A Generalized Accelerated Proximal Gradient Approach for Total Variation-Based Image Restoration

Wangmeng Zuo, *Member, IEEE,* and Zhouchen Lin, *Senior Member, IEEE*

***Abstract*—This paper proposes a generalized accelerated proximal gradient (GAPG) approach for solving total variation (TV) based image restoration problems. The GAPG algorithm generalizes the original APG algorithm by replacing the Lipschitz constant with an appropriate positive definite matrix, resulting in faster convergence. For TV-based image restoration problems, we further introduce two auxiliary variables that approximate the partial derivatives. Constraints on the variables can be easily imposed without modifying the algorithm much, and the TV regularization can be either isotropic or anisotropic. Compared with the recently developed APG-based methods for TV-based image restoration, i.e., monotone version of the two-step iterative shrinkage/thresholding algorithm (MTwIST) and monotone version of the fast iterative shrinkage/thresholding algorithm (MFISTA), our GAPG is much simpler as it does not require to solve an image denoising subproblem. Moreover, the convergence rate of $O(k^{-2})$ is maintained by our GAPG, where k is the number of iterations; the cost of each iteration in GAPG is also lower. As a result, in our experiments our GAPG approach can be much faster than MTwIST and MFISTA. The experiments also verify that our GAPG converges faster than the original APG and MTwIST when they solve identical problems.**

***Index Terms*—Image restoration, total variation, regularization, proximal gradient algorithm, convex optimization.**

I. Introduction

MANY image processing problems can be formalized as estimating the original image $\mathbf{x}$ from a corrupted observation $\mathbf{b}$ produced by first applying a linear operator $\mathcal{A}$ to $\mathbf{x}$ and then adding noise. Subsampling may also follow, resulting in missing values. A typical linear operator $\mathcal{A}$ is usually ill-conditioned or even singular. Thus image restoration is a classical linear inverse problem [1].

To solve the linear inverse problem, one needs to involve a regularization term in the objective function to utilize the prior knowledge to recover the original image. Image restoration is usually formulated as the following convex minimization problem:

$$\min_{\mathbf{x}\in B_{l,u}} F(\mathbf{x}) = \frac{1}{2}\left\|\mathcal{A}(\mathbf{x})-\mathbf{b}\right\|_F^2 + \lambda\Phi(\mathbf{x}), \tag{1}$$

W. Zuo is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, P.R. China. E-mail: cswmzuo@gmail.com.

Z. Lin is with Microsoft Research Asia, Beijing, 100190, P.R. China. E-mail: zhoulin@microsoft.com.

Manuscript received XX, XX 2010. This work is partially supported by Microsoft Research Asia, the National Natural Science Foundation of China under Contract No.s 60902099 and 60872099, and the Education Ministry Young Teacher Foundation of China under No. 200802131025.

where $\Phi(\mathbf{x})$ is a convex regularizer, λ is a regularization parameter [2], and $B_{l,u}$ are the boundedness constraints on the restored image $\mathbf{x}$:

$$B_{l,u} = \left\{\mathbf{x}\in\mathbb{R}^{m\times n} | l \le x_{i,j} \le u, \forall i,j\right\}, \tag{2}$$

in which l and u are the lower and upper bounds, respectively, and $m \times n$ is the size of the image.

Several regularizers, e.g., total variation (TV) [3], wavelet-based sparsity [4] [5], and non-local graph regularization [6] [7], have been proposed for image restoration. In this paper, we focus on TV-based image restoration. The TV model was first introduced in [3] as an effective regularizer. To date, this model has been widely adopted in many image processing problems, such as image denoising [8], blind deconvolution [9], compressed MR imaging [10], and microarray processing [11].

Because of the non-smooth and non-differentiable property of the TV regularizer, it is difficult to solve TV-based image restoration by conventional optimization methods. To date numerous methods, e.g., gradient-based methods [12][13][14][15], dual methods [8], graph cut [16], and second-order cone programming [17], have been developed in different contexts.

Nowadays images acquired by digital cameras usually contain tens of megapixels, which makes TV-based image restoration inherently a large scale optimization problem. Recently continuous efforts have been spent on designing effective algorithms with less requirements on computational load and memory. One gradient-based method, the iterative shrinkage/thresholding algorithm (IST), is developed for wavelet-based deconvolution [15]. Other researchers have also independently proposed IST in different contexts. Rigorous proofs of the convergence of IST have been provided in [18] and [19]. Subsequently, a generalized expectation maximization algorithm [13], also named as the iterative reweighted shrinkage method in [14], was proposed for image deconvolution.

Although simple and effective, IST has been known as a slow method, especially under some assumptions on the operator $\mathcal{A}$ [20]. Recently, a number of accelerated IST algorithms [14] [12] [21] have been proposed, and several of them, e.g., the two-step iterative shrinkage/thresholding (TwIST) algorithm and the fast iterative shrinkage/thresholding algorithm (FISTA), have been successfully applied to image restoration. FISTA is also known as the accelerated proximal gradient (APG) based method [22][23], which has an attractive

convergence rate of $O(k^{-2})$, where k is the number of iterations. APG has also been applied to matrix completion [24] and robust principal component analysis (RPCA) [22]. In this paper, we extend the original APG method to a more general and efficient class, called the generalized accelerated proximal gradient (GAPG) method, which maintains the convergence rate of $O(k^{-2})$.

When applied to TV-based image restoration, the TwIST [14] and FISTA [21] algorithms involve both outer and inner iterations, where the inner iteration is to approximately solve an image denoising subproblem. Recently, several image denoising methods, e.g., the dual approach by Chambolle [8] and the maximum flow algorithm by Goldfarb and Yin [25], have been proposed for solving the TV-based denoising problem. We expect that the efficiency of image restoration would be further improved if the image denoising subproblem could be avoided.

In this paper, by generalizing the FISTA method with a constant step size [12], we propose a GAPG method with a proven $O(k^{-2})$ convergence rate. Motivated by the variable splitting method [26][27][28], we further introduce two auxiliary variables to approximate the partial derivatives and reformulate the TV-based image restoration problem as an unconstrained convex optimization problem. Moreover, the problem reformulation is more suitable for improving the efficiency of GAPG, and we can also avoid solving the image denoising subproblem. Finally the GAPG framework is combined with the continuation technique [22][24][27][29][30] to solve the resulting optimization problem. Our method works for both anisotropic and isotropic discrete TV-based image restoration. Numerical results demonstrate the efficiency of the proposed method: our algorithms are much faster than the monotone version of TwIST (MTwIST) and the monotone version of FISTA (MFISTA) for TV-based image restoration. Further, our GAPG converges faster than the original APG and MTwIST when solving the same optimization problem.

The remainder of this paper is organized as follows. Section II introduces some background knowledge that is necessary for our paper, including the TV-based image restoration models, the original APG method, and a brief sketch of recent fast gradient-based methods for TV-based image restoration. Section III presents our GAPG method. Sections IV and V introduce the problem formalizations and algorithms for anisotropic and isotropic discrete TV-based image restoration, respectively. Then Section VI presents the experimental results. Finally Section VII concludes the paper.

II. Prerequisites and Related Work

In this section, we first describe the discrete TV-based image restoration models. Then we introduce the original APG method and the shrinkage operators which will be used in our TV-based image restoration algorithm. Finally, we briefly survey the recent APG-based methods for image restoration.

A. The Discrete TV-based Image Restoration Model

In the TV-based image restoration model, the linear operator $\mathcal{A}$ is an affine mapping and the regularization term in (1) is chosen as the total variation of the image $\mathbf{x}$:

$$\Phi(\mathbf{x}) = \mathrm{TV}(\mathbf{x}) \equiv \int_{\Omega} \|\nabla \mathbf{x}\| \mathrm{d}\Omega, \tag{3}$$

where Ω is the region that the image $\mathbf{x}$ occupies. When applying the model to discrete images, we have to discretize $\mathrm{TV}(\mathbf{x})$.

In the literature there are two kinds of discrete total variation: l_2-based isotropic TV and l_1-based anisotropic TV. Given a discrete image $\mathbf{x} \in \mathbb{R}^{m\times n}$, isotropic TV [21] is defined by

$$\mathrm{TV}_{iso}(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{(x_{i+1,j}-x_{i,j})^2+(x_{i,j+1}-x_{i,j})^2}\ , \tag{4}$$

and anisotropic TV [21] is defined by

$$\mathrm{TV}_{aniso}(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(|x_{i+1,j}-x_{i,j}|+|x_{i,j+1}-x_{i,j}|\right), \tag{5}$$

where we assume the reflexive boundary condition for $\mathbf{x}$:

$$\begin{array}{ll} x_{m+1,j} = x_{m,j}, & j = 1,\cdots,n; \\ x_{i,n+1} = x_{i,n}, & i = 1,\cdots,m. \end{array} \tag{6}$$

If we define the vertical and horizontal forward difference operators $\mathcal{D}_v$ and $\mathcal{D}_h$ as:

$$\mathbf{p} = \mathcal{D}_v(\mathbf{x}) : \mathbb{R}^{m\times n} \to \mathbb{R}^{(m-1)\times n}:$$

$$p_{i,j} = x_{i+1,j} - x_{i,j}, i = 1,\cdots,m-1, j = 1,\cdots,n, \tag{7}$$

and

$$\mathbf{q} = \mathcal{D}_h(\mathbf{x}) : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times(n-1)}:$$

$$q_{i,j} = x_{i,j+1} - x_{i,j}, i = 1,\cdots,m, j = 1,\cdots,n-1, \tag{8}$$

respectively, and isotropic TV induced norm as:

$$\|(\mathbf{p}\ \mathbf{q})\|_{iT} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{p_{i,j}^2+q_{i,j}^2}, \tag{9}$$

we may rewrite anisotropic and isotropic TVs as:

$$\mathrm{TV}_{iso}(\mathbf{x}) = \|(\mathcal{D}_v(\mathbf{x})\ \mathcal{D}_h(\mathbf{x}))\|_{iT}, \tag{10}$$

and

$$\mathrm{TV}_{aniso}(\mathbf{x}) = \|\mathcal{D}_v(\mathbf{x})\|_{l_1} + \|\mathcal{D}_h(\mathbf{x})\|_{l_1}, \tag{11}$$

respectively, where the matrix l_1-norm here refers to the sum of the absolute values.

The adjoint operators $\mathcal{D}_v^*$ and $\mathcal{D}_h^*$ of $\mathcal{D}_v$ and $\mathcal{D}_h$ can be found as:

$$\mathbf{z} = \mathcal{D}_v^*(\mathbf{p}) : \mathbb{R}^{(m-1)\times n} \to \mathbb{R}^{m\times n}:$$

$$z_{i,j} = p_{i-1,j} - p_{i,j}, i = 1,\cdots,m, j = 1,\cdots,n, \tag{12}$$

and

$$\mathbf{z} = \mathcal{D}_h^*(\mathbf{q}) : \mathbb{R}^{m\times(n-1)} \to \mathbb{R}^{m\times n}:$$

$$z_{i,j} = q_{i,j-1} - q_{i,j}, i = 1,\cdots,m, j = 1,\cdots,n, \tag{13}$$

respectively, where we assume a zero boundary condition for $\mathbf{p}$ and $\mathbf{q}$:

$$\begin{array}{ll} p_{0,j} = p_{m,j} = 0, & j = 1,\cdots,n; \\ q_{i,0} = q_{i,n} = 0, & i = 1,\cdots,m. \end{array} \tag{14}$$

If we rearrange $\mathbf{x}$, $\mathbf{p}$ and $\mathbf{q}$ row by row into vectors, still written as $\mathbf{x}$, $\mathbf{p}$ and $\mathbf{q}$ in order to save notations[1], respectively, then there are matrices $\mathbf{D}_v$ and $\mathbf{D}_h$ such that $\mathbf{p} = \mathbf{D}_v\mathbf{x}$ and $\mathbf{q} = \mathbf{D}_h\mathbf{x}$. Accordingly, the adjoint operators $\mathcal{D}_v^*$ and $\mathcal{D}_h^*$ are associated with matrices $\mathbf{D}_v^T$ and $\mathbf{D}_h^T$, respectively, where the superscript T denotes the transpose.

B. The Original APG Method

While the traditional first-order methods for differentiable objective functions can only achieve a convergence rate of $O(k^{-1})$, Nesterov [31] showed that for convex objective functions with Lipschitz continuous gradients:

$$\|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\|_F \leq L_f \|\mathbf{x}_1 - \mathbf{x}_2\|_F, \tag{15}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, the convergence rate can be dramatically improved to $O(k^{-2})$. Beck and Teboulle [12] further generalized Nesterov's method to the following type of objective functions:

$$F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \tag{16}$$

where g is a convex function, which may not be differentiable, and f is a smooth and convex function with Lipschitz continuous gradients. The method, called APG, can be described as follows.

Instead of directly minimizing $F(\mathbf{x})$, APG finds the optimal solution by minimizing a sequence of quadratic approximations, denoted as $Q(\mathbf{x}, \mathbf{y})$, of $F(\mathbf{x})$ at specially chosen points $\mathbf{y}$:

$$Q(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{L_f}{2} \|\mathbf{x} - \mathbf{y}\|_F^2 + g(\mathbf{x}). \tag{17}$$

For any $\mathbf{y}$, $Q(\mathbf{x}, \mathbf{y})$ upper bounds $F(\mathbf{x})$ [12]. Then $\mathbf{x}$ can be updated as the unique minimizer of $Q(\mathbf{x}, \mathbf{y})$:

$$\begin{aligned} \mathbf{x}_k &= \operatorname*{argmin}_{\mathbf{x}} Q(\mathbf{x}, \mathbf{y}_k) \\ &= \operatorname*{argmin}_{\mathbf{x}} \left\{ g(\mathbf{x}) + \tfrac{L_f}{2} \|\mathbf{x} - \mathbf{z}_k\|_F^2 \right\}, \end{aligned} \tag{18}$$

where $\mathbf{z}_k = \mathbf{y}_k - \frac{1}{L_f}\nabla f(\mathbf{y}_k)$.

One natural choice of the point $\mathbf{y}_k$ is $\mathbf{x}_k$, which is adopted in the iterative shrinkage/thresholding (IST) algorithm [15]. The convergence rate of such an update scheme is no worse than $O(k^{-1})$ [20] but no theoretical analysis can guarantee that higher convergence can be achieved.

In the smooth case, i.e., $g(\mathbf{x}) \equiv 0$, Nesterov [31] showed that the convergence rate can be improved to $O(k^{-2})$ by choosing

$$\mathbf{y}_{k+1} = \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1}), \tag{19}$$

for a sequence $\{t_k\}$ satisfying $t_{k+1}^2 - t_{k+1} = t_k^2$. In the nonsmooth case, i.e., $g(\mathbf{x}) \not\equiv 0$, Beck and Teboulle [12] proved that the updating schemes (18) and (19) are still valid to ensure the $O(k^{-2})$ convergence rate. More precisely, they proved the following theorem.

[1]In this paper, we use the same bold face small letter to denote an image and its vectorization. We believe that this should not cause ambiguity by referring to the context.

Theorem 1: Let $\{\mathbf{x}_k\}$ be generated by the APG method and $\mathbf{x}^*$ be any optimal solution, then

$$F(\mathbf{x}_k) - F(\mathbf{x}^*) \leq \frac{2L_f\|\mathbf{x}_0 - \mathbf{x}^*\|_F^2}{(k+1)^2}, \quad \forall k \geq 1. \tag{20}$$

Algorithm 1 summarizes the APG method.

Algorithm 1 The Original Accelerated Proximal Gradient Algorithm

1: **while** not converged **do**
2: $\mathbf{z}_k \leftarrow \mathbf{y}_k - \frac{1}{L_f}\nabla f(\mathbf{y}_k)$,
3: $\mathbf{x}_k \leftarrow \arg\min_{\mathbf{x}} \left\{ g(\mathbf{x}) + \frac{L_f}{2} \|\mathbf{x} - \mathbf{z}_k\|_F^2 \right\}$,
4: $t_{k+1} \leftarrow \frac{1+\sqrt{1+4t_k^2}}{2}$,
5: $\mathbf{y}_{k+1} \leftarrow \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1})$,
6: $k \leftarrow k + 1$.
7: **end while**

C. Solving the Subproblem in APG

One may notice that when $g \not\equiv 0$ the APG method involves a subproblem like

$$\min_{\mathbf{x}} \left\{ \varepsilon g(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{w}\|_F^2 \right\}. \tag{21}$$

For a general convex g, such a subproblem may not be simpler than the original problem. Fortunately, for many problems, e.g., our image restoration problems, g is often special (usually a norm or a characteristic function) and there may exist a closed-form solution to this subproblem, enabling the APG method to be practical.

When g is a norm, as in this paper we will only use the l_1-norm and the $l_{2,1}$-norm, below we only give the closed-form solutions to subproblems with such special g's.

Proposition 1: When $g(\mathbf{x}) = \|\mathbf{x}\|_1$, then the solution to problem (21) is:

$$\mathbf{x} = \mathcal{T}_\varepsilon(\mathbf{w}), \tag{22}$$

where $\mathcal{T}_\varepsilon(w) = \max(|w| - \varepsilon, 0)\operatorname{sgn}(w)$ is the shrinkage operator and $\mathcal{T}_\varepsilon(\mathbf{w})$ applies the shrinkage operator to $\mathbf{w}$ entrywise.

This result is well known in the literature, e.g., [4], [5].

Proposition 2: When $g(\mathbf{x}) = \|\mathbf{x}\|_{2,1} = \sum_{j=1}^n \|\mathbf{x}_j\|_2$, then the j-th column of the solution is:

$$\mathbf{x}_j = \mathcal{T}_\varepsilon(\|\mathbf{w}_j\|_2)\frac{\mathbf{w}_j}{\|\mathbf{w}_j\|_2}. \tag{23}$$

This result can be found as Lemma 3.3 in [29].

When g is a characteristic function $\chi_C(\mathbf{x})$ of a convex set C:

$$\chi_C(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in C, \\ +\infty, & \text{otherwise,} \end{cases}$$

the closed-form solution to the subproblem is simply projecting $\mathbf{w}$ onto C:

$$\mathbf{x} = \mathcal{P}_C(\mathbf{w}), \tag{24}$$

where $\mathcal{P}$ is the projection operator.

D. Previous Work

In this section, we review some related work on TV-based image restoration.

1) TV-based Image Denoising: TV-based image denoising is a special case of image restoration where the linear operator $\mathcal{A}$ is an identity matrix. Denote $\Psi_\lambda(\mathbf{b})$ as the image denoising operator:

$$\Psi_\lambda(\mathbf{b}) = \underset{\mathbf{x}\in B_{l,u}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{x}-\mathbf{b}\|_F^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{25}$$

where the TV regularizer can be either isotropic or anisotropic. The nonsmoothness of the TV regularizer makes it very difficult to solve the image denoising problem directly. Chambolle [8] showed that the dual problem of TV-based image denoising is a convex quadric programming problem and proposed a gradient-based method for solving the dual problem. By incorporating the boundedness constraints into the model Beck and Tebouble [21] proposed a fast gradient projection method to solve the dual problem. Goldfarb and Yin [25] also reported a parametric maximum flow algorithm to solve the image denoising problem with anisotropic TV.

The image denoising operator $\Psi_\lambda(\mathbf{b})$ is essential for several algorithms for general TV-based image restoration problems. For example, the iterative shrinkage/thresholding (IST) algorithm [15] for image restoration has the form

$$\mathbf{x}_{k+1} = \Psi_\lambda\left\{\mathbf{x}_k - \mathcal{A}^*(\mathcal{A}(\mathbf{x}_k)-\mathbf{b})\right\}, \tag{26}$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$ and the denoising problem is usually solved by Chambolle's dual approach [8].

2) Two-Step Iterative Shrinkage/Thresholding: The convergence of IST is known to be very slow [20]. Bioucas-Dias and Figueredo [14] proposed a two-step iterative shrinkage/thresholding (TwIST) method to improve the convergence rate of IST, where $\mathbf{x}_{k+1}$ is updated by involving both $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$:

$$\begin{aligned}\mathbf{x}_{k+1} &= (1-\alpha)\mathbf{x}_{k-1} + (\alpha-\beta)\mathbf{x}_k \\ &\quad +\Psi_\lambda\left\{\mathbf{x}_k - \mathcal{A}^*(\mathcal{A}(\mathbf{x}_k)-\mathbf{b})\right\},\end{aligned} \tag{27}$$

where α and β are two parameters. Bioucas-Dias and Figueredo [14] also proved the convergence of TwIST. From (27), a TV-based denoising problem should be solved for each iteration of TwIST. In real applications, this subproblem can be solved only approximately, resulting in non-monotonic decrease of the objective function value. To remedy this issue, Bioucas-Dias and Figueredo [14] further suggested a monotone version of TwIST (MTwIST). Let $\mathbf{z}$ be given by the right hand side of (27), then MTwIST updates $\mathbf{x}_{k+1}$ as

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{z}, & \text{if } F(\mathbf{z}) \le F(\mathbf{x}_k), \\ \Psi_\lambda\{\mathbf{x_k} - \mathcal{A}^*(\mathcal{A}(\mathbf{x}_k)-\mathbf{b})\}, & \text{if } F(\mathbf{z}) > F(\mathbf{x}_k). \end{cases} \tag{28}$$

3) Fast Iterative Shrinkage/Thresholding Algorithm: Most recently, Beck and Teboulle [21] proposed another fast two-step method, called FISTA, for TV-based image restoration. It applies the APG method to solve both the outer and inner iterations, where the inner iterations solve an image denoising subproblem via duality. Analogous to TwIST [14], FISTA also has the non-monotonicity problem when the image denoising subproblem is solved only approximately. So Beck and Teboulle [21] also suggested a monotone version of FISTA (MFISTA) to guarantee the monotonicity property of the minimization algorithm.

III. The Generalized Accelerated Proximal Gradient Method

In this section, we propose a generalized accelerated proximal gradient (GAPG) method to solve the minimization problem with an objective function in (16). We notice that the following inequality ((2.7) of [12])

$$f(\mathbf{x}) \le f(\mathbf{y}) + \langle \mathbf{x}-\mathbf{y}, \nabla f(\mathbf{y})\rangle + \frac{L_f}{2}\|\mathbf{x}-\mathbf{y}\|_F^2, \forall \mathbf{x},\mathbf{y}, \tag{29}$$

is the key to proving the $O(k^{-2})$ convergence rate of the original APG method, and the Lipschitz gradient condition (15) is just to ensure that (29) holds.

Given a positive definite matrix $\mathbf{L}$, we may introduce the $\mathbf{L}$-inner product: $\langle \mathbf{x},\mathbf{y}\rangle_{\mathbf{L}} = \mathbf{x}^T\mathbf{L}\mathbf{y}$ and the $\mathbf{L}$-norm $\|\mathbf{x}\|_{\mathbf{L}} = \sqrt{\langle \mathbf{x},\mathbf{x}\rangle_{\mathbf{L}}}$. Then the inequality (29) can be generalized as

$$f(\mathbf{x}) \le f(\mathbf{y}) + \langle \mathbf{x}-\mathbf{y}, \nabla f(\mathbf{y})\rangle + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_{\mathbf{L}_f}^2, \forall \mathbf{x},\mathbf{y}, \tag{30}$$

where $\mathbf{L}_f$ is a positive definite matrix. Such $\mathbf{L}_f$ exists for a broad class of f. For example, if f satisfies the Lipschitz gradient condition (15), then $\mathbf{L}_f$ can be chosen as $L_f\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. We will give other choices of $\mathbf{L}_f$ in our image restoration problems. The motivation to replace the inequality (29) with (30) is that a smaller Lipschitz constant may lead to faster convergence (c.f. (20)).

Analogously, GAPG updates $\mathbf{x}$ by minimizing a quadratic approximation $Q_{\mathbf{L}_f}(\mathbf{x},\mathbf{y})$ of $F(\mathbf{x})$ at specially chosen points $\mathbf{y}$, where

$$Q_{\mathbf{L}_f}(\mathbf{x},\mathbf{y}) = f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_{\mathbf{L}_f}^2 + g(\mathbf{x}) \tag{31}$$

and $\mathbf{y}_k$ is still chosen as (19). For the subproblem $p_{\mathbf{L}_f}(\mathbf{y}_k) = \operatorname{argmin}_{\mathbf{x}} Q_{\mathbf{L}_f}(\mathbf{x},\mathbf{y}_k)$ to be easy to solve, we usually choose a diagonal $\mathbf{L}_f$. Then we straightforwardly generalize the original APG method with a constant step size [12] by the GAPG algorithm as in Algorithm 2.

Algorithm 2 The Generalized Accelerated Proximal Gradient Algorithm

1: **while** not converged **do**
2: $\quad \mathbf{x}_k \leftarrow p_{\mathbf{L}_f}(\mathbf{y}_k)$,
3: $\quad t_{k+1} \leftarrow \frac{1+\sqrt{1+4t_k^2}}{2}$,
4: $\quad \mathbf{y}_{k+1} \leftarrow \mathbf{x}_k + \frac{t_k-1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1})$,
5: $\quad k \leftarrow k+1$.
6: **end while**

By replacing $L_f\langle\cdot,\cdot\rangle$ and $L_f\|\cdot\|_F^2$ in the proofs in [12] with $\langle\cdot,\cdot\rangle_{\mathbf{L}_f}$ and $\|\cdot\|_{\mathbf{L}_f}^2$, respectively, we can easily prove the $O(k^{-2})$ convergence rate of GAPG as stated in the following theorem.

Theorem 2: Let $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ be generated by GAPG. Then for any $k \geq 1$,

$$F(\mathbf{x}_k) - F(\mathbf{x}^*) \leq \frac{2\left\|\mathbf{x}_0 - \mathbf{x}^*\right\|_{\mathbf{L}_f}^2}{(k+1)^2}, \tag{32}$$

where $\mathbf{x}^*$ is any optimal solution.

One can easily see that if $\mathbf{L}_f = L_f\mathbf{I}$, then GAPG reduces to the original APG. However, as one will see $\mathbf{L}_f$ can have other choices such that (30) holds and $\|\mathbf{x}\|_{\mathbf{L}_f} \leq L_f\|\mathbf{x}\|_F$ for any $\mathbf{x}$, resulting in faster convergence.

IV. Anisotropic TV-based Image Restoration

As anisotropic TV-based image restoration is simpler than the isotropic case, we introduce our algorithm for it first. In this section, we first formulate the problem such that it can be easily solved by GAPG, then present the details of the solution by GAPG, and finally summarize the algorithm.

A. Problem Formalization

The anisotropic TV-based image restoration problem is formalized as

$$\min_{\mathbf{x}\in B_{l,u}} \frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_F^2 + \lambda \mathrm{TV}_{aniso}(\mathbf{x}), \tag{33}$$

where $\mathbf{A}$ is associated with the linear mapping $\mathcal{A}$ when $\mathbf{x}$ and $\mathbf{b}$ are treated as vectors, and $\mathrm{TV}_{aniso}(\mathbf{x})$ is defined as (5). By introducing two auxiliary variables $\mathbf{d}_v$ and $\mathbf{d}_h$ [27], [28], [29], and using (11), the problem can be written as

$$\begin{array}{c}\min\limits_{\mathbf{x}\in B_{l,u},\mathbf{d}_v,\mathbf{d}_h} \frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_F^2 + \lambda\left(\|\mathbf{d}_v\|_1 + \|\mathbf{d}_h\|_1\right), \\ s.t.\ \mathbf{d}_v = \mathbf{D}_v\mathbf{x} \text{ and } \mathbf{d}_h = \mathbf{D}_h\mathbf{x}.\end{array} \tag{34}$$

Like [27], [28], [29], [22], we further relax the equality constraint by changing (34) to

$$\begin{array}{l}\min\limits_{\mathbf{x}\in B_{l,u},\mathbf{d}_v,\mathbf{d}_h} \frac{\mu}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_F^2 + \frac{1}{2}\left\|\mathbf{d}_v-\mathbf{D}_v\mathbf{x}\right\|_F^2 \\ \qquad + \frac{1}{2}\left\|\mathbf{d}_h-\mathbf{D}_h\mathbf{x}\right\|_F^2 + \lambda\mu\left(\|\mathbf{d}_v\|_1+\|\mathbf{d}_h\|_1\right).\end{array} \tag{35}$$

The above can be further rewritten as

$$\begin{array}{l}\min\limits_{\mathbf{x},\mathbf{d}_v,\mathbf{d}_h} \frac{\mu}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_F^2 + \frac{1}{2}\left\|\mathbf{d}_v-\mathbf{D}_v\mathbf{x}\right\|_F^2 + \frac{1}{2}\left\|\mathbf{d}_h-\mathbf{D}_h\mathbf{x}\right\|_F^2 \\ \qquad + \lambda\mu(\|\mathbf{d}_v\|_1 + \|\mathbf{d}_h\|_1 + \chi_{B_{l,u}}(\mathbf{x})),\end{array} \tag{36}$$

and be solved by using the GAPG approach. When the relaxation parameter $\mu \to 0$, the solution to (36) approaches the solution to (33). Compared with the original problem formalization (33), the formalization (36) is more suitable to be solved by GAPG-based methods. With the formalization (36), the TV-based image denoising subproblem (25) in MTwIST and MFISTA is now unnecessary.

B. Anisotropic TV-based Image Restoration via GAPG

To utilize GAPG for anisotropic TV-based image restoration, we readily see that the decomposition in (16) should be chosen as

$$\begin{array}{rl} & f(\hat{\mathbf{x}}) \\ = & \frac{\mu}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|_F^2 + \frac{1}{2}\left\|\mathbf{d}_v-\mathbf{D}_v\mathbf{x}\right\|_F^2 + \frac{1}{2}\left\|\mathbf{d}_h-\mathbf{D}_h\mathbf{x}\right\|_F^2, \\ & g(\hat{\mathbf{x}}) \\ = & \lambda\mu\left(\chi_{B_{l,u}}(\mathbf{x}) + \|\mathbf{d}_v\|_1 + \|\mathbf{d}_h\|_1\right),\end{array} \tag{37}$$

where $\hat{\mathbf{x}} = (\mathbf{x}^T, \mathbf{d}_v^T, \mathbf{d}_h^T)^T$.

1) Choosing $\mathbf{L}_f$: Note that as f is a quadratic function, (30) can be an identity (the $\mathbf{x}$ therein should be replaced with $\hat{\mathbf{x}}$) with $\mathbf{L}_f = \mathbf{H}(f)$, where $\mathbf{H}(f)$ is the Hessian of f:

$$\mathbf{H}(f) = \begin{pmatrix} \mu\mathbf{A}^T\mathbf{A} + \mathbf{D}_v^T\mathbf{D}_v + \mathbf{D}_h^T\mathbf{D}_h & -\mathbf{D}_v^T & -\mathbf{D}_h^T \\ -\mathbf{D}_v & \mathbf{I} & \mathbf{0} \\ -\mathbf{D}_h & \mathbf{0} & \mathbf{I} \end{pmatrix}. \tag{38}$$

As $\mathbf{H}(f)$ is not diagonal, which may result in expensive computation to solve the subproblem like (21) in GAPG, we had better seek a diagonal $\mathbf{L}_f$ such that (30) still holds.

Observe that

$$\begin{array}{rl} & \hat{\mathbf{x}}^T\mathbf{H}(f)\hat{\mathbf{x}} \\ = & \mu\|\mathbf{A}\mathbf{x}\|^2 + \|\mathbf{D}_h\mathbf{x} - \mathbf{d}_h\|^2 + \|\mathbf{D}_v\mathbf{x} - \mathbf{d}_v\|^2 \\ \leq & \mu\|\mathbf{A}\mathbf{x}\|^2 + \eta\left(\|\mathbf{D}_h\mathbf{x}\|^2 + \|\mathbf{D}_v\mathbf{x}\|^2 + \|\mathbf{d}_v\|^2 + \|\mathbf{d}_h\|^2\right) \\ \leq & \lambda_{\max}\|\mathbf{x}\|_F^2 + \eta\|\mathbf{d}_v\|^2 + \eta\|\mathbf{d}_h\|^2,\end{array} \tag{39}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mu\mathbf{A}^T\mathbf{A} + \eta\mathbf{D}_v^T\mathbf{D}_v + \eta\mathbf{D}_h^T\mathbf{D}_h$ and the above inequality is always true if $\eta \geq 2$. Then we see that a diagonal $\mathbf{L}_f$ can be chosen as:

$$\mathbf{L}_f = \mathrm{diag}\left(\lambda_{\max}\mathbf{I}, \eta\mathbf{I}, \eta\mathbf{I}\right). \tag{40}$$

$\lambda_{\max}$ can be conveniently estimated as $\|\mu\mathbf{A}^T\mathbf{A} + \eta\mathbf{D}_v^T\mathbf{D}_v + \eta\mathbf{D}_h^T\mathbf{D}_h\|_F$ due to the fact that $\lambda_{\max}(\mathbf{X}) \leq \|\mathbf{X}\|_F$ holds for any nonnegative definite matrix $\mathbf{X}$. However, this is a severe overestimate when the size of image is large. So we provide a sharper estimate below.

Note that $\lambda_{\max}$ is actually the square of the largest singular value of

$$\mathbf{W} = \begin{pmatrix} \sqrt{\mu}\mathbf{A} \\ \sqrt{\eta}\mathbf{D}_v \\ \sqrt{\eta}\mathbf{D}_h \end{pmatrix}.$$

And we have the following inequalities.

Proposition 3:

$$\left\|\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\right\|_2 \leq \|\mathbf{X}\|_2 + \|\mathbf{Y}\|_2, \text{ where } \|\cdot\|_2 \text{ is the spectral norm.}$$

Proof:

$$\begin{array}{rcl} \left\|\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\right\|_2 & = & \left\|\begin{pmatrix}\mathbf{X}\\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\mathbf{0}\\ \mathbf{Y}\end{pmatrix}\right\|_2 \\ & \leq & \left\|\begin{pmatrix}\mathbf{X}\\ \mathbf{0}\end{pmatrix}\right\|_2 + \left\|\begin{pmatrix}\mathbf{0}\\ \mathbf{Y}\end{pmatrix}\right\|_2 \\ & = & \|\mathbf{X}\|_2 + \|\mathbf{Y}\|_2.\end{array}$$

■

Proposition 4: If matrix $\mathbf{K}$ is associated with a shift-invariant convolution operator $\mathcal{K}$ that applies a convolution kernel $\mathbf{k}$, then

$$\|\mathbf{K}\|_2 \le \|\mathbf{k}\|_{l_1}. \tag{41}$$

Proof: It is well known in matrix analysis [32] that

$$\|\mathbf{X}\|_2 \le \sqrt{\|\mathbf{X}\|_1\|\mathbf{X}\|_\infty}, \tag{42}$$

where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are the largest l_1-norm of the columns and rows of $\mathbf{X}$, respectively.

Suppose $\mathbf{Y} = \mathcal{K}(\mathbf{X})$. Then

$$\mathbf{Y}_{ij} = \sum_{p,q} \mathbf{X}_{pq}\mathbf{k}_{i-p,j-q},$$

where the indices $(i-p, j-q)$ of $\mathbf{k}$ should be understood as modulo the size of kernel. So the l_1-norm of the row of matrix $\mathbf{K}$ that corresponds to the (i,j)-th entry of $\mathbf{Y}$ is

$$\sum_{p,q} |\mathbf{k}_{i-p,j-q}| = \|\mathbf{k}\|_{l_1}.$$

Hence $\|\mathbf{K}\|_\infty = \|\mathbf{k}\|_{l_1}$.

On the other hand, the adjoint operator $\mathcal{K}^* : \mathbf{Y} \mapsto \mathbf{X}$ is as follows:

$$\mathbf{X}_{ij} = \sum_{p,q} \mathbf{Y}_{pq}\mathbf{k}_{p-i,q-j}.$$

So the l_1-norm of the row of matrix $\mathbf{K}^T$, which is the same as the l_1-norm of the corresponding column of matrix $\mathbf{K}$, that corresponds to the (i,j)-th entry of $\mathbf{X}$ is

$$\sum_{p,q} |\mathbf{k}_{p-i,q-j}| = \|\mathbf{k}\|_{l_1}.$$

Hence $\|\mathbf{K}\|_1 = \|\mathbf{k}\|_{l_1}$.

Thus the inequality is proven. ■

In particular, we have

Corollary 1:

$$\|\mathbf{D}_v\|_2 \le 2 \text{ and } \|\mathbf{D}_h\|_2 \le 2,$$

and if a nonnegative convolution kernel satisfies $\sum_{ij}\mathbf{k}_{ij} = 1$, then

$$\|\mathbf{K}\|_2 \le 1.$$

Based on the above analysis we can estimate $\lambda_{\max}$ more precisely as:

$$\lambda_{\max} \quad \le \quad (\sqrt{\mu}\|\mathbf{A}\|_2 + \sqrt{\eta}\|\mathbf{D}_v\|_2 + \sqrt{\eta}\|\mathbf{D}_h\|_2)^2, \tag{43}$$

where $\|\mathbf{A}\|_2$ can be estimated as

$$\|\mathbf{A}\|_2 \le \min(\|\mathbf{A}\|_F, \sqrt{\|\mathbf{A}\|_1\|\mathbf{A}\|_\infty}), \tag{44}$$

due to the fact that $\|\mathbf{X}\|_2 \le \|\mathbf{X}\|_F$ [32] and inequality (42). If $\mathbf{A}$ is associated with a shift-invariant blurring kernel, the estimate on $\|\mathbf{A}\|_2$ can be dramatically improved as $\|\mathbf{A}\|_2 \le 1$ thank to the second part of Corollary 1. Moreover, since the auxiliary variables $\mathbf{d}_v$ and $\mathbf{d}_h$ approximate $\mathbf{D}_v\mathbf{x}$ and $\mathbf{D}_h\mathbf{x}$, in our image restoration problems, we may also choose $\eta = 1$. By using smaller $\lambda_{\max}$ and η, which are the "Lipschitz constants", our GAPG will converge faster.

Here we would like to highlight the difference between the original APG and our GAPG. The original APG chooses a single "Lipschitz constant" L_f for all variables of the function f, while GAPG allows different "Lipschitz constants" for different variables. In the original APG, for the inequality (29) to be true, L_f has to be chosen as the largest "Lipschitz constant" of all variables. In our GAPG, this is unnecessary. So by comparing (32) and (20) GAPG is expected to converge faster than the original APG.

2) Solving the Subproblem: Then the subproblem of GAPG for anisotropic TV-based image restoration is:

$$\begin{aligned}\min_{\hat{\mathbf{x}}} \quad & \lambda\mu(\chi_{B_{l,u}}(\mathbf{x}) + \|\mathbf{d}_v\|_1 + \|\mathbf{d}_h\|_1) \\ & + f(\hat{\mathbf{y}}_k) + \langle \mathbf{x} - \mathbf{y}_{\mathbf{x}}^k, \mathbf{A}^T(\mathbf{A}\mathbf{y}_{\mathbf{x}}^k - \mathbf{b}) \\ & + \mathbf{D}_v^T(\mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_v}^k) + \mathbf{D}_h^T(\mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_h}^k)\rangle \\ & + \langle \mathbf{d}_v - \mathbf{y}_{\mathbf{d}_v}^k, \mathbf{y}_{\mathbf{d}_v}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k\rangle \\ & + \langle \mathbf{d}_h - \mathbf{y}_{\mathbf{d}_h}^k, \mathbf{y}_{\mathbf{d}_h}^k - \mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k\rangle \\ & + \frac{\lambda_{\max}}{2}\|\mathbf{x} - \mathbf{y}_{\mathbf{x}}^k\|_F^2 \\ & + \frac{\eta}{2}\|\mathbf{d}_v - \mathbf{y}_{\mathbf{d}_v}^k\|_F^2 + \frac{\eta}{2}\|\mathbf{d}_h - \mathbf{y}_{\mathbf{d}_h}^k\|_F^2,\end{aligned} \tag{45}$$

where $\hat{\mathbf{y}}_k = ((\mathbf{y}_{\mathbf{x}}^k)^T, (\mathbf{y}_{\mathbf{d}_v}^k)^T, (\mathbf{y}_{\mathbf{d}_h}^k)^T)^T$.

With a little algebra we can see that problem (45) can be decomposed into three independent smaller subproblems:

$$\begin{aligned}\min_{\mathbf{x}} \quad & \lambda\mu\chi_{B_{l,u}}(\mathbf{x}) \\ & + \frac{\lambda_{\max}}{2}\|\mathbf{x} - \{\mathbf{y}_{\mathbf{x}}^k - \lambda_{\max}^{-1}[\mathbf{A}^T(\mathbf{A}\mathbf{y}_{\mathbf{x}}^k - \mathbf{b}) \\ & + \mathbf{D}_v^T(\mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_v}^k) + \mathbf{D}_h^T(\mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_h}^k)]\}\|_F^2,\end{aligned} \tag{46}$$

$$\begin{aligned}\min_{\mathbf{d}_v} \quad & \lambda\mu\|\mathbf{d}_v\|_1 \\ & + \frac{\eta}{2}\|\mathbf{d}_v - [\mathbf{y}_{\mathbf{d}_v}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_v}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k)]\|_F^2,\end{aligned} \tag{47}$$

$$\begin{aligned}\min_{\mathbf{d}_h} \quad & \lambda\mu\|\mathbf{d}_h\|_1 \\ & + \frac{\eta}{2}\|\mathbf{d}_h - [\mathbf{y}_{\mathbf{d}_h}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_h}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k)]\|_F^2.\end{aligned} \tag{48}$$

Moreover, in every subproblem, the unknowns can be solved *independently* due to the separable nature of the subproblems. This property makes our GAPG algorithm *highly parallelizable*. Using the results in Section II-C, we can easily have

$$(\mathbf{x}_k)_{i,j} = \begin{cases} l, & \text{if } (\tilde{\mathbf{x}}_k)_{i,j} < l, \\ (\tilde{\mathbf{x}}_k)_{i,j}, & \text{if } l \le (\tilde{\mathbf{x}}_k)_{i,j} \le u, \\ u, & \text{if } (\tilde{\mathbf{x}}_k)_{i,j} > u, \end{cases} \tag{49}$$

$$\begin{aligned}(\mathbf{d}_v^k)_{i,j} &= \mathcal{T}_{\lambda\mu/\eta}((\tilde{\mathbf{d}}_v^k)_{i,j}), \\ (\mathbf{d}_h^k)_{i,j} &= \mathcal{T}_{\lambda\mu/\eta}((\tilde{\mathbf{d}}_h^k)_{i,j}),\end{aligned} \tag{50}$$

where

$$\begin{aligned}\tilde{\mathbf{x}}_k &= \mathbf{y}_{\mathbf{x}}^k - \lambda_{\max}^{-1}[\mathbf{A}^T(\mathbf{A}\mathbf{y}_{\mathbf{x}}^k - \mathbf{b}) \\ &\quad + \mathbf{D}_v^T(\mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_v}^k) + \mathbf{D}_h^T(\mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_h}^k)], \\ \tilde{\mathbf{d}}_v^k &= \mathbf{y}_{\mathbf{d}_v}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_v}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k), \\ \tilde{\mathbf{d}}_h^k &= \mathbf{y}_{\mathbf{d}_h}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_h}^k - \mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k).\end{aligned} \tag{51}$$

3) Continuation Technique: The relaxation parameter μ has to be chosen very small so that the solution to (36) is close enough to that to (33). However, a small μ will yield bad-conditioned problems and make the convergence slow[2]. To remedy this issue, the continuation technique has been

[2]This is because the thresholding operators in (50) will be ineffective.

adopted in [27][29][30] by setting a much larger initial value μ_0 of μ and gradually decreasing its value when the iteration goes on. The continuation technique has also been applied to other optimization problems, e.g., matrix completion [24] and robust principal component analysis [22]. Another way to tackle this issue is to jointly use the variable splitting and augmented Lagrangian method (ALM), specifically the alternating direction method of multipliers (ADMM) [26] [33] [34]. Other splitting methods, e.g., split-Bregman [28] and Douglas-Rachford Splitting (DRS) [35], have also been developed to address this difficulty. Actually, split-Bregman with a single inner iteration is equivalent to ADMM [36]. Please see [36] and [37] for more discussions on the relationship between ALM, split-Bregman, and DRS.

In our implementation, we adopt the continuation technique because it can be naturally integrated with the GAPG algorithm. We set $\mu_0 = \|\mathbf{b}\|_F$ and let it decrease gradually until the target value $\bar{\mu} = \delta\mu_0$ is reached. We empirically choose $\delta = 10^{-3}$ and decrease μ by

$$\mu_{k+1} = \max\left(\left(0.9 + 0.1\left(\frac{2k-2}{2k-1}\right)^{\frac{1}{8}}\right)\mu_k, \bar{\mu}\right).$$

When μ decreases, $\lambda_{\max}$ may also be reduced according to (43). Our numerical results show that such heuristic rules are sufficient for most applications.

C. Algorithm

We now summarize our GAPG-based anisotropic TV-based image restoration algorithm as Algorithm 3 [3]. Note that in real computation the matrix-vector products in Algorithm 3 should be replaced by operators acting on images. For example, $\mathbf{A}^T(\mathbf{A}\mathbf{y}_{\mathbf{x}}^k - \mathbf{b})$ is actually computed as $\mathcal{A}^*(\mathcal{A}(\mathbf{y}_{\mathbf{x}}^k) - \mathbf{b})$.

Compared with MTwIST [14] and MFISTA [21], the proposed algorithm is significantly faster with better or at least comparable restoration quality. The proposed algorithm is more suitable for fast image restoration thanks to three techniques. The first is the problem formalization (36) that involves auxiliary variables to decouple the problem, with which we need not solve the TV-based denoising problem (25) in each iteration. The second is the GAPG method, which is expected to be more efficient than the original APG by choosing an appropriate matrix $\mathbf{L}_f$ such that different variables can have different "Lipschitz constants". The third is the continuation technique to provide good initial solutions for our GAPG method.

It is also worth noting that both MFISTA and MTwIST need to evaluate the objective function values during the iterations in order to ensure that the objective function values decrease with iterations (c.f. (28)). In contrast, our GAPG algorithm does not require such extra computation.

V. Isotropic TV-based Image Restoration

In this section, we introduce our algorithm for isotropic TV-based image restoration. As it is very similar to the case of anisotropic TV, we only sketch some key differences in the problem formulation and the subproblems.

[3]To terminate the iteration, one may check whether the difference in the objective function value or the iterative solution is below a threshold.

Algorithm 3 Anisotropic TV-based Image Restoration via GAPG

Input: Observed image $\mathbf{b}$, λ, μ_0, $\mathbf{A}$, $\mathbf{D}_v$, and $\mathbf{D}_h$
Output: $\mathbf{x} \leftarrow \mathbf{x}_k$.

1: $\mathbf{x}_0 \leftarrow \mathcal{P}_{B_{l,u}}(\mathbf{b})$, $\mathbf{d}_v^0 \leftarrow \mathbf{D}_v\mathbf{x}_0$, $\mathbf{d}_h^0 \leftarrow \mathbf{D}_h\mathbf{x}_0$, $t_0 \leftarrow 1$, $\bar{\mu} \leftarrow \delta\mu_0$, $\mathbf{y}_{\mathbf{x}}^1 \leftarrow \mathbf{x}_0$, $\mathbf{y}_{\mathbf{d}_v}^1 \leftarrow \mathbf{d}_v^0$, $\mathbf{y}_{\mathbf{d}_h}^1 \leftarrow \mathbf{d}_h^0$, $k \leftarrow 1$.
2: **while** not converged **do**
3: // Line 4 solves subproblem (46).
4: $\mathbf{x}_k \leftarrow \mathcal{P}_{B_{l,u}}\{\mathbf{y}_{\mathbf{x}}^k - \lambda_{\max}^{-1}[\mathbf{A}^T(\mathbf{A}\mathbf{y}_{\mathbf{x}}^k - \mathbf{b}) + \mathbf{D}_v^T(\mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_v}^k) + \mathbf{D}_h^T(\mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k - \mathbf{y}_{\mathbf{d}_h}^k)]\}$,
5: // Lines 6-7 solve subproblems (47) and (48).
6: $\mathbf{d}_v^k \leftarrow \mathcal{T}_{\lambda\mu_k/\eta}(\mathbf{y}_{\mathbf{d}_v}^k - \frac{1}{\eta}(\mathbf{y}_{\mathbf{d}_v}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k))$,
7: $\mathbf{d}_h^k \leftarrow \mathcal{T}_{\lambda\mu_k/\eta}(\mathbf{y}_{\mathbf{d}_h}^k - \frac{1}{\eta}(\mathbf{y}_{\mathbf{d}_h}^k - \mathbf{D}_h\mathbf{y}_{\mathbf{x}}^k))$,
8: $t_{k+1} \leftarrow \frac{1+\sqrt{1+4(t_k)^2}}{2}$,
9: // Lines 10-12 update $\mathbf{y}_{\mathbf{x}}^{k+1}, \mathbf{y}_{\mathbf{d}_v}^{k+1}, \mathbf{y}_{\mathbf{d}_h}^{k+1}$.
10: $\mathbf{y}_{\mathbf{x}}^{k+1} \leftarrow \mathbf{x}_k + \frac{t_k-1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1})$,
11: $\mathbf{y}_{\mathbf{d}_v}^{k+1} \leftarrow \mathbf{d}_v^k + \frac{t_k-1}{t_{k+1}}(\mathbf{d}_v^k - \mathbf{d}_v^{k-1})$,
12: $\mathbf{y}_{\mathbf{d}_h}^{k+1} \leftarrow \mathbf{d}_h^k + \frac{t_k-1}{t_{k+1}}(\mathbf{d}_h^k - \mathbf{d}_h^{k-1})$,
13: Update μ_k to μ_{k+1},
14: $k \leftarrow k+1$
15: **end while**

A. Problem Formalization

The isotropic TV-based image restoration problem is formalized as

$$\min_{\mathbf{x}\in B_{l,u}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_F^2 + \lambda \mathrm{TV}_{iso}(\mathbf{x}), \tag{52}$$

where $\mathrm{TV}_{iso}(\mathbf{x})$ is defined as (4). Like the anisotropic case, by introducing two new variables $\mathbf{d}_v$ and $\mathbf{d}_h$ [28], imposing the boundedness constraints on $\mathbf{d}_v$ and $\mathbf{d}_h$ and using isotropic TV induced norm (9), the problem can be rewritten as:

$$\begin{aligned}\min_{\mathbf{x},\mathbf{d}_v,\mathbf{d}_h} \quad & \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_F^2 \\ & + \frac{1}{2}\|\mathbf{d}_v - \mathbf{D}_v\mathbf{x}\|_F^2 + \frac{1}{2}\|\mathbf{d}_h - \mathbf{D}_h\mathbf{x}\|_F^2 \\ & + \lambda\mu\left(\|(\mathbf{d}_v\ \mathbf{d}_h)\|_{iT} + \chi_{B_{l,u}}(\mathbf{x})\right).\end{aligned} \tag{53}$$

B. Isotropic TV-based Image Restoration via GAPG

We may still choose $f(\hat{\mathbf{x}})$ as the one in (37) and the rest in (53) as $g(\hat{\mathbf{x}})$, so the matrix $\mathbf{L}_f$ is still chosen as (40). Similarly, problem (53) can be decomposed into two smaller subproblems. The first one is still (46), hence $\mathbf{x}$ is still updated as (49). The second subproblem is:

$$\begin{aligned}\min_{\mathbf{d}_v,\mathbf{d}_h} \quad & \lambda\mu\|(\mathbf{d}_v\ \mathbf{d}_h)\|_{iT} \\ & + \frac{\eta}{2}\|\mathbf{d}_v - [\mathbf{y}_{\mathbf{d}_v}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_v}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k)]\|_F^2 \\ & + \frac{\eta}{2}\|\mathbf{d}_h - [\mathbf{y}_{\mathbf{d}_h}^k - \eta^{-1}(\mathbf{y}_{\mathbf{d}_h}^k - \mathbf{D}_v\mathbf{y}_{\mathbf{x}}^k)]\|_F^2,\end{aligned} \tag{54}$$

which can be further separated into mn independent subproblems:

$$\begin{aligned}\min_{(\mathbf{d}_v)_{i,j},(\mathbf{d}_h)_{i,j}} \quad & \lambda\mu\sqrt{(\mathbf{d}_v)_{i,j}^2 + (\mathbf{d}_h)_{i,j}^2} \\ & + \frac{\eta}{2}[((\mathbf{d}_v)_{i,j} - (\tilde{\mathbf{d}}_v^k)_{i,j})^2 \\ & + ((\mathbf{d}_h)_{i,j} - (\tilde{\mathbf{d}}_h^k)_{i,j})^2],\end{aligned} \tag{55}$$

Fig. 1. Isotropic TV-based image restoration from a blurred and noisy Lena image. top left: blurred and noisy Lena; top right: restored image using MTwIST; bottom left: restored image using MFISTA; bottom right: restored image using GAPG.

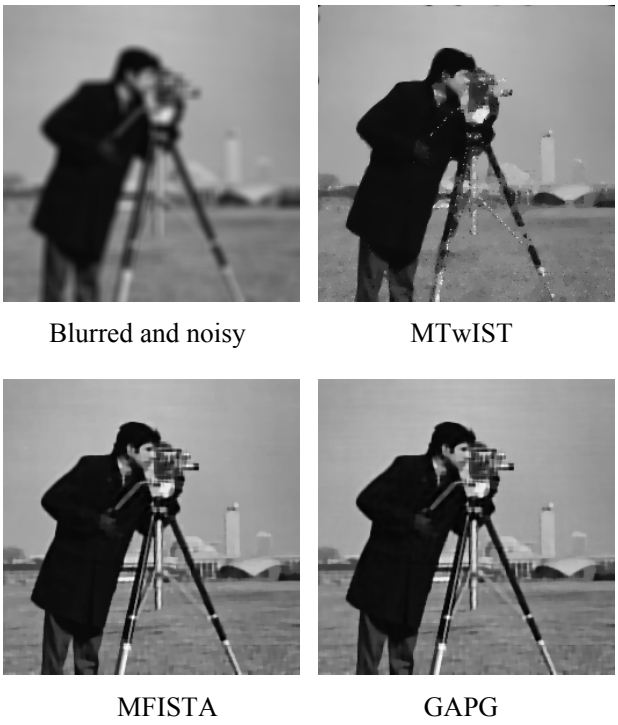

Fig. 2. Isotropic TV-based image restoration from a blurred and noisy Cameraman image. top left: blurred and noisy Cameraman; top right: restored image using MTwIST; bottom left: restored image using MFISTA; bottom right: restored image using GAPG.

where the definitions of $\tilde{\mathbf{d}}_v^k$ and $\tilde{\mathbf{d}}_h^k$ can be found in (51).

Then by Proposition 2, $\mathbf{d}_v^k$ and $\mathbf{d}_h^k$ are updated by:

$$\begin{aligned}(\mathbf{d}_v^k)_{i,j} &= \mathcal{T}_{\lambda\mu/\eta}\left(\sqrt{(\tilde{\mathbf{d}}_v^k)_{i,j}^2+(\tilde{\mathbf{d}}_h^k)_{i,j}^2}\right)\\ &\quad\times\frac{(\tilde{\mathbf{d}}_v^k)_{i,j}}{\sqrt{(\tilde{\mathbf{d}}_v^k)_{i,j}^2+(\tilde{\mathbf{d}}_h^k)_{i,j}^2}},\\ (\mathbf{d}_h^k)_{i,j} &= \mathcal{T}_{\lambda\mu/\eta}\left(\sqrt{(\tilde{\mathbf{d}}_v^k)_{i,j}^2+(\tilde{\mathbf{d}}_h^k)_{i,j}^2}\right)\\ &\quad\times\frac{(\tilde{\mathbf{d}}_h^k)_{i,j}}{\sqrt{(\tilde{\mathbf{d}}_v^k)_{i,j}^2+(\tilde{\mathbf{d}}_h^k)_{i,j}^2}},\end{aligned} \tag{56}$$

Finally, by combining the continuation technique, our GAPG-based algorithm for isotropic TV-based image restoration is almost the same as that in Algorithm 3, except that the updating rules for $\mathbf{d}_v^k$ and $\mathbf{d}_h^k$ are changed to (56). So we omit the pseudo codes.

VI. Experimental Results

In this section, we present the results of the proposed method for image restoration from blurred and noisy images and from incomplete samples (inpainting). In our experiments, we only test the algorithms for solving isotropic TV-based image restoration problems. Then we compare the computational time and the peak signal-to-noise ratio (PSNR) of the proposed method with two recent methods, MTwIST [14] and MFISTA [21]. In order to verify the efficiency of GAPG, we also compare the convergence speed of GAPG, the original APG, and MTwIST using the same image restoration problem formulation (53).

All the methods are implemented in MATLAB and are tested on a computer with a Core 2 Quad Q6600 processor running at 2.40GHz. The source codes of MTwIST (version 2) and MFISTA are downloaded from http://www.lx.it.pt/~bioucas/code/TwIST_v2.zip and http://ie.technion.ac.il/~becka/papers/tv_fista.zip, respectively[4]. When approximately solving the TV-based denoising subproblem that these two algorithms require (see Section II-D), we use the default number $k_{inner} = 10$ of the inner iterations.

A. Image Restoration

In this subsection, we present the numerical results of recovering image $\mathbf{x}$ from the blurred and noisy observed image $\mathbf{b}$ given by

$$\mathbf{b} = \mathbf{x} * \mathbf{k} + \epsilon, \tag{57}$$

where $\mathbf{k}$ is a blurring kernel, ϵ is an additive normally distributed noise, and $*$ is the convolution operator.

In our experiments, two 256×256 images, Lena and Cameraman, are used to test the effectiveness and efficiency of the GAPG method for isotropic TV-based image restoration. To obtain the blurred and noisy images, the original images are first blurred with a 9×9 Gaussian blurring kernel with mean zero and standard deviation 4 and then added with normally distributed noise with mean zero and standard deviation 10^{-3}

[4]We will also make our source code available online if the paper is accepted.

TABLE I
IMAGE RESTORATION: CPU TIMES AND PSNR VALUES OBTAINED USING MTWIST, MFISTA, AND GAPG FOR THE LENA IMAGE

	PSNR (dB)	CPU times (s)
MTwIST (k=100) [14]	29.00	20.21
MFISTA (k=100) [21]	29.13	30.79
GAPG (k=150)	29.23	4.86

TABLE II
IMAGE RESTORATION: CPU TIMES AND PSNR VALUES OBTAINED USING MTWIST, MFISTA, AND GAPG FOR THE CAMERAMAN IMAGE

	PSNR (dB)	CPU times (s)
MTwIST (k=100) [14]	27.33	20.83
MFISTA (k=100) [21]	27.56	31.38
GAPG (k=150)	27.66	4.91

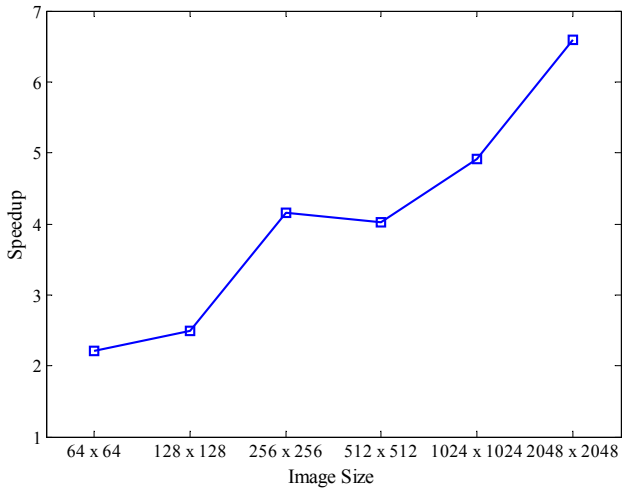

Fig. 3. The speedup of GAPG against MTwIST for image restoration with different image sizes.

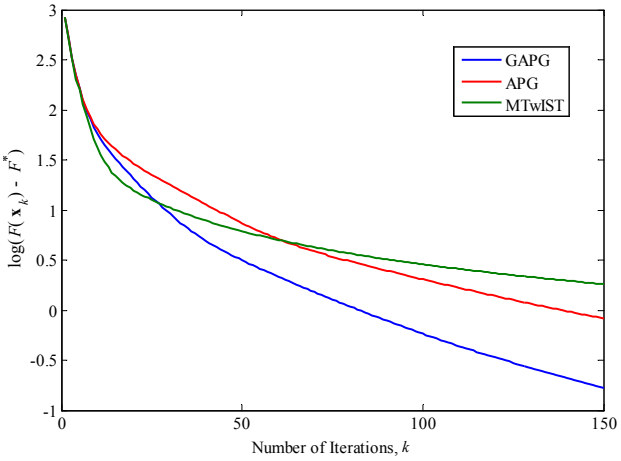

Fig. 4. Function values of GAPG, MTwIST, and the original APG in solving problem (53), where the image is the blurred and noisy Lena image and μ is fixed at $\|\mathbf{b}\|_F$.

(the graylevels are normalized to $[0, 1]$). Like [21], the regularization parameter $\lambda = 10^{-4}$ is chosen, which corresponds to the best PSNR value.

One may be interested in verifying whether the proposed GAPG method would achieve competitive restoration performance and better computational efficiency while compared with other solutions, e.g., MTwIST [14] and MFISTA [21]. Fig. 1 shows the blurred and noisy Lena image and the restoration results obtained by MTwIST (100 iterations) [14], MFISTA (100 iterations) [21], and GAPG (150 iterations)[5]. Fig. 2 shows the results on the Cameraman image. Tables I and II list the CPU times and the PSNR values obtained using MTwIST, MFISTA, and GAPG. The PSNR values of the three methods are similar, and that of GAPG is slightly higher. Although the number of iterations of GAPG ($k = 150$) is larger than those of MTwIST and MFISTA, thanks to the simpler problem formalization (53) and the GAPG method, the proposed algorithm can be about four times faster than MTwIST and MFISTA.

For the problem of recovering an image from blurred and noisy observation, we further evaluate the speedup of GAPG against MTwIST for Lena images with different sizes, as shown in Fig. 3. We adopt the simple nearest neighbor interpolation to generate the Lena images in sizes other than 256×256, and apply the same configurations of the blurring kernel and regularization parameters used in the previous experiments. From Fig. 3, GAPG can be at least two times faster than MTwIST for images of different sizes. Moreover, if the image size is not less than 256×256, GAPG can be at least four times faster than MTwIST.

It is informative to compare the convergence rates of GAPG, MTwIST, and the original APG in solving the same optimization problem. We do this by testing them on the same problem (53) with fixed $\mu = \|\mathbf{b}\|_F$. Fig. 4 shows their objective function values on the blurred and noisy Lena image. The minimum objective function value F^* is approximated by $F(\mathbf{x}_k)$ with $k = 2000$ using GAPG. At the 150th iteration, the function value of GAPG is 22.51 while those of the original APG and MTwiST are 23.17 and 24.18, respectively. Experiments on the Cameraman image also show similar results. Thus GAPG indeed converges faster than MTwIST and the original APG.

B. Image Inpainting

The target of image inpainting is to fill in missing pixels in a damaged image, which can be defined as

$$\mathcal{A}(\mathbf{x})(i,j) = \begin{cases} x(i,j), & \text{if } (i,j) \in \Omega, \\ 0, & \text{if } (i,j) \notin \Omega, \end{cases} \tag{58}$$

where Ω is the index set of known pixels. We only consider one special class of inpainting, where the missing pixels form small holes. In our experiments, we uniformly randomly sample 20% of pixels from two 256×256 images, Lena and Cameraman, to evaluate the inpainting methods using MTwIST, MFISTA, and GAPG. For this problem, it is easy to prove that $\|\mathbf{A}\|_2 = 1$. So we also have a tight estimate of $\lambda_{\max}$. In our experiments, the regularization parameter is chosen as $\lambda = 10^{-2}$, which corresponds to the best PSNR value.

[5]Here we run different methods with fixed number of iterations in order to obtain solutions with similar PSNRs, so that the convergence speed can be correctly compared.

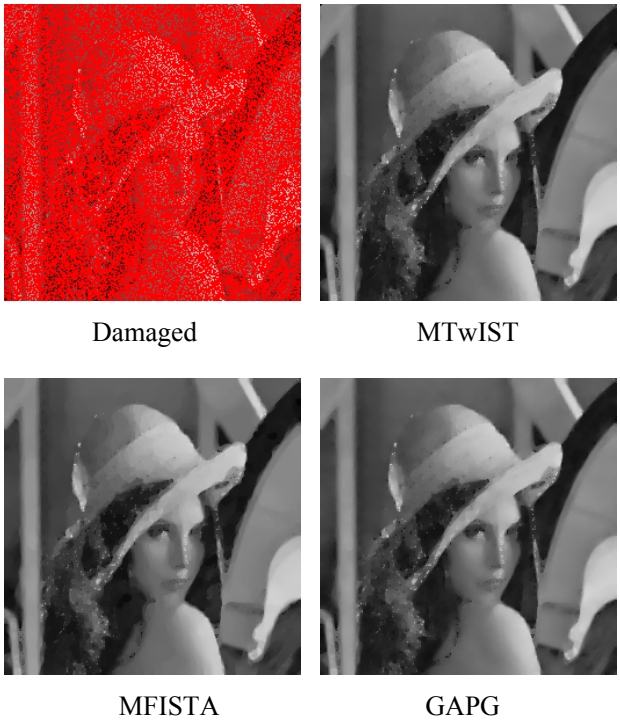

Fig. 5. The restoration results of image inpainting. top left: damaged Lena image; top right: restored image using MTwIST; bottom left: restored image using MFISTA; bottom right: restored image using GAPG.

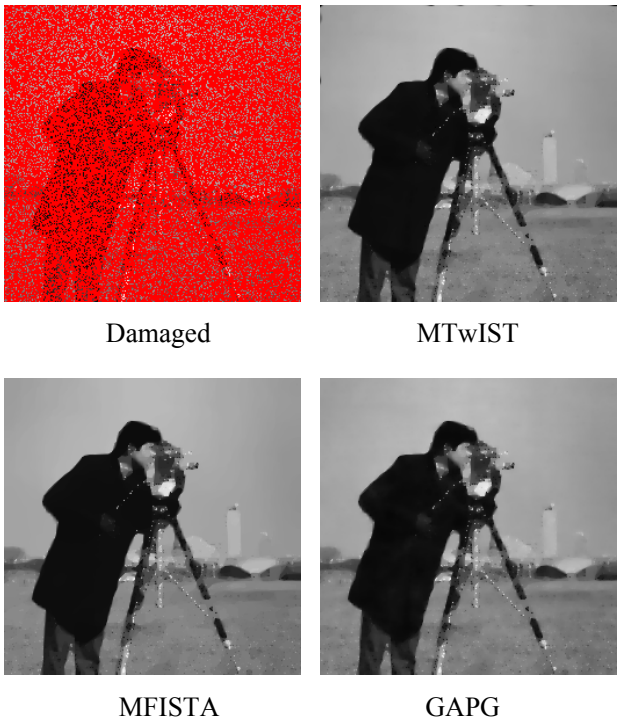

Fig. 6. The restoration results of image inpainting. top left: damaged Cameraman image; top right: restored image using MTwIST; bottom left: restored image using MFISTA; bottom right: restored image using GAPG.

First, using the Lena and Cameraman images, we compare the performance and efficiency of GAPG, MTwIST, and M-FISTA. Figures 5 and 6 show the inpainting results obtained using MTwIST ($k = 100$), MFISTA ($k = 100$), and GAPG ($k = 150$). Tables III and IV list the CPU times and the PSNR values obtained using MTwIST, MFISTA, and GAPG. As shown in Table III, for the Lena image, the PSNR of GAPG is 25.35 dB, which is slightly higher than those of MTwIST (25.17 dB) and MFISTA (24.98 dB). The CPU time of GAPG is $2.85s$, which is also much less than those of MTwIST ($12.41s$) and MFISTA ($33.51s$). As shown in Table IV, similar results are also obtained on the Cameraman image. Thus, for image inpainting, GAPG is also much more efficient than MTwIST and MFISTA.

TABLE III
IMAGE INPAINTING: CPU TIMES AND PSNR VALUES OBTAINED USING MTWIST, MFISTA, AND GAPG FOR THE LENA IMAGE

	PSNR (dB)	CPU times (s)
MTwIST (k=100) [14]	25.17	12.41
MFISTA (k=100) [21]	24.98	33.51
GAPG (k=150)	25.35	2.85

TABLE IV
IMAGE INPAINTING: CPU TIMES AND PSNR VALUES OBTAINED USING MTWIST, MFISTA, AND GAPG FOR THE CAMERAMAN IMAGE

	PSNR (dB)	CPU times (s)
MTwIST (k=100) [14]	22.94	12.44
MFISTA (k=100) [21]	22.88	35.54
GAPG (k=150)	23.38	2.86

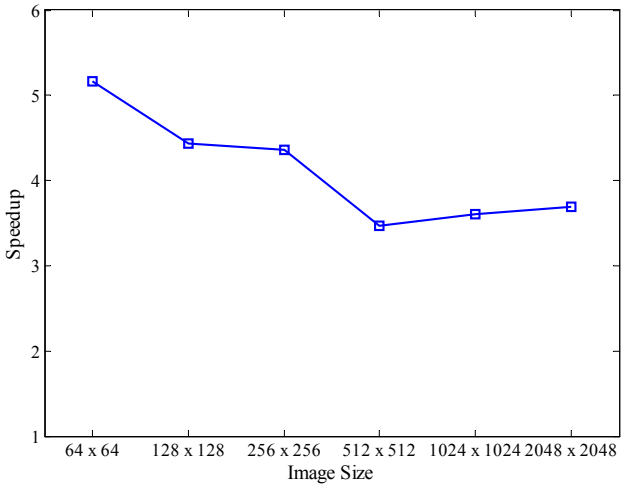

Fig. 7. The speedup of GAPG against MTwIST for inpainting with different image sizes.

Fig. 7 shows the speedup of GAPG against MTwIST in inpainting the Lena images with different sizes. One can see that GAPG is more than three times faster than MTwIST for the sizes of images tested.

Finally, we compare the objective function values of GAPG, MTwIST, and the original APG on the damaged Lena image. The experimental setting are the same as before. The results are shown in Fig. 8. At the 150th iteration, the function value of GAPG is 86.39 while those of the original APG and MTwIST are 87.93 and 88.51, respectively. Experiments on the Cameraman image also show similar results. So GAPG also converges faster than MTwIST and the original APG on

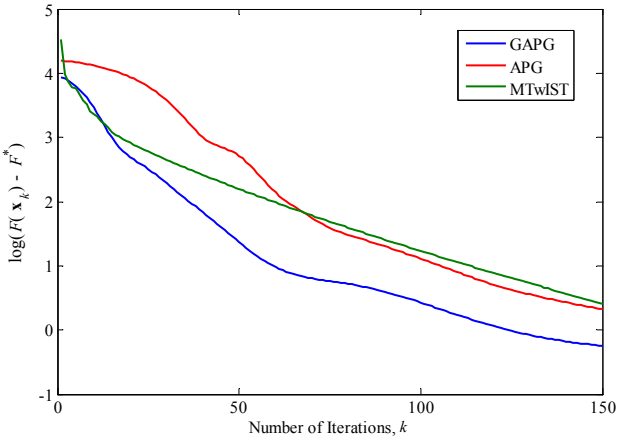

Fig. 8. Function values of GAPG, MTwIST, and the original APG in solving problem (53), where the image is the Lena image with missing pixels and μ is fixed at $\|\mathbf{b}\|_F$.

the same inpainting problem.

VII. Conclusions

In this paper, we propose a novel approach to solve anisotropic and isotropic TV-based image restoration. First, we extend the original APG method with a constant step size to propose a generalized accelerated proximal gradient method, such that different variables can have different "Lipschitz constants". The appealing convergence rate $O(k^{-2})$ is maintained by the GAPG method and our numerical results show that GAPG converges faster than the original APG. Second, by introducing two auxiliary variables, we are able to decompose the problem into much smaller problems that can be solved with high parallelism. Finally, we also adopt the common continuation technique to gradually reduce the relaxation parameter in order to provide good initial solutions. As a result, the proposed method is much faster than two recently developed APG-based methods, MTwIST [14] and MFISTA [21], with at least comparable restoration performance. Compared with the mathematical deductions of MTwIST and MFISTA, which require approximately solving image denoising subproblems, the deductions of GAPG are also much simpler.

Acknowledgment

The authors would like to thank the Associate Editor and the three reviewers for their constructive suggestions, and thank Dr. Jose M. Bioucas Dias and Dr. Amir Beck for making their source code public available.

References

[1] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, HJ: Prentice-Hall, 1977.
[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Bristol, U.K.: IOP, 1998.
[3] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.
[4] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[5] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.
[6] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *in Proc. European Conference on Computer Vision*, 2008, pp. 57–68.
[7] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Processing*, vol. PP, no. 99, p. 1, Jan. 2011.
[8] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Visualization*, vol. 20, no. 1–2, pp. 89–97, 2004.
[9] T. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 370–375, Mar 1998.
[10] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed mr imaging using total variation and wavelets," in *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
[11] W. Yin, T. Chen, X. Zhou, and A. Chakraborty, "Background correction for cDNA microarray image using the TV+l_1 model," *Bioinformatics*, vol. 21, no. 10, pp. 2410–2416, 2005.
[12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal of Imaging Science*, vol. 2, no. 1, pp. 183–202, Mar 2009.
[13] J. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Processing*, vol. 15, no. 4, pp. 937–951, Apr 2006.
[14] J. Bioucas-Das and M. Figueiredo, "A new twist: two step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2992–3004, Dec 2007.
[15] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 906–916, Aug 2003.
[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov 2001.
[17] D. Goldfarb and W. Yin, "Second-order cone programming methods for total variation-based image restoration," *SIAM Journal of Scientific Computing*, vol. 27, no. 2, pp. 622–645, 2005.
[18] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal of Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
[19] I. Daubechies, M. Defriese, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
[20] K. Bredies and D. Lorenz, "Linear convergence of iterative soft-thresholding," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5–6, pp. 813–837, Dec 2008.
[21] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2419–2434, Nov 2009.
[22] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, , and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009.
[23] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," submitted to *SIAM Journal on Optimization*, May 2008.
[24] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *preprint, National University of Singapore*, Apr. 2009.
[25] D. Goldfarb and W. Yin, "Parametric maximum flow algorithms for fast total variation minimization," *SIAM Journal on Scientific Computing*, vol. 31, no. 5, pp. 3712–3743, 2009.
[26] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2345–2356, Sept. 2010.
[27] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal of Imaging Science*, vol. 1, no. 3, pp. 248–272, July 2008.
[28] T. Goldstein and S. Osher, "The split bregman method for l_1 regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
[29] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2008.
[30] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, Apr. 2008, pp. 3373 – 3376.

[31] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.
[32] G. Golub and C. Loan, *Matrix Computation, 3rd Edition*. The John Hopkins University Press, 1996.
[33] M. Figueiredo and J. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE Trans. Image Processing*, vol. 19, no. 12, pp. 3133 – 3145, Dec. 2010.
[34] 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," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681 – 695, Mar. 2011.
[35] P. Combettes and J.-C. Pesquet, "A douglas-rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
[36] E. Esser, "Applications of lagrangian-based alternating direction methods and connections to split-bregman," *Technical Report 09-31, Computational and Applied Mathematics, University of California, Los Angeles*, Apr. 2009.
[37] S. Setzer, "Operator splittings, bregman methods and frame shrinkage in image processing," *International Journal of Computer Vision*, 2011.