

# A Primal-Dual Method for Total Variation-Based Wavelet Domain Inpainting

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

## Abstract

Loss of information in a wavelet domain can occur during storage or transmission when the images are formatted and stored in terms of wavelet coefficients. This calls for image inpainting in wavelet domains. In this paper, a variational approach is used to formulate the reconstruction problem. We propose a simple but very efficient iterative scheme to calculate an optimal solution and prove its convergence. Numerical results are presented to show the performance of the proposed algorithm.

## Index Terms

Total variation, inpainting, wavelet, primal-dual.

## I. INTRODUCTION

Image inpainting is used to correct corrupted data in images. It is an important task in image restoration with a wide range of applications: texts and scratches removal [4], [5], [16], zooming [9], [32], impulse noise removal [8], [13], [29], high-resolution image reconstruction [12], [24], [30], [35], parallel magnetic resonance imaging [7] among many others. Inpainting can be done in the image domain or in a transformed domain, depending on where the corrupted data live. For example, in the image compression standard JPEG 2000, images are formatted and stored in terms of wavelet coefficients. If some of the coefficients are lost or corrupted during storage or transmission, the degraded image may contain some artifacts such as blurred edges and black holes. This prompts the need of restoring corrupted information in wavelet domains [3], [27], [36], [39].

Inpainting in the image domain, such as zooming, is similar to interpolation, where the pixel values to be filled are approximated by using the pixel values of the available pixels. Methods for inpainting in the image domain usually exploit the property that neighboring pixel values are correlated, so that the corrupted pixel values can be estimated from neighboring information. Inpainting in a transformed domain is a different matter, because the corrupted transformed coefficients can affect the whole image and it is not easy to localize the affected regions in the image domain.

In digital image processing, an image is represented by a matrix or by a vector formed by stacking up the columns of the matrix. In the latter representation of an  $n \times n$  image, the  $(r, s)$ -th pixel becomes the  $((r-1)n+s)$ -th entry of the vector.

The acquired image  $\mathbf{g}$  is modeled as the original image  $\mathbf{f}$  plus an additive Gaussian noise  $\boldsymbol{\eta}$ :

$$\mathbf{g} = \mathbf{f} + \boldsymbol{\eta}.$$

Wen is with Faculty of Science, Kunming University of Science and Technology, Yunnan, China. E-mail: wenyouwei@gmail.com. Research supported in part by NSFC Gran, NSF of Guangdong Grant No. 9251064201000009.

Chan is with Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong. The research was supported in part by HKRGC Grant CUHK400510 and DAG Grant 2060408. E-mail: rchan@math.cuhk.edu.hk.

Yip is with Department of Mathematics, National University of Singapore, Block S17, 10, Lower Kent Ridge Road, 119076, Singapore. E-mail: andyyip@nus.edu.sg. Research supported in part by Academic Research Grant R146-000-116-112 from NUS, Singapore.

In the following, we denote by  $f_{r,s}$  the  $((r-1)n+s)$ -th entry of  $\mathbf{f}$ . In the JPEG 2000 format, images are transformed from the pixel domain to a wavelet domain through a wavelet transform. Wavelet coefficients of an image under an orthogonal wavelet transform  $\mathbf{W}$  are designated by a hat  $\widehat{\cdot}$ . Thus, the wavelet coefficients of the noisy image  $\mathbf{g}$  are given by  $\widehat{\mathbf{g}} = \mathbf{W}\mathbf{g}$ . The observed wavelet coefficients  $\widehat{\mathbf{u}}$  are given by  $\widehat{\mathbf{g}}$  but with the missing ones set to zero:

$$\widehat{\mathbf{u}}_i = \begin{cases} \widehat{\mathbf{g}}_i, & i \in \mathbf{I}, \\ 0, & i \in \Omega \setminus \mathbf{I}. \end{cases}$$

Here  $\Omega$  is the complete index set in the wavelet domain and  $\mathbf{I} \subset \Omega$  is the set of available wavelet coefficients. Then an observed image  $\mathbf{u}$  is obtained by the inverse wavelet transform  $\mathbf{u} = \mathbf{W}^T \widehat{\mathbf{u}}$ . The aim of wavelet domain inpainting is to reconstruct the original image  $\mathbf{f}$  from the given coefficients  $\widehat{\mathbf{u}}$ .

Durand and Froment considered to reconstruct the wavelet coefficients using total variation (TV) norm [22]. Rane *et al.* in [36] considered wavelet domain inpainting in wireless networks. The method classifies the corrupted blocks of coefficients into texture blocks and structure blocks, which are then reconstructed by using texture synthesis and wavelet domain inpainting respectively. To fill in the corrupted coefficients in the wavelet domain in such a way that sharp edges in the image domain are preserved, T. Chan, Shen and Zhou proposed in [17] to minimize the objective function

$$\mathcal{J}(\mathbf{f}) = \frac{1}{2} \sum_{i \in \mathbf{I}} (\widehat{\mathbf{f}}_i - \widehat{\mathbf{u}}_i)^2 + \lambda \|\mathbf{f}\|_{TV}, \quad (1)$$

where

$$\|\mathbf{f}\|_{TV} \equiv \sum_{(r,s) \in \Omega} \left( |(D_x \mathbf{f})_{r,s}|^2 + |(D_y \mathbf{f})_{r,s}|^2 \right)^{1/2}$$

is TV norm,  $D_z$  is the forward difference operator in the  $z$ -direction for  $z \in \{x, y\}$  and  $\lambda$  is a regularization parameter. This is a hybrid method that combines coefficients fitting in the wavelet domain and regularization in the image domain. In [23], it was shown that such a hybrid method can reduce the stair-case effect of the TV norm. Various alternative objective functions with distinctive properties were proposed by Yau, Tai and Ng [45], Zhang, Peng and Wu [46] and Zhang and Chan [47].

A number of numerical methods have been proposed for solving TV minimization problems in the primal setting. They include time marching scheme [41], fixed point iteration method [43], majorization-minimization approach [34] and many others. Solving TV regularization problems using these methods faces a numerical difficulty due to the non-differentiability of the TV norm. This difficulty can be overcome by introducing a small positive parameter  $\epsilon$  in the TV norm that makes the TV norm differentiable, i.e.,

$$\|\mathbf{f}\|_{TV,\epsilon} \equiv \sum_{(r,s) \in \Omega} \left( |(D_x \mathbf{f})_{r,s}|^2 + |(D_y \mathbf{f})_{r,s}|^2 + \epsilon^2 \right)^{1/2}.$$

The corresponding objective function then becomes

$$\mathcal{J}_\epsilon(\mathbf{f}) = \frac{1}{2} \sum_{i \in \mathbf{I}} (\widehat{\mathbf{f}}_i - \widehat{\mathbf{u}}_i)^2 + \lambda \|\mathbf{f}\|_{TV,\epsilon}. \quad (2)$$

In [17], Chan, Shen and Zhou derived the Euler-Lagrange equation and used an explicit gradient descent scheme to solve (2) in the primal setting. In [28], numerical results showed that as  $\epsilon$  decreases, the reconstructed edges become sharper. Since the exact TV norm (i.e.  $\epsilon = 0$ ) can often produce better images than its smoothed version, one would prefer the original formulation (1) to the smoothed one (2). In [14], we proposed an efficient optimization transfer algorithm to solve the original problem (1). The method introduces an auxiliary variable  $\zeta$  yielding a new objective function:

$$\mathcal{J}_2(\zeta, \mathbf{f}) = \frac{1 + \tau}{2\tau} \left( \sum_{i \in \mathbf{I}} (\zeta_i - \widehat{\mathbf{u}}_i)^2 + \tau \|\zeta - \widehat{\mathbf{f}}\|_2^2 \right) + \lambda \|\mathbf{f}\|_{TV}, \quad (3)$$

where  $\tau$  is a positive control parameter. We showed in [14] that  $\mathcal{J}(\mathbf{f}) = \min_{\zeta} \mathcal{J}_2(\zeta, \mathbf{f})$  for any positive  $\tau$ . An alternating minimization algorithm is used to find a solution of the bivariate functional. For the auxiliary variable  $\zeta$ , the minimizer has a closed-form solution; for the original variable  $\mathbf{f}$ , the minimization problem is a classical TV denoising problem:

$$\mathcal{S}_{TV}(\mathbf{W}^T \zeta) \equiv \arg \min_{\mathbf{f}} \left[ \frac{1}{2} \left\| \mathbf{W}^T \zeta - \mathbf{f} \right\|_2^2 + \gamma \left\| \mathbf{f} \right\|_{TV} \right], \quad (4)$$

where  $\gamma = \frac{\lambda}{1+\tau}$  and  $\mathcal{S}_{TV}(\cdot)$  is known as the Moreau proximal map [21], which can be computed efficiently using dual-based approaches [10], [11], [33], primal-dual hybrid methods [50], [51] and splitting schemes [26].

In Chambolle's approach [11], the TV norm is reformulated using duality:

$$\left\| \mathbf{f} \right\|_{TV} = \max_{\mathbf{p} \in \mathbf{A}} \mathbf{f}^T \operatorname{div} \mathbf{p}, \quad (5)$$

where  $\mathbf{A} \subset \mathbb{R}^{2n^2}$  is defined by

$$\begin{aligned} \mathbf{A} = \left\{ \mathbf{p} = ((\mathbf{p}^x)^T, (\mathbf{p}^y)^T)^T \in \mathbb{R}^{2n^2} : \mathbf{p}^x, \mathbf{p}^y \in \mathbb{R}^{n^2}, \right. \\ \left. (\mathbf{p}_{r,s}^y)^2 + (\mathbf{p}_{r,s}^x)^2 \leq 1 \text{ for all } r, s \right\}. \end{aligned}$$

The discrete divergence of  $\mathbf{p}$  is a vector 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  $r, s = 1, \dots, n$ . Chambolle showed that solving (4) requires the orthogonal projection of the observed image onto the convex constraints appearing in (5). Hence computing the solution of (4) hinges on computing a nonlinear projection.

In [14], we applied Chambolle's projection algorithm to solve the problem in (4). Since each iteration of algorithm in [14] requires solving a TV denoising problem, the speed of the method in [14] largely depends on how fast we can solve TV denoising problems. In practice, we found that it is not desirable to solve each TV denoising problem to full convergence; a few iterations of Chambolle's algorithm are sufficient. This approach is quite effective. However, we will show that the computational cost can be further reduced by avoiding inner iterations.

In this paper, we apply a primal-dual approach to solve wavelet inpainting problems. By using the dual formulation of the TV norm, the objective function  $\mathcal{J}(\mathbf{f})$  in (1) can be written as

$$\mathcal{J}(\mathbf{f}) = \max_{\mathbf{p} \in \mathbf{A}} \mathcal{Q}(\mathbf{f}, \mathbf{p}),$$

where

$$\mathcal{Q}(\mathbf{f}, \mathbf{p}) = \frac{1}{2} \sum_{i \in I} \left( \widehat{\mathbf{f}}_i - \widehat{\mathbf{u}}_i \right)^2 + \lambda \mathbf{f}^T \operatorname{div} \mathbf{p}. \quad (6)$$

Therefore, the minimization of (1) becomes the minimax problem

$$\min_{\mathbf{f}} \max_{\mathbf{p} \in \mathbf{A}} \mathcal{Q}(\mathbf{f}, \mathbf{p}). \quad (7)$$

We notice that the function in (6) is convex (in fact quadratic) in  $\mathbf{f}$  and concave (in fact linear) in  $\mathbf{p}$ . By using Proposition 2.6.1 in [6], an optimal solution  $(\mathbf{f}^*, \mathbf{p}^*)$  of (1) can be obtained through a saddle point  $(\mathbf{f}^*, \mathbf{p}^*)$  of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ . Since the data fitting term does not contain the complete index set, the minimax problem of (7) cannot be reformulated as a pure dual problem. Thus, Chambolle's dual approach cannot be directly applied to solve the minimax problem of (6). In this paper, we derive a simple iterative scheme to calculate a saddle point of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ .

We highlight the major differences between our method and the previous primal-dual methods to solve TV problems. First of all, none of the existing primal-dual methods [15], [50], [51] address the wavelet inpainting problem. In terms of numerical algorithms, the method in [15] requires solving a large linear system in each of its Newton's iteration and uses an approximated TV norm to ensure the Newton's iteration is well-defined. Our method does not require solving any linear system and uses the exact (discrete) TV norm. The method in [50], [51] employs a gradient descent method to the primal and dual variables alternatively. While our method resembles in some way the alternating direction optimization method [25], it uses an optimization transfer method to the primal variable and a predictor-corrector scheme to the dual variable. Moreover, in [50], [51], the conditions on the step size of the primal and dual variables under which the algorithm converges have not been given. But here we provide a convergence analysis of our method together with a restriction to the step size.

The paper is organized as follows. In Section II, the optimality conditions of (6) are studied and the main iterative scheme is given. We also give the convergence analysis. Numerical results are given in Section III. Finally, some concluding remarks are given in Section IV.

## II. PRIMAL-DUAL BASED ITERATIVE ALGORITHM

Numerous methods have been proposed to solve saddle point problems. They include finite envelope method, primal-dual steepest descent algorithm, alternating direction method and so on, see [18], [25], [38], [42], [44], [48], [49]. In this section, we develop an iterative algorithm to calculate a saddle point of (7) and show the convergence of the proposed algorithm.

Before deriving the optimality conditions of (7), we introduce some definitions. The discrete gradient operator  $\nabla$  is defined such that  $(\nabla \mathbf{f})_{r,s} = ((D_x \mathbf{f})_{r,s}, (D_y \mathbf{f})_{r,s})^T$ . The projection of  $\mathbf{x}$  onto the set  $A$  is defined as

$$\mathcal{P}_A(\mathbf{x}) = \operatorname{argmin}\{\|\mathbf{z} - \mathbf{x}\|_2 \mid \mathbf{z} \in A\}.$$

for each  $\mathbf{x} \in \mathbb{R}^{2n^2}$ . For  $\mathbf{x} = (x_{1,1}^x, \dots, x_{n,n}^x, x_{1,1}^y, \dots, x_{n,n}^y)^T \in \mathbb{R}^{2n^2}$ , let  $\alpha_{r,s} = \max \left\{ \sqrt{(x_{r,s}^x)^2 + (x_{r,s}^y)^2}, 1 \right\}$ . Then, the projection is given by

$$\mathcal{P}_A(\mathbf{x}) = \left( \frac{x_{1,1}^x}{\alpha_{1,1}}, \dots, \frac{x_{n,n}^x}{\alpha_{n,n}}, \frac{x_{1,1}^y}{\alpha_{1,1}}, \dots, \frac{x_{n,n}^y}{\alpha_{n,n}} \right)^T.$$

For convenience, we assume that two variables  $\mathbf{f}$  and  $\widehat{\mathbf{f}}$  can be transformed into each other by  $\widehat{\mathbf{f}} = \mathbf{W}\mathbf{f}$  and  $\mathbf{f} = \mathbf{W}^T\widehat{\mathbf{f}}$ , when there is no ambiguity. Let  $\mathbf{M}$  be a diagonal matrix with  $M_{ii} = 1$  if  $i \in \mathbf{I}$  and  $M_{ii} = 0$  if  $i \in \Omega \setminus \mathbf{I}$ . The objective function in (6) can be re-written as:

$$\mathcal{Q}(\mathbf{f}, \mathbf{p}) = \frac{1}{2} \left\| \mathbf{M}\widehat{\mathbf{f}} - \widehat{\mathbf{u}} \right\|_2^2 + \lambda \mathbf{f}^T \operatorname{div} \mathbf{p}. \quad (8)$$

By using Proposition 2.6.1 in [6], a pair  $(\mathbf{f}^*, \mathbf{p}^*)$  is a saddle point of (8) if and only if  $\mathbf{f}^*$  and  $\mathbf{p}^*$  are optimal solutions of the problems

$$\begin{aligned} \mathbf{f}^* &= \operatorname{argmin}_{\mathbf{f}} \mathcal{Q}(\mathbf{f}, \mathbf{p}^*) \\ \mathbf{p}^* &= \operatorname{argmin}_{\mathbf{p} \in \mathbf{A}} \mathcal{Q}(\mathbf{f}^*, \mathbf{p}), \end{aligned}$$

respectively. Notice that  $\mathbf{f}^*$  is an optimal solution of the problem  $\min_{\mathbf{f}} \mathcal{Q}(\mathbf{f}, \mathbf{p}^*)$  if and only if

$$\mathbf{M}\widehat{\mathbf{f}}^* - \widehat{\mathbf{u}} + \lambda \mathbf{W} \operatorname{div} \mathbf{p}^* = 0. \quad (9)$$

Here we use the fact that  $\widehat{\mathbf{u}} = \mathbf{M}\widehat{\mathbf{u}}$ . Since the matrix  $\mathbf{M}$  is rank deficient, it is impossible to represent the primal variable  $\mathbf{f}^*$  using the dual variable  $\mathbf{p}^*$ . The equation (9) can be modified as

$$\widehat{\mathbf{f}}^* = \mathbf{K} \left( \tau \widehat{\mathbf{f}}^* + \widehat{\mathbf{u}} - \lambda \mathbf{W} \operatorname{div} \mathbf{p}^* \right). \quad (10)$$

where  $\tau > 0$  is a parameter and  $\mathbf{K} = (\mathbf{M} + \tau \mathbf{I})^{-1}$ . We remark that  $\mathbf{K}$  is a diagonal matrix. By definition, for any  $\mathbf{p} \in A$ , we have

$$\mathcal{Q}(\mathbf{f}^*, \mathbf{p}^*) \geq \mathcal{Q}(\mathbf{f}^*, \mathbf{p}).$$

Substituting the expression of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$  in (8) into the above inequality, we obtain

$$(\mathbf{f}^*)^T (\operatorname{div} \mathbf{p}^* - \operatorname{div} \mathbf{p}) \geq 0.$$

Using the fact that  $\operatorname{div} = -\nabla^T$ , we obtain

$$(\mathbf{p} - \mathbf{p}^*)^T \nabla \mathbf{f}^* \geq 0, \quad (11)$$

which is the same as

$$(\mathbf{p} - \mathbf{p}^*)^T (\mathbf{p}^* - (\mathbf{p}^* - s\lambda \nabla \mathbf{f}^*)) \geq 0,$$

for any  $s > 0$ . Recall that  $\mathbf{v}^* \in A$  is the projection of a point  $\mathbf{v}$  onto a convex set  $A$  if and only if

$$(\mathbf{p} - \mathbf{v}^*)^T (\mathbf{v}^* - \mathbf{v}) \geq 0 \quad (12)$$

for any  $\mathbf{p} \in A$  [2], [6]. It follows that

$$\mathbf{p}^* = \mathcal{P}_A(\mathbf{p}^* - s\lambda \nabla \mathbf{f}^*) \quad (13)$$

for any  $s > 0$  if and only if (11) holds.

The equations (10) and (13) suggest that the computation of a saddle point of (6) can be done by performing an iterative scheme. Starting from some initial pair  $(\mathbf{f}^{(0)}, \mathbf{p}^{(0)})$ , the iteration scheme has the form

$$\mathbf{p}^{(k+\frac{1}{2})} = \mathcal{P}_A(\mathbf{p}^{(k)} - s\lambda \nabla \mathbf{f}^{(k)}), \quad (14)$$

$$\mathbf{f}^{(k+1)} = \mathbf{W}^T \mathbf{K} \left( \tau \widehat{\mathbf{f}}^{(k)} + \widehat{\mathbf{u}} - \lambda \mathbf{W} \operatorname{div} \mathbf{p}^{(k+\frac{1}{2})} \right), \quad (15)$$

$$\mathbf{p}^{(k+1)} = \mathcal{P}_A(\mathbf{p}^{(k)} - s\lambda \nabla \mathbf{f}^{(k+1)}), \quad (16)$$

The main computational complexity of the proposed iterative scheme comes from forward and backward wavelet transformation, the gradient operation and the divergence operator. Hence, the cost of each iterative step is  $O(N)$ , where  $N$  is the number of pixels. In order to guarantee the convergence of the algorithm, the dual variable is computed twice at each iteration: the predictor step  $\mathbf{p}^{(k+\frac{1}{2})}$  and the corrector step  $\mathbf{p}^{(k+1)}$ . Performing two steps to the dual variables allows for computing a lower bound of the difference between the primal variable at two consecutive iterations. We state the following property and give the proof in the Appendix.

*Lemma 1:* Let  $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)})$  be the sequence generated by (14)–(16). Then

$$(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)})^T \operatorname{div} (\mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})}) \leq 8s\lambda \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} \right\|_2^2. \quad (17)$$

For the dual variable, the predictor-corrector scheme used can be understood as a primal point method [18], [37], that is,  $\mathbf{p}^{(k+\frac{1}{2})}$  and  $\mathbf{p}^{(k+1)}$  are the minimizers of  $\mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p})$  and  $\mathcal{Q}_p(\mathbf{f}^{(k+1)}, \mathbf{p})$  respectively, where

$$\mathcal{Q}_p(\mathbf{f}, \mathbf{p}) = -\mathcal{Q}(\mathbf{f}, \mathbf{p}) + \frac{1}{2s} \|\mathbf{p}^{(k)} - \mathbf{p}\|_2^2. \quad (18)$$

We state this property below and give the proof in the Appendix.



Fig. 1. *Boat* and *Goldhill* images.

*Lemma 2:* Let  $\mathcal{Q}_p(\mathbf{f}, \mathbf{p})$  be defined in (18). Then, we have

$$\begin{aligned}\mathbf{p}^{(k+\frac{1}{2})} &= \arg \min_{\mathbf{p} \in \mathbf{A}} \mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p}), \\ \mathbf{p}^{(k+1)} &= \arg \min_{\mathbf{p} \in \mathbf{A}} \mathcal{Q}_p(\mathbf{f}^{(k+1)}, \mathbf{p}).\end{aligned}$$

The following theorem states that the proposed iterative scheme (14)–(16) converges to an optimal primal-dual solution, under an assumption on the step size of the primal and dual variable. The proof is given in the Appendix.

*Theorem 1:* The sequence  $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)})$  generated by (14)–(16) converges to a saddle point  $(\mathbf{f}^*, \mathbf{p}^*)$  of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$  provided that  $s < \frac{\tau}{16\lambda^2}$ .

Since  $\min_{\mathbf{f}} \mathcal{J}(\mathbf{f}) = \min_{\mathbf{f}} \max_{\mathbf{p}} \mathcal{Q}(\mathbf{f}, \mathbf{p})$ , we deduce that  $\mathbf{f}^*$  is a minimizer of  $\mathcal{J}(\mathbf{f})$ , i.e., the limit point of the primal variable is a solution of the original wavelet inpainting problem (1).

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 [19], redundant transforms [31] and tight frames [40]. In these cases,  $\mathbf{W}$  is not orthogonal, but still has full rank, so that  $\mathbf{W}^T \mathbf{W}$  is invertible. From  $\widehat{\mathbf{f}} = \mathbf{W} \mathbf{f}$ , we get  $\mathbf{f} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \widehat{\mathbf{f}} = \mathbf{W}^\dagger \widehat{\mathbf{f}}$ . Therefore, the iteration for the primal variable is modified to

$$\mathbf{f}^{(k+1)} = \mathbf{W}^\dagger \mathbf{K} \left( \tau \widehat{\mathbf{f}}^{(k)} + \widehat{\mathbf{u}} - \lambda \mathbf{W} \operatorname{div} \mathbf{p}^{(k+\frac{1}{2})} \right).$$

To guarantee convergence, the parameter  $s$  and  $\tau$  should satisfy  $s < \tau / (16\lambda^2 \|(\mathbf{W}^\dagger)^T \mathbf{W}^\dagger\|_2^2)$ .

### III. NUMERICAL RESULTS

In this section, we report some experimental results to illustrate the performance of the proposed approach. The experiments were performed under Windows Vista Premium and MATLAB v7.6 (R2008a) running on a desktop machine with an Intel Core 2 Duo CPU (E6750) at 2.67 GHz and 3 GB of memory.

The wavelet transform used is the Daubechies 7-9 bi-orthogonal wavelets with symmetric extensions at the boundaries [1], [19]. The initial guess of  $\mathbf{f}$  is set to the observed image in all the experiments reported below. As is usually done, the quality

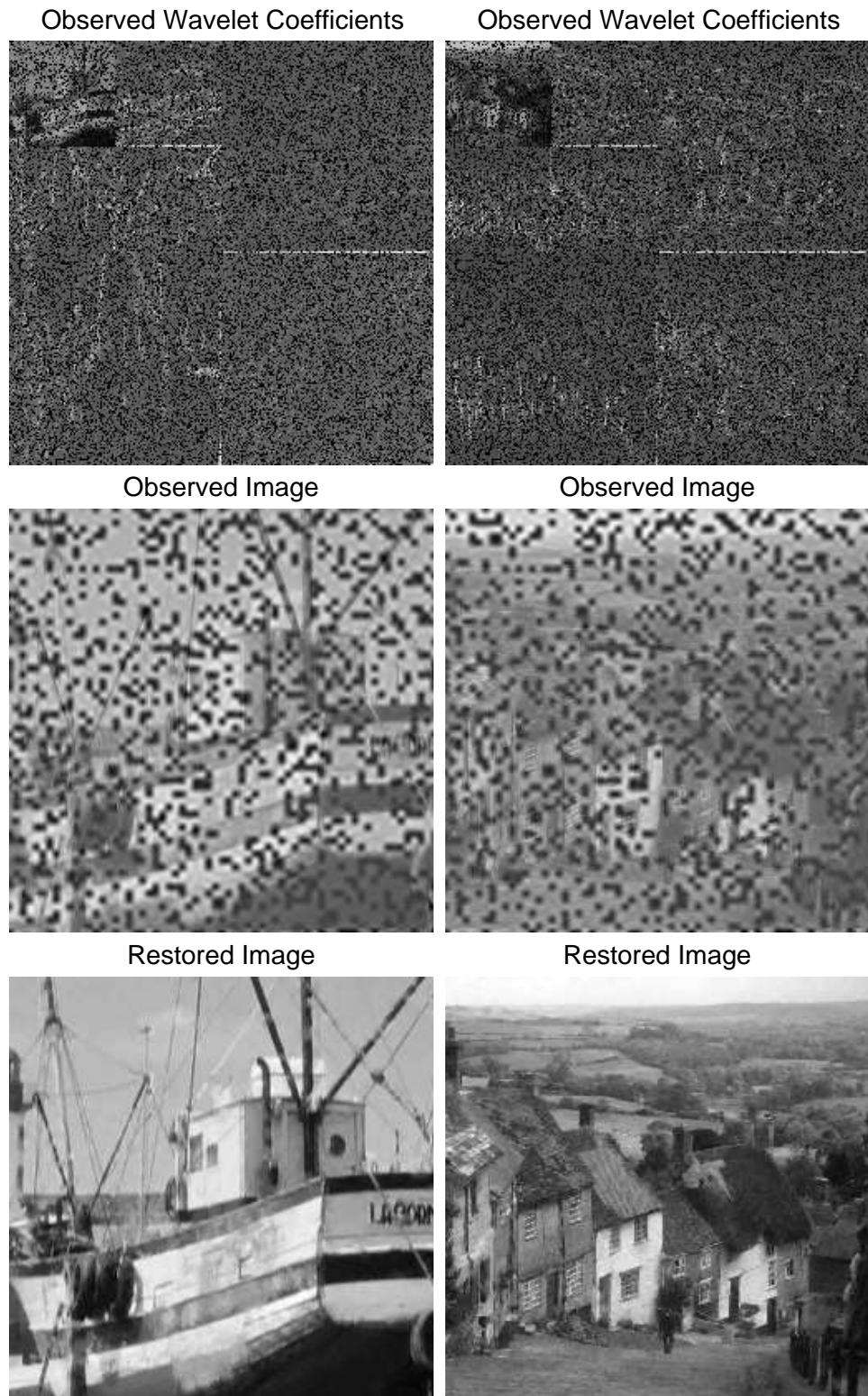


Fig. 2. Reconstruction results (25% lost). First row: the available wavelet coefficients. The original images were corrupted by an additive white Gaussian noise with standard deviation  $\sigma = 10$ . 25% of their wavelet coefficients were lost randomly. Second row: observed images in the spatial domain. Third row: restored images using the proposed algorithm.

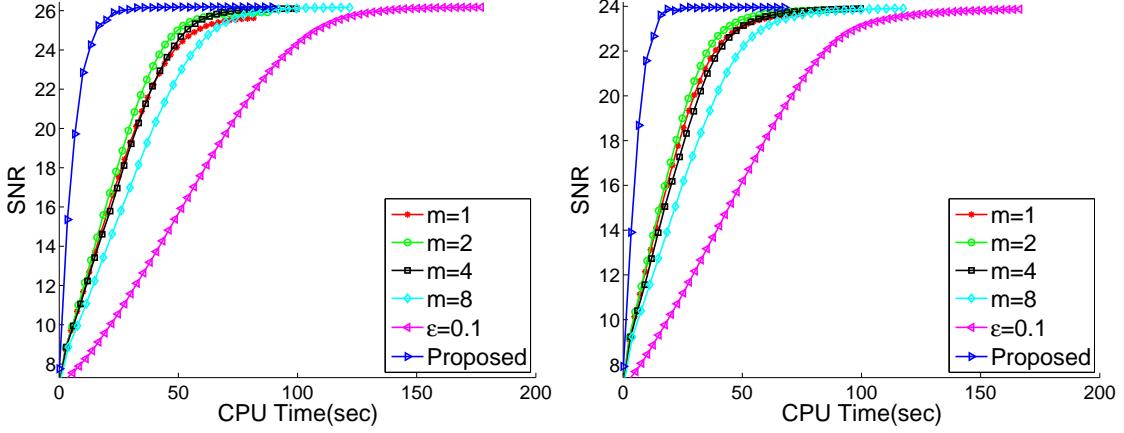


Fig. 3. Evolution of SNR against CPU time for the *Boat* and *Goldhill* images.

of a restored image is assessed using the Signal-to-Noise Ratio (SNR) defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{\|\mathbf{f}_{true}\|_2^2}{\|\mathbf{f}_{true} - \mathbf{f}\|_2^2} \right),$$

where  $\mathbf{f}_{true}$  and  $\mathbf{f}$  are the original image and the restored image respectively.

We consider the original clean  $256 \times 256$  *Boat* and *Goldhill* images shown in Figure 1. They are corrupted by an additive white Gaussian noise with standard deviation  $\sigma = 10$ . We compare the performance of the proposed method with the gradient descent algorithm [17] and the fast optimization transfer algorithm (OTA) [14]. The former algorithm simply applies the steepest descent iteration to the minimization problem (2). The latter uses an alternating minimization method to the bivariate functional (3). OTA requires solving TV denoising problems (4), which are done by Chambolle's projection algorithm [11]. In practice, only a few iterations of Chambolle's algorithm are run.

In the first experiment, we consider that the noisy images lost 25% of their wavelet coefficients randomly. The observed images in the wavelet domain and the spatial domain are shown in first row and second row of Figure 2, respectively. The restored images by using the proposed approach are shown in the third row of Figure 2.

We investigate how the SNR is improved by the proposed method and the OTA method (with various number of inner iterations) as time increases. The plots in Figure 3 show the evolution of SNR against CPU time for the *Boat* and *Goldhill* images. We set the TV smoothing parameter  $\varepsilon = 0.1$  and the number of inner iterations of the OTA to  $m = 1, 2, 4, 8$ . From the plots, we observe that the proposed method produces the best SNRs. We remark that the computational complexity per iteration of both the proposed method and the OTA is  $O(N)$ . However, the proposed method requires fewer iterations than the OTA to reach convergence.

In the second experiment, we consider that the noisy images lost 25% of their wavelet coefficients in  $8 \times 8$  blocks randomly. Figure 4 shows the observed images in the wavelet domain and in the spatial domain, and the restored images by the proposed method. The plots in Figure 5 show the evolution of SNR against CPU time. Once again the proposed method produces the best SNRs.

In order to reduce the size of an image, the JPEG 2000 standard can reduce the resolution automatically through its multi-resolution decomposition structure. The wavelet coefficients of all high-frequency sub-bands are discarded and only that of low-frequency sub-bands are kept. In this scheme, we obtain a low resolution image. The reconstruction of the original image from this low resolution image is reported in the third experiment.

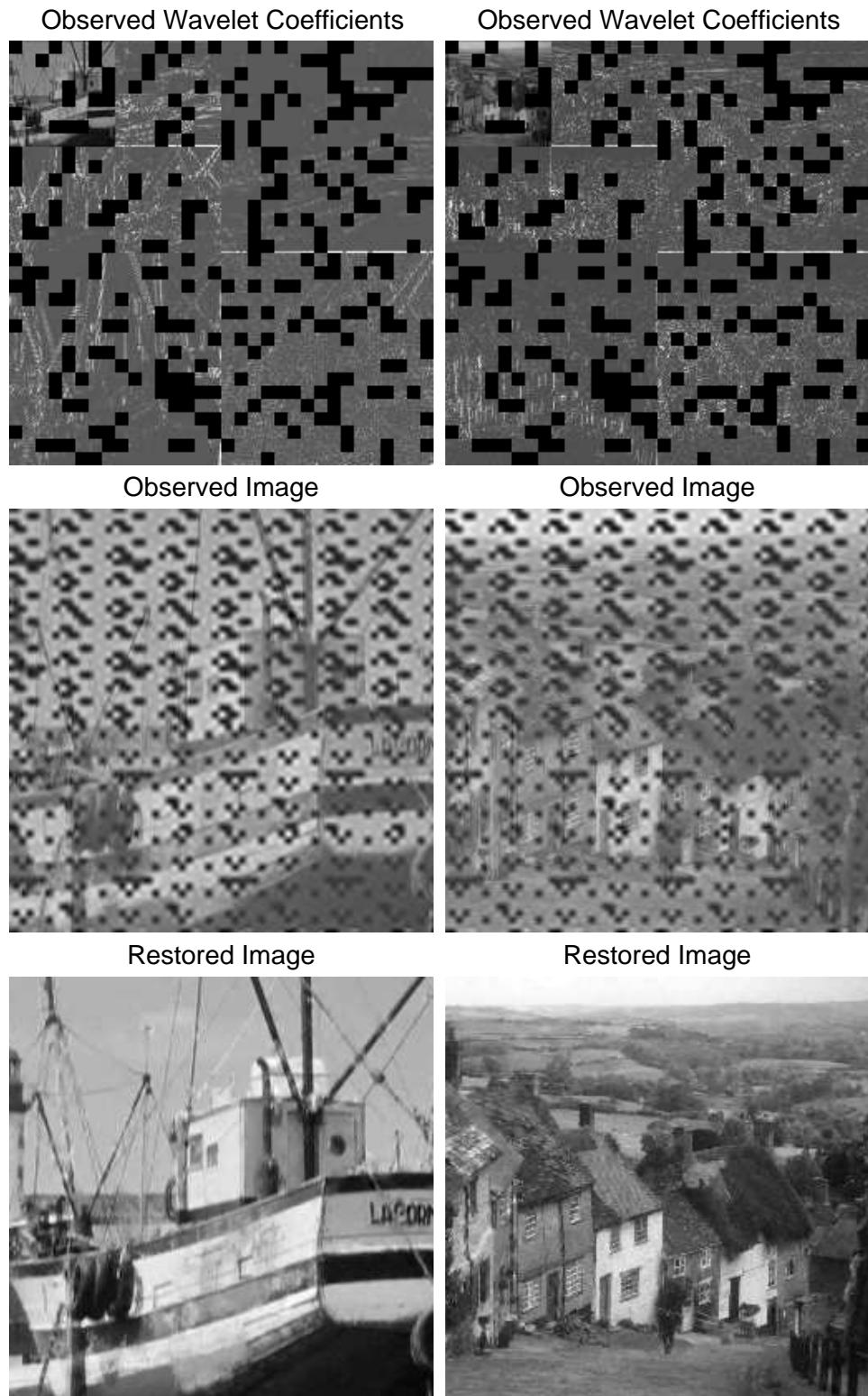


Fig. 4. Reconstruction results (25% lost). First row: the available wavelet coefficients. The original images were corrupted by an additive white Gaussian noise with standard deviation  $\sigma = 10$ . 25% of their wavelet coefficients in  $8 \times 8$  blocks were lost randomly. Second row: observed images in the spatial domain. Third row: restored images using the proposed algorithm.

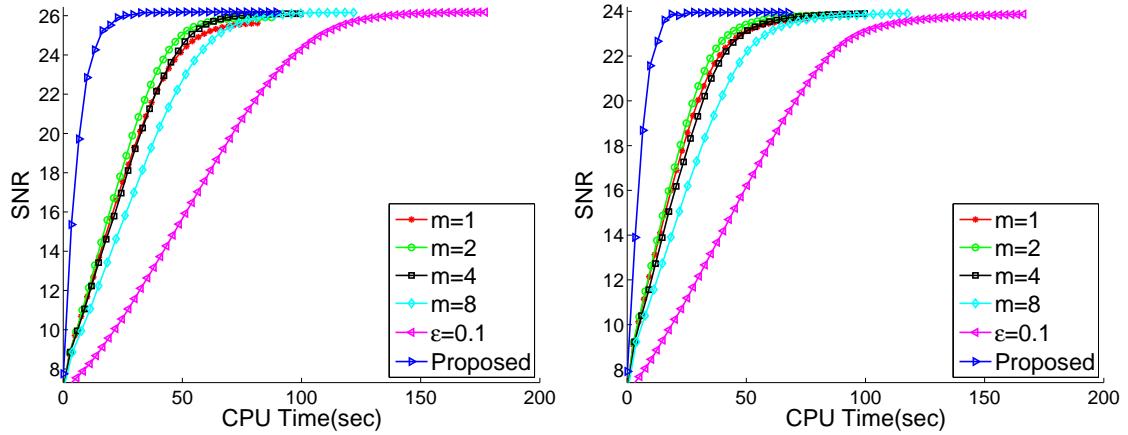


Fig. 5. Evolution of SNR against CPU time with *Boat* and *Goldhill* images.

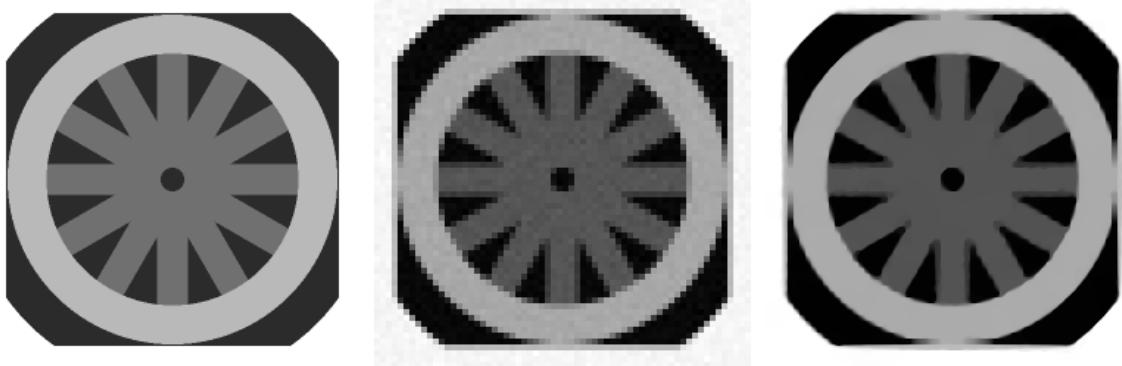


Fig. 6. Left: *Wheel* image. Middle: The observed image was corrupted by a white Gaussian noise with standard deviation  $\sigma = 15$  and all high-frequency sub-bands missing. Right: The reconstructed image.

In Figure 6, the original clean *Wheel* image of size  $256 \times 256$  is corrupted by a white Gaussian noise with standard deviation  $\sigma = 15$ . Then all its high-frequency sub-bands are discarded and only a low resolution image of size  $64 \times 64$  is kept. The reconstructed image is shown in the right of Figure 6. We observe that the sharp edges are reconstructed very well.

#### IV. CONCLUSION

We proposed an approach to total variation wavelet domain inpainting by exploiting primal-dual techniques. The original TV norm is represented by the dual formulation. The minimization problem (1) is transformed into a minimax problem, where a saddle point is sought. We show that each saddle point is a zero of the equations (10) and (13) and vice versa. However, equation (13) involves an expression of a non-differentiable projection operator, which prevents us from using higher order numerical methods that involve the use of derivatives. Instead, we show that a simple fixed point iteration without inner iterations is very efficient. We also prove that the sequence generated by the iterative scheme converges to a saddle point of the saddle problem. Numerical results show state-of-the-art performance and speed.

#### V. APPENDIX

##### A. Proof of Lemma 1

*Proof:* Let  $\mathbf{q}_1 = \mathbf{p}^{(k)} - s\lambda\nabla f^{(k+1)}$  and  $\mathbf{q}_2 = \mathbf{p}^{(k)} - s\lambda\nabla f^{(k)}$ . According to the definition of the operators  $\text{div}$  and  $\nabla$ , we have  $\text{div} = -\nabla^T$ . Thus we obtain

$$\begin{aligned}
& s\lambda(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)})^T \operatorname{div}(\mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})}) \\
&= -\left(s\lambda\nabla(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)})\right)^T (\mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2)) \\
&= (\mathbf{q}_1 - \mathbf{q}_2)^T (\mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2)).
\end{aligned}$$

Using the classical inequality  $2a^T b \leq \|a\|_2^2 + \|b\|_2^2$  for any vector  $a$  and  $b$ , we have

$$\begin{aligned}
& (\mathbf{q}_1 - \mathbf{q}_2)^T (\mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2)) \\
& \leq \frac{1}{2} (\|\mathbf{q}_1 - \mathbf{q}_2\|_2^2 + \|\mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2)\|_2^2) \\
& \leq \|\mathbf{q}_1 - \mathbf{q}_2\|_2^2 \\
& = \left\| s\lambda\nabla(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}) \right\|_2^2.
\end{aligned}$$

The last inequality uses the fact that the projection operator is nonexpansive [20]. By definition, for any  $\mathbf{f}$ , we have

$$\begin{aligned}
\|\nabla \mathbf{f}\|^2 &= \sum_{r,s} ((f_{r+1,s} - f_{r,s})^2 + (f_{r,s+1} - f_{r,s})^2) \\
&\leq 2 \sum_{r,s} (f_{r+1,s}^2 + 2f_{r,s}^2 + f_{r,s+1}^2) \\
&\leq 8\|\mathbf{f}\|^2.
\end{aligned}$$

Hence, the result holds. ■

### B. Proof of Lemma 2

*Proof:* By using Proposition 4.7.2 in [6],  $\mathbf{p}^{(k+\frac{1}{2})}$  minimizes  $\mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p})$  over the set  $\mathbf{A}$  if and only if there exists a  $\mathbf{d} \in \partial_p \mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p}^{(k+\frac{1}{2})})$  such that

$$(\mathbf{z} - \mathbf{p}^{(k+\frac{1}{2})})^T \cdot \mathbf{d} \geq 0,$$

for all  $\mathbf{z} \in \mathbf{A}$ . It is easy to check that

$$\begin{aligned}
\mathbf{d} &= \lambda\nabla\mathbf{f}^{(k)} + \frac{1}{s}(\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k)}) \\
&= \frac{1}{s} \left( \mathbf{p}^{(k+\frac{1}{2})} - (\mathbf{p}^{(k)} - s\lambda\nabla\mathbf{f}^{(k)}) \right) \\
&= \frac{1}{s} (\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{v}).
\end{aligned}$$

Here  $\mathbf{v} = \mathbf{p}^{(k)} - s\lambda\nabla\mathbf{f}^{(k)}$ .

From (14), we know  $\mathbf{p}^{(k+\frac{1}{2})} = \mathcal{P}_A(\mathbf{v})$ . Hence for any  $\mathbf{z} \in \mathbf{A}$ , we have

$$(\mathbf{z} - \mathbf{p}^{(k+\frac{1}{2})})^T \cdot \mathbf{d} = \frac{1}{s}(\mathbf{z} - \mathcal{P}_A(\mathbf{v}))^T (\mathcal{P}_A(\mathbf{v}) - \mathbf{v}) \geq 0,$$

where the inequality follows from the property (12) of the projection operator. Hence the result holds for  $\mathbf{p}^{(k+\frac{1}{2})}$ . Similarly, we can show that the result also holds for  $\mathbf{p}^{(k+1)}$ . ■

### C. Preliminary Lemmas

Before the proof of Theorem 1, we start with the following lemma, which gives an upper bound of the difference between the iterate  $\mathbf{f}^{(k+1)}$  and the optimal solution  $\mathbf{f}^*$ .

*Lemma 3:* Let  $(\mathbf{f}^*, \mathbf{p}^*)$  be a saddle point of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ , and  $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)})$  be the sequence generated by (14)–(16). Then, we have

$$\begin{aligned} & \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^* \right\|_2^2 \\ \leq & \tau \left\| \mathbf{f}^{(k)} - \mathbf{f}^* \right\|_2^2 - \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} \right\|_2^2 \\ & - 2\lambda (\mathbf{f}^{(k+1)})^T \operatorname{div}(\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^*). \end{aligned} \quad (19)$$

*Proof:* Notice that by (15), we have

$$\mathbf{M}\widehat{\mathbf{f}}^{(k+1)} - \widehat{\mathbf{u}} + \lambda \mathbf{W} \operatorname{div} \mathbf{p}^{(k+\frac{1}{2})} = \tau \left( \widehat{\mathbf{f}}^{(k)} - \widehat{\mathbf{f}}^{(k+1)} \right).$$

Hence

$$\begin{aligned} & \mathcal{Q}(\mathbf{f}^*, \mathbf{p}^{(k+\frac{1}{2})}) - \mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^{(k+\frac{1}{2})}) \\ = & \frac{1}{2} \left\| \mathbf{M}(\widehat{\mathbf{f}}^* - \widehat{\mathbf{f}}^{(k+1)}) \right\|_2^2 + \tau \left( \widehat{\mathbf{f}}^* - \widehat{\mathbf{f}}^{(k+1)} \right)^T \left( \widehat{\mathbf{f}}^{(k)} - \widehat{\mathbf{f}}^{(k+1)} \right) \\ \geq & \tau \left( \mathbf{f}^* - \mathbf{f}^{(k+1)} \right)^T \left( \mathbf{f}^{(k)} - \mathbf{f}^{(k+1)} \right). \end{aligned} \quad (20)$$

Since  $(\mathbf{f}^*, \mathbf{p}^*)$  is a saddle point of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ , we have

$$\mathcal{Q}(\mathbf{f}^*, \mathbf{p}^{(k+\frac{1}{2})}) \leq \mathcal{Q}(\mathbf{f}^*, \mathbf{p}^*) \leq \mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^*).$$

Thus we obtain

$$\mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^*) - \mathcal{Q}(\mathbf{f}^*, \mathbf{p}^{(k+\frac{1}{2})}) \geq 0. \quad (21)$$

Adding the inequalities (20) and (21), we have

$$\mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^*) - \mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^{(k+\frac{1}{2})}) \geq \tau \left( \mathbf{f}^* - \mathbf{f}^{(k+1)} \right)^T \left( \mathbf{f}^{(k)} - \mathbf{f}^{(k+1)} \right). \quad (22)$$

By the definition of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ , we have

$$\mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^*) - \mathcal{Q}(\mathbf{f}^{(k+1)}, \mathbf{p}^{(k+\frac{1}{2})}) = \lambda (\mathbf{f}^{(k+1)})^T \operatorname{div}(\mathbf{p}^* - \mathbf{p}^{(k+\frac{1}{2})}). \quad (23)$$

Also, we have trivially that

$$\left\| \mathbf{f} - \mathbf{f}^{(k)} \right\|_2^2 = \left\| \mathbf{f} - \mathbf{f}^{(k+1)} \right\|_2^2 + \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} \right\|_2^2 - 2 \left( \mathbf{f} - \mathbf{f}^{(k+1)} \right)^T \left( \mathbf{f}^{(k)} - \mathbf{f}^{(k+1)} \right). \quad (24)$$

Rearranging (22), (23) and (24), we obtain the result.  $\blacksquare$

### D. Proof of Theorem 1

*Proof:* Notice that  $\mathcal{Q}_p(\mathbf{f}, \mathbf{p})$  in (18) is a quadratic function and the Hessian matrix of  $\mathcal{Q}_p(\mathbf{f}, \mathbf{p})$  for the variable  $\mathbf{p}$  is  $\frac{1}{s} \mathbf{I}$ . Hence  $\mathcal{Q}_p(\mathbf{f}, \mathbf{p})$  is a strongly convex function with respect to the variable  $\mathbf{p}$ , i.e.,

$$\mathcal{Q}_p(\mathbf{f}, \mathbf{p}) \geq \mathcal{Q}_p(\mathbf{f}, \mathbf{q}) + \frac{1}{2s} \|\mathbf{p} - \mathbf{q}\|_2^2$$

for any  $\mathbf{p} \in \mathbf{A}$  and  $\mathbf{q} = \operatorname{argmin}_{\mathbf{p}} \mathcal{Q}_p(\mathbf{f}, \mathbf{p})$ . By Lemma 2, we have

$$\mathcal{Q}_p(\mathbf{f}^{(k+1)}, \mathbf{p}) \geq \mathcal{Q}_p(\mathbf{f}^{(k+1)}, \mathbf{p}^{(k+1)}) + \frac{1}{2s} \|\mathbf{p} - \mathbf{p}^{(k+1)}\|_2^2.$$

In particular, when  $\mathbf{p} = \mathbf{p}^*$ , we have

$$\begin{aligned} & \lambda \left( \mathbf{f}^{(k+1)} \right)^T \operatorname{div}(\mathbf{p}^{(k+1)} - \mathbf{p}^*) + \frac{1}{2s} \|\mathbf{p}^{(k)} - \mathbf{p}^*\|_2^2 \\ & \geq \frac{1}{2s} \|\mathbf{p}^{(k)} - \mathbf{p}^{(k+1)}\|_2^2 + \frac{1}{2s} \|\mathbf{p}^{(k+1)} - \mathbf{p}^*\|_2^2. \end{aligned} \quad (25)$$

Similarly, by Lemma 2 again, we have

$$\mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p}) \geq \mathcal{Q}_p(\mathbf{f}^{(k)}, \mathbf{p}^{(k+\frac{1}{2})}) + \frac{1}{2s} \|\mathbf{p} - \mathbf{p}^{(k+\frac{1}{2})}\|_2^2.$$

In particular, when  $\mathbf{p} = \mathbf{p}^{(k+1)}$ , we have

$$\begin{aligned} & \lambda \left( \mathbf{f}^{(k)} \right)^T \operatorname{div}(\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k+1)}) + \frac{1}{2s} \|\mathbf{p}^{(k)} - \mathbf{p}^{(k+1)}\|_2^2 \\ & \geq \frac{1}{2s} \|\mathbf{p}^{(k)} - \mathbf{p}^{(k+\frac{1}{2})}\|_2^2 + \frac{1}{2s} \|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})}\|_2^2. \end{aligned} \quad (26)$$

Multiplying the inequalities (25) and (26) by 2 on both sides, then adding them to the inequality (19), we have

$$\begin{aligned} & \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k+1)} - \mathbf{p}^* \right\|_2^2 \\ & \leq \tau \left\| \mathbf{f}^{(k)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^* \right\|_2^2 - \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^{(k+\frac{1}{2})} \right\|_2^2 \\ & \quad - \frac{1}{s} \left\| \mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})} \right\|_2^2 - \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} \right\|_2^2 \\ & \quad + 2\lambda (\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)})^T \operatorname{div}(\mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})}). \end{aligned}$$

According to Lemma 1, we have

$$\begin{aligned} & \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k+1)} - \mathbf{p}^* \right\|_2^2 \\ & \leq \tau \left\| \mathbf{f}^{(k)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^* \right\|_2^2 - \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^{(k+\frac{1}{2})} \right\|_2^2 \\ & \quad - \frac{1}{s} \left\| \mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})} \right\|_2^2 + (16s\lambda^2 - \tau) \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} \right\|_2^2. \end{aligned}$$

When  $s < \frac{\tau}{16\lambda^2}$ , we have

$$\begin{aligned} & \tau \left\| \mathbf{f}^{(k+1)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k+1)} - \mathbf{p}^* \right\|_2^2 \\ & \leq \tau \left\| \mathbf{f}^{(k)} - \mathbf{f}^* \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^* \right\|_2^2. \end{aligned} \quad (27)$$

Thus the sequence  $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)})$  is bounded and hence contains a subsequence  $(\mathbf{f}^{(k_i)}, \mathbf{p}^{(k_i)})$  converging to some limit point  $(\mathbf{f}^\infty, \mathbf{p}^\infty)$ .

It is easy to check that  $(\mathbf{f}^\infty, \mathbf{p}^\infty)$  satisfies the equations (10) and (13), hence  $(\mathbf{f}^\infty, \mathbf{p}^\infty)$  is a saddle point of  $\mathcal{Q}(\mathbf{f}, \mathbf{p})$ . Let  $(\mathbf{f}^*, \mathbf{p}^*) = (\mathbf{f}^\infty, \mathbf{p}^\infty)$ . Since the subsequence  $(\mathbf{f}^{(k_i)}, \mathbf{p}^{(k_i)})$  converges, for any  $\varepsilon > 0$ , there exists a constant  $M$  such that

$$\tau \left\| \mathbf{f}^{(k_i)} - \mathbf{f}^\infty \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k_i)} - \mathbf{p}^\infty \right\|_2^2 < \varepsilon,$$

for any  $i > M$ . According to (27), for any given  $\varepsilon > 0$ , there is

$$\tau \left\| \mathbf{f}^{(k)} - \mathbf{f}^\infty \right\|_2^2 + \frac{1}{s} \left\| \mathbf{p}^{(k)} - \mathbf{p}^\infty \right\|_2^2 < \varepsilon$$

for any  $k > k_M$ . Hence the result holds. ■

## REFERENCES

- [1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Trans. Image Process.*, 1(2):205–220, APR 1992.
- [2] M. Avriel. *Nonlinear programming: analysis and methods*. Dover Publications, 2003.
- [3] I. Bajic. Adaptive MAP error concealment for dispersively packetized wavelet-coded images. *IEEE Trans. Image Process.*, 15(5):1226–1235, 2006.
- [4] 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.
- [5] 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.
- [6] D. Bertsekas, A. Nedic, and E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, 2003.
- [7] M. Blaimer, F. Breuer, M. Mueller, R. Heidemann, M. Griswold, and P. Jakob. SMASH, SENSE, PILS, GRAPPA. How to choose the optimal method. *Top. Magnetic Reson. Imaging*, 15:223–236, 2004.
- [8] J. Cai, R. Chan, and M. Nikolova. Two-phase methods for deblurring images corrupted by impulse plus gaussian noise. *AIMS Journal on Inverse Problems and Imaging*, 2(2):187–204, 2008.
- [9] J. Cai, R. Chan, and Z. Shen. A framelet-based image inpainting algorithm. *Appl. Comput. Harmon. Anal.*, 24:131–149, 2008.
- [10] J. Carter. *Dual methods for total variation-based image restoration*. PhD thesis, Department of Mathematics, UCLA, 2002.
- [11] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004.
- [12] R. Chan, T. Chan, L. Shen, and Z. Shen. Wavelet algorithms for high-resolution image reconstruction. *SIAM J. Sci. Comput.*, 24:1408–1432, 2003.
- [13] R. Chan, C. Ho, and M. Nikolova. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *IEEE Trans. Image Process.*, 14(10):1479–1485, 2005.
- [14] R. Chan, Y. Wen, and A. Yip. A fast optimization transfer algorithm for image inpainting in wavelet domains. *IEEE Trans. Image Process.*, 18(7):1467–1476, 2009.
- [15] T. Chan, G. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.
- [16] T. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.
- [17] T. Chan, J. Shen, and H. Zhou. Total variation wavelet inpainting. *J. Math. Imaging Vision*, 25(1):107–125, 2006.
- [18] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Programming, Ser. A*, 64(1):81–101, 1994.
- [19] A. Cohen, I. Daubechies, and J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.
- [20] P. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.
- [21] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005.
- [22] S. Durand and J. Froment. Reconstruction of wavelet coefficients using total variation minimization. *SIAM J. Sci. Comput.*, 24:1754–1767, 2003.
- [23] S. Durand and M. Nikolova. Denoising of frame coefficients using  $l_1$  data-fidelity term and edge-preserving regularization. *SIAM Multiscale Model. Simul.*, 6(2):547–576, 2007.
- [24] M. Elad and A. Feuer. Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. *IEEE Trans. Image Process.*, 6(12):1646–1658, 1997.
- [25] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2(1):17–40, 1976.
- [26] T. Goldstein and S. Osher. The split Bregman method for L1 regularized problems. *SIAM J. Imaging Sci.*, 2:323–343, 2009.
- [27] S. Hemami and R. Gray. Subband-coded image reconstruction for lossy packet networks. *IEEE Trans. Image Process.*, 6(4):523–539, 1997.
- [28] Y. Huang, M. Ng, and Y. Wen. A fast total variation minimization method for image restoration. *Multiscale Model. Simul.*, 7(2):774–795, 2008.
- [29] Y. Huang, M. Ng, and Y. Wen. Fast image restoration methods for impulse and Gaussian noise removal. *IEEE Signal Process. Lett.*, 16:457–460, 2009.
- [30] S. Kim, N. Bose, and H. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframe. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 38(6):1013–1027, 1990.
- [31] N. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. Harmon. Anal.*, 10(3):234–253, 2001.
- [32] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM Journal on Numerical Analysis*, 39(1):1–37, 2002.
- [33] M. Ng, L. Qi, Y. Yang, and Y. Huang. On semismooth Newton's methods for total variation minimization. *J. Math. Imaging Vision*, 27:265–276, 2007.
- [34] J.P. Oliveira, J.M. Bioucas-Dias, and M.A.T. Figueiredo. Adaptive total variation image deblurring: A majorization–minimization approach. *Signal Processing*, 89(9):1683–1693, 2009.
- [35] S.C. Park, M.K. Park, and M.G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Mag.*, 20(3):21–36, 2003.

- [36] S. Rane, G. Sapiro, and M. Bertalmio. Structure and texture filling-in of missing image blocks in wireless transmission and compression applications. *IEEE Trans. Image Process.*, 12:296–303, 2003.
- [37] R. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.
- [38] R. Rockafellar. Computational schemes for large-scale problems in extended linear-quadratic programming. *Math. Programming, Ser. B*, 48(3):447–474, 1990.
- [39] J. Rombaut, A. Pizurica, and W. Philips. Optimization of packetization masks for image coding based on an objective cost function for desired packet spreading. *IEEE Trans. Image Process.*, 17(10):1849–1863, 2008.
- [40] 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.
- [41] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [42] P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM J. Optim.*, 7(4):951–965, 1997.
- [43] C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.
- [44] S. E. Wright. A general primal-dual envelope method for convex programming problems. *SIAM J. Optim.*, 10(2):405–414, 2000.
- [45] A. Yau, X.C. Tai, and M. Ng. L 0-Norm and Total Variation for Wavelet Inpainting. In *The 2nd international conference, SSVM 2009, Lecture Notes in Computer Science*, pages 539–551. Springer, 2009.
- [46] H. Zhang, Q. Peng, and Y. Wu. Wavelet inpainting based on  $p$ -laplace operator. *Acta Automatica Sinica*, 33(5):546–549, 2007.
- [47] X. Zhang and T. Chan. Wavelet inpainting by nonlocal total variation. *Inverse Problems and Imaging*, 4(1):191–210, FEB 2010.
- [48] C.Y. Zhu. Solving large-scale minimax problems with the primal-dual steepest descent algorithm. *Math. Programming*, 67(1, Ser. A):53–76, 1994.
- [49] C.Y. Zhu. On the primal-dual steepest descent algorithm for extended linear-quadratic programming. *SIAM J. Optim.*, 5(1):114–128, 1995.
- [50] M. Zhu. *Fast Numerical Algorithms for Total Variation Based Image Restoration*. PhD thesis, University of California, Los Angeles, 2008.
- [51] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report 08-34*, 2007.