Chapter 17
Minimizing the Moreau Envelope of Nonsmooth Convex Functions over the Fixed Point Set of Certain Quasi-Nonexpansive Mappings

Isao Yamada, Masahiro Yukawa and Masao Yamagishi

Summary: The first aim of this paper is to present a useful toolbox of quasi-nonexpansive mappings for convex optimization from the viewpoint of using their fixed point sets as constraints. Many convex optimization problems have been solved through elegant translations into fixed point problems. The underlying principle is to operate a certain quasi-nonexpansive mapping T iteratively and generate a convergent sequence to its fixed point. However, such a mapping often has infinitely many fixed points, meaning that a selection from the fixed point set $\mathrm{Fix}(T)$ should be of great importance. Nevertheless, most fixed point methods can only return an "unspecified" point from the fixed point set, which requires many iterations. Therefore, based on common sense, it seems unrealistic to wish for an "optimal" one from the fixed point set. Fortunately, considering the collection of quasi-nonexpansive mappings as a toolbox, we can accomplish this challenging mission simply by the *hybrid steepest descent method*, provided that the cost function is smooth and its derivative is Lipschitz continuous. A question arises: *how can we deal with "nonsmooth" cost functions* ?
The second aim is to propose a nontrivial integration of the ideas of the *hybrid steepest descent method* and the *Moreau-Yosida regularization*, yielding a useful approach to the challenging problem of nonsmooth convex optimization over $\mathrm{Fix}(T)$. The key is the use of smoothing of the original nonsmooth cost function by its *Moreau-Yosida regularization* whose derivative is always Lipschitz continuous. The field of application of hybrid steepest descent method can be extended to the mini-

---

Isao Yamada
Department of Communications and Integrated Systems, Tokyo Institute of Technology, S3-60, Tokyo, 152-8550 Japan, e-mail: isao@sp.ss.titech.ac.jp

Masahiro Yukawa
Department of Electrical and Electronic Engineering, Niigata University, 8050 Ikarashi Nino-cho, Nishi-ku, Niigata, 950-2181 Japan, e-mail: yukawa@eng.niigata-u.ac.jp

Masao Yamagishi
Department of Communications and Integrated Systems, Tokyo Institute of Technology, S3-60, Tokyo, 152-8550 Japan, e-mail: myamagi@sp.ss.titech.ac.jp

Pages 345 -- 390

mization of the ideal smooth approximation over $\mathrm{Fix}(T)$. We present the mathematical ideas of the proposed approach together with its application to a combinatorial optimization problem: the minimal antenna-subset selection problem under a highly nonlinear capacity-constraint for efficient MIMO (Multiple Input Multiple Output) communication systems.

Key words: Nonsmooth convex optimization, Moreau envelope, Hybrid steepest descent method
AMS 2010 Subject Classification: 47H10, 47H09, 49M20, 65K10

17.1 Introduction

How can we exploit various types of information efficiently in convex optimization? This has been one of the fundamental questions of paramount importance from both practical and theoretical viewpoints. We present a new insight into this question with (i) *fixed point characterizations* of constraint sets and (ii) the *Moreau-Yosida regularization* of a nonsmooth convex function. In order to contrast our contribution with existing approaches, let us briefly introduce a stream of research developments, including classical and state-of-the-art techniques, for treating (multiple) constraints.

17.1.1 Treatments of Constraints in Convex Optimization

A general convex optimization problem is formulated as follows: minimize a convex function $f \in \Gamma_0(\mathscr{H})$ over a closed convex subset C of a real Hilbert space $\mathscr{H}$. Here, $\Gamma_0(\mathscr{H})$ stands for the class of all lower semicontinuous convex functions from $\mathscr{H}$ to $(-\infty, \infty]$ which are not identically equal to $+\infty$. Suppose for instance that f is differentiable with its derivative Lipschitz continuous and P_C, the metric projection onto C (see Fact 17.2(c)), can be computed *efficiently*. In this special case, we may use Goldstein's *projected gradient method* [71]. However, this classical approach cannot satisfy the increasing demand for nonsmooth convex optimization under more general constraints.

A couple of unified approaches covering many existing schemes involve the following formulation [103, 89, 66, 57, 125, 46, 42]: minimize $f_1 + f_2$ for $f_i \in \Gamma_0(\mathscr{H})$, $i = 1, 2$. For example, under a certain *qualification condition* on f_1 and f_2, the *Douglas-Rachford splitting* type algorithm (see Examples 17.6(c) and 17.12(f))[89, 57, 42] approximates a minimizer of $f_1 + f_2$ with successive use of

$$\mathrm{prox}_{\gamma f_i} : \mathscr{H} \to \mathscr{H} : x \mapsto \arg\min_{y \in \mathscr{H}} \left\{ f_i(y) + \frac{1}{2\gamma} \|x - y\|^2 \right\}, \qquad (17.1)$$

which is well-defined as a single valued mapping called the *proximity operator* or *proximal mapping* [97, 98, 108, 46] of index $\gamma \in (0,\infty)$ of f_i $(i = 1,2)$ (see Section 17.2.1). This approach can handle the problem considered in the previous paragraph by letting $f_1 := f$ and $f_2 := i_C$ which denotes the indicator function

$$(\forall x \in \mathscr{H}) \quad i_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ \infty, & \text{otherwise.} \end{cases}$$

In fact, the proximity operator of i_C for any $\gamma \in (0,\infty)$ coincides with P_C. We emphasize however that the approach in [57, 125, 46, 42] practically requires an efficient scheme to compute the proximity operators, and obtaining such a scheme itself is often a challenging issue to address for each application individually.

This certainly motivates the recent active studies on computational schemes for proximity operators of various types of functions in $\Gamma_0(\mathscr{H})$ [46, 42, 62], which include the pre-composition of $g \in \Gamma_0(\mathscr{H})$ with a frame synthesis affine operator [31, 42, 62, 117]. Another development is the extension of the Douglas-Rachford splitting type scheme to the case of multiple convex functions [43, 44, 67]; i.e., minimize $\sum_{i=1}^m f_i$ for $m > 2$ and $f_i \in \Gamma_0(\mathscr{H})$ $(i = 1,2,\cdots,m)$, through the Pierra-type product-space reformulation [104, 105]. This extension enables us to deal with the case where a constraint set C can be expressed as the intersection of a finite number of closed convex sets C_i ($i \in \mathscr{I}$, *assuming that P_{C_i} can be computed efficiently*). Indeed, we can minimize a nonsmooth convex function $f := \sum_{j \in \mathscr{J}} f_j$ over C by applying the extended scheme to $\sum_{i \in \mathscr{I}} i_{C_i} + \sum_{j \in \mathscr{J}} f_j$. The use of the expression $C = \bigcap_{i \in \mathscr{I}} C_i$ shares similarity with the commonly used strategy in the simpler contexts of the convex feasibility problems (see, e.g., [18, 34, 7, 30, 48]). However, again, this approach has an obvious limitation, as there are many applications, including the one addressed in this work, in which the constraint set $C \subset \mathscr{H}$ can hardly be expressed as the intersection of (a finite number of) *simple* closed convex sets. The *fixed point characterization* throws us a rope to escape from the dilemma, as explained in the following.

17.1.2 Fixed Point Characterizations of Closed Convex Sets

A mapping $T : \mathscr{H} \to \mathscr{H}$ is called *quasi-nonexpansive* if this mapping has its nonempty fixed point set $\mathrm{Fix}(T) := \{x \in \mathscr{H} \mid T(x) = x\} \neq \emptyset$ and $\|T(x) - z\| \leq \|x - z\|$ $(\forall x \in \mathscr{H}, \forall z \in \mathrm{Fix}(T))$. In this case, the fixed point set $\mathrm{Fix}(T)$ is guaranteed to be closed convex in $\mathscr{H}$ (see Proposition 17.3 in Section 17.2.2). In the context of recent studies on the convex feasibility problems as well as the unified treatment of certain nonsmooth optimization schemes, many powerful ideas have been found to deal with a closed convex set C as *the fixed point set of an efficiently-computable quasi-nonexpansive mapping* [7, 35, 8, 131, 136, 135].

In the final manuscript, the following reference has been added into the square brackets underlined above:
Gandy, S., Recht, B., Yamada, I.: Tensor completion and low-n-rank tensor recovery via convex optimization. Inverse Probl. 27(2), 025010

For example, if the set $S := \arg\min_{x\in\mathscr{H}} \{f_1(x)+f_2(x)\}$ is nonempty in the above context of minimizing f_1+f_2 for $f_i \in \Gamma_0(\mathscr{H})$, $i=1,2$, the set S is a closed convex set which is usually hard to be expressed as the intersection of (a finite number of) simple closed convex sets. On the other hand, in a variety of scenarios, the set S can be expressed as the fixed point set of a nonexpansive mapping [46] or as the image of a proximity operator of the fixed point set of another nonexpansive mapping [42], where these nonexpansive mappings can be computed efficiently (see Section 17.2.2 for basic ideas to design a mapping that has a desirable fixed point set).

Another quite useful example is found in the characterization of the nonempty level set $\mathrm{lev}_{\leq 0}(g) := \{x \in \mathscr{H} \mid g(x) \leq 0\}$ of $g \in \Gamma_0(\mathscr{H})$ as the fixed point set of the subgradient projection $T_{\mathrm{sp}(g)}$ relative to g. The subgradient projection operator is *firmly quasi-nonexpansive* ([126, Lemma 2.8], [8, 135]) and has been playing important roles as a low complexity approximation of the metric projection onto $\mathrm{lev}_{\leq 0}(g)$ in many scenarios; e.g., in signal and image processing applications [139, 37, 41], the metric projection is often hard to compute (see Proposition 17.7 and Example 17.9 for designing better approximations than the subgradient projection). In [139, 134, 115, 112, 123], the subgradient projection was used to elude from the load for solving large scale systems of equations in an adaptive signal processing or adaptive online classification problems. In [41], the subgradient projection was used to suppress the *total variation* of the restored image.

The idea of dealing with a closed convex set as the fixed point set of a nonexpansive mapping has been applied successfully in creations of many powerful optimization schemes with the strong support of the innovative discovery of the *Mann iterative process* [93, 54, 74], which is an extremely simple algorithm to generate a (weakly) convergent sequence to a fixed point of a general nonexpansive mapping. Moreover recent notable extensions, e.g., [39], of the algorithm have a guarantee of convergence under much weaker conditions than those found in [93, 54, 74] and applied in the unifications, e.g., in [46, 42]. In short, these previous studies aim to find an *arbitrary* point in the fixed point set of a nonexpansive mapping. The next stage which we should clear is the following: find an *optimal* point in some sense in the fixed point set. The following subsection introduces some existing methods for this problem with a touch of motivation of the current study.

17.1.3 Existing Methods on the Advanced Stage

We now consider the problem of minimizing a convex function over the fixed point set of a certain quasi-nonexpansive mapping. There seems to be only few types of algorithms that can deal with this problem in a computationally manageable way. Among others, the *hybrid steepest descent method* (see, e.g., [137, 49, 138, 136, 100, 101, 129, 135, 120, 91, 83, 150, 33]) has been developed as an algorithm to achieve such a goal originally by extending a fixed point iteration [77, 88, 128, 36, 6]; the so-called *Halpern-type iteration* or *anchor method*, which is able to find from a given point the nearest fixed point of a nonexpansive mapping. The hybrid

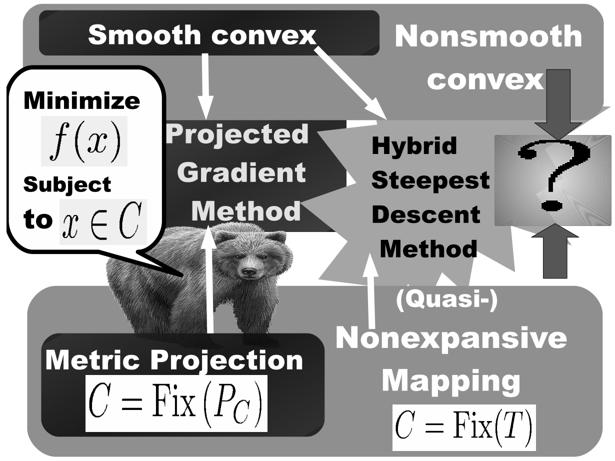


Fig. 17.1 Treatment of constraint sets as fixed point sets of nonlinear mappings.

steepest descent method has two distinguished features. First, it has a mathematical guarantee of convergence to the solution to the convex optimization over the fixed point set. Second, it only requires at each iteration simple computation of a gradient descent operator and a quasi-nonexpansive mapping, of which the fixed point set defines the constraint set of the optimization problem. Indeed, the method has been applied successfully to signal and image processing problems (see, e.g., [116, 78, 114, 120, 151, 117]).

By extending the ideas in [79], another algorithm, which we refer to as the *generalized Haugazeau's algorithm*, was developed for minimizing a *strictly convex* function in $\Gamma_0(\mathscr{H})$ over the fixed point set of a certain quasi-nonexpansive mapping [38]. In particular, this algorithm was specialized in a clear way for finding the nearest fixed point of a certain quasi-nonexpansive mapping [8] and applied successfully to an image recovery problem [41]. If we focus on the case of a nonstrictly convex function, the generalized Haugazeau's algorithm is not applicable, while some convergence theorems of the hybrid steepest descent method suggest its sound applicability *provided that the derivative of the function is Lipschitzian*. Due to the Lipschitz-continuity assumption, however, it still remains an open problem to minimize a *nonsmooth* convex function over the fixed point set of a quasi-nonexpansive mapping (see Fig.17.1).

17.1.4 Contributions of This Paper

So far we do not have in general any promising (computationally manageable) algorithm for the solution to the minimization problem of a *nonsmooth* convex function over the fixed point set of a quasi-nonexpansive mapping. We therefore present a nontrivial application of the hybrid steepest descent method to approach the problem. Our attention is to the notable fact that any function $f \in \Gamma_0(\mathscr{H})$ can be approximated with any accuracy by

$$ {}^{\gamma}f : \mathscr{H} \to \mathbb{R} : \quad x \mapsto \min_{y \in \mathscr{H}} \left\{ f(y) + \frac{1}{2\gamma}\|x-y\|^2 \right\} $$
$$ = f\left(\mathrm{prox}_{\gamma f}(x)\right) + \frac{1}{2\gamma}\left\|x - \mathrm{prox}_{\gamma f}(x)\right\|^2, \qquad (17.2) $$

which is called the *Moreau envelope* [1] (or the *Moreau-Yosida regularization*[2]) of index $\gamma \in (0,\infty)$ of f. The Moreau envelope ${}^{\gamma}f$ is a smooth approximation of f with surprisingly beautiful properties. In particular, the most attractive property for us is that the Moreau envelope ${}^{\gamma}f$ has a Lipschitz continuous gradient over $\mathscr{H}$ (see Section 17.3.1). Moreover, if $\arg\min_{x \in \mathscr{H}} f(x) \neq \emptyset$, the set of all global minimizers of f is equal to that of the Moreau envelope (see Fact 17.2). These distinctive features suggest that the Moreau-Yosida regularization and the proximity operator are the keys bridging the gap between the analyses of smooth and nonsmooth convex functions. For example, these features have been utilized to develop efficient algorithms specialized for unconstrained nonsmooth convex optimization problems (see, e.g., [107, 65]). In addition to this direct use, the practical value of the Moreau envelope has been examined implicitly or explicitly as a smooth relaxation of the absolute value function in many applications (see Section 17.3.1).

In this study, we propose to approach the nonsmooth optimization problem

$$ \text{minimize } f(x) \text{ subject to } x \in \mathrm{Fix}(T) \qquad (17.3) $$

by solving its smooth relaxation

$$ \text{minimize } {}^{\gamma}f(x) \text{ subject to } x \in \mathrm{Fix}(T) \qquad (17.4) $$

with the hybrid steepest descent method. Here, $f \in \Gamma_0(\mathscr{H})$ (which in particular we consider to be nonsmooth) and $T : \mathscr{H} \to \mathscr{H}$ is a quasi-nonexpansive mapping (Note: The solution sets for (17.3) and (17.4) are not the same in general although they coincide specially in the simplest unconstrained case, i.e., $\mathrm{Fix}(T) = \mathscr{H}$). Thanks to (i) the beautiful properties of the Moreau envelope and (ii) the flexibility in expressing a constraint set as the fixed point set of a quasi-nonexpansive mapping, the proposed approach enjoys wide applicability.

The rest of this paper is organized as follows. For readers' convenience, Section 17.2 presents a short tour in computational convex analysis which contains (i) elements of convex analysis, (ii) the fixed point theory of quasi-nonexpansive mapping including a basic algorithm to approximate a fixed point of the mapping, and (iii) elements of the variational inequality problems. It also introduces briefly one

[1] Nice introductions to the Moreau envelope are found, e.g., in [108, 46].

[2] As will be seen in (17.23), the derivative $\nabla^{\gamma} f$ is given as the *Yosida approximation* [141] of the subdifferential ∂f of f.

role of quasi-nonexpansive mapping in signal processing. In Section 17.3, we will introduce the essence of the Moreau-Yosida regularization and the hybrid steepest descent method. Then we will show how to join the two concepts to approach the minimization problem of a nonsmooth convex function over the fixed point set of certain quasi-nonexpansive mappings. In Section 17.4, we demonstrate the effectiveness of the proposed approach in its application to the minimal antenna-subset selection problem under a highly nonlinear capacity-constraint for efficient MIMO (Multiple Input Multiple Output) communication systems; the convex relaxation of the problem is the ℓ_1 norm minimization under the constraint. Finally, in Section 17.5, we conclude this paper with some remarks on other possible advanced applications of the hybrid steepest descent method.

17.2 A Short Tour in Computational Convex Analysis

17.2.1 Selected Elements of Convex Analysis

In the following, we list minimum notions in convex analysis, which are necessary for our discussion (see, e.g., [7, 46, 48, 59, 81, 108, 122, 147, 133, 10] for detailed account on these notions). Let $\mathscr{H}$ be a real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$.

Definition 17.1. (Basics in Convex Analysis)

(a) (Convex set) A set $C \subset \mathscr{H}$ is called convex if $\lambda x + (1-\lambda)y \in C$ for every $x, y \in C$ and every $\lambda \in [0,1]$. If a set $C \subset \mathscr{H}$ is closed as well as convex, it is called closed convex.

(b) (Convex function, Proper function) A function $f : \mathscr{H} \to (-\infty, \infty] := \mathbb{R} \cup \{\infty\}$ is called convex if

$$(\forall x, y \in \mathscr{H}, \forall \lambda \in (0,1)) \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \quad (17.5)$$

In particular, a convex function $f : \mathscr{H} \to (-\infty, \infty]$ is called proper if

$$\text{dom}(f) := \{x \in \mathscr{H} \mid f(x) < \infty\} \neq \emptyset.$$

A function $f \in \Gamma_0(\mathscr{H})$ is called strictly convex if

$$(x \neq y, \lambda \in (0,1)) \Rightarrow f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

(c) (Lower semicontinuous function) A function $f : \mathscr{H} \to (-\infty, \infty]$ is called lower semicontinuous if the set $\text{lev}_{\leq \alpha}(f) := \{x \in \mathscr{H} \mid f(x) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$ (Note: If f is continuous over $\mathscr{H}$, f is lower semicontinuous). The set of all proper lower semicontinuous convex functions is denoted by $\Gamma_0(\mathscr{H})$.

(d) (Coercivity) A function $f \in \Gamma_0(\mathscr{H})$ is called *coercive* if

$$\|x\| \to \infty \Rightarrow f(x) \to \infty.$$

In this case, the existence of a minimizer of f, i.e., $\{x^\star \in \mathscr{H} \mid f(x^\star) \le f(x) \quad (\forall x \in \mathscr{H})\} \neq \emptyset$, is guaranteed.

Fact 17.2. (Fundamental Tools for Convex Optimization)

(a) (Subgradient, Subdifferential, Legendre-Fenchel conjugate) Given $f \in \Gamma_0(\mathscr{H})$, the *subdifferential* of f at x is defined as the set of all *subgradients* of f at x:

$$\partial f(x) := \{u \in \mathscr{H} \mid \langle y - x, u\rangle + f(x) \le f(y), \forall y \in \mathscr{H}\}.$$

Therefore $0 \in \partial f(x) \Leftrightarrow f(x) = \min_{y\in\mathscr{H}} f(y)$. If f is continuous at $x \in \mathscr{H}$, $\partial f(x)$ is a nonempty closed convex set. Moreover, if f is Gâteaux differentiable[3] at x, the subdifferential at x is a singleton as $\partial f(x) = \{\nabla f(x)\}$ [81, 133, 10]. The subdifferential is regarded as a set-valued mapping $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$, which is called bounded if it maps bounded sets to bounded sets [14] (Note: $2^{\mathscr{H}}$ stands for the collection of all subsets of $\mathscr{H}$).
Remark that the subdifferential of f at $x \in \mathscr{H}$ can be defined alternatively as $\partial f(x) := \{u \in \mathscr{H} \mid f(x) + f^*(u) = \langle x, u\rangle\}$, where $f^* \in \Gamma_0(\mathscr{H})$ is defined by

$$(\forall u \in \mathscr{H}) \quad f^*(u) := \sup_{x\in\mathscr{H}} \{\langle x, u\rangle - f(x)\}$$

and it is called the *conjugate* (also named *Legendre-Fenchel conjugate*, or *Legendre-Fenchel transform*) of f.

(b) (Proximity operator) The proximity operator of index $\gamma \in (0,\infty)$ of $f \in \Gamma_0(\mathscr{H})$ is defined (as in (17.1)) by

$$\operatorname{prox}_{\gamma f} : \mathscr{H} \to \mathscr{H} : x \mapsto \arg\min_{y\in\mathscr{H}} \left\{ f(y) + \frac{1}{2\gamma}\|x-y\|^2 \right\}, \tag{17.6}$$

where the existence and the uniqueness of the minimizer are guaranteed respectively by the coercivity and the strict convexity of $f(\cdot) + \frac{1}{2\gamma}\|x - \cdot\|^2$. Equivalently, for every $x \in \mathscr{H}$, $\operatorname{prox}_{\gamma f}(x)$ is characterized as a unique point satisfying

---

[3] (Gâteaux and Fréchet derivatives of function) Let U be an open subset of $\mathscr{H}$. Then a function $f : U \to \mathbb{R}$ is called Gâteaux differentiable at $x \in U$ if there exists $a(x) \in \mathscr{H}$ such that $\lim_{\delta\to 0} \frac{f(x+\delta h) - f(x)}{\delta} = \langle a(x), h\rangle$ $(\forall h \in \mathscr{H})$. In this case, $\nabla f(x) := a(x)$ is called Gâteaux derivative (or gradient) of f at x.

On the other hand, a function $f : U \to \mathbb{R}$ is called *Fréchet* differentiable over U if for each $u \in U$ there exists $a(u) \in \mathscr{H}$ such that

$$f(u+h) = f(u) + \langle a(u), h\rangle + o(\|h\|) \text{ for all } h \in \mathscr{H},$$

where $r(h) = o(\|h\|)$ means $\lim_{h\to 0} r(h)/\|h\| = 0$. In this case, $\nabla f : U \to \mathscr{H}$ defined by $\nabla f(u) = a(u)$ is called Fréchet derivative of f over U. If f is Fréchet differentiable over U, f is also Gâteaux differentiable over U and both derivatives coincide. Moreover, if f is Gâteaux differentiable with continuous derivative ∇f over U, then f is also Fréchet differentiable over U.

$$\{\mathrm{prox}_{\gamma f}(x)\} = \{z \in \mathscr{H} \mid z + \gamma \partial f(z) \ni x\}, \tag{17.7}$$

i.e.,

$$\mathrm{prox}_{\gamma f}(x) = (I + \gamma \partial f)^{-1}(x), \tag{17.8}$$

which is again equivalent to

$$(\forall y \in \mathscr{H}) \quad \left\langle y - \mathrm{prox}_{\gamma f}(x), \frac{x - \mathrm{prox}_{\gamma f}(x)}{\gamma} \right\rangle + f(\mathrm{prox}_{\gamma f}(x)) \le f(y).$$

The proximity operator is firmly nonexpansive, i.e., $\mathrm{rprox}_{\gamma f} := 2\mathrm{prox}_{\gamma f} - I \colon \mathscr{H} \to \mathscr{H}$ is nonexpansive (see Section 17.2.2 for the definition of nonexpansivity of a mapping):

$$(\forall x, y \in \mathscr{H}) \quad \|(2\mathrm{prox}_{\gamma f} - I)x - (2\mathrm{prox}_{\gamma f} - I)y\| \le \|x - y\|.$$

Moreover, if $\arg\min_{x \in \mathscr{H}} f(x) \neq \emptyset$, the set of all minimizers of f is equal to that of the Moreau envelope and also expressed as the fixed point set of $\mathrm{prox}_{\gamma f} : \mathscr{H} \to \mathscr{H}$; i.e.,

$$\arg\min_{x \in \mathscr{H}} f(x) = \arg\min_{x \in \mathscr{H}} {}^{\gamma}f(x) = \mathrm{Fix}\left(\mathrm{prox}_{\gamma f}\right).$$

(c) (Metric projection onto closed convex sets) Given a nonempty closed convex set $C \subset \mathscr{H}$ and any point $x \in \mathscr{H}$, there exists a unique point $P_C(x) \in C$ satisfying

$$d_C(x) := \min_{z \in C} \|x - z\| = \|x - P_C(x)\|.$$

The mapping $\mathscr{H} \ni x \mapsto P_C(x) \in C$ is called the metric projection (or convex projection) onto C and obviously $P_C(x) = \mathrm{prox}_{\gamma \iota_C}(x)$ $(\forall \gamma \in (0,\infty), \forall x \in \mathscr{H})$, hence P_C is firmly nonexpansive with $\mathrm{Fix}(P_C) = C \neq \emptyset$ (see Example 17.6(a) and Fig.17.2). Moreover $P_C : \mathscr{H} \to C$ is characterized by

$$x^\star \in C \text{ satisfies } \langle x - x^\star, z - x^\star \rangle \le 0 \ (\forall z \in C) \quad \Leftrightarrow \quad x^\star \in C \text{ satisfies } x^\star = P_C(x). \tag{17.9}$$

(d) (Expression of a closed convex set I) Given (possibly infinitely many) closed convex sets $C_i \subset \mathscr{H}$ ($i \in \mathscr{I}$: an index set), their intersection $\bigcap_{i \in \mathscr{I}} C_i$ is again a closed convex set (Note: This property is a natural nonlinear generalization of the elementary fact that the intersection of multiple subspaces is again a subspace in a vector space).

(e) (Expression of a closed convex set II) Given a function $f \in \Gamma_0(\mathscr{H})$, the set $\mathrm{lev}_{\le 0}(f)$, which is called the (zero-)level set of f, is closed convex. Conversely, given a closed convex set $C \subset \mathscr{H}$, there exists a continuous convex function $f : \mathscr{H} \to \mathbb{R}$ satisfying $C = \mathrm{lev}_{\le 0}(f)$. The function $d_C : \mathscr{H} \to [0,\infty)$ in (c) is obviously such an example.

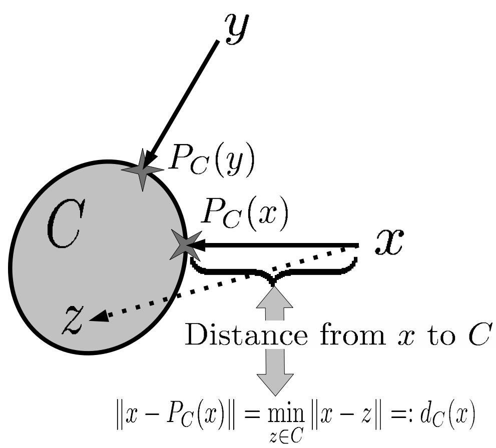


Fig. 17.2 Convex Projection: Metric projection onto a closed convex set C.

17.2.2 Quasi-Nonexpansive Mappings and Their Fixed Point Sets

Suppose that a mapping $T : \mathscr{H} \to \mathscr{H}$ has at least one fixed point. Then the mapping $T : \mathscr{H} \to \mathscr{H}$ is called quasi-nonexpansive (or Fejér) [7, 8, 54, 126] if T satisfies for every $x \in \mathscr{H}$ and every $z \in \mathrm{Fix}(T)$

$$\|T(x) - z\| \leq \|x - z\|. \tag{17.10}$$

The identity operator $I : \mathscr{H} \to \mathscr{H}$ is also a quasi-nonexpansive mapping which satisfies of course $\mathrm{Fix}(I) = \mathscr{H}$.

We introduce special subclasses of quasi-nonexpansive mappings below (see also Fig.17.3). A quasi-nonexpansive mapping T is said to be attracting if T satisfies for every $x \notin \mathrm{Fix}(T)$ and every $z \in \mathrm{Fix}(T)$

$$\|T(x) - z\| < \|x - z\|.$$

In particular, an attracting mapping T is called α-strongly attracting if there exists some $\alpha > 0$ satisfying for every $x \in \mathscr{H}$ and every $z \in \mathrm{Fix}(T)$

$$\alpha\|x - T(x)\|^2 \leq \|x - z\|^2 - \|T(x) - z\|^2.$$

The above inequality offers a lower bound for improvement by T of approximation accuracy of a point x to all fixed points z of T.

A quasi-nonexpansive mapping $T : \mathscr{H} \to \mathscr{H}$ is said to be α-averaged [3, 7] if there exist some $\alpha \in (0, 1)$ and some quasi-nonexpansive mapping N such that $T = (1 - \alpha)I + \alpha N$. In this case, T satisfies an obvious relation $\mathrm{Fix}(T) = \mathrm{Fix}(N)$. Moreover T is strongly attracting (see Proposition 17.3(b) below). In particular, if

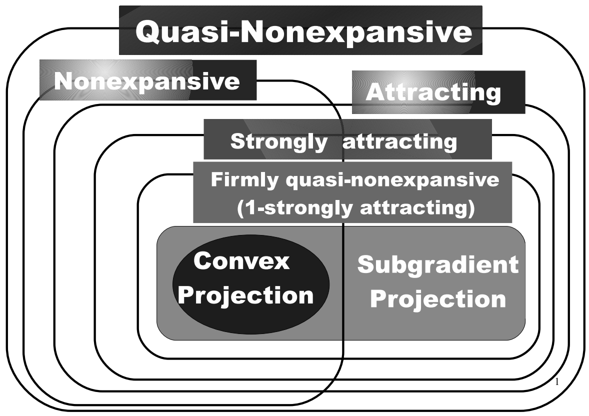


Fig. 17.3 Quasi-nonexpansive mapping and its subclasses (A nonexpansive mapping is also quasi-nonexpansive if this mapping has at least one fixed point).

T is $\frac{1}{2}$-averaged, T is called a firmly quasi-nonexpansive mapping [135] (the class of firmly quasi-nonexpansive mappings is specially denoted by $\mathfrak{T}$ [8]).

On the other hand, a mapping $T : \mathscr{H} \to \mathscr{H}$ is called Lipschitz continuous with a Lipschitz constant κ or shortly κ-Lipschitzian if there exists some $\kappa > 0$ satisfying for every $x, y \in \mathscr{H}$

$$\|T(x) - T(y)\| \leq \kappa \|x - y\|.$$

In particular, if there exists some $\kappa < 1$, T is called a contraction (or a strictly contractive) mapping. In this case, the Banach-Picard's contraction mapping theorem guarantees the unique existence of the fixed point of T, and it is not hard to see that T is α-averaged for any $\alpha \in \left[\frac{\kappa+1}{2}, 1\right)$. If the mapping T is 1-Lipschitzian, T is called a nonexpansive mapping [7, 69, 70, 122] and in this case, T is also quasi-nonexpansive if $\mathrm{Fix}(T) \neq \emptyset$. In contrast to the case of the existence of $\kappa < 1$, the existence of $\kappa = 1$ is insufficient to guarantee the existence of a fixed point in view of the following example: $T : \mathbb{R} \ni x \mapsto x + 1 \in \mathbb{R}$.

The following Proposition 17.3(a) guarantees that the closedness and convexity of the fixed point set of any quasi-nonexpansive mapping. This property is very fortunate to express a constraint set, in convex optimization, as the fixed point set of a quasi-nonexpansive mapping. For example, Proposition 17.3(a) together with Fact 17.2(d),(e) suggests that a closed convex set can be expressed as the intersection of possibly infinitely many simpler closed convex sets, each of which can be expressed as the fixed point set of an efficiently computable quasi-nonexpansive mapping. Moreover, by Proposition 17.3(b), given a quasi-nonexpansive mapping $N : \mathscr{H} \to \mathscr{H}$, we can construct a strongly attracting quasi-nonexpansive mapping $T := (1-\alpha)I + \alpha N$ ($\alpha \in (0,1)$) with $\mathrm{Fix}(T) = \mathrm{Fix}(N)$. Therefore the quasi-nonexpansive mapping (or even more specifically the attracting mapping) has a great deal of potential not only as an computational tool for monotone approximation to the closed convex set but also as an alternative mathematical expression of the closed convex set as its fixed point set.

Proposition 17.3. (Fundamental Properties of Quasi-Nonexpansive Mapping)

(a) Let $T : \mathscr{H} \to \mathscr{H}$ be a quasi-nonexpansive mapping. Then $\mathrm{Fix}(T)$ can be expressed as (see for example [8, 135]):

$$\mathrm{Fix}(T) = \bigcap_{y \in \mathscr{H}} \left\{ x \in \mathscr{H} \mid \langle y - T(y), x \rangle \leq \frac{\|y\|^2 - \|T(y)\|^2}{2} \right\}.$$

This tells us that $\mathrm{Fix}(T)$ can be expressed as the intersection of infinitely many *closed half spaces*, hence the closedness and convexity of $\mathrm{Fix}(T)$ are guaranteed by Fact 17.2(d).

(b) A quasi-nonexpansive mapping $T : \mathscr{H} \to \mathscr{H}$ is α-averaged for some $\alpha \in (0,1)$ if and only if T is $\left(\frac{1-\alpha}{\alpha}\right)$-strongly attracting [135]. Therefore a quasi-nonexpansive mapping T is $\frac{1}{2}$-averaged if and only if it is 1-strongly attracting.

In Proposition 17.4 below, (a) and (b) are slight refinement of similar results in [7, Props 2.10 & 2.12]. By applying these properties, we can construct a new quasi-nonexpansive mapping whose fixed point set is the intersection of the fixed point sets of given multiple quasi-nonexpansive mappings in Examples 17.6 and 17.9 in Section 17.2.3. Note that Proposition 17.4(c) holds even when $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2) = \emptyset$.

Proposition 17.4. (Algebraic Properties of Quasi-Nonexpansive Mapping)

(a) (Convex combination [135]) Suppose that $T_i : \mathscr{H} \to \mathscr{H}$ $(i = 1,2)$ are quasi-nonexpansive mappings satisfying $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2) \neq \emptyset$. Then for any $w \in (0,1)$ the mapping $T := wT_1 + (1-w)T_2$ is quasi-nonexpansive and satisfies $\mathrm{Fix}(T) = \mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2)$. In particular, if each T_i $(i = 1,2)$ is $\alpha_i(> 0)$-strongly attracting, then T is $\left(\frac{(\alpha_1+1)(\alpha_2+1)}{(1-w)\alpha_1 + w\alpha_2 + 1} - 1\right)$-strongly attracting.

(b) (Composition [135]) Let $T_1 : \mathscr{H} \to \mathscr{H}$ be a quasi-nonexpansive mapping and $T_2 : \mathscr{H} \to \mathscr{H}$ an attracting quasi-nonexpansive mapping satisfying $\mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2) \neq \emptyset$. Then $T := T_2 T_1$ is quasi-nonexpansive and $\mathrm{Fix}(T) = \mathrm{Fix}(T_1) \cap \mathrm{Fix}(T_2)$. In particular, if each T_i $(i = 1,2)$ is $\alpha_i(> 0)$-strongly attracting, then T is $\left(\frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}\right)$-strongly attracting.

(c) (Operations for averaged nonexpansive mappings [100, 135]) Suppose that each $T_i : \mathscr{H} \to \mathscr{H}$ $(i = 1,2)$ is α_i-averaged nonexpansive for some $\alpha_i \in [0,1)$. Then for every $w \in [0,1]$, the mapping $(1-w)T_1 + wT_2$ is $\{(1-w)\alpha_1 + w\alpha_2\}$-averaged nonexpansive. Moreover $T_1 T_2$ is α-averaged nonexpansive for $\alpha := \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in [0,1)$.

Finally, for intuitive understanding, we explain briefly how the attracting mapping is connected in essence with signal processing.

Remark 17.5. (A Role of Attracting Mapping in Signal Processing) Monotone approximation to an unknown desirable information to be estimated, say *estimandum*, is one of the most favorable properties for signal processing algorithms. In particular, in adaptive filtering or adaptive system identification problems (e.g., adaptive channel equalization, adaptive echo cancellation, etc), the algorithms are required, at each time, to offer a tentative approximation of an estimandum. By utilizing a priori knowledge as well as the latest statistical knowledge obtained from observed data, the algorithm is desired to update the previous estimate to a better one which is closer to the *estimandum*. A practical scenario to realize such a monotone approximation is divided into the following two steps: (Step 1) define a set, say a *target set*, which is sufficiently small but contains candidates consistent with all available knowledge on the *estimandum*, and (Step 2) realize a mapping T which shifts any point not in the target set strictly closer to every point in the target set and does not move any point in the target set (see Fig.17.4). If the *estimandum* surely belongs to the target set, the above scenario automatically realizes a monotone approximation to the *estimandum*. The mapping satisfying the condition in Step 2 is called *attracting mapping*. Obviously a point does not move by the mapping if and only if it is already in the target set. Therefore the target set must be the fixed point set of the attracting mapping. This observation suggests that *a key to realize a successful signal processing algorithm is how to design an attracting mapping of which the fixed point set is the target set*. On the other hand, as seen in Proposition 17.3(a), *the fixed point set of any attracting mapping is a closed convex set*. This simple but valuable observation tells us that for realizing monotone approximation, the attracting mapping is certainly ideal, and in this case, the target set is ensured to be a closed convex set. Moreover, if multiple attracting mappings with a common fixed point are given, we can define in constructive ways a new attracting mapping whose fixed point set is the intersection of the fixed point sets of the given mappings (see Proposition 17.4), which is extremely fortunate for the refinement of the target set in Step 1. Therefore the attracting mapping has a great deal of potential to be not only a computational tool for monotone approximation to the closed convex set but also an alternative mathematical expression of the closed convex set as its fixed point set.

In [139], it has been clarified that the adaptive filtering algorithms based on orthogonal projections [127, 80, 111] exploit the above feature of the attracting mapping implicitly. This discovery leads to a unified scheme called the *Adaptive Projected Subgradient Method* (APSM) [132, 134, 115]; this scheme is a time-varying extension of the Polyak's subgradient algorithm, which was developed for a nonsmooth convex optimization problem with a fixed target value, to the case where the convex objective itself keeps changing in the whole process. Under this simple umbrella of the APSM, a unified convergence analysis has been established for a wide range of adaptive algorithms. Moreover the APSM has been serving as a guiding principle to create various powerful adaptive algorithms for acoustic systems [145, 144], wireless communication systems [143, 25, 26], distributed learning for diffusion network [27], online learning in Reproducing Kernel Hilbert Spaces [112, 113, 123], etc. Moreover, a steady-state mean-square performance analysis of a simplest example of the APSM has been established in [121]; the analysis is based

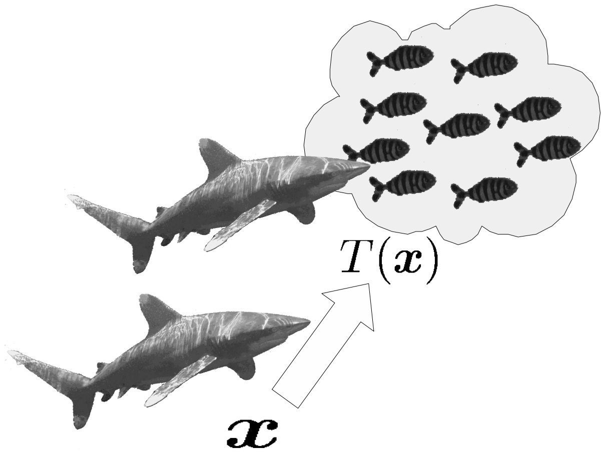


Fig. 17.4 What is the best possible strategy for a starving shark ? Maximal satisfaction is expected by approaching monotonically every fish. This is realized by an attracting mapping.

on the *energy conservation argument* [111] developed specially for performance analyses of adaptive filtering algorithms.

17.2.3 Toolbox of Quasi-Nonexpansive Mapping

We list particularly useful quasi-nonexpansive mappings called in this paper *design tool mappings*. With the aid of Proposition 17.4, the design tool mappings can be used as tools to design a new quasi-nonexpansive mapping whose fixed point set is the intersection of their fixed point sets.

Example 17.6. (Design Tool Mappings)

(a) (Metric projection / Convex projection) Given a nonempty closed convex set C in $\mathscr{H}$, the metric projection $P_C : \mathscr{H} \to C$ is a firmly nonexpansive mapping with $\mathrm{Fix}(P_C) = C$ (see Fact 17.2(c)). The firm nonexpansivity of P_C implies that P_C is also a 1-strongly attracting nonexpansive mapping (see Proposition 17.3(b)). Furthermore, the function $\varphi_1 : x \mapsto d_C^2(x) := \|x - P_C(x)\|^2$ is convex and Gâteaux differentiable over $\mathscr{H}$ with its derivative $\nabla\varphi_1(x) = 2(x - P_C(x))$ $(\forall x \in \mathscr{H})$.

(b) (Proximal forward-backward splitting operator [103, 66, 125, 46]) Suppose that

$$S := \arg\min_{x\in\mathscr{H}} \{f_1(x) + f_2(x)\}$$

is nonempty for $f_1, f_2 \in \Gamma_0(\mathscr{H})$, where f_2 is Gâteaux differentiable on $\mathscr{H}$ with its gradient $\nabla f_2 : \mathscr{H} \to \mathscr{H}$. Then $x^\star \in \mathscr{H}$ satisfies $x^\star \in S$ if and only if $x^\star \in \mathscr{H}$ is a fixed point of the *proximal forward-backward splitting operator* : $\mathrm{prox}_{\mu f_1}(I-\mu\nabla f_2)$ for any $\mu > 0$, i.e., $x^\star = \mathrm{prox}_{\mu f_1}(I-\mu\nabla f_2)(x^\star)$. If in addition ∇f_2 is κ-Lipschitzian for some $\kappa > 0$, the proximal forward-backward splitting operator $\mathrm{prox}_{\mu f_1}(I-\mu\nabla f_2)$ with $\mu \in \left(0, \frac{2}{\kappa}\right]$ is nonexpansive. Moreover, this operator is $\frac{1}{2-\gamma}$-averaged nonexpansive if $\mu \in \left(0, \frac{2\gamma}{\kappa}\right] \subset \left(0, \frac{2}{\kappa}\right)$ (Note: (i) The nonexpansivity of the proximal forward-backward splitting operator with $\mu \in \left(0, \frac{2}{\kappa}\right]$ is confirmed by the nonexpansivity of $\mathrm{prox}_{\mu f_1}$ and the nonexpansivity of $I-\mu\nabla f_2 = \left(1-\frac{\mu}{\frac{2}{\kappa}}\right)I + \frac{\mu}{\frac{2}{\kappa}}\left(I-\frac{2}{\kappa}\nabla f_2\right)$ [see Fact 17.15 in Section 17.2.5]. (ii) The averaged nonexpansivity of the operator with $\mu \in \left(0, \frac{2\gamma}{\kappa}\right] \subset \left(0, \frac{2}{\kappa}\right)$ is confirmed by applying Proposition 17.4(c) to the firm nonexpansivity of $\mathrm{prox}_{\mu f_1}$ and the γ-averaged nonexpansivity of $I-\mu\nabla f_2 = (1-\gamma)I + \gamma\left(I-\frac{\mu}{\gamma}\nabla f_2\right)$). In particular, setting $f_1 := i_C$ for a closed convex set $C \subset \mathscr{H}$ reproduces the characterization of the minimizers of f_2 over C by the fixed point set of the $\frac{1}{2-\gamma}$-averaged nonexpansive mapping $P_C(I-\mu\nabla f_2)$ for $\mu \in \left(0, \frac{2\gamma}{\kappa}\right] \subset \left(0, \frac{2}{\kappa}\right)$ [40, 136, 20]. This is essentially same as the fixed point characterization of the variational inequality problem as found in Fact 17.14 in Section 17.2.5.

(c) (Douglas-Rachford splitting operator [89, 57, 42]) Let $f_1, f_2 \in \Gamma_0(\mathscr{H})$ satisfy

$$S := \arg\min_{x\in\mathscr{H}} \{f_1(x) + f_2(x)\} \neq \emptyset.$$

Under the following *qualification condition*:

$$\left.\begin{array}{l} \mathrm{cone}(\mathrm{dom}(f_1)-\mathrm{dom}(f_2)) := \displaystyle\bigcup_{\lambda>0}\{\lambda x \mid x \in \mathrm{dom}(f_1)-\mathrm{dom}(f_2)\} \\ \text{is a closed subspace of } \mathscr{H}\text{, where} \\ \mathrm{dom}(f_1)-\mathrm{dom}(f_2) := \{x_1 - x_2 \in \mathscr{H} \mid x_i \in \mathrm{dom}(f_i)\ (i=1,2)\}, \end{array}\right\} \tag{17.11}$$

the Douglas-Rachford splitting type algorithm uses in principle the following characterization: for any $\gamma \in (0,\infty)$

$$x^\star \in \mathscr{H} \text{ minimizes } f_1 + f_2 \quad \Leftrightarrow \begin{cases} x^\star = \mathrm{prox}_{\gamma f_2}(y), \\ y \in \mathrm{Fix}\left(\mathrm{rprox}_{\gamma f_1}\mathrm{rprox}_{\gamma f_2}\right), \end{cases} \tag{17.12}$$

which means that S can be expressed as the image of $\mathrm{prox}_{\gamma f_2}$ of the fixed point set of the nonexpansive mapping $\mathrm{rprox}_{\gamma f_1}\mathrm{rprox}_{\gamma f_2}$ (Note: The firm nonexpansivity of $\mathrm{prox}_{\gamma f_i}$, $i = 1, 2$, guarantees the nonexpansivity of $\mathrm{rprox}_{\gamma f_i}$ (see Fact 17.2(b))).

(d) (Subgradient projection) Suppose that a continuous convex function $f : \mathscr{H} \to \mathbb{R}$ satisfies $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$. Let $f'(x) \in \partial f(x)$ $(\forall x \in \mathscr{H})$ be a selection from the subdifferential $\partial f(x)$ (Note: In this paper, we use the notation $\nabla f(x)$ for a Gâteaux

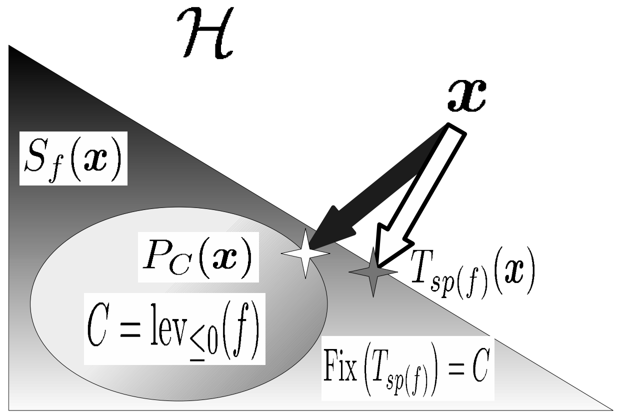


Fig. 17.5 Subgradient projection as an approximation of metric projection (see (17.13) for the definition of $S_f(x)$).

differentiable function f to distinguish from $f'(x)$ for a nonsmooth one). Then a mapping $T_{\mathrm{sp}(f)} : \mathscr{H} \to \mathscr{H}$ defined by

$$T_{\mathrm{sp}(f)} : x \mapsto \begin{cases} x - \frac{f(x)}{\|f'(x)\|^2} f'(x), & \text{if } f(x) > 0; \\ x, & \text{otherwise,} \end{cases}$$

is called a *subgradient projection relative to* f. For $f(x) > 0$, $T_{\mathrm{sp}(f)}(x)$ is given by the metric projection of x onto the closed half-space $\{y \in \mathscr{H} \mid \langle y - x, f'(x) \rangle + f(x) \leq 0\} \supset \mathrm{lev}_{\leq 0}(f)$. Therefore $T_{\mathrm{sp}(f)}$ is a 1-strongly attracting quasi-nonexpansive mapping with $\mathrm{Fix}(T_{\mathrm{sp}(f)}) = \mathrm{lev}_{\leq 0}(f)$ (see, e.g., [126, Lemma 2.8],[8] and Fig.17.5), hence Proposition 17.3(b) implies that $2T_{\mathrm{sp}(f)} - I$ is quasi-nonexpansive. The metric projection onto a closed convex set C can also be interpreted as a subgradient projection relative to a continuous convex function $d_C : x \mapsto \|x - P_C(x)\|$, i.e., $T_{\mathrm{sp}(d_C)} = P_C$. This fact is confirmed by

$$\partial d_C(x) = \begin{cases} \{z \in \mathscr{H} \mid \|z\| \leq 1, \langle z, y - x \rangle \leq 0, \forall y \in C\} \ni 0, & \text{if } x \in C; \\ \frac{x - P_C(x)}{d(x, C)}, & \text{otherwise.} \end{cases}$$

If we can use more information on the function $f \in \Gamma_0(\mathscr{H})$, we may define other strongly attracting mappings that realize better approximation to the set $\mathrm{lev}_{\leq 0}(f)$, as shown below.

Proposition 17.7. (A Generalization of Subgradient Projection [102]) *Let* $f : \mathscr{H} \to \mathbb{R}$ *be a continuous convex function with* $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$ *and* $f' : \mathscr{H} \to \mathscr{H}$ *a selection of the subdifferential* $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$*, i.e.,* $f'(x) \in \partial f(x), \forall x \in \mathscr{H}$*. Let* $\xi : \mathscr{H} \to \mathbb{R}$ *be a function satisfying* $\xi(x) \geq f(x), \forall x \in \mathscr{H}$*. Suppose that*

$$S_\xi(x) := \begin{cases} \{y \in \mathscr{H} \mid \langle y-x, f'(x)\rangle + \xi(x) \le 0\}, & \text{if } f(x) \ge 0; \\ \mathscr{H}, & \text{otherwise,} \end{cases} \tag{17.13}$$

satisfies (O-i) $S_\xi(x) \supset \mathrm{lev}_{\le 0}(f)$, *and (O-ii)* $x \notin \mathrm{lev}_{\le 0}(f) \Rightarrow x \notin S_\xi(x)$. *Then the projection onto* $S_\xi(x)$, *i.e.,*

$$T_{\mathrm{dsp},\xi} : x \mapsto \begin{cases} x - \frac{\xi(x)}{\|f'(x)\|^2} f'(x), & \text{if } f(x) > 0; \\ x, & \text{otherwise,} \end{cases} \tag{17.14}$$

is firmly quasi-nonexpansive with $\mathrm{Fix}(T_{\mathrm{dsp},\xi}) = \mathrm{lev}_{\le 0}(f)$.

Remark 17.8. ($T_{\mathrm{dsp},\xi}$ as a Deeper Outer Approximation) By the definition of subgradient, $S_f(x)$ satisfies the conditions (O-i) and (O-ii) in Proposition 17.7. In this special case, we have $T_{\mathrm{dsp},f} = T_{\mathrm{sp}(f)}$, hence $T_{\mathrm{dsp},\xi} : \mathscr{H} \to \mathscr{H}$ is a generalization of the *subgradient projection relative to* f. If $\xi(x) > f(x) > 0$, we have $S_\xi(x) \subsetneq S_f(x)$, i.e., $S_\xi(x)$ is a *deeper* outer approximation (of $\mathrm{lev}_{\le 0}(f)$ w.r.t. x) than $S_f(x)$. Several constructions of such $\xi(x) (> f(x))$ have been discussed for example in [84, Example 3.4], [102, 140].

Example 17.9. (Deepest Outer Approximation with Available Information)

(a) (Best quadratic lower bound with Lipschitz constant of gradient operator [140]) *Suppose that (i)* $f \in \Gamma_0(\mathscr{H})$ *is Gâteaux differentiable on* $\mathscr{H}$ *with its gradient* $\nabla f : \mathscr{H} \to \mathscr{H}$ *which is* κ*-Lipschitzian over* $\mathscr{H}$, *and (ii)* $\mathrm{lev}_{\le 0}(f) \neq \emptyset$ *and* $f(x) \ge -\rho$ ($\exists \rho \ge 0, \forall x \in \mathscr{H}$). *Fix* $z \in \mathscr{H} \setminus \mathrm{lev}_{\le 0}(f)$ *arbitrarily, and let* $g_{0,z}(x) := \langle x - z, \nabla f(z)\rangle + f(z)$ ($\forall x \in \mathscr{H}$). *Then the function* $g_{1,z} : \mathscr{H} \to \mathbb{R}$:

$$g_{1,z}(x) := \begin{cases} g_{0,z}(x), & \text{if } a \le g_{0,z}(x); \\ \frac{1}{2}\frac{(g_{0,z}(x)-b)^2}{a-b}, & \text{if } b \le g_{0,z}(x) \le a; \\ -\rho, & \text{if } g_{0,z}(x) \le b, \end{cases} \tag{17.15}$$

where $a := -\rho + \frac{\|\nabla f(z)\|^2}{2\kappa}$ *and* $b := -\rho - \frac{\|\nabla f(z)\|^2}{2\kappa}$, *satisfies* $g_{0,z}(x) \le g_{1,z}(x) \le f(x)$ ($\forall x \in \mathscr{H}$). *This implies* $\xi(y) := g_{1,z}(y) - \langle y - z, \nabla f(z)\rangle \ge g_{0,z}(y) - \langle y - z, \nabla f(z)\rangle = f(z)$ ($\forall y \in \mathscr{H}$), *hence*

$$\begin{aligned} \mathrm{lev}_{\le 0}(f) \subset \mathrm{lev}_{\le 0}(g_{1,z}) &= \{y \in \mathscr{H} \mid \langle y-z, \nabla f(z)\rangle + \xi(z) \le 0\} = S_\xi(z) \\ &\subset \{y \in \mathscr{H} \mid \langle y-z, \nabla f(z)\rangle + f(z) \le 0\} = \mathrm{lev}_{\le 0}(g_{0,z}) \\ &= S_f(z). \end{aligned} \tag{17.16}$$

Moreover $g_{1,z}$ *satisfies*

(i) $g_{1,z}(x)|_{x=z} = f(z)$ *and* $\nabla g_{1,z}(x)|_{x=z} = \nabla f(z)$,
(ii) $f(x) \ge g_{1,z}(x) \ge -\rho$ ($\forall x \in \mathscr{H}$) *and* $\|\nabla g_{1,z}(x) - \nabla g_{1,z}(y)\| \le \kappa \|x - y\|$ ($\forall x, y \in \mathscr{H}$).

(b) (Deepest outer approximating half-space of level set of a quadratic function [102]) *Suppose that a quadratic function* $f(x) := \|Ax - b\|^2 - \rho$ ($\forall x \in \mathscr{H}$)

satisfies $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$*, where* $A : \mathscr{H} \to \mathscr{H}'$ *is a bounded linear operator (*$\mathscr{H}'$ *is a real Hilbert space whose inner product and its induced norm are also denoted by* $\langle \cdot, \cdot \rangle$ *and* $\|\cdot\|$ *respectively),* $b \in \mathscr{H}'$ *and* $\rho \in \mathbb{R}$*. Fix* $z \in \mathscr{H} \setminus \mathrm{lev}_{\leq 0}(f)$ *arbitrarily and let* $\xi_\tau(z) := 2\left(f(z) - \tau - \sqrt{(f(z)-\tau)(-\tau)}\right)$ *for any* $\tau \in [-\rho, \inf_{y \in \mathscr{H}} f(y)]$*. Then* $S_{\xi_\tau}(z) \subset \mathscr{H}$ *satisfies*

(i) $\mathrm{lev}_{\leq 0}(f) \subset S_{\xi_\tau}(z) \subsetneq S_f(z)$ *for any* $\tau \in [-\rho, \inf_{y\in\mathscr{H}} f(y)]$,
(ii) $\widetilde{S}_{\xi_{\tau_{\min}}}(z) \cap \mathrm{lev}_{\leq 0}(f) \neq \emptyset$ *for* $\tau_{\min} := \min_{y\in\mathscr{H}} f(y)$*, where* $\widetilde{S}_{\xi_{\tau_{\min}}}(z)$ *is the boundary hyperplane of* $S_{\xi_{\tau_{\min}}}(z)$*.*

17.2.4 Iterative Approximation of a Fixed Point of Quasi-Nonexpansive Mapping

By introducing a real number sequence $(\alpha_n)_{n\geq 0} \subset [0,1]$, the algorithm in the Banach-Picard's contraction mapping theorem has been extended to

$$x_{n+1} := (1-\alpha_n)x_n + \alpha_n T(x_n), \tag{17.17}$$

where T is a quasi-nonexpansive mapping. To guarantee the weak convergence[4] of $(x_n)_{n\geq 0}$ to a fixed point of T, the demiclosedness of $I - T$ at $0 \in \mathscr{H}$ is required in addition to some condition on $(\alpha_n)_{n\geq 0}$, where, in general, a mapping $G : \mathscr{H} \to \mathscr{H}$ is said to be *demiclosed* at $y \in \mathscr{H}$ if weak convergence of a sequence $(x_n)_{n\geq 0} \subset \mathscr{H}$ to $x \in \mathscr{H}$ and strong convergence of $(G(x_n))_{n\geq 0}$ to $y \in \mathscr{H}$ imply $G(x) = y$. It is well-known [19] that the mapping $I - T$ is demiclosed at every point $y \in \mathscr{H}$ if $T : \mathscr{H} \to \mathscr{H}$ is nonexpansive. Moreover, if a continuous convex function $f : \mathscr{H} \to \mathbb{R}$ satisfies $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$ and its subdifferential $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$ is bounded in the sense of Fact 17.2(a), the mapping $I - T_{\mathrm{sp}(f)}$ is demiclosed at $0 \in \mathscr{H}$ (see [126, Lemma 2.9],[8]).

The convergence theorem of the algorithm (17.17), which is called the *Mann iterative process*, is summarized as follows.

Proposition 17.10. (Mann Iterative Process)
Let $T : \mathscr{H} \to \mathscr{H}$ be a quasi-nonexpansive mapping. Then for any initial point $x_0 \in \mathscr{H}$, the sequence $(x_n)_{n\geq 0} \subset \mathscr{H}$, generated by (17.17), converges weakly to a point in $\mathrm{Fix}(T)$, which depends on the choices of $x_0 \in \mathscr{H}$ and the real number sequence $(\alpha_n)_{n\geq 0} \subset [0,1]$, under either of the following conditions.

---

[4] (Strong and weak convergences) A sequence $(x_n)_{n\geq 0}$ in a real Hilbert space $\mathscr{H}$ is said to converge strongly to a point $x \in \mathscr{H}$ if the real number sequence $(\|x_n - x\|)_{n\geq 0}$ converges to 0, and to converge weakly to $x \in \mathscr{H}$ if the real number sequence $(\langle x_n - x, y\rangle)_{n\geq 0}$ converges to 0 for every $y \in \mathscr{H}$. If $(x_n)_{n\geq 0}$ converges strongly to x, then $(x_n)_{n\geq 0}$ converges weakly to x. The converse is true if $\mathscr{H}$ is finite dimensional, hence in finite dimensional case we do not need to distinguish these convergences.

(a) $I-T$ is demiclosed at $0 \in \mathscr{H}$ and $(\alpha_n)_{n\geq 0}$ is bounded away from 0 and 1, i.e., there exist $\varepsilon_1, \varepsilon_2 > 0$ satisfying $(\alpha_n)_{n\geq 0} \subset [\varepsilon_1, 1-\varepsilon_2]$ [54].
(b) T is nonexpansive and $\sum_{n\geq 0} \alpha_n(1-\alpha_n) = \infty$ [74].

Remark 17.11. (Several Forms of Mann-type Iterates)

(a) The iterative algorithm shown in (17.17) is commonly referred to as "Mann iterative process" because this has an alternative expression of

$$x_{n+1} := \sum_{j=1}^{n} a_{n,j} u_j \quad \text{and} \quad u_{n+1} := T(x_n) \tag{17.18}$$

given in [93] if $(a_{n,j})_{0\leq j\leq n, n\geq 0} \subset [0,1]$ satisfies $a_{n+1,j} = (1-a_{n+1,n+1})a_{n,j}$ and $\alpha_n = a_{n+1,n+1}$ $(n = 0,1,2,\ldots)$.
(b) Suppose in particular (i) that T is α-averaged quasi-nonexpansive for some $\alpha \in (0,1)$, i.e., $N := \frac{T-(1-\alpha)I}{\alpha}$ is quasi-nonexpansive, and (ii) that $I-T$ is demiclosed at $0 \in \mathscr{H}$. Then the sequence $(x_n)_{n=0}^{\infty} \subset \mathscr{H}$ generated by any initial point $x_0 \in \mathscr{H}$ and

$$x_{n+1} := T(x_n) = (1-\alpha)x_n + \alpha N(x_n)$$

converges weakly to a point in $\mathrm{Fix}(N) = \mathrm{Fix}(T)$.
(c) If $N : \mathscr{H} \to \mathscr{H}$ is firmly nonexpansive, i.e., $T := 2N - I$ is nonexpansive with $\mathrm{Fix}(T) = \mathrm{Fix}(N)$, the iteration (17.17) can be expressed equivalently as

$$x_{n+1} := \left(1 - \frac{t_n}{2}\right) x_n + \frac{t_n}{2}(2N-I)(x_n) = (1-t_n)x_n + t_n N(x_n),$$

where the conditions for $(\alpha_n)_{n\geq 0} \subset [0,1]$ in Proposition 17.10(a),(b) are replaced respectively by $(t_n)_{n\geq 0} = (2\alpha_n)_{n\geq 0} \subset [2\varepsilon_1, 2-2\varepsilon_2]$ and $\sum_{n\geq 0} t_n(2-t_n) = \infty$. This is a simplest case of a weak convergence theorem shown in [39] under much weaker conditions in order to cope with the numerical errors possibly unavoidable in the iterative computations.
(d) (Elsner-Koltracht-Neumann [60]) Suppose that $T : \mathbb{R}^m \to \mathbb{R}^m$ is continuous as well as attracting (In this case, the mapping T is said to be paracontractive). Then for any initial point $x_0 \in \mathbb{R}^m$, the sequence $(x_n)_{n\geq 0}$ generated by $x_{n+1} := T(x_n)$ converges to a point in $\mathrm{Fix}(T)$ (Note: This idea has been extended to the case of *Bregman distance* [29]).

We have found many useful algorithms whose primitive convergence properties can be examined simply by Proposition 17.10.

Example 17.12. (Mann Iterative Process Found in Applications)

(a) (POCS: Projections onto convex sets [18, 76, 142, 118]) Suppose that $C_i \subset \mathscr{H}$ $(i = 1,2,\ldots,m)$ are closed convex sets satisfying $\bigcap_{i=1}^{m} C_i \neq \emptyset$. Define $\frac{\lambda_i}{2}$-averaged nonexpansive mappings $T_i : \mathscr{H} \to \mathscr{H}$ $(i = 1,2,\ldots,m)$, with $\lambda_i \in$

$(0,2)$, by $T_i := I + \lambda_i(P_{C_i} - I) = \left(1 - \frac{\lambda_i}{2}\right) I + \frac{\lambda_i}{2}(2P_{C_i} - I)$, which obviously satisfy $\mathrm{Fix}(T_i) = C_i$ (see Example 17.6(a)). Moreover, by Proposition 17.4 (c) and (b), $T = T_m T_{m-1} \cdots T_1$ is averaged nonexpansive with $\mathrm{Fix}(T) = \bigcap_{i=1}^{m} C_i \neq \emptyset$. Applying Remark 17.11(b) to T, we verify that the sequence $(x_n)_{n\geq 0}$ generated by any $x_0 \in \mathscr{H}$ and $x_{n+1} := T(x_n)$ $(n = 0,1,2,\ldots)$ converges weakly to a point in $\bigcap_{i=1}^{m} C_i \neq \emptyset$. This scheme is the so-called *projections onto convex sets (POCS)* and applicable to *convex feasibility problems*.

(b) (Proximal forward-backward splitting method [103, 66, 125, 46]) Suppose that

$$S := \arg\min_{x\in\mathscr{H}} \{f_1(x) + f_2(x)\}$$

is nonempty for $f_1, f_2 \in \Gamma_0(\mathscr{H})$, where f_2 is Gâteaux differentiable on $\mathscr{H}$ with its κ-Lipschitzian gradient $\nabla f_2 : \mathscr{H} \to \mathscr{H}$. Then, for any $\mu \in \left(0, \frac{2}{\kappa}\right)$, the sequence $(x_n)_{n\geq 0}$ generated by any initial point $x_0 \in \mathscr{H}$ and $x_{n+1} := \mathrm{prox}_{\mu f_1}(I - \mu\nabla f_2)(x_n)$ converges weakly to a point in S. This scheme is the so-called *proximal forward-backward splitting method* which can be interpreted as a direct application of Remark 17.11(b) to Example 17.6(b).

(c) (Projected gradient method [71, 86]) Let $C \subset \mathscr{H}$ be a closed convex set and $f : \mathscr{H} \to \mathbb{R}$ a Gâteaux differentiable convex function satisfying $\arg\min_{x\in C} f(x) \neq \emptyset$. Suppose that the derivative $\nabla f : \mathscr{H} \to \mathscr{H}$ is κ-Lipschitzian over $\mathscr{H}$ for some $\kappa > 0$. Then for any $\mu \in \left(0, \frac{2}{\kappa}\right)$, the sequence $(x_n)_{n\geq 0}$ generated by any initial point $x_0 \in \mathscr{H}$ and $x_{n+1} := P_C(x_n - \mu\nabla f(x_n))$ converges weakly to a point in $\arg\min_{x\in C} f(x)$. This scheme is the so-called *projected gradient method* which can be interpreted as a direct application of Example 17.12(b) to $f_1 = i_C$ and $f_2 := f$.

(d) (PPM: Parallel projection method [34, 40]) Suppose that $K \subset \mathscr{H}$ and $C_i \subset \mathscr{H}$ $(i = 1,2,\ldots,m)$ are nonempty closed convex sets possibly having $K \cap (\bigcap_{i=1}^{m} C_i) = \emptyset$. Suppose also that the *mean squared distance function*: $\Phi_{\mathrm{ms}}(x) := \frac{1}{2}\sum_{i=1}^{m} w_i d_{C_i}^2(x)$ has its minimizer over K, i.e., $K_{\Phi_{\mathrm{ms}}} := \arg\min_{x\in K} \Phi_{\mathrm{ms}}(x) \neq \emptyset$, where $w_i > 0$ $(i = 1,2,\ldots,m)$ and $\sum_{i=1}^{m} w_i = 1$. Then the sequence $(x_n)_{n=0}^{\infty}$ generated by any $\mu \in (0,2)$, any $x_0 \in \mathscr{H}$ and

$$x_{n+1} := P_K\left((1-\mu)x_n + \mu\sum_i w_i P_{C_i}(x_n)\right)$$

converges weakly to a point in $K_{\Phi_{\mathrm{ms}}}$. This scheme is the so-called *parallel projection method (PPM)* and applicable to *inconsistent convex feasibility problems*. The PPM can be interpreted as a direct application of Example 17.12(c) to $f(x) = \Phi_{\mathrm{ms}}(x)$.

(e) (Projected Landweber method [58, 75]/ CQ-algorithm [20, 21]) Let $\mathscr{H}_o$ be a real Hilbert space equipped with an inner product $\langle\cdot,\cdot\rangle_o$ and its induced norm

$\|\cdot\|_o$. Suppose that the operator $A : \mathscr{H} \to \mathscr{H}_o$ is linear and bounded, i.e., $\|A\| := \sup_{x \in \mathscr{H} \setminus \{0\}} \frac{\|A(x)\|_o}{\|x\|} < \infty$, and that a closed convex set $C \subset \mathscr{H}$ and $b \in \mathscr{H}_o$ satisfy $\mathscr{S}_1 := \arg\min_{x \in C} \|A(x) - b\|_o^2 \neq \emptyset$. Then for any $\mu \in \left(0, 2\|A\|^{-2}\right)$, the sequence $(x_n)_{n \geq 0}$ generated by any point $x_0 \in \mathscr{H}$ and

$$x_{n+1} := P_C\left(x_n - \mu A^* A(x_n) + \mu A^*(b)\right)$$

converges weakly to a point in $\mathscr{S}_1$, where $A^* : \mathscr{H}_o \to \mathscr{H}$ is the adjoint operator of A [141, 48, 85, 133, 10]. This scheme is the so-called *projected Landweber method* and applicable to convexly constrained inverse problems. The projected Landweber method can be interpreted as a direct application of Example 17.12(c) to $f(x) = \frac{1}{2}\|A(x) - b\|_o^2$.
On the other hand, for given a pair of closed convex sets $C \subset \mathscr{H}$ and $Q \subset \mathscr{H}_o$, the problem for finding a point $x \in \mathscr{H}$ satisfying $x \in C$ and $A(x) \in Q$ is called the *split feasibility problem* (SFP). Since the SFP is reduced to a problem for finding a point in

$$\mathscr{S}_2 := \arg\min_{x \in C} \|P_Q A(x) - A(x)\|_o^2 \neq \emptyset,$$

a direct application of Example 17.12(c) to $f(x) = \frac{1}{2}\|P_Q A(x) - A(x)\|_o^2$ leads to the algorithm: $x_{n+1} := P_C\left(x_n - \mu A^*(I - P_Q)A(x_n)\right)$, which generates a weakly convergent sequence $(x_n)_{n \geq 0}$ to a point in $\mathscr{S}_2$ for any $\mu \in \left(0, 2\|A\|^{-2}\right)$ and any point $x_0 \in \mathscr{H}$. This scheme is the so-called *CQ-algorithm* and applicable to *split feasibility problems* (Note: The Mann iterative process has been applied to many other types of inverse problems. For example, an elliptic Cauchy problem was solved in [61] with Proposition 17.10(b) as a fixed point problem for a nonexpansive affine operator in a Hilbert space).

(f) (Douglas-Rachford splitting method [89, 57, 42]) Let $f_1, f_2 \in \Gamma_0(\mathscr{H})$ satisfy

$$S := \arg\min_{x \in \mathscr{H}} \{f_1(x) + f_2(x)\} \neq \emptyset.$$

Under the condition (17.11), the sequence $(x_n)_{n=0}^{\infty}$ generated by

$$x_{n+1} := (1 - \alpha_n)x_n + \alpha_n \mathrm{rprox}_{\gamma f_1} \mathrm{rprox}_{\gamma f_2}(x_n), \tag{17.19}$$

for any $x_0 \in \mathscr{H}$, any $\gamma \in (0, \infty)$ and any $(\alpha_n)_{n \geq 0} \subset [0, 1]$ satisfying $\sum_{n \geq 0} \alpha_n(1 - \alpha_n) = \infty$, converges weakly to a point in $(\mathrm{prox}_{\gamma f_2})^{-1}(S)$. The scheme (17.19) can be interpreted as a direct application of Proposition 17.10(b) to Example 17.6(c). Moreover, with use of $(t_n)_{n \geq 0} := (2\alpha_n)_{n \geq 0} \subset [0, 2]$ satisfying $\sum_{n \geq 0} t_n(2 - t_n) = \infty$, the scheme (17.19) can be expressed equivalently as

$$x_{n+1} := x_n + t_n \left\{ \mathrm{prox}_{\gamma f_1}\left(2\mathrm{prox}_{\gamma f_2}(x_n) - x_n\right) - \mathrm{prox}_{\gamma f_2}(x_n) \right\}, \tag{17.20}$$

which is a simplest example of the so-called *Douglas-Rachford splitting type algorithm* in [42, Theorem 20]. In particular, if $\dim(\mathscr{H}) < \infty$, the nonexpansivity of $\mathrm{prox}_{\gamma f_2}$ and the weak convergence of $(x_n)_{n=0}^{\infty}$ by (17.19) [or by (17.20)] to a point, say

$$y^\star \in (\mathrm{prox}_{\gamma f_2})^{-1}(S)(\Leftrightarrow \mathrm{prox}_{\gamma f_2}(y^\star) \in S),$$

guarantee

$$\|\mathrm{prox}_{\gamma f_2}(x_n) - \mathrm{prox}_{\gamma f_2}(y^\star)\| \leq \|x_n - y^\star\| \to 0 \quad (n \to \infty).$$

(g) (Subgradient method [106]) Let $f : \mathscr{H} \to \mathbb{R}$ be a continuous convex function satisfying $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$. Define a sequence $(x_n)_{n\geq 0} \subset \mathscr{H}$ with any initial point $x_0 \in \mathscr{H}$ and

$$x_{n+1} := \begin{cases} x_n - \lambda_n \frac{f(x_n)}{\|f'(x_n)\|^2} f'(x_n), & \text{if } f(x_n) > 0; \\ x_n, & \text{otherwise}, \end{cases} \tag{17.21}$$

where $f'(x_n) \in \partial f(x_n)$ for $f(x_n) > 0$, and $(\lambda_n)_{n\geq 0} \in (0,2)$ is bounded away from 0 and 2. Then the iteration (17.21) can be expressed as

$$x_{n+1} = \left[\left(1 - \frac{\lambda_n}{2}\right) I + \frac{\lambda_n}{2} T\right](x_n),$$

where $T := 2T_{\mathrm{sp}(f)} - I$. In particular, if the subdifferential $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$ is bounded in the sense of Fact 17.2(a), $I - T$ is demiclosed at $0 \in \mathscr{H}$ (see the first paragraph of Section 17.2.4), hence Proposition 17.10(a) guarantees the weak convergence of $(x_n)_{n=0}^{\infty}$ to a point in $\mathrm{Fix}(T) = \mathrm{Fix}(T_{\mathrm{sp}(f)}) = \mathrm{lev}_{\leq 0}(f)$.
This method is very useful for the following convex feasibility problems. Suppose that continuous convex functions $f_i : \mathscr{H} \to \mathbb{R}$ $(i = 1, 2, \ldots, m)$ satisfy $\bigcap_{i=1}^m \mathrm{lev}_{\leq 0}(f_i) \neq \emptyset$. Then, by defining a single convex function $f : \mathscr{H} \to \mathbb{R}$ satisfying $\mathrm{lev}_{\leq 0}(f) = \bigcap_{i=1}^m \mathrm{lev}_{\leq 0}(f_i)$, for example by $f(x) := \max_{i=1}^m f_i(x)$ or by $f(x) := \sum_{i=1}^m w_i f_i^+(x)$ with $f_i^+(x) = \max\{f_i(x), 0\}$ and $w_i > 0$ $(i = 1, 2, \ldots, m)$, we can reformulate the problem of finding a point in the nonempty intersection of the closed convex sets $\mathrm{lev}_{\leq 0}(f_i)$ to the problem of finding a point in $\mathrm{lev}_{\leq 0}(f)$. Indeed, if $f_i'(x_n) \in \partial f_i(x_n)$ $(i = 1, 2, \ldots, m)$ are available to compute $f'(x_n) \in \partial f(x_n)$ with the well-known calculus rules [81] and $\partial f : \mathscr{H} \to 2^{\mathscr{H}}$ is bounded, we can generate a weakly convergent sequence to a point in $\mathrm{lev}_{\leq 0}(f)$ by applying (17.21) to f.
Moreover, if $P_{\mathrm{lev}_{\leq 0}(f_i)} : \mathscr{H} \to \mathrm{lev}_{\leq 0}(f_i)$ $(i = 1, 2, \ldots, m)$ are available, an application of (17.21) with the aid of Example 17.6(a) to

$$f(x) := \frac{1}{2} \sum_{i=1}^m w_i d^2_{\mathrm{lev}_{\leq 0}(f_i)}(x) = \frac{1}{2} \sum_{i=1}^m w_i \|x - P_{\mathrm{lev}_{\leq 0}(f_i)}(x)\|^2$$

leads immediately to a version of the parallel projection algorithm [7, 30, 34] for convex feasibility problems.

17.2.5 Monotonicity of Derivatives of Convex Functions, Variational Inequality Problems

A mapping $F : \mathscr{H} \to \mathscr{H}$ is called (i) *monotone* over $S \subset \mathscr{H}$ if $\langle F(u) - F(v), u - v\rangle \geq 0$ for all $u, v \in S$. In particular, a mapping F which is monotone over $S \subset \mathscr{H}$ is called (ii) *paramonotone* over S if $\langle F(u) - F(v), u - v\rangle = 0 \Leftrightarrow F(u) = F(v)$ for all $u, v \in S$; (iii) η*-inverse strongly monotone* (or *firmly monotone*) over S if there exists $\eta > 0$ such that $\langle F(u) - F(v), u - v\rangle \geq \eta \|F(u) - F(v)\|^2$ for all $u, v \in S$ [90]; (iv) η*-strongly monotone* over S if there exists $\eta > 0$ such that $\langle F(u) - F(v), u - v\rangle \geq \eta \|u - v\|^2$ for all $u, v \in S$ [148].

Given $F : \mathscr{H} \to \mathscr{H}$ which is monotone over a nonempty closed convex set $C \subset \mathscr{H}$, the *variational inequality problem* $VIP(F,C)$ is defined as follows: find $u^\star \in C$ such that $\langle u - u^\star, F(u^\star)\rangle \geq 0$ for all $u \in C$. If a function $f \in \Gamma_0(\mathscr{H})$ is Gâteaux differentiable over an open set $U \supset C$, then the derivative ∇f is paramonotone over C [28]. In this case, the solution set of $VIP(\nabla f, C)$ is nothing but the set $\arg\min_{x\in C} f(x)$ provided that it is nonempty (see, e.g., [59, Proposition II.2.1] and [133, Theorem 7.7]).

The following facts are quite useful for translating a convex optimization problem into a fixed point problem.

Fact 17.13. (Properties of Variational Inequality Problem) [28, 59] *Let $F : \mathscr{H} \to \mathscr{H}$ be monotone and continuous over a nonempty closed convex set $C \subset \mathscr{H}$. Then*

(a) $u^\star$ is a solution of $VIP(F,C)$ if and only if, for all $u \in C$, $\langle F(u), u - u^\star\rangle \geq 0$.
(b) Suppose that (i) F is paramonotone over C, (ii) $u^\star \in C$ is a solution of $VIP(F,C)$ and (iii) $u \in C$ satisfies $\langle F(u), u - u^\star\rangle = 0$. Then u is also a solution of $VIP(F,C)$.

The characterization in (17.9) of the convex projection P_C yields at once an alternative interpretation of the VIP as a fixed point problem.

Fact 17.14. (VIP as a Fixed Point Problem) *Given $F : \mathscr{H} \to \mathscr{H}$ which is monotone over a nonempty closed convex set C, the following three statements are equivalent.*

(a) *$u^\star \in C$ is a solution of VIP(F,C); i.e.,*

$$\langle v - u^\star, F(u^\star)\rangle \geq 0 \text{ for all } v \in C.$$

(b) *For an arbitrarily fixed $\mu > 0$, $u^\star \in C$ satisfies*

$$\langle v - u^\star, (u^\star - \mu F(u^\star)) - u^\star\rangle \leq 0 \text{ for all } v \in C.$$

(c) *For an arbitrarily fixed* $\mu > 0$,

$$u^\star \in \mathrm{Fix}\left(P_C\left(I - \mu F\right)\right). \tag{17.22}$$

Fact 17.15. (Baillon-Haddad Theorem [56, 72, 4, 90, 9]) *Let* $f \in \Gamma_0(\mathscr{H})$ *be Gâteaux differentiable with its gradient* $\nabla f : \mathscr{H} \to \mathscr{H}$. *Then the following three statements are equivalent.*

(a) ∇f *is* κ*-Lipschitzian over* $\mathscr{H}$.
(b) ∇f *is* $1/\kappa$*-inverse strongly monotone over* $\mathscr{H}$.
(c) $I - \frac{2}{\kappa}\nabla f : \mathscr{H} \to \mathscr{H}$ *is nonexpansive over* $\mathscr{H}$.

Remark 17.16. (On Fact 17.15)

(a) The equivalence of Fact 17.15(b) and Fact 17.15(c) is confirmed by a simple algebra.
(b) Fact 17.15(c) guarantees that κ-Lipschitz continuity of ∇f implies $\frac{\mu\kappa}{2}$-averaged nonexpansivity of $I - \mu\nabla f = \left(1 - \frac{\mu\kappa}{2}\right) I + \frac{\mu\kappa}{2}\left(I - \frac{2}{\kappa}\nabla f\right)$ for any $\mu \in \left(0, \frac{2}{\kappa}\right)$.

17.3 Minimizing Moreau Envelope by Hybrid Steepest Descent Method

17.3.1 Moreau Envelope and Its Derivative

The Moreau envelope has surprisingly nice properties as follows.

Fact 17.17. (Distinctive Properties of Moreau Envelope (see, e.g., [97, 98, 108, 46])) *Given a function* $f \in \Gamma_0(\mathscr{H})$, *the Moreau envelope* ${}^\gamma f : \mathscr{H} \to \mathbb{R}$ *of* f *of index* $\gamma \in (0, \infty)$ *in (17.2) satisfies the following.*

(a) (Lower bound) $(\forall \gamma \in (0, \infty), \forall x \in \mathscr{H})$ $f(x) \geq {}^\gamma f(x)$.
(b) (Convergence) The function ${}^\gamma f$ *converges pointwise to* f *on* $\mathrm{dom}(f)$ *as* $\gamma \to 0$, *i.e.,*

$$\lim_{\gamma \downarrow 0} {}^\gamma f(x) = f(x) \ \ (\forall x \in \mathrm{dom}(f)).$$

Moreover, if f *is uniformly continuous on a bounded set* $S \subset \mathrm{dom}(f)$, ${}^\gamma f$ *converges uniformly to* f *on* S, *i.e.,* $\lim_{\gamma \downarrow 0} \sup_{x \in S} |{}^\gamma f(x) - f(x)| = 0$. *In particular, if* f *is continuous on a compact set* $S \subset \mathrm{dom}(f)$, *the* Heine's theorem *[1, Theorem 4.47] guarantees the uniform convergence of* ${}^\gamma f$ *to* f *on* S.
(c) (Lipschitz continuity of Fréchet derivative) ${}^\gamma f : \mathscr{H} \to \mathbb{R}$ *is Fréchet differentiable and its derivative is given by*

$$\nabla^\gamma f(x) = \frac{x - \mathrm{prox}_{\gamma f}(x)}{\gamma} = \frac{x - (I + \gamma \partial f)^{-1}(x)}{\gamma}, \tag{17.23}$$

hence $\nabla^{\gamma} f(x)$ is $\frac{1}{\gamma}$-Lipschitzian (Note: The firm nonexpansivity of $I - \mathrm{prox}_{\gamma f}$ is guaranteed by the nonexpansivity of $2(I - \mathrm{prox}_{\gamma f}) - I = -\mathrm{rprox}_{\gamma f}$).

The benefits of the Moreau envelope in applied sciences have been examined for the absolute value function $|\cdot| : \mathbb{R} \to [0,\infty)$. By a simple algebra, we verify that the Moreau envelope of the absolute value function is given explicitly by

$$ {}^{\gamma}|t| := \begin{cases} \frac{1}{2\gamma}t^2, & \text{if } |t| \leq \gamma \text{ ;} \\ |t| - \frac{1}{2}\gamma, & \text{otherwise.} \end{cases} \tag{17.24} $$

As pointed out in [15, 95], this is clearly equal, up to a scaling factor γ, to the so-called *Huber's M cost function* [82]

$$ \rho : \mathbb{R} \to \mathbb{R},\ t \mapsto \begin{cases} \frac{1}{2}t^2, & \text{if } |t| \leq \gamma \text{ ;} \\ \gamma|t| - \frac{1}{2}\gamma^2, & \text{otherwise,} \end{cases} \tag{17.25} $$

in the context of robust linear estimation theory. The Huber's M cost function has been used in an estimation problem:

$$ \text{find } x^{\star} \in \arg\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \rho\left((Ax - b)_i\right), \tag{17.26} $$

where $A \in \mathbb{R}^{m \times n}$ represents the underlying linear model, $b \in \mathbb{R}^m$ is the data vector, and $x \in \mathbb{R}^n$ is the parameter vector. A solution to (17.26), often referred to as an *M-estimator*, is known as a robust alternative to the Least Squares (LS) estimator that is unfortunately sensitive against occurrence of outliers in the ill-conditioned linear regression systems. Computational algorithms for the problem (17.26) are found for example in [95, 87, 16, 92]. In particular, a computational algorithm was given in [16] to a convexly constrained version of the problem (17.26) provided that the metric projection onto the constraint set is possible to compute efficiently.

The Huber's M cost function has also been used in many inverse problems [99, 24, 2, 78] as an excellent *robust convex penalty function* that grows linearly for t far from zero, hence it achieves least sensitivity to large outliers of large residual. We can also observe that the derivative of ${}^{\gamma}|\cdot|$ is always $\frac{1}{\gamma}$-Lipschitzian over $\mathbb{R}$ as mentioned in Fact 17.17(c), while the derivative of a straightforward smooth convex approximation $|\cdot|^p : \mathbb{R} \to [0,\infty)$ for any $1 < p < 2$ can never be Lipschitz continuous over $\mathbb{R}$ due to

$$ \lim_{t \downarrow 0} \left(p(p-1)t^{p-2}\right) = \infty. $$

This means that the Moreau-Yosida regularization offers a unified systematic strategy to realize a beautiful parametrized smooth convex approximation for general convex functions in $\Gamma_0(\mathscr{H})$. Nevertheless, the use of the Moreau envelope and the proximity operator has been very limited for many years in real world applications. This is mainly due to the evident computational difficulty in the definition (17.2), i.e., we have to minimize a possibly nonsmooth convex function $f(\cdot) + \frac{1}{2\gamma}\|x - \cdot\|^2$

for each $x \in \mathscr{H}$ to obtain the $\operatorname{prox}_{\gamma f}(x) \in \mathscr{H}$. Although this computational difficulty has never been resolved in general, the effectiveness of the proximity operator has been confirmed in relatively simple finite dimensional scenarios where $f \in \Gamma_0(\mathbb{R}^n)$ can be expressed in terms of $f_i \in \Gamma_0(\mathbb{R})$ $(i = 1, 2, \ldots, n)$ by

$$f : \mathbb{R}^n \to \mathbb{R} : (x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} f_i(x_i), \tag{17.27}$$

hence

$$\operatorname{prox}_{\gamma f}(x_1, \ldots, x_n) = \left(\operatorname{prox}_{\gamma f_1}(x_1), \ldots, \operatorname{prox}_{\gamma f_n}(x_n)\right).$$

In such a case, the computation of $\operatorname{prox}_{\gamma f}(x_1, \ldots, x_n)$ is reduced to finding the unique minimizer $\operatorname{prox}_{\gamma f_i}(x_i)$ of each univariate convex function $f_i(\cdot) + \frac{1}{2\gamma}|x_i - \cdot|^2$ $(i = 1, \ldots, n)$.

Next we list such useful examples including the soft-thresholding operator, which was developed originally for denoising [53]. Fortunately, the proximity operators of these examples have closed form expressions (Note: Many other useful formulae on the proximity operator are found for example in [46, 45]).

Example 17.18. (Closed Form Expressions of Some Proximity Operators [46])

(a) If $f \in \Gamma_0(\mathbb{R})$ is defined by

$$f : x \mapsto \begin{cases} -\ln(x) & \text{if } x > 0; \\ \infty & \text{if } x \leq 0, \end{cases}$$

we have for any $\gamma \in (0, \infty)$

$$\operatorname{prox}_{\gamma f}(x) = \frac{1}{2}\left(x + \sqrt{x^2 + 4\gamma}\right).$$

(b) Let $\{e_k\}_{k=1}^n$ be an orthonormal basis of $\mathbb{R}^n$ where the standard inner product is defined. Define a function $f \in \Gamma_0(\mathbb{R}^n)$ by $f : \mathbb{R}^n \ni x \mapsto \sum_{k=1}^n f_k(\langle x, e_k \rangle) \in (-\infty, \infty]$, where $f_k \in \Gamma_0(\mathbb{R})$ satisfies $f_k(x_k) \geq 0$ $(\forall x_k \in \mathbb{R})$ and $f_k(0) = 0$ $(k = 1, 2, \ldots, n)$. Then we have

$$\operatorname{prox}_f(x) = \sum_{k=1}^{n} \left(\operatorname{prox}_{f_k}(\langle x, e_k \rangle)\right) e_k \quad (x \in \mathbb{R}^n).$$

(c) In particular, if we define, as a special example of (b),

$$f : \mathbb{R}^n \ni x \mapsto \sum_{k=1}^{n} \omega_k |\langle x, e_k \rangle| \in \mathbb{R},$$

with constant weights $\omega_k > 0$ $(k = 1, 2, \ldots, n)$, we have

$$\mathrm{prox}_f(x) = \sum_{k=1}^{n} \mathrm{sgn}(\langle x, e_k \rangle) \max\{|\langle x, e_k \rangle| - \omega_k, 0\} e_k \quad (x \in \mathbb{R}^n).$$

The proximity operator in Example 17.18(c) is called the *soft-thresholding / shrinkage* [53, 51] and has been used widely for example in noise removal problems and in sparse matrix completion problems [32, 47, 22]. As seen from Example 17.18(c) for $n = 1$, the derivative of the Moreau envelope ${}^{\gamma}f(x)$ of the absolute value function $f(x) = |x|$ can be computed with lower complexity than the derivative of $f_\varepsilon(x) := \sqrt{x^2 + \varepsilon}$ ($\varepsilon > 0$) of which the use as a smooth approximation of f has been found in the literature.

To compute the proximity operator of $f \in \Gamma_0(\mathbb{R}^n)$ in more complex cases where the decomposition of f as in (17.27) is hard, fundamental theorems in convex analysis have been utilized implicitly or explicitly; e.g., the *Fenchel-Rockafeller* duality theorem [108] was used in [31, 46, 62] to compute the proximity operator of certain functions such as the *total variation function*. The following proposition and its corollary explain directly such strategies.

Proposition 17.19. (Expression of Proximity Operator by Legendre-Fenchel Transform) Let $\varphi \in \Gamma_0(\mathbb{R}^m)$, $L \in \mathbb{R}^{m \times n}$ and $d \in \mathrm{int}(S)$, where

$$S := L(\mathbb{R}^n) - \mathrm{dom}(\varphi) := \{Lx - y \in \mathbb{R}^m \mid x \in \mathbb{R}^n \text{ and } y \in \mathrm{dom}(\varphi)\}$$

and $\mathrm{int}(S)$ stands for the interior of S. Define $\widetilde{\varphi} \in \Gamma_0(\mathbb{R}^n)$ by $\widetilde{\varphi} : x \mapsto \varphi(Lx - d)$. Then for arbitrarily fixed $x \in \mathbb{R}^n$ and $\gamma \in (0, \infty)$,

$$\mathrm{prox}_{\gamma\widetilde{\varphi}}(x) := \arg\min_{z \in \mathbb{R}^n} \left(\widetilde{\varphi}(z) + \frac{1}{2\gamma}\|x - z\|^2 \right) = \arg\min_{z \in \mathbb{R}^n} \left(\varphi(Lz - d) + \frac{1}{2\gamma}\|x - z\|^2 \right)$$

can be expressed, with $\bar{y} \in \arg\min_{y \in \mathbb{R}^m} \left(\varphi^*(y) + \langle d, y \rangle + \frac{1}{2\gamma}\|\gamma L^t y - x\|^2 \right)$, by

$$\mathrm{prox}_{\gamma\widetilde{\varphi}}(x) = x - \gamma L^t \bar{y},$$

where $L^t \in \mathbb{R}^{n \times m}$ denotes the transpose of a matrix L and φ^* the (Legendre-Fenchel) conjugate of φ (see Fact 17.2(a)).

Proof. Clearly $\phi(z) := \frac{1}{2\gamma}\|x - z\|^2$ ($\forall z \in \mathbb{R}^n$) has $\mathrm{dom}(\phi) = \mathbb{R}^n$, which implies $d \in \mathrm{int}(S) \Leftrightarrow -d \in \mathrm{int}(-L(\mathrm{dom}(\phi)) + \mathrm{dom}(\varphi))$, where $-L(\mathrm{dom}(\phi)) + \mathrm{dom}(\varphi) = \{-L(x) + y \in \mathbb{R}^m \mid x \in \mathrm{dom}(\phi) \text{ and } y \in \mathrm{dom}(\varphi)\}$. It is also obvious that the conjugate of $\phi \in \Gamma_0(\mathbb{R}^n)$ is given by $\phi^*(u) = \frac{1}{2\gamma}(\|\gamma u + x\|^2 - \|x\|^2)$ ($\forall u \in \mathbb{R}^n$) with $\mathrm{dom}(\phi^*) = \mathbb{R}^n$, which implies

$$\begin{aligned} 0 &\in \mathrm{int}\left(-L^t(\mathrm{dom}(\varphi^*)) - \mathrm{dom}(\phi^*)\right) \\ &= \{-L^t(y) - u \mid y \in \mathrm{dom}(\varphi^*) \text{ and } u \in \mathrm{dom}(\phi^*)\} \\ &= \mathbb{R}^n. \end{aligned} \tag{17.28}$$

Therefore, by applying the *Fenchel-type duality scheme* (see for example [108, Example 11.41]), we deduce

$$-L^*\bar{y} \in \partial\phi(\mathrm{prox}_{\gamma\widetilde{\varphi}}(x)) = \left\{\nabla\phi(\mathrm{prox}_{\gamma\widetilde{\varphi}}(x))\right\} = \left\{\frac{1}{\gamma}\left(\mathrm{prox}_{\gamma\widetilde{\varphi}}(x) - x\right)\right\},$$

where

$$\begin{aligned}\bar{y} &\in \arg\max_{y\in\mathbb{R}^m}\left\{\langle -d, y\rangle - \varphi^*(y) - \phi^*(-L^t y)\right\} \\ &= \arg\min_{y\in\mathbb{R}^m}\left\{\varphi^*(y) + \langle d, y\rangle + \frac{1}{2\gamma}(\|-\gamma L^t y + x\|^2 - \|x\|^2)\right\}.\end{aligned}$$

■

Corollary 17.20. (Proximity Operator of Affinely Pre-composed ℓ_1-Norm Function) Let $\varphi \in \Gamma_0(\mathbb{R}^m)$ be defined by $\varphi : (y_1, \ldots, y_m) \mapsto \sum_{i=1}^m |y_i|$. Then for any $L \in \mathbb{R}^{m\times n}$ and $d \in \mathbb{R}^m$, the proximity operator of the function $\widetilde{\varphi} : x \mapsto \varphi(Lx - d)$ of index $\gamma \in (0, \infty)$ is given by $\mathrm{prox}_{\gamma\varphi} : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto x - \gamma L^t\bar{y}$, where

$$\bar{y} \in \arg\min_{y\in C}\left(\langle d, y\rangle + \frac{1}{2\gamma}\|\gamma L^t y - x\|^2\right) \tag{17.29}$$

with

$$C := \{y = (y_1, \ldots, y_m) \in \mathbb{R}^m \mid |y_i| \le 1 \quad (i = 1, \ldots, m)\}. \tag{17.30}$$

Proof. By $\mathrm{dom}(\varphi) = \mathbb{R}^m$, we have $S = L(\mathbb{R}^n) - \mathrm{dom}(\varphi) = \mathbb{R}^m$ and $d \in \mathrm{int}(S) = \mathbb{R}^m$. Moreover, by [17, Example 3.26], the conjugate of φ is given by $\varphi^* = i_C$. Therefore, $\bar{y} \in \mathbb{R}^m$ in Proposition 17.19 can be characterized by

$$\begin{aligned}\bar{y} &\in \arg\min_{y\in\mathbb{R}^m}\left(i_C(y) + \langle d, y\rangle + \frac{1}{2\gamma}\|\gamma L^t y - x\|^2\right) \\ &= \arg\min_{y\in C}\left(\langle d, y\rangle + \frac{1}{2\gamma}\|\gamma L^t y - x\|^2\right).\end{aligned}$$

■

The computation of the proximity operator $\mathrm{prox}_{\gamma i_C} = P_C$ is immediate for C in (17.30), i.e.,

$$P_C : \mathbb{R}^m \to C, \;\; (x_1, \ldots, x_m) \mapsto (y_1, \ldots, y_m), \quad \text{where} \quad y_i := \begin{cases} x_i & \text{if} |x_i| \le 1 \\ \frac{x_i}{|x_i|} & \text{if} |x_i| > 1 \end{cases}, \tag{17.31}$$

which implies that the solution of the smooth minimization problem (17.29) can be approximated efficiently for example by the *projected gradient method* [71] or many other improved algorithms (see, e.g., [11, 12]).

17.3.2 Hybrid Steepest Descent Method

As seen in Section 17.3.1, minimization of the Moreau-Yosida regularization of a possibly nonsmooth convex function $\Phi \in \Gamma_0(\mathscr{H})$ can be reduced to minimization of a smooth convex function whose gradient is Lipschitz continuous. In this section, we consider the following problem for minimizing such a smooth convex function over the fixed point set of certain quasi-nonexpansive mappings.

Problem 17.21. (Convex Optimization over the Fixed Point Set of Nonlinear Mapping) *Let* $T : \mathscr{H} \to \mathscr{H}$ *be a quasi-nonexpansive mapping whose fixed point set* $\mathrm{Fix}(T) = \{x \in \mathscr{H} \mid T(x) = x\}$ *is nonempty. Suppose that* $\Theta \in \Gamma_0(\mathscr{H})$ *is Gâteaux differentiable with the gradient* $\nabla\Theta$ *which is* κ*-Lipschitzian over* $T(\mathscr{H}) := \{T(x) \in \mathscr{H} \mid x \in \mathscr{H}\}$. *Then the problem is: find a point in the solution set*

$$\begin{aligned}\Omega &:= \left\{x^\star \in \mathrm{Fix}(T) \mid \Theta(x^\star) = \min_{x\in\mathrm{Fix}(T)} \Theta(x)\right\}\\ &= \{x^\star \in \mathrm{Fix}(T) \mid \langle x - x^\star, \nabla\Theta(x^\star)\rangle \geq 0 \quad (\forall x \in \mathrm{Fix}(T))\} \neq \emptyset. \quad (17.32)\end{aligned}$$

The *hybrid steepest descent method* (see, e.g., [137, 49, 138, 136, 100, 101, 129, 135, 120, 91, 83, 150, 33]) :

$$u_{n+1} := T(u_n) - \lambda_{n+1}\nabla\Theta\left(T(u_n)\right), \quad (17.33)$$

is an extremely simple algorithmic solution to Problem 17.21, where $(\lambda_n)_{n\geq 1} \subset [0,\infty)$ is a slowly decreasing nonnegative sequence. Among many convergence analyses on the algorithm (17.33), we introduce the following simple ones.

Theorem 17.22. (Hybrid Steepest Descent Method for Quasi-Nonexpansive Mappings)

I. (Strong convergence for nonexpansive mapping [131, 136]) *Let* $T : \mathscr{H} \to \mathscr{H}$ *be a nonexpansive mapping with* $\mathrm{Fix}(T) \neq \emptyset$. *Suppose that the gradient* $\nabla\Theta$ *is* κ*-Lipschitzian and* η*-strongly monotone over* $T(\mathscr{H})$, *which guarantees* $|\Omega| = 1$. *Then, by using any sequence* $(\lambda_n)_{n\geq 1} \subset [0,\infty)$ *satisfying (W1)* $\lim_{n\to\infty} \lambda_n = 0$, *(W2)* $\sum_{n\geq 1} \lambda_n = \infty$, *(W3)* $\sum_{n\geq 1} |\lambda_n - \lambda_{n+1}| < \infty$ *[or* $(\lambda_n)_{n\geq 1} \subset (0,\infty)$ *satisfying (L1)* $\lim_{n\to\infty} \lambda_n = 0$, *(L2)* $\sum_{n\geq 1} \lambda_n = \infty$, *(L3)* $\lim_{n\to\infty}(\lambda_n - \lambda_{n+1})\lambda_{n+1}^{-2} = 0$*], the sequence* $(u_n)_{n\geq 0}$ *generated, for arbitrary* $u_0 \in \mathscr{H}$, *by (17.33) converges strongly to the uniquely existing point* $u^\star \in \Omega$ *in (17.32).*

II. (Nonstrictly convex optimization I [100, 101]) *Assume* $\dim(\mathscr{H}) < \infty$. *Suppose that (i)* $T : \mathscr{H} \to \mathscr{H}$ *is an attracting nonexpansive mapping with bounded* $\mathrm{Fix}(T) \neq \emptyset$, *(ii)* $\nabla\Theta$ *is* κ*-Lipschitzian over* $T(\mathscr{H})$. *If the following condition (a) or (b) is fulfilled, then* $\Omega \neq \emptyset$ *automatically holds and the sequence* $(u_n)_{n\geq 0}$ *generated by (17.33), for arbitrary* $u_0 \in \mathscr{H}$, *satisfies* $\lim_{n\to\infty} d(u_n, \Omega) = 0$.

   (a) *The nonnegative sequence* $(\lambda_n)_{n\geq 1}$ *in (17.33) satisfies (W1), (W2) and* $(\lambda_n)_{n\geq 1} \in \ell_2$, *i.e.,* $\sum_{n\geq 1} \lambda_n^2 < \infty$.
   (b)

(i) T *is asymptotically shrinking; i.e., there exists* $R > 0$ *satisfying*

$$\sup_{\|u\| \geq R} \frac{\|T(u)\|}{\|u\|} < 1$$

(In this case, the nonemptiness and boundedness of $\mathrm{Fix}(T)$ *automatically hold (see [100])), and*

(ii) *the nonnegative sequence* $(\lambda_n)_{n\geq 1}$ *in (17.33) satisfies (W1) and (W2).*

III. (Nonstrictly convex optimization II [135]) *Assume* $\dim(\mathscr{H}) < \infty$. *Suppose* $f : \mathscr{H} \to \mathbb{R}$ *is a continuous convex function with* $\mathrm{lev}_{\leq 0}(f) \neq \emptyset$. *Let* $f' : \mathscr{H} \to \mathscr{H}$ *be a selection of the subdifferential* ∂f *and let* f' *be bounded on any bounded set. Assume (i)* $\xi(x) \geq f(x), \forall x \in \mathscr{H}$, *and (ii)* $S_\xi(x)$ *(in Proposition 17.7) satisfies (O-i) and (O-ii) for all* $x \in \mathscr{H}$. *Let* $T_\alpha := (1-\alpha)I + \alpha T_{\mathrm{dsp},\xi}, \forall \alpha \in (0,2)$, *where* $T_{\mathrm{dsp},\xi}$ *is defined in (17.14). Let* $K \subset \mathscr{H}$ *be a bounded closed convex set satisfying* $K \cap \mathrm{lev}_{\leq 0}(f) \neq \emptyset$, *which implies that* $T := P_K T_\alpha$ *satisfies* $\mathrm{Fix}(T) = K \cap \mathrm{lev}_{\leq 0}(f) \neq \emptyset$. *Suppose that* $\Theta \in \Gamma_0(\mathscr{H})$ *is Gâteaux differentiable over* K *where the gradient* $\nabla\Theta$ *is* κ*-Lipschitzian. Then* $\Omega \neq \emptyset$ *automatically holds and the sequence* $(u_n)_{n\geq 0}$ *generated by (17.33), for any* $u_0 \in \mathscr{H}$ *and* $\alpha \in (0,2)$, *satisfies* $\lim_{n\to\infty} d(u_n, \Omega) = 0$ *if* $(\lambda_n)_{n\geq 1} \subset [0,\infty)$ *is chosen to satisfy (W1) and (W2).*

The algorithm (17.33) was established originally as a generalization of the following fixed point iteration [77, 88, 128, 6] so-called *Halpern-type iteration* (or *anchor method*):

$$u_{n+1} := \lambda_{n+1} a + (1-\lambda_{n+1}) T(u_n), \tag{17.34}$$

which converges strongly to $P_{\mathrm{Fix}(T)}(a)$ for a nonexpansive mapping $T : \mathscr{H} \to \mathscr{H}$ and $a \in \mathscr{H}$.

Remark 17.23. (Conditions on $(\lambda_n)_{n\geq 1} \subset [0,\infty)$ in (17.33))

(a) (Necessary condition [77]) $\lim_{n\to\infty} \lambda_n = 0$ *and* $\sum_{n\geq 1} \lambda_n = \infty$ *are necessary to ensure the convergence of* $(u_n)_{n\geq 0}$ *to a point in* Ω. *Indeed, in the simple case of* $\mathscr{H} := \mathbb{R}$, $T(x) := 1$ *(*$\forall x \in \mathbb{R}$*) and* $\Theta(x) = \frac{1}{2}x^2$ *(*$\forall x \in \mathbb{R}$*), the method (17.33) is reduced to*

$$u_{n+1} := (1-\lambda_{n+1}) T(u_n) = 1 - \lambda_{n+1}, \quad n = 0,1,2,\ldots,$$

hence $\lim_{n\to\infty} \lambda_n = 0$ *is necessary for* $\lim_{n\to\infty} u_n = 1 \in \mathrm{Fix}(T) = \{1\}$. *Moreover, in the case of* $\mathscr{H} := \mathbb{R}$, $T(x) := -x$ *(*$\forall x \in \mathbb{R}$*) and* $\Theta(x) = \frac{1}{2}x^2$ *(*$\forall x \in \mathbb{R}$*), the method (17.33), for* $u_0 = 1$, *is reduced to*

$$u_{n+1} := (1-\lambda_{n+1}) T(u_n) = (-1)^n \prod_{i=0}^{n} (1-\lambda_{i+1}), \quad n = 0,1,2,\ldots,$$

from which $\prod_{i=0}^{\infty}(1-\lambda_{i+1}) = 0$ *(*$\Leftrightarrow \sum_{n=1}^{\infty} \lambda_n = \infty$ *when* $\lim_{n\to\infty} \lambda_n = 0$ *and* $(\lambda_n)_{n\geq 1} \subset [0,1)$*) is necessary for* $\lim_{n\to\infty} u_n = 0 \in \mathrm{Fix}(T) = \{0\}$.

(b) (Sufficient condition) *For the formula (17.34), the set of conditions (L1)–(L3) for* $(\lambda_n)_{n\geq 1} \subset (0,1]$ *was introduced in [88] while (W1)–(W3) for* $(\lambda_n)_{n\geq 1} \subset [0,1]$ *was introduced in [128]. [Note:* $\lambda_n := 1/n^{\rho}$ *for* $0<\rho<1$ *is a simple example of the sequence* $(\lambda_n)_{n\geq 1}$ *satisfying (L1)–(L3). The set of conditions (W1)–(W3) allows the case* $\lambda_n = \frac{1}{n}$*]. The condition (L3) was relaxed to* $\lim_{n\to\infty} \frac{\lambda_n}{\lambda_{n+1}} = 1$ *in [129], which allows the case* $\lambda_n = \frac{1}{n}$*. Moreover, if* T *is an averaged nonexpansive mapping, it was shown in [83] that the (W1) and (W2) for* $(\lambda_n)_{n\geq 1}$ *are sufficient to guarantee the strong convergence of (17.33) to the unique point in* Ω *under the scenario of Theorem 17.22-I (The sufficiency of (W1) and (W2) to guarantee the strong convergence of (17.34) to* $P_{\mathrm{Fix}(T)}(a)$ *was shown in [119]).*

Remark 17.24. (Hybrid Steepest Descent Method as an Extension of the Proximal Forward-Backward Splitting)

(a) *Under the same conditions imposed in Theorem 17.22-I, the sequence* $v_n := T(u_n)$ $(n = 0,1,2,\ldots)$ *generated, for any* $v_0 := T(u_0) \in T(\mathscr{H})$*, by*

$$v_{n+1} := T\left(I - \lambda_{n+1}\nabla\Theta\right)(v_n), \tag{17.35}$$

satisfies

$$0 \leq \|v_n - u^\star\| = \|T(u_n) - u^\star\| \leq \|u_n - u^\star\| \to 0 \quad (n\to\infty). \tag{17.36}$$

The formula (17.35) is regarded as a (partial) generalization of the proximal forward-backward splitting *in Example 17.12(b). Moreover, we can deduce from (17.35) a generalization [136, Remark 2.17(a)] of an algorithm in [109] (a version of projected Landweber method [13, 58, 75]) developed for the convexly constrained least-squares problems.*

(b) *If the strict convexity of* $\Theta \in \Gamma_0(\mathscr{H})$ *is assumed additionally in Theorems 17.22-II and III, the solution set becomes a singleton* $\Omega = \{u^\star\}$*. In such a case,* $\lim_{n\to\infty} d(u_n, \Omega) = 0$ *in Theorems 17.22-II and III is equivalent to*

$$\lim_{n\to\infty} \|u_n - u^\star\| = 0,$$

hence the relation (17.36) is again applicable to the sequence $(v_n)_{n\geq 0}$ *generated by (17.35), which guarantees the convergence of* $(v_n)_{n\geq 0}$ *to* $u^\star$*. In [64, Theorem 2], a similar algorithm to (17.35) is found specially for* $T = T_{\mathrm{sp}(f)}$*.*

Clearly, we can apply the hybrid steepest descent method (17.33) in Theorem 17.22 (or its alternative form (17.35)) to minimization of $\Theta : \mathscr{H} \to \mathbb{R} : x \mapsto {}^{\gamma}\Phi(x)$, which is the Moreau-Yosida regularization of a possibly nonsmooth convex function $\Phi \in \Gamma_0(\mathscr{H})$ of the index $\gamma > 0$, over the fixed point set of a certain quasi-nonexpansive mapping $T : \mathscr{H} \to \mathscr{H}$. In such a scenario, the $\frac{1}{\gamma}$-Lipschitz continuity of the gradient $\nabla\Theta : \mathscr{H} \to \mathscr{H}$ is guaranteed automatically by Fact 17.17(c), which is the only requirement for $\nabla\Theta$ in Theorem 17.22-II & III. By applying Propositions 17.3, 17.4 and 17.7, to various mappings in Examples 17.6 and 17.9, we can design many efficiently computable quasi-nonexpansive mappings as T whose fixed point set $\mathrm{Fix}(T)$ is desirable as the constraint set.

17.4 Application to Minimal Antenna-Subset Selection Problem for MIMO Communication Systems

We have proposed a promising approach by an integration of the ideas of the *hybrid steepest descent method* and the *Moreau-Yosida regularization* to the challenging nonsmooth convex optimization over the fixed point set of certain quasi-nonexpansive mappings. We present in this section its nontrivial application to a minimal antenna-subset selection problem for efficient MIMO systems (Note: The contents of this section have partially been presented in [146]).

17.4.1 Backgrounds and Motivations

Multiple antenna systems, broadly-termed MIMO (multiple-input multiple-output) systems, have given significant impacts to a wide range of research fields including communications, signal processing, and information theory because of its potential to increase the data rate without additional bandwidth [63, 124]. The gain, however, comes at the price of hardware and signal processing complexity, power consumption etc. [94]. One of the main causes for the complexity-increase is the cost of multiple RF (radio frequency) chains. Antenna selection has been considered as an attractive approach to reduce the hardware complexity without severely losing the advantages of MIMO systems (see [96, 110, 68, 55] and references therein). In particular, it has been shown that the antenna selection retains the diversity degree compared to the full-complexity system [96, 73]. The complexity reduction is achieved by equipping fewer RF chains than the antenna elements at the receiver/transmitter, and the same number of antennas as the RF chains are selected so that the achieved channel capacity is maximized.

Differently from the prior works, we consider power-limited systems in which it is desired to consume the minimum amount of power with the designated channel capacity achieved. At the receiver, for instance, each antenna element requires a 'power-consuming' RF chain that comprises a low noise amplifier, a frequency down-converter (a mixer), and an analog-to-digital converter. Also the signal processing complexity may seriously increase with the number of antenna elements.[5] Therefore, it would be a natural requirement to select the minimal antenna subset that achieves the designated channel capacity; the cardinality of such a subset depends highly on the channel state, signal to noise ratio (SNR) etc. Unfortunately, the

[5] When the multiple antennas are exploited for spatial multiplexing or the space-time trellis codes are adopted, the complexity increases sometimes exponentially [94].

problem of minimal antenna-subset selection is regarded as ℓ_0-norm [6] minimization under highly nonlinear constraint, hence it is hard to solve the problem directly because of its combinatorial nature when the number of antennas increases.

In this section, we present an alternative algorithmic solution for reaching an approximate solution by relaxing twice the ℓ_0-norm cost function in the original problem. The first relaxation is the standard ℓ_1-relaxation of ℓ_0-norm found widely in the recent approximation techniques for sparse optimization problems. Indeed, although the first relaxed problem can be handled as a convex optimization, it is still hard to solve directly due to the *nonsmoothness* of the new ℓ_1-norm cost function coupled with the *highly nonlinear* capacity-constraint. Therefore, the second relaxation is the Moreau envelope of ℓ_1-norm, which is a computationally manageable cost function under the capacity constraint.

The proposed algorithm is based on an application of Theorem 17.22-III (a version of the hybrid steepest descent method (HSDM) for the subgradient projection operator [135]) to the doubly relaxed problem: minimize the Moreau envelope of the ℓ_1-norm subject to the capacity constraint.

17.4.2 System Model and Problem Statement

For a MIMO system with N_{T} transmit antennas and N_{R} receive antennas, the received signal can be represented as

$$r_i := \sqrt{E_s} G s_i + n_i \in \mathbb{C}^{N_{\mathrm{R}}}. \tag{17.37}$$

Here, r_i represents the ith sample of the signals measured at the N_{R} receive antennas, $s_i \in \mathbb{C}^{N_{\mathrm{T}}}$ the ith symbol transmitted from the N_{T} transmit antennas, $E_s > 0$ the average energy at each receive antenna, $G \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ the channel matrix whose (p,q)th component represents the channel characteristics between the pth receive antenna and the qth transmit antenna, and n_i the additive white Gaussian noise with energy $N_0/2$ per complex dimension. We make the standard assumptions that the channel has frequency-flat fading and G is perfectly known at the receiver.[7] Also we assume that G is totally unknown at the transmitter, therefore choosing s_i such that its covariance matrix is $I_{N_{\mathrm{T}}}/N_{\mathrm{T}}$ [55]; we denote by I_m the $m \times m$ identity matrix. In this case, it is known that the channel capacity (mutual information) is given as follows [63]:

$$c_{\mathrm{full}} := \log_2 \det \left(I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H G \right) \text{ bps/Hz}, \tag{17.38}$$

where $\rho := E_s/N_0$ is the average SNR; $(\cdot)^H$ stands for the Hermitian transpose.

---

[6] The cardinality of the nonzero components in $x := (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ is often denoted by $\|x\|_0 \in \mathbb{N}$ and called commonly the ℓ_0-norm of x (or the *Hamming weight* of x in Coding Theory) although $\|\cdot\|_0$ does not satisfy either the conditions for norm or quasinorm.

[7] The channel could be moderately frequency-selective [110, 96].

We focus on the receive antenna selection. Let $\underline{c} \in (0, c_{\text{full}})$ denote the designated channel capacity to be ensured. The problem is to select the minimal antenna subset that achieves the capacity $\underline{c}$. Let $x := [x_1, x_2, \cdots, x_{N_\text{R}}]^t \in \{0,1\}^{N_\text{R}}$ represent an antenna subset in such a way that $x_j = 1$ ($x_j = 0$) indicates that the jth antenna is selected (not selected). Then, the channel capacity with the antenna subset represented by x is given by

$$c(x) := \log_2 \det \left(I_{N_\text{T}} + \frac{\rho}{N_\text{T}} G^H X G \right) \text{ bps/Hz}, \tag{17.39}$$

where $X := \text{diag}(x)$. The minimal antenna-subset selection problem is thus formulated as follows:

$$\min_{x \in \{0,1\}^{N_\text{R}}} \|x\|_0 \quad \text{s.t. } c(x) \geq \underline{c}, \tag{17.40}$$

where $\|\cdot\|_0$ denotes the ℓ_0-norm that counts the number of nonzero components. The problem in (17.40) is mathematically challenging, because it is nonlinearly-constrained sparse optimization. In general, finding its optimal solution involves exhaustive search. In the following, we present an efficient algorithmic solution using convex and differentiable relaxations of the ℓ_0 norm.

17.4.3 Convex and Differentiable Relaxations

To alleviate the difficulty in the combinatorial nature of the problem, we reformulate (17.40) into

$$\min_{x \in [0,1]^{N_\text{R}}} \psi(x) := \|x\|_1 \quad \text{s.t. } \varphi(x) := \underline{c} - c(x) \leq 0, \tag{17.41}$$

which is $\|\cdot\|_1$ minimization[8]. Because the function c is concave on $\mathbb{R}_+^{N_\text{R}}$ [55, 17], φ is convex on $\mathbb{R}_+^{N_\text{R}}$; $\mathbb{R}_+$ denotes the set of all nonnegative real numbers.

Unfortunately, we can still not find any computationally efficient solver for the reformulated problem in (17.41) because (i) the function ψ is (convex but) neither smooth nor strictly-convex and (ii) the metric projection onto the constraint set (i.e., the zero level set of φ) is not efficiently computable (For instance, the generalized Haugazeau's scheme [38] cannot be applied directly because of the non-strict-convexity of ψ). Therefore, we reformulate (17.41), in $\mathscr{H} := \mathbb{R}^{N_\text{R}}$ where the standard inner product and its induced norm are defined, by using the Moreau-Yosida regularization. Defining $\psi_\omega : \mathbb{R}^{N_\text{R}} \to \mathbb{R}_+, x \mapsto \omega \|x\|_1$, for an arbitrary constant $\omega > 0$ ($\psi = \psi_{\omega|\omega=1}$), our optimization problem to solve is given as follows:

$$\min_{x \in [0,1]^{N_\text{R}}} {}^\gamma\psi_\omega(x) \quad \text{s.t. } \varphi(x) \leq 0 \quad (\gamma > 0). \tag{17.42}$$

---

[8] In recent years, it has been proven both theoretically and experimentally that *sparse recovery* is possible in many cases by means of the ℓ_1-norm [52, 23].

17.4.4 Proposed Antenna-Subset Selection Algorithm

The key of the previous subsection is the second relaxation which replaces the nonsmooth ψ by ${}^{\gamma}\psi_{\omega}(x)$ having a *Lipschitz continuous* derivative. Our basic strategy is the following: (i) compute the solution $x^{\star}$ to the problem in (17.42) by HSDM (Theorem 17.22-III) and (ii) choose the antenna subset associated with the indices of (the minimum number of) the largest components of $x^{\star}$ such that the designated capacity $\underline{c}$ is achieved. Letting $\mathscr{I} := \{1, 2, \cdots, N_{\mathrm{R}}\}$, the proposed algorithm is given as below.

Algorithm 17.25.

(i) For an initial vector $x_0 \in \mathbb{R}^{N_{\mathrm{R}}}$, generate $(x_k)_{k=1}^{Q}$ recursively by HSDM (Q: the prespecified number of iterations), and let $x_Q =: [x_Q^{(1)}, x_Q^{(2)}, \cdots, x_Q^{(N_{\mathrm{R}})}]^t$.

(ii) Compute the arithmetic mean $\bar{x}_Q$ of x_Q.

(iii) Choose the indices corresponding to the components no smaller than $\bar{x}_Q$ as a temporary antenna subset.

Let $\mathscr{J} := \emptyset$.
for $j \in \mathscr{I}$
  if $x_Q^{(j)} \geq \bar{x}_Q$
    $\mathscr{J} := \mathscr{J} \cup \{j\}$
  end
end

(iv) Choose the minimal antenna subset.

Let $x_{\mathscr{J}} \in \{0, 1\}^{N_{\mathrm{R}}}$ be the vector representing the antenna subset $\mathscr{J}$ (see Section 17.4.2).
if $c(x_{\mathscr{J}}) < \underline{c}$
  while $c(x_{\mathscr{J}}) < \underline{c}$
    $j \in \arg\max_{\iota \in \mathscr{I} \setminus \mathscr{J}} x_Q^{(\iota)}$
    $\mathscr{J} := \mathscr{J} \cup \{j\}$
  end
else
  $\widehat{\mathscr{J}} := \mathscr{J}$
  while $c(x_{\widehat{\mathscr{J}}}) \geq \underline{c}$
    $\mathscr{J} := \widehat{\mathscr{J}}$
    $j \in \arg\min_{\iota \in \widehat{\mathscr{J}}} x_Q^{(\iota)}$
    $\widehat{\mathscr{J}} := \widehat{\mathscr{J}} \setminus \{j\}$
  end
end

(v) Output $\mathscr{J}$ as the selected antenna subset. □

The following subsection is devoted to explain precisely how to solve the problem in (17.42) by HSDM.

17.4.5 Optimization by Hybrid Steepest Descent Method

The problem in (17.42) has two constraints: the *capacity constraint*

$$x \in \mathrm{lev}_{\leq 0}(\varphi) := \left\{ x \in \mathbb{R}^{N_{\mathrm{R}}} : \varphi(x) \leq 0 \right\}$$

and the *box constraint*

$$x \in \mathscr{K} := [0,1]^{N_{\mathrm{R}}} = \left\{ x \in \mathbb{R}^{N_{\mathrm{R}}} : 0 \leq x_j \leq 1, \ \forall j \in \mathscr{I} \right\},$$

where $\mathscr{K} \cap \mathrm{lev}_{\leq 0}(\varphi) \neq \emptyset$ is confirmed by $\varphi(1_{N_{\mathrm{R}}}) = \underline{c} - c_{\mathrm{full}} < 0$ for $1_{N_{\mathrm{R}}} := [1,1,\cdots,1] \in \mathscr{K}$.

Note that $P_{\mathscr{K}}$ can be computed easily while the computation of $P_{\mathrm{lev}_{\leq 0}(\varphi)}$ is not a simple task at all. Fortunately, an application of Theorem 17.22-III to $\Theta := {}^{\gamma}\psi_{\omega}$, $f := \varphi$ and $K := \mathscr{K}$ guarantees for any $x_0 \in \mathbb{R}^{N_{\mathrm{R}}}$ that the recursion

$$x_{k+1} := (I - \lambda_{k+1} \nabla^{\gamma} \psi_{\omega}) \left(\widehat{T}_{\alpha}(x_k) \right), \ k \geq 0, \tag{17.43}$$

with

$$\widehat{T}_{\alpha} := P_{\mathscr{K}} \left[(1-\alpha) I + \alpha T_{\mathrm{sp}(\varphi)} \right], \ \alpha \in (0,2), \tag{17.44}$$

generates a sequence of points converging to a solution to (17.42).

Since φ is differentiable on $\mathbb{R}_{+}^{N_{\mathrm{R}}}$, its gradient $\nabla \varphi(x) := \left[\frac{\partial \varphi(x)}{\partial x_1}, \frac{\partial \varphi(x)}{\partial x_2}, \cdots, \frac{\partial \varphi(x)}{\partial x_{N_{\mathrm{R}}}} \right]^t$ is the unique subgradient at any $x \in \mathbb{R}_{+}^{N_{\mathrm{R}}}$; i.e., $\partial \varphi(x) = \{\nabla \varphi(x)\}$. Letting $G^H =: [g_1 \ g_2 \ \cdots g_{N_{\mathrm{R}}}]$, we have

$$I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H X G = I_{N_{\mathrm{T}}} + \sum_{j=1}^{N_{\mathrm{R}}} x_j \left(\frac{\rho}{N_{\mathrm{T}}} g_j g_j^H \right), \tag{17.45}$$

which is positive definite. Therefore, $\forall x \in \mathbb{R}_{+}^{N_{\mathrm{R}}}, \forall j \in \mathscr{I}$, we have

$$\begin{aligned} \frac{\partial \varphi(x)}{\partial x_j} &= -\frac{1}{\ln 2} \mathrm{tr} \left[\left(I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H X G \right)^{-1} \frac{\rho}{N_{\mathrm{T}}} g_j g_j^H \right] \\ &= -\frac{\rho}{N_{\mathrm{T}} \ln 2} g_j^H \left(I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H X G \right)^{-1} g_j, \end{aligned} \tag{17.46}$$

where tr $[\cdot]$ stands for the trace of matrix. Note that, since $(I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H X G)^{-1}$ is positive definite, $g_j^H \left(I_{N_{\mathrm{T}}} + \frac{\rho}{N_{\mathrm{T}}} G^H X G \right)^{-1} g_j > 0, \forall j \in \mathscr{I}$, thus $\frac{\partial \varphi(x)}{\partial x_j} < 0, \forall j \in \mathscr{I}$; $g_j \neq 0$ is silently assumed without loss of generality.

Finally, $\nabla^{\gamma} \psi_{\omega} (= \frac{1}{\gamma}(I - \mathrm{prox}_{\gamma \psi_{\omega}}))$ is computed simply by

$$\mathrm{prox}_{\gamma\psi_\omega} : \mathbb{R}^{N_\mathrm{R}} \ni x \mapsto \sum_{j=1}^{N_\mathrm{R}} \mathrm{sgn}(\langle x, e_j \rangle) \max\{|\langle x, e_j \rangle| - \gamma\omega, 0\} e_j, \tag{17.47}$$

where e_j, $j = 1, 2, \cdots, N_\mathrm{R}$, specially denotes the unit vector that has only one nonzero element at the jth position.

Remark 17.26. (On the recursion (17.43)) The operator $I - \lambda_{k+1}\nabla^\gamma \psi_\omega$ in (17.43) can be written as $I + \frac{\lambda_{k+1}}{\gamma}(\mathrm{prox}_{\gamma\psi_\omega} - I)$. From (17.47), $\mathrm{prox}_{\gamma\psi_\omega}$ attracts to zero such components of $\widehat{T}_\alpha(x_k)$ that are no greater than $\gamma\omega$. Therefore, $I - \lambda_{k+1}\nabla^\gamma \psi_\omega$ also has a similar zero-attracting function, thereby promoting the sparsity. The parameters γ and λ_{k+1} should satisfy $\lambda_{k+1}/\gamma \le 1$ so that all the components of x_{k+1} are kept nonnegative. Also γ and ω should satisfy $\gamma\omega < 1$ for preventing the situation where all the components are attracted to zero. We mention that a constant value for all λ_ks (as shown below) may be used, because the strict convergence is not necessarily required in the proposed algorithm. The computational complexity of the proposed algorithm is given approximately by $QN_{\min}^2(2N_{\max} + \underline{\mathrm{N}})$, where $N_{\min} := \min\{N_\mathrm{R}, N_\mathrm{T}\}$, $N_{\max} := \max\{N_\mathrm{R}, N_\mathrm{T}\}$, and $\underline{\mathrm{N}} := \min\{N_{\min}, N_{\max}/2\}$. Hence, the proposed algorithm is efficient particularly when N_T is sufficiently small compared to N_R. Note that there exists no other method available for the minimal antenna-subset selection problem (17.40).

Remark 17.27. (Equivalent expression of the problem (17.41)) Noting the range of x, the problem (17.41) can equivalently be formulated as follows:

$$\min_{x \in [0,1]^{N_\mathrm{R}}} \widetilde{\psi}(x) := 1_{N_\mathrm{R}}^t x \quad \text{s.t. } \varphi(x) \le 0. \tag{17.48}$$

Unfortunately, although the gradient $\nabla\widetilde{\psi}(x) = 1_{N_\mathrm{R}}$ is surely Lipschitz continuous, it is *not* possible, unlike the case of (17.42), to conclude immediately that (17.48) can be solved by applying Theorem 17.22-III for the following reason. Indeed, the HSDM recursion for (17.48) is given by

$$x_{k+1} := (I - \lambda_{k+1}\nabla\widetilde{\psi}(x))\left(\widehat{T}_\alpha(x_k)\right) = \widehat{T}_\alpha(x_k) - \lambda_{k+1}1_{N_\mathrm{R}},\ k \ge 0. \tag{17.49}$$

A simple inspection of (17.49) clarifies that λ_{k+1} should be no smaller than the minimum component of $\widehat{T}_\alpha(x_k)$ because the function φ, which is included in the operator $\widehat{T}_\alpha$, is convex only on $\mathbb{R}_+^{N_\mathrm{R}}$. Therefore, to guarantee the convergence by Theorem 17.22-III, careful design of the step size parameter λ_k is required at each iteration step.

17.4.6 Numerical Examples

Simulations are performed to show the efficacy of the proposed minimal antenna-subset selection algorithm. We consider the Rayleigh channel where the elements of

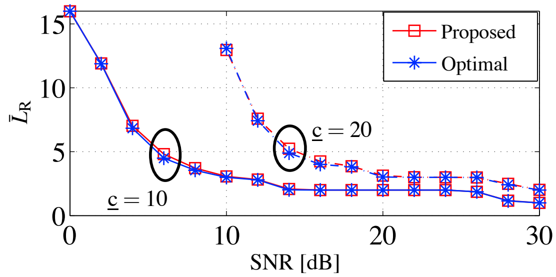
G are independently drawn from a complex zero-mean Gaussian distribution of the unit variance. In all the simulations, the HSDM parameters are set to $\alpha = 1, \gamma = 1.2$, $\omega = 0.8$, $Q = 20$, and $\lambda_k = 1$ ($\forall k = 1, 2, \cdots, Q$). In our experiments, the proposed algorithm is insensitive to the choice of the parameters within $\gamma\omega < 1$ and $\lambda_k/\gamma \leq 1$ (see Remark 17.26). All the simulated points are calculated by averaging over 2000 independent realizations of the channel matrix G.

First, Fig. 17.6 depicts the results for $N_{\mathrm{R}} = 16$, $N_{\mathrm{T}} = 4$, and $\underline{c} = 10, 20$. Figure 17.6.a describes the average number $\bar{L}_{\mathrm{R}}$ of antennas selected by the proposed algorithm. As a reference, we also plot the optimal solution to the original problem in (17.40); the optimal is computed by computationally-exhaustive full search. It is seen that the results of the proposed algorithm are comparable to the optimal; this suggests the reasonability of the relaxations introduced in Section 17.4.3. Figure 17.6.b describes the ergodic capacity of the proposed algorithm. With L_{R} denoting the number of antennas selected by the proposed algorithm, we also plot $C_{\max}(L_{\mathrm{R}})$, the maximum achievable capacity with the subset of L_{R} antennas, which is computed by exhaustive search. It is seen that the performance of the proposed algorithm is approximately the same as $C_{\max}(L_{\mathrm{R}})$; this is the side effects of the proposed algorithm. In summary, the results demonstrate that the proposed algorithm realizes (i) the near-minimal antenna subset and (ii) the near-maximum capacity achievable with the same number of antennas as selected by the algorithm.

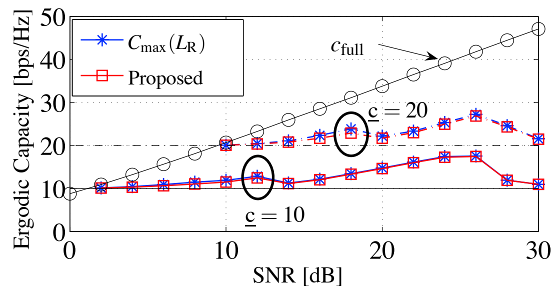
Second, Fig. 17.7 illustrates the results for (a) $N_{\mathrm{R}} = 16, N_{\mathrm{T}} = 16$, and $\underline{c} = 20, 40$ and (b) $N_{\mathrm{R}} = 64, N_{\mathrm{T}} = 16, 64$, and $\underline{c} = 60$. From Fig. 17.6.a and Fig. 17.7, it is seen that the number of antennas to be used can significantly be reduced particularly for high SNR. Moreover, in Fig. 17.7.b, we observe no distinct difference between $N_{\mathrm{T}} = 16$ and $N_{\mathrm{T}} = 64$ for SNR higher than 15 dB. Finally, Fig. 17.8 plots $\bar{L}_{\mathrm{R}}$ against N_{R} for SNR= 10 dB, $N_{\mathrm{T}} = 4$, and $\underline{c} = 20$. The result shows that an increase of the number of antenna elements equipped could yield reduction of the number of antennas used.

17.5 Concluding Remarks

In this paper, we have introduced the essence of the great applicability of the convex optimization over the fixed point set of quasi-nonexpansive mapping. First, we have shown that the fixed point characterization gives us the powerful toolbox to address the problem of finding an "optimal" point from the fixed point set. Second, we have proposed the integration of the *hybrid steepest descent method* and the *Moreau-Yosida regularization* by highlighting its distinctive properties as a smooth approximation of a nonsmooth convex function. The novel integration with the gifted toolbox has opened a path to dealing with the challenging nonsmooth convex optimization problems under the *cumbersome* constraint of the fixed point set, which are naturally desired yet have been unexplored in mathematical sciences and engineerings. We have demonstrated the effectiveness of the proposed approach in its application



(a) The average number of selected antennas $\bar{L}_R$.



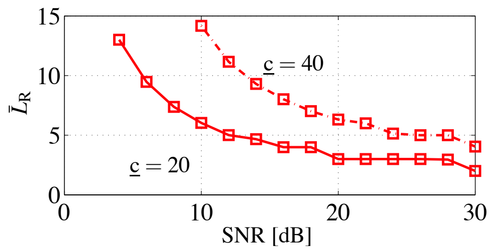
(b) Ergodic capacity.

Fig. 17.6 Comparisons with the optimal selection for $N_R = 16, N_T = 4$, and $\underline{c} = 10, 20$.

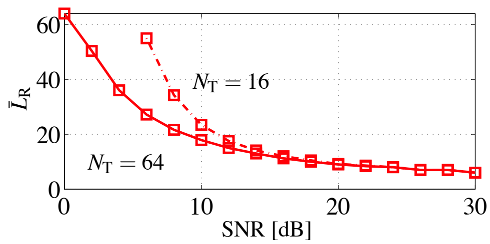
to the minimal antenna-subset selection problem under a highly nonlinear capacity constraint for efficient MIMO communication systems.

This paper has focused on the nonsmooth convex optimization problems over the fixed point set. We remark, however, that the hybrid steepest descent method has many other possible advanced applications. For example, by letting $T := \mathrm{rprox}_{\gamma f_1} \mathrm{rprox}_{\gamma i_C}$ for $f_1 \in \Gamma_0(\mathscr{H})$ and a closed convex set $C \subset \mathscr{H}$, we have the characterization: $G := \arg\min_{x \in C} f_1(x) = \{P_C(z) \mid z \in \mathrm{Fix}(T)\}$ (see Example 17.6(c)). This means that we can minimize a convex function $f_2 : C \to \mathbb{R}$ over the constraint set G by applying the hybrid steepest descent method to $\Theta : \mathscr{H} \to \mathbb{R} : x \mapsto f_2(P_C(x))$ and T provided that the derivative of Θ is Lipschitzian.

Acknowledgements The first author would like to thank Heinz Bauschke, Patrick Combettes and Russell Luke for their kind encouragement and invitation of the first author to the dream meeting: *The Interdisciplinary Workshop on Fixed-Point Algorithms for Inverse Problems in Science and Engineering* in November 1-6, 2009 at the Banff International Research Station.



(a) $N_{\mathrm{R}} = 16, N_{\mathrm{T}} = 16$, and $\underline{c} = 20, 40$.



(b) $N_{\mathrm{R}} = 64, N_{\mathrm{T}} = 16, 64$, and $\underline{c} = 60$.

Fig. 17.7 Performance for a large number of antennas.

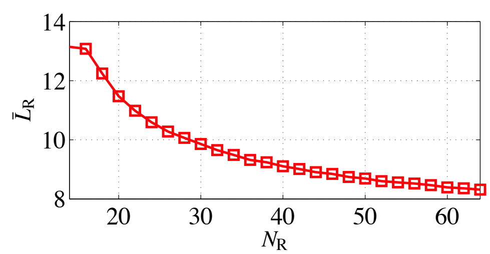


Fig. 17.8 N_{R} versus $\bar{L}_{\mathrm{R}}$ for SNR= 10 dB, $N_{\mathrm{T}} = 4$, and $\underline{c} = 20$.

References

1. Apostol, T.M.: Mathematical Analysis, 2nd ed. Addison-Wesley (1974)
2. Ascher, U.M., Haber, E., Huang, H.: On effective methods for implicit piecewise smooth surface recovery. SIAM J. Sci. Comput. **28**, 339–358 (2006)
3. Baillon, J.-B., Bruck, R.E., Reich, S.: On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. Houst. J. Math. **4**, 1–9 (1978)
4. Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et *n*-cycliquement monotones. Isr. J. Math. **26**, 137–150 (1977)
5. Barbu, V., Precupanu, Th.: Convexity and Optimization in Banach Spaces, 3rd Ed. D. Reidel Publishing Company (1986)
6. Bauschke, H.H.: The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. J. Math. Anal. Appl. **202**, 150–159 (1996)

7. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
8. Bauschke, H.H., Combettes, P.L.: A weak-to-strong convergence principle for Fejér monotone methods in Hilbert space. Math. Oper. Res. **26**, 248–264 (2001)
9. Bauschke, H.H., Combettes, P.L.: The Baillon-Haddad theorem revisited. J. Convex Anal. **17**, 781–787 (2010)
10. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2011)
11. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sciences **2**, 183–202 (2009)
12. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**, 2419–2434 (2009)
13. Bertero, M., Boccacci, P.: Introduction to Inverse Problems in Imaging. IOP (1998)
14. Borwein, J.M., Fitzpatrick, S., Vanderwerff, J.: Examples of convex functions and classifications of normed spaces. J. Convex Anal. **1**, 61–73 (1994)
15. Bougeard, M.L.: Connection between some statistical estimation criteria, lower-C_2 functions and Moreau-Yosida approximates. In: Bulletin International Statistical Institute 47th session **1**, INSEE Paris Press, pp.159–160 (1989)
16. Bougeard, M.L., Caquineau, C.D.: Parallel proximal decomposition algorithms for robust estimation. Ann. Oper. Res. **90**, 247–270 (1999)
17. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge Univ. Press (2004)
18. Bregman, L.M.: The method of successive projection for finding a common point of convex sets. Soviet Math. Dokl. **6**, 688–692 (1965)
19. Browder, F.E.: Convergence theorems for sequences of nonlinear operators in Banach spaces. Math. Z. **100**, 201–225 (1967)
20. Byrne, C.L.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Probl. **20**, 103–120 (2004)
21. Byrne, C.L.: Applied Iterative Methods. A K Peters, Ltd., Wellesley, Massachusettes (2007)
22. Cai, J.F., Candés, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**, 1956–1982 (2010)
23. Candés, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**, 21–30 (2008)
24. Capel, D., Zisserman, A.: Computer vision applied to super resolution. IEEE Signal Process. Mag. **20**, 75–86 (2003)
25. Cavalcante, R., Yamada, I.: Multiaccess interference suppression in orthogonal space-time block coded MIMO systems by adaptive projected subgradient method. IEEE Trans. Signal Process. **56**, 1028–1042 (2008)
26. Cavalcante, R., Yamada, I.: A flexible peak-to-average power ratio reduction scheme for OFDM systems by the adaptive projected subgradient method. IEEE Trans. Signal Process. **57**, 1456–1468 (2009)
27. Cavalcante, R., Yamada, I., Mulgrew, B.: An adaptive projected subgradient approach to learning in diffusion networks. IEEE Trans. Signal Process. **57**, 2762–2774 (2009)
28. Censor, Y., Iusem, A.N., Zenios, S.A.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Math. Program. **81**, 373–400 (1998)
29. Censor, Y., Reich, S.: Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization. Optimization **37**, 323–339 (1996)
30. Censor, Y., Zenios, S.A.: Parallel Optimization: Theory, Algorithm, and Optimization. Oxford University Press (1997)
31. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
32. Chambolle, A., DeVore, R.A., Lee, N.Y., Lucier, B.J.: Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. IEEE Trans. Image Process. **7**, 319–335 (1998)

33. Chidume, C.: Geometric Properties of Banach Spaces and Nonlinear Iterations (Chapter 7: Hybrid steepest descent method for variational inequalities). Lecture Notes in Mathematics **1965**, Springer (2009)
34. Combettes, P.L.: Foundation of set theoretic estimation. Proc. IEEE. **81**, 182–208 (1993)
35. Combettes, P.L.: Inconsistent signal feasibility problems: least squares solutions in a product space. IEEE Trans. Signal Process. **42**, 2955–2966 (1994)
36. Combettes, P.L.: Construction d'un point fixe commun à une famille de contractions fermes. C.R. Acad. Sci.Paris Sèr. I Math. **320**, 1385–1390 (1995)
37. Combettes, P.L.: Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. IEEE Trans. Image Process. **6**, 493–506 (1997)
38. Combettes, P.L.: Strong convergence of block-iterative outer approximation methods for convex optimization. SIAM J. Control Optim. **38**, 538–565 (2000)
39. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
40. Combettes, P.L., Bondon, P.: Hard-constrained inconsistent signal feasibility problems. IEEE Trans. Signal Process. **47**, 2460–2468 (1999)
41. Combettes, P.L., Pesquet, J.-C.: Image restoration subject to a total variation constraint. IEEE Trans. Image Process. **13**, 1213–1222 (2004)
42. Combettes, P.L., Pesquet, J.-C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Sel. Top. Signal Process. **1**, 564–574 (2007)
43. Combettes, P.L., Pesquet, J.-C.: A proximal decomposition method for solving convex variational inverse problems. Inverse Probl. **24** (2008)
44. Combettes, P.L., Pesquet, J.-C.: Split convex minimization algorithm for signal recovery. Proc. 2009 IEEE ICASSP (Taipei), 685–688 (2009)
45. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, H. Wolkowicz (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer-Verlag (2010)
46. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. SIAM Multiscale Model. Simul. **4**, 1168–1200 (2005)
47. Daubechies, I., Defrise, M., Mol, C.D.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm. Pure Appl. Math. **57**, 1413–1457 (2004)
48. Deutsch, F., Best Approximation in Inner Product Spaces. Springer-Verlag, New York (2001)
49. Deutsch, F., Yamada, I.: Minimizing certain convex functions over the intersection of the fixed point sets of nonexpansive mappings. Numer. Funct. Anal. Optim. **19**, 33–56 (1998)
50. Dolidze, Z.O.: Solutions of variational inequalities associated with a class of monotone maps. Ekonomika i Matem. Metody **18**, 925–927 (1982)
51. Donoho, D.L.: De-noising by soft-thresholding. IEEE Trans. Inf. Theory **41**, 613–627 (1995)
52. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**, 1289–1306 (2006)
53. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation via wavelet shrinkage. Biometrika **81**, 425–455 (1994)
54. Dotson, Jr, W.G.: On the Mann iterative process. Trans. Amer. Math. Soc. **149**, 65–73 (1970)
55. Dua, A., Medepalli, K., Paulraj, A.J.: Receive antenna selection in MIMO systems using convex optimization. IEEE Trans. Wirel. Commun. **5**, 2353–2357 (2006)
56. Dunn, J.C.: Convexity, monotonicity, and gradient processes. J. Math. Anal. Appl. **53**, 145–158 (1976)
57. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachfold splitting method and proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
58. Eicke, B.: Iteration methods for convexly constrained ill-posed problems in Hilbert space. Numer. Funct. Anal. Optim. **13**, 413–429 (1992)
59. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. Classics in Applied Mathematics **28**, SIAM (1999)
60. Elsner, L., Koltracht, L., Neumann, M.: Convergence of sequential and asynchronous nonlinear paracontractions. Numer. Math. **62**, 305–319 (1992)

61. Engle, H.W., Leitão, A.: A Mann iterative regularization method for elliptic Cauchy problems. Numer. Funct. Anal. Optim. **22**, 861–884 (2001)
62. Fadili, M.J., Starck, J.-L.: Monotone operator splitting for optimization problems in sparse recovery. Proc. 2009 IEEE ICIP, Cailo (2009)
63. Foschini, G.J., Gans, M.J.: On limits of wireless communications in a fading environment when using multiple antennas. Wirel. Pers. Commun. **6**, 311–335 (1998)
64. Fukushima, M.: A relaxed projection method for variational inequalities. Math. Program. **35**, 58–70 (1986)
65. Fukushima, M., Qi, L.: A globally and superlinearly convergent algorithm for nonsmooth convex minimization. SIAM J. Optim. **6**, 1106–1120 (1996)
66. Gabay, D.: Applications of the method of multipliers to variational inequalities. In : M. Fortin and R. Glowinski (eds.) Augmented Lagrangian Methods: Applications to the solution of boundary value problems, North-Holland, Amsterdam (1983)
67. Gandy, S., Yamada, I.: Convex optimization techniques for the efficient recovery of a sparsely corrupted low-rank matrix. Journal of Math-for-Industry **2**(2010B-5), 147–156 (2010)
68. Gharavi-Alkhansari, M., Gershman, A.B.: Fast antenna subset selection in MIMO systems. IEEE Trans. Signal Process. **52**, 339–347, (2004)
69. Goebel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory. Cambridge Univ. Press. (1990)
70. Goebel, K., Reich, S.: Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings. New York and Basel Dekker (1984)
71. Goldstein, A.A.: Convex programming in Hilbert space. Bull. Amer. Math. Soc. **70**, 709–710 (1964)
72. Golshtein, E.G., Tretyakov, N.V.: Modified Lagrangians & Monotone Maps in Optimization. Wiley and Sons (1996)
73. Gorokhov, A., Gore, D.A., Paulraj, A.J.: Receive antenna selection for MIMO spatial multiplexing: theory and algorithms. IEEE Trans. Signal Process. **51**, 2796–2807 (2003)
74. Groetsch, C.W.: A note on segmenting Mann iterates. J. Math. Anal. Appl. **40**, 369–372 (1972)
75. Groetsch, C.W.: Inverse Problems in Mathematical Sciences. Wiesbaden-Vieweg (1993)
76. Gubin, L.G., Polyak, B.T., Raik, E.V.: The method of projections for finding the common point of convex sets. USSR Comput. Maths. Phys. **7**, 1–24 (1967)
77. Halpern, B.: Fixed points of nonexpanding maps. Bull. Amer. Math. Soc. **73**, 957–961 (1967)
78. Hasegawa, H., Ohtsuka, T., Yamada, I., Sakaniwa, K.: An edge-preserving super-precision for simultaneous enhancement of spacial and grayscale resolutions. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E91-A**, 673–681 (2008)
79. Haugazeau, Y.: Sur les Inéquations variationnelles et la Minimisation de Fonctionnelles Convexes. Thèse, Universite de Paris (1968)
80. Haykin, S.: Adaptive Filter Theory, 4th edn. Prentice Hall (2002)
81. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer (1993)
82. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Statist. **35**, 73–101 (1964)
83. Iemoto, S., Takahashi, W.: Strong convergence theorems by a hybrid steepest descent method for countable nonexpansive mappings in Hilbert spaces. Sci. Math. Jpn. **69** (online: 2008-49), 227–240 (2009)
84. Kiwiel, K.C.: Block-iterative surrogate projection methods for convex feasibility problems. Linear Alg. Appl. **215**, 225–259 (1995)
85. Kreyszig, E.: Introductory Functional Analysis with Applications. Wiley Classics Library, Wiley, New York (1989)
86. Levitin, E.S., Polyak, B.T.: Constrained minimization method. USSR Comput. Maths. Phys. **6**, 1–50 (1966)
87. Li, W., Swetits, J.J.: The linear ℓ_1 estimator and the Huber M-estimator. SIAM J. Optim. **8**, 457–475 (1998)
88. Lions, P.L.: Approximation de points fixes de contractions. C. R. Acad. Sci. Paris Sèrie A-B **284**, 1357–1359 (1977)
89. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)

90. Liu, F., Nashed, M.Z.: Regularization of nonlinear ill-posed variational inequalities and convergence rates. Set-Valued Anal. **6**, 313–344 (1998)
91. Mainge, P.E.: Extension of the hybrid steepest descent method to a class of variational inequalities and fixed point problems with nonself-mappings. Numer. Funct. Anal. Optim. **29**, 820–834 (2008)
92. Mangasarian, O.L., Muscicant, D.R.: Robust linear and support vector regression. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 950–955 (2000)
93. Mann, W.: Mean value methods in iteration. Proc. Amer. Math. Soc. **4**, 506–510 (1953)
94. Mehta, N.B., Molisch, A.F.: MIMO System Technology for Wireless Communications, chapter 6, CRC Press (2006)
95. Michelot, C., Bougeard, M.L.: Duality results and proximal solutions of the Huber M-estimator problem. Appl. Math. Optim. **30**, 203–221 (1994)
96. Molisch, A.F., Win, M.Z.: MIMO systems with antenna selection. IEEE Microw. Mag. **5**, 46–56 (2004)
97. Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Ser. A Math. **255**, 2897–2899 (1962)
98. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
99. Nikolova, M.: Minimizing of cost functions involving nonsmooth data-fidelity terms — Application to the processing of outliers. SIAM J. Numer. Anal. **40**, 965–994 (2002)
100. Ogura, N., Yamada, I.: Non-strictly convex minimization over the fixed point set of the asymptotically shrinking nonexpansive mapping. Numer. Funct. Anal. Optim. **23**, 113–137 (2002)
101. Ogura, N., Yamada, I.: Non-strictly convex minimization over the bounded fixed point set of nonexpansive mapping. Numer. Funct. Anal. Optim. **24**, 129-135 (2003)
102. Ogura, N., Yamada, I.: A deep outer approximating half space of the level set of certain Quadratic Functions. J. Nonlinear Convex Anal. **6**, 187–201 (2005)
103. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**, 383–390 (1979)
104. Pierra, G.: Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique. Lecture Notes in Computer Science **41**, 200–218, Springer (1976)
105. Pierra, G.: Decomposition through formalization in a product space. Math. Program. **28**, 96–115 (1984)
106. Polyak, B.T.: Minimization of unsmooth functionals. USSR Comput. Maths. Phys. **9**, 14–29 (1969)
107. Rockafellar, R.T.: Monotone operators and proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
108. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, 1st edn. Springer (1998)
109. Sabharwal, A., Potter, L.C.: Convexly constrained linear inverse problems: Iterative least-squares and regularization. IEEE Trans. Signal Process. **46**, 2345–2352 (1998)
110. Sanayei, S., Nosratinia, A.: Antenna selection in MIMO systems. IEEE Commun. Mag. **42**, 68–73 (2004)
111. Sayed, A.H.: Fundamentals of Adaptive Filtering. Wiley-IEEE Press (2003)
112. Slavakis, K., Theodoridis, S., Yamada, I.: Online kernel-based classification using adaptive projection algorithms. IEEE Trans. Signal Process. **56**, 2781–2796 (2008)
113. Slavakis, K., Theodoridis, S., Yamada, I.: Adaptive constrained filtering in reproducing kernel Hilbert spaces: the beamforming case. IEEE Trans. Signal Process. **57**, 4744–4764 (2009)
114. Slavakis, K., Yamada, I.: Robust wideband beamforming by the hybrid steepest descent method. IEEE Trans. Signal Process. **55**, 4511–4522 (2007)
115. Slavakis, K., Yamada, I., Ogura, N.: The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings. Numer. Funct. Anal. Optim. **27**, 905–930 (2006)
116. Starck, J.-L., Murtagh, F.: Astronomical Image and Data Analysis, 2nd.edn. Springer (2006)
117. Starck, J.-L., Murtagh, F., Fadili, J.M.: Sparse Image and Signal Processing — Wavelets, Curvelets, Morphological Diversity. Cambridge Univ. Press (2010)

118. Stark, H., Yang, Y.: Vector Space Projections – A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics. John Wiley & Sons (1998)
119. Suzuki, T.: A sufficient and necessary condition for Halpern-type strong convergence to fixed points of nonexpansive mappings. Proc. Amer. Math. Soc. **135**, 99–106 (2007)
120. Takahashi, N., Yamada, I.: Parallel algorithms for variational inequalities over the Cartesian product of the intersections of the fixed point sets of nonexpansive mappings. J. Approx. Theory **153**, 139–160 (2008)
121. Takahashi, N., Yamada, I.: Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument. IEEE Trans. Signal Process. **57**, 3361–3372 (2009)
122. Takahashi, W.: Nonlinear Functional Analysis— Fixed Point Theory and its Applications. Yokohama Publishers (2000)
123. Theodoridis, S., Slavakis, K., Yamada, I.: Adaptive learning in a world of projections — A unifying framework for linear and nonlinear classification and regression tasks. IEEE Signal Processing Mag. **28**, (2011)
124. Telatar, I.E.: Capacity of multi-antenna Gaussian channels. Eur. Trans. Telecomm. **10**, 585–595 (1999)
125. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**, 119–138 (1991)
126. Vasin, V.V., Ageev, A.L.: Ill-Posed Problems with A Priori Information. VSP (1995)
127. Widrow, B., Stearns, S.D.: Adaptive Signal Processing. Prentice Hall (1985)
128. Wittmann, R.: Approximation of fixed points of nonexpansive mappings. Arch. Math. **58**, 486–491 (1992)
129. Xu, H.K., Kim, T.H.: Convergence of hybrid steepest descent methods for variational inequalities. J. Optim. Theory Appl. **119**, 185-201 (2003)
130. Yamada, I.: Approximation of convexly constrained pseudoinverse by Hybrid Steepest Descent Method. Proc. 1999 IEEE ISCAS, Florida (1999)
131. Yamada, I.: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: D. Butnariu, Y. Censor, S. Reich (eds.) Inherently Parallel Algorithm for Feasibility and Optimization and Their Applications, Elsevier, 473–504 (2001)
132. Yamada, I.: Adaptive projected subgradient method: A unified view for projection based adaptive algorithms. The Journal of IEICE **86**, 654–658 (2003) (in Japanese)
133. Yamada, I.: Kougaku no Tameno Kansu Kaiseki (Functional Analysis for Engineering), Suurikougaku-Sha / Saiensu-Sha (2009)
134. Yamada, I., Ogura, N.: Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions. Numer. Funct. Anal. Optim. **25**, 593–617 (2004)
135. Yamada, I., Ogura, N.: Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings. Numer. Funct. Anal. Optim. **25**, 619–655 (2004)
136. Yamada, I., Ogura, N., Shirakawa, N.: A numerically robust hybrid steepest descent method for the convexly constrained generalized inverse problems. In: Z. Nashed, O. Scherzer (eds.) Inverse Problems, Image Analysis, and Medical Imaging, *Contemporary Mathematics* **313**, 269–305 (2002)
137. Yamada, I., Ogura, N., Yamashita, Y., Sakaniwa, K.: An extension of optimal fixed point theorem for nonexpansive operator and its application to set theoretic signal estimation. Technical Report of IEICE **DSP96-106**, 63–70 (1996)
138. Yamada, I., Ogura, N., Yamashita, Y., Sakaniwa, K.: Quadratic optimization of fixed points of nonexpansive mappings in Hilbert space. Numer. Funct. Anal. Optim. **19**, 165–190 (1998)
139. Yamada, I., Slavakis, K., Yamada, K.: An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques. IEEE Trans. Signal Process. **50**, 1091–1101 (2002)
140. Yamagishi, M., Yamada, I.: A deep monotone approximation operator based on the best quadratic lower bound of convex functions. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E91-A**, 1858–1866 (2008)

141. Yosida, K.: Functional Analysis, 4th edn. Springer (1974)
142. Youla, D.C., Webb, H.: Image restoration by the method of convex projections: Part 1– Theory. IEEE Trans. Med. Imaging **1**, 81–94 (1982)
143. Yukawa, M., Cavalcante, R., Yamada, I.: Efficient blind MAI suppression in DS/CDMA by embedded constraint parallel projection techniques. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E88-A**, 2427–2435 (2005)
144. Yukawa, M., Slavakis, K., Yamada, I.: Adaptive parallel quadratic-metric projection algorithms. IEEE Trans. Audio Speech Lang. Process. **15**, 1665–1680 (2007)
145. Yukawa, M., Yamada, I.: Pairwise optimal weight realization — acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm. IEEE Trans. Signal Process. **54**, 4557–4571 (2006)
146. Yukawa, M., Yamada, I.: Minimal antenna-subset selection under capacity constraint for power-efficient MIMO systems: a relaxed ℓ_1-minimization approach. Proc. 2010 IEEE ICASSP, Dallas (2010)
147. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific (2002)
148. Zeidler, E.: Nonlinear Functional Analysis and its Applications, II/B - Nonlinear Monotone Operators. Springer (1990)
149. Zeidler, E.: Nonlinear Functional Analysis and its Applications, III - Variational Methods and Optimization. Springer (1985)
150. Zeng, L.C., Schaible, S., Yao, J.C.: Hybrid steepest descent methods for zeros of nonlinear operators with applications to variational inequalities. J. Optim. Theory Appl. **141**, 75–91 (2009)
151. Zhang, B., Fadili, J.M., Starck, J.-L.: Wavelet, ridgelet, and curvelets for Poisson noise removal. IEEE Trans. Image Process. **17**, 1093–1108 (2008)