where different optimality conditions are derived.

6.2 Mathematical Toolbox

Two basic properties of the Bregman distances (see (1.2.1)) of strictly convex functions $h\mathbb{R}^n \rightarrow \mathbb{R}$ are:

(i) $D_h(\mathbf{x}, \mathbf{y}) \geqslant 0$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(ii) $D_h(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$.

If, in addition h, is *strongly convex* with parameter $\sigma > 0$, then

$$D_h(\mathbf{x}, \mathbf{y}) \geqslant \frac{\sigma}{2} \|\mathbf{x} - \mathbf{y}\|^2 .$$

In particular, the strongly convex function ω defined in Section 6.1 whose prox center is $\mathbf{a}$ satisfies:

$$\omega(\mathbf{x}) = D_\omega(\mathbf{x}, \mathbf{a}) \geqslant \frac{\sigma}{2} \|\mathbf{x} - \mathbf{a}\|^2 \text{ for any } \mathbf{x} \in \mathbb{R}^n$$

and

$$D_\omega(\mathbf{x}, \mathbf{y}) \geqslant \frac{\sigma}{2} \|\mathbf{x} - \mathbf{a}\|^2 \text{ for any } \mathbf{x} \in \mathbb{R}^n. \tag{6.2.1}$$

6.2.1 The Gradient Mapping

We define the following two mappings which are essential in our analysis of the proposed algorithm for solving (MNP).

Definition 6.2.1 (Gradient mapping). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. For every $M > 0$ we define the following two mappings.*

(i) *The proj-grad mapping is defined by*

$$T_M(\mathbf{x}) := P_S\left(\mathbf{x} - \frac{1}{M}\nabla f(\mathbf{x})\right) \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

(ii) *The gradient mapping (see also [80]) is defined by*

$$G_M(\mathbf{x}) := M(\mathbf{x} - T_M(\mathbf{x})) = M\left[\mathbf{x} - P_S\left(\mathbf{x} - \frac{1}{M}\nabla f(\mathbf{x})\right)\right].$$

Remark 6.2.2 (Unconstrained case). *In the unconstrained setting, that is, when $S = \mathbb{R}^n$, the orthogonal projection is the identity operator.*

(i) *The proj-grad mapping T_M is equal to $I - \frac{1}{M}\nabla f$.*

(ii) *The gradient mapping G_M is equal to ∇f.* ◇

It is well known that $G_M(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} \in S^*$. Another important and known property of the gradient mapping is the monotonicity of its norm L (*cf.* [20, Lemma 2.3.1, page 236]).

Lemma 6.2.3 (Monotonicity of the gradient mapping). *For any $\mathbf{x} \in \mathbb{R}^n$, the function*

$$g(M) := \|G_M(\mathbf{x})\| \qquad M > 0$$

is monotonically increasing over $(0, \infty)$.

6.2.2 Cutting Planes

The notion of a *cutting plane* is a fundamental concept in optimization algorithms such as the ellipsoid and the analytic cutting plane methods. As an illustration, let us first consider the unconstrained setting in which $S = \mathbb{R}^n$. Given a point $\mathbf{x} \in \mathbb{R}^n$, the idea is to find a hyperplane which separates $\mathbf{x}$ from S^*. For example, it is well known that for any $\mathbf{x} \in S$, the following inclusion holds

$$S^* \subseteq \{\mathbf{z} \in \mathbb{R}^n : \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{z}\rangle \geqslant 0\}.$$

The importance of the above result is that it "eliminates" the open half-space

$$\{\mathbf{z} \in \mathbb{R}^n : \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{z} \rangle < 0\}.$$

The same cut is also used in the ellipsoid method where in the nonsmooth case the gradient is replaced with a subgradient (see, *e.g.*, [19, 79]). Note that $\mathbf{x}$ belongs to the cut, that is, to the hyperplane given by

$$H := \{\mathbf{z} \in \mathbb{R}^n : \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{z} \rangle = 0\},$$

which means that H is a so-called *neutral cut*. In a *deep cut*, the point $\mathbf{x}$ does not belong to the corresponding hyperplane. Deep cuts are at the core of the minimal norm-like gradient method that will be described in the sequel, and in this subsection we describe how to construct them in several scenarios (specifically, known/unknown Lipschitz constant, constrained/unconstrained versions). The half-spaces corresponding to the deep cuts are always of the form

$$Q_{M,\alpha,\mathbf{x}} := \left\{\mathbf{z} \in \mathbb{R}^n : \langle G_M(\mathbf{x}), \mathbf{x} - \mathbf{z} \rangle \geqslant \frac{1}{\alpha M} \|G_M(\mathbf{x})\|^2\right\}, \tag{6.2.2}$$

where the values of α and M depend on the specific scenario. Of course, in the unconstrained case, $G_M(\mathbf{x}) = \nabla f(\mathbf{x})$ (see Remark 6.2.2), and (6.2.2) reads as

$$Q_{M,\alpha,\mathbf{x}} := \left\{\mathbf{z} \in \mathbb{R}^n : \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{z} \rangle \geqslant \frac{1}{\alpha M} \|\nabla f(\mathbf{x})\|^2\right\}.$$

We will now split the analysis into two scenarios. In the first one, the Lipschitz constant L is known, while in the second, it is not.

Known Lipschitz Constant

In the unconstrained case ($S = \mathbb{R}^n$), and when the Lipschitz constant L is known, we can use the following known inequality (see, *e.g.*, [80, Theorem 2.1.5, page 56]):

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geqslant \frac{1}{L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \text{ for every } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{6.2.3}$$

By plugging $\mathbf{y} = \mathbf{x}^*$ for some $\mathbf{x}^* \in S^*$ in (6.2.3) and recalling that $\nabla f(\mathbf{x}^*) = \mathbf{0}$, we obtain that

$$\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \geqslant \frac{1}{L} \|\nabla f(\mathbf{x})\|^2 \tag{6.2.4}$$

for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x}^* \in S^*$. Thus, $S^* \subseteq Q_{L,1,\mathbf{x}}$ for any $\mathbf{x} \in \mathbb{R}^n$.

When S is not the entire space $\mathbb{R}^n$, the generalization of (6.2.4) is a bit intricate and in fact the result we can prove is the slightly weaker inclusion $S^* \subseteq Q_{L,\frac{4}{3},\mathbf{x}}$. The result is based on the following property of the gradient mapping G_L which was proven in the thesis [16] and is given here for the sake of completeness.

Lemma 6.2.4 (Property of the gradient mapping)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function whose gradient is Lipschitz with constant L. The gradient mapping G_L satisfies the following relation:*

$$\langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geqslant \frac{3}{4L} \left\| G_L(\mathbf{x}) - G_L(\mathbf{y}) \right\|^2 \tag{6.2.5}$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proof. From Corollary 1.2.39(i) it follows that

$$\left\langle T_L(\mathbf{x}) - T_L(\mathbf{y}), \left(\mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{L}\nabla f(\mathbf{y})\right) \right\rangle \geqslant \left\| T_L(\mathbf{x}) - T_L(\mathbf{y}) \right\|^2.$$

Since $T_L = I - \frac{1}{L}G_L$, we obtain that

$$\begin{aligned} &\left\langle \left(\mathbf{x} - \frac{1}{L}G_L(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{L}G_L(\mathbf{y})\right), \left(\mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{L}\nabla f(\mathbf{y})\right) \right\rangle \\ &\geqslant \left\| \left(\mathbf{x} - \frac{1}{L}G_L(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{L}G_L(\mathbf{y})\right) \right\|^2, \end{aligned}$$

which is equivalent to

$$\left\langle \left(\mathbf{x} - \frac{1}{L}G_L(\mathbf{x})\right) - \left(\mathbf{y} - \frac{1}{L}G_L(\mathbf{y})\right), \left(G_L(\mathbf{x}) - \nabla f(\mathbf{x})\right) - \left(G_L(\mathbf{y}) - \nabla f(\mathbf{y})\right) \right\rangle \geqslant 0.$$

Thence

$$\begin{aligned} \langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geqslant{} & \frac{1}{L} \left\| G_L(\mathbf{x}) - G_L(\mathbf{y}) \right\|^2 + \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \\ & - \frac{1}{L} \langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) \rangle. \end{aligned}$$

Now it follows from (6.2.3) that

$$\begin{aligned} L \langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geqslant{} & \left\| G_L(\mathbf{x}) - G_L(\mathbf{y}) \right\|^2 + \left\| \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) \right\|^2 \\ & \langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) \rangle. \end{aligned}$$

From the Cauchy-Schwarz inequality we get

$$L\langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle \geqslant \|G_L(\mathbf{x}) - G_L(\mathbf{y})\|^2 + \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 - \|G_L(\mathbf{x}) - G_L(\mathbf{y})\| \cdot \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|. \qquad (6.2.6)$$

By denoting $\alpha = \|G_L(\mathbf{x}) - G_L(\mathbf{y})\|$ and $\beta = \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|$, the right-hand side of (6.2.6) reads as $\alpha^2 + \beta^2 - \alpha\beta$ and satisfies

$$\alpha^2 + \beta^2 - \alpha\beta = \frac{3}{4}\alpha^2 + \left(\frac{\alpha}{2} - \beta\right)^2 \geqslant \frac{3}{4}\alpha^2,$$

which combined with (6.2.6) yields the inequality

$$L\langle G_L(\mathbf{x}) - G_L(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle \geqslant \frac{3}{4}\|G_L(\mathbf{x}) - G_L(\mathbf{y})\|^2.$$

Thus, (6.2.5) holds. □

By plugging $\mathbf{y} = \mathbf{x}^*$ for some $\mathbf{x}^* \in S^*$ in (6.2.5), we obtain that indeed

$$S^* \subseteq Q_{L,\frac{4}{3},\mathbf{x}}.$$

We summarize the above discussion in the following lemma which describes the deep cuts in the case when the Lipschitz constant is known.

Lemma 6.2.5 (Deep cuts - Lipschitz constant is known). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function whose gradient is Lipschitz with constant L. For any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x}^* \in S^*$, we have*

$$\langle G_L(\mathbf{x}), \mathbf{x}-\mathbf{x}^*\rangle \geqslant \frac{3}{4L}\|G_L(\mathbf{x})\|^2, \qquad (6.2.7)$$

that is,

$$S^* \subseteq Q_{L,\frac{4}{3},\mathbf{x}}.$$

If, in addition, $S = \mathbb{R}^n$ then

$$\langle \nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^*\rangle \geqslant \frac{1}{L}\|\nabla f(\mathbf{x})\|^2, \qquad (6.2.8)$$

that is,

$$S^* \subseteq Q_{L,1,\mathbf{x}}.$$

Unknown Lipschitz Constant

When the Lipschitz constant is not known, the following result is most useful.

Lemma 6.2.6 (Deep cuts - Lipschitz constant is not known). *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a continuously differentiable function. Let* $\mathbf{x} \in \mathbb{R}^n$ *be a vector satisfying the inequality*

$$f\left(T_M\left(\mathbf{x}\right)\right) \leqslant f\left(\mathbf{x}\right) + \left\langle \nabla f\left(\mathbf{x}\right), T_M\left(\mathbf{x}\right) - \mathbf{x} \right\rangle + \frac{M}{2}\left\|T_M\left(\mathbf{x}\right) - \mathbf{x}\right\|^2. \tag{6.2.9}$$

Then, for any $\mathbf{x}^* \in S^*$, *the inequality*

$$\left\langle G_M\left(\mathbf{x}\right), \mathbf{x} - \mathbf{x}^* \right\rangle \geqslant \frac{1}{2M}\left\|G_M\left(\mathbf{x}\right)\right\|^2 \tag{6.2.10}$$

holds true, that is,

$$S^* \subseteq Q_{M,2,\mathbf{x}}.$$

Proof. Let $\mathbf{x}^* \in S^*$. From (6.2.9) it follows that

$$0 \leqslant f\left(T_M\left(\mathbf{x}\right)\right) - f\left(\mathbf{x}^*\right) \leqslant f\left(\mathbf{x}\right) - f\left(\mathbf{x}^*\right) + \left\langle \nabla f\left(\mathbf{x}\right), T_M\left(\mathbf{x}\right) - \mathbf{x} \right\rangle + \frac{M}{2}\left\|T_M\left(\mathbf{x}\right) - \mathbf{x}\right\|^2. \tag{6.2.11}$$

Since f is convex, it follows from the subdifferential inequality (see (1.1.5)) that $f\left(\mathbf{x}\right) - f\left(\mathbf{x}^*\right) \leqslant \left\langle \nabla f\left(\mathbf{x}\right), \mathbf{x} - \mathbf{x}^* \right\rangle$, which combined with (6.2.11) yields

$$0 \leqslant \left\langle \nabla f\left(\mathbf{x}\right), T_M\left(\mathbf{x}\right) - \mathbf{x}^* \right\rangle + \frac{M}{2}\left\|T_M\left(\mathbf{x}\right) - \mathbf{x}\right\|^2. \tag{6.2.12}$$

In addition, from the definition of T_M (see Definition 6.2.1(i)) and Corollary 1.2.39(ii) we have the following inequality

$$\left\langle \mathbf{x} - \frac{1}{M}\nabla f\left(\mathbf{x}\right) - T_M\left(\mathbf{x}\right), T_M\left(\mathbf{x}\right) - \mathbf{x}^* \right\rangle \geqslant 0.$$

Summing up the latter inequality multiplied by M with (6.2.12) yields the inequality

$$M\left\langle \mathbf{x} - T_M\left(\mathbf{x}\right), T_M\left(\mathbf{x}\right) - \mathbf{x}^* \right\rangle + \frac{M}{2}\left\|T_M\left(\mathbf{x}\right) - \mathbf{x}\right\|^2 \geqslant 0,$$

which after some simple algebraic manipulation, can be shown to be equivalent to the desired result (6.2.10). □

When $M \geqslant L$, the inequality (6.2.9) is satisfied due to the so-called descent lemma, which is now recalled as it will also be essential in our analysis (see [20]).

Lemma 6.2.7 (Descent lemma). *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a continuously differentiable function whose gradient is Lipschitz with constant* L. *Then for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$f\left(\mathbf{x}\right) \leqslant f\left(\mathbf{y}\right) + \left\langle \nabla f\left(\mathbf{y}\right), \mathbf{x} - \mathbf{y} \right\rangle + \frac{L}{2}\left\|\mathbf{x} - \mathbf{y}\right\|^2. \tag{6.2.13}$$

Remark 6.2.8. *The inequality* (6.2.9) *for* $M \geqslant L$ *is well known, see for example [80].* $\diamond$

6.3 The Minimal Norm Gradient Algorithm

Before describing the algorithm, we require the following notation for the optimal solution of the problem consisting of the minimization ω over a given closed and convex set K.

$$\Omega(K) := \operatorname{argmin}_{\mathbf{x} \in K} \omega(\mathbf{x}). \tag{6.3.1}$$

From the optimality conditions (in this connection, see also Proposition 1.2.35) in problem (6.3.1), it follows that

$$\tilde{\mathbf{x}} = \Omega(K) \Leftrightarrow \langle \nabla\omega(\tilde{\mathbf{x}}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \geqslant 0 \text{ for all } \mathbf{x} \in K. \tag{6.3.2}$$

If $\omega(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{a}\|^2$, then $\Omega(K) = P_K(\mathbf{a})$. We are now ready to describe the algorithm in the case when the Lipschitz constant L is known.

The Minimal Norm Gradient Method (Known Lipschitz Constant)
Input: L - a Lipschitz constant of ∇f.
Initialization: $\mathbf{x}_0 = \mathbf{a}$.
General Step $(k = 1, 2, \ldots)$**:**

$$\mathbf{x}_k = \Omega(Q_k \cap W_k),$$

where

$$\begin{aligned} Q_k &= Q_{L,\beta,\mathbf{x}_{k-1}}, \\ W_k &= \{\mathbf{z} \in \mathbb{R}^n : \langle \nabla\omega(\mathbf{x}_{k-1}), \mathbf{z} - \mathbf{x}_{k-1} \rangle \geqslant 0\}, \end{aligned}$$

and β is equal to $\frac{4}{3}$ if $S \neq \mathbb{R}^n$ and to 1 if $S = \mathbb{R}^n$.

When the Lipschitz constant is unknown, then a backtracking procedure should be incorporated into the method.

The Minimal Norm Gradient Method (Unknown Lipschitz Constant)
Input: $L_0 > 0$ and $\eta > 1$.
Initialization: $\mathbf{x}_0 = \mathbf{a}$.
General Step $(k = 1, 2, \ldots)$**:**

(i) Find the smallest nonnegative integer number i_k such that with $\bar{L} = \eta_{i_k} L_{k-1}$ the inequality

$$(T_{\bar{L}}(\mathbf{x})) \leqslant f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), T_{\bar{L}}(\mathbf{x}) - \mathbf{x} \rangle + \frac{\bar{L}}{2} \|T_{\bar{L}}(\mathbf{x}) - \mathbf{x}\|^2$$

is satisfied and set $L_k = \bar{L}$.

(ii) Set

$$\mathbf{x}_k = \Omega(Q_k \cap W_k),$$

where

$$Q_k = Q_{L_k, 2, \mathbf{x}_{k-1}},$$
$$W_k = \{\mathbf{z} \in \mathbb{R}^n : \langle \nabla \omega(\mathbf{x}_{k-1}), \mathbf{z} - \mathbf{x}_{k-1} \rangle \geqslant 0\}.$$

To unify the analysis, in the constant step-size setting we will artificially define $L_k = L$ for any k and $\eta = 1$. In this notation the definition of the half-space Q^k in both the constant and backtracking step-size rules can be described as

$$Q_k = Q_{L_k, \beta, \mathbf{x}_{k-1}}, \tag{6.3.3}$$

where β is given by

$$\beta := \begin{cases} \frac{4}{3} & S \neq \mathbb{R}^n, \text{ known Lipschitz const.} \\ 1 & S = \mathbb{R}^n, \text{ known Lipschitz const.} \\ 2 & \text{unknown Lipschitz const.} \end{cases} \tag{6.3.4}$$

Remark 6.3.1. *From the definition of the backtracking rule it follows that*

$$L_0 \leqslant L_k \leqslant \eta L, \quad k = 0, 1, \ldots. \tag{6.3.5}$$

Therefore, it follows from Lemma 6.2.3 that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\|G_{L_0}(\mathbf{x})\| \leqslant \|G_{L_n}(\mathbf{x})\| \leqslant \|G_{\eta L}(\mathbf{x})\| . \tag{6.3.6}$$

◇

The following example shows that in the Euclidean setting, the main step has a simple and explicit formula.

Example 6.3.2. *In the Euclidean setting when* $\omega = \frac{1}{2}\|\cdot\|^2$, *we have* $\Omega(K) = P_K$ *and the computation of the main step*

$$\mathbf{x}_k = \Omega(Q_k \cap W_k)$$

boils down to finding the orthogonal projection onto an intersection of two half-spaces. This is a simple task, since the orthogonal projection onto the intersection of two half-spaces is given by an exact formula (see Example 1.2.41).

Note that the algorithm is well defined as long as the set $Q_k \cap W_k$ is nonempty. The latter property does hold true and we will now show a stronger result stating that in fact $S^* \subseteq Q_k \cap W_k$ for all k.

Lemma 6.3.3 (The intersection $Q_k \cap W_k$ is nonempty). *Let* $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ *be the sequence generated by the minimal norm gradient method with either a constant or a backtracking step-size rule. Then*

$$S^* \subseteq Q_k \cap W_k \tag{6.3.7}$$

for any $k \in \mathbb{N}$.

Proof. From Lemmata 6.2.5 and 6.2.6 it follows that $S^* \subseteq Q_k$ for every $k \in \mathbb{N}$ and we will now prove by induction on k that $S^* \subseteq W_k$. Since $W_1 = \mathbb{R}^n$, the claim is trivial for $k = 1$. Suppose that the claim holds for $k = n$, that is, we assume that $S^* \subseteq W_n$. To prove that $S^* \subseteq Q_{n+1} \cap W_{n+1}$, let us take $\mathbf{u} \in S^*$. Note that $S^* \subseteq Q_n \cap W_n$, and thus, since $\mathbf{x}_n = \Omega(Q_n \cap W_n)$, it follows from (6.3.2) that

$$\langle \nabla\omega(\mathbf{x}_n), \mathbf{x}_n - \mathbf{u}\rangle \geqslant 0.$$

This implies that $\mathbf{u} \in W_{n+1}$ and the claim that $S^* \subseteq Q_k \cap W_k$ for all $k \in \mathbb{N}$ is proven. □

6.4 Convergence Analysis

Our first claim is that the minimal norm gradient method generates a sequence $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ which converges to $\mathbf{x}^*_{\text{mn}} = \Omega(S^*)$.

Theorem 6.4.1 (Convergence result). *Let $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ be the sequence generated by the minimal norm gradient method with either a constant or a backtracking step-size rule. Then the following assertions are true.*

(i) *The sequence $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ is bounded.*

(ii) *The following inequality holds for any $k \in \mathbb{N}$:*

$$D_\omega(\mathbf{x}_k, \mathbf{x}_{k-1}) + D_\omega(\mathbf{x}_{k-1}, \mathbf{a}) \leqslant D_\omega(\mathbf{x}_k, \mathbf{a}). \tag{6.4.1}$$

(iii) $\mathbf{x}_k \to \mathbf{x}^*_{\mathrm{mn}}$ *as $k \to \infty$.*

Proof. (i) Since $\mathbf{x}_k = \Omega(Q_k \cap W_k)$, it follows that for any $\mathbf{u} \in Q_k \cap W_k$, and in particular for any $\mathbf{u} \in S^*$ we have

$$\omega(\mathbf{x}_k) \leqslant \omega(\mathbf{u}), \tag{6.4.2}$$

which combined with (6.1.2) establishes the boundedness of $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$.

(ii) From the three point identity (see (1.2.2)) we have

$$D_\omega(\mathbf{x}_k, \mathbf{x}_{k-1}) + D_\omega(\mathbf{x}_{k-1}, \mathbf{a}) - D_\omega(\mathbf{x}_k, \mathbf{a}) = \langle -\nabla\omega(\mathbf{x}_{k-1}), \mathbf{x}_k - \mathbf{x}_{k-1}\rangle.$$

From the definition of W_k we have $\mathbf{x}_{k-1} = \Omega(W_k)$. In addition, $\mathbf{x}_k \in W_k$, and hence from (6.3.2) it follows that

$$\langle \nabla\omega(\mathbf{x}_{k-1}), \mathbf{x}_k - \mathbf{x}_{k-1}\rangle \geqslant 0$$

and therefore (6.4.1) follows.

(iii) Recall that for any $\mathbf{x} \in \mathbb{R}^n$, we have $D_\omega(\mathbf{x}, \mathbf{a}) = \omega(\mathbf{x})$. From (6.4.1) it follows that the sequence $\{\omega(\mathbf{x}_k)\}_{k\in\mathbb{N}} = \{D_\omega(\mathbf{x}_k, \mathbf{a})\}_{k\in\mathbb{N}}$ is nondecreasing and bounded, and hence $\lim_{k\to\infty} \omega(\mathbf{x}_k)$ exists. This, together with (6.4.1) implies that

$$\lim_{k\to\infty} D_\omega(\mathbf{x}_k, \mathbf{x}_{k-1}) = 0,$$

and hence, since $D_\omega(\mathbf{x}_k, \mathbf{x}_{k-1}) \geqslant \frac{\sigma}{2}\|\mathbf{x}_k - \mathbf{x}_{k-1}\|^2$, it follows that

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}_{k-1}\| = 0. \tag{6.4.3}$$

Since $\mathbf{x}_k \in Q_k$ we have

$$\langle G_{L_k}(\mathbf{x}_{k-1}), \mathbf{x}_{k-1} - \mathbf{x}_k\rangle \geqslant \frac{1}{\beta L_k}\|G_{L_k}(\mathbf{x}_{k-1})\|^2,$$

which by the Cauchy-Schwarz inequality, implies that

$$\frac{1}{\beta L_k} \left\| G_{L_k} \left(\mathbf{x}_{k-1} \right) \right\| \leqslant \left\| \mathbf{x}_{k-1} - \mathbf{x}_k \right\| .$$

Now, from (6.3.5) and (6.3.6) it follows that

$$\frac{1}{\eta L} \left\| G_{L_0} \left(\mathbf{x}_{k-1} \right) \right\| \leqslant \left\| \mathbf{x}_{k-1} - \mathbf{x}_k \right\| . \tag{6.4.4}$$

To show that $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ converges to $\mathbf{x}^*_{\text{mn}}$, it is enough to show that any convergent subsequence converges to $\mathbf{x}^*_{\text{mn}}$. Let then $\{\mathbf{x}_{k_n}\}_{n \in \mathbb{N}}$ be a convergent subsequence whose limit is $\mathbf{w}$. From (6.4.3) and (6.4.4) along with the continuity of G_{L_0}, it follows that $G_{L_0}(\mathbf{w}) = 0$, so that $\mathbf{w} \in S^*$. Finally, we will prove that $\mathbf{w} = \Omega(S^*) = \mathbf{x}^*_{\text{mn}}$. Since $\mathbf{x}_{k_n} = \Omega(Q_{k_n} \cap W_{k_n})$, it follows from (6.3.2) that

$$\left\langle \nabla \omega \left(\mathbf{x}_{k_n} \right), \mathbf{z} - \mathbf{x}_{k_n} \right\rangle \geqslant 0 \text{ for all } \mathbf{z} \in Q_{k_n} \cap W_{k_n}.$$

Since $S^* \subseteq Q_{k_n} \cap W_{k_n}$ (see Lemma 6.3.3), we obtain that

$$\left\langle \nabla \omega \left(\mathbf{x}_{k_n} \right), \mathbf{z} - \mathbf{x}_{k_n} \right\rangle \geqslant 0 \text{ for all } \mathbf{z} \in S^*.$$

Taking the limit as $n \to \infty$, and using the continuity of $\nabla \omega$, we get

$$\left\langle \nabla \omega \left(\mathbf{w} \right), \mathbf{z} - \mathbf{w} \right\rangle \geqslant 0 \text{ for all } \mathbf{z} \in S^*.$$

Therefore, it follows from (6.3.2) that $\mathbf{w} = \Omega(S^*) = \mathbf{x}^*_{\text{mn}}$, and the result is proven. □

The next result shows that in the unconstrained case ($S = \mathbb{R}^n$), the function values of the sequence generated by the minimal norm gradient method, $\{f(\mathbf{x}_k)\}_{k \in \mathbb{N}}$, converges in a rate of $O\left(1/\sqrt{k}\right)$ (k being the iteration index) to the optimal value of the core problem. In the constrained case, the value $f(\mathbf{x}_k)$ is by no means a measure of the quality of the iterate $\mathbf{x}_k$ as it is not necessarily feasible. Instead, we will show that the rate of convergence of the function values of the *feasible* sequence $T_{L_k}(\mathbf{x}_k)$ (which in any case is computed by the algorithm), is also $O\left(1/\sqrt{k}\right)$. We also note that since the minimal norm gradient method is non-monotone, the convergence results are with respect to the minimal function value obtained until iteration k.

Theorem 6.4.2 (Rate of convergence)**.** *Let $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ be the sequence generated by the minimal norm gradient method with either a constant or backtracking step-size rules. Then for* *every* $k \geqslant 1$*, one has*

$$\min_{1 \leqslant n \leqslant k} f\left(T_{L_n}\left(\mathbf{x}_n\right)\right) - f^* \leqslant \frac{\beta \eta L \left\|\mathbf{a} - \mathbf{x}^*_{\mathrm{mn}}\right\|^2}{\sqrt{k}}, \tag{6.4.5}$$

where β *is given in* (6.3.4)*. If* $X = \mathbb{R}^n$*, then in addition*

$$\min_{1 \leqslant n \leqslant k} f\left(\mathbf{x}_n\right) - f^* \leqslant \frac{\beta \eta L \left\|\mathbf{a} - \mathbf{x}^*_{\mathrm{mn}}\right\|^2}{\sqrt{k}}. \tag{6.4.6}$$

Proof. Let n be a nonnegative integer. Since $\mathbf{x}_{n+1} \in Q_{n+1}$, we have by the Cauchy-Schwarz inequality

$$\left\|G_{L_{n+1}}\left(\mathbf{x}_n\right)\right\|^2 \leqslant \beta L_{n+1} \left\langle G_{L_{n+1}}\left(\mathbf{x}_n\right), \mathbf{x}_n - \mathbf{x}_{n+1}\right\rangle \leqslant \beta L_{n+1} \left\|G_{L_{n+1}}\left(\mathbf{x}_n\right)\right\| \left\|\mathbf{x}_n - \mathbf{x}_{n+1}\right\|.$$

Therefore,

$$\left\|G_{L_{n+1}}\left(\mathbf{x}_n\right)\right\| \leqslant \beta L_{n+1} \left\|\mathbf{x}_n - \mathbf{x}_{n+1}\right\|. \tag{6.4.7}$$

Squaring (6.4.7) and summing up over $n = 1, 2, \ldots, k$, one obtains

$$\sum_{n=1}^{k} \left\|G_{L_{n+1}}\left(\mathbf{x}_n\right)\right\|^2 \leqslant \beta^2 L_{n+1}^2 \sum_{n=1}^{N} \left\|\mathbf{x}_{n+1} - \mathbf{x}_n\right\|^2 \leqslant \beta^2 \eta^2 L^2 \sum_{n=1}^{N} \left\|\mathbf{x}_{n+1} - \mathbf{x}_n\right\|^2. \tag{6.4.8}$$

Taking into account (6.2.1) and (6.4.1), then from (6.4.8) we get

$$\begin{aligned}
\sum_{n=1}^{k} \left\|G_{L_{n+1}}\left(\mathbf{x}_n\right)\right\|^2 &\leqslant \beta^2 \eta^2 L^2 \sum_{n=1}^{k} \left\|\mathbf{x}_{n+1} - \mathbf{x}_n\right\|^2 \\
&\leqslant \frac{2\beta^2 \eta^2 L^2}{\sigma} \sum_{n=1}^{k} D_\omega\left(\mathbf{x}_{n+1}, \mathbf{x}_n\right) \\
&\leqslant \frac{2\beta^2 \eta^2 L^2}{\sigma} \sum_{n=1}^{k} \left(D_\omega\left(\mathbf{x}_{n+1}, \mathbf{a}\right) - D_\omega\left(\mathbf{x}_n, \mathbf{a}\right)\right) \\
&= \frac{2\beta^2 \eta^2 L^2}{\sigma} D_\omega\left(\mathbf{x}_{k+1}, \mathbf{a}\right) = \frac{2\beta^2 \eta^2 L^2}{\sigma} \omega\left(\mathbf{x}_{k+1}\right) \\
&\leqslant \frac{2\beta^2 \eta^2 L^2}{\sigma} \omega\left(\mathbf{x}^*_{\mathrm{mn}}\right).
\end{aligned} \tag{6.4.9}$$

From the definition of L_n, we obtain

$$f\left(T_{L_n}\left(\mathbf{x}_n\right)\right) - f^* \leqslant f\left(\mathbf{x}_n\right) - f^* + \left\langle \nabla f\left(\mathbf{x}_n\right), T_{L_n}\left(\mathbf{x}_n\right) - \mathbf{x}_n\right\rangle + \frac{L_n}{2} \left\|T_{L_n}\left(\mathbf{x}_n\right) - \mathbf{x}_n\right\|^2. \tag{6.4.10}$$

Since the function f is convex it follows from the subdifferential inequality (see (1.1.5)) that $f(\mathbf{x}_n) - f^* \leqslant \langle \nabla f(\mathbf{x}_n), \mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}} \rangle$, which combined with (6.4.10) yields

$$f(T_{L_n}(\mathbf{x}_n)) - f^* \leqslant \langle \nabla f(\mathbf{x}_n), T_{L_n}(\mathbf{x}_n) - \mathbf{x}^*_{\mathrm{mn}} \rangle + \frac{L_n}{2} \|T_{L_n}(\mathbf{x}_n) - \mathbf{x}_n\|^2. \tag{6.4.11}$$

By the characterization of the projection operator given in Proposition 1.2.38 with $\mathbf{x} = \mathbf{x}_n - \frac{1}{L_n}\nabla f(\mathbf{x}_n)$ and $\mathbf{y} = \mathbf{x}^*_{\mathrm{mn}}$, we have that

$$\left\langle \mathbf{x}_n - \frac{1}{L_n}\nabla f(\mathbf{x}_n) - T_{L_n}(\mathbf{x}_n), \mathbf{x}^*_{\mathrm{mn}} - T_{L_n}(\mathbf{x}_n) \right\rangle \leqslant 0,$$

which combined with (6.4.11) gives

$$\begin{aligned} f(T_{L_n}(\mathbf{x}_n)) - f^* &\leqslant L_n \langle \mathbf{x}_n - T_{L_n}(\mathbf{x}_n), T_{L_n}(\mathbf{x}_n) - \mathbf{x}_{\mathrm{mn}*} \rangle + \frac{L_n}{2} T_{L_n}(\mathbf{x}_n) - \mathbf{x}_n{}^2 \\ &= \langle G_{L_n}(\mathbf{x}_n), T_{L_n}(\mathbf{x}_n) - \mathbf{x}^*_{\mathrm{mn}} \rangle + \frac{1}{2L_n} \|G_{L_n}(\mathbf{x}_n)\|^2 \\ &= \langle G_{L_n}(\mathbf{x}_n), T_{L_n}(\mathbf{x}_n) - \mathbf{x}_n \rangle + \langle G_{L_n}(\mathbf{x}_n), \mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}} \rangle + \frac{1}{2L_n} \|G_{L_n}(\mathbf{x}_n)\|^2 \\ &= \langle G_{L_n}(\mathbf{x}_n), \mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}} \rangle - \frac{1}{2L_n} \|G_{L_n}(\mathbf{x}_n)\|^2 \\ &\leqslant \langle G_{L_n}(\mathbf{x}_n), \mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}} \rangle \\ &\leqslant \|G_{L_n}(\mathbf{x}_n)\| \|\mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}}\|. \end{aligned}$$

Squaring the above inequality and summing over $n = 1, 2, \ldots, k$, we get

$$\sum_{n=1}^{k} (f(T_{L_n}(\mathbf{x}_n)) - f^*)^2 \leqslant \sum_{n=1}^{k} \|G_{L_n}(\mathbf{x}_n)\|^2 \|\mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}}\|^2. \tag{6.4.12}$$

Now, from the three point identity (see (1.2.2)), we obtain that

$$D_\omega(\mathbf{x}^*_{\mathrm{mn}}, \mathbf{x}_n) + D_\omega(\mathbf{x}_n, \mathbf{a}) - D_\omega(\mathbf{x}^*_{\mathrm{mn}}, \mathbf{a}) = -\langle \nabla\omega(\mathbf{x}_n), \mathrm{xmn}^* - \mathbf{x}_n \rangle \leqslant 0$$

and hence

$$D_\omega(\mathbf{x}^*_{\mathrm{mn}}, \mathbf{x}_n) \leqslant D_\omega(\mathbf{x}^*_{\mathrm{mn}}, \mathbf{a}) = \omega(\mathbf{x}^*_{\mathrm{mn}}),$$

so that

$$\|\mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}}\|^2 \leqslant \frac{2\omega(\mathbf{x}^*_{\mathrm{mn}})}{\sigma}. \tag{6.4.13}$$

Combining (6.4.9) and (6.4.12) along with (6.4.13) we get that

$$\sum_{n=1}^{k} (f(T_{L_n}(\mathbf{x}_n)) - f^*)^2 \leqslant \frac{2\omega(\mathbf{x}^*_{\mathrm{mn}})}{\sigma} \sum_{n=1}^{k} \|G_{L_n}(\mathbf{x}_n)\|^2 \leqslant \frac{4\beta^2\eta^2 L^2}{\sigma^2}\omega(\mathbf{x}^*_{\mathrm{mn}})^2$$

from which we obtain that

$$k \min_{n=1,2,\ldots,k} \left(f\left(T_{L_n}\left(\mathbf{x}_n\right)\right) - f^*\right)^2 \leqslant \frac{4\beta^2\eta^2L^2}{\sigma^2}\omega\left(\mathbf{x}^*_{\mathrm{mn}}\right)^2,$$

proving the result (6.4.5). The result (6.4.6) in the case when $S = \mathbb{R}^n$ is established by following the same line of proof along with the observation that due to the convexity of f

$$f\left(\mathbf{x}_n\right) - f^* \leqslant \left\|\nabla f\left(\mathbf{x}_n\right)\right\| \left\|\mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}}\right\| = \left\|G_{L_n}\left(\mathbf{x}_n\right)\right\| \left\|\mathbf{x}_n - \mathbf{x}^*_{\mathrm{mn}}\right\|. \qquad \square$$

6.5 A Numerical Example - A Portfolio Optimization Problem

Consider the Markowitz portfolio optimization problem [73]. Suppose that we are given N assets numbered $1, 2, \ldots, N$ for which a vector of expected returns $\boldsymbol{\mu} \in \mathbb{R}^N$ and a positive semidefinite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ are known. In the Markowitz portfolio optimization problem we seek to find a minimum variance portfolio subject to the constraint that the expected return is greater or equal to a certain predefined minimal value r_0.

$$\begin{array}{ll} \min & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{s.t.} & \sum_{i=1}^N w_i = 1, \\ & \mathbf{w}^T\boldsymbol{\mu} \geqslant r_0, \\ & \mathbf{w} \geqslant \mathbf{0}. \end{array} \tag{6.5.1}$$

The decision variables vector $\mathbf{w}$ describes the allocation of the given resource to the different assets.

When the covariance matrix is rank deficient (that is, positive semidefinite but not positive definite), the optimal solution is not unique, and a natural issue in this scenario is to find one portfolio among all the optimal portfolios that is "best" with respect to an objective function different than the portfolio variance. This is, of course, a minimal norm-like solution optimization problem. We note that the situation in which the covariance matrix is rank deficient is quite common since the covariance matrix is usually estimated from the past trading price data and when the number of sampled periods is smaller than the number of assets, the covariance matrix is surely rank deficient. As a specific example, consider the portfolio optimization problem given by (6.5.1), where the expected returns vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are both estimated from real data on 8 types of assets ($N = 8$): US 3 month treasury bills, US government long bonds, SP 500, Wilshire 500, NASDAQ composite, corporate bond index, EAFE and Gold. The yearly returns are from 1973 to 1994.

The data can be found at http://www.princeton.edu/∼rvdb/ampl/nlmodels/markowitz/ and we have used the data between the years 1974 and 1977 in order to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ which are given below

$$\boldsymbol{\mu} = (1.0630, 1.0633, 1.0670, 1.0853, 1.0882, 1.0778, 1.0820, 1.1605)^T$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.0002 & -0.0005 & -0.0028 & -0.0032 & -0.0039 & -0.0007 & -0.0024 & 0.0048 \\ -0.0005 & 0.0061 & 0.0132 & 0.0136 & 0.0126 & 0.0049 & -0.0003 & -0.0154 \\ -0.0028 & 0.0132 & 0.0837 & 0.0866 & 0.0810 & 0.0196 & 0.0544 & -0.1159 \\ -0.0032 & 0.0136 & 0.0866 & 0.0904 & 0.0868 & 0.0203 & 0.0587 & -0.1227 \\ -0.0039 & 0.0126 & 0.0810 & 0.0868 & 0.0904 & 0.0192 & 0.0620 & -0.1232 \\ -0.0007 & 0.0049 & 0.0196 & 0.0203 & 0.0192 & 0.0054 & 0.0090 & -0.0261 \\ -0.0024 & -0.0003 & 0.0544 & 0.0587 & 0.0620 & 0.0090 & 0.0619 & -0.0900 \\ 0.0048 & -0.0154 & -0.1159 & -0.1227 & -0.1232 & -0.0261 & -0.0900 & 0.1725 \end{pmatrix}.$$

The sampled covariance matrix was computed via the following known formula for an unbiased estimator of the covariance matrix

$$\boldsymbol{\Sigma} := \frac{1}{T-1}\mathbf{R}\left(\mathbf{I}_T - \frac{1}{T}\mathbf{1}\mathbf{1}^T\right)\mathbf{R}^T.$$

Here $T = 4$ (number of periods) and $\mathbf{R}$ is the 8×4 matrix containing the assets' returns for each of the 4 years. The rank of the matrix $\boldsymbol{\Sigma}$ is at most 4, thus it is rank deficient. We have chosen the minimal return as $r_0 = 1.05$. In this case the portfolio problem (6.5.1) has multiple optimal solution, and we therefore consider problem (6.5.1) as the core problem and introduce a second objective function for the outer problem. Here we choose

$$\omega(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{a}\|^2.$$

Suppose that we wish to invest as much as possible in gold. Then we choose

$$\mathbf{a} = (0, 0, 0, 0, 0, 0, 0, 1)^T$$

and in this case the minimal norm gradient method gives the solution

$$(0.0000, 0.0000, 0.0995, 0.1421, 0.2323, 0.0000, 0.1261, 0.3999)^T.$$

If we wish a portfolio which is as dispersed as possible, then we choose

$$\mathbf{a} = (1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8)^T,$$

and in this case the algorithm produces the following optimal solution:

$$(0.1531, 0.1214, 0.0457, 0.0545, 0.1004, 0.1227, 0.1558, 0.2466)^T \ ,$$

which is very much different from the first optimal solution. Note that in the second optimal solution the investment in gold is much smaller and that the allocation of the resources is indeed much more scattered.

Index

References

[1] Alber, Y. I. and Butnariu, D. and Ryazantseva, I.: Regularization and reslution of monotone variational inequalities with operators given by hypomonotone approximations, *J. Nonlinear Convex Anal.* **6** (2005), 23–53. (Cited on page 61).

[2] Alber, Y. I. and Reich, S.: An iterative method for solving a class of nonlinear operator equations in Banach spaces, *Panamer. Math. J.* **4** (1994), 39–54. (Cited on page 64).

[3] Ambrosetti, A. and Prodi, G.: A Primer of Nonlinear Analysis, *Cambridge University Press*, Cambridge, 1993. (Cited on page 18).

[4] Attouch, H.: Variational Convergence for Functions and Operators, *Pitman*, Boston, 1984. (Cited on page 108).

[5] Bauschke, H. H. and Borwein, J. M.: On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996), 367–426. (Cited on page 156).

[6] Bauschke, H. H. and Borwein, J. M.: Legendre functions and the method of random Bregman projections, *J. Convex Anal.* **4** (1997), 27–67. (Cited on page 25).

[7] Bauschke, H. H., Borwein, J. M. and Combettes, P. L.: Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Comm. Contemp. Math.* **3** (2001), 615–647. (Cited on pages 24, 25, 49, 50, 106, and 123).

[8] Bauschke, H. H., Borwein, J. M. and Combettes, P. L.: Bregman monotone optimization algorithms, *SIAM J. Control Optim.* **42** (2003), 596–636. (Cited on pages 5, 9, 24, 56, 107, 109, and 111).

[9] Bauschke, H. H. and Combettes, P. L.: A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.* **26** (2001), 248–264. (Cited on pages 89, 95, and 128).

[10] Bauschke, H. H. and Combettes, P. L.: Construction of best Bregman approximations in reflexive Banach spaces, *Proc. Amer. Math. Soc.* **131** (2003), 3757–3766. (Cited on pages 89, 94, and 118).

[11] Bauschke, H. H. and Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, *Springer*, New York, 2011. (Cited on pages 5, 42, and 43).

[12] Bauschke, H. H., Combettes, P. L. and Luke, D. R.: A strongly convergent reflection method for finding the projection onto the intersection of two closed convex sets in a Hilbert space, *J. Approx. Theory* **141** (2006), 63-–69. (Cited on page 44).

[13] Bauschke, H. H., Combettes, P. L. and Reich, S.: The asymptotic behavior of the composition of two resolvents, *Nonlinear Anal.* **60** (2005), 283–301. (Cited on pages 108 and 131).

[14] Bauschke, H. H., Matoušková, E. and Reich, S.: Projection and proximal point methods: convergence results and counterexamples, *Nonlinear Anal.* **56** (2004), 715–738. (Cited on pages 89, 108, and 117).

[15] Bauschke, H. H., Wang, X. and Yao, L.: General resolvents for monotone operators: characterization and extension, *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, Medical Physics Publishing, Madison, WI, USA, 2010, 57–74. (Cited on pages 62, 81, 112, and 116).

[16] Beck, A.: Convergence Rate Analysis of Gradient Based Algorithms, School of Mathematical Sciences, *Tel Aviv Univerity*, 2003. (Cited on page 167).

[17] Beck, A. and Sabach, S.: A first order method for finding minimal norm-like solutions of convex optimization problems, submitted for publication 2011 (18 pp.). (Cited on page 11).

[18] Ben-Israel, A., Ben-Tal, A. and Zlobec, S.: Optimality in convex programming: a feasible directions approach, *Math. Programming Stud.* **19** (1982), 16–38. (Cited on page 164).

[19] Ben-Tal, A. and Nemirovski, A. S.: Lectures on Modern Convex Optimization, *MPS-SIAM Series on Optimization*, Philadelpia, 2001. (Cited on page 166).

[20] Bertsekas, D. P.: Nonlinear Programming, 2nd Edition, *Athena Scientific*, Belmont MA, 1999. (Cited on pages 165 and 169).

[21] Blum, E. and Oettli, W.: From optimization and variational inequalities to equilibrium problems, *Math. Student* **63** (1994), 123–145. (Cited on pages 137 and 138).

[22] Bonnans, J. F. and Shapiro, A.: Perturbation Analysis of Optimization Problems, *Springer*, New York, 2000. (Cited on page 20).

[23] Borwein, J. M., Guirao, A. J., Hájek, P. and Vanderwerff, J.: Uniformly convex functions on Banach spaces, *Proc. Amer. Math. Soc.* **137** (2009), 1081–1091. (Cited on page 28).

[24] Borwein, J. M. and Lewis, A. S.: Convex Analysis and Nonlinear Optimization. Theory and Examples, *CMS-Springer*, New York. Second extended edition, 2005. Paperback, 2009. (Cited on pages 32 and 113).

[25] Borwein, J. M., Reich, S. and Sabach, S.: A characterization of Bregman firmly nonexpansive operators using a new monotonicity concept, *J. Nonlinear Convex Anal.* **12** (2011), 161–184. (Cited on pages 32, 33, 35, 82, 83, 84, 85, and 86).

[26] Borwein, J. M. and Vanderwerff, J.: Constructions of uniformly convex functions, *Canadian Math. Bull.*, DOI:10.4153/CMB-2011-049-2, 2011. (Cited on pages 28, 32, and 113).

[27] Bregman, L. M.: A relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. and Math. Phys.* **7** (1967), 200–217. (Cited on page 24).

[28] Browder, F. E.: Nonexpansive nonlinear operators in Banach space, *Proc. Nat. Acad. Sci. USA* **54** (1965), 1011–1014. (Cited on page 54).

[29] Browder, F. E.: Existence and approximation of solutions of nonlinear variational inequalities, *Proc. Nat. Acad. Sci. U.S.A* **56** (1966), 1080–1086. (Cited on page 8).

[30] Browder, F. E.: Convergence of approximants to fixed points of non-expansive non-linear maps in Banach spaces, *Arch. Rational. Mech. Anal.* **24** (1967), 82–90. (Cited on page 105).

[31] Browder, F. E.: Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces, *Proceedings of Symposia in Pure Mathematics*, Vol. XVIII, Part 2, American Mathematical Society, Rhode Island, 1976. No cited.

[32] Bruck, R. E. and Reich, S.: Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston J. Math.* **3** (1977), 459–470. (Cited on pages 58, 108, and 131).

[33] Butnariu, D., Censor, Y. and Reich, S.: Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Comput. Optim. Appl.* **8** (1997), 21–39. (Cited on pages 24, 26, 27, 32, and 33).

[34] Butnariu, D. and Iusem, A. N.: Local moduli of convexity and their application to finding almost common fixed points of measurable families of operators, *Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995)*, Contemp. Math. 204, American Mathematical Society, Providence, RI, 1997, 61–91. (Cited on page 28).

[35] Butnariu, D. and Iusem, A. N.: Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization, *Kluwer Academic Publishers*, Dordrecht, 2000. (Cited on pages 14, 15, 16, 17, 24, 26, 27, 28, 29, 32, 37, and 47).

[36] Butnariu, D., Iusem, A. N. and Resmerita, E.: Total convexity for powers of the norm in uniformly convex Banach spaces, *J. Convex Anal.* **7** (2000), 319–334. (Cited on pages 29, 37, 38, and 42).

[37] Butnariu, D., Iusem, A. N. and Zălinescu, C.: On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces, *J. Convex Anal.* **10** (2003), 35–61. (Cited on pages 27, 28, and 108).

[38] Butnariu, D. and Kassay, G.: A proximal-projection method for finding zeroes of set-valued operators, *SIAM J. Control Optim.* **47** (2008), 2096–2136. (Cited on pages 64, 65, 108, and 148).

[39] Butnariu, D., Reich S. and Sabach, S.: A strong convergence theorem for resolvents of monotone operators, *J. Convex Anal.* **17** (2010), 991–1006. (Cited on page 108).

[40] Butnariu, D., Reich S. and Zaslavski, A.: There are many totally convex functions, *J. Convex Anal.* **13** (2006), 623–632. (Cited on page 31).

[41] Butnariu, D. and Resmerita, E.: Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces, *Abstr. Appl. Anal.* **2006** (2006), Art. ID 84919, 1–39. (Cited on pages 9, 24, 29, 39, 40, 107, and 108).

[42] Butnariu, D. and Resmerita, E.: Mosco stability of proximal operators in reflexive Banach spaces, *J. Nonlinear Convex Anal.* **8** (2007), 1–10. (Cited on page 108).

[43] Calvert, B.: The range of T-monotone operators, *Boll. Un. Math. Ital.* **4** (1971), 132–144. (Cited on page 64).

[44] Censor, Y., Gibali, A., Reich S. and Sabach, S.: The common variational inequality point problem, *Set-Valued and Variational Analysis*, **20** (2012), 229–247. (Cited on page 156).

[45] Censor, Y. and Lent, A.: An iterative row-action method for interval convex programming, *J. Optim. Theory Appl.* **34** (1981), 321–353. (Cited on pages 24 and 37).

[46] Censor, Y. and Reich, S.: Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization, *Optimization* **37** (1996), 323–339. (Cited on pages 9, 57, 59, 91, and 107).

[47] Cioranescu, I.: Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, *Kluwer Academic Publishers*, Netherlands, 1990. (Cited on pages 4, 21, 23, and 62).

[48] Combettes, P. L.: Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.* **38** (2000), 538–565. (Cited on page 89).

[49] Combettes, P. L. and Hirstoaga, S. A.: Equilibrium programming in Hilbert spaces, *J. Nonlinear Convex Anal.* **6** (2005), 117–136. (Cited on page 137).

[50] Chen, G. and Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM J. Optim.* **3** (1993), 538–543. (Cited on page 24).

[51] Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with application to convex programming, *Math. Oper. Res.* **18** (1993), 202–226. (Cited on pages 9, 107, and 108).

[52] Facchinei, F. and Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. 1 and 2, *Springer*, New York, 2003. (Cited on page 146).

[53] Ferris, M. C. and Mangasarian, O. L.: Finite perturbation of convex programs, *Appl. Math. Optim.* **23** (1991), 263–273. (Cited on page 162).

[54] Gárciga Otero, R. and Svaiter, B. F.: A strongly convergent hybrid proximal method in Banach spaces, *J. Math. Anal. Appl.* **289** (2004), 700–711. (Cited on pages 89, 118, and 123).

[55] Genel, A. and Lindenstrauss, J.: An example concerning fixed points, *Israel J. Math.* **22** (1975), 81–86. (Cited on pages 89 and 117).

[56] Goebel, K. and Reich, S.: Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings, *Marcel Dekker*, New York, 1984. (Cited on pages 4, 54, and 90).

[57] Güler, O.: On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.* **29** (1991), 403–419. (Cited on page 117).

[58] Halpern, B.: Fixed points of non-expanding maps, *Bull. Amer. Math. Soc.* **73** (1967), 957–961. (Cited on pages 89 and 105).

[59] Hartman, P. and Stampacchia, G.: On some non-linear elliptic diferential-functional equations, *Acta Mathematica* **115** (1966), 271–310. (Cited on page 146).

[60] Haugazeau, Y.: Sur les inéquations variationnelles et la minimisation de fonctionnelles convexes, Thèse, Université de Paris, Paris, France, 1968. (Cited on page 89).

[61] Iusem, A. N., Kassay, G. and W. Sosa.: On certain conditions for the existence of solutions of equilibrium problems, *Math. Program.* **116** (2009), 259–273. (Cited on page 137).

[62] Iusem, A. N. and Nasri, M. : Inexact proximal point methods for equilibrium problems in Banach spaces, *Numer. Funct. Anal. Optim.* **28** (2007), 1279–1308. (Cited on page 137).

[63] Iusem, A. N. and W. Sosa.: Iterative algorithms for equilibrium problems, *Optimization* **52** (2003), 301–316. (Cited on page 137).

[64] Kamimura, S. and Takahashi, W.: Strong convergence of a proximal-type algorithm in a Banach space, *SIAM J. Optim.* **13** (2002), 938–945. (Cited on pages 118 and 123).

[65] Kanzow, C., Qi, H. and Qi, L.: On the minimum norm solution of linear programs, *J. Optim. Theory Appl.* **116** (2003), 333–345. (Cited on page 163).

[66] Kassay, G., Reich, S. and Sabach, S.: Iterative methods for solving systems of variational inequalities in reflexive Banach spaces, *SIAM J. Optim.* **21** (2011), 1319–1344. (Cited on pages 65, 146, 148, 149, 150, and 151).

[67] Kinderlehrer, D. and Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications, *Academic Press*, New York, 1980. (Cited on page 146).

[68] Klir, G. J. and Folger, T. A.: Fuzzy Sets, Uncertainty, and Information, *Prentice Hall*, Englewood Cliffs, NJ, 1988. (Cited on page 32).

[69] Kohsaka, F. and Takahashi, W.: Fixed point theorems for a class of nonlinear mappings related to maximal monotone operators in Banach spaces, *Archiv der Mathematik* **21** (2008), 166–177. (Cited on pages 71 and 106).

[70] Krasnoselski, M. A.: Two observations about the method of successive approximations, *Uspehi Math. Nauk.* **10** (1955),123–127. (Cited on page 88).

[71] Mangasarian, O. L. and Meyer, R. R.: Nonlinear perturbation of linear programs, *SIAM J. Control and Optim.* **17** (1979), 745-–757. (Cited on page 162).

[72] Mann, W. R.: Mean value methods in iteration, *Proc. Amer. Math. Soc.* **4** (1953), 506–510. (Cited on pages 88 and 89).

[73] Markowitz, H.: Portfolio selection, *J. Finance* **7** (1952), 77-–91. (Cited on page 177).

[74] Martín-Márquez, V,. Reich, S. and Sabach, S.: Iterative methods for approximating fixed points of Bregman nonexpansive operators, *Discrete and Continuous Dynamical Systems*, accepted for publication. (Cited on pages 44, 45, 46, 48, 51, 75, 76, 77, 79, 80, 89, and 92).

[75] Martinet, B.: Régularisation d'inéquations variationelles par approximations successives, *Revue Française d'Informatique et de Recherche Opérationelle* **4** (1970), 154–159. (Cited on page 117).

[76] Minty, G. J.: Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.* **29** (1962), 341–346. (Cited on page 108).

[77] Minty, G. J.: On the monotonicity of the gradient of a convex function, *Pacific J. Math.* **14** (1964), 243–247. (Cited on page 4).

[78] Moreau, J. J.: Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France* **93** (1965), 273–299. (Cited on page 108).

[79] Nemirovski, A. S. and Yudin, D. B.: Problem Complexity and Method Efficiency in Optimization, *John Wiley & Sons Inc.*, New York, 1983. (Cited on page 166).

[80] Nesterov, Y.: Introductory Lectures on Convex Optimization, *Kluwer*, Boston, 2004. (Cited on pages 165, 166, and 170).

[81] Nevanlinna, O. and Reich, S.: Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces, *Israel J. Math.* **32** (1979), 44–58. (Cited on page 108).

[82] Pascali, D. and Sburlan, S.: Nonlinear Mappings of Monotone Type, *Sijthoff&Nordhoff International Publishers*, Alphen aan den Rijn, 1978. (Cited on pages 14, 62, and 63).

[83] Phelps, R. R.: Convex Functions, Monotone Operators, and Differentiability, Lecture Notes in Mathematics **1364**, 2nd Edition, *Springer*, Berlin, 1993. (Cited on pages 17 and 18).

[84] Reich, S.: Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.* **67** (1979), 274–276. (Cited on pages 71 and 93).

[85] Reich, S.: Strong convergence theorems for resolvents of accretive operators in Banach spaces, *J. Math. Anal. Appl.* **75** (1980), 287–292. (Cited on page 105).

[86] Reich, S.: On the asymptotic behavior of nonlinear semigroups and the range of accretive operators II, *J. Math. Anal. Appl.* **87** (1982), 134–146. (Cited on page 23).

[87] Reich, S.: Book review of "Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems" by Ioana Cioranescu, *Bull. Amer. Math. Soc.* **26** (1992), 367-370. (Cited on page 21).

[88] Reich, S.: A weak convergence theorem for the alternating method with Bregman distances, *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, Marcel Dekker, New York, 1996, 313-318. (Cited on pages 6, 9, 56, 57, 72, 90, 107, and 131).

[89] Reich, S. and Sabach, S.: A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces, *J. Nonlinear Convex Anal.* **10** (2009), 471–485. (Cited on pages 18, 108, 119, 122, 123, and 124).

[90] Reich, S. and Sabach, S.: Two strong convergence theorems for a proximal method in reflexive Banach spaces, *Numer. Funct. Anal. Optim.* **31** (2010), 22–44. (Cited on pages 48, 51, and 130).

[91] Reich, S. and Sabach, S.: Existence and approximation of fixed points of Bregman firmly nonexpansive mappings in reflexive Banach spaces, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, 2010, 299–314. (Cited on pages 6, 19, 66, 67, 69, 71, and 103).

[92] Reich, S. and Sabach, S.: Two strong convergence theorems for Bregman strongly nonexpansive operators in reflexive Banach spaces, *Nonlinear Analysis* **73** (2010), 122–135. (Cited on pages 102, 137, 138, and 139).

[93] Reich, S. and Sabach, S.: A projection method for Bregman firmly nonexpansive operators in reflexive Banach spaces, *J. Fixed Point Theory Appl.* **9** (2011), 101–116. (Cited on pages 8 and 137).

[94] Reich, S. and Sabach, S.: Three strong convergence theorems for iterative methods for solving equilibrium problems in reflexive Banach spaces, *Optimization Theory and Related Topics*, Contemporary Mathematics, Vol. 568, Amer. Math. Soc., Providence, RI, 2012, 225–240. (Cited on page 137).

[95] Reich, S. and Zaslavski, A. J.: Infinite products of resolvents of accretive operator, *Topol. Methods Nonlinear Anal.* **15** (2000), 153–168. (Cited on page 131).

[96] Resmerita, E.: On total convexity, Bregman projections and stability in Banach spaces, *J. Convex Anal.* **11** (2004), 1–16. (Cited on pages 29 and 47).

[97] Rockafellar, R. T.: Convex Analysis, *Princeton University Press*, Princeton 1970. (Cited on page 12).

[98] Rockafellar, R. T.: On the maximal monotonicity of subdifferential mappings, *Pacific J. Math.* **33** (1970), 209–216. (Cited on page 62).

[99] Rockafellar, R. T.: On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.* **149** (1970), 75–88. (Cited on page 155).

[100] Rockafellar, R. T.: Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14** (1976), 877–898. (Cited on pages 9, 81, 107, 108, and 117).

[101] Rockafellar, R. T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.* **1** (1976), 97–116. (Cited on pages 9 and 108).

[102] Sabach, S.: Resolvent and Applications, Department of Mathematics, *University of Haifa*, 2008. (Cited on pages 20 and 21).

[103] Sabach, S.: Products of finitely many resolvents of maximal monotone mappings in reflexive Banach spaces, *SIAM J. Optim.* **21** (2011), 1289–1308. (Cited on pages 9 and 131).

[104] Showalter, R. E.: Monotone Operators in Banach Space and Nonlinear Partial Differential Equations, Volume 49 of Mathematical Surveys and Monographs, *AMS*, Providence, 1997. (Cited on page 146).

[105] Solodov, M. V. and Svaiter, B. F.: Forcing strong convergence of proximal point iterations in a Hilbert space, *Math. Program.* **87** (2000), 189–202. (Cited on pages 89, 117, and 123).

[106] Takahashi, W. and Zembayashi, K.: Strong convergence theorem by a new hybrid method for equilibrium problems and relatively nonexpansive mappings, *Fixed point Theory Appl.* **2008** (2008), Art. ID 528476, 1–11. (Cited on page 137).

[107] Takahashi, W. and Zembayashi, K.: Strong and weak convergence theorems for equilibrium problems and relatively nonexpansive mappings in Banach spaces, *Nonlinear Anal.* **70** (2009), 45–57. (Cited on page 137).

[108] Teboulle, M.: Entropic Proximal Mappings with Applications to Nonlinear Programming, *Math. Oper. Res.* **17** (1992), 670–690. (Cited on pages 9 and 107).

[109] Tikhonov, A. N. and Arsenin V. Y., Solution of Ill-Posed Problems, *V. H. Winston*, Washington, 1977. (Cited on page 162).

[110] Vladimirov, A. A., Nesterov, Y. E. and Chekanov, Y. N.: On uniformly convex functionals (Russian), *Vestnik Moskovskaya Universiteta, Seria Matematika y Kibernetika* **3** (1978), 12–23. (Cited on page 28).

[111] Wei, L. and Zhou, H. Y.: Projection scheme for zero points of maximal monotone operators in Banach spaces, *J. Math. Res. Exposition* **28** (2008), 994–998. (Cited on page 129).

[112] Zălinescu, C.: Convex Analysis in General Vector Spaces, *World Scientific Publishing*, Singapore, 2002. (Cited on pages 21, 27, 28, and 29).

[113] Zeidler, E.: Nonlinear Functional Analysis and its Applications, Vol. II/B, *Springer*, Berlin, 1990. (Cited on pages 4 and 146).

משותף למספר סופי של ההעתקות מונוטוניות ובהתאמת האלגוריתם למצב שבו ישנן שגיאות חישוביות.

בפרק החמישי מובאים יישומיים של השיטות האיטרטיביות שפותחו בפרקים השלישי והרביעי לפתרון בעיות אופטימיזציה נוספות כגון: בעיות שיווי משקל, אי שיוויונים וריאציוניים ובעיית הבציעות הקמורה במרחבי בנך רפלקסיביים. אני מראה לגבי כל אחת מהבעיות כיצד לבנות אופרטורים מתאימים (כדוגמת הרזולונטה בבעיית השורשים) ובעלי התכונה שנקודת שבת שלהם היא פתרון לבעיה הנתונה ולהפך.

לצורך השימוש באופרטורים הללו באלגוריתמים שפותחו בפרק השלישי נדרש שהאופרטורים יקימו את ההנחות שמובאות במשפטי ההתכנסות. אני מוכיח שבכל אחד מן המקרים האופרטורים הנבנים אכן מקימים את התכונות ולכן השימוש בהם באלגוריתמים מוביל לשיטות איטרטיביות למציאת פתרון לבעיות אלו.

הפרק השישי והאחרון הוא הפרק היישומי של העבודה. האלגוריתמים שהוצגו בפרקים הקודמים הם בעלי מבנה מיוחד, הנותן לנו יכולת להתגבר על הבעיה המפורסמת של מציאת פתרונות לבעיות אופטימיזציה קמורה בעלי נורמה מינימאלית. בעיה זאת העסיקה חוקרים רבים בעשרות השנים האחרונות ונמצאו מספר שיטות לפתרון הבעיה. כל השיטות האלה מבוססות על שיטת הרגולריזציה של טיחונוב. בגלל קושי הבעיה השיטות הקיימות לפתרונה הן בעלות חסרונות גדולים. בפרק זה אני מציג לראשונה אלגוריתם מסדר ראשון הפותר את הבעיה הזו בצורה ישירה. אחד החסרונות של השיטות הקיימות הוא שבכל שלב בשיטה נדרש פתרון תת-בעיה, כלומר בכל שלב נדרשת הרצה של אלגוריתם נוסף, ובאלגוריתם המוצג בעבודה זו אין צורך בכך כלל.

מוכח משפט התכנסות עבור האלגוריתם וכמו כן מוכח קצב ההתכנסות של האלגוריתם. לבסוף מובאות תוצאות נומריות בהקשר של בעיית תיק ההשקעות.

בעיית תיק ההשקעות היא בעיה מרכזית בכלכלה והיא מאוד מסובכת. בהינתן סכום כסף להשקעה הבעיה היא לבנות תיק השקעות כך שהרווח ממנו יהיה כבקשתו של הלקוח, אבל הסיכון בו יהיה מינימאלי בהשוואה ליתר התיקים. בפועל בגלל ריבוי הנתונים נלקחים בחשבון נתוני העבר של המניות לאורך מספר מועט של שנים. עובדה זאת משפיעה על פתרון הבעיה ולמעשה יתכנו הרבה תיקי השקעות המקימים את מינימאליות הסיכון והרווח מהם הוא זהה. בפועל יועץ ההשקעות מוסיף אילוצים מדומים על הבעיה כדי לקבל תיק השקעות יחיד. האלגוריתם שלנו בא לפתור את הבעיה הזאת באופן שמבין כל תיקי ההשקעות הטובים (כלומר המקימים את בקשות הלקוח) הוא ימצא את הטוב ביותר ביחס להעדפה נוספת של הלקוח. כלומר בשלב זה יועץ ההשקעות מתייעץ עם הלקוח לצורך קבלת העדפה נוספת (למשל מעדיף להשקיע בזהב יותר מאשר בדולר) ומציאת התיק הספציפי שמקיים את כל בקשותיו של הלקוח והוא הטוב ביותר מבין כל התיקים הטובים ביחס להעדפתו האישית.

השיטה הראשונה, ובודאי השיטה הידועה ביותר לקרוב נקודות שבת, היא שיטת פיקרד. במקור שיטה זאת מיועדת לקרוב נקודות שבת של אופרטורים מכווצים במובן הצר ואף ידוע כי עבור אופרטורים מכווצים מסוימים השיטה אינה מתכנסת לנקודת שבת. כדי לקבל התכנסות נדרשת הנחה נוספת על האופרטור (פרט לעובדה שהוא מכווץ) ו/או הנחה נוספת על המרחב.
בגלל בעיתיות זו פותחה שיטה נוספת המכלילה את שיטת פיקרד ונקראת שיטת מאן. שיטה זאת מקרבת נקודות שבת של אופרטורים מכווצים אך באופן כללי מובטחת רק התכנסות חלשה לנקודת שבת ולא התכנסות חזקה.
בחלק השלישי של הפרק השלישי אני מראה כיצד ניתן לקבל התכנסות חזקה. בחלק זה ישנם שני אלגוריתמים הבונים סדרת נקודות המתכנסת חזק לנקודת שבת של האופרטור הנתון ובנוסף ידוע לנו בדיוק לאיזו נקודת שבת. נקודת השבת המתקבלת על ידי הסדרה היא הקרובה ביותר לנקודת ההתחלה ביחס למרחק ברגמן. בנוסף אני מוכיח שהכללת התוצאות למקרה של מספר סופי של אופרטורים אפשרית ואז מובטחת התכנסות חזקה לנקודת שבת משותפת של כל האופרטורים.
בזמן הרצת האלגוריתמים מופיעים אי דיוקים חישוביים המשפיעים על התנהגותו של האלגוריתם. לשם כך שיפרתי את האלגוריתמים כך שיוכלו לטפל בשגיאות החישוביות הללו. עוד הראיתי שתחת הנחות מסוימות על השגיאות האלגוריתם עדיין מתכנס לנקודת שבת.
לבסוף אני מוכיח התכנסות שיטה סתומה לקרוב נקודות שבת של אופרטורים מכווצים מסוימים ביחס למרחקי ברגמן.
בפרק הרביעי אני מתרכז בשיטות איטרטיביות למציאת שורשים של העתקות מונוטוניות. בעיית השורשים הינה בעיה מרכזית באופטימיזציה מפני שבהינתן פונקצית מטרה נאותה, קמורה ורציפה מלרע למחצה (לא גזירה בהכרח) אנו יודעים כי העתקת תת-הדיפרנציאל שלה היא העתקה מונוטונית. כמו כן ידוע כי כל נקודת מינימום של הפונקציה היא שורש של תת-הדיפרנציאל ולהפך.
אחת הדרכים הנפוצות ביותר למציאת שורשים של אופרטורים מונוטוניים במרחבי הילברט היא לבנות מההעתקה המונוטונית אופרטור שנקרא רזולוונטה וחשיבותו היא בכך שכל נקודת שבת שלו היא בהכרח שורש של ההעתקה ולהפך. בדרך זו אנו ממירים בעיית שורשים לבעיית נקודת שבת.
כאשר מנסים להשתמש בשיטה זאת במרחבי בנך נתקלים במספר בעיות, בין היתר הרזולוונטה הקלאסית איננה מתאימה ומאבדת את כל חשיבותה. מאחר והמעבר לבעיית נקודת שבת הוא מאוד נפוץ - נדרשת רזולונטה מוכללת. בשנת 2003 אכן נמצאה הרזולוונטה הזו והוכחו כל התכונות הנדרשות להכללת האלגוריתמים הידועים במרחבי הילברט למרחבי בנך.
את הפרק הרביעי אני מתחיל בהצגת תכונות ידועות של הרזולוונטה הכללית ולאחר מכן אני מציג דוגמאות לרזולוונטות במרחבים אוקלידיים, במרחבי הילברט ובמרחבי בנך. בשני החלקים האחרונים של הפרק מובאים ארבעה אלגוריתמים למציאת שורשים של העתקות מונוטוניות במרחבי בנך רפלקסיביים. בדומה לפרק השלישי אני דן בשתי הכללות: מציאת שורש

תקציר

תזה מחקרית זאת מתמקדת בשיטות איטרטיביות לפתרון מגוון רחב של בעיות אופטימיזציה במרחבים בעלי מימד אינסופי וגם במרחבים אוקלידיים.
בתזה ישנם שישה פרקים המסכמים את התרומות שלי לתחומים הבאים: אופטימיזציה, אנליזה לא ליניארית ושיטות נומריות. התרומות שלי לתחומים אלו נפרשות על המנעד שבין שיטות ישימות לפתרון בעיות אופטימיזציה המגיעות מהעולם המדעי לבין אלגוריתמים איטרטיביים לקרוב פתרונות של בעיות אופטימיזציה במרחבים בעלי מימד אינסופי.
בפרק הראשון ניתן למצוא את כל המושגים הבסיסים הנדרשים להבנת התוצאות המובאות בפרקים שלאחר מכן. פרק זה נועד לנוחות הקורא ולמען שלמות העבודה. בהתחלה, מובאות הגדרות ותוצאות קלאסיות באנליזה קמורה הקשורות לפונקציות המוגדרות ממרחב בנך (בעל מימד אינסופי) רפלקסיבי לישר הממשי המורחב (הערך פלוס אינסוף יכול להתקבל). לאחר מכן המושג המרכזי ביותר בתזה - מרחקי ברגמן – מוצג בצורה מפורטת. בנוסף, בחלק זה מוכחות מספר תוצאות בסיסיות, הקשורות למרחקים אלו, אשר תהווינה בסיס לתוצאות המרכזיות.
בשני החלקים האחרונים של הפרק הראשון מרוכזים כל המושגים הנדרשים לגבי אופרטורים ממרחב בנך רפלקסיבי לעצמו וכמו כן העתקות ממרחב בנך למרחב תת הקבוצות של המרחב הצמוד.
הפרק השני מוקדש לתוצאות חדשות בתחום של תורת נקודות השבת. במשך שישים השנה האחרונות תורת נקודות השבת תופסת חלק מרכזי באנליזה פונקציונאלית והיא מהווה בסיס חשוב למספר תיאוריות.
בעבודה זאת אני מוכיח לראשונה תוצאות עבור נקודות שבת של אופרטורים מכווצים המוגדרים ביחס למרחקי ברגמן. אופרטורים מכווצים, המוגדרים ביחס לנורמה, הם מאוד ידועים ובעלי חשיבות רבה. לפני כעשור התחילו להשתמש באופרטורים מכווצים המוגדרים ביחס למרחק ברגמן ולא ביחס לנורמה. אופרטורים אלו הם כלליים יותר ובשנים האחרונות השימוש בהם מתרחב מאוד לתחומי הלמידה הממוחשבת, עיבוד תמונה ועוד.
אני אציג בפרק זה שני חלקים כאשר בראשון אוכיח אפיון של אופרטורים מכווצים באופן הדוק המוגדרים ביחס למרחקי ברגמן. אפיון זה חשוב משום שהוא נותן לנו כלים לייצר דוגמאות רבות של אופרטורים מסוג זה במרחבים אוקלידיים. אני אציג בעיקר דוגמאות לאופרטורים כאלה כאשר הפונקציות המגדירות את מרחק ברגמן הן שתי אנטרופיות.
בחלק השני, אציג תוצאות לגבי תנאי הכרחי ותנאי מספיק לקיום נקודת שבת לאופרטורים מסוג זה וכמו כן אוכיח מספר תכונות של קבוצת נקודות השבת של אופרטורים כאלה.
הפרק השלישי הוא המשך ישיר של הפרק השני והוא מתמקד בשיטות איטרטיביות למציאת נקודות שבת של אופרטורים מכווצים המוגדרים ביחס למרחק ברגמן. אני אציג שלוש שיטות מפורשות לקרוב נקודות שבת.

עבודת המחקר נעשתה בהנחיית פרופסור
שמעון רייך בפקולטה למתמטיקה

אני מודה לטכניון על התמיכה הכספית
הנדיבה בהשתלמותי

שיטות איטרטיביות לפתרון בעיות אופטימיזציה

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר
דוקטור לפילוסופיה

שהם סבאח

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
אייר תשע"ב   חיפה   מאי 2012