

A Fast Optimization Transfer Algorithm for Image Inpainting in Wavelet Domains

Raymond H. Chan, You-Wei Wen, Andy M. Yip

Abstract

A wavelet inpainting problem refers to the problem of filling in missing wavelet coefficients in an image. A variational approach was used in Chan, Shen and Zhou (Total variation wavelet inpainting, *J. Math. Imaging Vision*, 25(1):107–125, 2006). The resulting functional was minimized by the gradient descent method. In this paper, we use an optimization transfer technique which involves replacing their univariate functional by a bivariate functional by adding an auxiliary variable. Our bivariate functional can be minimized easily by alternating minimization: for the auxiliary variable, the minimum has a closed form solution; and for the original variable, the minimization problem can be formulated as a classical total variation (TV) denoising problem, and hence can be solved efficiently using a dual formulation. We show that our bivariate functional is equivalent to the original univariate functional. We also show that our alternating minimization is convergent. Numerical results show that the proposed algorithm is very efficient and outperforms that in Chan, Shen and Zhou.

Index Terms

Total variation, image inpainting, wavelet, alternating minimization, optimization transfer.

I. INTRODUCTION

Image inpainting is used in repairing damaged pictures or removing unnecessary elements from pictures. It is an important image processing task in many real-life applications such as film restoration, text removal, scratch removal, and special effects in movies [14].

Inpainting may be done in the pixel domain or in a transformed domain. In 2000, Bertalmio *et al.* [4] considered restoring damaged paintings and photographs in the pixel domain. In general, the observed image \mathbf{g} can be described

Chan is with Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong. The research was supported in part by HKRGC Grant 400505 and CUHK DAG 2060257. E-mail: rchan@math.cuhk.edu.hk.

Wen is with Faculty of Science, South China Agricultural University, Wushan, Guangzhou, P. R. China. Present affiliation: Centre for Wavelets, Approximation and Information Processing, Temasek Laboratories, National University of Singapore, 3, Science Drive 2, 117543, Singapore. E-mail: tslwyw@nus.edu.sg. Research supported in part by NSFC Grant No. 60702030 and the wavelets and information processing program under a grant from DSTA, Singapore.

Yip is with Department of Mathematics, National University of Singapore, 3, Science Drive 2, 117543, Singapore. E-mail: andyyip@nus.edu.sg. Research supported in part by Academic Research Grant R146-000-116-112 from NUS, Singapore.

by:

$$g_{r,s} = \begin{cases} f_{r,s} + n_{r,s}, & (r,s) \in I, \\ 0, & (r,s) \in \Omega \setminus I. \end{cases}$$

Here f and n are the original noise-free image and the Gaussian white noise respectively, Ω is the complete index set, I is the set of observed pixels and $I \subset \Omega$ is the inpainting domain. Without loss of generality, we assume that the size of the image is $n \times n$, but all discussions can be equally applied to images of size $n \times m$. Bertalmio *et al.* [4] used partial differential equations (PDE) to smoothly propagate information from the surrounding areas along the isophotes into the inpainting domain. Ballester *et al.* developed a variational inpainting model based on a joint cost functional on the gradient vector field and gray values [2]. Chan and Shen considered the Total Variational (TV) inpainting model [12] and Curvature Driven Diffusion (CDD) model [13]. The TV inpainting model stems from the well-known Rudin-Osher-Fatemi's image model [33] and it fills in the missing regions such that the TV is minimized. Chan, Kang and Shen introduced an inpainting technique using an Euler's elastica energy-based variational model [11]. All these researches focused on inpainting in the pixel domain.

Inpainting in wavelet domains is a completely different problem since there are no well-defined inpainting regions in the pixel domain. After the release of the new image compression standard JPEG2000, many images are formatted and stored in terms of wavelet coefficients. During storing or transmitting, some wavelet coefficients may be lost or corrupted. This prompts the need of restoring the missing information in the wavelet domain. Inspired by the practical need, Chan, Shen and Zhou studied image inpainting problems in wavelet domains [15]. Let us denote the standard orthogonal wavelet expansion of g and f by

$$\mathbf{g}(\boldsymbol{\alpha}) = \sum_{j,k} \alpha_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2,$$

and

$$\mathbf{f}(\boldsymbol{\beta}) = \sum_{j,k} \beta_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2,$$

where $\{\psi_{j,k}\}$ denotes the wavelet basis [21], and $\{\alpha_{j,k}\}, \{\beta_{j,k}\}$ are the wavelet coefficients of g and f given by

$$\alpha_{j,k} = \langle \mathbf{g}, \psi_{j,k} \rangle \quad \text{and} \quad \beta_{j,k} = \langle \mathbf{f}, \psi_{j,k} \rangle, \quad (1)$$

for $j \in \mathbb{Z}, k \in \mathbb{Z}^2$. For convenience, we denote $\mathbf{f}(\boldsymbol{\beta})$ by f when there is no ambiguity. Assume that the wavelet coefficients in the index region I are known, that is, the available wavelet coefficients are given by

$$\xi_{j,k} = \begin{cases} \alpha_{j,k}, & (j,k) \in I, \\ 0, & (j,k) \in \Omega \setminus I. \end{cases}$$

The aim of wavelet domain inpainting is to reconstruct the wavelet coefficients of f using the given coefficients ξ . It is well-known that the inpainting problem is ill-posed, i.e. it admits more than one solution. There are many different ways to fill in the missing coefficients, and therefore we have different reconstructions in the pixel domain. A regularization method can be used to fill in the missing wavelet coefficients. Since TV minimization can preserve

sharp edges while reducing noises and other oscillations in the reconstruction, Rudin, Osher and Fatemi [33] proposed to use TV regularization to solve the denoising problem in the pixel domain:

$$\min_{\mathbf{f}} \|\mathbf{f} - \mathbf{g}\|_2^2 + \lambda \|\mathbf{f}\|_{TV}. \quad (2)$$

Here $\|\cdot\|_{TV}$ is the discrete TV norm. Let us define the discrete gradient operator $\nabla_{\pm} : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$ by

$$(\nabla_{\pm} \mathbf{f})_{r,s} = \left((\nabla_{\pm}^x \mathbf{f})_{r,s}, (\nabla_{\pm}^y \mathbf{f})_{r,s} \right)$$

with

$$(\nabla_{\pm}^x \mathbf{f})_{r,s} = \pm (f_{r\pm 1,s} - f_{r,s})$$

and

$$(\nabla_{\pm}^y \mathbf{f})_{r,s} = \pm (f_{r,s\pm 1} - f_{r,s})$$

for $r, s = 1, \dots, n$. We use the reflective boundary condition such that $f_{0,s} = f_{1,s}$, $f_{n+1,s} = f_{n,s}$, $f_{r,0} = f_{r,1}$ and $f_{r,n+1} = f_{r,n}$ for $r, s = 1, \dots, n$. Here $f_{r,s}$ refers to the $((r-1)n+s)$ -th entry of the vector \mathbf{f} (it is the (r, s) -th pixel location of the image). The discrete TV of \mathbf{f} is defined by

$$\begin{aligned} \|\mathbf{f}\|_{TV} &= \sum_{1 \leq r, s \leq n} |(\nabla_{+} \mathbf{f})_{r,s}| \\ &= \sum_{1 \leq r, s \leq n} \sqrt{\left|(\nabla_{+}^x \mathbf{f})_{r,s}\right|^2 + \left|(\nabla_{+}^y \mathbf{f})_{r,s}\right|^2}, \end{aligned}$$

where $|\cdot|$ is the Euclidean norm in \mathbb{R}^2 .

TV norm has become very popular for regularization approaches. In [15], Chan, Shen and Zhou selected TV norm to facilitate the wavelet inpainting problem so that the missing or damaged coefficients can be filled in faithfully to preserve sharp edges in the pixel domain. Precisely, they considered the following minimization problem

$$\min_{\beta} \mathcal{J}(\beta) \equiv \min_{\beta} \sum_{j,k} \chi_{j,k} (\xi_{j,k} - \beta_{j,k})^2 + \lambda \|\mathbf{f}\|_{TV}, \quad (3)$$

with $\chi_{j,k} = 1$ if $(j, k) \in I$ and $\chi_{j,k} = 0$ if $(j, k) \in \Omega \setminus I$, and λ is the regularization parameter. The first term in $\mathcal{J}(\beta)$ is the data-fitting term and the second is the TV regularization term.

There are many methods available in the literature to find the minimizer of total variation denoising problems in the pixel domain. These methods include PDE based methods such as explicit [33], semi-implicit [27], operator splitting [28], and lagged diffusivity fixed point iterations [34]. However, solving TV regularization problems using these methods poses a numerical difficulty due to the non-differentiability of the TV norm. This difficulty can be overcome by introducing a small positive parameter ϵ in the TV norm, which prevents the denominator from vanishing in numerical implementations, i.e.,

$$\|\mathbf{f}\|_{TV,\epsilon} = \sum_{1 \leq r, s \leq n} \sqrt{\left|(\nabla_{+}^x \mathbf{f})_{r,s}\right|^2 + \left|(\nabla_{+}^y \mathbf{f})_{r,s}\right|^2 + \epsilon}.$$

The specification of ϵ involves trade-offs between the quality of edges restored and the speed in converging to the fixed point. Precisely, the smaller ϵ is, the higher quality of the restoration on the edges will be. However, the

convergence of fixed point iteration becomes slower as the modified TV norm $\|f\|_{TV,\epsilon}$ is closer to the original non-differentiable TV norm in (2). An alternative way to get around the non-differentiability of TV norm is to reformulate the TV denoising problem as a minimum graph cut problem [20]. In this approach, the original TV norm is replaced by an anisotropic TV norm.

To use the original TV norm in the formulation and avoid the numerical difficulty, Carter [8] and Chambolle [9] studied a dual formulation of (2). Chambolle showed that the solution of (2) is given by the orthogonal projection of the observed image onto a convex set derived in the dual formulation of (2). Therefore computing the solution of (2) hinges on computing a nonlinear projection. Chambolle further developed a semi-implicit gradient descent algorithm to solve the constrained minimization problem arising from the nonlinear projection. Based on the theories in semismooth operators, Ng *et al.* [29] studied semismooth Newton's methods for computing the nonlinear projection. Multilevel optimization methods are considered in [10], [16]. The convergence and numerical results have shown that these duality based algorithms are quite effective. However, since the first term in (3) is rank deficient, it is difficult to directly extend the dual approach to find the minimizer of the wavelet inpainting problem in (3).

In this paper, we propose a new efficient optimization transfer algorithm to solve (3). An auxiliary variable ζ is introduced and the new objective function is

$$\mathcal{J}_2(\zeta, \beta) = \frac{1+\tau}{\tau} (\|\chi(\zeta - \xi)\|_2^2 + \tau \|\zeta - \beta\|_2^2) + \lambda \|f(\beta)\|_{TV},$$

where χ denotes a diagonal matrix with diagonal entries $\chi_{j,k}$ and τ is a positive regularization parameter. The function $\mathcal{J}_2(\zeta, \beta)$ is a quadratic majorizing function [25] of $\mathcal{J}(\beta)$. We will show in the following section that for any positive regularization parameter τ , we have

$$\mathcal{J}(\beta) = \min_{\zeta} \mathcal{J}_2(\zeta, \beta).$$

Thus the minimization of \mathcal{J}_2 w.r.t ζ and β is equivalent to the minimization of \mathcal{J} w.r.t β . Moreover, the equivalency holds for any $\tau > 0$. It has been a popular approach to introduce an auxiliary variable into the original optimization problem. For example, half-quadratic algorithms [22], [23], [30], majorization-minimization algorithms [6], [25] were proposed to linearize the non-linear terms and separate the parameters in the optimization problem. In these algorithms, the subproblem for each variable is linear and therefore simple to solve. However, these methods cannot be readily extended to inpainting problems.

In our model, the subproblem for the variable ζ is linear, but that for the original variable β is still non-linear. One key observation is that, by using the unitary invariance property of the L_2 norm, computing a minimizer β of the subproblem is equivalent to solving a TV denoising problem in (2). Therefore, we can use a dual approach to find the minimizer. Since we use an alternating minimization approach to compute a minimizer of the objective function \mathcal{J}_2 , we do not need to solve each subproblem exactly. Instead, we simply need to reduce the associated objective function, which can be achieved by running a few steps of the non-linear projection.

Recently, several authors proposed the use of bivariate functionals and alternating minimization for solving various TV minimization problems [7], [24]. However, convergence theories there do not hold in the case of TV inpainting. Moreover, their bivariate functionals are not exactly equivalent to the original functional.

This paper is arranged as follows. In Section II, we review the numerical algorithm on total variation wavelet inpainting presented by Chan, Shen and Zhou [15]. In Section III, we derive a quadratic majorizing function for the data-fitting term in $\mathcal{J}(\beta)$ and propose a bivariate functional together with an alternating minimization algorithm to find its minimizer. In Section IV, numerical results are given to demonstrate the efficiency of our algorithm. In Section V, we give a conclusion. In the appendix, we prove the convergence of our alternating minimization algorithm.

II. EXPLICIT GRADIENT DESCENT ALGORITHM

In this section, we review the numerical algorithm on total variation wavelet inpainting presented in [15]. In their paper, Chan, Shen and Zhou derived the Euler Lagrange equation for the wavelet inpainting problem. An explicit gradient descent scheme has been used to solve (3) in a primal setting. The cost function in (3) achieves the global minimum when its gradient equals zero, i.e.,

$$\nabla_{\beta} \mathcal{J}(\beta) = 0. \quad (4)$$

The gradient of the objective function in the continuous setting can be calculated as follows:

$$\begin{aligned} & \frac{\partial \mathcal{J}(\beta)}{\partial \beta_{j,k}} \\ &= 2\chi_{j,k}(\beta_{j,k} - \xi_{j,k}) + \lambda \int_{\mathbb{R}^2} \frac{\nabla f(\beta)}{|\nabla f(\beta)|_\epsilon} \cdot \frac{\partial \nabla f(\beta)}{\partial \beta_{j,k}} dx \\ &= 2\chi_{j,k}(\beta_{j,k} - \xi_{j,k}) + \lambda \int_{\mathbb{R}^2} \frac{\nabla f(\beta)}{|\nabla f(\beta)|_\epsilon} \cdot \nabla \psi_{j,k} dx \end{aligned}$$

for $j \in \mathbb{Z}, k \in \mathbb{Z}^2$. Denote the regularized mean curvature of the image f by

$$\kappa = \nabla \cdot \frac{\nabla f}{|\nabla f|_\epsilon}. \quad (5)$$

If we assume that the mother wavelet ψ is compactly supported and at least Lipschitz continuous, then using integration by parts yields

$$\begin{aligned} & \frac{\partial \mathcal{J}(\beta)}{\partial \beta_{j,k}} \\ &= 2\chi_{j,k}(\beta_{j,k} - \xi_{j,k}) - \lambda \int_{\mathbb{R}^2} \left(\nabla \cdot \frac{\nabla f(\beta)}{|\nabla f(\beta)|_\epsilon} \right) \psi_{j,k} dx \\ &= 2\chi_{j,k}(\beta_{j,k} - \xi_{j,k}) - \lambda \langle \kappa, \psi_{j,k} \rangle. \end{aligned}$$

Note that the term $\langle \kappa, \psi_{j,k} \rangle$ is the curvature projected on the wavelet basis. The problem in (4) can be solved by the following gradient flow

$$\frac{d\beta}{dt} = -\nabla_{\beta} \mathcal{J}(\beta). \quad (6)$$

The system of differential equations (6) can be easily solved by the following explicit scheme:

$$\beta_{j,k}^{(i+1)} = \beta_{j,k}^{(i)} - \delta t (2\chi_{j,k}(\beta_{j,k} - \xi_{j,k}) - \lambda \langle \kappa, \psi_{j,k} \rangle) \quad (7)$$

for $j \in \mathbb{Z}, k \in \mathbb{Z}^2$, where δt is a time-step parameter to guarantee both the stability of the numerical solutions and an appropriate convergence speed.

Notice that the curvature is defined in the pixel domain. In practice, we compute it directly by transforming the coefficients to the pixel domain, and then transforming it back to the wavelet domain. More precisely, we first calculate

$$\mathbf{f}(\boldsymbol{\beta}^{(i)}) = \sum_{j,k} \beta_{j,k}^{(i)} \psi_{j,k}(x), \quad (8)$$

Next, we obtain the curvature of \mathbf{f} by using the standard finite difference method:

$$\begin{aligned} \kappa_{r,s} &= \nabla_-^x \cdot \left(\frac{(\nabla_+^x f)_{r,s}}{\sqrt{|(\nabla_+^x f)_{r,s}|^2 + |(\nabla_+^y f)_{r,s}|^2 + \epsilon}} \right) + \\ &\quad \nabla_-^y \cdot \left(\frac{(\nabla_+^y f)_{r,s}}{\sqrt{|(\nabla_+^x f)_{r,s}|^2 + |(\nabla_+^y f)_{r,s}|^2 + \epsilon}} \right). \end{aligned}$$

Finally, we compute the projection of the curvature projection onto the wavelet basis by

$$v_{j,k} = \langle \boldsymbol{\kappa}, \psi_{j,k} \rangle = \sum_{r,s} \kappa_{r,s} \psi_{j,k}(x_{r,s}). \quad (9)$$

We may rewrite the iterative formula (7) in matrix form:

$$\boldsymbol{\beta}^{(i)} = \mathbf{K}^{(i-1)} \boldsymbol{\beta}^{(i-1)} + \mathbf{r},$$

where $\mathbf{K}^{(i-1)}$ is a matrix depending on $\boldsymbol{\beta}^{(i-1)}$ and δt , and \mathbf{r} is a fixed vector depending on ξ . The time step δt should be chosen such that the spectral radius of $\mathbf{K}^{(i-1)}$ is as small as possible to ensure fast convergence. Thus, we must first estimate the spectrum $\rho(\mathbf{K}^{(i-1)})$ for each i and optimize it w.r.t. δt . This is difficult as the matrix $\mathbf{K}^{(i-1)}$ is very large and $\rho(\mathbf{K}^{(i-1)})$ is highly non-linear w.r.t. δt . In practice, the parameter δt is usually chosen to be very small in order to guarantee the decrease of the cost function. Another approach is to incorporate backtracking line search [31] into the gradient descent. This method is to find a parameter δt_i such that

$$\mathcal{J}(\boldsymbol{\beta}^{(i)} + \delta t_i \Delta \boldsymbol{\beta}) > \mathcal{J}(\boldsymbol{\beta}^{(i)}) + \mu \delta t_i \nabla \mathcal{J}^T \Delta \boldsymbol{\beta},$$

where $\Delta \boldsymbol{\beta} = -\nabla \mathcal{J}(\boldsymbol{\beta}^{(i)})$ and μ is a fixed positive parameter.

III. AN EFFICIENT TV MINIMIZATION ALGORITHM

A. A quadratic majorizing function for the data-fitting term

Now we derive a new quadratic majorizing function for data-fitting term in the objective function (3). Notice that the data-fitting term is already quadratic. Our hope is to reformulate the wavelet inpainting problem as a denoising problem, which can then be solved by classical numerical schemes efficiently.

We begin with the quadratic function

$$q(\zeta_{j,k}, \beta_{j,k}) = \chi_{j,k} (\zeta_{j,k} - \xi_{j,k})^2 + \tau (\zeta_{j,k} - \beta_{j,k})^2,$$

which is strictly convex. By differentiating $q(\zeta_{j,k}, \beta_{j,k})$ w.r.t $\zeta_{j,k}$, we find the minimizer $\bar{\zeta}_{j,k}$ of $q(\zeta_{j,k}, \cdot)$:

$$\begin{aligned}\bar{\zeta}_{j,k} &= \frac{1}{\chi_{j,k} + \tau} (\chi_{j,k} \xi_{j,k} + \tau \beta_{j,k}) \\ &= \begin{cases} \frac{1}{1+\tau} (\xi_{j,k} + \tau \beta_{j,k}), & (j, k) \in I, \\ \beta_{j,k}, & (j, k) \in \Omega \setminus I. \end{cases} \end{aligned} \quad (10)$$

Hence the minimum value of the function $q(\zeta_{j,k}, \beta_{j,k})$ for a fixed $\beta_{j,k}$ is given by

$$\begin{aligned}q(\bar{\zeta}_{j,k}, \beta_{j,k}) &= \begin{cases} \frac{\tau}{1+\tau} (\xi_{j,k} - \beta_{j,k})^2, & (j, k) \in I, \\ 0, & (j, k) \in \Omega \setminus I, \end{cases} \\ &= \frac{\tau}{1+\tau} \chi_{j,k} (\xi_{j,k} - \beta_{j,k})^2, \quad (j, k) \in \Omega.\end{aligned}$$

Denote by $\mathcal{Q}(\zeta, \beta)$ the sum of the function $q(\zeta_{j,k}, \beta_{j,k})$ over j, k . Therefore we have

$$\begin{aligned}\mathcal{Q}(\zeta, \beta) &:= \sum_{j,k} q(\zeta_{j,k}, \beta_{j,k}) \\ &= \sum_{j,k} \left[\chi_{j,k} (\zeta_{j,k} - \xi_{j,k})^2 + \tau (\zeta_{j,k} - \beta_{j,k})^2 \right] \\ &\geq \frac{\tau}{1+\tau} \sum_{j,k} \chi_{j,k} (\xi_{j,k} - \beta_{j,k})^2.\end{aligned}$$

The equality holds if and only if $\zeta_{j,k}$ satisfies (10) for all j, k . Hence, we obtain the following identity

$$\sum_{j,k} \chi_{j,k} (\xi_{j,k} - \beta_{j,k})^2 = \min_{\zeta} \frac{1+\tau}{\tau} \mathcal{Q}(\zeta, \beta).$$

Now we introduce a new objective function

$$\mathcal{J}_2(\zeta, \beta) := \frac{1+\tau}{\tau} \mathcal{Q}(\zeta, \beta) + \lambda \|f(\beta)\|_{TV}. \quad (11)$$

By the convexity of \mathcal{J} and \mathcal{J}_2 , the minimization problem for $\mathcal{J}(\beta)$ is equivalent to $\mathcal{J}_2(\zeta, \beta)$, i.e.,

$$\min_{\beta} \mathcal{J}(\beta) = \min_{\zeta, \beta} \mathcal{J}_2(\zeta, \beta). \quad (12)$$

The main difference between the two minimization problems in (12) is that a new variable ζ is introduced in \mathcal{J}_2 , so that the minimization w.r.t each variable is simple. In the index region $(j, k) \in I$, $\zeta_{j,k}$ is a weighted average of the noisy wavelet coefficient $\xi_{j,k}$ and the restored wavelet coefficient $\beta_{j,k}$, while in the index region $(j, k) \in \Omega \setminus I$, $\zeta_{j,k}$ is the restored wavelet coefficient $\beta_{j,k}$. Hence ζ can be regarded as an average of ξ and β . We remark that a new regularization parameter τ is introduced in the new minimization problem, but (12) holds for any τ . In Section IV, we will show that the quality of the restored image and CPU running time is insensitive to the parameter τ .

B. Alternating minimization method

We propose an alternating minimization algorithm to find a minimizer of $\mathcal{J}_2(\zeta, \beta)$. Starting from an initial guess $(\zeta^{(0)}, \beta^{(0)})$, we use an alternating minimization algorithm to generate the sequence

$$\begin{cases} \zeta^{(i)} := \mathcal{S}_1(\beta^{(i-1)}) = \operatorname{argmin}_{\zeta} \mathcal{J}_2(\zeta, \beta^{(i-1)}), \\ \beta^{(i)} := \mathcal{S}_2(\zeta^{(i)}) = \operatorname{argmin}_{\beta} \mathcal{J}_2(\zeta^{(i)}, \beta). \end{cases} \quad (13)$$

From (10), we obtain

$$\zeta^{(i)} = (\tau \mathbf{I} + \chi)^{-1} (\chi \xi + \tau \beta^{(i-1)}).$$

The diagonal matrix $\tau \mathbf{I} + \chi$ is non-singular for any $\tau > 0$. Next, we find the minimizer of $\mathcal{J}_2(\zeta^{(i)}, \cdot)$:

$$\beta^{(i)} = \operatorname{argmin}_{\beta} \left\{ (1 + \tau) \left\| \zeta^{(i)} - \beta \right\|_2^2 + \lambda \|f(\beta)\|_{TV} \right\}. \quad (14)$$

Here we focus on wavelet inpainting problems whose wavelet transform matrix \mathbf{W} is orthogonal.

Defining $\mathbf{u}^{(i)} = \mathbf{W}^{-1} \zeta^{(i)}$ and noticing that $\beta = \mathbf{W} f$, the minimization problem (14) becomes

$$\beta^{(i)} = \mathbf{W} \cdot \operatorname{argmin}_f \left\{ (1 + \tau) \left\| \mathbf{W} (\mathbf{u}^{(i)} - f) \right\|_2^2 + \lambda \|f\|_{TV} \right\}.$$

Exploiting the unitary invariance property of the L_2 norm, we can discard the multiplication by \mathbf{W} in the first term in the argmin and get

$$\beta^{(i)} = \mathbf{W} \cdot \operatorname{argmin}_f \left\{ (1 + \tau) \left\| \mathbf{u}^{(i)} - f \right\|_2^2 + \lambda \|f\|_{TV} \right\}.$$

This is exactly the standard TV denoising problem, which can be solved by many TV minimization methods mentioned in Section I. In this paper, we employ the Chambolle's projection algorithm in the denoising step because of its simplicity and efficiency. In this scheme, we solve the following dual constrained minimization problem:

$$\min_{\mathbf{p}} \left\| \mathbf{u}^{(i)} - \frac{2\lambda}{1 + \tau} \operatorname{div} \mathbf{p} \right\|_2^2 \quad (15)$$

subject to

$$|p_{r,s}| \leq 1, \quad \forall 1 \leq r, s \leq n.$$

Here

$$p_{r,s} = \begin{bmatrix} p_{r,s}^x \\ p_{r,s}^y \end{bmatrix}$$

is the dual variable of the (r, s) -th pixel, \mathbf{p} is the concatenation of all $p_{r,s}$, and the discrete divergence of \mathbf{p} is defined as:

$$(\operatorname{div} \mathbf{p})_{r,s} \equiv p_{r,s}^x - p_{r-1,s}^x + p_{r,s}^y - p_{r,s-1}^y$$

with $p_{0,s}^x = p_{r,0}^y = 0$ for $r, s = 1, \dots, n$. The vector $\operatorname{div} \mathbf{p}$ is the concatenation of all $(\operatorname{div} \mathbf{p})_{r,s}$. When the minimizer \mathbf{p}^* of the constrained optimization problem in (15) is determined, the denoised image $f^{(i)}$ can be obtained as follows:

$$f^{(i)} = \mathbf{u}^{(i)} - \gamma \operatorname{div} \mathbf{p}^*, \quad (16)$$

where $\gamma = 2\lambda/(1 + \tau)$. In [9], the iterative scheme for computing the optimal solution \mathbf{p} is given by:

$$p_{r,s}^{l+1} = \frac{p_{r,s}^l + \delta \gamma \nabla (\gamma \operatorname{div} \mathbf{p}^l - \mathbf{u}^{(i)})_{r,s}}{1 + \delta \gamma \left| \nabla (\gamma \operatorname{div} \mathbf{p}^l - \mathbf{u}^{(i)})_{r,s} \right|}, \quad \forall 1 \leq r, s \leq n, \quad (17)$$

where $p_{r,s}^l$ is the l -th iteration for the minimizer, and $\delta \leq \frac{1}{8}$ is the step size introduced in the projection gradient method, see [9] for details.

There is a natural weakness in using a dual formulation, because (16) only holds when the sequence $\{\mathbf{p}^l\}$ converges. Moreover, (17) does not correspond to a gradient descent process of the original primal objective function. Thus an early termination of (17) may increase the value of the primal objective function. Thus, after some l iterations on (17), we apply a relaxed step of the form

$$\mathbf{f}^{(i)} = \mathbf{f}^{(i-1)} + \delta t \left(\mathbf{u}^{(i)} - \gamma \operatorname{div} \mathbf{p}^l - \mathbf{f}^{(i-1)} \right),$$

where $\delta t \leq 1$ is the step size. We can apply a backtracking line search to find a proper choice for the value of δt .

We remark that the preceding discussion is based on an orthogonal wavelet transform. This method can be generalized to non-orthogonal wavelet transforms, for example, bi-orthogonal wavelet transforms [17], redundant transforms [26] and tight frames [32]. In these cases, \mathbf{W} is not orthogonal, but still has full rank, and hence $\mathbf{W}^T \mathbf{W}$ is invertible. Therefore the iteration can be modified to:

$$p_{r,s}^{l+1} = \frac{p_{r,s}^l + \delta \gamma \nabla \left(\gamma (\mathbf{W}^T \mathbf{W})^{-1} (\operatorname{div} \mathbf{p}^l - \mathbf{W}^T \zeta^{(i)}) \right)_{r,s}}{1 + \delta \gamma \left| \nabla \left(\gamma (\mathbf{W}^T \mathbf{W})^{-1} (\operatorname{div} \mathbf{p}^l - \mathbf{W}^T \zeta^{(i)}) \right)_{r,s} \right|}.$$

In order to guarantee the convergence, the step size δ should satisfy $8\delta \left\| (\mathbf{W}^T \mathbf{W})^{-1} \right\|_2 \leq 1$.

The convergence of the alternating minimization method is given in the Appendix.

IV. NUMERICAL RESULTS

We illustrate the performance of the proposed algorithm for image inpainting problems and compare it with the gradient descent method proposed by Chan, Shen and Zhou [15]. Our codes are written in MATLAB R2008a. Signal-to-Noise Ratio (SNR) is used to measure the quality of the restored images. It is defined as follows:

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\mathbf{f}\|_2}{\|\tilde{\mathbf{f}} - \mathbf{f}\|_2} \right),$$

where \mathbf{f} and $\tilde{\mathbf{f}}$ are the original image and the restored image respectively. We compare the proposed method with the gradient descent method in [15].

In all the tests, all images were corrupted by white Gaussian noise with standard deviation $\sigma = 10$. We set the initial guess of β to the available wavelet coefficients ζ . We used Daubechies 7-9 bi-orthogonal wavelets with symmetric extensions at the boundaries [1], [17]. To reduce the time in searching for a good regularization parameter τ , we fix τ to 2×10^{-2} . The gradient descent method is employed using the modified TV with $\varepsilon = 10^{-3}$. For the regularization parameter, we tried $\lambda = 2, 4, 6, 8, 10$, and pick the one such that the SNR of the restored image is optimized.

In the first example, we consider the ‘‘Cameraman’’ image which is corrupted by an additive noise. The corrupted image with 50% wavelet coefficients randomly missing is shown in Fig.1(b). Since some low-frequency wavelet coefficients are missing, there are large corrupted regions in the pixel domain. It is difficult to define an inpainting region in the pixel domain. Fig.1 shows the original image, the damaged image, and the images restored by using gradient descent with time step $\delta t = 0.32, 0.16, 0.08, 0.04$ with line search, and our method. Since all methods

correspond to the same optimization problem, the restored images have little visible differences when the number of iteration goes to infinity.

In Fig. 2, we plot the SNRs of the restored images and the value of the objective function computed within the iterations. We observe that the proposed method produces the best SNRs, and the gradient descent method with a larger time step produces better SNRs. We also observe that the value of the objective function may stagnate at a high value when a large time step is chosen.

In the second and third examples, we use the “Lena” image and the “Barbara” image respectively. These images have a nice mixture of details, flat regions, shading area and textures. The original image, the damaged image and restored images are shown in Fig. 3 and Fig. 5. The plots of SNR and objective value versus CPU time are shown in Fig. 4 and 6.

In these experiments, we also test how the parameter τ in \mathcal{J}_2 affects the CPU time. We use $\tau = 1 \times 10^{-j}$ ($j = -4, -3, \dots, 2$) and plot the CPU running time versus the parameter for the “Cameraman”, “Lena” and “Barbara” images in Fig. 7 respectively. We find that for a broad range of values of the parameter $\tau < 1 \times 10^{-1}$, it has no significant influence on the CPU time. Thus any small values τ can be used to obtain a good performance.

V. CONCLUSION

In this paper, we present an efficient algorithm to fill in missing or damaged wavelet coefficients due to lossy image transmission or communication. This algorithm is derived from the state-of-the-art optimization transfer approach, in which an auxiliary variable is introduced to the objective function. We have employed an alternating minimization method to solve the proposed bivariate problem. The subproblem of the auxiliary variable has a closed form solution, which is a weighted average of the noisy wavelet coefficient and the restored wavelet coefficient. By using the unitary invariance property of the L_2 norm, the subproblem of the original variable becomes a classical TV denoising problem which can be solved efficiently using nonlinear projection. The proposed method avoids differentiating the TV term as in the gradient descent method [15]. It allows us to use the original TV without modifying it to $\|\mathbf{f}\|_{TV,\epsilon}$. Our experimental results show that the proposed algorithm is very efficient and outperforms the gradient descent method.

APPENDIX A CONVERGENCE ANALYSIS

In this section, we study the convergence of the alternating minimization algorithm. We remark that convergence of alternating minimization is non-trivial in the case of non-differentiable objective functions even if they are strictly convex. Although our bivariate objective function is non-differentiable, it is constructed such that the algorithm still converges to a minimum. Starting from an arbitrary $\boldsymbol{\beta}^{(0)}$, we consider the sequence $\{\boldsymbol{\beta}^{(i)}\}$ generated by (13). We first show that it is convergent.

Theorem 1: The sequence $\{\boldsymbol{\beta}^{(i)}\}$ generated by $\boldsymbol{\beta}^{(i)} = \mathcal{S}_2(\mathcal{S}_1(\boldsymbol{\beta}^{(i-1)}))$ converges to a $\boldsymbol{\beta}^* \in \mathbb{R}^{n^2}$ for any $\boldsymbol{\beta}^{(0)} \in \mathbb{R}^{n^2}$.



Fig. 1. (a) Original cameraman image. (b) Damaged image with 50% wavelet coefficients randomly lost. (c)–(f) Restored image by the gradient descent method with $\delta t = 0.32, 0.16, 0.08, 0.04$ respectively. (g) Restored image by the gradient descent method with linear search. (h) Restored image by the proposed method.

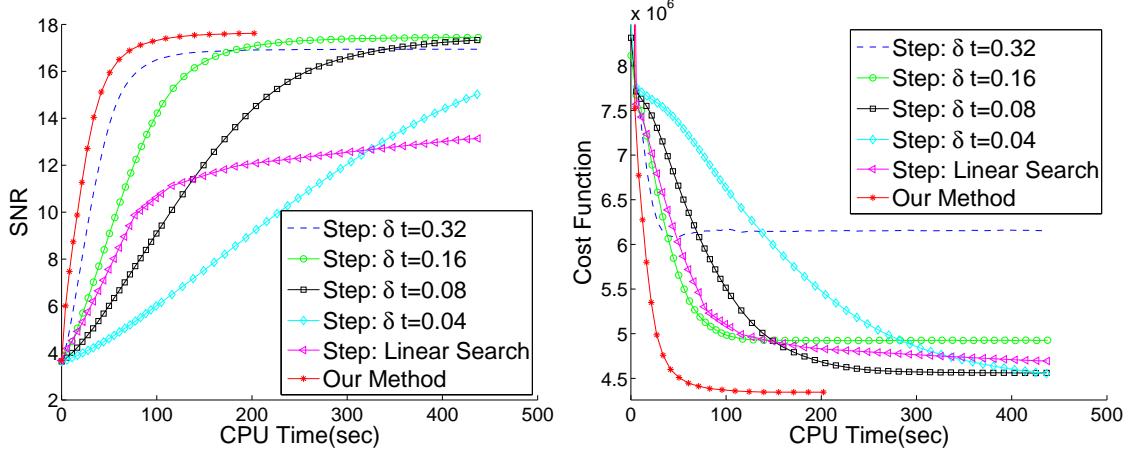


Fig. 2. Convergence profiles for the Cameraman image. (a) SNR versus CPU time. (b) The objective value versus CPU time.

Proof: Recall that

$$\mathcal{J}_2(\zeta, \beta) = \frac{1+\tau}{\tau} \mathcal{Q}(\zeta, \beta) + \lambda \|\mathbf{f}(\beta)\|_{TV}.$$

Then

$$\begin{aligned} & \mathcal{J}_2(\zeta^{(i)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \\ &= \frac{1+\tau}{\tau} \mathcal{Q}(\zeta^{(i)}, \beta^{(i-1)}) - \frac{1+\tau}{\tau} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) \\ & \quad + \lambda \|\mathbf{f}(\beta^{(i-1)})\|_{TV} - \lambda \|\mathbf{f}(\beta^{(i)})\|_{TV}. \end{aligned} \quad (18)$$

Since $\|\mathbf{f}(\beta)\|_{TV}$ is convex w.r.t β , we deduce that

$$\|\mathbf{f}(\beta^{(i-1)})\|_{TV} - \|\mathbf{f}(\beta^{(i)})\|_{TV} \geq (\beta^{(i-1)} - \beta^{(i)})^T \mathbf{s}, \quad (19)$$

where \mathbf{s} is an arbitrary element of the subgradient $\frac{\partial}{\partial \beta} \|\mathbf{f}(\beta)\|_{TV}$. Next, by Taylor's expansion on the quadratic function $\mathcal{Q}(\zeta^{(i)}, \cdot)$, we obtain

$$\begin{aligned} & \mathcal{Q}(\zeta^{(i)}, \beta^{(i-1)}) - \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) \\ &= (\beta^{(i-1)} - \beta^{(i)})^T \nabla_{\beta} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) + \\ & \quad \frac{1}{2} (\beta^{(i-1)} - \beta^{(i)})^T \nabla_{\beta}^2 \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) (\beta^{(i-1)} - \beta^{(i)})^T \\ &= (\beta^{(i-1)} - \beta^{(i)})^T \nabla_{\beta} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) + \\ & \quad \frac{\tau+1}{2} \|\beta^{(i-1)} - \beta^{(i)}\|_2^2. \end{aligned}$$



Fig. 3. (a) Original Lena image. (b) Damaged image with 50% wavelet coefficients randomly lost. (c)–(f) Restored image by the gradient descent method with $\delta t = 0.32, 0.16, 0.08, 0.04$ respectively. (g) Restored image by the gradient descent method with linear search. (h) Restored image by the proposed method.

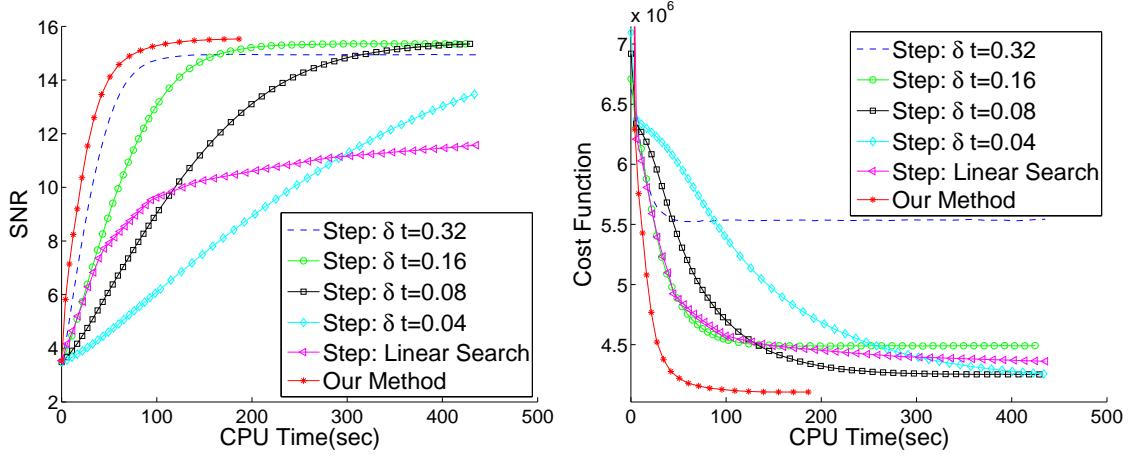


Fig. 4. Convergence profiles for the Lena image. (a) SNR versus CPU time. (b) The objective value versus CPU time.

Using (18) and (19), we have

$$\begin{aligned} & \mathcal{J}_2(\zeta^{(i)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \\ \geq & (\beta^{(i)} - \beta^{(i)})^T \left(\frac{1+\tau}{\tau} \nabla_{\beta} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) + \lambda s \right) + \\ & \frac{(\tau+1)^2}{2\tau} \|\beta^{(i-1)} - \beta^{(i)}\|_2^2. \end{aligned}$$

Since $\beta^{(i)}$ is the minimizer of the cost function $\mathcal{J}_2(\zeta^{(i)}, \cdot)$, we have

$$0 \in \frac{\partial}{\partial \beta} \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}),$$

where

$$\frac{\partial}{\partial \beta} \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) = \frac{1+\tau}{\tau} \frac{\partial}{\partial \beta} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) + \lambda \frac{\partial}{\partial \beta} \|f(\beta^{(i)})\|_{TV}.$$

By choosing $s \in \frac{\partial}{\partial \beta} \|f(\beta^{(i)})\|_{TV}$ such that $\frac{1+\tau}{\tau} \frac{\partial}{\partial \beta} \mathcal{Q}(\zeta^{(i)}, \beta^{(i)}) + \lambda s = 0$, we have

$$\mathcal{J}_2(\zeta^{(i)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \geq \frac{(\tau+1)^2}{2\tau} \|\beta^{(i-1)} - \beta^{(i)}\|_2^2.$$

Notice that $\zeta^{(i)} = \operatorname{argmin}_{\zeta} \mathcal{J}_2(\zeta, \beta^{(i-1)})$, we deduce $\mathcal{J}_2(\zeta^{(i-1)}, \beta^{(i-1)}) \geq \mathcal{J}_2(\zeta^{(i)}, \beta^{(i-1)})$. Hence

$$\begin{aligned} & \mathcal{J}_2(\zeta^{(i-1)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \\ \geq & \mathcal{J}_2(\zeta^{(i)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \\ \geq & \frac{(\tau+1)^2}{2\tau} \|\beta^{(i-1)} - \beta^{(i)}\|_2^2. \end{aligned}$$



Fig. 5. (a) Original Barbara image. (b) Damaged image with 50% wavelet coefficients randomly lost. (c)–(f) Restored image by the gradient descent method with $\delta t = 0.32, 0.16, 0.08, 0.04$ respectively. (g) Restored image by the gradient descent method with linear search. (h) Restored image by the proposed method.

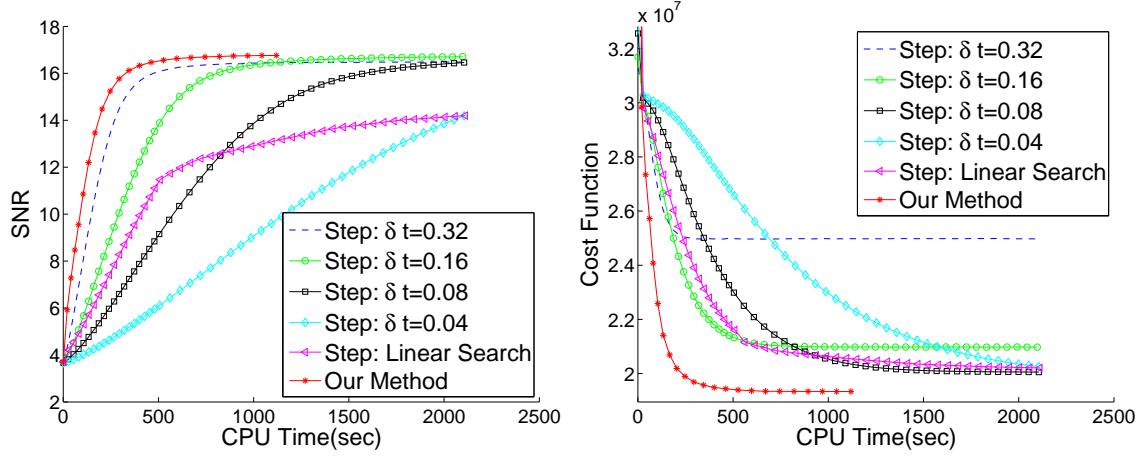


Fig. 6. Convergence profiles for the Barbara image. (a) SNR versus CPU time. (b) The objective value versus CPU time.

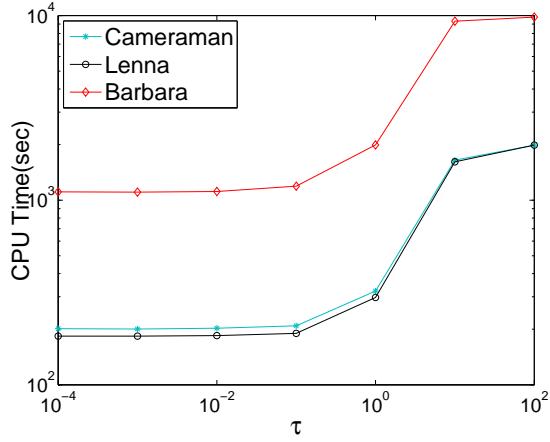


Fig. 7. CPU time versus the parameter τ .

Summing the above inequality from $i = 1$ to p , we obtain

$$\begin{aligned} & \mathcal{J}_2(\zeta^{(0)}, \beta^{(0)}) - \mathcal{J}_2(\zeta^{(p)}, \beta^{(p)}) \\ & \geq \sum_{i=1}^p \mathcal{J}_2(\zeta^{(i-1)}, \beta^{(i-1)}) - \mathcal{J}_2(\zeta^{(i)}, \beta^{(i)}) \\ & \geq \frac{(\tau + 1)^2}{2\tau} \sum_{i=1}^p \|\beta^{(i-1)} - \beta^{(i)}\|_2^2, \end{aligned}$$

for all p . Let $p \rightarrow \infty$, we know

$$\sum_{i=1}^p \|\beta^{(i-1)} - \beta^{(i)}\|_2^2 < \infty,$$

so that $\{\beta^{(i)}\}$ is a convergent sequence. ■

Next, we aim to show that the limit $\beta^* = \lim_{i \rightarrow \infty} \beta^{(i)}$ is a fixed point of $\mathcal{S}_2 \circ \mathcal{S}_1$. To do this, we first show that $\mathcal{S}_2 \circ \mathcal{S}_1$ is non-expansive and hence continuous.

Definition 1: [18] An operator $\mathbf{P} : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$ is called *non-expansive* if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{n^2}$, we have

$$\|\mathbf{P}(\mathbf{x}_1) - \mathbf{P}(\mathbf{x}_2)\|_2 \leq \|\mathbf{x}_1 - \mathbf{x}_2\|_2.$$

Lemma 1: [19, Lemma 2.4] Let φ be convex and semi-continuous and $\alpha > 0$. Let \mathcal{T} be an operator defined by

$$\mathcal{T}(\mathbf{y}) := \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \alpha \varphi(\mathbf{x}) \right\}. \quad (20)$$

Then \mathcal{T} is non-expansive.

Theorem 2: The operators \mathcal{S}_1 , \mathcal{S}_2 , $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ are non-expansive and continuous.

Proof: We first prove that both \mathcal{S}_1 and \mathcal{S}_2 are non-expansive and continuous. Then the non-expansiveness of the composition $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ follows easily.

For the operator \mathcal{S}_1 , we have

$$\begin{aligned} \mathcal{S}_1(\beta^{(i-1)}) &= \arg \min_{\zeta} \mathcal{J}_2(\zeta, \beta^{(i-1)}) \\ &= \arg \min_{\zeta} \left\{ \|\chi(\zeta - \xi)\|_2^2 + \tau \|\zeta - \beta^{(i-1)}\|_2^2 \right\}. \end{aligned}$$

Since $\|\zeta - \beta^{(i-1)}\|_2^2$ is convex and semi-continuous, by Lemma 1, \mathcal{S}_1 is non-expansiveness. The continuity of \mathcal{S}_1 simply follows from

$$\|\mathcal{S}_1(\mathbf{y}) - \mathcal{S}_1(\mathbf{x})\|_2 \leq \|\mathbf{y} - \mathbf{x}\|_2 \rightarrow 0$$

as $\mathbf{x} \rightarrow \mathbf{y}$.

For the operator \mathcal{S}_2 , by noting that $\phi(\mathbf{f}) = \|\mathbf{f}\|_{TV}$ is convex and continuous, and

$$\beta^{(i)} = \mathbf{W} \cdot \arg \min_{\mathbf{f}} \left\{ \|\mathbf{u}^{(i)} - \mathbf{f}\|_2^2 + \frac{\lambda}{1+\tau} \|\mathbf{f}\|_{TV} \right\},$$

we have that the operator \mathcal{S}_3 defined by

$$\mathcal{S}_3(\mathbf{u}^{(i)}) = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{u}^{(i)} - \mathbf{f}\|_2^2 + \frac{\lambda}{1+\tau} \|\mathbf{f}\|_{TV} \right\}$$

is non-expansive. Notice that $\mathbf{u}^{(i)} = \mathbf{W}^{-1} \zeta^{(i)}$, thus $\mathcal{S}_2 = \mathbf{W} \circ \mathcal{S}_3 \circ \mathbf{W}^{-1}$ is also non-expansive. It follows that the compositions $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ are non-expansive. The continuity of \mathcal{S}_2 , $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ follows directly from the definition of a non-expansive operator. \blacksquare

Next, we aim to show the existence of a fixed point of the operators $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$.

Theorem 3: Let $\beta^* = \lim_{i \rightarrow \infty} \beta^{(i)}$ and $\zeta^* = \lim_{i \rightarrow \infty} \zeta^{(i)}$, where $\beta^{(i)} = \mathcal{S}_2 \circ \mathcal{S}_1(\beta^{(i-1)})$ and $\zeta^{(i)} = \mathcal{S}_1 \circ \mathcal{S}_2(\zeta^{(i-1)})$. Then β^* and ζ^* are fixed points of $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ respectively. Moreover, we have $\beta^* = \mathcal{S}_2(\zeta^*)$ and $\zeta^* = \mathcal{S}_1(\beta^*)$.

Proof: Since $\beta^{(i)} = \mathcal{S}_2 \circ \mathcal{S}_1(\beta^{(i-1)}) = \mathcal{S}(\beta^{(i-1)})$, by continuity of \mathcal{S} , we have

$$\beta^* = \lim_{i \rightarrow \infty} \beta^{(i)} = \lim_{i \rightarrow \infty} \mathcal{S}(\beta^{(i-1)}) = \mathcal{S} \left(\lim_{i \rightarrow \infty} \beta^{(i-1)} \right) = \mathcal{S}(\beta^*).$$

Similarly, we have $\zeta^* = \mathcal{S}_1 \circ \mathcal{S}_2(\zeta^*)$, $\beta^* = \mathcal{S}_2(\zeta^*)$ and $\zeta^* = \mathcal{S}_1(\beta^*)$. ■

Notice that $\mathcal{J}_2(\zeta, \beta)$ is convex and coercive (i.e., $\mathcal{J}_2(\zeta, \beta) \rightarrow \infty$ as $\|\zeta\|_2^2 + \|\beta\|_2^2 \rightarrow \infty$), and therefore, a minimum of $\mathcal{J}_2(\zeta, \beta)$ exists [5, Proposition 2.1.1]. We are now in the position to prove the following theorem. Our proof is based on Proposition 2 in [3] but it uses a different bivariate functional.

Theorem 4: Let ζ^* and β^* be fixed points of $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ respectively, then (ζ^*, β^*) minimizes $\mathcal{J}_2(\zeta, \beta)$.

Proof: As ζ^* and β^* are fixed points of $\mathcal{S}_1 \circ \mathcal{S}_2$ and $\mathcal{S}_2 \circ \mathcal{S}_1$ respectively, and $\beta^* = \mathcal{S}_2(\zeta^*)$ and $\zeta^* = \mathcal{S}_1(\beta^*)$, we deduce that

$$\begin{cases} \zeta^* = \operatorname{argmin}_{\zeta} \mathcal{J}_2(\zeta, \beta^*), \\ \beta^* = \operatorname{argmin}_{\beta} \mathcal{J}_2(\zeta^*, \beta). \end{cases}$$

It implies that

$$\begin{cases} 0 = \frac{\partial}{\partial \zeta} \mathcal{J}_2(\zeta^*, \beta^*) = \frac{1+\tau}{\tau} \nabla_{\zeta} \mathcal{Q}(\zeta^*, \beta^*), \\ 0 \in \frac{\partial}{\partial \beta} \mathcal{J}_2(\zeta^*, \beta^*) = \frac{1+\tau}{\tau} \nabla_{\beta} \mathcal{Q}(\zeta^*, \beta^*) + \lambda \frac{\partial}{\partial \beta} \|\mathbf{f}(\beta^*)\|_{TV}, \end{cases} \quad (21)$$

where $\frac{\partial}{\partial \zeta} \mathcal{J}_2$ and $\frac{\partial}{\partial \beta} \mathcal{J}_2$ are the subgradients of the univariate functions $\mathcal{J}_2(\cdot, \beta)$ and $\mathcal{J}_2(\zeta, \cdot)$ respectively.

Let $\partial \mathcal{J}_2$ be the subgradient of the bivariate function $\mathcal{J}_2(\cdot, \cdot)$. In general, it is not necessarily true that $\partial \mathcal{J}_2(\zeta, \beta) = [\frac{\partial}{\partial \zeta} \mathcal{J}_2, \frac{\partial}{\partial \beta} \mathcal{J}_2]^T$. However, as we will show next, our bivariate function is constructed to possess such a nice property. Since \mathcal{J}_2 is the sum of two convex continuous functions, we have

$$\partial \mathcal{J}_2 = \frac{1+\tau}{\tau} \partial \mathcal{Q} + \lambda \partial \|\mathbf{f}\|_{TV}.$$

Notice that \mathcal{Q} is differentiable and $\|\mathbf{f}\|_{TV}$ does not depend on ζ , we deduce that

$$\partial \mathcal{J}_2 = \frac{1+\tau}{\tau} \begin{bmatrix} \nabla_{\zeta} \mathcal{Q} \\ \nabla_{\beta} \mathcal{Q} \end{bmatrix} + \lambda \begin{bmatrix} \{0\} \\ \frac{\partial}{\partial \beta} \|\mathbf{f}\|_{TV} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \zeta} \mathcal{J}_2 \\ \frac{\partial}{\partial \beta} \mathcal{J}_2 \end{bmatrix}.$$

Therefore, $0 \in \partial \mathcal{J}_2(\zeta^*, \beta^*)$ if and only if (21) holds. This implies that (ζ^*, β^*) is a minimizer of \mathcal{J}_2 . ■

REFERENCES

- [1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Trans. Image Process.*, 1(2):205–220, 1992.
- [2] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.*, 10(8):1200–1211, 2001.
- [3] J. Bect, L. Blanc Féraud, G. Aubert, and A. Chambolle. A l_1 -unified variational framework for image restoration. In *Proc. European Conference on Computer Vision*, Lecture Notes in Computer Sciences, vol. 3024, 2004.
- [4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In Kurt Akeley, editor, *SIGGRAPH 2000, Computer Graphics Proceedings*, pages 417–424. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
- [5] D. Bertsekas, A. Nedic, and E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, 2003.
- [6] J. Bioucas-Dias, M. Figueiredo, and R. Nowak. Total variation-based image deconvolution: a majorization-minimization approach. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006*, volume 2, 14–19 May 2006.
- [7] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Fast global minimization of the active contour/snake Model. *J. Math. Imaging Vision*, 28(2):151–167, 2007.
- [8] J. Carter. *Dual methods for total variation-based image restoration*. PhD thesis, Department of Mathematics, UCLA, 2002.
- [9] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004.

- [10] T. Chan, K. Chen, and X. Tai. Nonlinear multilevel schemes for solving total variation image minimization problem. In *Image Processing based on Partial Differential Equations*, Tai, Lie, Chan and Osher eds., Springer, Heidelberg, 265–285, 2006.
- [11] T. Chan, S. Kang, and J. Shen. Euler’s elastica and curvature-based inpainting. *SIAM J. Appl. Math.*, 63(2):564–592, 2002.
- [12] T. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.
- [13] T. Chan and J. Shen. Nontexture inpainting by curvature-driven diffusions. *J. Vis. Commun. Image Represent.*, 12(4):436–449, 2001.
- [14] T. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia, PA, 2005.
- [15] T. Chan, J. Shen, and H. Zhou. Total variation wavelet inpainting. *J. Math. Imaging Vision*, 25(1):107–125, 2006.
- [16] K. Chen and X. Tai. A nonlinear multigrid method for total variation minimization from image restoration. *J. Sci. Comput.*, 33(2):115–138, 2007.
- [17] A. Cohen, I. Daubechies, and J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.
- [18] P. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.
- [19] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005.
- [20] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part 1: Fast and exact optimization. *J. Math. Imaging and Vision*, 26(3):261–276, 2006.
- [21] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996, 1988.
- [22] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383, 1992.
- [23] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 4(7):932–946, 1995.
- [24] Y. Huang, M. Ng, and Y. Wen. A fast total variation minimization method for image restoration. *Multiscale Model. Simul.*, 7(2):774–795, 2008.
- [25] D. Hunter and K. Lange. A tutorial on MM algorithms. *Am. Stat.*, 58(1):30–37, 2004.
- [26] N. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. Harmon. Anal.*, 10(3):234–253, 2001.
- [27] D. Krishnan, P. Lin, and X. Tai. An efficient operator splitting method for noise removal in images. *Comm. Comput. Phys.*, 1:847–858, 2006.
- [28] M. Lysaker, S. Osher, and X. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Trans. Image Process.*, 13:1345–1357, 2004.
- [29] M. Ng, L. Qi, Y. Yang, and Y. Huang. On semismooth Newton’s methods for total variation minimization. *J. Math. Imaging and Vision*, 27:265–276, 2007.
- [30] M. Nikolova and M. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.*, 2005.
- [31] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Verlag, New York, 1999.
- [32] A. Ron and Z. Shen. Affine systems in $l^2(r^d)$: The analysis of the analysis operator. *J. Func. Anal.*, 148:408–447, 1997.
- [33] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physics D*, 60:259–268, 1992.
- [34] C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.