East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.150413.260913a

Vol. 4, No. 1, pp. 21-34
February 2014

An Algorithm for the Proximity Operator in Hybrid TV-Wavelet Regularization, with Application to MR Image Reconstruction

Yu-Wen Fang, Xiao-Mei Huo and You-Wei Wen*

Faculty of Science, Kunming University of Science and Technology, Yunnan, China.

Received 15 April 2013; Accepted (in revised version) 26 September 2013

Available online 24 February 2014

Abstract. Total variation (TV) and wavelet L_1 norms have often been used as regularization terms in image restoration and reconstruction problems. However, TV regularization can introduce staircase effects and wavelet regularization some ringing artifacts, but hybrid TV and wavelet regularization can reduce or remove these drawbacks in the reconstructed images. We need to compute the proximal operator of hybrid regularizations, which is generally a sub-problem in the optimization procedure. Both TV and wavelet L_1 regularisers are nonlinear and non-smooth, causing numerical difficulty. We propose a dual iterative approach to solve the minimization problem for hybrid regularizations and also prove the convergence of our proposed method, which we then apply to the problem of MR image reconstruction from highly random under-sampled k-space data. Numerical results show the efficiency and effectiveness of this proposed method.

AMS subject classifications: 65K10, 68U10

Key words: Total Variation (TV), wavelet, regularization, MR image.

1. Introduction

In many image restoration or reconstruction problems, we need to solve a linear inverse problem of the form

$$\boldsymbol{g} = K\boldsymbol{f} + \boldsymbol{n}\,,$$

where $\boldsymbol{g}$ is the observed data, K is the system operator, $\boldsymbol{f}$ is the original image with size $m \times n$ and $\boldsymbol{n}$ is the random noise. It is well known that restoring an image is a very ill-conditioned process, and to alleviate this a regularization approach is generally used. The approach is to minimise the objective function, which is the weighted sum of the data-fitting term and the term containing some prior information about the original image.

In many image processing problems, an image can be modelled as a piecewise smooth function, and simultaneously sparsely represented by a wavelet basis — e.g. Lustig *et*

*Corresponding author. *Email address:* `wenyouwei@gmail.com` (Y.-W. Wen)

al [13] illustrate such sparsity in the transform domain of MR images and piecewise smoothness in the spatial domain of angiogram images. The images consequently have both small total variation (TV) norm [16] and small L_1 norm, and the reconstructed image $\boldsymbol{f}$ is a minimizer of the objective function

$$\min_{\boldsymbol{f}} D(\boldsymbol{g},\boldsymbol{f}) + \lambda_1 \mathrm{TV}(\boldsymbol{f}) + \lambda_2 \left\|\boldsymbol{W}\boldsymbol{f}\right\|_1 , \tag{1.1}$$

where $D(\boldsymbol{g},\boldsymbol{f})$ is the data-fitting term that denotes a discrepancy measure between the observed data $\boldsymbol{g}$ and the solution $\boldsymbol{f}$, and $\lambda_i (i = 1,2)$ is the regularization parameter. The term $\mathrm{TV}(\boldsymbol{f})$ denotes the TV norm of the image $\boldsymbol{f}$, which can preserve edges in the image due to the piecewise smooth regularization property of the TV norm, but it may over-smooth image details and introduce staircase effects. While wavelet L_1 regularization can keep local image features and details through sparse representation of the image, it may introduce some ringing artifacts along image contours. The main advantage in combining TV regularization with the L_1 norm of wavelet coefficients is to reduce or remove staircase effects caused by TV regularization and ring effects caused by wavelet regularization.

The chief challenge in solving the problem (1.1) is that the TV and L_1 regularisers are both nonlinear and non-smooth. The minimizer of (1.1) can be computed by the conjugate gradient method [13] or PDE approach [12], but the main drawback is that the convergence is very slow in practice. When the data fitting term $D(\boldsymbol{g},\boldsymbol{f})$ has a Lipschitz-continuous gradient, it is possible to use the forward-backward splitting proximal algorithm to solve the optimization problem [8]. The proximity operator of the function $\psi(\boldsymbol{f})$ is defined as

$$\mathrm{prox}_{\psi}(\boldsymbol{u}) = \arg\min_{\boldsymbol{f}} \frac{1}{2} \left\|\boldsymbol{f} - \boldsymbol{u}\right\|_2^2 + \psi(\boldsymbol{f}) , \tag{1.2}$$

where $\psi(\boldsymbol{f}) = \lambda_1 \mathrm{TV}(\boldsymbol{f}) + \lambda_2 \left\|\boldsymbol{W}\boldsymbol{f}\right\|_1$. Applying forward-backward splitting proximal algorithm, the solution of the problem (1.1) is given by

$$\boldsymbol{f} = \mathrm{prox}_{\alpha\psi} \left(\boldsymbol{f} - \alpha \nabla_{\boldsymbol{f}} D(\boldsymbol{g},\boldsymbol{f})\right)$$

where $\alpha > 0$, which suggests that the minimizer $\boldsymbol{f}$ might be achieved by performing an iterative scheme with an initial solution.

However, an important task in forward-backward splitting proximal algorithm is to compute the proximal operator of the regularisers. Chambolle [3] proposed a project algorithm to compute the proximal operator of a TV regulariser, and it is well known that the proximal operator of a wavelet L_1 regulariser is a shrinkage operator [9]. Combettes & Pesquet developed an iterative method to compute the proximity operator of composite regularisers, by performing the proximity operator of each regulariser independently [7]. Recently, we obtained a formulation to compute the proximal operator when the function ψ is a linear combination of a TV norm and wavelet L_1 norm [2], but the relevant convergence analysis was not given there. In this article, we reconsider how to compute the proximal operator of the linear combination of the TV and wavelet L_1 norms — i.e. we study the minimization problem

$$\min_{\boldsymbol{f}} \mathbf{Q}(\boldsymbol{f}) \equiv \frac{1}{2} \left\|\boldsymbol{f} - \boldsymbol{g}\right\|_2^2 + \lambda_1 \mathrm{TV}(\boldsymbol{f}) + \lambda_2 \left\|\boldsymbol{W}\boldsymbol{f}\right\|_1 . \tag{1.3}$$

This problem (1.3) differs from the one considered in Ref. [18], where the wavelet basis was used to represent the texture part of the image. We solve our problem (1.3) by re-writing the TV norm in the dual formulation proposed by Chambolle [3], when it becomes a min-max problem with an optimal solution that can be obtained by computing a saddle point, which can be calculated by computing the dual variable first. An iterative method is proposed to seek the dual variable, and the convergence of our proposed method is shown. We then apply this method to solve the problem of MR image reconstruction from highly under-sampled data, and present some consequent numerical results.

This article is organised as follows. In Section 2, we develop the dual iterative algorithm for hybrid TV and wavelet L_1 regularised problems. In Section 3, we introduce the problem of MR image reconstruction from high under-sampled data and apply our proposed method to solve it. Our consequent numerical results are in Section 4, and some concluding remarks are drawn in Section 5.

2. Dual Iterative Algorithm for Hybrid TV and L_1 Regularized Problems

Let us first provide notation used throughout this paper. We denote by X the Euclidean space $\mathbb{R}^{m\times n}$, and $Y{=}X\times X$. For $\boldsymbol{g}{\in}X$, $g_{i,j}{\in}\mathbb{R}$ denotes the $((i{-}1)m{+}j)$-th component of $\boldsymbol{g}$. For $\boldsymbol{p}\in Y$, $p_{i,j}=(p_{i,j,1},p_{i,j,2})\in\mathbb{R}^2$ denotes the $((i-1)m+j)$-th component of $\boldsymbol{p}$. The inner product is$\langle \boldsymbol{p},\boldsymbol{q}\rangle_Y=\sum_{i,j}\sum_{k=1}^2 p_{i,j,k}q_{i,j,k}$, and the norm $\|\boldsymbol{p}\|_\infty=\max_{i,j}\{|p_{i,j}|\}$ with $|p_{i,j}|=\sqrt{p_{i,j,1}^2+p_{i,j,2}^2}$ and $\|\boldsymbol{p}\|_2=\sqrt{\langle \boldsymbol{p},\boldsymbol{p}\rangle}$. In order to define a discrete Total Variation (TV), we introduce a discrete version of the gradient operator. For any $\boldsymbol{f}\in X$, the gradient ∇ is a linear operator from X to Y, $\nabla \boldsymbol{f}$ is a vector in Y given by $(\nabla \boldsymbol{f})_{r,s}=((\nabla_x \boldsymbol{f})_{r,s},(\nabla_y \boldsymbol{f})_{r,s})$ with $(\nabla_x \boldsymbol{f})_{r,s}=\boldsymbol{f}_{r+1,s}-\boldsymbol{f}_{r,s}$ and $(\nabla_y \boldsymbol{f})_{r,s}=\boldsymbol{f}_{r,s+1}-\boldsymbol{f}_{r,s}$. The discrete TV of the image $\boldsymbol{f}$ is then defined as

$$\mathrm{TV}(\boldsymbol{f})\equiv \big\||\nabla \boldsymbol{f}|\big\|_1=\sum_{r,s}\sqrt{(\nabla_x \boldsymbol{f})_{r,s}^2+(\nabla_y \boldsymbol{f})_{r,s}^2}\,;$$

and the discrete version of the divergence operator is defined by $\mathrm{div}=-\nabla^T$ where ∇^T is the adjoint of ∇, such that for every $\boldsymbol{p}\in Y$ and $\boldsymbol{f}\in X$ we have $\boldsymbol{f}^T\mathrm{div}\boldsymbol{p}=-\langle \boldsymbol{p},\nabla \boldsymbol{f}\rangle_Y$. We also define the set

$$A\equiv\{\boldsymbol{p}\in Y:\|\boldsymbol{p}\|_\infty\le 1\}, \tag{2.1}$$

and the characteristic function δ_A of A as $\delta_A(\boldsymbol{p})=0$ for $\boldsymbol{p}\in A$ and $\delta_A(\boldsymbol{p})=+\infty$ if $\boldsymbol{p}\notin A$. (It is obvious that A is a convex set.) The discrete TV of the image $\boldsymbol{f}$ is also the Legendre-Fenchel conjugate of δ [3,4] — i.e.

$$\mathrm{TV}(\boldsymbol{f})=\max_{\boldsymbol{p}\in Y}\Big\{\boldsymbol{f}^T\mathrm{div}\boldsymbol{p}-\delta_A(\boldsymbol{p})\Big\}=\max_{\boldsymbol{p}\in A}\boldsymbol{f}^T\mathrm{div}\boldsymbol{p}. \tag{2.2}$$

We represent the TV norm using the dual formulation and define the objective function $\mathrm{J}(\boldsymbol{f},\boldsymbol{p})$ as

$$\mathrm{J}(\boldsymbol{f},\boldsymbol{p})=\frac{1}{2}\|\boldsymbol{f}-\boldsymbol{g}\|^2+\lambda_1\boldsymbol{f}^T\mathrm{div}\boldsymbol{p}+\lambda_2\|\boldsymbol{W}\boldsymbol{f}\|_1\,, \tag{2.3}$$

hence $\mathbf{Q}(\boldsymbol{f}) = \max_{\boldsymbol{p}\in A} \mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ and the minimization problem becomes a min-max problem

$$\min_{\boldsymbol{f}} \max_{\boldsymbol{p}\in A} \mathbf{J}(\boldsymbol{f},\boldsymbol{p})\,.$$

By using arguments of duality theory for convex programming, a pair $(\boldsymbol{f}^*,\boldsymbol{p}^*)$ is a saddle point for $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ if and only if

$$\mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}) \le \mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}^*) \le \mathbf{J}(\boldsymbol{f},\boldsymbol{p}^*) \tag{2.4}$$

for any $\boldsymbol{f}$ and $\boldsymbol{p} \in A$, which means that

$$\min_{\boldsymbol{f}} \max_{\boldsymbol{p}\in A} \mathbf{J}(\boldsymbol{f},\boldsymbol{p}) = \mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}^*) = \max_{\boldsymbol{p}\in A} \min_{\boldsymbol{f}} \mathbf{J}(\boldsymbol{f},\boldsymbol{p}). \tag{2.5}$$

Thus to compute the minimizer of $\mathbf{Q}(\boldsymbol{f})$, we seek to compute the saddle point of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$. However, before describing our iterative method to calculate the saddle point, let us define the so-called soft-thresholding and projection operators.

Definition 2.1. For scalar b and λ, consider the operator

$$S_\lambda(b) = \begin{cases} b-\lambda\,, & b \ge \lambda \\ 0\,, & |b| < \lambda \\ b+\lambda\,, & b \le -\lambda \end{cases}.$$

For the vector $\boldsymbol{b}$, the soft-thresholding operator $\mathbf{S}_\lambda(\boldsymbol{b})$ is then defined by $(\mathbf{S}_\lambda(\boldsymbol{b}))_i = S_\lambda(\boldsymbol{b}_i)$, when we have

$$\mathbf{S}_\lambda(\boldsymbol{b}) = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{b}\|^2 + \lambda\|\boldsymbol{x}\|_1.$$

Definition 2.2. The projection of a vector $\boldsymbol{q}$ onto the convex set A is defined as

$$\mathbf{P}_A(\boldsymbol{q}) = \arg\min_{\boldsymbol{p}\in A} \left\|\boldsymbol{p}-\boldsymbol{q}\right\|_2^2. \tag{2.6}$$

We can apply the Lagrangian method to calculate the projection operator $\mathbf{P}_A$. The Lagrangian function associated with (2.6) is

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

where $\beta_{i,j} \ge 0$ is the Lagrangian multiplier associated with the constraint $|p_{i,j}|^2 \le 1$. Its complementarity conditions implies that for the optimal $\beta_{i,j}$ either $\beta_{i,j} = 0$ with $|p_{i,j}| \le 1$, or $\beta_{i,j} > 0$ with $|p_{i,j}| = 1$ and $|q_{i,j}| > 1$. In the first case, we have $p_{i,j} = q_{i,j}$; and in the second case, the KKT conditions [1] yield

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

Thus we have $\beta_{i,j} = |q_{i,j}| - 1$ and therefore $p_{i,j} = q_{i,j}/|q_{i,j}|$, so that

$$\left(\mathbf{P}_A(\boldsymbol{q})\right)_{i,j} = \frac{q_{i,j}}{\max(1,|q_{i,j}|)}\,. \tag{2.7}$$

We now establish the following theorem.

Theorem 2.1. *$(\boldsymbol{f}^*,\boldsymbol{p}^*)$ is a saddle point of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ if and only if $(\boldsymbol{f}^*,\boldsymbol{p}^*)$ satisfies*

$$\boldsymbol{f}^* = \boldsymbol{W}^T \mathbf{S}_{\lambda_2}(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^*)) \tag{2.8}$$

and

$$\boldsymbol{p}^* = \mathbf{P}_A(\boldsymbol{p}^* - \tau\lambda_1 \nabla \boldsymbol{f}^*). \tag{2.9}$$

where τ is a parameter. Moreover, Eqs. (2.8) *and* (2.9) *can be reformulated in a more compact form:*

$$\boldsymbol{p}^* = \mathbf{P}_A(\boldsymbol{p}^* - \tau\lambda_1 \nabla \boldsymbol{W}^T \mathbf{S}_{\lambda_2}\left(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^*))\right) . \tag{2.10}$$

Proof. According the constrained optimality condition [1], $(\boldsymbol{f}^*,\boldsymbol{p}^*)$ is the saddle point of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ if and only if there exists the pair $(0,\boldsymbol{v}_p)$ with $0 \in \partial_f \mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}^*)$ and $\boldsymbol{v}_p \in -\partial_p \mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}^*)$ such that for any $\boldsymbol{p} \in A$ there is $\boldsymbol{v}_p^T(\boldsymbol{p} - \boldsymbol{p}^*) \geq 0$.

We first show that Eq. (2.8) holds. With $\phi(\boldsymbol{p}) = \boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}$, we rewrite the formula in Eq. (2.3) as

$$\mathbf{J}(\boldsymbol{f},\boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{f} - \phi(\boldsymbol{p})\|^2 + \lambda_2\|\boldsymbol{W}\boldsymbol{f}\|_1 + \frac{1}{2}\|\boldsymbol{g}\|^2 - \frac{1}{2}\|\phi(\boldsymbol{p})\|^2 ,$$

and for simplicity denote $\boldsymbol{f}_w = \boldsymbol{W}\boldsymbol{f}$. Once $\boldsymbol{f}_w$ is available, $\boldsymbol{f}$ can be reconstructed by $\boldsymbol{f} = \boldsymbol{W}^T\boldsymbol{f}_w$. Exploiting the unitary invariance property of the L_2 norm we have

$$\|\boldsymbol{f} - \phi(\boldsymbol{p})\|^2 = \|\boldsymbol{f}_w - \boldsymbol{W}\phi(\boldsymbol{p})\|^2 ,$$

whence $\boldsymbol{f}^* = \mathrm{argmin}_f \mathbf{J}(\boldsymbol{f},\boldsymbol{p}^*)$ or $0 \in \partial_f \mathbf{J}(\boldsymbol{f}^*,\boldsymbol{p}^*)$ if and only if

$$\boldsymbol{f}_w^* = \mathbf{S}_{\lambda_2}(\boldsymbol{W}\cdot\phi(\boldsymbol{p}^*))$$

such that Eq. (2.8) holds.

On the other hand, $(\boldsymbol{f}^*,\boldsymbol{p}^*)$ is a saddle point of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ if and only if the equalities Eq. (2.4) hold for any $\boldsymbol{f}$ and $\boldsymbol{p} \in A$. Substituting the expression of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ into Eq. (2.4), we have the following inequality

$$(\boldsymbol{f}^*)^T(\mathrm{div}\boldsymbol{p} - \mathrm{div}\boldsymbol{p}^*) \leq 0 ,$$

and noting that $\mathrm{div} = -\nabla^T$ and $(\boldsymbol{f}^*)^T\mathrm{div}\boldsymbol{p} = -\langle \nabla\boldsymbol{f},\boldsymbol{p}\rangle_Y$ we obtain

$$\langle \boldsymbol{p} - \boldsymbol{p}^*, \boldsymbol{p}^* - (\boldsymbol{p}^* - \tau\lambda_1\nabla\boldsymbol{f}^*)\rangle_Y \geq 0 .$$

Then on recalling that $\boldsymbol{q}^* \in A$ is the projection of a point $\boldsymbol{q}$ onto a convex set A if and only if

$$\langle \boldsymbol{p} - \boldsymbol{q}^*, \boldsymbol{q}^* - \boldsymbol{q}\rangle \geq 0$$

for any $\boldsymbol{p} \in A$, it follows that Eq. (2.9) holds. $\square$

2.1. Dual Iterative Method

Theorem 2.1 states that the saddle point of $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ can be calculated by seeking the dual variable $\boldsymbol{p}$ that satisfies Eq. (2.10), and then replacing the dual variable in Eq. (2.8) to compute the primal variable. However, Eq. (2.10) involving the expression for the soft thresholding operator and the projection operation is non-differentiable, which poses serious difficulty for its numerical solution. To address this difficulty, Chen *et.al.* [6] introduced the concept of slant differentiability in view of its Lipschitz continuity, and proposed semi-smooth methods to solve this type of equation.

We describe a simple iterative method in this subsection, as Theorem 2.1 suggests that the dual variable $\boldsymbol{p}$ in (2.10) might be computed via an iterative scheme. Thus starting from some initial pair $\boldsymbol{p}^{(0)}$, we consider the iteration scheme of form

$$\boldsymbol{p}^{(k+1)} = \mathbf{P}_A\left(\boldsymbol{p}^{(k)} - \tau\lambda_1\nabla \boldsymbol{W}^T\mathbf{S}_{\lambda_2}(\boldsymbol{W}(\boldsymbol{g} - \lambda_1\mathrm{div}\boldsymbol{p}^{(k)}))\right), \tag{2.11}$$

and show that the sequence of $\boldsymbol{p}^{(k)}$ converges to some point $\boldsymbol{p}^*$ satisfying Eq. (2.9) under the assumption on the stepsize τ. Once $\boldsymbol{p}^*$ is obtained, the primal variable $\boldsymbol{f}^*$ can be computed from Eq. (2.8). The iterative scheme in Eq. (2.11) can also be interpreted as an alternating direction method to minimise and maximise with respect to $\boldsymbol{f}$ and $\boldsymbol{p}$, respectively — i.e.

$$\begin{aligned}\boldsymbol{f}^{(k+1)} =&\mathrm{argmin}_{\boldsymbol{f}}\mathbf{J}\left(\boldsymbol{f},\boldsymbol{p}^{(k)}\right),\\ \boldsymbol{p}^{(k+1)} =&\mathrm{argmin}_{\boldsymbol{p}\in A}\mathbf{J}\left(\boldsymbol{f}^{(k+1)},\boldsymbol{p}\right) - \frac{1}{2\tau}\left\|\boldsymbol{p}-\boldsymbol{p}^{(k)}\right\|_2^2.\end{aligned}$$

It is easy to check that

$$\boldsymbol{f}^{(k+1)} =\boldsymbol{W}^T\mathbf{S}_{\lambda_2}\left(\boldsymbol{W}(\boldsymbol{g} - \lambda_1\mathrm{div}\boldsymbol{p}^{(k)})\right), \tag{2.12}$$

$$\boldsymbol{p}^{(k+1)} =\mathbf{P}_A\left(\boldsymbol{p}^{(k)} - \tau\lambda_1\nabla\boldsymbol{f}^{(k+1)}\right). \tag{2.13}$$

The first step is to minimise the primal variable by fixing the dual variable. The second step is to solve the constrained maximization problem, which can be via the proximal method [14,15] that is widely applied to solve convex minimization problems [5,7,10,11,17].

The following Theorem states that the sequence $\boldsymbol{p}^{(k)}$ generated from Eq. (2.13) is bounded when $0<\tau<1/(4\lambda_1^2)$.

Theorem 2.2. *If* $(\boldsymbol{f}^*,\boldsymbol{p}^*)$ *is a saddle point of* $\mathbf{J}(\boldsymbol{f},\boldsymbol{p})$ *and the sequence* $(\boldsymbol{f}^{(k)},\boldsymbol{p}^{(k)})$ *is generated from Eqs.* (2.12) *and* (2.13)*, then*

$$\left\|\boldsymbol{p}^{(k+1)}-\boldsymbol{p}^*\right\|_2^2 \le \left\|\boldsymbol{p}^{(k)}-\boldsymbol{p}^*\right\|_2^2 - 2\tau\left(1-4\tau\lambda_1^2\right)\left\|\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right\|_2^2. \tag{2.14}$$

In particular, when $0<\tau<1/(4\lambda_1^2)$ *the sequence* $(\boldsymbol{f}^{(k)},\boldsymbol{p}^{(k)})$ *converges to some limit point* $(\boldsymbol{f}^\dagger,\boldsymbol{p}^\dagger)$.

Proof. It is well known that the projection operator is non-expansive — i.e. for all $\boldsymbol{p}$ and $\boldsymbol{q}$, we have

$$\left\|\mathbf{P}_A(\boldsymbol{q})-\mathbf{P}_A(\boldsymbol{p})\right\|_2^2 \le \left\|\boldsymbol{q}-\boldsymbol{p}\right\|_2^2 .$$

Eqs. (2.9) and (2.13) yield

$$\begin{aligned}
&\left\|\boldsymbol{p}^{(k+1)}-\boldsymbol{p}^*\right\|_2^2\\
=&\left\|\mathbf{P}_A(\boldsymbol{p}^{(k)}-\tau\lambda_1\nabla\boldsymbol{f}^{(k+1)})-\mathbf{P}_A(\boldsymbol{p}^*-\tau\lambda_1\nabla\boldsymbol{f}^*)\right\|_2^2\\
\le&\left\|\left(\boldsymbol{p}^{(k)}-\boldsymbol{p}^*\right)-\tau\lambda_1\nabla\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right)\right\|_2^2\\
=&\left\|\boldsymbol{p}^{(k)}-\boldsymbol{p}^*\right\|_2^2+\tau^2\lambda_1^2\left\|\nabla\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right)\right\|_2^2-2\tau\lambda_1\left\langle\boldsymbol{p}^{(k)}-\boldsymbol{p}^*,\nabla\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right)\right\rangle_Y .
\end{aligned}$$

By definition,

$$\left\|\nabla\boldsymbol{f}\right\|_2^2=\sum_{r,s}\left((f_{r+1,s}-f_{r,s})^2+(f_{r,s+1}-f_{r,s})^2\right)=2\sum_{r,s}\left(f_{r+1,s}^2+2f_{r,s}^2+f_{r,s+1}^2\right)\le 8\|\boldsymbol{f}\|^2 .$$

From $\langle\boldsymbol{p},\nabla\boldsymbol{f}\rangle_Y=-\boldsymbol{f}^T\mathrm{div}\boldsymbol{p}$, we then obtain

$$\begin{aligned}
\left\|\boldsymbol{p}^{(k+1)}-\boldsymbol{p}^*\right\|_2^2\le\left\|\boldsymbol{p}^{(k)}-\boldsymbol{p}^*\right\|_2^2&+8\tau^2\lambda_1^2\left\|\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right\|_2^2\\
&+2\tau\lambda_1\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right)^T\mathrm{div}\left(\boldsymbol{p}^{(k)}-\boldsymbol{p}^*\right).
\end{aligned}\tag{2.15}$$

Now consider the bound of the last term in the above inequality. For simplicity, let us denote $\phi(\boldsymbol{p})=\boldsymbol{g}-\lambda_1\mathrm{div}\boldsymbol{p}$. According to the optimality condition, $\boldsymbol{f}^{(k+1)}$ minimises $\mathbf{J}(\boldsymbol{f},\boldsymbol{p}^{(k)})$ if and only if there exists $\boldsymbol{v}_f\in\partial_f\mathbf{J}(\boldsymbol{f}^{(k+1)},\boldsymbol{p}^{(k)})$ such that

$$\boldsymbol{v}_f^T\left(\boldsymbol{f}-\boldsymbol{f}^{(k+1)}\right)\ge 0$$

for any $\boldsymbol{f}$. Noting that $\boldsymbol{f}^{(k+1)}-\phi(\boldsymbol{p}^{(k)})+\lambda_2\boldsymbol{W}^T\mathrm{sign}(\boldsymbol{f}_w^{(k+1)})\in\partial_f\mathbf{J}(\boldsymbol{f}^{(k+1)},\boldsymbol{p}^{(k)})$, we have

$$\left(\boldsymbol{f}^{(k+1)}-\phi(\boldsymbol{p}^{(k)})+\lambda_2\boldsymbol{W}^T\mathrm{sign}(\boldsymbol{f}_w^{(k+1)})\right)^T\left(\boldsymbol{f}-\boldsymbol{f}^{(k+1)}\right)\ge 0 .\tag{2.16}$$

Similarly, on using $\boldsymbol{f}^*=\mathrm{argmin}_f\mathbf{J}(\boldsymbol{f},\boldsymbol{p}^*)$ we obtain

$$\left(\boldsymbol{f}^*-\phi(\boldsymbol{p}^*)+\lambda_2\boldsymbol{W}^T\mathrm{sign}(\boldsymbol{f}_w^*)\right)^T\left(\boldsymbol{f}-\boldsymbol{f}^*\right)\ge 0 .\tag{2.17}$$

Setting $\boldsymbol{f}=\boldsymbol{f}^*$ in (2.16) and $\boldsymbol{f}=\boldsymbol{f}^{(k+1)}$ in (2.17), and then summing the resulting inequalities gives

$$\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*\right)^T\left(\boldsymbol{f}^{(k+1)}-\boldsymbol{f}^*+\lambda_1\mathrm{div}(\boldsymbol{p}^{(k)}-\boldsymbol{p}^*)+\lambda_2\boldsymbol{b}\right)\le 0\tag{2.18}$$

where $\boldsymbol{b} = \boldsymbol{W}^T \left(\mathrm{sign}(\boldsymbol{f}_w^{(k+1)}) - \mathrm{sign}(\boldsymbol{f}_w^*)\right)$. It is easy to check that

$$\left(\boldsymbol{f}^{(k+1)} - \boldsymbol{f}^*\right)^T \boldsymbol{b} = \left(\boldsymbol{f}_w^{(k+1)} - \boldsymbol{f}_w^*\right)^T \left(\mathrm{sign}(\boldsymbol{f}_w^{(k+1)}) - \mathrm{sign}(\boldsymbol{f}_w^*)\right) \geq 0\,. \tag{2.19}$$

From (2.18) and (2.19), we obtain

$$\lambda_1 \left(\boldsymbol{f}^{(k+1)} - \boldsymbol{f}^*\right)^T \mathrm{div}\left(\boldsymbol{p}^{(k)} - \boldsymbol{p}^*\right) \leq -\left\|\boldsymbol{f}^{(k+1)} - \boldsymbol{f}^*\right\|_2^2,$$

and summing this inequality and (2.15) then obtain the desired inequality (2.14).

When $0 < \tau < 1/(4\lambda_1^2)$, the sequence $\boldsymbol{p}^{(k)}$ is bounded and therefore contains a subsequence $\boldsymbol{p}^{(k_i))}$ converging to some limit point $\boldsymbol{p}^\dagger$. Since the subsequence $\boldsymbol{p}^{(k_i))}$ is convergent, for any given $\epsilon > 0$ there exists a constant i_0 such that

$$\left\|\boldsymbol{p}^{(k_i)} - \boldsymbol{p}^\dagger\right\|_2^2 < \epsilon, \quad \forall i > i_0\,.$$

Corresponding to (2.14), when $0 < \tau < 1/(4\lambda_1^2)$ there exists a constant j_0 (for example $j_0 = k_{i_1}$) for any given $\epsilon > 0$ such that

$$\left\|\boldsymbol{p}^{(j)} - \boldsymbol{p}^\dagger\right\|_2^2 \leq \left\|\boldsymbol{p}^{(j_0)} - \boldsymbol{p}^\dagger\right\|_2^2 < \epsilon\,, \quad \forall j > j_0\,,$$

so the sequence $\boldsymbol{p}^{(k)}$ converges to $\boldsymbol{p}^\dagger$. Finally, on writing $\boldsymbol{f}^\dagger = \boldsymbol{W}^T \mathbf{S}_{\lambda_2}(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^\dagger))$ we have

$$\begin{aligned}\left\|\boldsymbol{f}^{(j+1)} - \boldsymbol{f}^\dagger\right\|_2^2 &= \left\|\mathbf{S}_{\lambda_2}(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^{(j)})) - \mathbf{S}_{\lambda_2}(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^\dagger))\right\|_2^2 \\ &\leq \|\mathrm{div}\|_2^2 \left\|\boldsymbol{p}^{(j)} - \boldsymbol{p}^\dagger\right\|_2^2\end{aligned}$$

such that the sequence $\boldsymbol{f}^{(j)}$ is also convergent, and hence the sequence $(\boldsymbol{f}^{(k)}, \boldsymbol{p}^{(k)})$ converges to $(\boldsymbol{f}^\dagger, \boldsymbol{p}^\dagger)$. □

The next Theorem states that the sequence $\boldsymbol{f}^{(k)}$ generated by Eq. (2.12) converges to a minimizer of (3.1) below.

Theorem 2.3. *When $0 < \tau < 1/(4\lambda_1^2)$, the sequence $(\boldsymbol{f}^{(k)}, \boldsymbol{p}^{(k)})$ generated by Eqs.* (2.12) *and* (2.13) *converges to a saddle point of* $\mathbf{J}(\boldsymbol{f}, \boldsymbol{p})$*. Moreover, $\boldsymbol{f}^{(k)}$ converges to a minimizer of* (3.1).

Proof. According to Theorem 2.2, the sequence $(\boldsymbol{f}^{(k)}, \boldsymbol{p}^{(k)})$ converges to some limit point $(\boldsymbol{f}^\dagger, \boldsymbol{p}^\dagger)$, so it remains to show that $(\boldsymbol{f}^\dagger, \boldsymbol{p}^\dagger)$ is a saddle point of $\mathbf{J}(\boldsymbol{f}, \boldsymbol{p})$. The limit point $(\boldsymbol{f}^\dagger, \boldsymbol{p}^\dagger)$ satisfies

$$\begin{aligned}\boldsymbol{f}^\dagger &= \boldsymbol{W}^T \mathbf{S}_{\lambda_2}\left(\boldsymbol{W}(\boldsymbol{g} - \lambda_1 \mathrm{div}\boldsymbol{p}^\dagger)\right), \\ \boldsymbol{p}^\dagger &= \mathbf{P}_A\left(\boldsymbol{p}^\dagger - \tau\lambda_1 \nabla \boldsymbol{f}^\dagger\right)\,,\end{aligned}$$

so from Theorem 2.1 it is a saddle point of $\mathbf{J}(\boldsymbol{f}, \boldsymbol{p})$. □

3. MR Image Reconstruction

Due to its non-invasive manner and excellent depiction of soft tissue changes, Magnetic Resonance (MR) imaging is widely used in radiology to visualise internal structure and functions of the body. Sparsity in the transform domain and piecewise smoothness in the spatial domain make it possible to reconstruct MR images from under-sampled k-space data. Lustig *et al* [13] proposed to reconstruct the MR image by performing hybrid regularised optimization via

$$\boldsymbol{f}_0 = \underset{\boldsymbol{u}}{\mathrm{argmin}} \mathbf{Q}_0(\boldsymbol{f}) , \tag{3.1}$$

where

$$\mathbf{Q}_0(\boldsymbol{f}) = \frac{1}{2} \left\| \boldsymbol{F}_p \boldsymbol{f} - \boldsymbol{b} \right\|_2^2 + \lambda_1 \mathrm{TV}(\boldsymbol{f}) + \lambda_2 \left\| \boldsymbol{W} \boldsymbol{f} \right\|_1 , \tag{3.2}$$

with $\boldsymbol{b}$ the observed data in k-space and $\mathbf{F}_p$ a Fourier transform evaluated only at a subset of frequency domain samples (corresponding to one of the k-space under-sampling schemes). It is notable that the under-sampled Fourier transform matrix $\boldsymbol{F}_p$ can be rewritten as $\boldsymbol{F}_p = \boldsymbol{P}\boldsymbol{F}$, where $\boldsymbol{P}$ is a sampling matrix and $\boldsymbol{F}$ the full Fourier transform, so the MR image reconstruction problem can be regarded as a Fourier domain in-painting problem (filling in the missing sampled points in the k-space).

Since the rows of the down-sampling matrix are orthogonal, we have $\boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}$ while $\boldsymbol{F}$ is a Fourier transform matrix, implying that $\|\boldsymbol{u} - \boldsymbol{F}\boldsymbol{f}\|_2^2 = \|\boldsymbol{F}^T\boldsymbol{u} - \boldsymbol{f}\|_2^2$. Using a certain matrix operation and optimization transform, we derive a quadratic majorizing function for the data fitting term in the objective function (3.2):

$$\left\| \boldsymbol{F}_p \boldsymbol{f} - \boldsymbol{b} \right\|_2^2 = \frac{1+\alpha}{\alpha} \min_{\boldsymbol{u}} \|\boldsymbol{P}\boldsymbol{u} - \boldsymbol{b}\|_2^2 + \alpha \left\| \boldsymbol{F}^T \boldsymbol{u} - \boldsymbol{f} \right\|_2^2 .$$

On defining the bivariate function $\mathbf{Q}_1(\boldsymbol{x}, \boldsymbol{u})$ as

$$\mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f}) = \frac{1+\alpha}{2\alpha} \|\boldsymbol{P}\boldsymbol{u} - \boldsymbol{b}\|_2^2 + \frac{1+\alpha}{2} \left\| \boldsymbol{F}^T \boldsymbol{u} - \boldsymbol{f} \right\|_2^2 + \lambda_1 \mathrm{TV}(\boldsymbol{f}) + \lambda_2 \|\boldsymbol{W}\boldsymbol{f}\|_1 ,$$

from the convexity of $\mathbf{Q}_0(\boldsymbol{f})$ and $\mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f})$ the minimization problem for $\mathbf{Q}_0(\boldsymbol{f})$ is equivalent to $\mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f})$ — i.e. $\min_{\boldsymbol{f}} \mathbf{J}(\boldsymbol{f}) = \min_{\boldsymbol{f},\boldsymbol{u}} \mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f})$. Consequently, instead of computing the minimizer of the objective function $\mathbf{Q}(\boldsymbol{f})$, we calculate the minimizer of the objective function $\mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f})$ via an alternating minimization algorithm. Thus starting from an initial guess $\boldsymbol{f}^{(0)}$, we use the alternating minimization algorithm to generate the sequence

$$\begin{cases} \boldsymbol{u}^{(k)} = \mathrm{argmin}_{\boldsymbol{u}}\, \mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f}^{(k-1)}) , \\ \boldsymbol{f}^{(k)} = \mathrm{argmin}_{\boldsymbol{u}}\, \mathbf{Q}_1(\boldsymbol{u}^{(k)}, \boldsymbol{f}) . \end{cases}$$

We observe that the objective function $\mathbf{Q}_1(\boldsymbol{u}, \boldsymbol{f})$ is quadratic with respect to $\boldsymbol{u}$, so $\boldsymbol{u}^{(k)}$ can easily be computed from the formula

$$\boldsymbol{u}^{(k)} = \left(\boldsymbol{P}^T\boldsymbol{P} + \alpha \boldsymbol{I}\right)^{-1} \left(\boldsymbol{P}^T\boldsymbol{b} + \alpha \boldsymbol{F}\boldsymbol{f}^{(k-1)}\right) . \tag{3.3}$$

Since $\boldsymbol{P}$ is a random under-sampled matrix, $\boldsymbol{P}^T\boldsymbol{P}$ is a diagonal matrix with diagonal entries are 0 and 1, so we can compute $\boldsymbol{u}^{(k)}$ easily. We remark that $\boldsymbol{u}^{(k)}$ can be regarded as an average of $\boldsymbol{P}^T\boldsymbol{b}$ and $\boldsymbol{F}\boldsymbol{f}^{(k-1)}$. If the point in k-space is unsampled, its value is filled by the Fourier coefficient of the restored image. For the variable $\boldsymbol{f}$, we have

$$\boldsymbol{f}^{(k)} = \underset{\boldsymbol{f}}{\operatorname{argmin}}\, \mathbf{Q}_1(\boldsymbol{u}^{(k)}, \boldsymbol{f}) = \underset{\boldsymbol{f}}{\operatorname{argmin}}\, \frac{1}{2}\left\|\boldsymbol{F}^T\boldsymbol{u} - \boldsymbol{f}\right\|_2^2 + \beta_1 \mathrm{TV}(\boldsymbol{f}) + \beta_2\|\boldsymbol{W}\boldsymbol{f}\|_1\,,$$

where $\beta_i = \lambda_i/(1+\alpha)$ for $i = 1, 2$. We can apply the dual proximal based iterative method (cf. Ref. (2.11)) to calculate the dual variable $\boldsymbol{p}$, and then compute $\boldsymbol{f}$ from Eq. (2.12).

Let us now summarise the resulting algorithm.

Algorithm 3.1 Iterative Method for MR Image Reconstruction.

Ensure: $\boldsymbol{u} = \mathrm{IterMethod}(\boldsymbol{P}, \boldsymbol{b}, \lambda_1, \lambda_2)$.
Require: $\boldsymbol{P}, \boldsymbol{b}, \beta_1, \beta_2$.

```
Initialize u^(0). Set the parameter α, τ.
β_i = λ_i/(1+α) for i = 1,2.
while stopping criterion is not satisfied do
    u^(k) = (P^T P + αI)^(-1) (P^T b + α f^(k-1));
    Initialize p^(k,0).
    while stopping criterion is not satisfied do
        p^(k,j) = P_A(p^(k,j-1) − τ β_1 ∇ W^T S_{β_2}(W(F^T u^(k) − β_1 div p^(k,j-1))));
    end while
    f^(k) = W^T S_{β_2}(W(F^T u^(k) − β_1 div p^(k,j)));
end while
return f = f^(k).
```

4. Numerical Results

We now illustrate the performance of our proposed algorithm for sparse MRI reconstruction problems, and compare it with both the gradient descent method (GD) and the nonlinear conjugate gradient descent method with backtracking linear search (Nonlinear CG) — cf. Ref. [13]. Our codes are written in Matlab R2009a.

The experiments were performed under Mac OS X10.7.3 and MATLAB R2011a on a MacBook Air Laptop with a 1.7GHz Intel Core i5 processor and 4GB of RAM. The Signal-to-Noise Ratio (SNR) used to measure the quality of the restoration results is defined as follows: $\mathrm{SNR} = 10\log_{10}(\|\boldsymbol{u}\|_2^2/\|\boldsymbol{u} - \widehat{\boldsymbol{u}}\|_2^2)$, where $\boldsymbol{u}$ and $\widehat{\boldsymbol{u}}$ are the original image and the restored image, respectively. The observed image was chosen as the initial image. The sample rate in the tests was 20%, selected with polynomial variable density sampling, and Gaussian white noise with standard deviation $\sigma = 0.01$ was added in the sampling data. We chose the parameters $\lambda_1 = \lambda_2 = 0.01$ and $\tau = 0.248$.

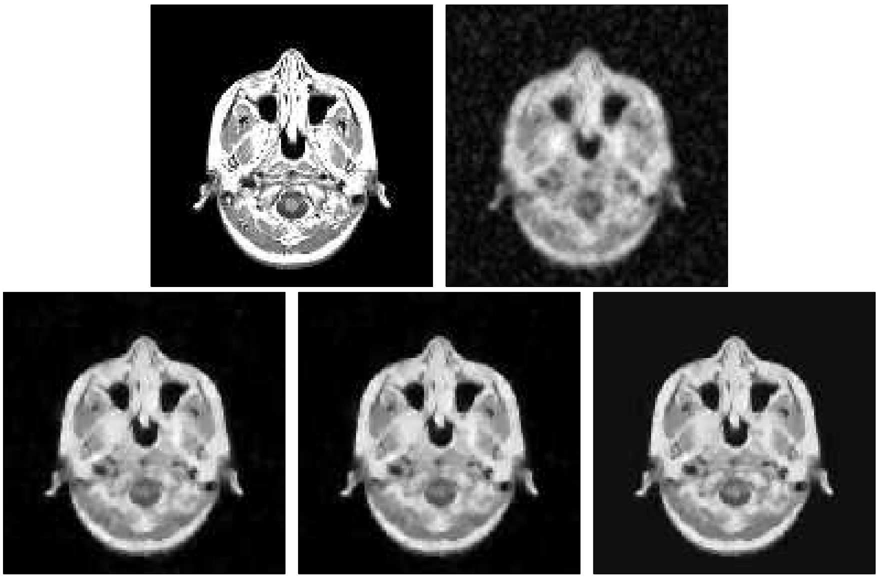

Figure 1: The original 128×128 image (Image1), the observed image with 20% k-space data, and the restored image obtained from the conjugate gradient method, the nonlinear conjugate gradient method and our proposed method, respectively.

Two MR images of brain with size 128×128 were used in the test. Fig. 2 and Fig. 4 show the original images, the observed image with 20% k-space data, and the restored image obtained by the conjugate gradient method, the nonlinear conjugate gradient method and our proposed method, respectively. The SNRs of the restored images versus CPU times are shown in Fig. 3, and we observe that our proposed method produced the best SNRs.

Next, we investigated our proposed method scaling with the image size. We used the "Shepp-Logan" phantom image generated by the MATLAB command `phantom(n)` with $n = 64, 128, 256, 512$. Fig. 4 shows the original "Shepp-Logan" phantom image with $n = 64$, its observed image with 20% k-space data, and the restored image obtained from conjugate gradient method, nonlinear conjugate gradient method and our proposed method, respectively. The SNRs of the restored images with size n versus CPU times are shown in Fig. 5, and we again observe that our proposed method produces the best SNRs.

5. Conclusion

We have considered the proximal operator of both hybrid TV and wavelet L_1 regularization, which can reduce or remove staircase effects caused by TV regularization and ring effects caused by wavelet regularization. In order to overcome the numerical difficulty caused by the nonlinearity and non-smoothness of the TV and wavelet regularizations, we represented the TV-norm in a dual formulation. We proposed an iterative method to compute the dual variable and analyzed the convergence of this method. We then applied our method to the problem of reconstructing MR images from highly random under-sampled

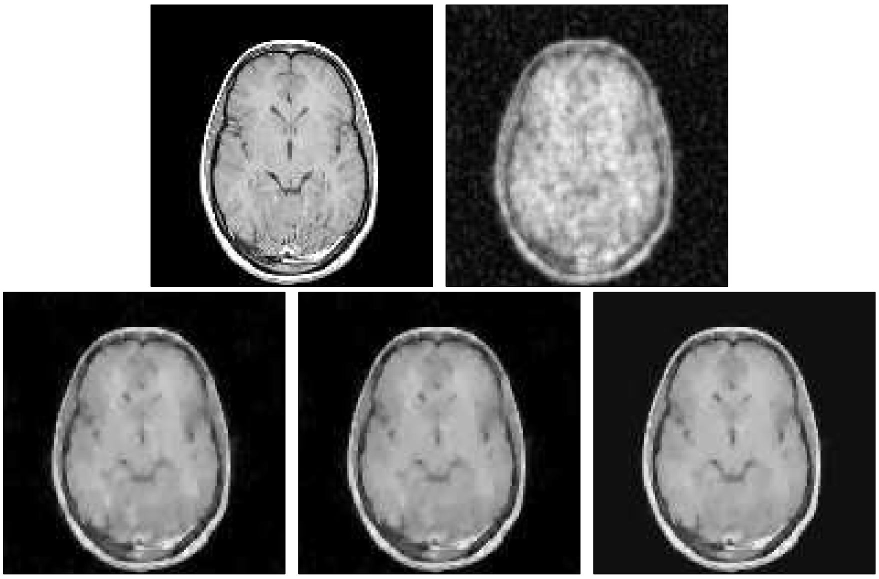

Figure 2: The original 128×128 image (Image2), the observed image with 20% k-space data, and the restored image obtained from the conjugate gradient method, the nonlinear conjugate gradient method and our proposed method, respectively.

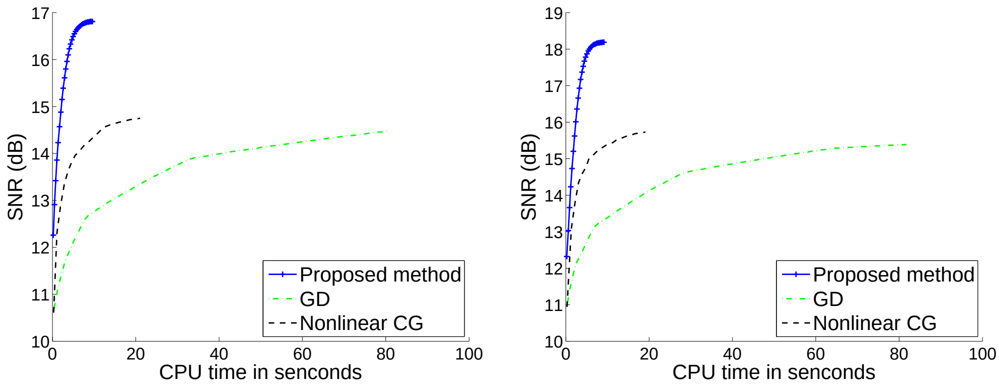

Figure 3: SNR versus CPU time for Image1 (left) and Image2 (right).

k-space data, using an optimization transfer technique that involves replacing the original univariate functional in the MR image reconstruction by a bivariate functional on adding an auxiliary variable. Our bivariate functional can be minimised by alternating minimization, where the minimum of the auxiliary variable has a closed form solution, and the minimization problem for the original variable is equivalent to solving the proximal operator of the hybrid regularization. Our experimental results indicate that our proposed algorithm is very efficient relative to gradient descent methods.

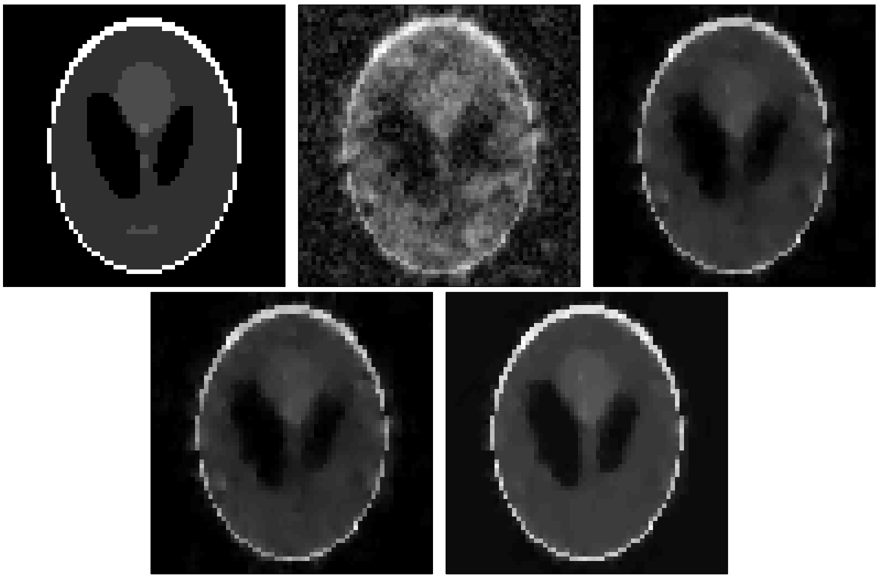

Figure 4: The original 64×64 phantom, the observed image with 20% k-space data, and the restored image obtained from the conjugate gradient method, the nonlinear conjugate gradient method, and our proposed method respectively.

Acknowledgments

Our research is supported in part by NSFC Grant 11361030 and a SRF for ROCS of SEM.

References

[1] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Dover Publications, 2003.

[2] Z. Cao and Y. Wen, *A Splitting Algorithm for MR Image Reconstruction from Sparse Sampling*, pp. 329–334, CRC PressI Llc, 2012.

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

[4] T. Chan, G. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput. **20**(6), 1964–1977 (1999).

[5] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Math. Program. Ser. A **64**(1), 81–101 (1994).

[6] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal. **38**(4), 1200–1216 (2000).

[7] P. Combettes and J. Pesquet, *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems **24**(6), 065014 (2008).

[8] P.L. Combettes and V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul. **4**(4), 1168–1200 (2005).

[9] D. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions*, Appl. Comput. Harmon. Anal. **1**, 100–115 (1995).

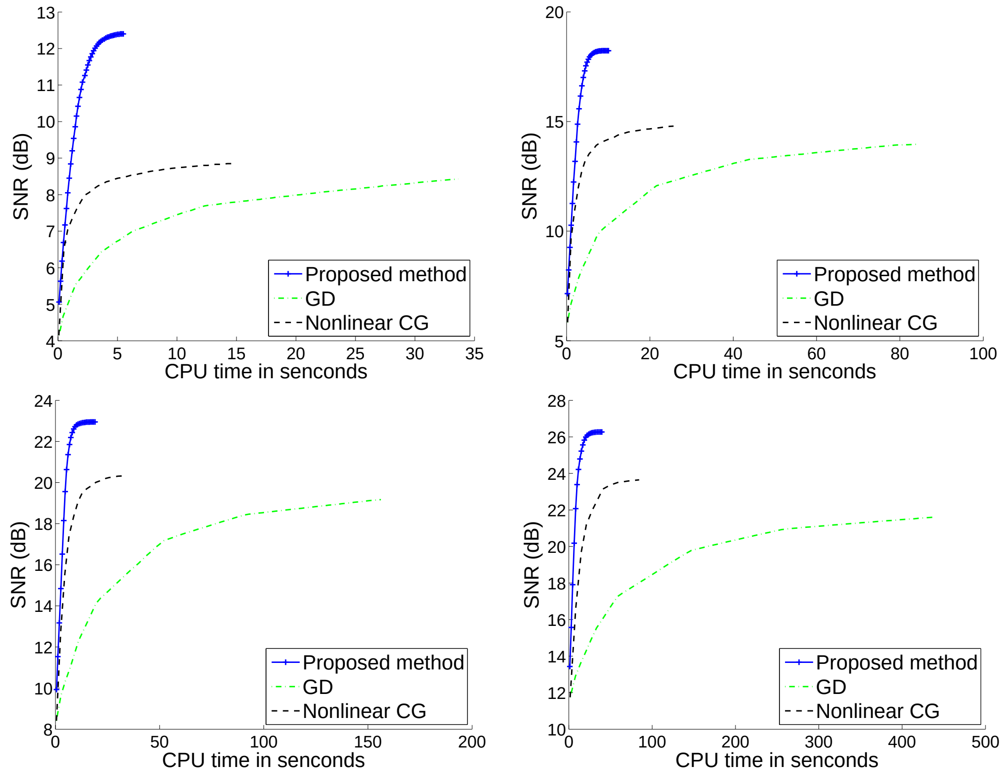

Figure 5: SNR versus CPU time for phantom images with size $n = 64, 128, 256, 256$(from left to right).

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

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

[12] L. He, T.C. Chang, S. Osher, T. Fang, and P. Speier, *MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods*, UCLA CAM Report **6**(35), 2006.

[13] M. Lustig, D. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007).

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

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

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

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

[18] T. Zeng, X. Li, and M. Ng, *Alternating minimization method for total variation based wavelet shrinkage model* Commun. Comput. Phys. **8**(5), 976–994 (2010).