SIGNAL RECOVERY BY MINIMIZING THE MOREAU ENVELOPE OVER THE FIXED POINT SET OF NONEXPANSIVE MAPPINGS

Isao Yamada and Shunsuke Ono

Department of Communications and Computer Engineering, Tokyo Institute of Technology

ABSTRACT

Many inverse problems in signal processing have been reduced to a certain nonsmooth convex optimization problem of which the solution set is a closed convex set in a real Hilbert space. This fact indicates that each point in the solution set may differ too largely in some aspects to be ignored. However, almost all existing algorithms designed specially for such a convex optimization problem converge to only an unspecified point in the solution set and cannot pick up a best one, in some additional strategic criterion, among the set. In this paper, we present a flexible utilization of the hybrid steepest descent method for the strategic point selection among the solution set of nonsmooth convex optimization problems. The proposed method can minimize sum of the Moreau envelope of nonsmooth convex functions over the fixed point set of a variety of nonexpansive mappings (e.g., Proximal Forward-Backward splitting operator, Douglas-Rachford splitting operator and Primal-Dual splitting operator) that characterize precisely the solution sets of the nonsmooth convex optimization problems.

Index Terms— Signal recovery, Convex Optimization, Moreau envelope, Hybrid steepest descent method

1. INTRODUCTION

The goal of classical estimation in signal processing has been to find a unique estimate by minimizing a certain objective function designed heavily based on the solvability of the resulting optimization problem, which may lead to unreliable solutions inconsistent with the observed data and a priori knowledge. On the other hand, the set theoretic estimation [4] aims to characterize the set of all acceptable estimates that are best consistent with all available information arising from the observed data and a priori knowledge. In a simplest case where each piece of available information is associated with a set in a common solution space and these sets are consistent, the set of all acceptable estimates is given as the intersection of these sets. Unfortunately, in many practical situations, the translation of each piece of information into a set in a common solution space is not simple task. Moreover, such pieces of information become often inconsistent mainly due to the unavoidable uncertainties caused, e.g., by the model mismatch and the randomness of the noise process, etc. To cope with such general situations, it is natural to define a generalized feasible set [5] that is the set of all points satisfying maximally a certain acceptable consistency measure for all available information, and hence the generalized set theoretic estimation problems are automatically reduced to optimization problems.

In the last decade, with the remarkable advance of computational algorithms for minimization of nonsmooth convex functions, significant progress has been made in inverse problems and signal processing [3, 6, 7, 8, 10, 11]. The set of all minimizers of a nonstrictly convex function is a closed convex set which is in general an infinite point set. Although this set can be interpreted as a generalized convex feasible set [13] in a wider sense, we call this set *the 1st stage solution set* in this paper. In the scenarios of signal processing, each point in the set may differ too largely in some aspects to be ignored.

This situation demands strongly for a certain strategic selection of a point exactly among the 1st stage solution set. Note that, if we minimize the sum of the above nonstrictly convex function and a certain new convex function introduced for a strategic selection, such a minimizer has no longer guarantee to be an element of the 1st stage solution set, and hence it is often hard to be accepted, as an improvement from the elements in the set, e.g., for the purpose of the medical diagnostic imaging.

An ideal strategic selection is realized if we have some algorithmic solution to a hierarchical convex optimization, i.e., minimization of an additionally introduced convex function, called the 2nd stage cost function, over the 1st stage solution set. Unfortunately, almost all existing convex optimization schemes cannot be applied to such a hierarchical convex optimization because these schemes converge to only an unspecified point in the 1st stage solution set.

In this paper, observing the fact that many existing algorithms [1, 6, 7, 8, 10, 11] for the nonsmooth convex optimization have been established based on elegant translations of convex optimization problems into certain fixed point problems of nonexpansive mappings designed in terms of the proximity operators in a real Hilbert space (see, e.g., Proximal forward backward splitting in Fact 3(a), Douglas-Rachford splitting in Fact 3(b) and Primal-Dual splitting in Fact 4), we demonstrate that such a hierarchical convex optimization can be solved simply by applying *the hybrid steepest descent method* [13, 14, 15] to the nonexpansive mapping and the gradient of *the Moreau envelope of nonsmooth convex functions*. Moreover, to show the effectiveness of the proposed strategic selection, we present an application of the proposed strategy to an image restoration problem by integrating the hybrid steepest descent method and a Primal-Dual splitting [8].

2. CONVEX OPTIMIZATION AND FIXED POINT OF NONEXPANSIVE MAPPINGS

In the following, we list minimum notions in convex analysis and fixed point theory of nonexpansive mapping, which are necessary for our discussion (See, e.g., [1, 15] for detailed account on these notions). Let $\mathcal{H}$ be a real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$.

This work is supported in part by JSPS Grants-in-Aids (B-21300091) and (24・2522).

2.1. Tools for Convex Optimization

Fact 1 (Elements of Convex Analysis)

(a) (Convex set) A set $C \subset \mathcal{H}$ 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 $C \subset \mathcal{H}$ is closed as well as convex, it is called closed convex.

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

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

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

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

for all $\boldsymbol{x} \neq \boldsymbol{y}$ and all $\lambda \in (0,1)$.

(c) (Proper lower semicontinuous convex function) A function $f : \mathcal{H} \to (-\infty, \infty]$ is called lower semicontinuous if the set $\mathrm{lev}_{\leq \alpha}(f) := \{\boldsymbol{x} \in \mathcal{H} \mid f(\boldsymbol{x}) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$ (Note: If f is continuous over $\mathcal{H}$, f is lower semicontinuous). A function $f : \mathcal{H} \to (-\infty, \infty]$ is called proper if

$$\mathrm{dom}(f) := \{\boldsymbol{x} \in \mathcal{H} \mid f(\boldsymbol{x}) < \infty\} \neq \emptyset.$$

The set of all proper lower semicontinuous convex functions is denoted by $\Gamma_0(\mathcal{H})$.

(d) (Proximity operator and Moreau envelope) The proximity operator of index $\gamma \in (0,\infty)$ of $f \in \Gamma_0(\mathcal{H})$ is defined by

$$\mathrm{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H} : \boldsymbol{x} \mapsto \arg\min_{\boldsymbol{y} \in \mathcal{H}} \left[f(\boldsymbol{y}) + \frac{1}{2\gamma}\|\boldsymbol{x} - \boldsymbol{y}\|^2 \right],$$

where the existence and the uniqueness of the minimizer are guaranteed. The function

$$\begin{aligned} {}^{\gamma}f : \mathcal{H} \to \mathbb{R} : \boldsymbol{x} &\mapsto \min_{\boldsymbol{y} \in \mathcal{H}} \left[f(\boldsymbol{y}) + \frac{1}{2\gamma}\|\boldsymbol{x} - \boldsymbol{y}\|^2 \right] \\ &= f\left(\mathrm{prox}_{\gamma f}(\boldsymbol{x})\right) + \frac{1}{2\gamma}\left\|\boldsymbol{x} - \mathrm{prox}_{\gamma f}(\boldsymbol{x})\right\|^2 \end{aligned}$$

is called the *Moreau envelope* of index $\gamma \in (0,\infty)$. ${}^{\gamma}f$ satisfies (i) $f(\boldsymbol{x}) \geq {}^{\gamma}f(\boldsymbol{x})$ ($\forall \gamma \in (0,\infty)$, $\forall \boldsymbol{x} \in \mathcal{H}$), (ii) $\lim_{\gamma \downarrow 0} {}^{\gamma}f(\boldsymbol{x}) = f(\boldsymbol{x})$ ($\forall \boldsymbol{x} \in \mathrm{dom}(f)$) and (iii) ${}^{\gamma}f$ is Gâteaux differentiable[1] and its derivative is given by

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

and hence $\nabla {}^{\gamma}f : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\gamma}$-Lipschitzian.

(e) (Fenchel-Rockafellar conjugate) For any $f \in \Gamma_0(\mathcal{H})$, the function $f^* : \mathcal{H} \ni \boldsymbol{y} \mapsto \sup_{\boldsymbol{x} \in \mathcal{H}} (\langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x})) \in (-\infty, \infty]$ satisfies $f^* \in \Gamma_0(\mathcal{H})$. f^* is called the Fenchel-Rockafellar conjugate of f. For any $\gamma \in (0,\infty)$, the proximity operators $\mathrm{prox}_{\gamma f}$ and $\mathrm{prox}_{\gamma^{-1} f^*}$ are related by

$$I = \mathrm{prox}_{\gamma f} + \gamma \mathrm{prox}_{\gamma^{-1} f^*} \circ \gamma^{-1} I,$$

where I stands for the identity operator.

[1] A function $g : \mathcal{H} \to \mathbb{R}$ is called Gâteaux differentiable at $x \in \mathcal{H}$ if there exists $a(x) \in \mathcal{H}$ such that $\lim_{\delta \to 0} \frac{g(x+\delta h) - g(x)}{\delta} = \langle a(x), h \rangle$ ($\forall h \in \mathcal{H}$). In this case, $\nabla g(x) := a(x)$ is called Gâteaux derivative (or gradient) of g at x.

Example 1 (Proximity Operators)

(a) (A principle of splitting) For multiple real Hilbert spaces

$$(\mathcal{X}_j, \langle \cdot, \cdot \rangle_j, \|\cdot\|_j) \quad (j = 1, \ldots, p),$$

define the standard product space

$$(\mathcal{H} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_p, \langle \cdot, \cdot \rangle, \|\cdot\|)$$

which is a real Hilbert space equipped with the inner product $\langle (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p), (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p) \rangle := \sum_{j=1}^p \langle \boldsymbol{x}_j, \boldsymbol{y}_j \rangle_j$ and its induced norm $\|\cdot\|$. Suppose that a function $f \in \Gamma_0(\mathcal{H})$ is defined by $f := \sum_{j=1}^p \varphi_j \in \Gamma_0(\mathcal{H})$ with $\varphi_j \in \Gamma_0(\mathcal{X}_j)$ $(j = 1, \ldots, p)$. Then we have

$$\mathrm{prox}_{\gamma f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) = \left(\mathrm{prox}_{\gamma \varphi_1}(\boldsymbol{x}_1), \ldots, \mathrm{prox}_{\gamma \varphi_p}(\boldsymbol{x}_p) \right).$$

(b) (Metric projection) For a given nonempty closed convex set $C \subset \mathcal{H}$ and $\boldsymbol{x} \in \mathcal{H}$, there exists a unique $P_C(\boldsymbol{x}) \in C$ satisfying

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

The mapping which assigns $\boldsymbol{x} \in \mathcal{H}$ to $P_C(\boldsymbol{x}) \in C$ is called the metric projection onto C. The metric projection P_C satisfies $P_C(\boldsymbol{x}) = \mathrm{prox}_{\gamma i_C}(\boldsymbol{x})$ ($\forall \gamma \in (0,\infty), \forall \boldsymbol{x} \in \mathcal{H}$), where

$$i_C : \mathcal{H} \ni \boldsymbol{x} \mapsto \begin{cases} 0 & \text{if } \boldsymbol{x} \in C \\ \infty & \text{if } \boldsymbol{x} \notin C. \end{cases}$$

(c) (Soft-thresholding operator [9, 6]) For a given nonempty closed convex set $C \subset \mathcal{H}$ and $\gamma \in (0,\infty)$,

$$\begin{aligned} &\mathrm{prox}_{\gamma d_C}(\boldsymbol{x}) \\ = &\begin{cases} \boldsymbol{x} + \gamma(P_C(\boldsymbol{x}) - \boldsymbol{x})/d_C(\boldsymbol{x}) & \text{if } d_C(\boldsymbol{x}) > \gamma \\ P_C(\boldsymbol{x}) & \text{otherwise.} \end{cases} \end{aligned}$$

2.2. Translation of Convex Optimization Problem into Fixed Point Problem

Fact 2 *(Nonexpansive Mapping and Fixed points [1, 15])*

(a) A mapping $T : \mathcal{H} \to \mathcal{H}$ is called nonexpansive if $\|T(\boldsymbol{x}) - T(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$ ($\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$). If the fixed point set $\mathrm{Fix}(T) := \{\boldsymbol{x} \in \mathcal{H} \mid T(\boldsymbol{x}) = \boldsymbol{x}\}$ is nonempty, $\mathrm{Fix}(T)$ is guaranteed to be closed convex in $\mathcal{H}$. In particular, if there exist $\alpha \in [0,1)$ and a nonexpansive mapping $N : \mathcal{H} \to \mathcal{H}$ satisfying $T = (1-\alpha)I + \alpha N$, T is called α-averaged nonexpansive.

(b) If $T_i : \mathcal{H} \to \mathcal{H}$ ($i = 1,2$) are α_i-averaged nonexpansive, $T := T_1 T_2$ is $\frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}$-averaged nonexpansive. Moreover, for any nonexpansive mapping T satisfying $\mathrm{Fix}(T) \neq \emptyset$, T is α-averaged for $\alpha \in (0,1)$ if and only if T is $\alpha^{-1}(1-\alpha)$-strongly attracting, i.e.,

$$(\forall \boldsymbol{x} \in \mathcal{H}, \forall \boldsymbol{z} \in \mathrm{Fix}(T))$$
$$\alpha^{-1}(1-\alpha)\|\boldsymbol{x} - T(\boldsymbol{x})\|^2 \leq \|\boldsymbol{x} - \boldsymbol{z}\|^2 - \|T(\boldsymbol{x}) - \boldsymbol{z}\|^2.$$

(c) For any $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in (0,\infty)$, $\mathrm{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$ is 1/2-averaged, and hence $\nabla({}^{\gamma}f) : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\gamma}$-Lipschitzian.

Fact 3 *(Fixed Point Characterization I [1, 15])*
Suppose that $f_1, f_2 \in \Gamma_0(\mathcal{H})$ *satisfy*

$$S := \arg\min_{\boldsymbol{x}\in\mathcal{H}} [f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})] \neq \emptyset. \tag{2}$$

(a) *(Proximal forward-backward splitting operator) Suppose that* f_1 *is Gâteaux differentiable on* $\mathcal{H}$ *with its gradient* $\nabla f_1 : \mathcal{H} \to \mathcal{H}$. *Then* $\boldsymbol{x}^\star \in \mathcal{H}$ *satisfies* $\boldsymbol{x}^\star \in S$ *if and only if* $\boldsymbol{x}^\star \in \mathcal{H}$ *is a fixed point of the* proximal forward-backward splitting operator *:* $\mathrm{prox}_{\mu f_2}(I - \mu\nabla f_1)$ *for any* $\mu > 0$*, i.e.,* $\boldsymbol{x}^\star = \mathrm{prox}_{\mu f_2}(I - \mu\nabla f_1)(\boldsymbol{x}^\star)$. *If in addition* ∇f_1 *is* κ*-Lipschitzian for some* $\kappa > 0$*, the proximal forward-backward splitting operator* $\mathrm{prox}_{\mu f_2}(I - \mu\nabla f_1)$ *with* $\mu \in \left(0, \frac{2}{\kappa}\right]$ *is nonexpansive. Moreover, this operator is* $\frac{1}{2-\gamma}$*-averaged nonexpansive if* $\mu \in \left(0, \frac{2\gamma}{\kappa}\right] \subset \left(0, \frac{2}{\kappa}\right)$ *In particular, setting* $f_2 := i_C$ *for a closed convex set* $C \subset \mathcal{H}$ *reproduces the characterization of the minimizers of* f_1 *over* C *by the fixed point set of the* $\frac{1}{2-\gamma}$*-averaged nonexpansive mapping* $P_C(I - \mu\nabla f_1)$ *for* $\mu \in \left(0, \frac{2\gamma}{\kappa}\right] \subset \left(0, \frac{2}{\kappa}\right)$.

(b) *(Douglas-Rachford splitting operator) Suppose that*

$$\bigcup_{\lambda>0} \{\lambda\boldsymbol{x} \mid \boldsymbol{x} \in \mathrm{dom}(f_1) - \mathrm{dom}(f_2)\}$$

is a closed subspace of $\mathcal{H}$. *Then by using nonexpansive mappings* $\mathrm{rprox}_{\gamma f_i} := 2\mathrm{prox}_{\gamma f_i} - I$ *(*$i = 1, 2$*) and* $\mathrm{rprox}_{\gamma f_1}\mathrm{rprox}_{\gamma f_2}$ *for* $\gamma \in (0, \infty)$*, we have*

$$\begin{aligned} &\boldsymbol{x}^\star \in \mathcal{H} \text{ minimizes } f_1 + f_2 \\ \Leftrightarrow\ &\begin{cases} \boldsymbol{x}^\star = \mathrm{prox}_{\gamma f_2}(\boldsymbol{y}), \\ \boldsymbol{y} \in \mathrm{Fix}\left(\mathrm{rprox}_{\gamma f_1}\mathrm{rprox}_{\gamma f_2}\right), \end{cases} \end{aligned}$$

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* $\mathrm{rprox}_{\gamma f_1}\mathrm{rprox}_{\gamma f_2}$.

Fact 4 *(Fixed Point Characterization II [8])*
For a pair of real Hilbert spaces $\mathcal{X}_1$ *and* $\mathcal{X}_2$*, consider the solution set*

$$S_p := \arg\min_{\boldsymbol{x}\in\mathcal{X}_1} [f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + f_3(L(\boldsymbol{x}))] \neq \emptyset, \tag{3}$$

of a convex optimization problem, where $L : \mathcal{X}_1 \to \mathcal{X}_2$ *is a bounded linear operator,* $f_1 \in \Gamma_0(\mathcal{X}_1)$ *is Gâteaux differentiable with* β*-Lipschitzian gradient* $\nabla f_1 : \mathcal{X}_1 \to \mathcal{X}_1$*, and* $f_2 \in \Gamma_0(\mathcal{X}_1)$*,* $f_3 \in \Gamma_0(\mathcal{X}_2)$. *Suppose that*

$$\bigcup_{\lambda>0} \{\lambda\boldsymbol{x} \mid \boldsymbol{x} \in L(\mathrm{dom}(f_2)) - \mathrm{dom}(f_3)\}$$

is a closed subspace of $\mathcal{X}_1$. *Then it is well known that the solution set* S_d *of the dual formulation is nonempty [1], i.e.,*

$$S_d := \arg\min_{\boldsymbol{y}\in\mathcal{X}_2} [(f_1 + f_2)^*(-L^*\boldsymbol{y}) + f_3^*(\boldsymbol{y})] \neq \emptyset,$$

where $L^* : \mathcal{X}_2 \to \mathcal{X}_1$ *is the adjoint operator of* L*, i.e.,* L^* *is the unique bounded linear operator satisfying* $\langle L(\boldsymbol{x}), \boldsymbol{y}\rangle_2 = \langle \boldsymbol{x}, L^*(\boldsymbol{y})\rangle_1$ *(*$\forall \boldsymbol{x} \in \mathcal{X}_1$ *and* $\forall \boldsymbol{y} \in \mathcal{X}_2$*).*

Moreover, the set $S_p \times S_d \subset \mathcal{X}_1 \times \mathcal{X}_2$ *can be expressed as the fixed point set of a nonexpansive mapping* $T : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_1 \times \mathcal{X}_2$ *defined, together with the inner product* $\langle\cdot,\cdot\rangle_{\mathcal{P}}$ *and its induced norm* $\|\cdot\|_{\mathcal{P}}$*, as follows [8].*

For the standard product space $(\mathcal{H} := \mathcal{X}_1 \times \mathcal{X}_2, \langle\cdot,\cdot\rangle, \|\cdot\|)$ *(see Example 1(a)), define first a strongly positive bounded linear operator*

$$\mathcal{P} : (\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto \left(\frac{1}{\tau}\boldsymbol{x}_1 - L^*(\boldsymbol{x}_2), -L(\boldsymbol{x}_1) + \frac{1}{\sigma}\boldsymbol{x}_2\right),$$

where $\tau > 0$*,* $\sigma > 0$ *are chosen to satisfy* $\frac{1}{\tau} - \sigma\|L\|^2 \geq \frac{\beta}{2}$. *Define also a new Hilbert space* $(\mathcal{H}_{\mathcal{P}} := \mathcal{X}_1 \times \mathcal{X}_2, \langle\cdot,\cdot\rangle_{\mathcal{P}}, \|\cdot\|_{\mathcal{P}})$*, where* $\langle(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{y}_1, \boldsymbol{y}_2)\rangle_{\mathcal{P}} := \langle\mathcal{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{y}_1, \boldsymbol{y}_2)\rangle$ *and* $\|(\boldsymbol{x}_1, \boldsymbol{x}_2)\|_{\mathcal{P}} := \sqrt{\langle(\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{x}_1, \boldsymbol{x}_2)\rangle_{\mathcal{P}}}$. *Then the mapping* $T : (\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto (\boldsymbol{y}_1, \boldsymbol{y}_2)$ *defined by*

$$\begin{cases} \boldsymbol{y}_1 := \mathrm{prox}_{\tau f_2}[\boldsymbol{x}_1 - \tau(\nabla f_1(\boldsymbol{x}_1) + L^*(\boldsymbol{x}_2))] \\ \boldsymbol{y}_2 := \mathrm{prox}_{\sigma f_3^*}[\boldsymbol{x}_2 + \sigma L(2\boldsymbol{y}_1 - \boldsymbol{x}_1)] \end{cases}$$

is nonexpansive (averaged nonexpansive if $\frac{1}{\tau} - \sigma\|L\|^2 > \frac{\beta}{2}$*) in* $(\mathcal{H}_{\mathcal{P}} := \mathcal{X}_1 \times \mathcal{X}_2, \langle\cdot,\cdot\rangle_{\mathcal{P}}, \|\cdot\|_{\mathcal{P}})$*, and satisfies*

$$\mathrm{Fix}(T) = S_p \times S_d \subset \mathcal{H}_{\mathcal{P}}.$$

Remark 1 *(Expressions of Solution Sets with Computable Nonexpansive Mappings)*

(a) *From Fact3(a), we see that the set* S *in (2) can be expressed as the fixed point set of a computable nonexpansive mapping if the lipschitzian gradient* ∇f_1 *and the proximity operator* $\mathrm{prox}_{\mu f_2}$ *are available.*

(b) *From Fact3(b), we see that the set* S *in (2) can be expressed in terms of the fixed point set of a computable nonexpansive mapping if the proximity operators* $\mathrm{prox}_{\gamma f_i}$ *(*$i = 1, 2$*) are are available.*

(c) *From Fact 4, we see that the set* S_p *in (3) can be expressed in terms of the fixed point set of a computable nonexpansive mapping if the lipschitzian gradient* ∇f_1 *and the proximity operators* $\mathrm{prox}_{\gamma f_i}$ *(*$i = 2, 3$*) are available. Different fixed point characterizations of* $S_p \times S_d$ *could be reproduced, e.g., from the results in [7].*

3. ITERATIVE APPROXIMATION OF A FIXED POINT OF NONEXPANSIVE MAPPING

As seen in Remark 1, the solution set of a variety of convex optimization problems can be expressed precisely in terms of the fixed point set of a certain computable nonexpansive mapping. This fact implies that we can approximate iteratively a point in the solution set by applying the following algorithms.

Fact 5 *(Fixed Point Algorithms for Nonexpansive Mappings) Let* $T : \mathcal{H} \to \mathcal{H}$ *be a nonexpansive mapping having* $\mathrm{Fix}(T) \neq \emptyset$. *Then we can apply the following algorithms to approximate a point in* $\mathrm{Fix}(T)$.

I. *(Krasnosel'skii-Mann iterative process (see, e.g., [1])) For any initial point* $\boldsymbol{x}_0 \in \mathcal{H}$ *and any real sequence* $(\alpha_n)_{n\geq 0} \subset [0, 1]$ *satisfying* $\sum_{n\geq 0}\alpha_n(1 - \alpha_n) = \infty$*, the sequence* $(\boldsymbol{x}_n)_{n\geq 0} \subset \mathcal{H}$*, generated by*

$$\boldsymbol{x}_{n+1} := (1 - \alpha_n)\boldsymbol{x}_n + \alpha_n T(\boldsymbol{x}_n),$$

converges weakly to a point in $\mathrm{Fix}(T)$*, which depends on the choices of* $\boldsymbol{x}_0 \in \mathcal{H}$ *and the real number sequence* $(\alpha_n)_{n\geq 0} \subset [0, \infty)$.

II. (Hybrid steepest descent method (see, e.g., [13, 14, 15])) Suppose that $\Theta \in \Gamma_0(\mathcal{H})$ is Gâteaux differentiable with the gradient $\nabla\Theta$ which is κ-Lipschitzian over $T(\mathcal{H}) := \{T(\boldsymbol{x}) \in \mathcal{H} \mid \boldsymbol{x} \in \mathcal{H}\}$. The hybrid steepest descent method :

$$\boldsymbol{x}_{n+1} := T(\boldsymbol{x}_n) - \alpha_{n+1}\nabla\Theta\left(T(\boldsymbol{x}_n)\right) \tag{4}$$

has been developed to approximate a point in

$$\Omega := \left\{\boldsymbol{x}^\star \in \mathrm{Fix}(T) \mid \Theta(\boldsymbol{x}^\star) = \min_{\boldsymbol{x}\in\mathrm{Fix}(T)} \Theta(\boldsymbol{x})\right\} \neq \emptyset.$$

(a) Suppose that the gradient $\nabla\Theta$ is η-strongly monotone over $T(\mathcal{H})$, which guarantees $|\Omega| = 1$. Then, by using any sequence $(\alpha_n)_{n\geq 1} \subset [0,\infty)$ satisfying (W1) $\lim_{n\to\infty}\alpha_n = 0$, (W2) $\sum_{n\geq 1}\alpha_n = \infty$, (W3) $\sum_{n\geq 1}|\alpha_n - \alpha_{n+1}| < \infty$ or $(\alpha_n)_{n\geq 1} \subset (0,\infty)$ satisfying (L1) $\lim_{n\to\infty}\alpha_n = 0$, (L2) $\sum_{n\geq 1}\alpha_n = \infty$, (L3) $\lim_{n\to\infty}(\alpha_n - \alpha_{n+1})\alpha_{n+1}^{-2} = 0]$, the sequence $(\boldsymbol{x}_n)_{n\geq 0}$ generated, with arbitrary $\boldsymbol{x}_0 \in \mathcal{H}$, by (4) converges strongly to the uniquely existing point $u^\star \in \Omega$

(b) Assume $\dim(\mathcal{H}) < \infty$ and $T : \mathcal{H} \to \mathcal{H}$ is an averaged nonexpansive mapping with bounded $\mathrm{Fix}(T) \neq \emptyset$. Then the sequence $(\boldsymbol{x}_n)_{n\geq 0}$ generated by (4), for arbitrary $\boldsymbol{x}_0 \in \mathcal{H}$, satisfies $\lim_{n\to\infty} d_\Omega(\boldsymbol{x}_n) = 0$ if $(\alpha_n)_{n\geq 1} \subset (0,\infty)$ satisfies (W1), (W2) and $(\alpha_n)_{n\geq 1} \in \ell_2$, i.e., $\sum_{n\geq 1}\alpha_n^2 < \infty$.

4. APPLICATION TO SIGNAL RECOVERY

4.1. Problem Formulation

Many signal recovery problems, such as signal reconstruction, image restoration, compressed sensing, and tensor completion, can be cast as an inverse problem where the goal is to estimate an unknown signal $\bar{\mathbf{u}} \in \mathbb{R}^N$ from an observation

$$\mathbf{v} = \mathcal{N}(\mathbf{\Phi}\bar{\mathbf{u}}), \tag{5}$$

where $\mathbf{\Phi} \in \mathbb{R}^{K\times N}$ $(K \leq N)$ is a linear degradation operator, and $\mathcal{N}$ some noise contamination process being not necessarily additive. To achieve satisfactory signal recovery, such an inverse problem is often translated into the following optimization problem: find $\mathbf{u}^\star$ in

$$\arg\min_{\mathbf{u}\in\mathbb{R}^N} \left[\phi(\mathbf{M}_1\mathbf{u}) + \psi(\mathbf{\Phi}\mathbf{u}) + i_D(\mathbf{u})\,(=: J(\mathbf{u}))\right], \tag{6}$$

where $\phi \in \Gamma_0(\mathcal{Y}_1)$ is a possibly nonsmooth convex prior, ($\mathcal{Y}_1$ is a finite dimensional real Hilbert space), $\psi \in \Gamma_0(\mathbb{R}^K)$ a data-fidelity function, $\mathbf{M}_1 : \mathbb{R}^N \to \mathcal{Y}_1$ a bounded linear operator, and D a closed convex set which represents some possible numerical range of $\bar{\mathbf{u}}$. Here we assume that the proximity operators of ϕ, ψ and the metric projection onto D are available.

Although this approach brings us a reasonable estimate, the non-strict convexity of the 1st stage cost function often results in infinitely-many minimizers (i.e., acceptable estimates) to the problem, and they are considerably different in another measure. Hence it is natural to characterize a more reasonable solution as the following hierarchical convex optimization problem: find $\mathbf{u}^{\star\star}$ in

$$\arg\min_{\mathbf{u}^\star\in\arg\min J(\mathbb{R}^N)} {}^{\gamma}\varphi(\mathbf{M}_2\mathbf{u}), \tag{7}$$

where $\varphi \in \Gamma_0(\mathcal{Y}_2)$ ($\mathcal{Y}_2$ is another finite dimensional real Hilbert space) is another possibly nonsmooth convex prior which reflects some desired property discarded in ϕ so as to select a higher quality signal among all solutions to (6).

4.2. Solver via The Hybrid Steepest Descent Method

In the following, the set of minimizers in (6) is characterized exactly in terms of the fixed point set of a nonexpansive mapping, which makes the hybrid steepest descent method applicable to the hierarchical convex optimization problem (7). By letting $\mathcal{X}_1 := \mathbb{R}^N$, $\mathcal{X}_2 := \mathcal{Y}_1 \times \mathbb{R}^K$, $f_1 : \mathbb{R}^N \to \mathbb{R} : \mathbf{u} \mapsto 0$, $f_2 : \mathbb{R}^N \to [0,\infty] : \mathbf{u} \mapsto i_D(\mathbf{u})$, $L : \mathbb{R}^N \to \mathcal{Y}_1 \times \mathbb{R}^K : \mathbf{u} \mapsto (\mathbf{M}_1\mathbf{u}, \mathbf{\Phi}\mathbf{u})$, and $f_3 : \mathcal{X}_2 \to [0,\infty] : (\mathbf{z}_1,\mathbf{z}_2) \mapsto \phi(\mathbf{z}_1) + \psi(\mathbf{z}_2)$, we see that the problem (6) is interpreted as the problem (3), and hence the problem (7) can be expressed as: find $(\mathbf{u}^{\star\star}, \mathbf{z}_1^{\star\star}, \mathbf{z}_2^{\star\star})$ in

$$\arg\min_{(\mathbf{u}^\star,\mathbf{z}_1^\star,\mathbf{z}_2^\star)\in S_p\times S_d} \Theta(\mathbf{u}^\star, \mathbf{z}_1^\star, \mathbf{z}_2^\star), \tag{8}$$

where $\Theta : \mathcal{H}_{\mathcal{P}} \to \mathbb{R} : (\mathbf{u},\mathbf{z}_1,\mathbf{z}_2) \mapsto {}^{\gamma}\varphi(\mathbf{M}_2\mathbf{u})$. Since T is computable from the assumption, and the gradient of Θ is given by

$$\nabla\Theta : \mathcal{H}_{\mathcal{P}} \to \mathcal{H}_{\mathcal{P}} : (\mathbf{u},\mathbf{z}_1,\mathbf{z}_2) \mapsto (\mathbf{M}_2^* \circ (\nabla^{\gamma}\varphi) \circ \mathbf{M}_2(\mathbf{u}), \mathbf{0}, \mathbf{0})$$

and obviously Lipschitzian, we can apply the hybrid steepest descent method to the problem (8), resulting in

$$\begin{aligned}&(\mathbf{u}^{(n+1)}, \mathbf{z}_1^{(n+1)}, \mathbf{z}_2^{(n+1)}) := \\ &T(\mathbf{u}^{(n)}, \mathbf{z}_1^{(n)}, \mathbf{z}_2^{(n)}) - \alpha_n\nabla\Theta(T(\mathbf{u}^{(n)}, \mathbf{z}_1^{(n)}, \mathbf{z}_2^{(n)})).\end{aligned}$$

4.3. Numerical Examples

We examined the recovery performance of the hierarchical convex optimization (7) on image restoration. In this experiment, the linear degradation operator $\mathbf{\Phi}$ in (5) was set to a masking operator that consists of a subset of randomly chosen K rows of the identity matrix, so that the degraded observation $\mathbf{v}$ is remaining pixels. We consider to restore a full image from it. The test image "Barbara" were randomly missing about 40% of their original pixels (i.e., $K \approx 0.6N$), and contaminated by an additive white Gaussian noise with the standard deviation 5.

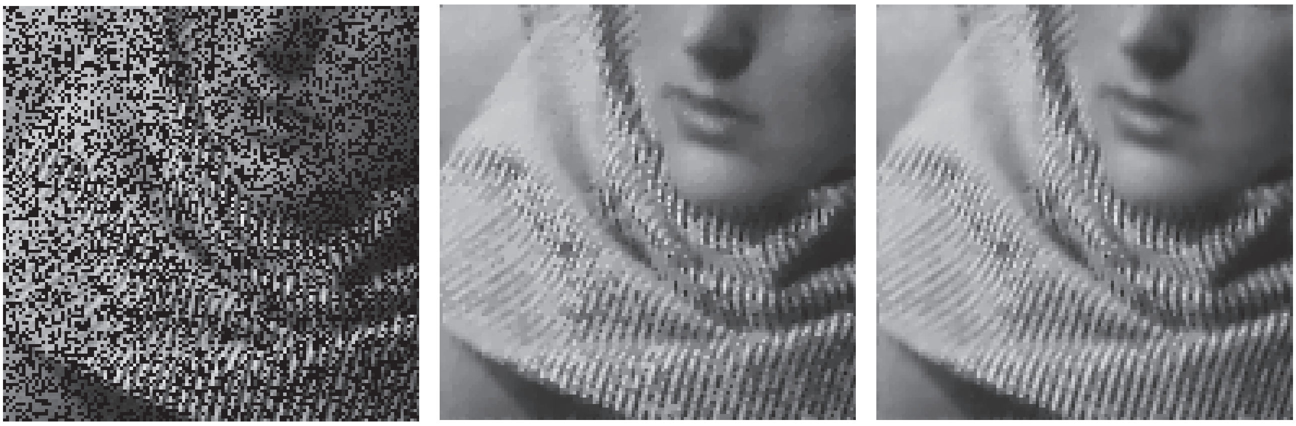
For $\phi(\mathbf{M}_1\cdot)$, we used the Total Variation, where $\phi : \mathbb{R}^N \times \mathbb{R}^N \to (0,\infty) : (\mathbf{x},\mathbf{y}) \mapsto \sum_{i=1}^N \sqrt{x_i^2 + y_i^2}$, $\mathbf{M}_1 : \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N : \mathbf{u} \mapsto (\mathbf{D}_v\mathbf{u}, \mathbf{D}_h\mathbf{u})$, $\mathbf{D}_v, \mathbf{D}_h$ are vertical and horizontal discrete gradient operators. In this case, the proximity operator of ϕ is given by Example 1(a)(c). For $\varphi(\mathbf{M}_2\cdot)$, we adopted the ℓ^1 norm of curvelet coefficients, where $\varphi : \mathbb{R}^{\tilde{N}} \to (0,\infty) : \mathbf{x} \mapsto \sum_{i=1}^{\tilde{N}} |x_i|$, $\mathbf{M}_2 : \mathbb{R}^N \to \mathbb{R}^{\tilde{N}} : \mathbf{u} \mapsto \mathbf{C}\mathbf{u}$, and $\mathbf{C}$ is the discrete curvelet transform with four scales and 16 angles [2]. Although the closed form expression of the proximity operator of $\varphi(\mathbf{M}_2\cdot)$ is unavailable due to the non-semi-orthogonality of $\mathbf{M}_2(= \mathbf{C})$, $\nabla\left({}^{\gamma}\varphi \circ \mathbf{M}_2\right)$ is available as long as $\mathrm{prox}_{\gamma\varphi}$ is available (see (1)). The proximity operator of φ is given by Example 1(a)(c).

This hierarchical formulation is designed for restoration of an image having clear edges and contours among all possible smooth images in the set of minimizers in (6).

The data-fidelity function ψ was set to $i_{C_{\mathbf{v}}^{\varepsilon}}$, where $C_{\mathbf{v}}^{\varepsilon} := \{\mathbf{x} \in \mathbb{R}^K \mid \|\mathbf{x} - \mathbf{v}\| \leq \varepsilon\}$, and the metric projection onto $C_{\mathbf{v}}^{\varepsilon}$ is given by

$$P_{C_{\mathbf{v}}^{\varepsilon}} : \mathbb{R}^K \to \mathbb{R}^K : \mathbf{x} \mapsto \begin{cases}\mathbf{x}, & \text{if } \|\mathbf{x} - \mathbf{v}\| \leq \rho, \\ \mathbf{v} + \varepsilon\frac{\mathbf{x}-\mathbf{v}}{\|\mathbf{x}-\mathbf{v}\|}, & \text{otherwise.}\end{cases}$$

The set D was designed as $D := \{\mathbf{x} \in \mathbb{R}^N \mid x_i \in [0,255]\ i = 1,\ldots,N\}$ which represents the range of pixel intensity. The metric



(a) Degraded. (b) Conventional. PSNR=23.06 dB, SSIM=0.8077 (c) Proposed. PSNR=24.22 dB, SSIM=0.8437

Fig. 1. The resulting images.

projection onto D is given by

$$P_D : \mathbb{R}^N \to \mathbb{R}^N : x_i \mapsto \begin{cases} 0, & \text{if } x_i < 0, \\ x_i, & \text{if } 0 \leq x_i \leq 255, \\ 255, & \text{if } x_i > 255. \end{cases}$$

The parameters are set as $(\varepsilon, \tau, \sigma, \alpha_n) = (\|\mathbf{\Phi}\bar{\mathbf{u}} - \mathbf{v}\|, 0.4, 0.4, \frac{1}{n})$.

Figure 1 depicts the resulting images. Compared with the restored image in Fig. 1(b) obtained by solving (6) (conventional), the restored image in Fig. 1(c), which is restored by solving the problem (7) (proposed), has clear striped patterns and indicates higher PSNR [dB] and SSIM [12].

5. CONCLUSION

This paper has presented a flexible utilization of the hybrid steepest descent method for a strategic point selection among the solution set of nonsmooth convex optimization problems. The numerical example for an image restoration application demonstrates the effectiveness of the proposed formulation as a hierarchical convex optimization.

6. REFERENCES

[1] H. H. Bauschke, and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer-Verlag, 2011.

[2] E. Candès, L. Demanet, D. L. Donoho, and L. Ying, “Fast discrete curvelet transforms,” SIAM J. Multi. Model. Simul., vol. 5, no. 3, pp. 861-899, 2006.

[3] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” J. Mathematical Imaging and Vision, vol.40, no.1, pp.120-145, 2011.

[4] P. L. Combettes, “Foundation of set theoretic estimation,” *Proc. IEEE*, vol.81, pp.182-208, 1993.

[5] P. L. Combettes, “Inconsistent signal feasibility problems: least squares solutions in a product space,” *IEEE Trans. Signal Process.*, vol.42, pp.2955-2966, 1994.

[6] P. L. Combettes and J.-C. Pesquet, “A proximal decomposition method for solving convex variational inverse problems,” “Proximal splitting methods in signal processing,” pp.185-212, In: Fixed Point Algorithms for Inverse Problems in Science and Engineering, Springer-Verlag, 2011.

[7] P. L. Combettes and J. C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,” Set-Valued and Variational Analysis, 20 (2), pp.307-330, 2012.

[8] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” J. Optimization Theory and Applications, DOI 10.1007/s10957-012-0245-9 (Online: 29 Dec., 2012).

[9] D. L. Donoho, De-noising by soft-thresholding, IEEE Trans. Information Theory, vol. 41, no.3, pp.613-627, 1995.

[10] S. Gandy, B. Recht, and I. Yamada, “Tensor completion and low-n-rank tensor recovery via convex optimization,” Inverse Probl., vol.27, no.2, 025010, 2011.

[11] J.-L. Starck, F. Murtagh, and J. M. Fadili, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity, Cambridge University Press, 2010.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[13] I. Yamada, “The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings,” Studies in Computational Mathematics, 8, pp.473-504, 2001.

[14] I. Yamada and N. Ogura, “Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings,” Numer. Funct. Anal. Optim., vol.25, no.7&8, pp. 619-655, 2004.

[15] I. Yamada, M. Yukawa, M. Yamagishi, “Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings,” pp.345-390, In: Fixed Point Algorithms for Inverse Problems in Science and Engineering, Springer-Verlag, 2011.