Parameter Selection for Total Variation Based Image Restoration Using Discrepancy Principle

You-Wei Wen, Raymond H. Chan

***Abstract*—The key issues in solving image restoration problem successfully are: the estimation of the regularization parameter which balances the data-fidelity with the regularity of the solution; and the development of efficient numerical techniques for computing the solution. In this paper, we derive a fast algorithm that simultaneously estimates the regularization parameter and restores the image. The new approach is based on total-variation (TV) regularized strategy and Morozov discrepancy principle. The TV norm is represented by the dual formulation that changes the minimization problem into a minimax problem. A proximal point method is developed to compute the saddle point of the minimax problem. By adjusting the regularization parameter adaptively in each iteration, the solution is guaranteed to satisfy the discrepancy principle. We will give the convergence proof of our algorithm and show numerically that it is better than some state-of-the-art methods in both speed and accuracy.**

***Index Terms*—Regularization parameter, discrepancy principle, primal-dual, total variation (TV), constrained/unconstrained problem.**

I. Introduction

Image restoration is an important image processing task with many real-world applications, such as surveillance, microscopy imaging, and remote sensing. During acquisition and transmission, digital images are often degraded due to sensor noise, the relative motion between the camera and the original scene, defocusing of the lens system, and the physical size of the sensor elements. In general, the degradation process of a static scene can be modeled with a spatially linear shift-invariant system, where the original image is convolved with a spatially invariant point spread function and added with Gaussian white noise [2]. 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 $n_1 \times n_2$ image, the (r,s)-th pixel becomes the $((r-1)n_1+s)$-th entry of the vector. The discrete imaging model of the degradation process can be expressed as follows:

$$\mathbf{g} = \boldsymbol{H}\mathbf{f}_{\text{clean}} + \mathbf{n}. \tag{1}$$

Here $\mathbf{f}_{\text{clean}}$ and $\mathbf{g}$ are the original image and the observed image respectively, $\boldsymbol{H}$ is the blurring matrix which we assume to be known, and $\mathbf{n}$ is a vector of zero-mean Gaussian white noise with variance σ^2. The task of image restoration is to recover the original image $\mathbf{f}_{\text{clean}}$ from the observed image $\mathbf{g}$ with unknown $\mathbf{n}$ such that $\mathbf{f} \approx \mathbf{f}_{\text{clean}}$.

The simple approach in image restoration is inverse filtering, which solves the least squares problem $\min_{\mathbf{f}} \|\boldsymbol{H}\mathbf{f} - \mathbf{g}\|_2^2$. Obviously $\mathbf{f} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\mathbf{g}$. However, the approach is not feasible either because $(\boldsymbol{H}^T\boldsymbol{H})^{-1}$ does not exist or it is very ill-conditioned that a small perturbation in the observed image $\mathbf{g}$ can produce a large perturbation in the restored image $\mathbf{f}$.

The ill-conditioning can be alleviated by using Total Variation (TV) regularization [55]. The main advantage for the TV formulation is the ability to preserve edges in the image due to the piecewise smooth regularization property of the TV norm. The objective function of the TV image restoration problem is given by

$$\min_{\mathbf{f}} \Phi(\mathbf{f};\lambda) \equiv \left\{ \mathrm{TV}(\mathbf{f}) + \frac{\lambda}{2}\|\boldsymbol{H}\mathbf{f} - \mathbf{g}\|_2^2 \right\}, \tag{2}$$

where $\lambda > 0$ is a fixed, given regularization parameter and $\mathrm{TV}(\mathbf{f}) = \|\nabla\mathbf{f}\|_1$ is the TV norm of $\mathbf{f}$. A number of numerical methods have been proposed for solving (2). They include time marching schemes [40], [55], fixed point iteration method [59], primal-dual Newton method [17], multilevel optimization methods [16], [19], splitting schemes [38], [63], [64] and Nesterov's algorithm [6]. When the TV norm is approximated by an anisotropic TV norm, i.e., $\mathrm{TV}(\mathbf{f}) = \|\nabla_x\mathbf{f}\|_1 + \|\nabla_y\mathbf{f}\|_1$ with ∇_z being the forward difference operator in the z-direction for $z \in \{x, y\}$, there are some other efficient solvers [22], [23], [30].

The objective function in (2) is a weighted sum of two terms: the regularization term and the data fidelity term. The regularization parameter λ plays an important role. By adjusting λ, a compromise is achieved to suppress the noise and preserve the nature of the original image. The appropriate compromise highly depends on the choice of λ. If λ is too large, the regularized solution is under-smoothed, while if λ is too small, the regularized solution does not fit the given data properly. A good recovered image can be obtained by choosing a suitable λ. According to the implicit functions theorem, the minimizer of (2) is a continuous function with respect to λ. Given a λ, we will use $\mathbf{f}(\lambda)$ to denote the optimal solution of the problem (2) for that λ. When there is no ambiguity, we will simply denote $\mathbf{f}(\lambda)$ by $\mathbf{f}$. Usually, λ is determined manually by trial-and-error method, the generalized cross validation (GCV) method [32], [33], the L-curve method [37], the discrepancy principle [45], or the variational Bayes approach [4], [5], [48].

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 Grant No. 11101195, and NSF Grant of KMUST.

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.

The GCV evaluation formula can be derived when the regularization term has a quadratic form. However, due to the non-linearity of TV-norm, it is impossible to derive the GCV evaluation formula when TV-norm is used as the regularization term. When lagged diffusivity fixed point iterations are applied to solve the Euler-Lagrange equation derived from (2), the TV term is linearized by a quadratic formulation in each iteration. Lin *et al.* [43] have applied the unbiased predictive risk estimator (UPRE) to compute the λ of the quadratic approximation. Liao *et al.* [42] incorporated the GCV technique into the splitting-and-regularization framework to handle the parameter estimation for TV-based image reconstruction problem. More precisely, they applied GCV approach to choose the optimal λ for the corresponding Tikhonov problem in each iteration when the alternating minimization method was applied to calculate the minimizer. However, it is usually difficult to calculate the minimizer of the GCV function. Also, it is well known that the GCV method tends to under-smooth the solution. Sometimes, the GCV function can have multiple minimizers [39].

L-curve method is another method to choose the regularization parameter [37], [41]. The L-curve is a parametric plot of $(\log \mathrm{TV}(\mathbf{f}(\lambda)), \log \|\boldsymbol{H}\mathbf{f}(\lambda) - \mathbf{g}\|_2^2)$. Basically, the L-curve is made up of a "flat" part and a "vertical" part. The chosen regularization parameter is the corner point of the L-curve. It is shown that the corner point produces the point of maximum curvature [36]. The main difficulty with L-curve method is that we need to solve (2) many times for different λ's and therefore the algorithm is computational expensive. Sometimes, it is difficult to locate the corner or there does not exist a corner. Also the regularized solutions obtained by the L-curve approach fail to converge to the original image when the noise variance $\sigma \to 0$, see [27], [35], [60].

Another regularization parameter method is Morozov's discrepancy principle which selects λ by matching the norm of the residual to some upper bound, i.e., a good regularized solution $\mathbf{f}$ should lie in the set

$$D = \left\{ \mathbf{f} : \|\boldsymbol{H}\mathbf{f} - \mathbf{g}\|_2^2 \leq c^2 \right\}, \tag{3}$$

where c is a constant which depends on the noise level [3], [10], [32], [47], [62]. When the variance σ^2 of the noise is available, the upper bound is given by $c^2 = \tau n_1 n_2 \sigma^2$ with τ being a pre-determined parameter. In general, one sets $\tau = 1$ [32]. If the variance σ^2 of the noise is unknown, it can be estimated using the median rule [44].

Under the discrepancy principle, the image restoration problem can be represented as solving a constrained optimization problem described as

$$\min_{\mathbf{f} \in D} \mathrm{TV}(\mathbf{f}). \tag{4}$$

A common method [8] to solve (4) is to apply Lagrangian method to convert the constrained minimization problem into an unconstrained minimization problem (2). Mathematically, problems (2) and (4) are equivalent. Assume that there exists a solution for problem (4), it will also be a solution of (2) for a particular choice of $\lambda \geq 0$, which is the Lagrange multiplier corresponding to the constraint $\mathbf{f} \in D$ in (4).

Complementarity condition can be used to show that for the minimizer of (4), we have either $\mathbf{f}(0) \in D$ or

$$\|\boldsymbol{H}\mathbf{f}(\lambda) - \mathbf{g}\|_2^2 = c^2 \tag{5}$$

for $\lambda > 0$. If $\lambda = 0$, then minimizing (2) is equivalent to minimizing $\mathrm{TV}(\mathbf{f})$ and so the solution $\mathbf{f}$ is a constant image, which is not the situation that happens for real world images. Thus the discrepancy principle is trying to seek a $\lambda > 0$ such that (5) holds when the minimizer $\mathbf{f}$ in (4) is not a constant image.

Notice that there does not exist a closed-form solution for (2), and hence it is difficult to find a solution of λ in (5). Blomgren and T. Chan [10] developed a modular solver to update λ in order to make use of existing methods of the unconstrained problem to compute the corresponding constrained one. In [3], Aujol and Gilboa considered to automate the choice of λ for denoising problems. Their approach was to compute the optimal solution of (2) for a given parameter λ by applying Chambolle's projection algorithm. If (5) does not hold for the solution $\mathbf{f}(\lambda)$, an updating rule is applied to adjust the parameter λ. This procedure is applied iteratively until (5) is satisfied. However, the approaches in [3] and [10] need to solve problem (2) many times for a sequence of λ's. Hence the computational cost is expensive.

In [11], [49], the authors used the iterated refinement method to solve TV image restoration problem (2). The residual $\|\boldsymbol{H}\mathbf{f} - \mathbf{g}\|_2^2$ would decrease monotonically during the iterative procedure, and hence (5) was used as a stopping criterion. In fact, the iteration number plays the role of regularization in the iterative procedure. However, the main aim of [11], [49] was to use Bregman distances to design an iterative regularization procedure in order to improve TV restoration results rather than to find the minimizer of (4). The iterated refinement method also requires to solve a series of problems (2) for different $\mathbf{g}$'s.

In [47], [62], projected gradient descent method was applied to solve (4). In each iteration, in order to ensure that the current iterant is a feasible one, it is projected on the feasible set D. Finding the projection onto the set D is equivalent to solving a constrained least square problem. Its solution can be computed efficiently when the blur matrix $\boldsymbol{H}$ can be diagonalized. This is the case under the assumption of periodic boundary condition or for symmetric point spread function with Neumann boundary condition [46].

In this paper, we apply the discrepancy principle to estimate the regularization parameter λ. A proximal-based primal-dual method, where Legendre-Fenchel's duality is used to represent the TV norm, is used to solve (4). The minimization problem is solved by finding a saddle point of the primal-dual function. Proximal point iterations are applied to the sub-differential of the primal-dual function alternately with the primal variable and the dual variable fixed alternatively. In each iteration, λ is updated in order to guarantee that the primal variable is in the feasible set D. Numerical results show that our algorithm is very effective in finding a good λ, and it is even faster than the methods that solve the unconstrained problem (2) in [56], [61].

We will provide a convergence proof of our algorithm. We emphasize that the convergence property is unknown in the parameter selection methods using GCV approach [42], [43]. We now highlight the major differences between our method and previous methods using the discrepancy principle to solve the TV problems. The methods in [3], [10], [15], [11], [49], require to solve a series of unconstrained minimization problems (2) where each problem corresponds to a different λ. Our method just needs to solve one minimization problem where λ is changing adaptively during the iterations. Also the methods in [1], [15], [47] were more complex than our method since several variables were introduced while our algorithm only introduces a dual variable and hence is easier to implement. In fact, all the parameters in our algorithm are either fixed constants or computed automatically by the algorithm itself, so the algorithm can be run without any supervision.

The outline of the paper is as follows. In Section II, we present a primal-dual model for TV-based image restoration problem where the TV norm is presented by Fenchel's duality. In Section III, we apply the discrepancy principle to choose the regularization parameter λ in each iteration. We also give the convergence proof for our algorithm. Experiment results are reported in Section IV to demonstrate the effectiveness of our algorithm in choosing λ. Finally, a short conclusion is given in Section V.

II. Primal-dual Model

A. Notations

Let us describe the notations we will be using throughout this paper. We denote by X the Euclidean space $\mathbb{R}^{n_1\times n_2}$, and $Y = X \times X$. For $\mathbf{g} \in X$, $g_{i,j} \in \mathbb{R}$ denotes the $((i-1)n_1+j)$-th component of $\mathbf{g}$. For $\mathbf{p} \in Y$, $p_{i,j} = (p_{i,j,1}, p_{i,j,2}) \in \mathbb{R}^2$ denotes the $((i-1)n_1+j)$-th component of $\mathbf{p}$. Define the inner product $\langle \mathbf{g}, \mathbf{f}\rangle_X = \sum_{i,j} g_{i,j} f_{i,j}$, $\langle \mathbf{p}, \mathbf{q}\rangle_Y = \sum_{i,j}\sum_{k=1}^{2} p_{i,j,k} q_{i,j,k}$ and the norm $\|\mathbf{f}\|_2 = \sqrt{\langle \mathbf{f}, \mathbf{f}\rangle_X}$, $\|\mathbf{p}\|_2 = \sqrt{\langle \mathbf{p}, \mathbf{p}\rangle_Y}$ and $\|\mathbf{p}\|_\infty = \max_{i,j}\{|p_{i,j}|\}$ with $|p_{i,j}| = \sqrt{p_{i,j,1}^2 + p_{i,j,2}^2}$.

To define a discrete total variation, we introduce a discrete version of the gradient operator. For any $\mathbf{f} \in X$, the gradient ∇ is a linear operator from X to Y, $\nabla\mathbf{f}$ is a vector in Y given by $(\nabla\mathbf{f})_{i,j} = ((\nabla_x\mathbf{f})_{i,j}, (\nabla_y\mathbf{f})_{i,j})$. The discrete version of the divergence operator is defined by $\mathrm{div} = -\nabla^T$ where ∇^T is the adjoint of ∇. Therefore, for every $\mathbf{p} \in Y$ and $\mathbf{f} \in X$, we have $\langle -\mathrm{div}\mathbf{p}, \mathbf{f}\rangle_X = \langle \mathbf{p}, \nabla\mathbf{f}\rangle_Y$.

Define the set

$$A \equiv \{\mathbf{p} \in Y : \|\mathbf{p}\|_\infty \le 1\}, \tag{6}$$

and the characteristic function δ_A of A as $\delta_A(\mathbf{p}) = 0$ for $\mathbf{p} \in A$ and $\delta_A(\mathbf{p}) = +\infty$ if $\mathbf{p} \notin A$. The discrete TV of the image $\mathbf{f}$ is also the Legendre-Fenchel conjugate of δ [13], [17]

$$\mathrm{TV}(\mathbf{f}) = \max_{\mathbf{p}\in Y}\{\langle \mathrm{div}\mathbf{p}, \mathbf{f}\rangle_X - \delta_A(\mathbf{p})\} = \max_{\mathbf{p}\in A}\langle \mathrm{div}\mathbf{p}, \mathbf{f}\rangle_X. \tag{7}$$

B. Primal-dual Model

We present a primal-dual model for problem (2). Using Legendre-Fenchel's duality (7) to represent the TV norm, problem (2) can be written as

$$\min_{\mathbf{f}} \max_{\mathbf{p}\in A} \mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda) \tag{8}$$

with

$$\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda) \equiv \langle \mathbf{f}, \mathrm{div}\mathbf{p}\rangle_X + \frac{\lambda}{2}\|\boldsymbol{H}\mathbf{f} - \mathbf{g}\|_2^2. \tag{9}$$

To emphasize the dependence of $\mathcal{J}$ on the parameter λ, we explicitly write λ out in its arguments. We note that $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$ in (9) is convex in $\mathbf{f}$ and concave in $\mathbf{p}$. We call a pair $(\mathbf{f}^*, \mathbf{p}^*)$ a saddle point for $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$ if

$$\mathcal{J}(\mathbf{f}^*, \mathbf{p}; \lambda) \le \mathcal{J}(\mathbf{f}^*, \mathbf{p}^*; \lambda) \le \mathcal{J}(\mathbf{f}, \mathbf{p}^*; \lambda), \quad \forall \mathbf{f}, \mathbf{p}.$$

Notice that the null space of $\boldsymbol{H}$ does not contain any constant vectors since $\boldsymbol{H}$ is a blur matrix generated by some point spread function and we have $\boldsymbol{H}\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector of all ones. On the contrary, the null space of ∇ is the set of all constant vectors. Thus the intersection of the null space of $\boldsymbol{H}$ and the null space of ∇ is an empty set. Hence there exists a saddle point $(\mathbf{f}^*, \mathbf{p}^*)$ of $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$ [7]. Using the existence of the saddle point of $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$, convex analysis can be applied to show that the minimum and the maximum in (8) can be swapped, i.e.

$$\min_{\mathbf{f}} \max_{\mathbf{p}\in A} \mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda) = \mathcal{J}(\mathbf{f}^*, \mathbf{p}^*; \lambda) = \max_{\mathbf{p}\in A} \min_{\mathbf{f}} \mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda).$$

Using primal-dual model, an optimal solution $\mathbf{f}^*$ of (2) can be obtained by calculating the saddle point of $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$. By using Proposition 2.6.1 in [7], a pair $(\mathbf{f}^*, \mathbf{p}^*)$ is a saddle point of (8) if and only if $\mathbf{f}^*$ and $\mathbf{p}^*$ are the optimal solutions of the problems

$$\mathbf{f}^* = \operatorname*{argmin}_{\mathbf{f}} \mathcal{J}(\mathbf{f}, \mathbf{p}^*; \lambda) \tag{10}$$

$$\mathbf{p}^* = \operatorname*{argmax}_{\mathbf{p}\in A} \mathcal{J}(\mathbf{f}^*, \mathbf{p}; \lambda), \tag{11}$$

respectively. From (8) and (10), we obtain

$$\mathrm{div}\mathbf{p}^* + \lambda\boldsymbol{H}^T(\boldsymbol{H}\mathbf{f}^* - \mathbf{g}) = 0. \tag{12}$$

The KKT necessary conditions of the dual optimality for (11) yield the existence of the Lagrange multipliers $\gamma_{i,j} \ge 0$, associated with the constraint $\mathbf{p} \in A$, such that we have

$$\big(\nabla\mathbf{f}^*\big)_{i,j} + \gamma_{i,j}p_{i,j}^* = 0, \quad \forall i, j$$

where either $\gamma_{i,j} > 0$ with $|p_{i,j}| = 1$, or $\gamma_{i,j} = 0$ with $|p_{i,j}| < 1$. We see that in either case $\gamma_{i,j} = |\big(\nabla\mathbf{f}^*\big)_{i,j}|$. Therefore, if $\mathbf{p}^*$ is a solution of (11), we have

$$\big(\nabla\mathbf{f}^*\big)_{i,j} + |\big(\nabla\mathbf{f}^*\big)_{i,j}|p_{i,j}^* = 0, \quad \forall i, j. \tag{13}$$

Hence we have the following lemma.

Lemma 1. *Assume that* $(\mathbf{f}^*, \mathbf{p}^*)$ *is the saddle point of* $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda)$ *with* $\mathbf{p} \in A$, *see* (8). *Then the equations* (12) *and* (13) *hold.*

C. Proximal Point Method

Seeking the saddle point of (8) is equivalent to solving the system (10)–(11). Note that (11) involves a non-differentiable function which poses a serious computational difficulty. To deal with it, we use iterative methods. We apply primal-dual proximal point method [18], [21], [25], [26], [53], [54], [58] to compute the saddle point of (8) alternately with the primal variable $\mathbf{f}$ and the dual variable $\mathbf{p}$ fixed alternatively. More specifically, starting at a point $(\mathbf{f}^{(0)}, \mathbf{p}^{(0)})$, the sequence is generated successively according to the iteration

$$\mathbf{p}^{(k+\frac{1}{2})} = \underset{\mathbf{p}\in A}{\operatorname{argmin}}\, \psi_k(\mathbf{p}; \mathbf{f}^{(k)}, \lambda), \tag{14}$$

$$\mathbf{f}^{(k+1)} = \underset{\mathbf{f}}{\operatorname{argmin}}\, \phi_k(\mathbf{f}; \mathbf{p}^{(k+\frac{1}{2})}, \lambda), \tag{15}$$

$$\mathbf{p}^{(k+1)} = \underset{\mathbf{p}\in A}{\operatorname{argmin}}\, \psi_k(\mathbf{p}; \mathbf{f}^{(k+1)}, \lambda), \tag{16}$$

where

$$\psi_k(\mathbf{p}; \mathbf{f}, \lambda) \equiv -\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda) + \frac{1}{2s}\big\|\mathbf{p} - \mathbf{p}^{(k)}\big\|_2^2, \tag{17}$$

$$\phi_k(\mathbf{f}; \mathbf{p}, \lambda) \equiv \mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda) + \frac{1}{2t}\big\|\mathbf{f} - \mathbf{f}^{(k)}\big\|_2^2 \tag{18}$$

with the constants $s, t > 0$. We remark that our method is different from the primal-dual method proposed in [65], [66] where a gradient descent method is employed to the primal and dual variables alternatively. Our method resembles in some way the dual method in [31] which uses a predictor-corrector scheme [18] in the alternating direction iterations for the dual variable. We remark that in order to prove convergence of our method, we iterate $\mathbf{p}$ twice and $\mathbf{f}$ once in each iteration and it suffices to set $t = 1$ and $s = 1/16$, see Sections III-A and IV-B.

Next we discuss how to solve subproblems (14)–(16).

1) Subproblems for the dual variable $\mathbf{p}$*:* In order to solve problems (14) and (16), we first define the projection operator onto the set A given by (6). The projection of a vector $\mathbf{q}$ onto A can be conveniently expressed as

$$\mathcal{P}_A(\mathbf{q}) = \underset{\mathbf{p}\in A}{\operatorname{argmin}} \|\mathbf{p} - \mathbf{q}\|_2^2. \tag{19}$$

Lagrangian method can be applied to calculate the projection operator $\mathcal{P}_A$. The Lagrangian function associated with (19) is

$$\|\mathbf{p} - \mathbf{q}\|_2^2 + \sum_{i,j} \beta_{i,j}\big(|p_{i,j}|^2 - 1\big),$$

where $\beta_{i,j} \geq 0$ is the Lagrangian multiplier associated with the constraint $|p_{i,j}|^2 \leq 1$. Its complementarity conditions implies that for the optimal $\beta_{i,j}$, either $\beta_{i,j} = 0$ with $|p_{i,j}|, |q_{i,j}| < 1$, or $\beta_{i,j} > 0$ with $|p_{i,j}| = 1$ and $|q_{i,j}| \geq 1$. In the former case, we have $p_{i,j} = q_{i,j}$. In the latter case, the KKT conditions yields

$$p_{i,j} - q_{i,j} + \beta_{i,j} p_{i,j} = 0, \quad \forall i, j.$$

Therefore, we have $\beta_{i,j} = |q_{i,j}| - 1$, and thus $p_{i,j} = q_{i,j}/|q_{i,j}|$. Hence, we obtain

$$\big(\mathcal{P}_A(\mathbf{q})\big)_{i,j} = \frac{q_{i,j}}{\max(1, |q_{i,j}|)}. \tag{20}$$

We consider the solution of (14) now. By (17), problem (14) can be written as

$$\begin{aligned}\mathbf{p}^{(k+\frac{1}{2})} &= \underset{\mathbf{p}\in A}{\operatorname{argmin}} \left\{ \big\langle \mathbf{p}, \nabla \mathbf{f}^{(k)} \big\rangle_Y + \frac{1}{2s}\big\|\mathbf{p} - \mathbf{p}^{(k)}\big\|_2^2 \right\} \\ &= \underset{\mathbf{p}\in A}{\operatorname{argmin}} \frac{1}{2s} \left\{ 2\big\langle \mathbf{p} - \mathbf{p}^{(k)}, s\nabla \mathbf{f}^{(k)} \big\rangle_Y + \big\|\mathbf{p} - \mathbf{p}^{(k)}\big\|_2^2 \right\} \\ &= \underset{\mathbf{p}\in A}{\operatorname{argmin}} \big\|\mathbf{p} - \big(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k)}\big)\big\|_2^2.\end{aligned}$$

The minimization problem is equivalent to computing the projection of $\big(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k)}\big)$ onto the set A. Therefore

$$\mathbf{p}^{(k+\frac{1}{2})} = \mathcal{P}_A\big(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k)}\big). \tag{21}$$

Similarly, $\mathbf{p}^{(k+1)}$ in (16) can be computed by

$$\mathbf{p}^{(k+1)} = \mathcal{P}_A(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k+1)}). \tag{22}$$

2) Subproblem for the primal variable $\mathbf{f}$*:* From (18), we see that the objective function in (15) is quadratic with respect to $\mathbf{f}$. Hence $\mathbf{f}^{(k+1)}$ can easily be computed by the formula

$$\mathbf{f}^{(k+1)} = \big(\lambda t \boldsymbol{H}^T \boldsymbol{H} + \boldsymbol{I}\big)^{-1}\big(\lambda t \boldsymbol{H}^T \mathbf{g} + \mathbf{u}_k\big), \tag{23}$$

where

$$\mathbf{u}_k = \mathbf{f}^{(k)} - t\mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}. \tag{24}$$

We remark that in image restoration $\boldsymbol{H}$ is a block-circulant with circulant-block (BCCB) matrix when periodic boundary conditions are applied to the image boundary. The matrix $\boldsymbol{H}$ can be diagonalized by fast Fourier transform matrix [14]. Therefore, (24) can be solved using three Fast Fourier Transforms (FFT) in $O(n_1 n_2 \log(n_1 n_2))$ operations for an $n_1 \times n_2$ restored image, see for instance [46]. In the following, we will use periodic boundary conditions.

The resulting algorithm for TV image restoration using the primal-dual model is summarized in Algorithm 1.

Algorithm 1 Primal-Dual Model for TV Image Restoration Algorithm (PDM-TV)

Function: $\mathbf{f} = \text{PDM-TV}(\mathbf{g}, \boldsymbol{H})$.
Input: $\mathbf{g}$, $\boldsymbol{H}$.
1: Initialize $\mathbf{f}^{(0)}$ and $\mathbf{p}^{(0)}$. Set the step sizes s and t.
2: **while** stopping criterion is not satisfied **do**
3: $\quad \mathbf{p}^{(k+\frac{1}{2})} = \mathcal{P}_A\big(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k)}\big)$;
4: $\quad \mathbf{u}_k = \mathbf{f}^{(k)} - t\mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}$;
5: $\quad \mathbf{f}^{(k+1)} = \big(\lambda t \boldsymbol{H}^T \boldsymbol{H} + \boldsymbol{I}\big)^{-1}\big(\lambda t \boldsymbol{H}^T \mathbf{g} + \mathbf{u}_k\big)$;
6: $\quad \mathbf{p}^{(k+1)} = \mathcal{P}_A(\mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k+1)})$;
7: **end while**
8: **return** $\mathbf{f} = \mathbf{f}^{(k+1)}$.

III. Adaptive Regularization Parameter Selection

In this section, the discrepancy principle strategy is used to determine the regularization parameter λ in each iteration step. Note that Algorithm 1 is a proximal-based primal-dual method applied to solve the unconstrained problem (2). We will modify it to generate a sequence $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)}, \lambda_k)$ where λ_k will converge to the Lagrange multiplier λ^* corresponding to the constraint $\mathbf{f} \in D$ given in (3) and $(\mathbf{f}^*, \mathbf{p}^*)$ will converge

to the saddle point of $\mathcal{J}(\mathbf{f},\mathbf{p};\lambda^*)$ in (8). The convergence proof will be given in Section III-A. Our strategy is to adjust λ adaptively during the iterations so that the restored image is always in the feasible set D. More specifically, we replace the iteration step (15) with the following one:

$$\mathbf{f}^{(k+1)} = \underset{\mathbf{f}\in D}{\operatorname{argmin}}\, \phi_k(\mathbf{f};\mathbf{p}^{(k+\frac{1}{2})},\lambda_{k+1}). \tag{25}$$

Here $\lambda_{k+1}\geq 0$ is the Lagrange multiplier corresponding to the constraint $\mathbf{f}\in D$. Hence the updating rule in Step 5 of Algorithm 1, i.e. (23), is modified as

$$\mathbf{f}^{(k+1)}(\lambda_{k+1}) = \big(\lambda_{k+1}t\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I}\big)^{-1}\big(\lambda_{k+1}t\boldsymbol{H}^T\mathbf{g}+\mathbf{u}_k\big). \tag{26}$$

Since $\mathbf{f}^{(k+1)}$ has a closed-form solution, it is possible to analyse the discrepancy $\mathbf{e}_{k+1}$ given by

$$\begin{aligned}\mathbf{e}_{k+1} &= \boldsymbol{H}\mathbf{f}^{(k+1)}(\lambda_{k+1})-\mathbf{g}\\ &= \boldsymbol{H}\big(\lambda_{k+1}t\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I}\big)^{-1}\big(\lambda_{k+1}t\boldsymbol{H}^T\mathbf{g}+\mathbf{u}_k\big)-\mathbf{g}\end{aligned}$$

Using the identity $\mu\boldsymbol{H}\big(\mu\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I}\big)^{-1} = \big(\mu\boldsymbol{H}\boldsymbol{H}^T+\boldsymbol{I}\big)^{-1}\boldsymbol{H}^T$, we obtain

$$\mathbf{e}_{k+1} = \big(\lambda_{k+1}t\boldsymbol{H}\boldsymbol{H}^T+\boldsymbol{I}\big)^{-1}\big(\boldsymbol{H}\mathbf{u}_k-\mathbf{g}\big). \tag{27}$$

Define the function $\mathcal{K}(\lambda,\mathbf{u})$ as

$$\mathcal{K}(\lambda,\mathbf{u}) \equiv \Big\|\big(\lambda t\boldsymbol{H}\boldsymbol{H}^T+\boldsymbol{I}\big)^{-1}\big(\boldsymbol{H}\mathbf{u}-\mathbf{g}\big)\Big\|_2^2. \tag{28}$$

It is obviously that $\mathcal{K}(\lambda_{k+1},\mathbf{u}_k) = \|\mathbf{e}_{k+1}\|_2^2$. The following lemma states that there must exist a unique regularization λ_{k+1} such that $\big\|\boldsymbol{H}\mathbf{f}^{(k+1)}(\lambda_{k+1})-\mathbf{g}\big\|_2^2 = c^2$ when $\big\|\boldsymbol{H}\mathbf{u}_k-\mathbf{g}\big\|_2^2 > c^2$. Similar results can be found in [12], [51].

Lemma 2. *Let* $\mathbf{r}=\boldsymbol{H}\mathbf{u}-\mathbf{g}$ *and* $\mathcal{K}(\lambda,\mathbf{u})$ *be defined by* (28). *Then* $\mathcal{K}(\lambda,\mathbf{u})$ *is a strictly positive and strictly monotonically decreasing convex function of* λ. *Moreover, the equation*

$$\mathcal{K}(\lambda,\mathbf{u}) = b^2 \tag{29}$$

has a ***unique*** *solution* $\lambda>0$ *for any* b *satisfying* $\|\mathbf{r}_0\|_2^2 \leq b^2 < \|\mathbf{r}\|_2^2$, *where* $\mathbf{r}_0$ *denotes the orthogonal projection of* $\mathbf{r}$ *onto the null space of* $\boldsymbol{H}\boldsymbol{H}^T$.

Lemma 2 can be proven by directly computing the first and second order derivatives of $\mathcal{K}(\lambda,\mathbf{u})$ with respect to λ and the unique solution of λ is guaranteed by the strict convexity of $\mathcal{K}$. From Lemma 2, we know that when $\|\boldsymbol{H}\mathbf{u}_k-\mathbf{g}\|_2^2 > c^2$, i.e., when $\mathbf{u}_k\notin D$, there exists a unique solution to the equation $\mathcal{K}(\lambda,\mathbf{u}_k)=c^2$. It means that we can find a unique $\lambda_{k+1}>0$ such that $\|\boldsymbol{H}\mathbf{f}^{(k+1)}(\lambda_{k+1})-\mathbf{g}\|_2^2 = c^2$. When $\mathbf{u}_k\in D$, we can simply choose $\lambda_{k+1}=0$. This leads to $\mathbf{f}^{(k+1)}=\mathbf{u}_k\in D$, see (26). Therefore, there always exists a unique $\lambda_{k+1}\geq 0$ such that $\mathbf{f}^{(k+1)}\in D$. Notice that problem (29) is nonlinear, we therefore apply Newton method to solve it. In order to accelerate the convergence speed, we use λ_k as the initial value when computing λ_{k+1}.

The resulting algorithm is summarized in Algorithm 2.

Algorithm 2 Discrepancy Principle Based Primal-Dual Model for TV Image Restoration Algorithm (DP-PDM-TV)

Function: $(\mathbf{f},\lambda) = \text{DP-PDM-TV}(\mathbf{g},\boldsymbol{H},c^2)$.
Input: $\mathbf{g}$, $\boldsymbol{H}$, c^2.
1: Initialize $\mathbf{f}^{(0)}$ and $\mathbf{p}^{(0)}$. Set the step sizes s and t.
2: **while** stopping criterion is not satisfied **do**
3: $\quad\mathbf{p}^{(k+\frac{1}{2})} = \mathcal{P}_A\big(\mathbf{p}^{(k)}-s\nabla\mathbf{f}^{(k)}\big)$;
4: $\quad\mathbf{u}_k = \mathbf{f}^{(k)}-t\text{div}\mathbf{p}^{(k+\frac{1}{2})}$;
5: $\quad$**if** $\mathbf{u}_k\in D$; **then**
6: $\quad\quad\lambda_{k+1}=0$;
7: $\quad$**else**
8: $\quad\quad$Solve $\mathcal{K}(\lambda_{k+1},\mathbf{u}_k)=c^2$;
9: $\quad$**end if**
10: $\quad\mathbf{f}^{(k+1)} = \big(\lambda_{k+1}t\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I}\big)^{-1}\big(\lambda_{k+1}t\boldsymbol{H}^T\mathbf{g}+\mathbf{u}_k\big)$;
11: $\quad\mathbf{p}^{(k+1)} = \mathcal{P}_A(\mathbf{p}^{(k)}-s\nabla\mathbf{f}^{(k+1)})$;
12: **end while**
13: **return** $\mathbf{f}=\mathbf{f}^{(k+1)}$ and $\lambda=\lambda_{k+1}$.

A. Convergence Analysis

We now show that the sequence $\mathbf{f}^{(k)}$ generated by Algorithm 2 converges to the minimizer $\mathbf{f}^*$ of the problem (4) and the sequence λ_k converges to the Lagrange multiplier λ^* corresponding to the constraint $\mathbf{f}\in D$, i.e., the solution of (4) is also a solution of (2) for the choice of $\lambda=\lambda^*$. Note that by (2) and (8), we have $\Phi(\mathbf{f};\lambda^*) = \max_{\mathbf{p}\in A}\mathcal{J}(\mathbf{f},\mathbf{p};\lambda^*)$. Thus we need to show that the sequence $(\mathbf{f}^{(k)},\mathbf{p}^{(k)})$ generated by Algorithm 2 is convergent to the saddle point of $\mathcal{J}(\mathbf{f},\mathbf{p};\lambda^*)$ with $\mathbf{p}\in A$.

We need two Lemmas and their proofs are given in Appendix. The first Lemma gives a lower bound of the difference between the primal variables at two consecutive iterations. The bound can be obtained precisely because we have performed two proximal steps to the dual variable in each iteration. The second Lemma states that the sequence $(\mathbf{f}^{(k)},\mathbf{p}^{(k)})$ generated by (14), (16), and (25) (i.e., Algorithm (2)) is bounded when $st\leq 1/16$.

Lemma 3. *Let* $(\mathbf{f}^{(k)},\mathbf{p}^{(k)},\lambda_k)$ *be the sequence generated by Algorithm 2, see* (14), (16), *and* (25). *Then*

$$8s\big\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\big\|_2^2 \geq \big\langle \mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}, \text{div}(\mathbf{p}^{(k+1)}-\mathbf{p}^{(k+\frac{1}{2})})\big\rangle_X. \tag{30}$$

Lemma 4. *Assume that* $(\mathbf{f}^{(k)},\mathbf{p}^{(k)},\lambda_k)$ *be the sequence generated by Algorithm 2, see* (14), (16), *and* (25). *When* $st\leq 1/16$, *we have*

$$\begin{aligned}&\frac{1}{t}\|\mathbf{f}^*-\mathbf{f}^{(k)}\|_2^2+\frac{1}{s}\|\mathbf{p}^*-\mathbf{p}^{(k)}\|_2^2\\ \geq\ &\frac{1}{t}\|\mathbf{f}^*-\mathbf{f}^{(k+1)}\|_2^2+\frac{1}{s}\|\mathbf{p}^*-\mathbf{p}^{(k+1)}\|_2^2.\end{aligned} \tag{31}$$

In particular, the sequence $(\mathbf{f}^{(k)},\mathbf{p}^{(k)},\lambda_k)$ *converges to some limit point* $(\mathbf{f}^\dagger,\mathbf{p}^\dagger,\lambda^\dagger)$ *and either* $\lambda^\dagger=0$ *with* $\mathbf{f}^\dagger\in D$ *or* $\lambda^\dagger>0$ *is the solution to the equation* $\big\|\boldsymbol{H}\mathbf{f}^\dagger(\lambda)-\mathbf{g}\big\|_2^2 = c^2$.

We first show that the limit point $(\mathbf{f}^\dagger,\mathbf{p}^\dagger)$ is a saddle point of $\mathcal{J}(\mathbf{f},\mathbf{p};\lambda^\dagger)$ with $\mathbf{p}^\dagger\in A$ (see (8)). According the Steps 4 and 10 of Algorithm 2, we know that $(\mathbf{f}^\dagger,\mathbf{p}^\dagger)$ satisfies the

following equation:

$$\mathbf{f}^\dagger = \left(\lambda^\dagger t \boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{I}\right)^{-1}\left(\lambda^\dagger t \boldsymbol{H}^T\mathbf{g} + \mathbf{f}^\dagger - t\mathrm{div}\mathbf{p}^\dagger\right).$$

Multiplying $\left(\lambda^\dagger t\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{I}\right)$ on both sides, and re-arranging the terms, we have $\mathrm{div}\mathbf{p}^\dagger + \lambda^\dagger \boldsymbol{H}^T\left(\boldsymbol{H}\mathbf{f}^\dagger - \mathbf{g}\right) = 0$, which is exactly equation (12) with λ replaced by $\lambda^\dagger$.

According to Step 11 in Algorithm 2, we know that $(\mathbf{f}^\dagger, \mathbf{p}^\dagger)$ also satisfies the following equation:

$$\mathbf{p}^\dagger = \mathcal{P}_A(\mathbf{p}^\dagger - s\nabla\mathbf{f}^\dagger)$$

According to the property of the projection operator $\mathcal{P}_A$ (see (20)), we have

$$p^\dagger_{i,j} = \frac{p^\dagger_{i,j} - s\left(\nabla\mathbf{f}^\dagger\right)_{i,j}}{\max\left(1, \left|p^\dagger_{i,j} - s\left(\nabla\mathbf{f}^\dagger\right)_{i,j}\right|\right)}.$$

Therefore if $\left|p^\dagger_{i,j} - s\left(\nabla\mathbf{f}^\dagger\right)_{i,j}\right| \le 1$, we have $\left(\nabla\mathbf{f}^\dagger\right)_{i,j} = 0$ and $\left|p^\dagger_{i,j}\right| \le 1$. If $\left|p^\dagger_{i,j} - s\left(\nabla\mathbf{f}^\dagger\right)_{i,j}\right| > 1$, we have $\left|p^\dagger_{i,j}\right| = 1$ and $s\left(\nabla\mathbf{f}^\dagger\right)_{i,j} = (1 - \left|p^\dagger_{i,j} - s\left(\nabla\mathbf{f}^\dagger\right)_{i,j}\right|)p^\dagger_{i,j}$. Thus $\mathbf{p}^\dagger \in A$ and the KKT optimality condition for the dual variable, i.e. (13), holds.

Thus we have the following theorem.

Theorem 1. *Assume that $(\mathbf{f}^*, \mathbf{p}^*)$ is the saddle point of $\mathcal{J}(\mathbf{f}, \mathbf{p}; \lambda^*)$, where λ^* is the Lagrange multiplier corresponding to the constraint $\mathbf{f} \in D$. Then the sequence $(\mathbf{f}^{(k)}, \mathbf{p}^{(k)}, \lambda_k)$ generated by Algorithm 2 converges to $(\mathbf{f}^*, \mathbf{p}^*, \lambda^*)$ provided that $st \le 1/16$. In particular, $\mathbf{f}^{(k)}$ converges to the minimizer of* (4)*, λ_k converges the Lagrange multiplier corresponding to the constraint $\mathbf{f} \in D$ associated with the unconstrained problem* (2)*.*

B. Upper Bound c^2

A good choice for the upper bound c^2 should minimize the error in the restored image. However, since the original image $\mathbf{f}_{\mathrm{clean}}$ is unknown, the upper bound should be chosen according to the noise level [3], [10], [32], [47], [62]. If the knowledge of the noise variance σ^2 is not available, it can be estimated using the median rule [44]. Given a wavelet transform with $\widehat{\mathbf{g}}_{HH}$ being the high-high coefficients of the observed image $\mathbf{g}$ at the finest wavelet transform level, the noise variance can be estimated by

$$\sigma = \mathrm{median}(|\widehat{\mathbf{g}}_{HH}|)/0.6745.$$

Here $\mathrm{median}(|\widehat{\mathbf{g}}_{HH}|)$ is the median of the absolute value of $\widehat{\mathbf{g}}_{HH}$. Once we have σ, the upper bound c^2 is generally chosen as $c^2 = \tau n_1 n_2 \sigma^2$ for some constant τ. A typical choice is to set $\tau = 1$ [3], [10], [32], [47], [62]. It was observed however in [24], [32] that the choice of λ based on $\tau = 1$ usually yields an over-smooth solution, which implies that λ is too small. Lemma 2 states that the norm of the discrepancy is a monotonically decreasing function of λ. In order to obtain a larger λ, we should choose a τ less than 1.

In order to determine a good τ, we use the approach of equivalent Degrees of Freedom (DF) [32], [34], [57]. It provides an estimator of τ by solving the equation

$$\|\boldsymbol{H}\mathbf{f}(\lambda) - \mathbf{g}\|_2^2 = \mathrm{DF}\cdot\sigma^2,$$

where DF is the effective number of degrees of freedom. The main challenge for DF approach is the lack of an analytical expression of DF since there does not exist a closed-form formula for the solution $\mathbf{f}(\lambda)$. Our aim is to find an approximate one.

Let $(\mathbf{f}^\dagger, \mathbf{p}^\dagger, \lambda^\dagger)$ be the solution generated by Algorithm 2 for the parameter $\tau = 1$. The discrepancy $\mathbf{e}^\dagger = \boldsymbol{H}\mathbf{f}^\dagger - \mathbf{g}$ is given by

$$\mathbf{e}^\dagger = \boldsymbol{H}\left(\lambda^\dagger t\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{I}\right)^{-1}\left(\lambda^\dagger t\boldsymbol{H}^T\mathbf{g} + \mathbf{u}^\dagger\right) - \mathbf{g},$$

where $\mathbf{u}^\dagger = \mathbf{f}^\dagger - t\mathrm{div}\mathbf{p}^\dagger$, see (24) or Step 4 of Algorithm 1. Hence by (1) we obtain

$$\mathbf{e}^\dagger = \widehat{\mathbf{e}}_f - \left(\lambda^\dagger t\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{I}\right)^{-1}\mathbf{n},$$

where $\widehat{\mathbf{e}}_f = \boldsymbol{H}\left(\lambda^\dagger t\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{I}\right)^{-1}\mathbf{u}^\dagger - \left(\lambda^\dagger t\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{I}\right)^{-1}\boldsymbol{H}\mathbf{f}_{\mathrm{clean}}$ is a fixed vector. As a consequence, $\|\mathbf{e}^\dagger\|_2^2$ is χ^2-distributed with variance σ^2 and its degrees of freedom is equal to $\mathrm{trace}\left[(\lambda^\dagger t\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{I})^{-1}\right]$ [32], [34], [57]. Therefore, we choose the parameter τ as

$$\tau = \mathrm{trace}\left[(\lambda^\dagger t\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{I})^{-1}\right]/(n_1 n_2). \tag{32}$$

Notice that $\lambda^\dagger t > 0$, hence $\tau < 1$, as we would expect. In the experiments of Section IV-A, we can see that (32) performs quite well in approximating the optimal τ.

We remark that the parameter τ is dependent on the degrees of freedom. However, the degrees of freedom are difficult to derive analytically and how to compute them is still an open problem. What we have done above is to estimate the degrees by using $\tau = 1$ rather than to propose an iterative scheme to update τ. Thus it is enough to start with $\tau = 1$ and update τ only one time.

In Algorithm 3, we give our algorithm that selects the regularization parameter λ automatically. Note that by Theorem 1, both Steps 2 and 4 there are convergent. Hence Algorithm 3 is convergent too.

Algorithm 3 Automatic Regularization Parameter Selection (AutoRegSel) for TV-Based Image Restoration

Function: $(\mathbf{f}, \lambda) = \mathrm{AutoRegSel}(\mathbf{g}, \boldsymbol{H})$.
Input: $\mathbf{g}$, $\boldsymbol{H}$.
1: Estimate σ^2 and set $c_0^2 = n_1 n_2 \sigma^2$.
2: $(\mathbf{f}, \lambda_0) = \text{DP-PDM-TV}(\mathbf{g}, \boldsymbol{H}, c_0^2)$.
3: Compute τ by (32) and set $c^2 = \tau n_1 n_2 \sigma^2$.
4: $(\mathbf{f}, \lambda) = \text{DP-PDM-TV}(\mathbf{g}, \boldsymbol{H}, c^2)$.
5: **return** $\mathbf{f}, \lambda$.

IV. Numerical Results

In this section, experimental results are given to illustrate the performance of our proposed approach. The experiments were performed under Windows 7 and MATLAB v7.8 on a Thinkpad T400s Laptop with an Intel Core(TM)2 Duo P9400 processor and 2GB of RAM. The Blurred Signal-to-Noise Ratio (BSNR) and the Improved Signal-to-Noise Ratio (ISNR) are used to measure the quality of the observed images and the restoration results respectively. They are defined as follows: $\mathrm{BSNR} = 10\log_{10}\left(\|\mathbf{g}\|_2^2/\|\mathbf{n}\|_2^2\right)$ and $\mathrm{ISNR} = 10\log_{10}\left(\|\mathbf{g} - \mathbf{f}_{\mathrm{clean}}\|_2^2/\|\mathbf{f} - \mathbf{f}_{\mathrm{clean}}\|_2^2\right)$, see (1).

A. Experiment 1—The choice of τ

In the first experiment, we illustrate how to choose a suitable upper bound $c^2 = \tau n_1 n_2 \sigma^2$ for the discrepancy principle. In particular, we show that the choice of τ we derive in (32) is a good one. The test images are *Lena* and *Cameraman* images with size 256×256. MATLAB commands `fspecial('Gaussian',[9 9],3)` and `fspecial('average',9)` respectively are used to generate the Gaussian blur and uniform blur used in the experiment. We remark that many previous works for image restoration [5], [9], [42] have reported the ISNR performance for these two blur kernels. To each blurred image, a Gaussian noise is added such that the BSNR of the observed images are 10, 20, 30 and 40dB respectively. The stopping criterion is set to $\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2 / \|\mathbf{f}^{(k)}\|_2 < 10^{-4}$. According to Theorem 1, we fix the parameter $t = 1$ and $s = 1/16$.

The plots in Figure 1 show the changes of ISNR against τ. The ISNR's obtained by the general setting of $\tau = 1$ are marked by "∇", and those obtained by our choice of τ given in (32) are marked by "o". It is clearly seen that for $\tau = 1$, only when the noise is high (i.e. BSNR = 10dB), will the ISNR be close to the maximum. For other noise levels, its ISNR's are far from the maximum. However, the ISNR's obtained by using the τ in our Algorithm are always close to the maximum, and it is especially good when the noise are not high (BSNR $\geq$ 20dB).

The plots of the regularization parameter λ versus the iteration number are shown in Figure 2. We observe that the higher the level of the noise, the smaller the λ is, as one would expected. According to (32), the smaller the λ is, the closer τ is to 1, a fact reflected already in Figure 1. We remark that there is always a jump in λ in Figure 2 because the set $\{\lambda_k\}$ consists two parts: one from Step 2 in Algorithm 3 where we set $\tau = 1$ and another from Step 4 in the same algorithm where we set τ by (32). We observe that λ_k stabilizes within 100 iterations. We also plot the figure of the CPU time (seconds) versus the ISNR in Figure 3. We see that the whole algorithm usually converges within 5 seconds.

B. Experiment 2—Step size t

The convergence proof requires that $st \leq 1/16$. Here we investigate if their values affect the regularization parameter λ. We plot t versus λ in Figure 4 for $t = 0.1k$ with $k = 2, 3, \ldots, 20$ and $s = \frac{1}{16t}$. Though the value of λ is gradually decreasing when t increases, the absolute error between the maximum of λ and the minimum of λ divided by the minimum of λ is less than 4×10^{-3} in the experiment. This means that t has very little influence on λ. Therefore, in all our experiments, we fix the parameters $t = 1$ and $s = 1/16$.

C. Experiment 3—Is our λ good?

We now compare the results obtained by our method (which chooses the regularization parameter λ automatically) with those presented in [5], [9], [42]. Babacan *et al.* [5] considered the Gamma distribution for the hyperpriors of the regularization parameter. Bioucas-Dias *et al.* [9] adopted majorization-minimization approach to estimate the original image and λ which is assumed to follow the Jeffeys' distribution. Liao *et al.* [42] considered splitting the primal variable and then applying the GCV technique to choose λ in each restoration step. For simplicity, we call the approach of Bioucas-Dias *et al.* [9] as BFO, the approach of Babacan *et al.* [5] as BMA and the approach of Liao *et al.* [42] as LLN.

The ISNR values of the *Lena*, *Cameraman* and *Shepp-Logan* images blurred by the kernels in Experiment 1 are shown in Table I. The symbol "–" in the table means that the results are not given in the reference paper. The bold numbers in the table stand for the best ISNR's obtained among all the methods. The results show that the ISNR's of our method are better than the other methods, except in the Lena and Cameraman images when the noise is low (BSNR=40dB). Even in those two cases, our ISNR are close to best results by BMA than the other two methods. The results indicate that our λ's are good.

TABLE I
ISNRs OBTAINED BY DIFFERENT METHODS.

Gaussian blur with variance 9					
BSNR	Image	BFO [9]	BMA [5]	LLN[42]	Proposed
20	Lena	2.99	2.87	2.57	**3.12**
	Cameraman	2.21	1.72	1.82	**2.59**
	Shepp-Logan	4.24	1.85	–	**7.01**
30	Lena	3.82	3.87	4.17	**4.21**
	Cameraman	3.59	2.63	3.43	**4.05**
	Shepp-Logan	7.21	4.31	–	**9.07**
40	Lena	4.41	4.78	5.44	**5.84**
	Cameraman	5.78	3.39	5.02	**6.21**
	Shepp-Logan	10.27	6.69	–	**12.21**
9×9 uniform blur					
BSNR	Image	BFO [9]	BMA [5]	LLN[42]	Proposed
20	Lena	4.05	3.72	3.15	**4.28**
	Cameraman	3.27	2.42	2.88	**3.85**
	Shepp-Logan	6.25	3.01	–	**7.45**
30	Lena	5.43	5.89	4.43	**5.94**
	Cameraman	5.69	5.41	5.57	**5.86**
	Shepp-Logan	10.49	7.77	–	**11.49**
40	Lena	6.22	**8.42**	6.92	8.01
	Cameraman	8.46	**8.57**	7.86	8.46
	Shepp-Logan	16.39	13.69	–	**17.32**

D. Experiment 4—Comparison in speed and accuracy

In this experiment, we compare our algorithm (AutoRegSel) with the current state-of-the-art methods: alternating direction methods presented in [47] (Ng-Weiss-Yuan), FTVd-v4[1] [56], [61], and C-SALSA[2] [1], [28]. The stopping criterion of all the methods is that the relative difference between the successive iterate of the restored image should satisfy $\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 / \|\mathbf{f}^{(k)}\|_2^2 \leq 10^{-6}$ or the number of iterations is larger than 1000. We consider the three image restoration problems in [29]: in the first problem, the point spread function is a 9×9 uniform blur with the noise variance $\sigma^2 = 0.56^2$ (Prob. 1); and in the last two problems, the point spread function is given by $h_{ij} = 1/(1+i^2+j^2)$ for $i, j = -7, \ldots, 7$ with $\sigma^2 = 2$ (Prob. 2) and $\sigma^2 = 8$ (Prob. 3) respectively. The upper bound c^2 for AutoRegSel, Ng-Weiss-Yuan and C-SALSA is set to $c^2 = \tau n_1 n_2 \sigma^2$ where τ is computed by (32).

[1] http://www.caam.rice.edu/∼optimization/L1/ftvd/v4.1/

[2] http://cascais.lx.it.pt/∼mafonso/salsa.html

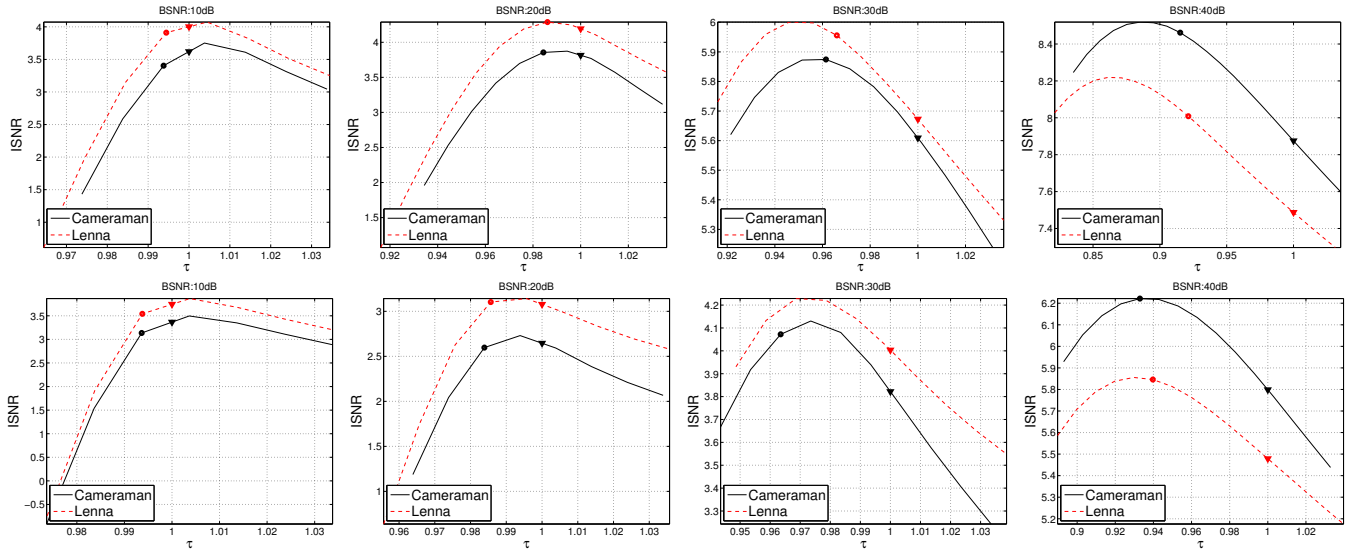

Fig. 1. The parameter τ versus ISNR for cameraman image and Lena image. The images are blurred by a uniform blur of size 9×9 (first row) and a Gaussian blur of size 9×9 with variance 9 (second row). The ISNRs obtained by the proposed method and by $\tau = 1$ are marked by "o" and "∇" respectively.

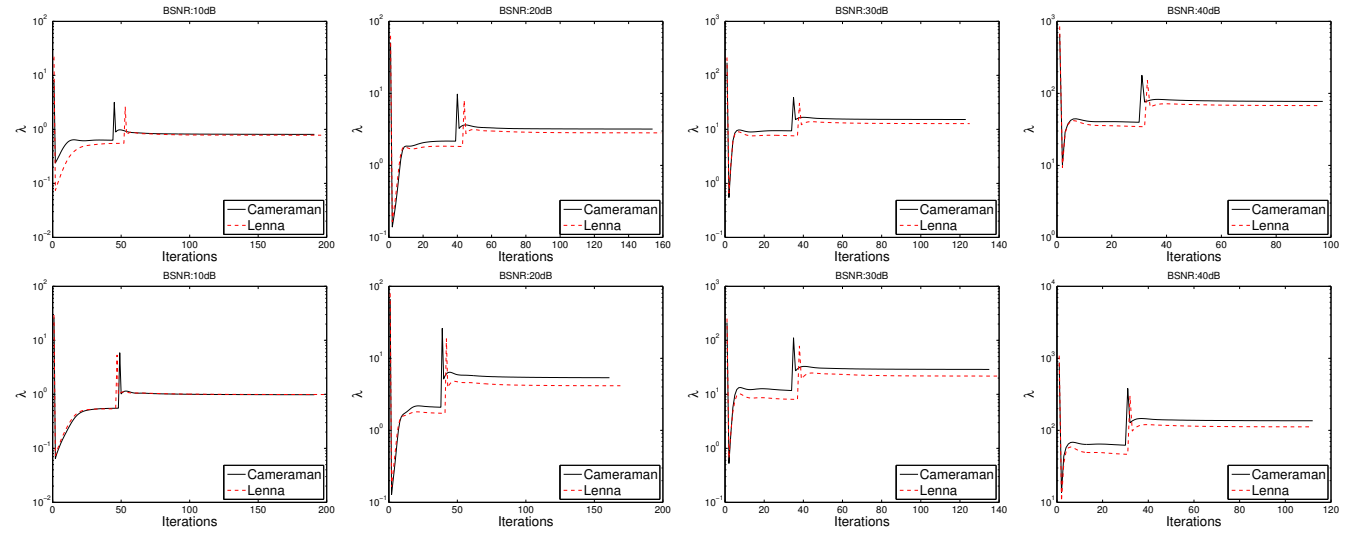

Fig. 2. The regularization parameter versus iterations for cameraman image (solid line) and Lena image (dash line). The images are blurred by a uniform blur of size 9×9 (first row) and a Gaussian blur of size 9×9 with variance 9 (second row).

Since λ should be predetermined in FTVd-v4, we choose the λ from our algorithm as the one in FTVd-v4. The plots of ISNR versus CPU time are shown in Figure 5. Table II shows the ISNR values, the number of iteration to reach convergence and the CPU running times. We emphasize that our method chooses λ automatically, while for method like FTVd-v4, one has to determine λ manually, for example, by running the algorithm many time to determine the best λ by trial-and-error. From the tables and the plots, we observe that our algorithm produces the best ISNRs when compared to those discrepancy principle based methods (i.e., Ng-Weiss-Yuan and C-SALSA).

V. Conclusion

We develop a primal-dual model and the accompanying algorithm for TV image restoration problem. Since the variance of the noise in any given observed image can be estimated easily, Morozov discrepancy principle is applied to find the best regularization parameter λ. During the iteration, λ is automatically updated to converge to the best λ. We gave a convergence proof for the algorithm and the numerical results show that the proposed algorithm performs competitively with the best state-of-the-art methods both in time and accuracy.

VI. Appendix

A. Proof of Lemma 3

Proof: Let $\mathbf{q}_1 = \mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k+1)}$ and $\mathbf{q}_2 = \mathbf{p}^{(k)} - s\nabla \mathbf{f}^{(k)}$. According to the definition of the operators div and ∇, we have $\mathrm{div} = -\nabla^T$. Thus by (21) and (22), we obtain

$$\begin{aligned} & s\langle \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}, \mathrm{div}(\mathbf{p}^{(k+1)} - \mathbf{p}^{(k+\frac{1}{2})}) \rangle_X \\ =& -\langle s\nabla(\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}), \mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2) \rangle_Y \\ =& \langle \mathbf{q}_1 - \mathbf{q}_2, \mathcal{P}_A(\mathbf{q}_1) - \mathcal{P}_A(\mathbf{q}_2) \rangle_Y. \end{aligned}$$

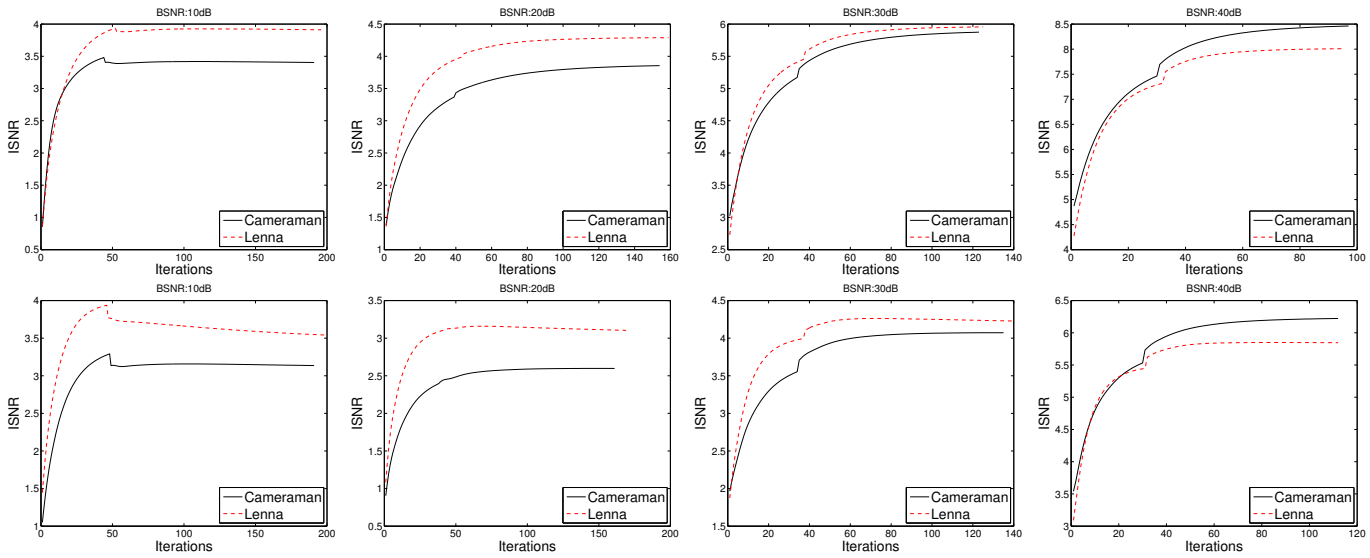

Fig. 3. ISNR versus iterations for cameraman image (solid line) and Lena image (dash line). The images are blurred by a uniform blur of size 9×9 (first row) and a Gaussian blur of size 9×9 with variance 9 (second row).

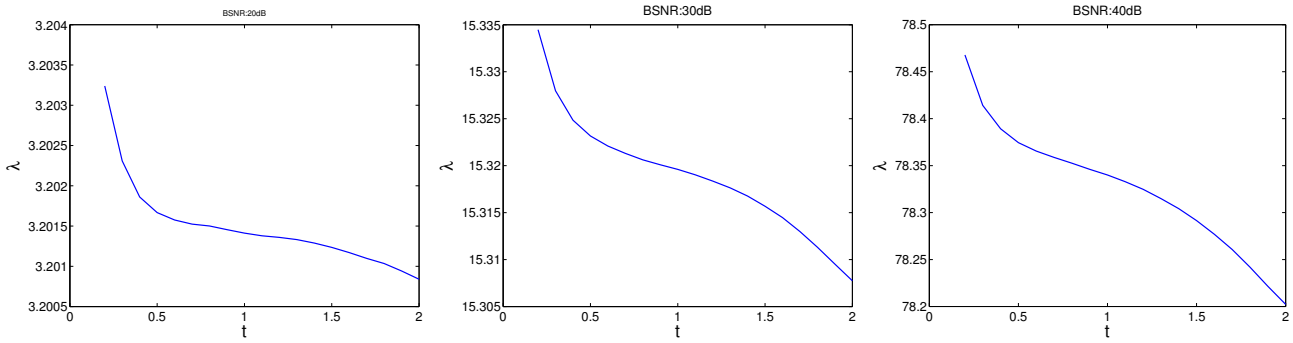

Fig. 4. λ versus the step length t for cameraman image. The images are blurred by a uniform blur of size 9×9 with BSNR=20dB, 30dB, 40dB respectively. The absolute error between the maximum of λ and the minimum of λ divided by the minimum of λ is less than 4×10^{-3}.

Using the classical inequality $2\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2$ for any vectors $\mathbf{x}$ and $\mathbf{y}$, we have

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

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

$$\begin{aligned} \|\nabla \mathbf{f}\|_2^2 &= \sum_{r,s} \left((f_{r+1,s} - f_{r,s})^2 + (f_{r,s+1} - f_{r,s})^2\right) \\ &\leq 2\sum_{r,s} \left(f_{r+1,s}^2 + 2f_{r,s}^2 + f_{r,s+1}^2\right) \leq 8\|\mathbf{f}\|_2^2. \end{aligned}$$

Hence, the result holds. ■

B. Proof of Lemma 4

Proof: Notice that $\psi_k(\cdot; \mathbf{f})$ in (17) and $\phi_k(\cdot; \mathbf{p}, \lambda)$ in (18) are strongly convex functions with respect to $\mathbf{p}$ and $\mathbf{f}$ respectively. Recall that a differentiable function φ is called strongly convex with modulus μ if [50]:

$$\varphi(\mathbf{w}) \geq \varphi(\mathbf{v}) + \nabla\varphi(\mathbf{v})^T(\mathbf{w}-\mathbf{v}) + \frac{\mu}{2}\|\mathbf{w}-\mathbf{v}\|_2^2, \quad \forall \mathbf{v}, \mathbf{w}. \tag{33}$$

A twice differentiable function φ is strongly convex with modulus μ if and only if $(\nabla^2\varphi - \mu \boldsymbol{I})$ is semi-positive definite, where $\nabla^2\varphi$ is the Hessian matrix [52, p268][53].

Consider $\varphi(\mathbf{p}) = \psi_k(\mathbf{p}; \mathbf{f}^{(k)})$. Notice that the Hessian matrix of $\varphi(\cdot)$ at $\mathbf{p}^{(k+\frac{1}{2})}$ is $\frac{1}{s}\boldsymbol{I}$, so $\varphi(\cdot)$ is a strongly convex function with modulus $1/s$. By using Proposition 4.7.2 in [7], $\mathbf{p}^{(k+\frac{1}{2})}$ minimizes $\psi_k(\mathbf{p}^{(k+\frac{1}{2})}; \mathbf{f}^{(k)}, \lambda_k)$ over A if and only if there exists a $\mathbf{d} \in \psi_k(\mathbf{p}^{(k+\frac{1}{2})}; \mathbf{f}^{(k)}, \lambda_k)$ such that

$$\left\langle \mathbf{d}, \mathbf{p} - \mathbf{p}^{(k+\frac{1}{2})} \right\rangle_Y \geq 0, \quad \forall \mathbf{p} \in A. \tag{34}$$

Thus by using (33), we have

$$\begin{aligned} & -\left\langle \mathbf{f}^{(k)}, \mathrm{div}\mathbf{p} \right\rangle_X + \frac{1}{2s}\left\|\mathbf{p} - \mathbf{p}^{(k)}\right\|_2^2 \geq -\left\langle \mathbf{f}^{(k)}, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})} \right\rangle_X \\ & + \frac{1}{2s}\left\|\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k)}\right\|_2^2 + \frac{1}{2s}\left\|\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}\right\|_2^2, \quad \forall \mathbf{p} \in A. \end{aligned}$$

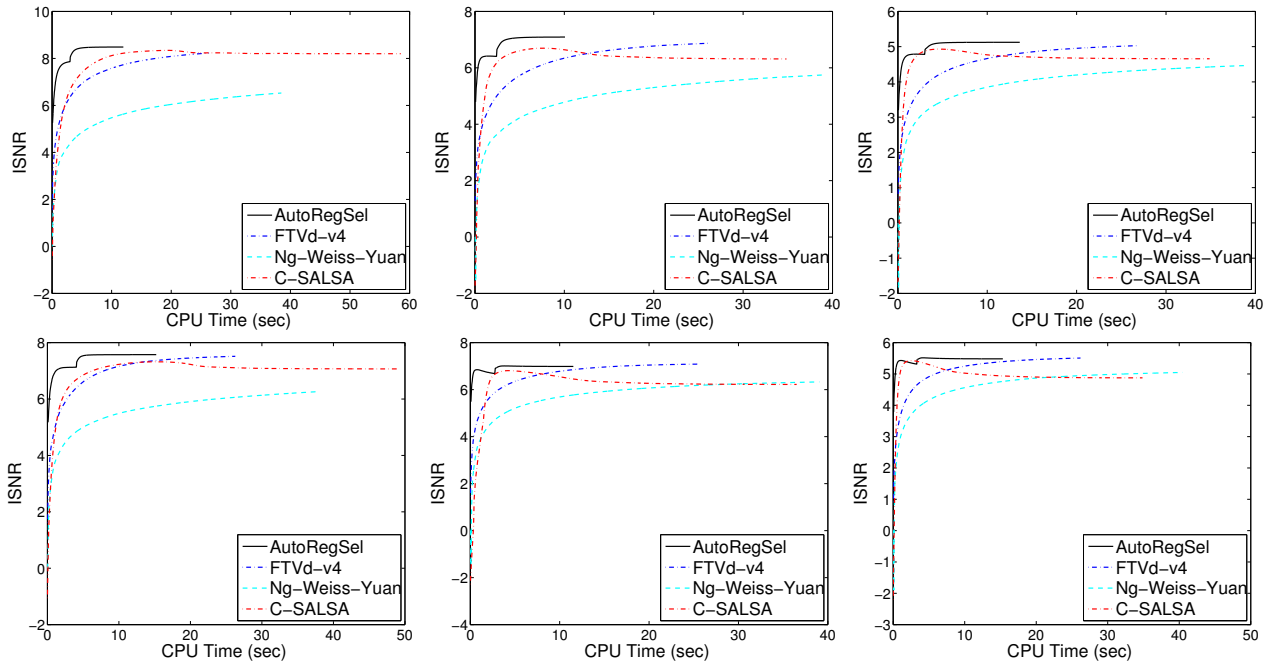

Fig. 5. ISNR versus CPU time (in seconds) for cameraman image (top row) and Lena image (bottom row). The images are blurred by a 9×9 uniform blur with noise variance $\sigma^2 = 0.56^2$ (first column) and a 15×15 point spread function given by $h_{ij} = 1/(1+i^2+j^2)$ with noise variance 2 (second column) and 8 (third column) respectively.

TABLE II
COMPARISON WITH OTHER METHODS. "TIMES" STANDS FOR CPU TIMES IN SECONDS.

		Cameraman			Lena		
Problem	Method	ISNR	Iterations	Times	ISNR	Iterations	Times
Prob.1	AutoRegSel	8.49	399	9.76	7.58	442	11.01
	FTVd-v4 [56], [61]	8.24	1000	19.54	7.52	1000	19.58
	Ng-Weiss-Yuan [47]	6.53	1000	29.02	6.26	1000	29.92
	C-SALSA [1], [28]	8.20	917	45.06	7.07	824	37.40
Prob.2	AutoRegSel	7.10	336	8.34	7.00	391	9.52
	FTVd-v4 [56], [61]	6.89	1000	19.94	7.09	1000	19.82
	Ng-Weiss-Yuan [47]	5.75	1000	29.54	6.33	1000	29.92
	C-SALSA [1], [28]	6.31	586	26.40	6.22	594	27.10
Prob.3	AutoRegSel	5.13	450	11.40	5.49	507	12.77
	FTVd-v4 [56], [61]	5.04	1000	19.88	5.51	1000	20.15
	Ng-Weiss-Yuan [47]	4.46	1000	29.81	5.05	1000	29.26
	C-SALSA [1], [28]	4.66	590	26.54	4.87	602	26.60

Substituting $\mathbf{p} = \mathbf{p}^{(k+1)}$, we obtain

$$\begin{aligned} &\big\langle \mathbf{f}^{(k)}, \mathrm{div}\big(\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k+1)}\big)\big\rangle_X + \frac{1}{2s}\left\|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\right\|_2^2 \\ \geq\ & \frac{1}{2s}\|\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k)}\|_2^2 + \frac{1}{2s}\|\mathbf{p}^{(k+\frac{1}{2})} - \mathbf{p}^{(k+1)}\|_2^2. \end{aligned} \quad (35)$$

Using similar arguments on (16), we have

$$\begin{aligned} &\big\langle \mathbf{f}^{(k+1)}, \mathrm{div}\big(\mathbf{p}^{(k+1)} - \mathbf{p}^{*}\big)\big\rangle_X + \frac{1}{2s}\|\mathbf{p}^{*} - \mathbf{p}^{(k)}\|_2^2 \\ \geq\ & \frac{1}{2s}\|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\|_2^2 + \frac{1}{2s}\|\mathbf{p}^{(k+1)} - \mathbf{p}^{*}\|_2^2. \end{aligned} \quad (36)$$

Next we let $\varphi(\mathbf{f}) = \phi_k(\mathbf{f}; \mathbf{p}^{(k+\frac{1}{2})}, \lambda_{k+1})$, see (18). We have $0 \in \partial\varphi(\mathbf{f}^{(k+1)})$ and $\mathbf{f}^{(k+1)} = \mathrm{argmin}_{\mathbf{f}\in D}\varphi(\mathbf{f})$. Similar to (34), we have $\nabla\phi(\mathbf{f}^{(k+1)})^T\big(\mathbf{f} - \mathbf{f}^{(k+1)}\big) \geq 0$, for all $\mathbf{f} \in D$. Using the fact that $\varphi(\mathbf{f})$ is a strongly convex function with modulus $1/t$ and applying the inequality (33), we obtain

$$\begin{aligned} &\big\langle \mathbf{f}^{*}, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}\big\rangle_X + \frac{\lambda_{k+1}}{2}\big\|\boldsymbol{H}\mathbf{f}^{*} - \mathbf{g}\big\|_2^2 + \frac{1}{2t}\|\mathbf{f}^{*} - \mathbf{f}^{(k)}\|_2^2 \\ \geq &\big\langle \mathbf{f}^{(k+1)}, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}\big\rangle_X + \frac{\lambda_{k+1}}{2}\big\|\boldsymbol{H}\mathbf{f}^{(k+1)} - \mathbf{g}\big\|_2^2 \\ &+ \frac{1}{2t}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 + \frac{1}{2t}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{*}\|_2^2. \end{aligned}$$

From Steps 5–9 of Algorithm 2, we have i) $\mathbf{u}_k \in D$ and $\lambda_{k+1} = 0$ or ii) $\mathbf{u}_k \notin D$, $\lambda_{k+1} > 0$ and $\big\|\boldsymbol{H}\mathbf{f}^{(k+1)} - \mathbf{g}\big\|_2^2 = c^2$. For Case i), we have

$$\begin{aligned} &\big\langle \mathbf{f}^{*} - \mathbf{f}^{(k+1)}, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}\big\rangle_X + \frac{1}{2t}\|\mathbf{f}^{*} - \mathbf{f}^{(k)}\|_2^2 \\ \geq\ & \frac{1}{2t}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_2^2 + \frac{1}{2t}\|\mathbf{f}^{(k+1)} - \mathbf{f}^{*}\|_2^2. \end{aligned} \quad (37)$$

For Case ii), using the fact that $\mathbf{f}^{*} \in D$, we derive that $\frac{\lambda_{k+1}}{2}\big(\big\|\boldsymbol{H}\mathbf{f}^{*} - \mathbf{g}\big\|_2^2 - c^2\big) \leq 0$. Hence the inequality (37) also holds.

Adding the inequalities (30), (35), (35) and (37), we obtain

$$\begin{aligned}&\left\langle \mathbf{f}^*, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}\right\rangle_X + (8s-\frac{1}{2t})\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\|_2^2\\ &\quad+\frac{1}{2t}\|\mathbf{f}^*-\mathbf{f}^{(k)}\|_2^2+\frac{1}{2s}\|\mathbf{p}^*-\mathbf{p}^{(k)}\|_2^2\\ \geq&\left\langle \mathbf{f}^{(k+1)}, \mathrm{div}\mathbf{p}^*\right\rangle_X+\frac{1}{2t}\|\mathbf{f}^{(k+1)}-\mathbf{f}^*\|_2^2+\frac{1}{2s}\|\mathbf{p}^{(k+1)}-\mathbf{p}^*\|_2^2\\ &+\frac{1}{2s}\Big(\|\mathbf{p}^{(k+\frac{1}{2})}-\mathbf{p}^{(k)}\|_2^2+\|\mathbf{p}^{(k+\frac{1}{2})}-\mathbf{p}^{(k+1)}\|_2^2\Big).\end{aligned}$$

Notice that $(\mathbf{f}^*, \mathbf{p}^*)$ is the saddle point of minimax function, therefore

$$\left\langle \mathbf{f}^*, \mathrm{div}\mathbf{p}^{(k+\frac{1}{2})}\right\rangle_X \leq \left\langle \mathbf{f}^*, \mathrm{div}\mathbf{p}^*\right\rangle_X \leq \left\langle \mathbf{f}^{(k+1)}, \mathrm{div}\mathbf{p}^*\right\rangle_X.$$

Then we obtain

$$\begin{aligned}&(16s-\frac{1}{t})\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\|_2^2+\frac{1}{t}\|\mathbf{f}^*-\mathbf{f}^{(k)}\|_2^2+\frac{1}{s}\|\mathbf{p}^*-\mathbf{p}^{(k)}\|_2^2\\ \geq&\frac{1}{t}\|\mathbf{f}^{(k+1)}-\mathbf{f}^*\|_2^2+\frac{1}{s}\|\mathbf{p}^{(k+1)}-\mathbf{p}^*\|_2^2.\end{aligned}$$

Therefore, the inequality (31) holds when $st \leq 1/16$. 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 $(\mathbf{f}^\dagger, \mathbf{p}^\dagger)$.

Since the subsequence $\left(\mathbf{f}^{(k_i)}, \mathbf{p}^{(k_i)}\right)$ converges, for any $\varepsilon > 0$, there exists a constant i_0 such that

$$\frac{1}{t}\left\|\mathbf{f}^{(k_i)}-\mathbf{f}^\dagger\right\|_2^2+\frac{1}{s}\left\|\mathbf{p}^{(k_i)}-\mathbf{p}^\dagger\right\|_2^2<\varepsilon,\quad \forall i>i_0$$

According to (31), for any given $\varepsilon > 0$, there is an k_0 such that

$$\frac{1}{t}\left\|\mathbf{f}^{(k)}-\mathbf{f}^\dagger\right\|_2^2+\frac{1}{s}\left\|\mathbf{p}^{(k)}-\mathbf{p}^\dagger\right\|_2^2<\varepsilon,\quad \forall k>k_0$$

Notice that $\mathbf{u}^\dagger = \mathbf{f}^\dagger - t\mathrm{div}\mathbf{p}^\dagger$. Thus according to Steps 5–9 of Algorithm 2, we have either i) $\mathbf{u}^\dagger \in D$ which implies $\lambda^\dagger = 0$ or ii) $\mathbf{u}^\dagger \notin D$ which implies $\lambda^\dagger > 0$ and $\mathcal{K}(\lambda^\dagger, \mathbf{u}^\dagger) = c^2$. Case i) implies that $\mathbf{f}^\dagger = \mathbf{u}^\dagger \in D$ while Case ii) implies that $\left\|\boldsymbol{H}\mathbf{f}^\dagger(\lambda^\dagger)-\mathbf{g}\right\|_2^2 = \mathcal{K}(\lambda^\dagger, \mathbf{u}^\dagger) = c^2$, *cf* (27). Hence the result holds. ■

ACKNOWLEDGMENT

The first author would like to thank Prof. P. Weiss for providing the source code of [47], and to the authors of FTVd-v4 and C-SALSA for making their source codes freely available.

REFERENCES

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2011.

[2] H. Andrew and B. Hunt. *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[3] J.F. Aujol and G. Gilboa. Constrained and SNR-based solutions for TV-Hilbert space image denoising. *Journal of Mathematical Imaging and Vision*, 26(1):217–237, 2006.

[4] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian blind deconvolution using a total variation prior. *IEEE Trans. Image Process.*, 18(1):12–26, 2009.

[5] S.D. Babacan, R. Molina, and A.K. Katsaggelos. Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Process.*, 17(3):326–339, 2008.

[6] S. Becker, J. Bobin, and E.J. Candes. Nesta: A fast and accurate first-order method for sparse recovery. *SIAM J. Imaging Sci.*, 4(1):1–39, 2011.

[7] D. Bertsekas, A. Nedic, and E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, 2003.

[8] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press Inc., New York, 1982.

[9] J.M. Bioucas-Dias, M.A.T. Figueiredo, and J.P. Oliveira. Adaptive total variation image deconvolution: A majorization-minimization approach. In *Proceedings of the European Signal Processing Conference (EUSIPCO2006)*, 2006.

[10] P. Blomgren and T. Chan. Modular solvers for image restoration problems using the discrepancy principle. *Numer. Linear Algebra Appl.*, 9(5):347–358, 2002.

[11] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Commun. Math. Sci.*, 4(1):179–212, 2006.

[12] D. Calvetti and L. Reichel. Tikhonov regularization of large linear problems. *BIT*, 43(2):263–283, 2003.

[13] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004.

[14] R. Chan and M. Ng. Conjugate gradient methods for Toeplitz systems. *SIAM Rev.*, 38:427–482, 1996.

[15] S.H. Chan, P.E. Gill, and T.Q. Nguyen. An augmented Lagrangian method for total variation image restoration. *submitted to IEEE Transactions on Image Processing*, 2010.

[16] T. Chan and K. Chen. An optimization-based multilevel algorithm for total variation image denoising. *Multiscale Model. Simul.*, 5(2):615–645, 2006.

[17] 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.

[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] K. Chen and X. Tai. A nonlinear multigrid method for total variation minimization from image restoration. *J. Sci. Comput.*, 33(2):115–138, 2007.

[20] P. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.

[21] P. Combettes and J. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6):Article No. 065014, 2008.

[22] J. Darbon and M. Sigelle. A fast and exact algorithm for total variation minimization. *Lecture Notes in Computer Science*, 3522:351–359, 2005.

[23] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation. I. Fast and exact optimization. *J. Math. Imaging Vision*, 26(3):261–276, 2006.

[24] G. Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12):2024–2036, 1989.

[25] F. Dupe, J. Fadili, and J. Starck. A proximal iteration for deconvolving poisson noisy images using sparse representations. *IEEE Trans. Image Process.*, 18(2):310–321, 2009.

[26] J. Eckstein and D. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming, Ser. A*, 55(3):293–318, 1992.

[27] H. Engl and W. Grever. Using the L-curve for determining optimal regularization parameters. *Numer. Math.*, 69(1):25–31, 1994.

[28] M. Figueiredo, J. M. Bioucas-Dias, and M. V. Afonso. Fast frame-based image deconvolution using variable splitting and constrained optimization. In *Proc. IEEE/SP 15th Workshop Statistical Signal Processing SSP '09*, pages 109–112, 2009.

[29] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8):906–916, 2003.

[30] H. Fu, M. Ng, M. Nikolova, and J. Barlow. Efficient minimization methods of mixed l_2-l_1 and l_1-l_1 norms for image restoration. *SIAM J. Sci. Comput.*, 27(6):1881–1902, 2006.

[31] 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.

[32] N. P. Galatsanos and A. K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. Image Process.*, 1(3):322–336, 1992.

[33] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[34] P. Hall and DM Titterington. Common structure of techniques for choosing smoothing parameters in regression problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(2):184–198, 1987.

[35] Martin Hanke. Limitations of the L-curve method in ill-posed problems. *BIT*, 36(2):287–301, 1996.

[36] P. Hansen and D. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, 1993.

[37] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[38] 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.

[39] JT Kent and M. Mohammadzadeh. Global optimization of the generalized cross-validation criterion. *Statistics and Computing*, 10(3):231–236, 2000.

[40] D. Krishnan, P. Lin, and X. Tai. An efficient operator splitting method for noise removal in images. *Commun. Comput. Phys.*, 1:847–858, 2006.

[41] C. Lawson and R. Hanson. *Solving least squares problems*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1974. Prentice-Hall Series in Automatic Computation.

[42] H. Liao, F. Li, and M. Ng. Selection of regularization parameter in total variation image restoration. *J. Opt. Soc. Am. A*, 26(11):2311–2320, Nov 2009.

[43] Y. Lin, B. Wohlberg, and H. Guo. UPRE method for total variation parameter selection. *Signal Processing*, 90(8):2546–2551, AUG 2010.

[44] S. Mallat. *A Wavelet Tour of Signal Processing*. 2nd edition. Academic Press: San Diego, 1999.

[45] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.

[46] M. Ng, R. Chan, and W. Tang. A fast algorithm for deblurring models with Neumann boundary conditions. *SIAM J. Sci. Comput.*, 21(3):851–866, 1999.

[47] M. Ng, P. Weiss, and X. Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *SIAM J. Sci. Comput.*, 32(5):2710–2736, 2010.

[48] 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.

[49] 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(2):460–489, 2005.

[50] B.T. Polyak. Existence theorems and convergence of minimizing sequences in extremum problems with restrictions. *Soviet Math. Dokl.*, 7:72–75, 1966.

[51] L. Reichel and A. Shyshkov. A new zero-finder for Tikhonov regularization. *BIT Numerical Mathematics*, 48:627–643, 2008.

[52] A.W. Roberts and D. E. Varberg. *Convex functions.* Pure and Applied Mathematics, 57. New York-London: Academic Press, a subsidiary of Harcourt Brace Jovanovich, Publishers. XX, 300 , 1973.

[53] R. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

[54] R. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

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

[56] M. Tao, J. Yang, and B. He. Alternating direction algorithms for total variation deconvolution in image reconstruction. *TR0918, Department of Mathmatics, Nanjing University, November*, 2009.

[57] DM Titterington. Choosing the regularization parameter in image restoration. *Lecture Notes-Monograph Series*, 20:392–402, 1991.

[58] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.*, 29(1):119–138, 1991.

[59] C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.

[60] C. R. Vogel. Non-convergence of the L-curve regularization parameter selection method. *Inverse Problems*, 12(4):535–547, 1996.

[61] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.*, 1(3):248–272, 2008.

[62] P. Weiss, L. Blanc-Féraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM J. Sci. Comput.*, 31(3):2047–2080, 2009.

[63] J. Yang, W. Yin, Y. Zhang, and Y. Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sci.*, 2(2):569–592, 2008.

[64] J. Yang, Y. Zhang, and W. Yin. An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise. *SIAM J. Sci. Comput.*, 31(4):2842–2865, 2009.

[65] M. Zhu. *Fast Numerical Algorithms for Total Variation Based Image Restoration*. PhD thesis, University of California, Los Angeles, 2008.

[66] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report 08-34*, 2007.

You-Wei Wen received the M.S. degree in computer science from Guangdong University of Technology in 2001, and the Ph.D degree in applied mathematics from the University of Hong Kong, in 2006. He is now a Full Professor at Kunming University of Science and Technology, P. R. China. His research interests include signal/image processing, scientific computing.

Raymond H. Chan was born in 1958 in Hong Kong. He received his B.Sc. degree in Mathematics from the Chinese University of Hong Kong and his M.Sc. and Ph.D. degree in Applied Mathematics from New York University. He is now a Chair Professor in the Department of Mathematics, The Chinese University of Hong Kong. His research interests include numerical linear algebra and image processing problems.