A Hierarchical Convex Optimization Approach for High Fidelity Solution Selection in Image Recovery

Shunsuke ONO* and Isao YAMADA*
* Dept. Communications and Integrated Systems, Tokyo Institute of Technology, Tokyo, Japan
E-mail: ono@sp.ss.titech.ac.jp; isao@sp.ss.titech.ac.jp Tel/Fax: +81-3-5734-2503/2905

***Abstract*—The aim of this paper is to propose a hierarchical convex optimization for selecting a high fidelity image from possible solutions of a convex optimization problem associated with existing image recovery methods. Image recovery problems have been cast in certain convex optimization problems which have infinitely many solutions in general. However, existing convex optimization algorithms are designed to reach one solution randomly, and hence can not select a solution corresponding to a high fidelity image from the possible solutions. In this paper, we propose to select a high fidelity image by solving a newly-formulated hierarchical convex optimization problem. This problem is a constrained minimization of a convex criteria over the solution set of all images which are optimal in the sense of any existing image recovery method. The hierarchical convex optimization problem is efficiently solved by a proposed iterative scheme based on the hybrid steepest descent method with the help of a nonexpansive mapping related to the Douglas-Rachford splitting type algorithms. Numerical results indicate that our method appropriately selects a recovered image of high fidelity in the case of inpainting and compressed sensing recovery.**

I. Introduction

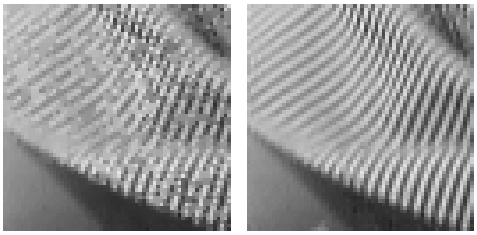
Many image recovery problems, such as denosing, deblurring, inpainting, and compressed sensing, are stated as an inverse problem where an unknown original image $\mathbf{u}_{\mathrm{org}} \in \mathbb{R}^{N_0}$ is estimated from an observation $\mathbf{v} \in \mathbb{R}^{K(\leq N_0)}$ which is degraded by a certain linear operator $\mathbf{\Phi} \in \mathbb{R}^{K \times N_0}$ and contaminated by a noise which is not necessarily additive. Such an inverse problem has been solved efficiently by formulating them as various convex optimization problems which exploit certain prior knowledge on the unknown original image.

In the last decade, convex optimization-based image recovery has gained considerable attention, and many methods have been studied, for example, in [1], [2], [3], [6], [8], [9], [10], [14], [15], [16], [17], [18], [20]. Their associated convex optimization problems can be included in the following problem.

Problem I.1 Let $\mathbf{A}_k$ $(k = 1, \ldots, p)$ be a linear operator from $\mathbb{R}^{N_0}$ to $\mathbb{R}^{N_k}$, and let $F_k \in \Gamma_0(\mathbb{R}^{N_0})$ $(k = 0, \ldots, p)$ be a proper lower semicontinuous convex function[1] from $\mathbb{R}^{N_k}$ to $(-\infty, \infty]$ whose proximity operator (see section II-A) is computable. Then the problem is finding

$$\mathbf{u}^\star \in S := \arg \min_{\mathbf{u} \in \mathbb{R}^{N_0}} \left\{ F_0(\mathbf{u}) + \sum_{k=1}^{p} F_k(\mathbf{A}_k \mathbf{u}) \right\}. \quad (1)$$

Generally, the solution of the problem (1) is not necessary unique because there are many cases where $F_0(\mathbf{u}) + \sum_{k=1}^{p} F_k(\mathbf{A}_k \mathbf{u})$ in (1) does not become strictly convex[2] over $\mathbb{R}^{N_0}$. This leads to the co-existence of solutions correspoinding to "ordinary" and "high fidelity" images. One of the typical cases can be found in the *total variation-based (TV-based)* recovery (e.g., [1], [3]). Although the TV-based recovery is an effective approach, there exist both types of solutions. We show an example in Fig. 1. Noticeable artifacts (disconnected pattern) remain in Fig. 1(a). In contrast, Fig. 1(b) preserves clear straight-lines.



(a) Ordinary (b) High fidelity

Fig. 1. (a) Ordinary case: a recovered image without our technique in which straight-line patterns are still corrupted. (b) High fidelity case: a selected recovered image by using our technique which has clear edges and lines compared to (a). Note that both of two are the solutions of the same problem.

Unfortunately, existing image recovery methods based on the formulation of type (1) are only designed to find one solution randomly, because it is even difficult (require some iterative scheme) to reach S. This fact implies that if there exists a technique which can always "select" a solution corresponding to a high fidelity image from S, we can improve the performance of all convex optimization-based image recovery methods based on (1) by applying the selection technique.

In this paper, we propose such a selection technique based on the following hierarchical convex optimization problem.

Problem I.2 (Solution selection of Problem I.1) Let $\varphi : \mathbb{R}^{N_0} \to \mathbb{R}$ be a differentiable convex function with β-Lipschitz

[1] Let $\mathcal{H}$ be a real Hilbert space with its inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. A function $f : \mathcal{H} \to (-\infty, \infty]$ is called *proper lower semicontinuous convex* if $\mathrm{dom}(f) := \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) < \infty\} \neq \emptyset$, $\mathrm{lev}_{\leq \alpha}(f) := \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$, and $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ and $\lambda \in (0, 1)$, respectively. The set of all proper lower semicontinuous convex functions is denoted by $\Gamma_0(\mathcal{H})$. Euclidean space $\mathbb{R}^{N_k}$ is an instance of a real Hilbert space.

[2] A function $f \in \Gamma_0(\mathcal{H})$ is called *strictly convex* if $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$ for every $\mathbf{x} \neq \mathbf{y}$ and $\lambda \in (0, 1)$.

continuous gradient for some $\beta \in (0, \infty)$, i.e., for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $\|\nabla\varphi(\mathbf{x}) - \nabla\varphi(\mathbf{y})\|_2 \leq \beta\|\mathbf{x} - \mathbf{y}\|_2$. Then the problem is finding

$$\mathbf{u}^{\star\star} \in \arg\min_{\mathbf{u}^\star \in S} \varphi(\mathbf{u}^\star). \tag{2}$$

This problem is equivalent to find a minimizer of the "selector" φ, which is designed to take a smaller value at an image of higher fidelity, among the solution set of Problem I.1. Hence, solving Problem I.2 corresponds to selecting a high fidelity image among all possible recovered images. To tackle this problem, we utilize the fact that the solution set S can be expressed in terms of the fixed point set of a certain nonexpansive mapping[3] derived from the *Douglas-Rachford splitting* type algorithms (e.g., [8], [11]). Thanks to this characterization, we can efficiently solve Problem I.2 by using the *hybrid steepest descent method* (e.g., [22], [23]) (see, Section II-C).

The rest of this paper is organized as follows. In the next section, we introduce basic mathematical tools used in our technique. Then, we propose the selection technique in Section III. We also present the case of the TV-based recovery in the same section where our technique works effectively. Corresponding numerical results are shown in Section IV. Finally, we conclude the paper in Section V with some remarks.

II. Preliminaries

A. Proximity Operator and Moreau envelope

We utilizes the notion of the proximity operator and the Moreau envelope [13], [25] in our technique.

Definition II.1 (Proximity operator) The proximity operator of index $\gamma \in (0, \infty)$ of $f \in \Gamma_0(\mathcal{H})$ is defined by

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

where the existence and the uniqueness of the minimizer are guaranteed respectively by the coercivity[4] and the strict convexity of $f(\cdot) + \frac{1}{2\gamma}\|\mathbf{x} - \cdot\|^2$.

Definition II.2 (Moreau envelope and its derivative) Every function $f \in \Gamma_0(\mathcal{H})$ can be approximated with any accuracy by

$$\begin{aligned} {}^\gamma f : \mathcal{H} \to \mathbb{R} : \ \mathbf{x} &\mapsto \min_{\mathbf{y} \in \mathcal{H}} \left\{ f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{x} - \mathbf{y}\|^2 \right\} \\ &= f\left(\mathrm{prox}_{\gamma f}(\mathbf{x})\right) + \frac{1}{2\gamma}\left\|\mathbf{x} - \mathrm{prox}_{\gamma f}(\mathbf{x})\right\|^2, \end{aligned} \tag{3}$$

which is called the Moreau envelope of index $\gamma \in (0, \infty)$ of f. ${}^\gamma f$ is differentiable and its gradient

$$\nabla {}^\gamma f(\mathbf{x}) : \mathcal{H} \to \mathcal{H} : \mathbf{x} \mapsto \frac{\mathbf{x} - \mathrm{prox}_{\gamma f}(\mathbf{x})}{\gamma}, \tag{4}$$

is γ^{-1}-Lipschitzian[5].

[3] A mapping $T : \mathcal{H} \to \mathcal{H}$ is called *nonexpansive* if $\|T(\mathbf{x}) - T(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$. $\mathrm{Fix}(T) := \{\mathbf{x} \in \mathcal{H} \mid T(\mathbf{x}) = \mathbf{x}\}$ is called the *fixed point set of* T.

[4] A function $f \in \Gamma_0(\mathcal{H})$ is called *coercive* if $\|\mathbf{x}\| \to \infty \Rightarrow f(\mathbf{x}) \to \infty$.

[5] A mapping $T : \mathcal{H} \to \mathcal{H}$ is called κ-Lipschitzian if $\|T(\mathbf{x}) - T(\mathbf{y})\| \leq \kappa\|\mathbf{x} - \mathbf{y}\|$ for some $\kappa > 0$ and every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$.

B. Douglas-Rachford Splitting

In recent years, the *Douglas-Rachford splitting* type algorithms (e.g., [8], [11]) have been studied and applied to convex optimization problems. Let f_1, $f_2 \in \Gamma_0(\mathcal{H})$ satisfy

$$U := \arg\min_{\mathbf{x} \in \mathcal{H}} \{f_1(\mathbf{x}) + f_2(\mathbf{x})\} \neq \emptyset. \tag{5}$$

The Douglas-Rachford splitting type algorithms use in principle the following characterization: for any $\delta \in (0, \infty)$

$$U = \left\{ \mathrm{prox}_{\delta f_2}(\mathbf{y}) \in \mathcal{H} |\ \mathbf{y} \in \mathrm{Fix}\left(\mathrm{rprox}_{\delta f_1}\mathrm{rprox}_{\delta f_2}\right) \right\} \tag{6}$$

which means that U can be expressed as the image of $\mathrm{prox}_{\delta f_2}$ of the fixed point set of the nonexpansive mapping $\mathrm{rprox}_{\delta f_1}\mathrm{rprox}_{\delta f_2}$, where $\mathrm{rprox}_{\delta f} := 2\mathrm{prox}_{\delta f} - I$ (I denotes the identity operator).

C. Hybrid Steepest Descent Method

The hybrid steepest descent method (e.g., [22], [23]) is known as an algorithm for solving the following convex optimization problem.

Problem II.1 Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping whose fixed point set is nonempty. Suppose that $\Theta \in \Gamma_0(\mathcal{H})$ is differentiable with gradient $\nabla\Theta$ which is κ-Lipschitzian over $T(\mathcal{H}) := \{T(\mathbf{x}) \in \mathcal{H} \mid \mathbf{x} \in \mathcal{H}\}$. Then the problem is finding

$$\mathbf{x}^\star \in \Omega := \arg\min_{\mathbf{x} \in \mathrm{Fix}(T)} \Theta(\mathbf{x}). \tag{7}$$

The hybrid steepest descent method

$$\mathbf{x}_{n+1} := T(\mathbf{x}_n) - \lambda_{n+1}\nabla\Theta(T(\mathbf{x}_n)) \tag{8}$$

is an simple algorithmic solution to Problem (7), where $(\lambda_n)_{n>1} \subset [0, \infty)$ is a slowly decreasing nonnegative sequence (e.g., $\lambda_n = \frac{1}{n}$, see Section 17.3.2 of [23] for details).

III. Proposed Method

A. General Case

First, we characterize the solution set S in (1) by the form of (6). Let $X = \mathbb{R}^{N_0} \times \cdots \times \mathbb{R}^{N_p}$ be a real Hilbert space where the inner product $\langle \cdot, \cdot \rangle_X : X \times X \to \mathbb{R}$ and its induced norm $\|\cdot\|_X$ are defined as $\langle (\mathbf{x}_0, \ldots, \mathbf{x}_p), (\mathbf{y}_0, \ldots, \mathbf{y}_p) \rangle_X := \mathbf{x}_0^t\mathbf{y}_0 + \cdots + \mathbf{x}_p^t\mathbf{y}_p$, $\|(\mathbf{x}_0, \ldots, \mathbf{x}_p)\|_X := \sqrt{\langle (\mathbf{x}_0, \ldots, \mathbf{x}_p), (\mathbf{x}_0, \ldots, \mathbf{x}_p) \rangle_X}$, for $(\mathbf{x}_0, \ldots, \mathbf{x}_p), (\mathbf{y}_0, \ldots, \mathbf{y}_p) \in X$, respectively. Define $D := \{(\mathbf{u}, \mathbf{A}_1\mathbf{u}, \ldots \mathbf{A}_p\mathbf{u}) \mid \mathbf{u} \in \mathbb{R}^{N_0}\} \subset X$ and

$$f_1(\mathbf{z}_0, \ldots, \mathbf{z}_p) := F_0(\mathbf{z}_0) + \sum_{k=1}^{p} F_k(\mathbf{z}_k), \tag{9}$$

$$f_2(\mathbf{z}_0, \ldots, \mathbf{z}_p) := \iota_D(\mathbf{z}_0, \ldots, \mathbf{z}_p), \tag{10}$$

where ι_D denotes the *indicator function* of D given by

$$\iota_D(\mathbf{x}) := \begin{cases} 0, & \text{if } \mathbf{x} \in D, \\ \infty, & \text{otherwise.} \end{cases} \tag{11}$$

Then by (5) for $\mathcal{H} := X$, we obtain

$$\mathbf{u}^\star \in S \Leftrightarrow (\mathbf{z}_0^\star, \ldots, \mathbf{z}_p^\star) \in U, \tag{12}$$

where

$$U := \arg \min_{(\mathbf{z}_0,\ldots,\mathbf{z}_p)\in\mathcal{H}} \{f_1(\mathbf{z}_0,\ldots,\mathbf{z}_p) + f_2(\mathbf{z}_0,\ldots,\mathbf{z}_p)\}, \tag{13}$$

and $\mathbf{z}_0^\star = \mathbf{u}^\star$. Hence by (6), we have the characterization:

$$U := \left\{\mathrm{prox}_{\delta f_2}(\mathbf{z}_0,\ldots,\mathbf{z}_p) \mid (\mathbf{z}_0,\ldots,\mathbf{z}_p) \in \mathrm{Fix}(T)\right\}, \tag{14}$$

where $T := \mathrm{rprox}_{\delta f_1}\mathrm{rprox}_{\delta f_2}$. Note that $\mathrm{prox}_{\delta f_2}$ is equivalent to the metric projection onto D given by

$$\begin{aligned} &P_D(\mathbf{z}_0,\ldots,\mathbf{z}_p)\\ &:= \arg \min_{(\mathbf{x}_0,\ldots,\mathbf{x}_p)\in D} \|(\mathbf{x}_0,\ldots,\mathbf{x}_p) - (\mathbf{z}_0,\ldots,\mathbf{z}_p)\|_X\\ &= (\mathbf{u}_D, \mathbf{A}_1\mathbf{u}_D, \ldots, \mathbf{A}_p\mathbf{u}_D), \end{aligned} \tag{15}$$

where

$$\begin{aligned} \mathbf{u}_D &= \arg \min_{\mathbf{u}\in\mathbb{R}^{N_0}} \left\{\|\mathbf{u}-\mathbf{z}_0\|_2^2 + \sum_{k=1}^{p}\|\mathbf{A}_k\mathbf{u}-\mathbf{z}_k\|_2^2\right\}\\ &= \left(\mathbf{I}_N + \sum_{k=1}^{p}\mathbf{A}_k^t\mathbf{A}_k\right)^{-1}\left(\mathbf{z}_0 + \sum_{k=1}^{p}\mathbf{A}_k^t\mathbf{z}_k\right) \end{aligned} \tag{16}$$

($(\cdot)^t$ denotes the transposition and $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ the identity matrix).

Next, let

$$\Theta(\mathbf{z}_0,\ldots,\mathbf{z}_p) := \varphi \circ L \circ P_D(\mathbf{z}_0,\ldots,\mathbf{z}_p), \tag{17}$$

where $L : X \to \mathbb{R}^N : (\mathbf{x}_0,\ldots,\mathbf{x}_p) \mapsto \mathbf{x}_0$. Then, we can reformulate our target hierarchical convex optimization problem (2) into the following problem which is finding

$$(\mathbf{z}_0^{\star\star},\ldots,\mathbf{z}_p^{\star\star}) \in \arg \min_{(\mathbf{z}_0^\star,\ldots,\mathbf{z}_p^\star)\in\mathrm{Fix}(T)} \Theta(\mathbf{z}_0^\star,\ldots,\mathbf{z}_p^\star), \tag{18}$$

where $L \circ P_D(\mathbf{z}_0^{\star\star},\ldots,\mathbf{z}_p^{\star\star}) = \mathbf{u}^{\star\star}$. We remark that since $L \circ P_D$ is a linear operator, $\nabla\Theta$ can be derived explicitly as long as φ is differentiable (also mentioned in the concluding remarks of [23]). Hence, we can solve Problem I.2 by applying the hybrid steepest descent method to the problem (18) on the condition that $\mathrm{prox}_{\delta f_1}(\mathbf{z}_1^\star,\ldots,\mathbf{z}_p^\star) = (\mathrm{prox}_{\delta F_1}(\mathbf{z}_1),\ldots,\mathrm{prox}_{\delta F_p}(\mathbf{z}_p))$ is computable. The obtained iterative scheme based on the hybrid steepest descent method for solving (18) is shown in Algorithm III.1.

Algorithm III.1 (Algorithm for solving (18))

Set $i, j = 0$, and choose $\mathbf{u}_0 \in \mathbb{R}^{N_0}$, $\gamma, \delta, \varepsilon \in (0,\infty)$.
$\mathbf{z}_0 = \mathbf{u}_0$
for $i = 1$ to p **do**
  $\mathbf{z}_i^{(0)} = \mathbf{A}_i\mathbf{u}_0$
end for
while a stop criterion is not satisfied **do**
  $\lambda = \frac{1}{j+1}$
  $\mathbf{u}_D^{(j)} = L \circ P_D(\mathbf{z}_0^{(j)},\ldots,\mathbf{z}_p^{(j)})$
  $\mathbf{t}_0^{(j)} = 2\mathbf{u}_D^{(j)} - \mathbf{z}_0^{(j)}$
  $\mathbf{t}_0^{(j)} = 2\mathrm{prox}_{\delta F_0}(\mathbf{t}_0^{(j)}) - \mathbf{t}_0^{(j)}$
  for $i = 1$ to p **do**
    $\mathbf{t}_i^{(j)} = 2\mathbf{A}_i\mathbf{u}_D^{(j)} - \mathbf{z}_i^{(j)}$
    $\mathbf{t}_i^{(j)} = 2\mathrm{prox}_{\delta F_i}(\mathbf{t}_i^{(j)}) - \mathbf{t}_i^{(j)}$
  end for
  $(\mathbf{g}_0^{(j)},\ldots,\mathbf{g}_p^{(j)}) = \nabla\Theta(\mathbf{t}_0^{(j)},\ldots,\mathbf{t}_p^{(j)})$
  for $i = 0$ to p **do**
    $\mathbf{z}_i^{(j+1)} = \mathbf{t}_i^{(k)} - \lambda\mathbf{g}_i^{(j)}$
  end for
  $j = j + 1$
end while
Output $L \circ P_D(\mathbf{z}_0^{(j)},\ldots,\mathbf{z}_p^{(j)})$

B. Application to Total Variation-Based Recovery

In this section, we present the proposed technique specialized for the TV-based recovery. We focus on the following hierarchical convex optimization problem.

Problem III.1 (Solution selection of Problem III.2)

$$\text{Find } \mathbf{u}^{\star\star} \in \arg \min_{\mathbf{u}^\star\in S^{\mathrm{TV}}} {}^{\gamma}\|\mathbf{F}\mathbf{u}^\star\|_1, \tag{19}$$

where S^{TV} is the solution set defined in Problem III.2, ${}^{\gamma}\|\cdot\|_1$ the Moreau envelope of the ℓ^1 norm (see, Definition II.2) and $\mathbf{F} \in \mathbb{R}^{M\times N}$ ($M \geq N$, $N(= n_h n_v)$ denotes the number of pixels) a certain tight frame satisfying $\mathbf{F}^t\mathbf{F} = \mathbf{I}_N$.

This problem is designed for selecting a high fidelity image from all the solutions of Problem III.2. As the selector φ, we employ the Moreau envelope of the ℓ^1 norm of the coefficients of a tight frame $\mathbf{F}$. This is because we expect to select a recovered image in S^{TV} with clear edges and contours (see Fig. 1(b)) by utilizing a tight frame $\mathbf{F}$ which provides a sparse representation of these components (e.g., curvelet [4]). Note that ${}^{\gamma}\|\cdot\|_1$ is differentiable and its gradient is γ^{-1}-Lipschitzian (see Definition II.2).

Problem III.2 (TV-based recovery [1], [3]) Let $\mathbf{v} = \mathbf{\Phi}\mathbf{u}_{\mathrm{org}} + \boldsymbol{\eta} \in \mathbb{R}^K$ be a degraded observation where $\mathbf{u}_{\mathrm{org}} \in \mathbb{R}^N$ is an original image, and $\boldsymbol{\eta}$ the additive white Gaussian noise. Define

$$C_{[0,255]} := \{\mathbf{x} \in \mathbb{R}^N \mid x_i \in [0,255] \text{ for } i = 1,\ldots,N\}, \tag{20}$$

$$C_{\mathbf{v},\varepsilon} := \{\mathbf{x} \in \mathbb{R}^K \mid \|\mathbf{x}-\mathbf{v}\|_2 \leq \varepsilon\}. \tag{21}$$

The TV-based recovery is finding

$$\mathbf{u}^\star \in S^{\mathrm{TV}} := \arg \min_{\mathbf{u}\in\mathbb{R}^N} \left\{\|\mathbf{u}\|_{\mathrm{TV}} + \iota_{C_{[0,255]}}(\mathbf{u}) + \iota_{C_{\mathbf{v},\varepsilon}}(\mathbf{\Phi}\mathbf{u})\right\}, \tag{22}$$

where $\|\cdot\|_{\mathrm{TV}} := \sum_{i,j}\sqrt{(\nabla_1\cdot)^2 + (\nabla_2\cdot)^2}$ is the *isotropic total variation (TV)*, $\nabla_1, \nabla_2 \in \mathbb{R}^{N\times N}$ the discrete vertical and horizontal gradient operators defined by

$$(\nabla_1\mathbf{x})_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < n_v,\\ x_{1,j} - x_{i,j}, & \text{if } i = n_v, \end{cases}$$

$$(\nabla_2\mathbf{x})_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j} & \text{if } j < n_h,\\ x_{i,1} - x_{i,j} & \text{if } j = n_h. \end{cases}$$

Now, let $L : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}^N : (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{x}_0$,

$$D := \{(\mathbf{u}, \mathbf{u}, \mathbf{\Phi u}) \mid \mathbf{u} \in \mathbb{R}^N\}, \tag{23}$$

$$f_1(\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2) := \|\mathbf{z}_0\|_{\text{TV}} + \iota_{C_{[0,255]}}(\mathbf{z}_1) + \iota_{C_{\mathbf{v},\varepsilon}}(\mathbf{z}_2), \tag{24}$$

$$f_2(\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2) := \iota_D(\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2), \tag{25}$$

$$\Theta(\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2) := {}^{\gamma}\|\mathbf{F}L(P_D(\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2))\|_1. \tag{26}$$

Then, Problem III.1 can be reformulated into a form of (18), and hence can be solved by the proposed algorithm. Algorithm III.1 specialized for Problem III.1 is shown in Algorithm III.2, and the implementation of each step of the algorithm is summarized in Remark III.1.

Algorithm III.2 (Algorithm for solving (19))

Set $i, j = 0$, and choose $\mathbf{u}_0 \in \mathbb{R}^N$, $\gamma, \delta, \varepsilon \in (0, \infty)$.
$\mathbf{z}_0^{(0)} = \mathbf{u}_0$
$\mathbf{z}_1^{(0)} = \mathbf{u}_0$
$\mathbf{z}_2^{(0)} = \mathbf{\Phi u}_0$
while a stop criterion is not satisfied **do**
  $\lambda = \frac{1}{j+1}$
  $(\mathbf{u}_D^{(j)}, \mathbf{u}_D^{(j)}, \mathbf{\Phi u}_D^{(j)}) = P_D(\mathbf{z}_0^{(j)}, \mathbf{z}_1^{(j)}, \mathbf{z}_2^{(j)})$
  $\mathbf{t}_0^{(j)} = 2\mathbf{u}_D^{(j)} - \mathbf{z}_0^{(j)}$
  $\mathbf{t}_0^{(j)} = 2\text{prox}_{\delta\|\cdot\|_{\text{TV}}}(\mathbf{t}_0^{(j)}) - \mathbf{t}_0^{(j)}$
  $\mathbf{t}_1^{(j)} = 2\mathbf{u}_D^{(j)} - \mathbf{z}_1^{(j)}$
  $\mathbf{t}_1^{(j)} = 2\text{prox}_{\delta\iota_{C_{[0,255]}}}(\mathbf{t}_1^{(j)}) - \mathbf{t}_1^{(j)}$
  $\mathbf{t}_2^{(j)} = 2\mathbf{\Phi u}_D^{(j)} - \mathbf{z}_2^{(j)}$
  $\mathbf{t}_2^{(j)} = 2\text{prox}_{\delta\iota_{C_{\mathbf{v},\varepsilon}}}(\mathbf{t}_2^{(j)}) - \mathbf{t}_2^{(j)}$
  $(\mathbf{g}_0^{(j)}, \mathbf{g}_1^{(j)}, \mathbf{g}_2^{(j)}) = \nabla({}^{\gamma}\|\cdot\|_1 \circ \mathbf{F} \circ L \circ P_D)(\mathbf{t}_0^{(j)}, \mathbf{t}_1^{(j)}, \mathbf{t}_2^{(j)})$
  $\mathbf{z}_0^{(j+1)} = \mathbf{t}_0^{(k)} - \lambda\mathbf{g}_0^{(j)}$
  $\mathbf{z}_1^{(j+1)} = \mathbf{t}_1^{(k)} - \lambda\mathbf{g}_1^{(j)}$
  $\mathbf{z}_2^{(j+1)} = \mathbf{t}_2^{(k)} - \lambda\mathbf{g}_2^{(j)}$
  $j = j + 1$
end while
Output $\mathbf{u}_D^{(j)}$

Remark III.1 (Notes of the implementation)

- $\text{prox}_{\delta\|\cdot\|_{\text{TV}}}$: We observe that the computation of the proximity operator of $\|\cdot\|_{\text{TV}}$ is equivalent to the computation of the minimizer of the well-known TV denoising model called the ROF [19]. The minimizer of the ROF model can be well approximated by the projected gradient type methods (e.g., [2], [5], [24]).
- $\text{prox}_{\delta\iota_{C_{[0,255]}}}$: The proximity operator of $\iota_{C_{[0,255]}}$ is the metric projection onto $C_{[0,255]}$ given by

$$[P_{C_{[0,255]}}(\mathbf{x})]_i = \begin{cases} 0, & \text{if } x_i < 0, \\ x_i, & \text{if } 0 \le x_i \le 255, \\ 255, & \text{if } x_i > 255, \end{cases} \tag{27}$$

  for $i = 1, \ldots, N$.
- $\text{prox}_{\delta\iota_{C_{\mathbf{v},\varepsilon}}}$: The proximity operator of $\iota_{C_{\mathbf{v},\varepsilon}}$ is the metric projection onto $C_{\mathbf{v},\varepsilon}$ given by

$$P_{C_{\mathbf{v},\varepsilon}}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \text{if } \|\mathbf{x} - \mathbf{v}\|_2 \le \varepsilon, \\ \mathbf{v} + \varepsilon\frac{\mathbf{x}-\mathbf{v}}{\|\mathbf{x}-\mathbf{v}\|_2}, & \text{otherwise.} \end{cases} \tag{28}$$

- $\nabla{}^{\gamma}\|\mathbf{F}L(P_D(\cdot))\|_1$: By (4), the gradient of ${}^{\gamma}\|\cdot\|_1$ is given by

$$\nabla{}^{\gamma}\|\mathbf{x}\|_1 = \frac{\mathbf{x} - \text{prox}_{\gamma\|\cdot\|_1}(\mathbf{x})}{\gamma}, \tag{29}$$

  where $\text{prox}_{\gamma\|\cdot\|_1}$ is given by the so-called *soft-thresholding* [12], i.e.,

$$[\text{prox}_{\gamma\|\cdot\|_1}(\mathbf{x})]_i = \begin{cases} x_i - \gamma, & \text{if } x_i > \gamma, \\ x_i + \gamma, & \text{if } x_i < -\gamma, \\ 0, & \text{if } |x_i| \le \gamma, \end{cases} \tag{30}$$

  for $i = 1, \ldots, M$. From the chain rule, we can derive the gradient of ${}^{\gamma}\|\mathbf{F}L(P_D(\cdot))\|_1$ explicitly as

$$\begin{aligned} &\nabla({}^{\gamma}\|\cdot\|_1 \circ \mathbf{F} \circ L \circ P_D)(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) \\ &= P_D\left(\frac{\mathbf{u}_D - \mathbf{F}^t\text{prox}_{\gamma\|\cdot\|_1}(\mathbf{F}\mathbf{u}_D)}{\gamma}, \mathbf{0}, \mathbf{0}\right), \end{aligned} \tag{31}$$

  where

$$\begin{aligned} \mathbf{u}_D &= L(P_D(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)) \\ &= \arg\min_{\mathbf{u}\in\mathbb{R}^N}\{\|\mathbf{u} - \mathbf{x}_0\|_2^2 + \|\mathbf{u} - \mathbf{x}_1\|_2^2 + \|\mathbf{\Phi u} - \mathbf{x}_2\|_2^2\} \\ &= \left(2\mathbf{I}_N + \mathbf{\Phi}^t\mathbf{\Phi}\right)^{-1}\left(\mathbf{x}_0 + \mathbf{x}_1 + \mathbf{\Phi}^t\mathbf{x}_2\right). \end{aligned} \tag{32}$$

IV. Numerical Experiments

In this section, we show the efficacy of the proposed selection technique in the case of the TV-based recovery described in Section III-B through inpainting and compressive sensing recovery. We use five standard images (256×256 [pixels]; thus $N = 65,536$) for test images. We employ the discrete curvelet transform with 4 scale and 16 angles [4] for a tight frame $\mathbf{F}$ in (19).

In the inpainting experiments, $\mathbf{v} \in \mathbb{R}^K$ is an image with missing pixels, and $\mathbf{\Phi} = \mathbf{M} \in \mathbb{R}^{K\times N}$ is a masking operator which comprises a subset of the rows of the identity matrix. The test images are corrupted with about 40% of their original pixels (i.e., $K \approx 0.6N$) without any noise, i.e, $\varepsilon = 0$. The parameters γ, δ are set to 1.0×10^{-3}.

In the compressive sensing recovery experiment, $\mathbf{v} \in \mathbb{R}^K$ is incomplete measurements, and $\mathbf{\Phi} \in \mathbb{R}^{K\times N}$ is a certain random sampling operator. We employ the noiselet random sampling matrix for $\mathbf{\Phi} = \mathbf{M\Psi}$ where $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ is the orthogonal noiselet transform matrix [7] satisfying $\mathbf{\Psi}^t\mathbf{\Psi} = \mathbf{\Psi\Psi}^t = \mathbf{I}_N$. The number of measurements is about 40% of the number of pixels of the original image (i.e., $K \approx 0.4N$) and the white Gaussian noise whose standard deviation $\sigma = 2.55$ is added. We choose $\varepsilon = \|\mathbf{v} - \mathbf{\Phi u}_{\text{org}}\|_2$. The parameter γ, δ are set to 1.0×10^{-3} same as the inpainting experiment.

For comparison, we consider the following two cases.

- Optimizing Problem III.2 without the proposed selection technique. This is to verify that the proposed method appropriately selects a solution corresponding to a high fidelity image from S^{TV}.
- Optimizing the following problem without the proposed selection technique.

TABLE I
COMPARISON OF PSNR[DB] AND SSIM OF THE INPAINTING EXPERIMENT.

Method \ Test image	Barbara		Boat		Cameraman		Lena		Mandrill	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TV-based without our technique	26.69	0.9050	33.53	0.9528	29.10	0.9410	32.89	0.9578	27.99	0.8694
TVC-based ($w = 1$)	30.26	0.9507	35.91	0.9713	30.33	0.9466	33.73	0.9646	28.40	0.8793
TVC-based (hand-optimized w)	32.68	0.9655	36.20	0.9733	30.46	**0.9477**	33.86	0.9657	28.50	**0.8820**
TV-based with our technique	**32.93**	**0.9674**	**36.48**	**0.9747**	**30.53**	0.9475	**33.93**	**0.9658**	**28.54**	0.8819

TABLE II
COMPARISON OF PSNR[DB] AND SSIM OF THE COMPRESSIVE SENSING RECOVERY EXPERIMENT.

Method \ Test image	Barbara		Boat		Cameraman		Lena		Mandrill	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TV-based without our technique	26.18	0.7854	33.54	0.9161	32.61	0.9065	32.80	0.9127	26.49	0.7286
TVC-based ($w = 1$)	29.25	0.8581	34.08	0.9266	31.25	0.8814	32.95	0.9217	26.54	0.7420
TVC-based (hand-optimized w)	29.86	0.8931	**34.32**	**0.9299**	31.53	0.8848	33.20	0.9244	26.92	**0.7565**
TV-based with our technique	**30.18**	**0.8996**	34.22	0.9288	**32.72**	**0.9096**	**33.33**	**0.9258**	**26.93**	0.7559

Problem IV.1 (TV-Curvelet(TVC)-based recovery) Let $w \in (0, \infty)$. The problem is finding

$$\tilde{\mathbf{u}}^\star \in S^{\mathrm{TVC}} := \arg\min_{\mathbf{u}\in\mathbb{R}^N} \{\|\mathbf{u}\|_{\mathrm{TV}} + w\|\mathbf{F}\mathbf{u}\|_1 + \iota_{C_{[0,255]}}(\mathbf{u}) + \iota_{C_{\mathbf{v},\varepsilon}}(\mathbf{\Phi}\mathbf{u})\}.$$

The objective function of the TVC-based recovery is the weighted sum of the form of the TV-based recovery (22) and the selector in (19) (without Moreau envelope). This is to check whether we can restore a higher quality image, by optimizing the hierarchical formulation (Problem III.1) or by optimizing the weighted-sum formulation (Problem IV.1). Note that the solution set of Problem IV.1 is varied depending on w (clearly $S^{\mathrm{TVC}} = S^{\mathrm{TV}}$ in case of $w = 0$). We should remark that the proposed technique can be applied to Problem IV.1: we can also select a high fidelity image from S^{TVC} by using our technique with a suitable design of the selector.

The optimization problems in the above two cases are solved by the Douglas-Rachford type algorithm. The iteration numbers of Algorithm III.2 and the Douglas-Rachford type algorithm are fixed to 500.

Tab. I and II show the comparison of PSNR [dB] and SSIM [21] of the results of the inpainting and compressive sensing recovery experiments, respectively. First, we found that all results restored by the TV-based recovery with our selection technique achieve higher PSNR and SSIM than those without. This fact implies that a high fidelity image can be selected appropriately by using our proposed selection technique. Second, from comparison between the results of the TVC-based recovery using $w = 1$ and those using a hand-optimized w, we found that the TVC-based recovery is sensitive to w (note that the hand-optimized w is different from each test image). On the other hand, the TV-based recovery with the proposed selection technique is free from the weight tuning because there does not exist any parameter corresponding to w. Moreover, as indicated in the tables, the recovery performance of the TV-based recovery with our selection technique is better for most of the images than that of the TVC-based recovery using the hand-optimized w in the both experiments. These observations imply that a simple (free from weight tuning) but very powerful (high recovery performance) image recovery method is obtained by formulating the hierarchical optimization and applying our selection technique rather than by formulating the weighted-sum of different convex priors.

We depict the result images of the inpainting experiment on the test image 'Barbara' in Fig. 2. Fig. 2(b) is restored by the TV-based recovery without our selection technique whose striped patterns are still corrupted. On the other hand, by using our selection technique, these patterns are seen to be restored efficiently as shown in Fig. 2(c). The result images of the compressive sensing recovery experiment on the test image 'Lena' are presented in Fig. 3. Note that Fig. 3(a) is the reconstructed image $\mathbf{\Phi}^\dagger\mathbf{v}$ ($\mathbf{\Phi}^\dagger$ denotes the Moore-Penrose pseudo inverse of $\mathbf{\Phi}$) instead of $\mathbf{v}$ itself, because $\mathbf{v}$ is random sampled noiselet coefficients and is meaningless to human eyes. We notice that there exist staircase effects in the recovered image of the TV-based recovery without our technique (Fig. 3(b)). By contrast, Fig. 3(b), where we used our selection technique, is seen to have clear edges and lines without any noticeable visual artifact.

V. CONCLUSION

We have proposed a technique for selecting a high fidelity image from the solution set of a convex optimization problem associated with existing image recovery methods. Our proposed hierarchical convex optimization problem has been designed for selecting a solution corresponding to a high fidelity image among all possible solutions of the convex optimization problem. Thanks to the characterization of the solution set by using the nonexpansive mapping derived from the Douglas-Rachford splitting type algorithms, we could utilize the hybrid steepest descent method for solving the hierarchical optimization problem. The numerical results have revealed not only that our technique appropriately selects a high fidelity image, but also that optimizing the hierarchical problem by using our technique achieves high recovery performance even it is free from any weight tuning.

Finally we remark again: our selection technique can be integrated with all image recovery methods whose convex optimization problems can be solved by the Douglas-Rachford type algorithms, and improve their performance by our technique with a suitable design of the selector, for example, the

(a) $\mathbf{v}$. (b) Without our technique. (c) **With our technique.**

Fig. 2. 'Barbara' images of the inpainting experiment: (a) shows the image with 60% pixels. There exist certain artifacts (disconnected patterns) in (b) which is the result of the TV-based recovery without our selection technique. In contrast, a recovered image with clear edges and contours is obtained by applying our selection technique to the TV-based recovery as shown in (c).

(a) $\mathbf{\Phi}^{\dagger}\mathbf{v}$. (b) Without our technique. (c) **With our technique.**

Fig. 3. 'Lena' images of the compressed sensing recovery experiment: (a) is the reconstructed image by using the pseudo-inverse of $\mathbf{\Phi}$ from the noisy randomly-sampled measurements $\mathbf{v}$. Compared to result (b), there is no staircasing effects in (c) selected by our technique.

regularization function proposed in [15], [17] may be also useful for selecting a high fidelity image with textures.

ACKNOWLEDGMENT

This work was partially supported by Grant-in-Aid for JSPS Fellows.

REFERENCES

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[2] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[3] S. Becker, J. Bobin, and E. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, pp. 1–39, 2011.

[4] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multi. Model. and Sim.*, vol. 5, no. 3, pp. 861–899, 2006.

[5] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. and Vis.*, vol. 20, pp. 89–97, 2004.

[6] C. Chaux, J. -C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imag. Sci.*, vol. 2, pp. 730–762, 2009.

[7] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Applied and Comput. Harmonic Anal.*, vol. 10, pp. 27–44, 2001.

[8] P. L. Combettes and J. -C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Topics in Signal Process.*, vol. 1, pp. 564–574, 2007.

[9] F. X. Dụbe, M. J. Fadili, and J. L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310–321, 2009.

[10] S. Durand, J. Fadili, and M. Nikolova, "Multiplicative noise removal using L1 fidelity on frame coefficients," *J. Math. Imag. Vis.*, vol. 36, pp. 201–226, 2010.

[11] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[12] S. Mallat, *A Wavelet Tour of Signal Processing, 2nd ed.*, Academic Press, 1999.

[13] J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Ser. A Math.*, vol. 255, pp. 2897–2899, 1965.

[14] S. Ono, T. Miyata, and K. Yamaoka, "Total Variation-Wavelet-Curvelet Regularized Optimization for Image Restoration," in *Proc. of IEEE ICIP*, pp. 2721–2724, 2011.

[15] S. Ono, T. Miyata, I. Yamada, and K. Yamaoka, "Missing Region Recovery by Promoting Blockwise Low-Rankness," in *Proc. of IEEE ICASSP*, pp. 1281–1284, 2012.

[16] S. Ono, T. Miyata, I. Yamada, and K. Yamaoka, "Image recovery by decomposition with component-wise regularization," *IEICE Trans. Fundamentals.*, vol. E95-A, no. 12, 2012 (to appear).

[17] S. Ono, T. Miyata, and I. Yamada, "Block nuclear norm: a convex regularization function with texture refinement capability," submitted for publication.

[18] N. Pustelnik, C. Chaux, and J. -C. Pesquet, "Parallel Proximal Algorithm for Image Restoration Using Hybrid Regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, 2011.

[19] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[20] G. Steidl and T. Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods," *J. Math. Imag. Vis.*, vol. 36, pp. 168–184, 2010.

[21] 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.

[22] I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," *Inherently Parralel Algorithm for Feasibility and Opt. and Their App.*, (D. Butnariu *et al*, eds.), pp. 473–504, Elsevier, 2001.

[23] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings," *Fixed-Point Algorithm for Inverse Problems in Science and Engineering* (H. H. Bauschke *et al*, eds.), pp. 345–390, Springer-Verlag, 2011.

[24] M. Yamagishi and I. Yamada, "Overrelaxation of the fast iterative shrinkage-thresholding algorithm with variable stepsize," *IOP Science Inverse Probl.*, vol. 27, 105008, 2011.

[25] K. Yosida, Functional Analysis, 4th edn. Springer, 1974.