Fast Total Variation Wavelet Inpainting via Approximated Primal-Dual Hybrid Gradient (PDHG)

Xiaojing Ye and Haomin Zhou *

Abstract

The primal-dual hybrid gradient (PDHG) algorithm has been successfully applied to a number of total variation (TV) based image reconstruction problems for fast numerical solutions. We show that PDHG can also effectively solve the computational problem of image inpainting in wavelet domain, where high quality images are to be recovered from incomplete wavelet coefficients due to lossy data transmission. In particular, as the original PDHG algorithm requires the orthogonality of encoding operators for optimal performance, we propose an approximated PDHG algorithm to tackle the non-orthogonality of Daubechies 7-9 wavelet which is widely used in practice. We show that this approximated version essentially alters the gradient descent direction in the original PDHG algorithm, but eliminates its orthogonality restriction and retains low computation complexity. Moreover, we prove that the sequences generated by the approximated PDHG algorithm always converge monotonically to an exact solution of the TV based image reconstruction problem starting from any initial guess. We demonstrate that the approximated PDHG algorithm also works on more general image reconstruction problems with total variation regularizations, and analyze the condition on the step sizes that guarantees the convergence.

Key words. total variation, wavelet, inpainting, PDHG.

1 Introduction

1.1 Total Variation Wavelet Inpainting

Image inpainting refers to a class of image processing tasks that recover high quality images from incomplete or corrupted data in the image domain or a transform domain. Suppose u is the image to be recovered, we follow the standard treatment to vectorize the image u into a vector in $\mathbb{R}^N$ by stacking its columns, where N is the number of pixels in u. Then the inpainting problem can be generally formulated as

$$f = SAu + n, \tag{1}$$

where $A \in \mathbb{R}^{N\times N}$ represents the transform under the domain of which data is acquired, $f \in \mathbb{R}^m$ is the acquired data, n is the observation noise in f, and $S \in \mathbb{R}^{m\times N}$ is a binary selection matrix that represents the subsampling pattern in the transformed domain of Au. Regions that are not selected by S are considered damaged or corrupted, and they need to be recovered.

When A is identity matrix, the damage happens to pixel values directly. This is the inpainting problem in the image domain, which was initially addressed by Bertalmio *et al.* in [3] and Mansou *et al.* [31]. The term image inpainting was first coined in [3]. Since then, it has received considerable attentions in the imaging community, see, e.g. [18, 17, 24, 7]. In [3], the authors proposed to smoothly propagate information from the surrounding areas into the inpainting domain. Later, Ballester *et*

*School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA. Tel: (404) 894-2712, Fax: (404) 894-4409. Email: {xiaojing.ye, hmzhou}@math.gatech.edu. This work is partially supported by NSF Faculty Early Career Development (CAREER) Award DMS-0645266 and DMS-1042998.

al. [1] developed a variational inpainting model based on a joint cost functional that incorporates the regularization on gradient vector field to interpolate miss pixel values in the image domain. In [18], Chan and Shen used total variation as regularization to recover image with missing pixel values. Chan *et al.* also introduced an inpainting technique using an Eulers elastica energy-based variational model for image inpainting in [16]. For better synthesis of textures in images, Efros and Leung suggested to incorporate information from pixels with similar neighborhood [23]. In [4], Bertalmio *et al.* proposed to separate the structure and texture components of an image and apply different techniques to reconstruct both. In recent years, frame-based regularization method, which uses the 1-norm for sparse frame coefficients, is also proposed for image inpainting [8, 9]. The methods mentioned above all focus on inpainting (1) in the image domain.

Image inpainting in the transform domain, however, is significantly different from the inpainting problem in the image domain mentioned above. In this case, the transform matrix A is no longer the identity matrix, and missing data in the transform domain can usually affect pixels in certain regions or even the entire image domain of u. Therefore, approaches based on diffusion or interpolation in the image domain are not appropriate since there are no well defined inpainting regions. In this paper, we consider a specific image inpainting problem in the wavelet domain, called wavelet inpainting [19]. Wavelet inpainting is an important imaging task in real world due to the increasing popularity of the JPEG2000 image compression standard. In JPEG2000 data format, images are stored and transferred in terms of wavelet coefficients. As data loss is inevitable during signal transmissions, it is important to recover images from incomplete wavelet coefficients. In other words, one needs to solve for a clean image $u \in \mathbb{R}^N$ from (1) given f, a subset of wavelet coefficients of u under the wavelet transform A.

The inpainting problem (1), either in image domain or transform domain, is underdetermined in general and hence presents infinitely many solutions. Therefore, one needs to employ appropriate regularization to obtain desirable images. Inspired by the great success of TV regularization in image reconstruction, Chan *et al.* proposed two models in [19] to recover image u, depending whether the collected wavelet coefficients f are contaminated by noise or not, respectively:

- Model I (Noisy data)
$$\min_u \left\{ \alpha TV(u) + \frac{1}{2}\|A_s u - f\|^2 \right\}. \tag{2}$$
- Model II (Noiseless data)
$$\min_u TV(u), \quad \text{subject to } A_s u = f. \tag{3}$$

In (2) and (3), we let $A_s := SA$ and $\|\cdot\| = \|\cdot\|_2$ for notation simplicity. The TV norm of u in the discrete setting can be represented by

$$TV(u) := \max_{p \in X} p^T D u = \sum_{i=1}^{N} \|D_i u\|, \tag{4}$$

where $D = (D_x^T, D_y^T)^T \in \mathbb{R}^{2N \times N}$ is the discrete gradient operator, and $D_x, D_y \in \mathbb{R}^{N \times N}$ are the discrete partial derivative operators along the x and y axes, respectively. The superscript ^{T} denotes the transpose (or conjugate) of a matrix (or an operator). The matrix $D_i \in \mathbb{R}^{2 \times N}$ constitutes the i-th rows of D_x and D_y. The variables u and p are then called the primal and dual variables, respectively. The admissible set X of the dual variable p is defined by

$$X := \{p = (p_1, \cdots, p_N, p_{N+1}, \cdots, p_{2N})^T \in \mathbb{R}^{2N} : \|(p_i, p_{N+i})^T\| \le 1, \forall i\}. \tag{5}$$

It has been shown that models (2) and (3) can efficiently recover images with well preserved edges from very limited wavelet coefficients. However, the computational difficulty hinders their applications since the objective functions in these two models are nonsmooth due to the TV norm and the constraint involved in model (3). Motivated by the computation challenge, the main goal of this paper is to design numerical algorithms that can efficiently find the solutions for the models.

1.2 Related Algorithms

There have been extensive researches conducted in recent years to solve minimization problems as in (2). We refer them to TV-L2 algorithms as the objective function in such minimization problems consists of a TV regularization term and a quadratic (L2) data fidelity term. In the pioneer work of TV based image restoration [34], Rudin *et al.* modified the TV norm by adding a small perturbation $\epsilon > 0$ such that $TV_\epsilon(u) := \sum_i (\|D_i u\|^2 + \epsilon)^{1/2}$, which becomes differentiable. Then the authors employed explicit gradient descent flow to the minimization problem with this modified TV norm to obtain an approximated solution. In [35], Vogel *et al.* proposed a fixed point iteration scheme that updates the image via a semi-implicit gradient descent scheme.

In recent years, variable splitting technique has become very popular. The key idea of variable splitting is introducing auxiliary variables to separate the computation difficulties, such that the subproblems of all variables can be effectively solved and the overall performance is optimized, see, e.g. [36, 38, 28, 37, 40, 21]. For instance, the split Bregman algorithm developed by Goldstein and Osher in [28] utilizes the form $TV(u) = \sum_i \|D_i u\|$ in (2), then introduces variables w_i to substitute $D_i u$ and convert the model (2) to a minimization problem of variables w_i and u subject to $w_i = D_i u$ for all i. The Bregman iterative regularization method [32] is adopted to tackle the constraints. The combined split Bregman algorithm involves only two-dimensional soft shrinkage to update w and a least squares problem to solve for u in each iteration. It was later shown to be equivalent to the alternating direction method (ADM) of multipliers applied to the augmented Lagrangian of the constrained minimization [5, 22, 26, 27].

However, solving the least squares subproblem of u in the split Bregman iteration algorithm requires the inverse of $D^T D + \xi A_s^T A_s$ for some constant $\xi > 0$. *Such computation has low complexity only if there are fast transforms available to diagonalize* $D^T D + \xi A_s^T A_s$. However, this condition is known to hold for only few types of image reconstruction problems such as denosing, deblurring and single-channel magnetic resonance image (MRI) reconstruction etc. In these cases, the matrix $D^T D + \xi A_s^T A_s$ can be diagonalized by fast Fourier transform (or discrete cosine transform) if the image u is assumed to have periodic (symmetric) boundary condition. For the wavelet inpainting problem (2), this condition is not satisfied. Later, variations of the split Bregman algorithms are developed to eliminate the restrictions and work on more general form of A_s [42, 25, 39, 21] where only computations of A_s and A_s^T are needed.

In recent years, algorithms based on the primal-dual formulation of TV norm in (4) has been extensively studied as they appear to outperform many previous methods in term of computation efficiency. The algorithms reformulate the problem as a minimax problem

$$\min_u \max_{p \in X} \left\{ \alpha p^T D u + \frac{1}{2} \|A_s u - f\|^2 \right\}, \tag{6}$$

and alternately solve for the primal variable u and dual variable p in each iteration. The first primal-dual algorithm for TV based image reconstruction was proposed by Chan *et al.* in [20] to solve Euler-Lagrange equations using Newton's method. This leads to a quadratic convergence rate and highly accurate solutions; however, the cost per iteration is high since the method explicitly uses second-order information and the inversion of a Hessian matrix is required. In [10], Chambolle used the dual formulation of the TV denoising problem, namely (2) with $A = S = I$, and provided an efficient semi-implicit gradient descent algorithm for the dual variable. However, the method does not naturally extend to more general cases as in (2) with other A_s. In [44], Zhu and Chan proposed a primal-dual hybrid gradient (PDHG) method. PDHG alternately updates the primal and dual variables u and p using gradient descent and gradient ascend schemes, respectively. The PDHG algorithm is shown to be more efficient than the split Bregman algorithm in many image reconstruction problems. Since [44], there have been a number of researchers exploited the PDHG algorithm and its convergence behavior. See, e.g. [25, 11, 29] and references therein. More importantly, PDHG only requires the inverse of $I + \xi A_s^T A_s$. We will show later that the PDHG algorithm naturally fits the wavelet inpainting problem (2) if the wavelet A is orthogonal, and it is computationally effective.

The study of TV-L2 algorithms has inspired many computation methods for TV wavelet inpainting problem (2). In [19], Chan *et al.* used the traditional gradient descend schemes to solve for the optimal solutions of (2) and (3). In [43], the authors proposed to solve the minimization problem (2) using operator splitting and the resulting algorithm involves a TV denoising solver in each iteration. Recently, Chan *et al.* developed a series of algorithms to tackle the TV wavelet inpainting problem [13, 15, 14]. In [13], a fast optimization transfer algorithm (OTA) is developed. OTA introduces an auxiliary variable v to substitute Au in (2), and form an unconstrained minimization problem by adding $\|v - Au\|^2$ to the cost function. Then the u and v are solved by alternating direction minimizations. Since OTA involves a TV denoising-type subproblem in each iteration, the authors employed Chambolle's method [10] as the subproblem solver. The authors also demonstrated that only a few inner iterations of the subproblem solver should be proceeded for overall efficiency.

The algorithm developed in [15] combines the split Bregman and the splitting fashion in [13] so that the subproblems in each iteration have closed form solutions and hence inner iterations are eliminated. The authors used the ADM algorithm to solve the resulting constrained minimization problem. In each iteration, the computational cost includes two wavelet transforms that migrate from image domain to wavelet domain and vice versa, and two fast Fourier transforms that solve the linear system involving the inverse of $I+\xi D^T D$. The algorithm in [14] is derived from the optimality condition of the minimax problem and can be viewed as a variation of the PDHG algorithm.

In this paper, we show that the original PDHG algorithm works immediately on the TV wavelet inpainting problem (2) if the wavelet transform A is orthogonal, and hence all existing results on the PDHG algorithm including step size selection can be applied. In this case, the PDHG algorithm only requires two wavelet transforms per-iteration. This is the same to the algorithm in [14]. On the other side, compared to the present work, the algorithm in [14] requires more computations due to an extra correction step of the dual variable p and a tighter restriction on the step sizes to ensure convergence.

Furthermore, for a general transform A which is not orthogonal, the PDHG can not be applied directly. We propose an approximated PDHG algorithm to solve the computational problem of the unconstrained minimization in (2) as well as the constrained one in (3) with A being the widely used biorthogonal wavelet transforms. In addition, we extend our results to the image reconstruction problems where A has arbitrary form, or even the data fidelity term in (2) is not quadratic.

The rest of this paper is organized as follows. In the next section, we show how to solve the TV wavelet inpainting problem using PDHG algorithm. Then we proposed an approximated PDHG algorithm to solve (2) and (3) with biorthogonal wavelet transform. Section 3 show the convergence of the approximated PDHG algorithm. In Section 4, we present the numerical results using several real image wavelet inpainting problems. We extend PDHG to more general image reconstruction problems with nonlocal TV and nonquadratic data fidelity term in Section 5. The last section concludes the paper.

2 Algorithms

2.1 The PDHG Algorithm

In this section, we first quote the original PDHG algorithm, and show that it works immediately on the wavelet inpainting problem (2) if A is orthogonal. The PDHG algorithm first employs the primal-dual formulation of TV norm (4) to (2), and solves for the saddle point of the following minimax problem

$$\min_{u} \max_{p \in X} \Phi(u,p) := \left\{ \alpha p^T Du + \frac{1}{2}\|A_s u - f\|^2 \right\}. \tag{7}$$

Here, Φ is the objective function that is convex with respect to the primal variable u and concave with respect to the dual variable p. The PDHG procedure iterates in the following scheme:

$$p^{k+1} = \arg\max_{p\in X}\left\{\Phi(u^k,p) - \frac{1}{2\tau_k}\|p-p^k\|^2\right\} \tag{8}$$

$$\tilde{p}^{k+1} = p^{k+1} + \theta(p^{k+1}-p^k) \tag{9}$$

$$u^{k+1} = \arg\min_{u}\left\{\Phi(u,\tilde{p}^{k+1}) + \frac{1}{2\sigma_k}\|u-u^k\|^2\right\} \tag{10}$$

starting from an initial guess (u^0,p^0). In other words, the PDHG algorithm alternately applies gradient ascend scheme to p and gradient descent scheme to u. In (8) and (10), τ_k and σ_k behave as the step sizes for p and u respectively. The combination parameter θ is set to 0 in the original PDHG paper [44], and then 1 in [12] for provable convergence. In view of the objective function $\Phi(u,p)$ in (7), we obtain closed form solutions for the subproblems in (8) and (10) as follows: the solution in (8) has closed form as

$$p^{k+1} = \Pi_X(p^k + \tau_k D u^k), \tag{11}$$

where Π_X is the projection to the admissible set X. Note that X is closed and convex set in $\mathbb{R}^{2N}$, the operation Π_X is given by

$$\left((\Pi_X(p))_i, (\Pi_X(p))_{N+i}\right)^T = \frac{(p_i, p_{N+i})^T}{\max\{1, \|\,(p_i,p_{N+i})^T\,\|\}}, \quad i=1,\cdots,N. \tag{12}$$

Namely, Π_X projects the 2-vector $(p_i,p_{N+i})^T$ onto the Euclidean unit ball $\{z\in\mathbb{R}^2 : \|z\|\le 1\}$ in $\mathbb{R}^2$ for $i=1,\cdots,N$. Therefore, the computation in (8) has very low complexity (linear in terms of N) and can be easily carried out in parallel.

The remaining problem is then to solve the problem (10) effectively. One can readily see that the objective function in (10) is least squares:

$$\alpha(\tilde{p}^{k+1})^T D u + \frac{1}{2}\|A_s u - f\|^2 + \frac{1}{2\sigma_k}\|u-u^k\|^2, \tag{13}$$

and the normal equation of u^{k+1} is

$$\left(I + \sigma_k A^T S^T S A\right) u^{k+1} = u^k - \alpha\sigma_k D^T \tilde{p}^{k+1} + \sigma_k A^T S^T f. \tag{14}$$

Hence the key is to invert the matrix $I+\sigma_k A^T S^T S A$ in an effective manner. If the encoding wavelet A used in the first place is orthogonal, namely $A^TA = I$, we have

$$I + \sigma_k A^T S^T S A = A^T(I + \sigma_k S^T S)A, \tag{15}$$

where the matrix $I+\sigma_k S^T S$ is diagonal since S is merely a binary selection matrix that strikes some rows of the identity matrix. Therefore the inverse of (15) can be easily obtained, and the solution to (10) is

$$u^{k+1} = A^T(I+\sigma_k S^T S)^{-1}\left(A(u^k - \alpha\sigma_k D^T\tilde{p}^{k+1}) + \sigma_k S^T f\right), \tag{16}$$

for which the main computation involves two (forward and backward) wavelet transforms A and A^T.

2.2 Approximated PDHG Algorithm

In this subsection, we consider solving the noiseless model (3) which has a linear constraint $A_s u = f$, with a biorthogonal wavelet transform A. Such considerations are necessary in practice since wavelet coefficients are either received or lost but not contaminated by noises during data transmission. Also, the widely used wavelet transform is Daubechies 7-9 which is biorthogonal, meaning that there exists

$B \neq A$ such that $B^T A = I$. It is obvious that the PDHG presented in Section 2.1 can not be applied directly because equations (15) and (16) are no longer true. Therefore, we need modifications to solve these two issues.

To tackle the constraint in (3), we first utilize the augmented Lagrangian method by introducing the multiplier λ. That is, we solve the saddle point problem

$$\max_{\lambda} \min_{u} \left\{ TV(u) - \lambda^T (A_s u - f) + \frac{\beta}{2} \|A_s u - f\|^2 \right\}. \tag{17}$$

To be consistent with the notations in parameters above, we let $1/\beta \to \alpha$ and $\lambda/\beta \to \lambda$. These changes of notations do not affect our solution to (3) and the final value of λ is not of interests. By completing the square and replacing $TV(u)$ by the primal dual formulation of TV as in (4), we obtain

$$\max_{\lambda} \min_{u} \max_{p \in X} \left\{ \Phi(u, p, \lambda) := \alpha p^T D u + \frac{1}{2} \|A_s u - f - \lambda\|^2 - \frac{1}{2} \|\lambda\|^2 \right\}. \tag{18}$$

In the setting of classical Lagrangian multiplier method, the inner minimax problem with respect to u and p needs to be solved thoroughly, then the obtained u is used to update the multiplier λ, followed by solving the inner minimax problem with respect to u and p again, etc. Obviously this requires extensive computations for inner iterations on (8) and (10), which is not optimal from the practical point of view. Alternatively, we propose to use the following scheme

$$p^{k+1} = \arg\max_{p \in X} \left\{ \Phi(u^k, p, \lambda^k) - \frac{1}{2\tau_k} \|p - p^k\|^2 \right\}, \tag{19}$$

$$\tilde{p}^{k+1} = p^{k+1} + \theta(p^{k+1} - p^k), \tag{20}$$

$$u^{k+1} = \arg\min_{u} \left\{ \Phi(u, \tilde{p}^{k+1}, \lambda^k) + \frac{1}{2\sigma_k} \|u - u^k\|_Q^2 \right\}, \tag{21}$$

$$\lambda^{k+1} = \lambda^k - (A_s u^{k+1} - f). \tag{22}$$

where $\|u\|_Q^2 := u^T Q u$ and Q is a symmetric positive definite matrix to be selected properly in different applications. This is as if only one inner iteration is applied to the minimax problem with respect to u and p in (18) for fixed λ^k, and then immediately updates λ^k. We will show in the next section that under certain conditions the scheme (19)–(22) converges to a solution of (3). Note that the minimax problem (18) reduces to (2) if λ is set to the constant 0 and not ever updated. Without loss of generality, we consider the case for noiseless model (3) with λ being updated.

The differences of the approximated PDHG algorithm (19)–(22) to the original one (8)–(10) are mainly in the update of λ in (22), and more importantly, the altered gradient descent direction $Q^{-1}\nabla_u \Phi$ in (21) instead of $\nabla_u \Phi$ in the original PDHG (10). As a consequence, the normal equation in (21) becomes

$$\left(Q + \sigma_k A_s^T A_s\right) u^{k+1} = Q u^k - \alpha \sigma_k D^T \tilde{p}^{k+1} + \sigma_k S^T A^T (f + \lambda^k). \tag{23}$$

We have seen from (15) that the original PDHG algorithm (10) requires the orthogonality of A to solve the u subproblem effectively and attain optimal performance overall. However, this condition does not hold in many practical applications since the most widely used Daubechies 7-9 wavelet in image encoding is biorthogonal. In this case, the wavelet transform matrix A is not orthogonal and hence $A^T A \neq I$. Instead, A has a dual basis matrix $B \neq A$ that $B^T A = I$. To overcome this difficulty, one can choose $Q = A^T A$. With the computations in (19) unchanged, we observe that the normal equation (21) of u becomes

$$(A^T A + \sigma_k A_s^T A_s) u = A^T (I + \sigma_k S^T S) A u = A^T A u^k - \alpha \sigma_k D^T \tilde{p}^{k+1} + \sigma_k A_s^T (f + \lambda^k). \tag{24}$$

Hence the solution can be obtained by

$$u^{k+1} = B^T (I + \sigma_k S^T S)^{-1} \left(A u^k - \alpha \sigma_k B D^T \tilde{p}^{k+1} + \sigma_k S^T (f + \lambda^k) \right), \tag{25}$$

where the computation complexity remains as two wavelet transforms B^T and B, since Au^k is a by-product of the previous iteration.

3 Convergence Analysis

Convergence of the original PDHG algorithm (8)–(10) has been explored by several groups [25, 11, 29]. In particular, He and Yuan in [29] rewrite the PDHG algorithm (8)–(10) in the form of a long-established proximal point algorithm (PPA) [30, 33] and hence convergence results easily follow as PPA is essentially in a contraction fashion. We use this idea to prove that the scheme in (19)–(22) with approximated term in Q-norm and multiplier λ is also convergent and the limit is a solution to the minimax problem (18).

Lemma 3.1. *Let D be the gradient operator and $Q, \tau_k, \sigma_k, \theta$ be defined as in (19)–(22), then*

$$\begin{pmatrix} \frac{1}{\tau_k} I & \alpha D \\ \alpha\theta D^T & \frac{1}{\sigma_k} Q \end{pmatrix} \tag{26}$$

is positive definite if

$$\tau_k \sigma_k < \frac{4 q_N}{(1+\theta)^2 \alpha^2 \|D^T D\|}, \tag{27}$$

where $q_N > 0$ is the minimal eigenvalue of Q.

Proof. For any $p \in \mathbb{R}^{2N}$ and $u \in \mathbb{R}^N$, we have

$$\begin{aligned} & (p^T, u^T) \begin{pmatrix} \frac{1}{\tau_k} I & \alpha D \\ \alpha\theta D^T & \frac{1}{\sigma_k} Q \end{pmatrix} \begin{pmatrix} p \\ u \end{pmatrix} \\ =& \frac{\|p\|^2}{\tau_k} + \frac{\|u\|_Q^2}{\sigma_k} + (1+\theta)\alpha p^T D u \\ \geq& \frac{\|p\|^2}{\tau_k} + \frac{q_N \|u\|^2}{\sigma_k} - (1+\theta)\alpha \sqrt{\|D^T D\|} \|p\| \|u\| \\ \geq& \left(2\sqrt{q_N/\tau_k \sigma_k} - (1+\theta)\alpha \sqrt{\|D^T D\|} \right) \|p\| \|u\|, \end{aligned} \tag{28}$$

where we used Cauchy-Schwartz inequality. Therefore, matrix defined in (26) is positive definite if (27) holds. □

Theorem 3.2. *Any sequence generated by the approximated PDHG algorithm (19)–(22) with $\theta = 1$ and step sizes satisfying (27) converges to a solution of the minimax problem (18).*

Proof. A solution (u^*, p^*, λ^*) to (18) is a saddle point that satisfies the optimality condition

$$\begin{aligned} -\alpha(p - p^*)^T D u^* \geq 0, \quad \forall p \in X \\ \alpha D^T p^* - A_s^T \lambda^* = 0, \\ -A_s u^* + f = 0. \end{aligned} \tag{29}$$

On the other hand, the sequence $\{(p^k, u^k, \lambda^k)\}$ generated by the approximated PDHG algorithm (19)–(22) solves the corresponding minimization problems in each iteration and hence satisfies

$$(p - p^{k+1})^T \left(-\alpha D u^k + \frac{1}{\tau_k}(p^{k+1} - p^k) \right) \geq 0, \quad \forall p \in X \tag{30}$$

$$\alpha D^T \tilde{p}^{k+1} + A_s^T (A_s u^{k+1} - f - \lambda^k) + \frac{1}{\sigma_k} Q(u^{k+1} - u^k) = 0, \tag{31}$$

$$\lambda^{k+1} - \lambda^k + A_s u^{k+1} - f = 0, \tag{32}$$

where $\tilde{p}^{k+1} = 2p^{k+1} - p^k$ as $\theta = 1$. We plug (32) into (31) and obtain

$$\alpha D^T \tilde{p}^{k+1} - A_s^T \lambda^{k+1} + \frac{1}{\sigma_k} Q(u^{k+1} - u^k) = 0. \tag{33}$$

For notation simplicity, we define

$$z = \begin{pmatrix} p \\ u \\ \lambda \end{pmatrix} \in X \times \mathbb{R}^N \times \mathbb{R}^m, \quad F(z) = \begin{pmatrix} -\alpha D u \\ \alpha D^T p - A_s^T \lambda \\ A_s u - f \end{pmatrix}, \quad M = \begin{pmatrix} \frac{1}{\tau_k} I & \alpha D & 0 \\ \alpha D^T & \frac{1}{\sigma_k} Q & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{34}$$

Then, the optimality condition (29) can be simply written as

$$(z - z^*)^T F(z^*) \geq 0, \quad \forall z \in X \times \mathbb{R}^N \times \mathbb{R}^m. \tag{35}$$

and (30), (33), and (32) can be written together as

$$(z - z^{k+1})^T \left(F(z^{k+1}) + M(z^{k+1} - z^k) \right) \geq 0, \quad \forall z \in X \times \mathbb{R}^N \times \mathbb{R}^m. \tag{36}$$

In addition, it can be easily verified that

$$(z - z^*)^T \left(F(z) - F(z^*) \right) = 0, \quad \forall z \in X \times \mathbb{R}^N \times \mathbb{R}^m. \tag{37}$$

Now we set $z \to z^*$ in (36) and $z \to z^{k+1}$ in (35) and (37), and then combine the three inequalities to obtain

$$(z^k - z^*)^T M \left(z^k - z^{k+1} \right) \geq (z^k - z^{k+1})^T M \left(z^k - z^{k+1} \right). \tag{38}$$

If the step sizes satisfy the condition in (27), M is symmetric positive definite and hence

$$\begin{aligned} \|z^{k+1} - z^*\|_M^2 &= \|z^k - z^*\|_M^2 - 2(z^k - z^*)^T M(z^k - z^{k+1}) + \|z^k - z^{k+1}\|_M^2 \\ &\leq \|z^k - z^*\|_M^2 - \|z^k - z^{k+1}\|_M^2. \end{aligned} \tag{39}$$

Hence, the norm $\|z^k - z^*\|_M$ is nonincreasing. Two immediate consequences are that z^k is bounded and $\|z^k - z^{k+1}\|_M^2 \to 0$. The first one implies that there exists a subsequence $\{z^{k_j}\}$ that converges to a limit point $\hat{z} < \infty$. The second one indicates that $\{z^{k_j+1}\}$ also converges to $\hat{z}$. Placing $\{z^{k_j}\}$ in (19)–(22), and let $j \to \infty$, we can see that $\hat{z}$ satisfy the optimality condition and hence is in the place of z^* in (35), (36) and (37). As a subsequence converge to $\hat{z}$ and $\|z^k - \hat{z}\|_M$ is non-increasing, the entire sequence $\{z^k\}$ also converges (in M-norm) to $\hat{z}$, which is a saddle point of (18) since it satisfies the optimality condition (35). □

Remark 3.3. In Theorem 3.2, we showed that the approximated PDHG algorithm (19)–(22) converges if $\theta = 1$. In the case that $\theta \in [-1, 1)$, M defined in (34) is still positive definite provided the step size condition (27), but not symmetric anymore. Nevertheless, the inequality (38) still holds, and one can choose any symmetric positive definite matrix W and obtain

$$\left(W(z^k - z^*) \right)^T \left(-W^{-1} M(z^k - z^{k+1}) \right) \leq -(z^k - z^{k+1})^T M \left(z^k - z^{k+1} \right) \leq -c\|z^k - z^{k+1}\|^2 < 0 \tag{40}$$

for some constant $c > 0$ due to the fact that M is positive definite (even not symmetric). This implies that $-W^{-1}M(z^k - z^{k+1})$ is a descent direction. Thus, one can correct the output z^{k+1} of the approximated PDHG algorithm by

$$\bar{z}^{k+1} = z^k - t_k W^{-1} M(z^k - z^{k+1}) \tag{41}$$

for some step size t_k, and use $\bar{z}^{k+1}$ as the input of the next iteration in (19)–(22). It can be readily show that

$$\begin{aligned} \|\bar{z}^{k+1} - z^*\|_W^2 &= \|z^k - z^*\|_W^2 - 2t_k(z^k - z^*)^T M(z^k - z^{k+1}) + t_k^2 \|W^{-1} M(z^k - z^{k+1})\|_W^2 \\ &\leq \|z^k - z^*\|_W^2 - 2ct_k\|z^k - z^{k+1}\|^2 + t_k^2 \|M^T W^{-1} M(z^k - z^{k+1})\|^2. \end{aligned} \tag{42}$$

Hence, there exists an interval of t_k that guarantees the decay $\|z^{k+1} - z^*\|_W^2 < \|z^k - z^*\|_W^2$. For instance, one can choose t_k as

$$t_k = c\|z^k - z^{k+1}\|^2 / \|M^T W^{-1} M(z^k - z^{k+1})\|^2. \tag{43}$$

More detailed discussion on t_k can be found in [29]. In summary, the corrected sequence $\{\bar{z}^k\}$ monotonically converges to z^* in W-norm. In practice, the matrix W can be set to $\mathrm{diag}(I/\tau_k, Q/\sigma_k, I)$ for ease of computation.

Remark 3.4. In the case of $\theta = 1$, there is no correction step needed for the approximated PDHG algorithm, and Theorem 3.2 indicates the convergence of scheme (19)–(22). Moreover, if A is orthogonal, the step size condition (27) yields $\tau_k\sigma_k < 1/8\alpha^2$ for the approximated PDHG algorithm due to the fact that $\|D^TD\| = 8$ and $\|A^TA\| = 1$. In contrast, the primal dual algorithm proposed in [14] requires an additional correction step of the dual variable p and more strict step size constraint $\tau_k\sigma_k < 1/16\alpha^2$. Moreover, the approximated PDHG algorithm proposed in this paper fits to more general TV based image reconstruction problems as shown later, and the theoretical results such as Theorem 3.2 can directly apply.

4 Numerical Results

In this section, we test the approximated PDHG (APDHG) algorithm (19)–(22) on a set of real images to see its efficiency in TV wavelet inpainting. All the algorithms are coded and tested in MATLAB computing environment on a Linux PC with AMD Athlon 5000+ Dual Core processor (only one core is used in computation) at 2.6GHz and 4GB of memory.

4.1 Comparison Algorithms

We compare the performance of APDHG with several recently developed TV image reconstruction algorithms on wavelet inpainting problem. As the matrix $A_s^TA_s$ cannot be diagonalized easily if A is not orthogonal, we select the algorithms that do not require the inverse of $I + \xi A_s^TA_s$. More specifically, we choose the Algorithm (A2) in [42] and the algorithm proposed in [39] for comparison. The algorithm in [42] utilizes the idea of Bregman operator splitting [41] and replace $\|A_su - f\|^2$ by $\|u - (u^k - \delta A_s^T(A_su^k - f))\|^2/\delta$ in the split Bregman algorithm [28]. We denote this algorithm by BOS. The algorithm in [39] which we refer to as SBB integrates the Barzilai-Borwein step size selection method [2] into the split Bregman scheme and achieves significant improvement over BOS. The schemes of BOS and SBB can be unified as follows,

$$\begin{aligned} w_i^{k+1} &= \max\{\|D_iu^k + b_i^k\| - 1/2\rho, 0\}(D_iu^k + b_i^k)/\|D_iu^k + b_i^k\| \quad i = 1, \cdots, N, \\ u^{k+1} &= \left(\rho D^TD + \lambda\delta_k I\right)^{-1}\left(\rho D^T(w^{k+1} - b^k) + \lambda\delta_k u^k - \lambda A_s^T(A_su^k - f)\right), \\ b_i^{k+1} &= b_i^k - (w_i^{k+1} - D_iu^{k+1}) \quad i = 1, \cdots, N, \end{aligned} \tag{44}$$

The main computation complexity in (44) is A_s and A_s^T, and two FFT to invert the matrix $\rho D^TD + \lambda\delta_k I$. In addition, BOS requires $\delta \le \|A_s^TA_s\| = \|S^TS\| = 1$, whereas SBB compute Barzilai-Borwein step sizes by

$$\delta_k = \|A_su^{k+1} - Au^k\|^2/\|u^{k+1} - u^k\|^2 \tag{45}$$

and hence $\delta_k \ge 1$. The SBB algorithm is shown to be much more efficient than the BOS algorithm as the former retains low computation complexity while significantly reduces the number of iterations to reach the same level of accuracy. For more details of these two comparison algorithms, we refer interested readers to [42, 39, 21]. It is worth pointing out that both BOS and SBB literally work for any operator A_s as long as A_s and A_s^T are computable.

4.2 Numerical Results

We test APDHG algorithm for models (2) and (3), denoted by APDHG (2) and APDHG (3), respectively. We also compare the performance with BOS and SBB on a set of natural images. The name and size of images are listed in Table 4.2. The intensities of images are scaled to $[0, 1]$. We used

Table 1: Test images

No.	Image	Size N
1	barbara	256×256
2	cameraman	256×256
3	boat	512×512
4	man	512×512

Daubechies 7-9 wavelet and discard 50% wavelet coefficients randomly for all test images. In numerical implementation, we simply use wavelet transform A to substitute the dual transform B, but still keep B^T whenever needed. The reconstruction results are shown in this section and also quantitatively evaluated by the peak signal to noise ratio (PSNR) defined by

$$\text{PSNR} := 20 \log_{10} \left(\frac{\sqrt{N} \|\bar{u}\|_\infty}{\|u - \bar{u}\|} \right), \tag{46}$$

where u and $\bar{u}$ represent the reconstructed and the original images, respectively. In addition, we plot the value of objective function (Obj) in (2) versus CPU time in seconds (CPU Time). Note that the Obj vs CPU Time is not meant to reflect the performance of the APDHG (3) algorithm for the constrained model (3) whose objective differs from (2).

The original cameraman image is shown in Figure 1(a). With 50% wavelet coefficients randomly missing, the corrupted image obtained by the inverse wavelet transform is shown in Figure 1(b), where the damages appear across the entire image domain. The reconstructed images by the BOS [42], SBB [39], and APDHG (2) and APDHG (3) are shown in Figures 1(c), 1(d), 1(e) and 1(f), respectively. Due to the over strict condition on step sizes, the BOS algorithm does not improve image quality much and hence there are still many corrupted regions in the reconstructed image, as shown in Figure 1(c). The SBB algorithm appears to be efficient as most of damaged regions are repaired, however, some stripe-shaped artifacts also present in the reconstructed image, as shown in Figure 1(d). The APDHG results in Figures 1(e) and 1(f) are more plausible as they recover the cameraman image well without obvious artifacts. In particular, the reconstructed image in Figure 1(f) looks more similar to the original image in Figure 1(a), especially that the natural shading in the sky is well preserved. On the contrary, the same region in Figure 1(e) obtained by APDHG (2) is completely smoothed out. Figures 2(a) and 2(b) show the evolution of objective function value (given in (2)) and the PSNR of the BOS, SBB, and APDHG algorithms. Due to the over strict limitation on step sizes, the BOS algorithm converges very slowly compared to the other algorithms. SBB algorithm converges very fast but exhibits significant oscillations during the computation process. The APDHG algorithm converges for both models (2) and (3).

The reconstruction of barbara image (Figure 3(a)) follows the same strategies as for the cameraman image. There are 50% wavelet coefficients randomly missing, resulting in a corrupted image shown in Figure 3(b). The reconstructed images by BOS, SBB, and APDHG for models (2) and (3) are shown in Figures 3(c), 3(d), 3(e), and 3(f), respectively. Again the SBB reconstruction has some artifacts appeared in the recovery, whereas the APDHG returns images much closer to the original one. Other tests in Figures 5 and 7 yield similar conclusions as the previous two tests.

5 Conclusion

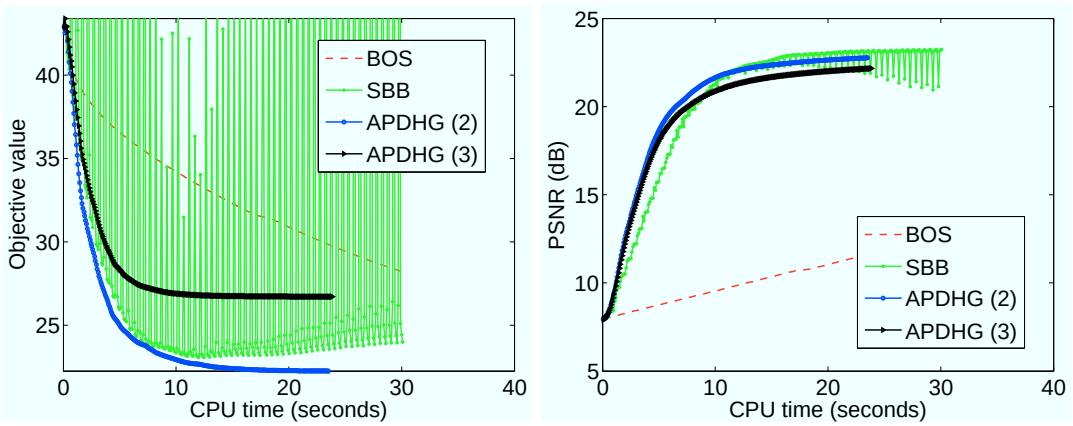
We propose the approximated PDHG algorithm (19)–(22) for TV based image reconstruction problems (2) and (3). We show that the algorithm works effectively for the TV wavelet inpainting problem. Moreover, the algorithm can be extended to work on more general TV based image reconstruction problems. Convergence of the proposed algorithm is also discussed.

(a) Original (b) Corrupted

(c) BOS (d) SBB

(e) APDHG (2) (f) APDHG (3)

Figure 1: Inpainting results of cameraman.



(a) Objective value (b) PSNR

Figure 2: Objective value and PSNR versus CPU time of cameraman image.

References

[1] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, *Filling-in by joint interpolation of vector fields and gray levels*, Image Processing, IEEE Transactions on, 10 (2001), pp. 1200 –1211.

[2] J. Barzilai and J. M. Borwein, *Two point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, in SIGGRAPH, 2000, p. 417424.

[4] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, *Simultaneous structure and texture image inpainting*, IEEE Trans. Image Process., 12 (2003), pp. 882–889.

[5] D. Bertsekas, *Parallel and Distributed Computation*, Prentice Hall, 1989.

[6] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530.

[7] M. Burger, L. He, and C.-B. Schönlieb, *Cahnhilliard inpainting and a generalization for grayvalue images*, SIAM J. Imag. Sci., 2 (2009), pp. 1129–1167.

[8] J.-F. Cai, R. Chan, and Z. Shen, *A framelet-based image inpainting algorithm*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 131–149.

[9] J.-F. Cai, H. Ji, F. Shang, and Z. Shen, *Inpainting for compressed images*, Appl. Comput. Harmon. Anal., 29 (2010), pp. 368–381.

[10] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20 (2004), pp. 89–97.

[11] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock, *An introduction to Total Variation for Image Analysis*, tech. rep., 2009.

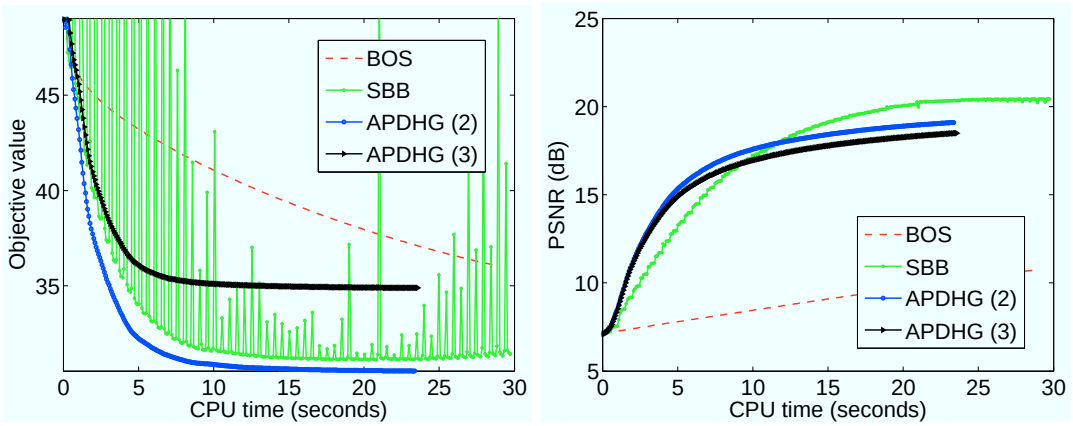
[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

(a) Original (b) Corrupted

(c) BOS (d) SBB

(e) APDHG (2) (f) APDHG (3)

Figure 3: Inpaiting results of barbara.



(a) Objective value (b) PSNR

Figure 4: Objective value and PSNR versus CPU time of barbara image.

[13] R. Chan, Y. Wen, and A. Yip, *A fast optimization transfer algorithm for image inpainting in wavelet domains*, IEEE Trans. Image Process., 18 (2009), pp. 1467–1476.

[14] R. Chan, Y. Wen, and A. Yip, *A primal-dual method for total variation-based wavelet domain inpainting*, IEEE Trans. Image Process., 21 (2012), pp. 106–114.

[15] R. H. Chan, J. Yang, and X. Yuan, *Alternating direction method for image inpainting in wavelet domains*, SIAM J. Imag. Sci., 4 (2012), pp. 807–826.

[16] T. Chan, S. Kang, and J. Shen, *Eulers elastica and curvature-based inpainting*, SIAM J. Appl. Math., 63 (2002), pp. 564–592.

[17] T. Chan, S. Osher, and J. Shen, *The digital tv filter and nonlinear denoising*, IEEE Trans. Image Process., 10 (2001), pp. 231–241.

[18] T. Chan and J. Shen, *Mathematical models for local nontexture inpaintings*, SIAM J. Appl. Math., 62 (2002), pp. 1019–1043.

[19] T. Chan, J. Shen, and H. Zhou, *Total variation wavelet inpainting*, J. Math. Imaging Vis., 25 (2006), pp. 107–125.

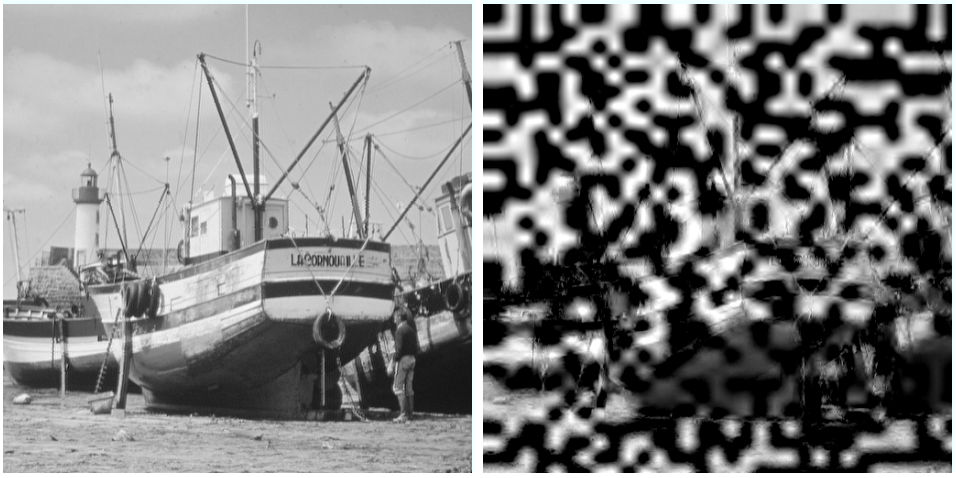
[20] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variationbased image restoration*, SIAM J. Optim., 20 (1999), pp. 1964–1977.

[21] Y. Chen, W. W. Hager, F. Huang, D. T. Phan, X. Ye, and W. Yin, *Fast algorithms for image reconstruction with application to partially parallel MR imaging*, SIAM J. Imag. Sci., 5 (2012), pp. 90–118.

[22] J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.

[23] A. Efros and T. Leung, *Texture synthesis by non-parametric sampling*, in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, 1999, pp. 1033 –1038 vol.2.

[24] S. Esedoglu and J. Shen, *Digital inpainting based on the mumford-shah-euler image model*, European J. Appl. Math., 13 (2002), pp. 353–370.

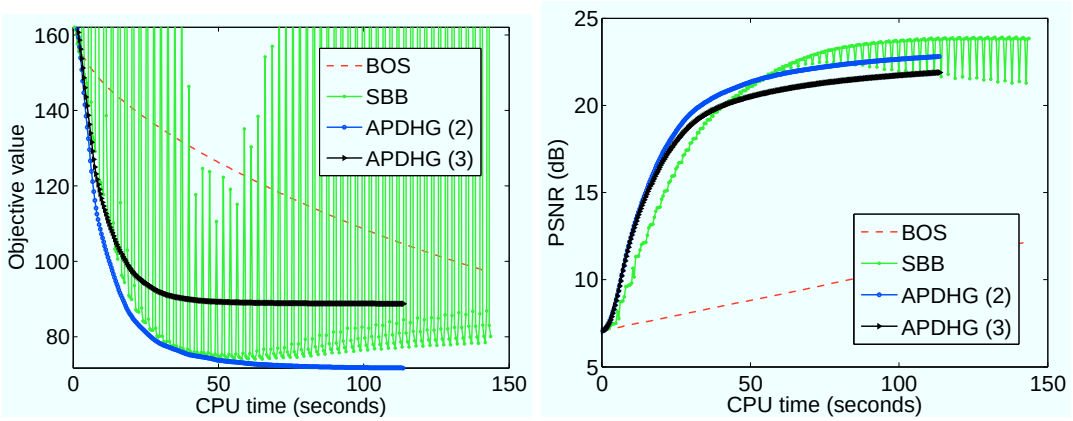


(a) Original (b) Corrupted

(c) BOS (d) SBB

(e) APDHG (2) (f) APDHG (3)

Figure 5: Inpaiting results of boat.



(a) Objective value (b) PSNR

Figure 6: Objective value and PSNR versus CPU time of boat image.

[25] E. Esser, X. Zhang, and T. Chan, *A general framework for a class of first order primal-dual algorithms for tv minimization*, SIAM J. Imag. Sci., 3 (2010), pp. 1015–1046.

[26] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[27] R. Glowinski and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un, et la résolution par pénalisation-dualité d'une classe de problèmes de dirichlet nonlinéaires*, RAIRO Analyse Numérique, 9 (1975), pp. 41–76.

[28] T. Goldstein and S. Osher, *The split Bregman method for L1 regularized problems*, SIAM J. Imag. Sci., 2 (2009), pp. 323–343.

[29] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imag. Sci., 5 (2011), pp. 119–149.

[30] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Francaise Inform. Rech. Oper. Ser. R-3, 4 (1970), pp. 154–158.

[31] S. Masnou and J.-M. Morel, *Level lines based disocclusion*, in Proc. of Intl. Conf. Imag. Proc., 1998.

[32] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489.

[33] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control, 14 (1976), pp. 877–898.

[34] L. Rudin, S. Osher, and E. Fatemi, *Non-linear total variation noise removal algorithm*, Physics D., 60 (1992), pp. 259–268.

[35] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.

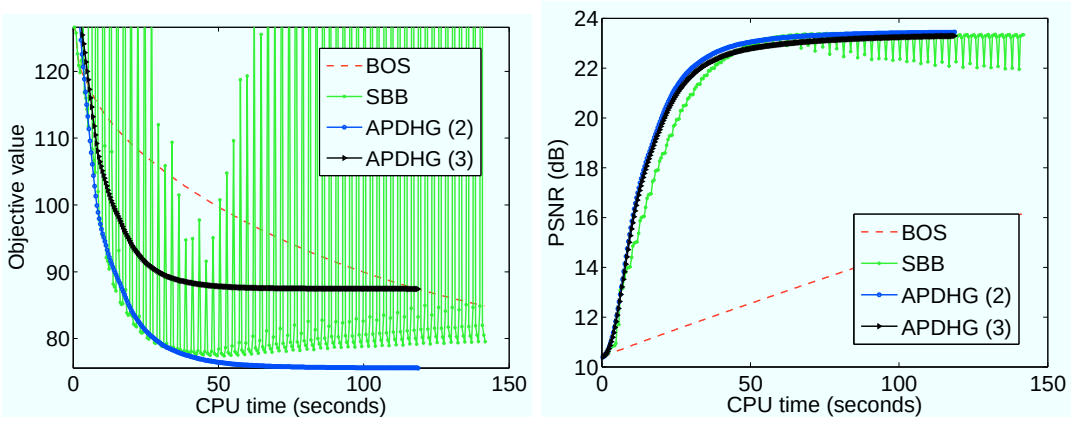
[36] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., 1 (2008), pp. 248–272.

(a) Original (b) Corrupted

(c) BOS (d) SBB

(e) APDHG (2) (f) APDHG (3)

Figure 7: Inpaiting results of man.



(a) Objective value (b) PSNR

Figure 8: Objective value and PSNR versus CPU time of man image.

[37] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split bregman iteration for ROF, vectorial TV, and high order models partial Fourier data*, SIAM J. Imag. Sci., 3 (2010), pp. 300–339.

[38] J. Yang, Y. Zhang, and W. Yin, *A fast TVL1-L2 minimization algorithm for signal reconstruction from partial Fourier data*, IEEE Journal of Selected Topics in Signal Processing, Special Issue on Compressed Sensing, 4 (2010), pp. 288–297.

[39] X. Ye, Y. Chen, and F. Huang, *Computational acceleration for MR image reconstruction in partially parallel imaging*, IEEE Trans. Med. Imag., 30 (2011), pp. 1055–1063.

[40] X. Ye, Y. Chen, W. Lin, and F. Huang, *Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories*, IEEE Trans. Med. Imag., 30 (2011), pp. 575–585.

[41] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imag. Sci., 3 (2010), pp. 253–276.

[42] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.

[43] X. Zhang and T. Chan, *Wavelet inpainting by nonlocal total variation*, Inverse Probl. Imag., 4 (2010), pp. 191–210.

[44] M. Zhu and T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Tech. Rep. 08-34, CAM UCLA, 2008.