SIAM J. NUMER. ANAL.
Vol. 50, No. 3, pp. 983–1003

AN ITERATIVE LAGRANGE MULTIPLIER METHOD FOR CONSTRAINED TOTAL-VARIATION-BASED IMAGE DENOISING*

JIANPING ZHANG†, KE CHEN‡, AND BO YU†

Abstract. Various effective algorithms have been proposed in the past two decades for nonlinear PDEs arising from the unconstrained total-variation-based image denoising problem regularizing the total variation constrained minimization model. Such algorithms can be used to obtain a satisfactory result as long as a suitable regularization parameter balancing the trade-off between a good fit to the data and a regular solution is given. However, it is generally difficult to obtain a suitable regularization parameter without which restored images can be unsatisfactory: if it is too large, then the resulting solution is still contaminated by noise, while if too small, the solution is a poor approximation of the true noise-free solution. To provide an automatic method for the regularization parameter when the noise level is known a priori, one way is to address the coupled Karush–Kuhn–Tucker (KKT) systems from the constrained total variation optimization problem. So far much less work has been done on this problem. This paper presents an iterative update algorithm for a Lagrange multiplier to solve the KKT conditions, and our proposed method can adaptively deal with noisy images with different variances σ^2. Numerical experiments show that our model can effectively find a highly accurate solution and produce excellent restoration results in terms of image quality.

Key words. constrained optimization, image denoising, total variation, partial differential equations, Lagrange multiplier

AMS subject classifications. 62H35, 65N22, 65N55, 74G65, 74G75

DOI. 10.1137/110829209

1. Introduction.

In the image formation process (e.g., recording, transmission), the generation of noise in an observed image is usually unavoidable. In many important cases, high noise levels can add much difficulty to further processing tasks such as automatic recognition of important features. Therefore, effective denoising plays an important role in the processing, and hence understanding of noisy images becomes a fundamental task of computer vision.

The goal of denoising is to remove noise or spurious details from a given, possibly corrupted, digital picture while maintaining essential features such as edges. Approaches to image denoising have been developed along three main directions: (i) wavelet-based methods, for example, methods based on wavelet thresholding introduced by Donoho and Johnstone in [28]; (ii) stochastic or statistical methods—a prominent example is the Markov random field approach introduced by Geman and Geman [31]; and (iii) PDE or variational approaches [2, 49]. In the past two decades, the third direction, based on the calculus of variations and PDEs, has seen great success [2, 22, 25, 24, 40, 45, 46]. In particular, it is observed that this approach reduces oscillations near discontinuities. One important reason for their success is that these models are particularly well suited to imposing geometric constraints (such as regular-

*Received by the editors March 31, 2011; accepted for publication (in revised form) December 22, 2011; published electronically May 3, 2012. This research was supported by the National Natural Science Foundation of China (11171051) and the Fundamental Research Funds for the Central Universities.

http://www.siam.org/journals/sinum/50-3/82920.html

†School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning 116024, People's Republic of China (j_p_zhang@sina.com, http://math.dlut.edu.cn/).

‡Corresponding authors. Centre for Mathematical Imaging Techniques and Department of Mathematical Sciences, University of Liverpool, Peach Street, Liverpool L69 7ZL, United Kingdom (k.chen@liverpool.ac.uk, www.liv.ac.uk/www/cmit).

ity) on the solutions sought. Among the best known and most influential examples is the Rudin–Osher–Fatemi (ROF) total-variation-based image denoising model [2, 40]. Denoising models such as the ROF model can be easily adapted to different ill-posed problem situations [3, 29] and hence have attracted a great deal of attentions in the image processing community.

In this paper, we focus our attention on the third direction, i.e., the PDE-based variational formulation. We are mainly concerned with the popular denoising technique with total variation (TV), introduced in [40]. The TV seminorm does not penalize discontinuities in the image and thus allows us to recover the edges of the original image. The idea of our study may be generalized to other models.

Let $z(x,y) : \Omega \mapsto R^+$ be a given image containing an unknown additive noise η, where Ω is a convex, bounded region of R^2 and u is a true image such that

$$z(x,y) = u(x,y) + \eta(x,y). \tag{1.1}$$

For this inverse problem (1.1), there exist many models restoring u [46, 24]. The ROF model [40] takes the form

$$\min_u TV(u) := \int_\Omega |\nabla u| \, dxdy \quad \text{s.t.} \quad F(u) := \frac{1}{2}(\|u-z\|^2_{L^2(\Omega)} - \sigma^2) = 0, \tag{1.2}$$

where we assume that the $L^2(\Omega)$-norm of the noise data defined by

$$\sigma = \Big(\int_\Omega |z(x,y) - u(x,y)|^2 \, dxdy \Big)^{\frac{1}{2}} \tag{1.3}$$

is known a priori. This problem has the form of a constrained optimization problem with a TV regularization seminorm as the objective function and the equality (1.3) as the constraint. Since the objective $TV(u)$ is not differentiable when $|\nabla u| = 0$, it is common to consider a modified minimization problem of the form (1.2)

$$\min_u TV_\epsilon(u) := \int_\Omega \sqrt{|\nabla u|^2 + \epsilon} \, dxdy \quad \text{s.t.} \quad F(u) = 0, \tag{1.4}$$

where $\epsilon > 0$ is a small parameter. To study the solution u, we need the space of functions of bounded variation on Ω defined by

$$BV(\Omega) = \Big\{ u \in L^1(\Omega) : TV(u) < +\infty \Big\}. \tag{1.5}$$

Clearly, $BV(\Omega)$ is a Banach space with respect to BV-norm: $\|u\|_{BV} = \|u\|_{L^1(\Omega)} + TV(u)$. We also have $W^{1,1}(\Omega) \subset BV(\Omega) \subset L^p(\Omega) \subset L^1(\Omega)$ for $1 \leqslant p \leqslant d/(d-1)$, where d denotes the dimension ($d = 2$ in this work) [1].

Denoting the Lagrangian functional for the constrained problem (1.4) by

$$E_\epsilon(\lambda, u) := \int_\Omega \sqrt{|\nabla u|^2 + \epsilon} \, dxdy + \frac{\lambda}{2}\Big(\int_\Omega |u-z|^2 \, dxdy - \sigma^2 \Big), \tag{1.6}$$

where λ is a Lagrange multiplier, the unconstrained problem can be written as

$$\min_{\lambda, u \in BV(\Omega)} E_\epsilon(\lambda, u). \tag{1.7}$$

Using the Gâteaux-differential we derive the following first-order necessary conditions, i.e., KKT conditions [38]:

$$L_\lambda(u) := -\nabla \cdot \Big(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon}} \Big) + \lambda(u - z) = 0 \tag{1.8a}$$

$$F(u) := \frac{1}{2}(\|u - z\|^2_{L^2(\Omega)} - \sigma^2) = 0 \tag{1.8b}$$

with

$$u \in \mathfrak{U} := \left\{ u(x,y) : \ (x,y) \in \Omega \text{ and } \frac{\partial u(x,y)}{\partial \vec{n}} = 0 \text{ on } \partial\Omega \right\}, \tag{1.9}$$

where $\vec{n}$ is the unit exterior normal to the boundary $\partial\Omega$.

We remark that many researchers have successfully solved the above model by fixing λ first. In particular, if λ is given, various effective algorithms can solve the nonlinear PDEs (1.8a) with the boundary condition from (1.9) [2, 14, 16, 22, 23, 37, 41, 45].

In works of [4, 38, 40], some discussion is given on attempts beyond the generic method of L-curve methods from inverse formulation to predict an optimal parameter. In general such a solution does not satisfy (1.8b), i.e., the constrained condition $F(u) = 0$; this means that the solution of the Euler–Lagrange equation (1.8a) with the fixed parameter λ is not a solution of the KKT system (1.8). Furthermore, such a regularization parameter λ balancing the trade-off between a good fit to the data and a regular solution is difficult to give, and the crucial issue in the unconstrained regularization (1.6) is how to choose such parameter λ in order to produce a regularized solution u^* close to the true noise-free solution u, satisfying the KKT system. If the value λ is too large, then the corresponding solution u^* is contaminated by noise, while if too small, u^* is a poor approximation of u. Much less work has been done to solve directly the coupled KKT system (1.8); if the system could be solved, the parameter choice for λ is resolved.

According to the discrepancy principle, a regularization parameter should be chosen so that the corresponding solution u^* has a residual L^2-norm equal to a known σ. Thus the KKT method provides only one (implicit) way of realizing the discrepancy principle. There exist some interesting nonlinear scale space methods in multiresolution framework [10, 39, 42] for choosing the parameter explicitly. Specifically, the scale space methods reconstruct an image from an noisy image by a time-continuous nonlinear diffusion equation, starting with z and gradually smoothing and evolving it toward the mean value image, while the inverse scale space methods start with the constant image $u(x, 0) = 0$ and evolve until the noisy image z is reached as time t increases. So the crucial point is to know when iterations should be stopped before convergence; if both methods are stopped at a suitable time t_* which satisfies the discrepancy principle $\|u(t_*) - z\|_{L^2} = \sigma$, we may expect that the approximating noise-free image is obtained, e.g., if the inverse scale space flow is stopped at t_*, then $u(t_*)$ approaches the noise-free image in the sense of Bregman distance (see [10]).

Our key goal in this paper is to investigate the KKT system and choose effectively the regularization parameter λ through solving the system. In such a KKT strategy we shall achieve the discrepancy principle implicitly. Our study will first establish the monotonicity property of the constrained condition with respect to the multiplier parameter, hence making sure that the discrepancy principle seeks the highly approximation solution of the Lagrange equation (1.8a) for different λ to meet with the KKT system (1.8) with the boundary condition (1.9). Further, a full multigrid method with Krylov space acceleration is employed to solve large-size discretization equations (1.8a).

The rest of the paper is organized as follows. In section 2 we review some methods for solving the ROF model as an unconstrained problem or as a PDE problem, and survey how to choose the corresponding regularization parameter. In section 3,

we build some mathematical theories to iteratively seek the solution of the KKT system from the constrained image denoising problem (or ROF model) by using several theorems, before proposing our multiplier iterative update algorithm. We give experimental results in section 4. Finally we conclude this work in section 5.

2. Review of existing algorithms. As noted, various ways of solving the ROF model have been proposed. Below we review a few of them.

2.1. Methods for a given parameter λ. The ROF model may be solved by three types of approaches: the parabolic approach, the elliptic approach, and the optimization approach.

The parabolic approach: Time marching method. In 1992, Rudin, Osher, and Fatemi [40] used this classic parabolic equation with time (the gradient descent method) to solve the above Euler–Lagrange equation from the constrained TV model, i.e.,

$$\frac{\partial u(\cdot,t)}{\partial t} = N(u) = \nabla\cdot\left(\frac{\nabla u(\cdot,t)}{\sqrt{|\nabla u(\cdot,t)|^2+\epsilon}}\right) - \lambda(u(\cdot,t)-u(\cdot,0)) \text{ for } t>0, \tag{2.1a}$$

$$u(x,y,0) = z(x,y), \quad (x,y)\in\Omega, \quad u(.,t)\in\mathfrak{U}\times R^+. \tag{2.1b}$$

In [37], (2.1) is refined by $\frac{\partial u(\cdot,t)}{\partial t} = |\nabla u| N(u)$. Instead of solving the steady flow directly, the equations are advanced in time until a steady state solution is achieved.

Because of the stability constraint of the numerical method, the time step Δt used in the time marching methods shall be restricted, i.e., Courant–Friedrichs–Lewy (CFL) stability condition: $\Delta t < ch^2$ and c is the CFL number (explicit methods generally require $c<1$, but implicit methods allow larger time steps); h is the grid size.

Nevertheless, it is well known that due to the small discretized scale h for high-accuracy simulation, only very small time step Δt is allowed and leads to slow evolution toward the steady-state solution.

Lagged diffusivity fixed point iteration. First Newton's linearization technique for (1.8a) was also considered by Vogel and Oman [45],

$$\begin{aligned} L'_\lambda(u^n)(u^{n+1}-u^n) :=& -\nabla\cdot\left(\frac{\nabla(u^{n+1}-u^n)}{\sqrt{|\nabla u^n|^2+\epsilon}}\right) \\ &+\nabla\cdot\left(\frac{\left(\nabla(u^{n+1}-u^n)\right)^T\nabla u^n}{(\sqrt{|\nabla u^n|^2+\epsilon})^3}\nabla u^n\right) + \lambda(u^{n+1}-u^n) \\ =& -L_\lambda(u^n), \end{aligned}$$

which may be simplified as

$$\lambda u^{n+1} - \nabla\cdot\left(\frac{\nabla u^{n+1}}{\sqrt{|\nabla u^n|^2+\epsilon}}\right) + \nabla\cdot\left(\frac{\left(\nabla(u^{n+1}-u^n)\right)^T\nabla u^n}{(\sqrt{|\nabla u^n|^2+\epsilon})^3}\nabla u^n\right) = \lambda z.$$

From the numerical results in [45], we can note that the above Newton method is divergent for even mildly small ϵ because it has a very small domain of convergence when ϵ is small and a very good initial guess should be chosen in order for the method to converge (also see [25]); note that a homotopy algorithm may be used to accelerate the Newton method [48].

To alleviate this drawback, Vogel and Oman [45] dropped the highest-order term in $L'_\lambda(u)\delta u$ from differentiating $L_\lambda(u)$ and proposed the use of an equivalent quasi-

Newton iterative technique to solve the nonlinear equation (1.8a), resulting in the commonly used lagged diffusivity fixed point algorithm, i.e.,

$$\lambda u^{n+1} - \nabla \cdot \left(\frac{\nabla u^{n+1}}{\sqrt{|\nabla u^n|^2 + \epsilon}} \right) = \lambda z. \tag{2.2}$$

A proof of the global convergence of the fixed point algorithm is given in [23], which gives an estimate for the rate of convergence in terms of the spectral structure of a preconditioned Hessian.

Primal-dual Newton method. From the above discussion, a main difficulty in solving (1.8a) by Newton's method is due to the high nonlinearity when ϵ is small, leading to a small domain of convergence, so a good initial guess for Newton's method with second-order convergence is necessary (which is practically impossible). To deal with this high nonlinearity, Chan, Golub, and Mulet [22] proposed a primal-dual Newton method to reduce the nonlinearity of the objective function before applying a linearization technique such as Newton's method. This technique is accomplished by introducing an additional variable for the flux quantity appearing in the gradient of the objective function, which can be interpreted as the unit normal to the level sets of the image function:

$$\begin{cases} g(u,\omega) := -\nabla \cdot \omega + \lambda(u - z) = 0, \\ f(u,\omega) := \omega \sqrt{|\nabla u|^2 + \epsilon} - \nabla u = 0. \end{cases} \tag{2.3}$$

We can linearize this (u, ω) system (2.3) by Newton's method,

$$\begin{pmatrix} \lambda I & -\nabla\cdot \\ -(I - \frac{\omega \nabla u^T}{|\nabla u|})\nabla & |\nabla u| \end{pmatrix} \begin{pmatrix} \delta u \\ \delta \omega \end{pmatrix} = - \begin{pmatrix} g(u,\omega) \\ f(u,\omega) \end{pmatrix}.$$

From the experimental results in [22], we can know that the new primal-dual method is globally convergent, whereas the primal Newton method has a small domain of convergence [25, 17]. We remark that the method can also be used to speed up the solution of the dual problem of the TV model [18].

Alternating split-Bregman methods. Although the minimization problem of the ROF model can be solved directly [19, 20], much simpler algorithms can be derived from introducing intermediate variables, especially for image deblurring problems. Instead of solving the deblurring problem

$$\min_u B_\epsilon(\lambda, u) := \int_\Omega \sqrt{|\nabla u|^2 + \epsilon}\, dxdy + \frac{\lambda}{2}\Big(\int_\Omega |Ku - z|^2\, dxdy - \sigma^2 \Big),$$

where K is a blur operator, an alternating split-Bregman method can be written as

$$\begin{cases} u = \arg\min_u \dfrac{\mu}{2}\|Ku - z + c\|_2^2 + \dfrac{\lambda}{2}\|Du - d + b\|_2^2, \\ d = \arg\min_d \|d\|_1 + \dfrac{\lambda}{2}\|d - Du - b\|^2, \\ b = b + \delta_b(Du - d), \\ c = c + \delta_c(Ku - z), \end{cases}$$

where $\mu > 0$, $0 < \delta_b \le 1$, and $0 < \delta_c < 2$. Here $Du = TV(u)$ or can be any other regularizer. Even when $K = I$, this formulation has advantages since, clearly, each subproblem is much simpler than before [32, 33, 11, 47].

Duality-based methods. Rather than solving the primal unconstrained minimization problem (1.7), i.e., $\min_u P(u) := \int_\Omega |\nabla u|\, dxdy + \frac{\lambda}{2}\int_\Omega |u-z|^2\, dxdy$, with the definition of the Legendre–Fenchel transform of $TV(u)$: $TV^*(v) = \sup_u \langle u, v\rangle - TV(u)$, Zhu, Wright, and Chan [51] have proposed several gradient projection algorithms that make use of the following dual formulation:

$$\min_{p\in C_0^1(\Omega),|p|\leqslant 1} D(p) := \frac{1}{2}\|\nabla\cdot p + \lambda z\|_2^2.$$

Similar derivations of the above dual formulation are presented in [12, 14, 18, 26]. In particular, a further regularization of p may be added to the formulation.

Multigrid methods. Nonlinear multigrid methods [30, 41, 26] based on the full approximation scheme and the Krylov space accelerated technique have been proposed to solve the Euler–Lagrange equation (1.8a), with a smoother using local adapted Gauss–Seidel relaxation to the linear equation at each fixed point iteration. To apply them directly to solve (1.7), refer to [19, 20, 16]. The goal of a multigrid method is to accelerate the convergence of a basic iterative method by global correction, accomplished by solving recursively a coarse problem [5, 44]; application of multigrid methods to other related models may be found in [6, 7, 8, 9]. In our numerical experiments, a multigrid method is used to solve the nonlinear equation (1.8a) due to its low computational cost.

2.2. Methods for updating the parameter λ. Needless to say, with a fixed regularization parameter λ, a solution from the above proposed methods to the ROF model or the nonlinear PDE (1.8a) violates the constrained condition (1.8b), i.e., it is not a zero point of the KKT system (1.8). Although standard techniques solving the nonlinear equations do not work well for this kind of mixed differential-integral system, there exist previous work on ways of updating the parameter λ.

The original ROF model [40] presented a Lagrange multiplier updating technique which merely multiplies (2.1a) by $(u-z)$ and integrates by parts over Ω. If a steady state of (1.8a) has been reached, the left side of (2.1a) vanishes. Then the Lagrange multiplier becomes

$$\lambda = -\frac{1}{\sigma^2}\int_\Omega \left[\frac{(\nabla u)^T\nabla(u-z)}{\sqrt{|\nabla u|^2+\epsilon}}\right] dxdy. \tag{2.4}$$

In fact, we note that the solution of system (1.8a) and (2.4) cannot solve the system (1.8a) and (1.8b); for example, the mean value image satisfies the former but not the latter.

A better method of updating λ is shown in [34], where it is assumed that the two-dimensional image u and the two-dimensional TV gradient matrix $V = \nabla u$ with $\nabla u_{i,j} = (\nabla_x u_{i,j}, \nabla_y u_{i,j})$ of size $n\times n$ are converted rowwise to sized n^2 vectors. Then the Lagrange multiplier λ of the constrained optimization problem is computed from solving $\min_u TV_\epsilon(u) = \frac{1}{n^2}\sum_{i,j=1}^n \sqrt{|\nabla u_{i,j}|^2+\epsilon}$ with $F(u) = 0$ leading to the initial choice of $\lambda = \frac{(F(u)-\nabla_u TV_\epsilon(u))^T\nabla_u F(u)}{(\nabla_u F(u))^T\nabla_u F(u)}$. In fact, the main method of [34, 35] is to solve the coupled KKT system by a quasi-Newton method with the initial guess for λ from the above formula; note that Newton-type methods may have convergence difficulties unless ϵ is large. This method is also used in [36, 38, 43] to update the Lagrange multiplier of the KKT system.

3. Iterative Lagrange multiplier method. Solving the coupled nonlinear system (1.8) consisting of a differential equation and an integral equation is a difficult task by using some classical iterative methods, for example, a Newton type linearization technique, the gradient descent method with time component, or a fixed point iteration. Therefore the dilemma is that one needs to solve the coupled KKT system to obtain a reliable regularization parameter λ (for the discrepancy principle), while such an effective solution method is hard to find.

We now turn to an alternative way of seeking the solution of this system that is guaranteed to yield a minimum of the constrained minimization problem (1.4). Our idea relies on establishing the analytical properties of the underlying functional. Then a computational procedure becomes apparent. Below we analyze some properties of the Lagrangian functional $E_\epsilon(\lambda, u)$ from (1.6) (refer to [1, 13] for other derivations of existence and uniqueness theories of the objective functional) and the constrained functional $F(u)$ before introducing our proposed method.

3.1. Convexity of the Lagrange functional $E_\epsilon(\lambda, u)$ and solution uniqueness. First, for a fixed Lagrange multiplier λ_0 and all $v \in BV(\Omega)$, the functional $E_\epsilon(\lambda_0, u)$ is G-differentiable or Gâteaux-differentiable at any $u \in BV(\Omega)$ and we denote this first-order G-differential in the direction v by $E'_\epsilon(\lambda_0, u)v$:

$$
\begin{aligned}
E'_\epsilon(\lambda_0, u)v &= \lim_{t\to 0} \frac{E_\epsilon(\lambda_0, u+tv) - E_\epsilon(\lambda_0, u)}{t} \\
&= \lim_{t\to 0} \int_\Omega \left(\frac{\nabla(u+tv)}{\sqrt{|\nabla(u+tv)|^2+\epsilon}} \right)^T \nabla v + \lambda_0(u-z)v \, dxdy \qquad (3.1)\\
&= \int_\Omega \left[-\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\epsilon}} \right) + \lambda_0(u-z) \right] v \, dxdy \\
&\equiv \langle L_{\lambda_0}(u), v \rangle,
\end{aligned}
$$

where the third equality holds due to satisfying the boundary condition $\frac{\partial u}{\partial \vec{n}} = 0$ on $\partial\Omega$. The continuous linear functional $E'_\epsilon(\lambda_0, u)$ or $L_{\lambda_0}(u)$ is called the G-derivative (of $E_\epsilon(\lambda_0, u)$) at $u(x) \in BV(\Omega)$.

Similarly, for any $v \in BV(\Omega)$ the function $\langle L_{\lambda_0}(u), v\rangle$ also is G-differentiable at any $u \in BV(\Omega)$ in the direction $\omega \in BV(\Omega)$. Also $E_\epsilon(\lambda_0, u)$ is second-order G-differentiable since the second-order G-derivative can be derived as follows:

$$
\begin{aligned}
E''_\epsilon(\lambda_0, u)v\omega &= \lim_{t\to 0} \frac{\langle L_{\lambda_0}(u+t\omega), v\rangle - \langle L_{\lambda_0}(u), v\rangle}{t} \\
&= \lim_{t\to 0} \int_\Omega \frac{\left(\frac{\nabla(u+t\omega)}{\sqrt{|\nabla u+t\omega)|^2+\epsilon}} - \frac{\nabla u}{\sqrt{|\nabla u|^2+\epsilon}} \right)^T \nabla v + t\lambda_0 \omega v}{t} \, dxdy \qquad (3.2)\\
&= \int_\Omega (\nabla\omega)^T \left(\frac{(\nabla u)^T \nabla u - \nabla u (\nabla u)^T}{(\sqrt{|\nabla u|^2+\epsilon})^3} + \frac{\epsilon}{(\sqrt{|\nabla u|^2+\epsilon})^3} \right) \nabla v + \lambda_0 \omega v \, dxdy \\
&\equiv J_{\lambda_0}(u)(v, \omega).
\end{aligned}
$$

Below we denote the second-order G-differential of objective functional $E_\epsilon(\lambda_0, u)$ in two directions v and ω by $E''_\epsilon(\lambda_0, u)v\omega$ or $J_{\lambda_0}(u)(v, \omega)$. We know that the eigenvalues of $(\nabla u)^T \nabla u - \nabla u (\nabla u)^T$ are 0 and $(\nabla u)^T \nabla u$, so the eigenvalues of $(\frac{(\nabla u)^T \nabla u - \nabla u (\nabla u)^T}{(\sqrt{|\nabla u|^2+\epsilon})^3} +$

$\frac{\epsilon}{(\sqrt{|\nabla u|^2+\epsilon})^3}$) are positive for all $u(x) \in BV(\Omega)$. Together with its symmetry, the inequality

$$J_{\lambda_0}(u)(p,p) \geqslant \lambda_0 \|p\|^2_{L^2(\Omega)} \text{ for any } u,\ p \in BV(\Omega)$$

is obtained and the above equality holds when $\|\nabla p\|_2 = 0$.

Therefore the basic convexity of the unconstrained optimization problem $E(\lambda_0, u)$ with the fixed Lagrange multiplier λ_0 is established as stated in the following theorem.

THEOREM 3.1. *If λ_0 is a fixed nonnegative constant, then $J_{\lambda_0}(u)(\cdot\ ,\cdot)$ is symmetric positive semidefinite and $E(\lambda_0, \cdot)$ is a convex functional with respect to $u(x)$, where $x \in \Omega$. Furthermore, if $\lambda_0 > 0$, then $J_{\lambda_0}(u)(\cdot\ ,\cdot)$ is uniformly positive definite and $E(\lambda_0, \cdot)$ is a strictly convex functional with respect to $u(x)$, where $x \in \Omega$.*

Proof. Since $J_{\lambda_0}(u)(p,p) \geqslant \lambda_0 \|p\|^2_{L^2(\Omega)} > 0$ for all $u, p \in BV(\Omega)$, $p \neq 0$, and $\lambda_0 > 0$, the strict convexity of $E(\lambda_0, \cdot)$ follows easily from [49, Corollary 42.8]. (Also see [1, 13] for other derivations of convexity of the objective functional.) ∎

In the following we discuss further properties of the minimization problem for the Lagrange functional $E_\epsilon(\lambda_0, u)$ with a given multiplier λ_0, i.e., the relation between the minimization problem $\min_u E_\epsilon(\lambda_0, u)$ and the equation $L_{\lambda_0}(u) = 0$. First we state a known result.

THEOREM 3.2 (see [49, Proposition 42.10]). *If $f : X \to \mathbb{R}$ is a convex G-differentiable functional on the real Banach space X and $f'(u)$ is its G-derivative, then*

$$f(u) \text{ has a minimum at } u \Leftrightarrow f'(u) = 0.$$

Proof. Refer to [49, Proposition 42.10]. ∎

From (3.1), (3.2), and Theorems 3.1 and 3.2, we conclude that $E(\lambda_0, u)$ is a convex G-differentiable functional on the real Banach space $BV(\Omega)$ and $E(\lambda_0, u)$ has a minimum at $u \Leftrightarrow L_{\lambda_0}(u) = 0$, so the solution of the G-derivative equation $L_{\lambda_0}(u) = 0$ is a minimizer of $E(\lambda_0, u)$.

Finally after convexity, the uniqueness of solution can be stated.

THEOREM 3.3 (see [49, Theorem 47.C]). *Suppose that the following conditions hold:*

(1) *$BV(\Omega)$ is a convex nonempty subset of the real locally convex space X;*
(2) *$f : BV(\Omega) \subseteq X \to \mathbb{R}$ is convex; $f(v)$ is extended to X by setting $f(v) = +\infty$ for $v \notin BV(\Omega)$.*

Then the structure of the solution set U for minimization problem $\min_{u \in BV(\Omega)} f(u)$ has these properties:

(i) *U is convex.*
(ii) *U is closed when $f(u)$ is lower semicontinuous and $BV(\Omega)$ is closed.*
(iii) *Every local minimum of $f(u)$ on $BV(\Omega)$ is also a global minimum of $f(u)$ on $BV(\Omega)$.*

Consequently, $f(u)$ has at most one minimizer when $f(u)$ is strictly convex on $BV(\Omega)$.

Proof. Refer to [49, Theorem 47.C]. ∎

According to Theorem 3.3, every local minimum of $E_\epsilon(\lambda_0, u)$ on $BV(\Omega)$ is also a global minimum of $E_\epsilon(\lambda_0, u)$ on $BV(\Omega)$ in the case of Lagrange multiplier $\lambda_0 \geqslant 0$. Furthermore, if $\lambda_0 > 0$, $E_\epsilon(\lambda_0, u)$ is strictly convex on convex set $BV(\Omega)$, so a solution of $L_{\lambda_0}(u) = 0$ is a unique global minimizer of the unconstrained objective functional; existence and uniqueness theories of the TV problem can be found in [1, 13].

3.2. Monotonicity of the constrained functional $F(u_\lambda)$. A crucial step in our study is to establish monotonicity of the constrained functional $F(u_\lambda)$, with which useful algorithms can be derived next.

THEOREM 3.4. *Let λ_1 and λ_2 be nonnegative constants, u_{λ_1} be the solution of $L_{\lambda_1}(u) = 0$, and u_{λ_2} be the solution of $L_{\lambda_2}(u) = 0$. Then*

$$(1)\ \ E_\epsilon(\lambda_1, u_{\lambda_1}) \leqslant E_\epsilon(\lambda_1, u_{\lambda_2}); \qquad (2)\ \ E_\epsilon(\lambda_2, u_{\lambda_2}) \leqslant E_\epsilon(\lambda_2, u_{\lambda_1}).$$

Furthermore, if $\lambda_1 < \lambda_2$, we have

$$(3)\ \ F(u_{\lambda_2}) \leqslant F(u_{\lambda_1}); \qquad (4)\ \ TV_\epsilon(u_{\lambda_1}) \leqslant TV_\epsilon(u_{\lambda_2}).$$

In addition, if $0 < \lambda_1 < \lambda_2$, all $\leqslant$ can be replaced by $<$.

Proof. Since u_{λ_1} is the solution of $L_{\lambda_1}(u) = 0$ and $\lambda_1 \geqslant 0$, then functional $E_\epsilon(\lambda_1, u)$ is convex and u_{λ_1} is a minimizer of $E_\epsilon(\lambda_1, u)$ from Theorem 3.2. According to Theorem 3.3, u_{λ_1} is also a global minimizer of $E_\epsilon(\lambda_1, u)$, so inequality $E_\epsilon(\lambda_1, u_{\lambda_1}) \leqslant E_\epsilon(\lambda_1, u_{\lambda_2})$ holds. Similarly (2) also holds.

We rewrite (1) and (2) as the following:

$$\begin{aligned} &TV_\epsilon(u_{\lambda_1}) + \lambda_1 F(u_{\lambda_1}) \leqslant TV_\epsilon(u_{\lambda_2}) + \lambda_1 F(u_{\lambda_2}),\\ &\text{i.e., } TV_\epsilon(u_{\lambda_1}) - TV_\epsilon(u_{\lambda_2}) \leqslant \lambda_1 (F(u_{\lambda_2}) - F(u_{\lambda_1})); \end{aligned} \tag{3.3}$$

$$\begin{aligned} &TV_\epsilon(u_{\lambda_2}) + \lambda_2 F(u_{\lambda_2}) \leqslant TV_\epsilon(u_{\lambda_1}) + \lambda_2 F(u_{\lambda_1}),\\ &\text{i.e., } TV_\epsilon(u_{\lambda_2}) - TV_\epsilon(u_{\lambda_1}) \leqslant \lambda_2 (F(u_{\lambda_1}) - F(u_{\lambda_2})). \end{aligned} \tag{3.4}$$

Adding (3.3) to (3.4) leads to $F(u_{\lambda_2}) \leqslant F(u_{\lambda_1})$; adding a multiple λ_2 of (3.3) to a multiple λ_1 of (3.4) yields $TV_\epsilon(u_{\lambda_1}) \leqslant TV_\epsilon(u_{\lambda_2})$. Thus we have proved (3) and (4).

Indeed if $0 < \lambda_1 < \lambda_2$, u_{λ_i} is the unique global minimizer of the objective functional $E_\epsilon(\lambda_i, u)$ for $i = 1, 2$, all inequalities hold strictly. ∎

Since $F(u)$ is continuous on $BV(\Omega)$ and u_λ is also continuous with respect to λ (from the *implicit function theorem* in Banach spaces; refer to [50, Theorem 4B] and [21]), we infer that $F(u_\lambda)$ is also continuous in λ. Therefore for $\lambda \geqslant 0$, according to Theorem 3.4, $F(u_\lambda)$ is strictly monotonous. So it only remains to identify two distinct points of nonnegative λ at which $F(u_\lambda)$ take opposite signs in order to show that the solution of (1.8a) can satisfy $F(u_\lambda) = 0$.

Let $U_0 := \{u \in BV(\Omega) : L_\lambda(u) = 0 \text{ for all } \lambda \geqslant 0\}$. In order to construct a sequence $\{u_{\lambda_i} \in U_0,\ i = 0, 1, 2, \dots\}$ such that $\lim_{i\to+\infty} F(u_{\lambda_i}) = 0$, we give the following theorem.

THEOREM 3.5. *For any given σ such that $0 < \sigma < \|z - m(z)\|_{L^2(\Omega)}$, where $m(z)$ is the mean of function $z(x)$ on convex Ω, i.e., $m(z) = \frac{\int_\Omega z(x)dx}{|\Omega|}$, there exist $u_{\lambda_0}(x)$, $u_{\lambda_1}(x) \in U_0$ such that*

$$F(u_{\lambda_0}(x)) > 0 \text{ and } F(u_{\lambda_1}(x)) < 0.$$

Proof. Recall that $u_\lambda(x) = \arg\min_u \{TV_\epsilon(u) + \lambda(\frac{1}{2}\|u - z\|^2_{L^2(\Omega)} - \sigma^2)\}$. Since $u_0(x) = \arg\min_u TV_\epsilon(u)$ has a constant solution, assuming the initial guess used is z, we obtain $u_0 = m(z)$. Similarly, $u_\infty(x) = \arg\min_u(\frac{1}{2}\|u - z\|^2_{L^2(\Omega)} - \sigma^2)$ has the solution z. Hence

$$\lim_{\lambda\to 0} F(u_\lambda(x)) = F(u_0(x)) = F(m(z)) > 0$$

and

$$\lim_{\lambda\to+\infty} F(u_\lambda(x)) = F(u_{+\infty}(x)) = F(z) < 0,$$

so there exists an $M > 0$ such that $F(u_{\lambda_0}(x)) > 0$ and $F(u_{\lambda_1}(x)) < 0$, if we take $\lambda_1 > M$, $u_{\lambda_0} = m(z) \in U_0$, and $u_{\lambda_1} \in U_0$.

Alternatively, define a zero mean value image $\bar{z} = z - m(z)$ and the associated problem $u(\lambda) = \arg\min_u\{TV_\epsilon(u) + \lambda(\frac{1}{2}\|u - \bar{z}\|^2_{L^2(\Omega)} - \sigma^2)\}$. Refer to [10, 21] for more details.) Since $u(0) = 0$ and $u(+\infty) = \bar{z}$, for the general z with $u_\lambda(x)$, we have $u_0 = 0 + m(z) = m(z)$ and $u_{+\infty} = \bar{z} + m(z) = z$. Then the same argument applies, i.e., we have $F(u_{\lambda_0}(x)) > 0$ and $F(u_{\lambda_1}(x)) < 0$. ∎

The above theorem shows that there exist such nonnegative parameters λ_0 and λ_1 such that $F(u_{\lambda_0}(x)) > 0$ and $F(u_{\lambda_1}(x)) < 0$. Thus according to continuity and monotonicity of the constrained function $F(u_\lambda)$ with respect to λ, there exists a unique $\lambda^* \in (\lambda_0, \lambda_1)$ such that $(\lambda^*, u_{\lambda^*})$ is the solution of the KKT system (1.8). Consequently, starting from two such initial values, it is not a hard task to locate the desired λ^*.

The basic idea of constructing an approximation sequence u_{λ_i} converging to u_{λ^*} is the following:

(a) Set $\lambda_0 = 0$, $\lambda_1 = M$. Solve the solution u_{λ_1} of (1.8a) by using a multigrid method.
(b) If $F(u_{\lambda_1}) > 0$, set $\lambda_0 := M$, take $M := 2M$, and solve recursively the equation (1.8a) until $F(u_{\lambda_1}) < 0$. Set $\lambda_1 := M$.
(c) Take $\lambda_2 \in (\lambda_0, \lambda_1)$, obtain u from (1.8a) by a multigrid method, and set $u_{\lambda_2} := u$.
(d) If $F(u_{\lambda_2}) < 0$, take $\lambda_3 \in (\lambda_0, \lambda_2)$; else if $F(u_{\lambda_2}) > 0$, take $\lambda_3 \in (\lambda_2, \lambda_1)$ and compute u_{λ_3}.
(e) Similar to (d), obtain $(\lambda_4, u_{\lambda_4})$, $(\lambda_5, u_{\lambda_5})$, $(\lambda_6, u_{\lambda_6})$, ...

3.3. Relationships of the solutions of constrained and unconstrained problems. Before we present our algorithm, it remains to connect the solutions of constrained and unconstrained problems because our idea is to solve the constrained problem (KKT system) through solving the relatively easier unconstrained problems.

The above sequence gives the solution $(\lambda^*, u_{\lambda^*})$ of the KKT system (1.8) with the boundary condition (1.9), which is also a stationary point of the unconstrained optimization problem (1.7). The following theorem shows that such a solution u_{λ^*} is also the global minimizer of the constrained optimization problem (1.4).

THEOREM 3.6. *Suppose that $\lambda^* > 0$ and $u_{\lambda^*}(x) \in BV(\Omega)$ is a stationary point of the Lagrangian functional $E_\epsilon(\lambda, u)$, i.e., solves the unconstrained optimization problem* (1.7). *Then the same function $u_{\lambda^*}(x)$ also solves the constrained problem* (1.4).

Proof. For given $\lambda^* > 0$, according to Theorem 3.1, $E_\epsilon(\lambda^*, u)$ is a strictly G-differentiable convex functional, $u_{\lambda^*}(x)$ is the stationary point of $E_\epsilon(\lambda^*, u)$, and hence $u_{\lambda^*}(x)$ is the unique and global minimizer of $E_\epsilon(\lambda^*, u)$ from Theorems 3.2–3.3, i.e.,

$$E_\epsilon(\lambda^*, u_{\lambda^*}) < E_\epsilon(\lambda^*, u) \quad \text{or} \quad TV_\epsilon(u_{\lambda^*}) + \lambda^* F(u_{\lambda^*}) < TV_\epsilon(u) + \lambda^* F(u),$$

where $u \in BV(\Omega)$ and $u \neq u_{\lambda^*}$. Since u_{λ^*} satisfies the constrained condition $F(u_{\lambda^*}) = 0$ from (1.8b), the above inequality is equivalent to

$$TV_\epsilon(u_{\lambda^*}) < TV_\epsilon(u) + \lambda^* F(u) \text{ for all } u \in BV(\Omega) \text{ and } u \neq u_{\lambda^*}$$

and

$$TV_\epsilon(u_{\lambda^*}) < TV_\epsilon(u) \text{ for all } u \in BV(\Omega) \text{ with } F(u) = 0 \text{ and } u \neq u_{\lambda^*}.$$

This means that $u_{\lambda^*}(x) \in BV(\Omega)$ minimizes the constrained problem (1.4). ∎

3.4. Algorithm descriptions. We are now ready to state our algorithm, which is essentially a realization of the intended idea of solving the constrained problem

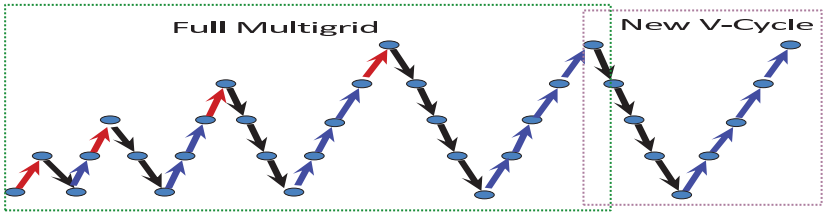


FIG. 3.1. *Structure of a full multigrid scheme. The left box presents a full multigrid cycle. Red arrows denote the high-order interpolation operator HR_{2h}^h, a common full multigrid interpolation for second-order accurate discretizations is cubic interpolation. Black and blue arrows denote the interpolation and restriction transfer within the multigrid cycle, respectively. After running the full multigrid cycle, the new multigrid cycle (in red box) beginning the previous approximation solution is implemented circularly until the stopping criterion is met.*

(KKT system) through solving unconstrained problems. Specifically, the goal of our algorithm is to seek the solution u^* of the nonlinear equation (1.8a) with different multipliers λ while examining if the constrained condition (1.8b) is satisfied; the latter condition is the discrepancy principle.

In order to obtain the solution of the KKT system (1.8) in an indirect way, first we present a multigrid algorithm to solve the nonlinear equation (1.8a). As in all efficient multigrid implementations, we use a full multigrid method (also see Algorithm 1 and Figure 3.1) to start from the coarsest grid Ω_0 in order to provide a good initial approximation on the finest grid Ω_ℓ through computation and interpolation of approximations on coarser grids. Each coarse grid solution serves as an initial approximation for the next finer grid. This process continues until it reaches the finest level where the solution of the problem is required for the usual multigrid. Typically for a full multigrid method, a transfer of higher accuracy than the interpolation used within normal multigrid cycles has to be chosen and then the full multigrid scheme is the most efficient [44].

ALGORITHM 1 (full multigrid algorithm). *Below our transfer operator HR_{2h}^h is a high-order interpolation operator and ℓ denotes the grid level number. FAS denotes a normal full approximation algorithm by Brandt* [41, 27, 44, 8]*, as in Figure* 3.1.

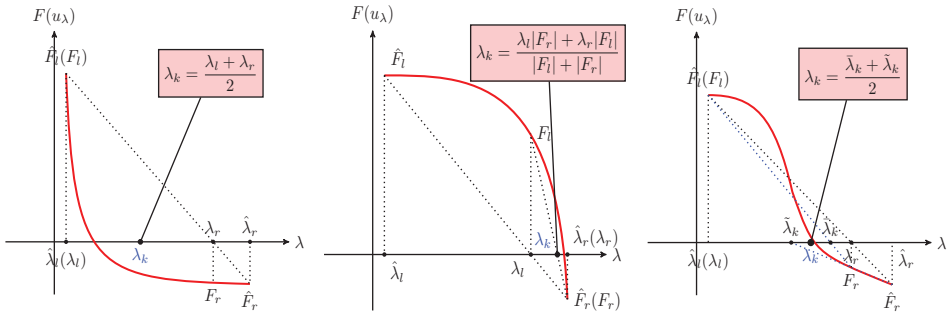
Step 1. *On level $i = 0$:*
  Solve exactly $L_\lambda^0(u) = 0$ to obtain u_0 and set $u_0^* = u_0$;

Step 2. *On level $i = 1, 2, \ldots, \ell - 1$:*
  $u_i := HR_{2h}^h u_{i-1}^*$;
  $u_i^* = FAS(i + 1, u_i)$;

Step 3. *On the finest level $i = \ell$, set $u_0 = u_{\ell-1}^*$.*
  Iterate on j for IT steps:
  $u_j = FAS(\ell, u_{j-1})$; *check stopping criterion;*

As indicated, our algorithm will start from two parameters λ_0, λ_1 and search for λ^*, where $F(\lambda^*) = 0$. Although the bisection method is an obvious choice, it may not be efficient (depending on the exact profile of $F(u_\lambda)$) and we found that a combined method of the secant method, the bisection method, and the alternating secant mean-value marching (ASMM) method is much better, as shown in Algorithm 2 for the multiplier parameter-update technique. Here the secant method, the bisection method, and the ASMM method are employed alternatively to seek the zero of the function $F(u_\lambda)$.

Our idea of acceleration aims to overcome slow convergence when two consecutive iterates are too close to each other. Suppose $\hat{F}_l$ and $\hat{F}_r$ are function values of the left



(a) Case 1: Bisection method (b) Case 2: Secant method (c) Case 3: ASMM method

FIG. 3.2. *Partial cases in iterative multiplier update scheme, where* $\bar{\lambda}_k = \lambda_r - \frac{\tau_r \eta_r}{\tau_r + \eta_r}(\hat{\lambda}_r - \lambda_r)$ *and* $\tilde{\lambda}_k = \lambda_r - \frac{\eta_r}{1-\eta_r}(\hat{\lambda}_r - \lambda_r)$ *with* $\eta_r = \frac{|F_r|}{|\hat{F}_r|}$ *and* $\tau_r = \frac{|\hat{F}_l|}{|\hat{F}_r|}$.

and right of a zero point in the previous iteration, while F_l and F_r are function values of the left and right of the zero point at the current iteration, for example, the current iteration may be obtained by $F_l = F(u_{\lambda_2})$ if $F(u_{\lambda_2}) > 0$ and $F_r = F(u_{\lambda_2})$ when $F(u_{\lambda_2}) < 0$, in the bisection method from the previous iteration $\hat{F}_l = F(u_{\lambda_0})$ and $\hat{F}_r = F(u_{\lambda_1})$. In general, since each in a pair of values F_l, F_r has opposite signs, the case of $|F_a|/|F_b| \geqslant |\hat{F}_a|/|\hat{F}_b| > \tau_0$ and $|F_b|/|\hat{F}_b| > 0.80$ (where $a \in A = \{r, l\}$, $b \in A \setminus a$) is naturally suited to the bisection method. Otherwise the secant and the ASMM method are better, as given in the Algorithm 2, where we take 0.8 as close enough to 1 and set $\tau_0 > 1$. We also illustrate three cases of $\frac{|\hat{F}_l|}{|\hat{F}_r|} > \tau_0$ in Figure 3.2, where we take $\tau_0 = 2$.

ALGORITHM 2 (a robust parameter choice method). *Given an integer* k*, two suitable parameters* $\hat{\lambda}_l$ *and* $\hat{\lambda}_r$*, the corresponding function values* $\hat{F}_l$ *and* $\hat{F}_r$ *in the previous iteration, two parameters* λ_l *and* λ_r*, the corresponding function values* F_l *and* F_r *in the current iteration, and some* $\tau_0 > 1$*:*

Step 1. *If* $k = 2$*, Secant* $\lambda_k = \frac{\hat{\lambda}_l|\hat{F}_r| + \hat{\lambda}_r|\hat{F}_l|}{|\hat{F}_l| + |\hat{F}_r|}$ *and finish the algorithm;*

Step 2. *For* $k > 2$*, do the following*

- *If the current ratio is similar to the previous:*
  $\tau_r = \frac{|\hat{F}_l|}{|\hat{F}_r|} > \tau_0$ & $\frac{|F_l|}{|F_r|} \geqslant \frac{|\hat{F}_l|}{|\hat{F}_r|}$,
  - *if two current end values are almost equal* $\eta_r = \frac{|F_r|}{|\hat{F}_r|} > 0.80$,
    $\lambda_k = (\lambda_l + \lambda_r)/2$; *(Bisection method)*
  - *else*
    $\lambda_k = \lambda_r - \frac{1}{2}(\frac{\tau_r \eta_r}{\tau_r + \eta_r} + \frac{\eta_r}{1-\eta_r})(\hat{\lambda}_r - \lambda_r)$; *(ASMM method)*
  - *end*
- *elseif the current reciprocal ratio is similar to the previous:*
  $\tau_l = \frac{|\hat{F}_r|}{|\hat{F}_l|} > \tau_0$ & $\frac{|F_r|}{|F_l|} \geqslant \frac{|\hat{F}_r|}{|\hat{F}_l|}$,
  - *if two current end values are almost equal* $\eta_l = \frac{|F_l|}{|\hat{F}_l|} > 0.80$,
    $\lambda_k = (\lambda_l + \lambda_r)/2$; *(Bisection method)*
  - *else*
    $\lambda_k = \lambda_l + \frac{1}{2}(\frac{\tau_l \eta_l}{\tau_l + \eta_l} + \frac{\eta_l}{1-\eta_l})(\lambda_l - \hat{\lambda}_l)$; *(ASMM method)*
  - *end*
- *else*
  $\lambda_k = \frac{\lambda_l|F_r| + \lambda_r|F_l|}{|F_l| + |F_r|}$; *(Secant method)*
- *End*

To demonstrate how this parameter algorithm can be superior, we consider the following generic example of seeking the zero for $F(x) = 10000 - x^3$ in [0, 101]. First the secant method is not yet convergent to the solution after 100 iterations and the bisection method reaches the residual $F(x) = -4.23e^{-4}$ after 27 iterations when $|x_k - x_{k-1}| < 1.0e^{-6}$ is quite small, while our ASMM scheme converges to solution $x = 21.5444$ with a much smaller residual $F(x) = 1.4188e^{-10}$ after only 13 iterations.

Finally we are ready to state our overall algorithm. In accordance with the above theoretical results, we assume $\eta(x, y)$ is zero mean value noise with variance σ^2 (i.e., $m(u) = m(z)$), and the condition $0 < \sigma < \|z - m(z)\|_{L^2(\Omega)}$ throughout the paper. Then $F(m(z)) > 0$ can be assured when $\lambda = 0$ is chosen and $m(z)$ is the solution of the ROF equation (1.8a). Thus the first initial λ satisfying $F(u_\lambda) > 0$ is $\lambda_0 = 0$, while for the second initial λ satisfying $F(u_\lambda) < 0$ we pick some large value M and set $\lambda_1 = M$. The final Algorithm 3 below uses the above full multigrid Algorithm 1 and the parameter choice Algorithm 2.

ALGORITHM 3 (iterative multiplier method). *Given two initial parameters $\lambda_0 = 0$ and $\lambda_1 = M > 0$, update λ_0, λ_1 (with respective solutions u_{λ_0} and u_{λ_1} of* (1.8a) *are such that $F(u_{\lambda_0}) > 0$ and $F(u_{\lambda_1}) < 0$). Let $F_l = F(u_{\lambda_0})$, $F_r = F(u_{\lambda_1})$, $\lambda_l = \lambda_0$, $\lambda_r = \lambda_1$, $k = 2$;*

Step 1. *Obtain λ_k by Algorithm* 2;

Step 2. *Set $\hat{F}_l = F_l$ and $\hat{F}_r = F_r$; $\hat{\lambda}_l = \lambda_l$ and $\hat{\lambda}_r = \lambda_r$; Solve the equation* (1.8a) *to obtain the unique solution u_{λ_k} by using the full multigrid method (*see *Algorithm* 1 *or Figure* 3.1);

- *If $\|F(u_{\lambda_k})\| < \eta_0$, then*
  u_{λ_k} solves the constrained optimization problem (1.4) *and return solution $u^* := u_{\lambda_k}$;*
- *else if $F(u_{\lambda_k}) \leqslant -\eta_0$*
  set $F_r = F(u_{\lambda_k})$, $\lambda_r = \lambda_k$;
- *else if $F(u_{\lambda_k}) \geqslant \eta_0$*
  set $F_l = F(u_{\lambda_k})$, $\lambda_l = \lambda_k$;
- *end*

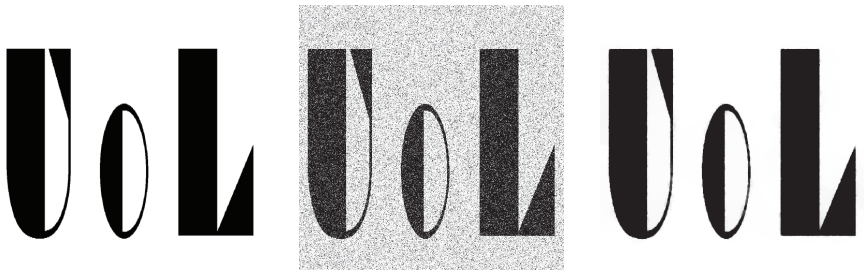
Step 3. *Check the step length;*

- *If $|\lambda_k - \lambda_{k-1}| < \eta_1$,*
  stop and return $u^* := u_{\lambda_k}$;
- *else*
  $k := k + 1$, go back to Step 1*;*
- *end*

Thus we have presented an algorithm for getting a solution to the KKT system without directly solving it.

4. Numerical experiments. This section presents some numerical results from experiments using our iterative multiplier Algorithm 3 for synthetic and real images. Four sets of noisy data will be used in our tests, called Example 1, 2, 3, 4, respectively, and an initial guess is the noisy image $z(x, y)$. The full multigrid process solving the diffusion equation (1.8a) is stopped after achieving a relative residual of 10^{-4} or relative error of 10^{-4} or 40 multigrid cycles, and the smoothing parameter $\epsilon = 1.0e^{-2}$ (except for comparisons) is fixed. Finally the tests for the Lagrange multiplier λ will be finished in achieving a constrained function residual of $TOL = 10^{-2}$(i.e., $F(u_\lambda) < TOL$).

Example 1. First, we test the performance of Algorithm 3 for the UoL binary noisy image in Figure 4.1. The experiment is carried out on a grayscale image with



(a) True image (b) Noise image z (c) Denoise image u

FIG. 4.1. *UoL binary image denoising result. From left to right: the original image, the noisy image with a zero mean Gaussian random noise of variance* $\sigma^2 = (\frac{100}{255})^2 \times 512^2$ *(*or *psnr* $= 8.13$*), where* $psnr = 10\log_{10}\frac{nm(\max u)^2}{\|u-u^*\|_F^2}$ *is the peak signal-to-noise ratio and* u^* *is an approximation to a noise-free image* u*, and the denoised image(*or *psnr* $= 25.81$*) of constrained optimization problem* (1.8) *by our iterative method.*

512×512 pixels and range $[0, 1]$ to which zero mean value Gaussian random noise with variance $\sigma^2 = (\frac{100}{255})^2 \times 512^2$. Let $u_{\lambda_0} = m(u) = m(z)$ and u_{λ_1} be the solutions of the diffusion equation (1.8a) with multiplier $\lambda_0 = 0$ and fixed $\lambda_1 = 10000$ (deliberately set quite large, unlike in Algorithm 3, where we set $M = 1000$ and adaptively choose λ_1), respectively, in these cases that the constrained functions $F(u_{\lambda_0}) = 53245.42$ and $F(u_{\lambda_1}) = -36798.63$ are examined. Because of continuity of $F(u_\lambda)$ with respect to λ, its unique zero point λ^* is achieved in interval (λ_0, λ_1) from our Algorithm 2.

We show this test in Figure 4.1. The original and noisy images are displayed at the left and the middle of Figure 4.1, $\lambda^* = 697.04$ is detected to reduce the constrained function $F(u_{\lambda^*})$ to -0.0075, and the solution image u_{λ^*} is depicted on the right of Figure 4.1.

To intuitively describe the denoising ability, we present noisy image surface $z(x, y)$, regularization solution image surface $u^*(x, y)$, and curves $u^*(x_0, \cdot)$ along the y-direction in Figure 4.2. Especially, the *bottom left* shows the result of $x_0 = \frac{256}{512}$ in the y-direction with a blue curve denoting the noisy curve $z(0.5, \cdot)$, a red denoting the true curve $u(0.5, \cdot)$, and a green one giving the solution $u^*(0.5, \cdot)$. Similarly, the *bottom right* presents the comparison of three curves when x_0 is $\frac{180}{512}$.

Example 2. Second, we compare the performance of our iterative Lagrange method (Algorithm 3) with Landi's Lagrange method in [35]. A texture image with 256×256 pixels in range $[0, 1]$ and zero mean value Gaussian random noise with variance $\sigma^2 = (\frac{50}{255})^2 \times 256^2$ is tested. Lagrange multipliers λ^* and solutions u^* are obtained by our iterative Lagrange method and Landi's Lagrange method in this example, and the results are presented in Figure 4.3. Let $u_{fmean} = m(u) = m(z)$ be the solution of the diffusion equation (1.8a) with multiplier $\lambda_0 = 0$, and $\lambda_1 = M = 1000$. The noise-free image and noisy image are respectively displayed at the top of Figure 4.3. Figure 4.3(c) presents a convergent restoration result using our Algorithm 3 after implementing six iterations for a fixed smoothing parameter $\epsilon = 1.0$; the optimal Lagrange multiplier $\lambda^* = 1196.62$ is detected. In contrast, with the Landi Lagrange method, Figure 4.3(d) displays a convergent denoised image after 249 iterations, and $\lambda^* = 1175.18$ is obtained. With $\epsilon = 1.0$, we note that both a Newton-type method and a multigrid method solve the diffusion equation with good performance in this case. However, with $\epsilon = 0.01$ (i.e., increased nonlinearity), Newton-type methods such as Landi's Lagrange method may have convergence difficulties; Figure 4.3(f) presents a bad result in the case of solving the KKT system using Landi's Lagrange method

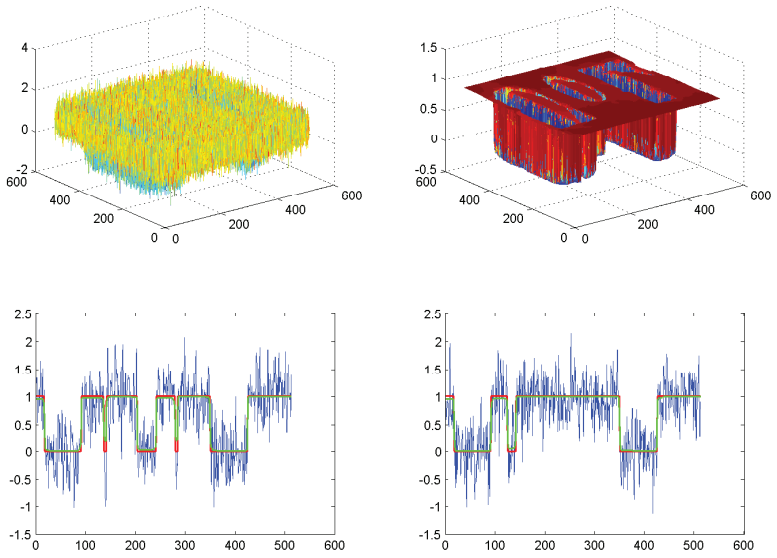
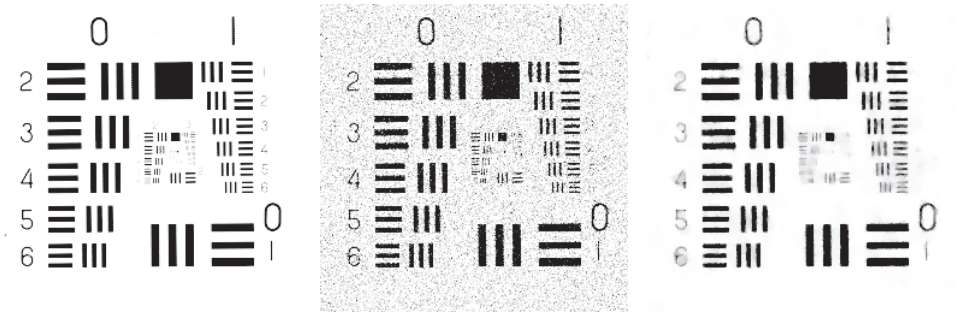
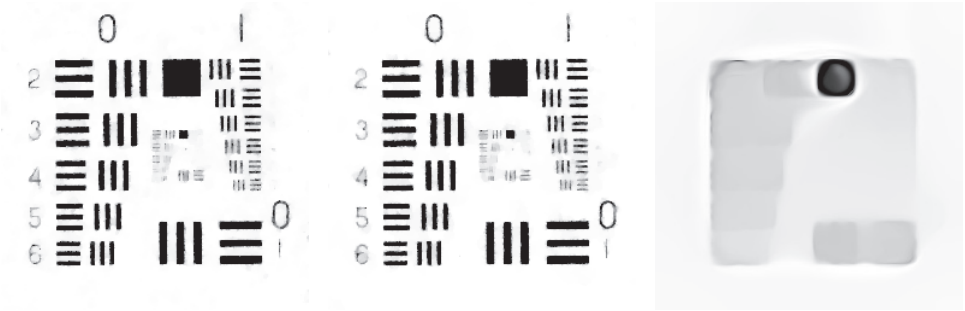


FIG. 4.2. *Surface and curve figures of results in Figure* 4.1. *From left to right (top row): noise image surface, restoration image surface. On the bottom row: superimposition of the noisy curve* $z(x_0,\cdot)$*, the true curve* $u(x_0,\cdot)$*, and the restored curve* $u^*(x_0,\cdot)$ *for* $x_0 = \frac{256}{512}$ *along the y-direction (left) and the same curves at* $x_0 = \frac{180}{512}$ *(right). Blue curve: noise data; red curve: true data; green curve: restoration data.*

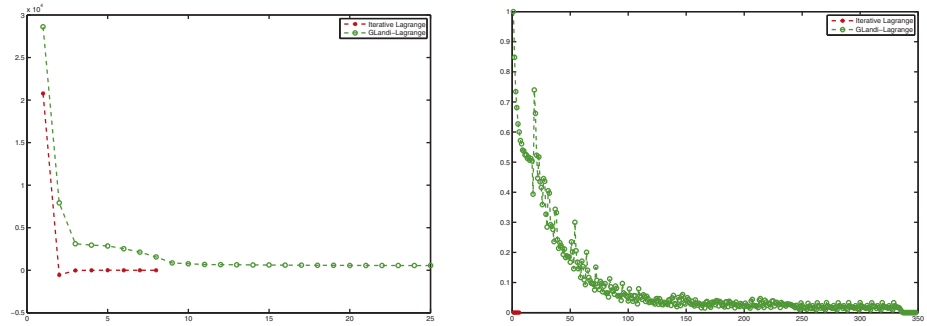


(a) True image (b) Noisy image ($\sigma = \frac{50}{255} \times 256$) (c) Our method ($\epsilon = 1$)



(d) Landi's method ($\epsilon = 1$) (e) Our method ($\epsilon = 0.01$) (f) Landi's method ($\epsilon = 0.01$)

FIG. 4.3. *Comparison of our iterative Lagrange multiplier method and Landi's Lagrange method.*



(a) Comparison of constraint functions $F(u)$.

(b) Comparison of the residuals of objective functions.

FIG. 4.4. *Comparison of the relative residuals of the Lagrange equation $L_\lambda(u)$ and the constraint functional $F(u)$.*

TABLE 4.1

Comparison of our method and Landi's method with different smoothing parameter ϵ for a noise image with variance $\sigma^2 = (\frac{50}{255})^2 \times 256^2$ (or psnr = 14.15) in Figure 4.3(b).

	Iterative multiplier				Landi's method		
ϵ	λ^*	O-It	I-It	psnr	λ^*	O-It	psnr
10^0	1196.62	4	3-3/3-3-3-3	22.36	1175.18	249	22.32
10^{-1}	1238.72	4	4-4/4-4-4-4	22.55	1217.18	329	22.51
10^{-2}	1254.12	4	10-5/5-5-5-5	22.61	**178.97**	**>3000**	**12.27**
10^{-3}	1259.15	4	40-5/12-12-12-12	22.63	**4.08**	**>3000**	**10.37**
10^{-4}	1261.21	4	40-5/40-40-40-40	22.64	**<1**	**>3000**	**6.76**

after 100 iterations; however, our Algorithm 3 still gives an excellent denoised image (see Figure 4.3(e)).

We also compare the relative residual history and the constraint functional $F(u)$ history of the two methods in Figure 4.4. Here we note that our iterative Lagrange multiplier method solves accurately the diffusion equation in each iteration (small relative residual) and can rapidly reduce the modulus of the objective function (a few iterations) with less oscillations.

Performance comparisons of our method and Landi's method for the different smoothing parameter ϵ are shown in Table 4.1 by finding the best regularization parameter λ^* and the corresponding *psnr* value of denoised images. We also report the numbers of iterations used by both methods in the third, fourth, and seventh columns. The first O-It denotes the numbers of iterations finding the optimal value λ^* in a feasible interval (λ_0, λ_1), and I-It denotes the number of multigrid cycles in the corresponding outer iterations (while numbers before / denote multigrid cycles used in seeking the feasible interval). The second O-It denotes the numbers of the Newton iterations in Landi's method. (The cases of convergence difficulties are given in boldface, where λ and *psnr* are shown after achieving 3000 iterations.)

Example 3. In the third example we demonstrate the robust process of Algorithm 3 for dealing with a larger resolution. A grayscale image with 512×512 pixels and zero mean value Gaussian random noise with variance $\sigma^2 = (\frac{50}{255})^2 \times 512^2$ is tested, and we show the obtained parameter sequence λ_k converging to the Lagrange multiplier λ^* and solution paths $\{u_k\}$ are followed in this example. Let $u_{fmean} = m(u) = m(z)$ be the solution of the diffusion equation (1.8a) with multiplier $\lambda_1 = 0$,

FIG. 4.5. *Camera image denoising result. From left to right and then top to bottom: the original image, the noisy image with a zero mean Gaussian random noise of variance* $\sigma^2 = (\frac{50}{255})^2 \times 512^2$ *(or* $psnr = 14.30$*), and eight restoration images: mean image by setting* $\lambda = 0$*, the denoised image from setting* $\lambda = M = 1000$*, and the denoised image sequence at multiplier* $\lambda_k (k = 1, \ldots, 6)$*; the psnr of the final image is* 27.17*.*

TABLE 4.2
Parameter choice λ_k *following.*

k	1	2	3	4	5	6
λ_k	2000	3000	2056.57	2069.22	2070.58	2070.65
u_k	u_1	u_2	u_3	u_4	u_5	u_6
$F(u_{\lambda_k})$	41.530	−692.598	8.300	0.844	0.043	0.001

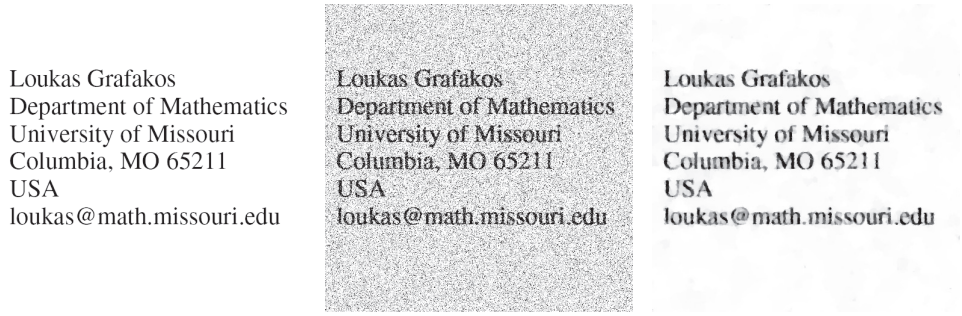
and set $\lambda_2 = M = 1000$ initially. In this case at the initial step, the constrained functions are $F(u_{fmean}) = 25561.225$, $F(u_{1000}) = 777.251$, $F(u_{2000}) = 41.530$, and $F(u_{3000}) = -692.598$. Because of the continuity of $F(u_\lambda)$ with respect to λ, its unique zero point $\lambda^* = 2070.65$ is achieved in interval (λ_1, λ_2) from our Algorithm 2.

We show in Figure 4.5 how efficiently our iterative multiplier method restores such a real image. Here only a few iterations are necessary to obtain the solution $(u^*(x, y), \lambda^*)$ of the KKT system (1.8), but other common penalty methods with fixed λ are not satisfactory in solving this coupled KKT system. The path of Lagrange multipliers λ_k is shown in Table 4.2 using Algorithm 2, where u_{λ_k} is the solution of the diffusion equation (1.8a) at $\lambda = \lambda_k$.

Here we observe that the sequence of approximations of our iterative scheme converges to the limit λ^*. Thus the use of our iterative multiplier approach is reasonable from both the numerical and the theoretical point of view.

Example 4. The fourth example illustrates the capability of Algorithm 3 in adaptively dealing with different noisy variance σ^2. The sets of noisy image $z(x, y)$ with 512×512 pixels and range $[0, 1]$ are presented in the first row of Figure 4.6 with a zero mean Gaussian random noise of variance, respectively, $\sigma_1^2 = (\frac{20}{255})^2 \times 512^2$, $\sigma_2^2 = (\frac{40}{255})^2 \times 512^2$, $\sigma_3^2 = (\frac{60}{255})^2 \times 512^2$, $\sigma_4^2 = (\frac{80}{255})^2 \times 512^2$, and $\sigma_5^2 = (\frac{100}{255})^2 \times 512^2$ from left to right. We report the associated visual results for different images with different σ^2 in the second row of Figure 4.6. For these images from left to right, different optimal Lagrange multipliers are detected from Algorithm 3 and are collected as follows: $\lambda_1^* = 5682.81$, $\lambda_2^* = 2521.23$, $\lambda_3^* = 1564.42$, $\lambda_4^* = 1117.23$, $\lambda_5^* = 863.31$. These multipliers also show that higher variance σ^2 requires smaller parameter λ to solve the KKT system (as expected). The key advantage of our algorithm lies in its

FIG. 4.6. *Numerical results of Lena test image with different variance* σ^2. *From left to right on the first row: noise image of variance, respectively,* $\sigma_1^2 = (\frac{20}{255})^2 \times 512^2$ *(*or $psnr = 21.87$*),* $\sigma_2^2 = (\frac{40}{255})^2 \times 512^2$ *(*or $psnr = 15.85$*),* $\sigma_3^2 = (\frac{60}{255})^2 \times 512^2$ *(*or $psnr = 12.33$*),* $\sigma_4^2 = (\frac{80}{255})^2 \times 512^2$ *(*or $psnr = 9.83$*), and* $\sigma_5^2 = (\frac{100}{255})^2 \times 512^2$ *(*or $psnr = 7.89$*). From left to right on the second row: the associated restoration results and corresponding* $psnrs$: $29.35, 26.50, 24.98, 23.95, 23.16$.



(a) True image (b) Noise image z (c) Denoise image u

FIG. 4.7. *Binary text image denoising result. From left to right: the original image, the noisy image with a zero mean Gaussian random noise of variance* $\sigma^2 = (\frac{90}{255})^2 \times 512^2$ *(*or $psnr = 9.05$*), and the denoised image(*or $psnr = 21.13$*) of constrained optimization problem* (1.8) *by our iterative method.*

ability of *adaptive* parameter choice for different image sizes and different levels of noise, in contrast to many common TV methods requiring a given λ by users; often it is difficult to provide a good value due to the level of noise and image scale even with many tests.

Finally in this section, we use two real images to illustrate the quality of restoration by our new solution method. Figure 4.7 shows an experiment for a binary text image with 512×512 size and zero mean Gaussian random noise of variance $\sigma^2 = (\frac{90}{255})^2 \times 512^2$, where $\lambda^* = 1313.80$ is found after four iterations when $M = 1000$ in Algorithm 3. Figure 4.8 shows the result for an aircraft image with 512×512 size and zero mean Gaussian random noise of variance $\sigma^2 = (\frac{80}{255})^2 \times 512^2$. These two experiments again demonstrate that our method can restore images quite effectively (in quality) and efficiently (in iterations).

(a) True image (b) Noise image z (c) Denoise image u

FIG. 4.8. *Aircraft natural image denoising result. From left to right: the original image, the noisy image with a zero mean Gaussian random noise of variance* $\sigma^2 = (\frac{80}{255})^2 \times 512^2$ *(*or *psnr* $= 10.16$*), and the denoised image (*or *psnr* $= 25.01$*) of constrained optimization problem* (1.8) *by our iterative method. The Lagrange multiplier detected is* $\lambda^* = 1062.23$.

5. Conclusions. In this paper we presented an efficient iterative Lagrange multiplier method to solve the KKT system for a constrained TV image denoising problem (ROF model), the full multigrid scheme with Krylov space acceleration is employed to improve the computational performance of the nonlinear diffusion equation. The main advantage of our algorithm is its ability of adaptive multiplier choice for any image size and level of noise to obtain an accurate solution of the KKT system. In contrast, many common methods for ROF require a given λ by the user, which is difficult to provide and can dramatically influence the restoration result. From the theoretical point of view, the novelty of our new algorithm comes from analyzing the convexity of the objective function and uniqueness condition of the solutions, the transfer between the constrained and unconstrained optimization problem; with these, we are able to propose an iterative multiplier scheme through solving unconstrained optimization problems iteratively to find an effective solution of the KKT system without having to solve it directly. Numerical results show that this resulting scheme can automatically deal with different synthetic and natural images, different image resolutions, and different noise levels. Future work will consider generalization of this work to other denoising models [15, 8].

Acknowledgment. The authors thank the anonymous reviewers for their insightful comments, which greatly improved this paper.

REFERENCES

[1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
[2] G. AUBERT AND P. KORNPROBST, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, 2nd ed., Springer, New York, 2006.
[3] J. M. BARDSLEY AND A. LUTTMAN, *Total variation-penalized Poisson likelihood estimation for ill-posed problems*, Adv. Comput. Math., 31 (2009), pp. 35–59.
[4] P. BLOMGREN AND T. F. CHAN, *Modular solvers for image restoration problems using the discrepancy principle*, Numer. Linear Algebra Appl., 9 (2002), pp. 347–358.
[5] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.
[6] C. BRITO-LOEZA AND K. CHEN, *Multigrid method for a modified curvature driven diffusion model for image inpainting*, J. Comput. Math., 26 (2008), pp. 856–875.
[7] C. BRITO-LOEZA AND K. CHEN, *Fast numerical algorithms for Euler's elastica inpainting model*, Internat. J. Modern Math., 5 (2010), pp. 157–182.
[8] C. BRITO-LOEZA AND K. CHEN, *Multigrid algorithm for high order denoising*, SIAM J. Imaging Sci., 3 (2010), pp. 363–389.

[9] C. Brito-Loeza and K. Chen, *On high-order denoising models and fast algorithms for vector-valued images*, IEEE Trans. Image Process., 19 (2010), pp. 1518–1527.
[10] M. Burger, G. Gilboa, S. Osher, and J. J. Xu, *Nonlinear inverse scale space methods*, Commun. Math. Sci., 4 (2006), pp. 179–212.
[11] J. F. Cai, S. Osher, and Z. W. Shen, *Split Bregman methods and frame based image restoration*, Multiscale Model. Simul., 8 (2009/10), pp. 337–369.
[12] J. L. Carter, *Dual Method for Total Variation-based Image Restoration*, Technical report 02-13, UCLA CAM, 2002.
[13] A. Chambolle and P. L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
[14] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[15] Q. S. Chang, X. C. Tai, and L. Xing, *A compound algorithm of denoising using second-order and fourth-order partial differential equations*, Numer. Math. Theory Methods Appl., 2 (2009), pp. 353–376.
[16] R. H. Chan and K. Chen, *A multilevel algorithm for simultaneously denoising and deblurring images*, SIAM J. Sci. Comput., 32 (2010), pp. 1043–1063.
[17] R. H. Chan, C. W. Ho, C. Y. Leung, and M. Nikolova, *Minimization of detail-preserving regularization functional by Newton's method with continuation*, in Proceedings of IEEE International Conference on Image Processing, Genova, Italy, 2005, pp. 125–128.
[18] T. F. Chan, K. Chen, and J. L. Carter, *Iterative methods for solving the dual formulation arising from image restoration*, Electron. Trans. Numer. Anal., 26 (2007), pp. 299–311.
[19] T. F. Chan and K. Chen, *On a nonlinear multigrid algorithm with primal relaxation for the image total variation minimisation*, Numer. Algorithms, 41 (2006), pp. 387–411.
[20] T. F. Chan and K. Chen, *An optimization-based multilevel algorithm for total variation image denoising*, Multiscale Model. Simul., 5 (2006), pp. 615–645.
[21] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.
[22] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.
[23] T. F. Chan and P. Mulet, *On the convergence of the lagged diffusivity fixed point method in total variation image restoration*, SIAM J. Numer. Anal., 36 (1999), pp. 354–367.
[24] T. F. Chan and J. H. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.
[25] T. F. Chan, H. M. Zhou, and R. H. Chan, *Continuation method for total variation denoising problems*, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series 2563, F. T. Luk, ed., 1995, pp. 314–325.
[26] K. Chen, Y. Q. Dong, and M. Hintermueller, *Nonlinear multigrid solver with line Gauss-Seidel-semismooth-Newton smoother for the Fenchel pre-dual in total variation based image restoration*, Inverse Prob. Imaging, 5 (2011), pp. 323–339.
[27] K. Chen, *Matrix Preconditioning Techniques and Applications*, Cambridge University Press, Cambridge, UK, 2005.
[28] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), pp. 1200–1224.
[29] B. G. Fitzpatrick and S. L. Keeling, *On approximation in total variation penalization for image reconstruction and inverse problems*, Numer. Funct. Anal. Optim., 18 (1997), pp. 941–958.
[30] C. Frohn-Schauf, S. Henn, and K. Witsch, *Nonlinear multigrid methods for total variation image denoising*, Comput. Vis. Sci., 7 (2004), pp. 199–206.
[31] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-6 (1984), pp. 721–741.
[32] Y. M. Huang, M. K. Ng, and Y. W. Wen, *A fast total variation minimization method for image restoration*, Multiscale Model. Simul., 7 (2008), pp. 774–795.
[33] Y. M. Huang, M. K. Ng, and Y. W. Wen, *A new total variation method for multiplicative noise removal*, SIAM J. Imaging Sci., 2 (2009), pp. 20–40.
[34] G. Landi, *A truncated Lagrange method for total variation-based image restoration*, J. Math. Imaging Vision, 28 (2007), pp. 113–123.
[35] G. Landi, *The Lagrange method for the regularization of discrete ill-posed problems*, Comput. Optim. Appl., 39 (2008), pp. 347–368.
[36] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Kluwer Academic, Boston, 2003.

[37] A. Marquina and S. Osher, *Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal*, SIAM J. Sci. Comput., 22 (2000), pp. 387–405.
[38] L. A. Melara, A. J. Kearsley, and R. A. Tapia, *Augmented Lagrangian homotopy method for the regularization of total variation denoising problems*, J. Optim. Theory Appl., 134 (2007), pp. 15–25.
[39] S. Osher, M. Burger, D. Goldfarb, J. J. Xu, and W. T. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489.
[40] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[41] J. Savage and K. Chen, *An improved and accelerated non-linear multigrid method for total-variation denoising*, Int. J. Comput. Math., 82 (2005), pp. 1001–1015.
[42] E. Tadmor, S. Nezzar, and L. Vese, *Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation*, Commun. Math. Sci., 6 (2008), pp. 281–307.
[43] R. A. Tapia, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J. Optim. Theory Appl., 22 (1977), pp. 135–194.
[44] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, San Diego, CA, 2001.
[45] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.
[46] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
[47] C. L. Wu and X. C. Tai, *Augmented Lagrangian method, dual methods and split-Bregman iterations for ROF, vectorial TV and higher order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.
[48] F. L. Yang, K. Chen, and B. Yu, *Homotopy curve tracking for total variation image restoration*, J. Comput. Math., 30 (2012), pp. 177–196.
[49] E. Zeidler, *Nonlinear Functional Analysis and Its Applications* III: *Variational Methods and Optimization*, Springer-Verlag, New York, 1985.
[50] E. Zeidler, *Nonlinear Functional Analysis and Its Applications* I: *Fixed-Point Theorems*, Springer-Verlag, New York, 1986.
[51] M. Q. Zhu, S. J. Wright, and T. F. Chan, *Duality-based algorithms for total-variation-regularized image restoration*, Comput. Optim. Appl., 47 (2010), pp. 377–400.