Computational Acceleration for MR Image Reconstruction in Partially Parallel Imaging

3 authors:

Xiaojing Ye
Georgia State University

29 PUBLICATIONS   **391** CITATIONS

SEE PROFILE

Yunmei Chen
University of Florida

90 PUBLICATIONS   **2,510** CITATIONS

SEE PROFILE

Feng Huang

64 PUBLICATIONS   **1,441** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   fluid mechanics View project

Computational Acceleration for MR Image Reconstruction in Partially Parallel Imaging

Xiaojing Ye*, Yunmei Chen, and Feng Huang

Abstract—In this paper, we present a fast numerical algorithm for solving total variation and ℓ_1 (TVL1) based image reconstruction with application in partially parallel MR imaging. Our algorithm uses variable splitting method to reduce computational cost. Moreover, the Barzilai-Borwein step size selection method is adopted in our algorithm for much faster convergence. Experimental results on clinical partially parallel imaging data demonstrate that the proposed algorithm requires much fewer iterations and/or less computational cost than recently developed operator splitting and Bregman operator splitting methods, which can deal with a general sensing matrix in reconstruction framework, to get similar or even better quality of reconstructed images.

Index Terms—L1 minimization, image reconstruction, convex optimization, partially parallel imaging.

I. INTRODUCTION

IN this paper we develop a novel algorithm to accelerate the computation of total variation (TV) and/or ℓ_1 based image reconstruction. The general form of such problems is

$$\min_u \left\{ \alpha \|u\|_{TV} + \beta \|\Psi^\top u\|_1 + \frac{1}{2}\|Au - f\|^2 \right\}, \quad (1)$$

where $\|\cdot\|_{TV}$ is the total variation, $\|\cdot\|_1$ and $\|\cdot\| \equiv \|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms, respectively. For notation simplicity we only consider two dimensional (2D) images in this paper, whereas the method can be easily extended to higher dimensional cases. Following the standard treatment we will vectorize an (2D) image u into one-dimensional column vector, i.e. $u \in \mathbb{C}^N$ where N is the total number of pixels in u. Then, the (isotropic) TV norm is defined by

$$\|u\|_{TV} = \int_\Omega |Du| = \sum_{i=1}^{N} \|D_i u\| \quad (2)$$

where for each $i = 1, \cdots, N$, $D_i \in \mathbb{R}^{2 \times N}$ has two nonzero entries in each row corresponding to finite difference approximations to partial derivatives of u at the i-th pixel along the coordinate axes. In (1), $\alpha, \beta \geq 0$ $(\alpha + \beta > 0)$ are parameters corresponding to the relative weights of the data fidelity term $\|Au - f\|^2$ and the terms $\|u\|_{TV}$ and $\|\Psi^\top u\|_1$. Model (1) has been widely applied to image reconstruction problems. Solving

(1) yields a restored clean image u from an observed noisy or blurred image f when $A = I$ or a blurring matrix, respectively. In compressive sensing (CS) applications, A is usually a large and ill-conditioned matrix depending on imaging devices or data acquisition patterns, and f represents the under-sampled data. In CS $\Psi = [\psi_1, \cdots, \psi_N] \in \mathbb{C}^{N \times N}$ is usually a proper orthogonal matrix (e.g. wavelet) that sparsifies the underlying image u.

A. Partially Parallel MR Imaging

The CS reconstruction via TVL1 minimization (1) has been successfully applied to an emerging MR imaging application known as partially parallel imaging (PPI). PPI uses multiple RF coil arrays with separate receiver channel for each RF coil. A set of multi-channel k-space data from each radiofrequency (RF) coil array is acquired simultaneously. The imaging is accelerated by acquiring a reduced number of k-space samples. Partial data acquisition increases the spacing between regular subsequent read-out lines, thereby reducing scan time. However, this reduction in the number of recorded Fourier components leads to aliasing artifacts in images. There are two general approaches for removing the aliasing artifacts and reconstructing high quality images: image domain-based methods and k-space based methods. Various models in the framework of (1) have been employed as image domain-based reconstruction methods in PPI [1], [2], [3], [4], [5], [6], [7], [8], [9]. Sensitivity encoding (SENSE) [4], [3] is one of the most commonly used methods of such kind. SENSE utilizes knowledge of the coil sensitivities to separate aliased pixels resulted from undersampled k-space.

The fundamental equation for SENSE is as follows: In a PPI system consisting of J coil arrays, the under-sampled k-space data f_j from the j-th channel relates to the underlying image u by $P\mathcal{F}(S_j \odot u) = f_j$, $j = 1, \cdots, J$, where $\mathcal{F}$ is the Fourier transform, P is a binary matrix representing the under-sampling pattern (mask), and $S_j \in \mathbb{C}^N$ is the sensitivity map of the j-th channel in the vector form as u. The symbol $\odot$ is the Hadamard (or componentwise) product between two vectors. In early works on SENSE, the reconstruction was obtained by solving a least squares problem

$$\min_{u \in \mathbb{C}^N} \sum_{j=1}^{J} \|\mathcal{F}_p(S_j \odot u) - f_j\|^2, \quad (3)$$

where $\mathcal{F}_p$ is the undersampled Fourier transform defined by $\mathcal{F}_p \triangleq P\mathcal{F}$. Denote

$$A = (\mathcal{F}_p \mathbf{S}_1; \mathcal{F}_p \mathbf{S}_2; \cdots; \mathcal{F}_p \mathbf{S}_J) \text{ and } f = (f_1; \cdots; f_J), \quad (4)$$

where $\mathbf{S}_j \triangleq \operatorname{diag}(S_j) \in \mathbb{C}^{N \times N}$ is the diagonal matrix with $S_j \in \mathbb{C}^N$ on the diagonal, $j = 1, \cdots, J$. Here $(\cdot; \cdot)$ stacks the arguments vertically to form a matrix. Then problem (3) can be expressed as

$$\min_{u \in \mathbb{C}^N} \|Au - f\|^2, \tag{5}$$

and then solved by conjugate gradient (CG) algorithm. However, due to the ill-conditioning of the encoding matrix A, it has been shown in [6] that the CG iteration sequence often exhibits a "semi-convergence" behavior, which can be characterized as initially converging toward the exact solution and later diverging. Moreover, the convergence speed is low, when the acceleration factor is high.

Recently, total variation (TV) based regularization has been incorporated into SENSE to improve reconstructed image quality and convergence speed over the un-regularized CG method ([1], [9]). TV based regularization can be also considered as forcing the reconstructed image to be sparse with respect to spatial finite differences. This sparsity along with the sparsity of MR signals under wavelet transforms have been exploited in [10], where the framework (1) has been employed to reconstruct MR images from under-sampled k-space data.

There have been several fast numerical algorithms for solving (1) that will be briefly reviewed in the next section. However, computational acceleration is still an important issue for certain medical applications, such as breath-holding cardiac imaging. For the application in PPI the computational challenge is not only from the lack of differentiability of the TV and ℓ_1 terms, but also the large size and severe ill-conditioning of the inversion matrix A in (4).

The main contribution of this paper is to develop a fast numerical algorithm for solving (1) with general A. The proposed algorithm incorporates the Barzilai-Borwein (BB) method into a variable splitting framework for optimal step size selection. The numerical results on partially parallel imaging (PPI) problems demonstrate much improved performance on reconstruction speed for similar image quality.

B. Previous Work

In reviewing the prior work on TVL1-based image reconstruction, we simplify (1) by taking $\beta = 0$. It is worth pointing out here that TV has much stronger practical performance than ℓ_1 in image reconstructions, yet harder to solve because the gradient operators involved are not invertible as Ψ in the ℓ_1 term. In [11], [12], a method is developed based on the following reformulation of (1) with $\beta = 0$:

$$\min_{u,w} \left\{ \alpha \sum_{i=1}^N \|w_i\| + \frac{1}{2} \|Au - f\|^2 : w_i = D_i u, \forall i \right\} \tag{6}$$

Then the linear constraint was treated with a quadratic penalty

$$\min_{u,w} \left\{ \alpha \sum_{i=1}^N \|w_i\| + \frac{\rho}{2} \|Du - w\|^2 + \frac{1}{2} \|Au - f\|^2 \right\}, \tag{7}$$

where $w \in \mathbb{C}^{2N}$ is formed by stacking the two columns of $(w_1, \cdots, w_N)^\top$, and $D = (D^x; D^y) \in \mathbb{C}^{2N \times N}$. D^x and D^y are the horizontal and vertical global finite difference matrices

(N-by-N), i.e. they consist of the first and second rows of all D_i's, respectively. For any fixed ρ, (7) can be solved by alternating minimizations. If both $D^\top D$ and $A^\top A$ can be diagonalized by the Fourier matrix, as they would if A is either the identity matrix or a blurring matrix with periodic boundary conditions, then each minimization involves shrinkage and two fast Fourier transforms (FFTs). A continuation method is used to deal with the slow convergence rate associated with a large value for ρ. The method, however, is not applicable to more general A.

In [13] Goldstein and Osher developed a split Bregman method for (6). The resulting algorithm has similar computational complexity to the algorithm in [11]; the convergence is fast and the constraints are exactly satisfied. Later the split Bregman method was shown to be equivalent to the alternating direction method of multipliers (ADMM) [14], [15], [16], [17] applied to the augmented Lagrangian $L(w, u; p)$ defined by

$$\alpha \sum_{i=1}^N \|w_i\| + \frac{1}{2} \|Au - f\|^2 + \langle p, Du - w \rangle + \frac{\rho}{2} \|Du - w\|^2, \tag{8}$$

where $p \in \mathbb{C}^{2N}$ is the Lagrangian multiplier. Nonetheless, the algorithms in [13], [11], [12] benefit from the special structure of A, and they lose efficiency if $A^\top A$ cannot be diagonalized by fast transforms. To treat a more general A, the Bregman operator splitting (BOS) method [18] replaces $\|Au - f\|^2$ by a proximal-like term $\delta \|u - (u^k - \delta^{-1} A^\top (Au^k - f))\|^2$ for some $\delta > 0$. BOS is an inexact Uzawa method that depends on the choice of δ. The advantage of BOS is that it can deal with general A and does not require the inversion of $A^\top A$ during the computation. However, BOS is relatively less efficient than the method presented in this paper, even if δ is chosen optimally. The comparison of our method with the BOS algorithm will be presented in Section IV.

There are also several methods developed to solve the associated dual or primal-dual problems of (1) based on the dual formulation of the TV norm:

$$\|u\|_{TV} = \max_{p \in X} \langle p, Du \rangle, \tag{9}$$

where $X = \{p \in \mathbb{C}^{2N} : p_i \in \mathbb{C}^2, \|p_i\| \leq 1, \forall i\}$ and p_i extracts the i-th and $(i+N)$-th entries of p. Consequently, (1) can be written as a minimax problem

$$\min_{u \in \mathbb{C}^N} \max_{p \in X} \left\{ \alpha \langle p, Du \rangle + \frac{1}{2} \|Au - f\|^2 \right\}. \tag{10}$$

In [19], Chan *et al.* proposed to solve the primal-dual Euler-Lagrange equations using Newton's method. This leads to a quadratic convergence rate and highly accurate solutions; however, the cost per iteration is much higher since the method explicitly uses second-order information and the inversion of a Hessian matrix is required. In [20], Chambolle used the dual formulation of the TV denoising problem (1) with $A = I$, and provided an efficient semi-implicit gradient descent algorithm for the dual. However, the method does not naturally extend to the case with more general A. Recently, Zhu and Chan [21] proposed a primal-dual hybrid gradient (PDHG) method. PDHG alternately updates the primal and dual variables u and p. Numerical results show that PDHG outperforms methods in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

3

[20], [13] for denoising and deblurring problems, but its efficiency again relies on the fact that $A^\top A$ can be diagonalized by fast transforms. Later, several variations of PDHG, referred to as projected gradient descent algorithms, were applied to the dual formulation of image denoising problem in [22] to make the method more efficient. Further enhancements involve different step-length rules and line-search strategies, including techniques based on the Barzilai-Borwein method [23].

Another approach that can be applied to (1) in the imaging context (1) with a general A is the operator splitting (OS) method. In [24] the OS idea of [25] is applied to image reconstruction in compressed magnetic resonance imaging. The OS scheme rewrites (1) as

$$\min_u \alpha \sum_i h(D_i u) + \frac{1}{2}\|Au - f\|^2 \qquad (11)$$

where $h(\cdot) \triangleq \|\cdot\|$. Then the optimal conditions for (11) are

$$w_i^* \in \partial h(D_i u^*), \quad \delta_1 \alpha D_i^\top w_i^* + \delta_1 A^\top(Au^* - f) = 0, \quad (12)$$

where $\partial h(z)$ is the subdifferential of h at point z defined by

$$\partial h(z) \triangleq \{d \in \mathbb{C}^N : h(y) - h(z) \geq \langle d, y - z\rangle, \forall y\}.$$

The theory of conjugate duality gives the equivalency $y \in \partial h(z) \Leftrightarrow z \in \partial h^*(y)$, $\forall y, z$, where $h^*(y) \triangleq \sup_v \{\langle y, v\rangle - h(v)\}$. Hence the first condition in (12) can be written as

$$0 \in \delta_2 h^*(w_i^*) + (w_i^* - t_i^*), \quad t_i^* = \delta_2 D_i u^* + w_i^* \quad (13)$$

and then the first one leads to

$$w_i^* \in \partial h\left((t_i^* - w_i^*)/\delta_2\right) = \partial h(t_i^* - w_i^*), \qquad (14)$$

where the equality is due to $h(\cdot) = \|\cdot\|$. (14) is equivalent to

$$w_i^* = \arg\min_{w_i}\left\{h(t_i^* - w_i) + \frac{1}{2}\|w_i\|^2\right\} \qquad (15)$$

that projects t_i^* onto the unit ball in $\mathbb{R}^2$. Then, combining (15) and the last equalities in (12) and (13), the OS scheme iterates the following for a fixed point (which is also a solution to (1)):

$$\begin{cases} t_i^{k+1} = w_i^k + \delta_2 D_i u^k, & \forall i \\ w_i^{k+1} = \arg\min_{w_i}\left\{\|t_i^{k+1} - w_i\| + \frac{1}{2}\|w_i\|^2\right\}, & \forall i \\ u^{k+1} = \delta_1 \alpha \sum_i D_i^\top w_i^{k+1} + \delta_1 A^\top(Au^k - f) + u^k \end{cases}$$

OS is efficient for solving (1) with general A when all the parameters are carefully chosen. However it is still not as efficient as our method even under its optimal settings. The comparison of our method with the OS algorithm [24] will be given in Section IV.

II. PROPOSED ALGORITHM

In this paper, we develop a fast algorithm to numerically solve problem (1). Note that the computational challenge of (1) comes from the combination of two issues: one is possibly huge size and of the inversion matrix A, and the other one is the non-differentiability of the TV and ℓ_1 terms.

As discussed earlier, despite that there were some fast algorithms proposed recently to solve image restoration problems similar to (1), their efficiency relies on a very special structure of A such that $A^\top A$ can be diagonalized by fast transforms, which is not the case in most medical imaging problems, such as that in (4) in PPI application.

To tackle the computational problem of (1), we first introduce auxiliary variables w_i and z_i to transform $D_i u$ and $\psi_i^\top u$ out of the non-differentiable norms:

$$\min_{w,z,u}\left\{\alpha \sum_i \|w_i\| + \beta \sum_i |z_i| + \frac{1}{2}\|Au - f\|^2\right\}, \qquad (16)$$
$$w_i = D_i u, z_i = \psi_i^\top u, \forall\, i = 1, \cdots, N,$$

which is clearly equivalent to the original problem (1) as they share the same solutions u. To deal with the constraints in (16) brought by variable splitting, we form the augmented Lagrangian defined by

$$L(w, z, u; b, c)$$
$$= \alpha \sum_i \left(\|w_i\| - \rho\langle b_i, w_i - D_i u\rangle + \frac{\rho}{2}\|w_i - D_i u\|^2\right)$$
$$+ \beta \sum_i \left(|z_i| - \rho c_i(z_i - \psi_i^\top u) + \frac{\rho}{2}|z_i - \psi_i^\top u|^2\right) \qquad (17)$$
$$+ \frac{1}{2}\|Au - f\|^2,$$

where $b \in \mathbb{C}^{2N}$ and $c = (c_1, \cdots, c_N)^\top \in \mathbb{C}^N$ are Lagrangian multipliers. Here $b_i \in \mathbb{C}^2$ extracts the i-th and $(i + N)$-th entries of b. For notation simplicity we used the same parameter $\rho > 0$ for all constraints in (17). The method of multipliers iterates the minimizations of Lagrangian L in (17) with respect to (w, z, u) and the updates of the multipliers b and c:

$$\begin{cases} (w^{k+1}, z^{k+1}, u^{k+1}) = \arg\min_{w,z,u} L(w, z, u; b^k, c^k) \\ b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), & \forall i \\ c_i^{k+1} = c_i^k - (z_i^{k+1} - \psi_i^\top u^{k+1}), & \forall i \end{cases} \qquad (18)$$

It is proved that the sequence $\{(w^k, z^k, u^k)\}_k$ generated by (18) converges to the solution of (16) with any $\rho > 0$.

Since the updates of b^k and c^k are merely simple calculations, we now focus on the minimization of $L(w, z, u; b^k, c^k)$ in (18). First we introduce functions

$$\phi_1(s, t) = |s| + (\rho/2)\cdot|s - t|^2, \quad s, t \in \mathbb{C},$$
$$\phi_2(\mathbf{s}, \mathbf{t}) = \|\mathbf{s}\| + (\rho/2)\cdot\|\mathbf{s} - \mathbf{t}\|^2, \quad \mathbf{s}, \mathbf{t} \in \mathbb{C}^2.$$

By completing the squares in (17), we find the equivalency

$$\arg\min_{w,z,u} L(w, z, u; b^k, c^k) \equiv$$

$$\arg\min_{w,z,u}\left\{\alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k)\right.$$
$$\left. + \frac{1}{2}\|Au - f\|^2\right\} \qquad (19)$$

because the objective functions in these two minimizations are equal up to a constant independent of (w, z, u).

To solve (19) we first rewrite the objective function in a simpler way. Let $x = (w; z; u)$ and $B = (0, 0, A)$, and define functions $J_k(x) \equiv J_k(w, z, u)$ by

$$J_k(x) \triangleq \alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k),$$

and data fidelity $H(x)$ by

$$H(x) = (1/2) \cdot \|Bx - f\|^2. \qquad (20)$$

Then problem (19) (or equivalently, the minimization subproblem in (18)) can be expressed as

$$x^{k+1} = \arg \min_x \left\{ J_k(x) + H(x) \right\}, \qquad (21)$$

We further introduce $Q_\delta(x, y)$ defined by

$$Q_\delta(x, y) \triangleq H(y) + \langle \nabla H(y), x - y \rangle + \frac{\delta}{2} \|x - y\|^2, \qquad (22)$$

which is a linearization of $H(x)$ at point y plus a proximity term $\|x - y\|^2/2$ penalized by parameter $\delta > 0$. It has been shown in [26] that the sequence $\{x^{k+1,l}\}_l$ generated by

$$x^{k+1,l+1} = \arg \min_x \left\{ J_k(x) + Q_{\delta_{k+1,l}}(x, x^{k+1,l}) \right\} \qquad (23)$$

converges to the solution x^{k+1} of (21) with any initial $x^{k+1,0}$ and proper choice of $\delta_{k+1,l}$ for $l = 0, 1, \cdots$ [1]. Interestingly, we found that in practice the optimal performance can be consistently achieved by only iterating (23) once to approximate the solution x^{k+1} in (21).

Therefore, we substitute the first subproblem in (18) by

$$x^{k+1} = \arg \min_x \left\{ J_k(x) + Q_{\delta_k}(x, x^k) \right\}, \qquad (24)$$

where δ_k is chosen based on the Barzilai-Borwein (BB) method as suggested in [26]. BB method handles ill-conditioning much better than gradient methods with a Cauchy step [27]. In the BB implementation, the Hessian of the objective function is approximated by a multiple of the identity matrix. We employ the approximation

$$\delta_k = \arg \min_\delta \left\| \left(\nabla H(x^k) - \nabla H(x^{k-1}) \right) - \delta(x^k - x^{k-1}) \right\|^2, \qquad (25)$$

and get

$$\delta_k = \langle \nabla H(x^k) - \nabla H(x^{k-1}), x^k - x^{k-1} \rangle / \|x^k - x^{k-1}\|^2. \qquad (26)$$

This makes the iteration (24) exhibit a certain level of quasi-Newton convergence behavior.

From the definition of J_k and Q_{δ_k}, (24) is equivalent to

$$(w^{k+1}, z^{k+1}, u^{k+1}) = \arg \min_{w,z,u} \Phi_k(w, z, u) \qquad (27)$$

where the objective $\Phi_k(w, z, u)$ is defined by

$$\begin{aligned} &\Phi_k(w, z, u) \\ &\triangleq \alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k) \\ &+ \frac{\delta_k}{2} \left(\|w - w^k\|^2 + \|z - z^k\|^2 + \left\| u - u^k + \delta_k^{-1} A^\top (Au^k - f) \right\|^2 \right) \end{aligned} \qquad (28)$$

Theoretically, an iterative scheme can be applied to obtain the solution $(w^{k+1}, z^{k+1}, u^{k+1})$ of (27). However, here we propose only to do *one* iteration followed by the updates of b^k, c^k and δ_k in (18). This is an analogue to the split Bregman method and ADMM applied to the augmented Lagrangians, and leads to the optimal performance of (18). In summary, we propose a scheme as in (29) for solving the minimization problem (16). The updates of b^{k+1} and c^{k+1} in (29) are merely simple calculations. In (29), δ_{k+1} is derived from (26) with H defined in (20), and also has an explicit form that can be quickly computed [2]. Next, we show that w_i^{k+1} and z_i^{k+1} can be obtained by soft shrinkages by the following theorem.

Theorem II.1. *For given d-vectors $\mathbf{t}_1, \mathbf{t}_2 \in \mathbb{R}^d$ and positive numbers $a_1, a_2 > 0$, the solution to minimization problem*

$$\min_{\mathbf{s} \in \mathbb{R}^d} \left\{ \|\mathbf{s}\| + \frac{a_1}{2} \|\mathbf{s} - \mathbf{t}_1\|^2 + \frac{a_2}{2} \|\mathbf{s} - \mathbf{t}_2\|^2 \right\} \qquad (30)$$

is given by the shrinkage of a weighted sum of $\mathbf{t}_1$ and $\mathbf{t}_2$:

$$\mathcal{S}_d(\mathbf{t}_1, \mathbf{t}_2; a_1, a_2) \triangleq \mathtt{shrink}_d \left(\frac{a_1 \mathbf{t}_1 + a_2 \mathbf{t}_2}{a_1 + a_2}, \frac{1}{a_1 + a_2} \right) \qquad (31)$$

where $\mathtt{shrink}_d$ is the d-dimensional soft shrinkage operator defined by

$$\mathtt{shrink}_d(\mathbf{t}, \mu) \triangleq \max\{\|\mathbf{t}\| - \mu, 0\} \cdot \frac{\mathbf{t}}{\|\mathbf{t}\|}. \qquad (32)$$

with convention $0 \cdot (\mathbf{0}/\|\mathbf{0}\|) = \mathbf{0}$.

Proof: By completing the squares, the minimization problem (30) is equivalent to

$$\min_{\mathbf{s} \in \mathbb{R}^d} \left\{ \|\mathbf{s}\| + \left(\frac{a_1 + a_2}{2} \right) \cdot \left\| \mathbf{s} - \frac{a_1 \mathbf{t}_1 + a_2 \mathbf{t}_2}{a_1 + a_2} \right\|^2 \right\}, \qquad (33)$$

because the objective functions are the same up to a constant independent of $\mathbf{s}$. Minimizations of form (33) have a well known explicit solver $\mathtt{shrink}_d$ and hence the conclusion follows. ∎

According to Theorem II.1, w_i^{k+1} and z_i^{k+1} in (29) can be obtained by

$$w_i^{k+1} = \mathcal{S}_2 \left(D_i u^k + b_i^k, w_i^k; \rho, \alpha_k \right) \qquad (34)$$

and

$$z_i^{k+1} = \mathcal{S}_1 \left(\psi_i^\top u^k + c_i^k, z_i^k; \rho, \beta_k \right) \qquad (35)$$

where $\alpha_k = \delta_k/\alpha$ and $\beta_k = \delta_k/\beta$. Therefore the computational costs for (34) and (35) are linear in terms of N.

The u-subproblem in (29) is a least squares problem. The optimal condition of this problem reads

$$L_k u = r_k \qquad (36)$$

where $L_k = \alpha \rho D^\top D + \beta \rho I + \delta_k I$ and

$$r_k = \alpha \rho D^\top w^{k+1} + \beta \rho \Psi^\top z^{k+1} + \delta_k u^k - A^\top (Au^k - f).$$

Under periodic boundary condition, the matrix $D^\top D$ is block circulant and hence can be diagonalized by Fourier matrix $\mathcal{F}$.

---

[1] e.g. for fixed k, any limit point of $\{x^{k+1,l}\}_l$ is a solution of (21) when $\delta_{k+1,l}$ was chosen such that the objective function $J_k(x^{k+1,l}) + H(x^{k+1,l})$ monotonically decreases as $l \to \infty$ [26].

[2] The main computations for updating δ_k are norm evaluations (no A operation needed since Au^k has been computed in the u-step and can be saved for use in δ-step in (29)).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

5

$$
\begin{cases}
w_i^{k+1} = \arg\min_{w_i} \left\{ \|w_i\| + \frac{\rho}{2}\|w_i - D_i u^k - b_i^k\|^2 + \frac{\delta_k}{2\alpha}\|w - w^k\|^2 \right\}, \quad \forall i; \\
z_i^{k+1} = \arg\min_{z_i} \left\{ |z_i| + \frac{\rho}{2}|z_i - \Psi_i^\top u^k - c_i^k|^2 + \frac{\delta_k}{2\beta}|z_i - z_i^k|^2 \right\}, \quad \forall i; \\
u^{k+1} = \arg\min_u \left\{ \alpha\rho\|Du - w^{k+1}\|^2 + \beta\rho\|\Psi^\top u - z^{k+1}\|^2 + \delta_k \left\| u - \left(u^k - \delta_k^{-1} A^\top (Au^k - f) \right) \right\|^2 \right\}; \\
b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), \quad \forall i; \\
c_i^{k+1} = c_i^k - (z_i^{k+1} - \psi_i u^{k+1}), \quad \forall i; \\
\delta_{k+1} = \|A(u^{k+1} - u^k)\|^2 / \left(\|w^{k+1} - w^k\|^2 + \|z^{k+1} - z^k\|^2 + \|u^{k+1} - u^k\|^2 \right).
\end{cases}
\tag{29}
$$

Let $\Lambda = \mathcal{F}^\top D^\top D \mathcal{F}$ which is a diagonal matrix, then apply $\mathcal{F}$ on both sides of (36) to obtain

$$\hat{L}_k \mathcal{F} u = \hat{r}_k \tag{37}$$

where $\hat{L}_k = \alpha\rho\Lambda + \beta\rho I + \delta_k I$ and $\hat{r}_k = \mathcal{F} r_k$. Note that $\hat{L}_k$ can be "trivially" inverted because it is diagonal and positive definite. Therefore, u^{k+1} can be easily obtained by

$$u^{k+1} = \mathcal{F}^\top (\hat{L}_k^{-1} \mathcal{F} r_k). \tag{38}$$

As all variables in (29) can be quickly solved, we propose Algorithm 1, called **TVL1rec**, to solve problem (16). As

---

Algorithm 1 TVL1 Reconstruction Algorithm (**TVL1rec**)

Input α, β, ϵ and ρ. Set $u^0 = c = 0$, $b = 0$, $\delta_0 = 1$, $k = 0$.
repeat
  Given u^k, compute w^{k+1} and z^{k+1} using (34) and (35);
  Given w^{k+1} and z^{k+1}, compute u^{k+1} using (38);
  Update b^k, c^k and δ_k as in (29);
  $k \leftarrow k + 1$
until $\|u^k - u^{k-1}\|/\|u^k\| < \epsilon$.
return u^k

---

discussed above, w and z can be updated using soft shrinkages and hence the computational costs are linear in terms of N. The update of u involves two fast Fourier transforms (FFTs) which have computational complexity $N \log N$ and two operations of A (one is $A^\top$). If $\beta > 0$ there are also two wavelet transforms (in z- and u- steps) involved which require similar computational cost as FFT. Therefore, unlike most recently developed algorithms, our algorithm can deal with arbitrary matrix A and even more general H with nonlinear constraint (as long as H is convex and ∇H is computable). Also, the per iteration computation of the proposed algorithm is very cheap, and thanks to the BB step size δ_k, the convergence speed is significantly improved compared to other two modern methods BOS and OS, as shown in Section IV.

III. METHOD

Experiments were designed to test the effectiveness of the proposed algorithm **TVL1rec** on PPI reconstructions. To demonstrate the potential in clinic applications, the three data sets used in the experiments were acquired by commercially available 8-element head coils. For comparison, we also implemented the Bregman operator splitting algorithm (BOS) [18] and a compressive MR image reconstruction algorithm based on operator splitting (OS) [24] for solving (1).

TABLE I
TESTS NUMBER, DATA INFORMATION AND PARAMETERS.

No.	Image	Abbrev.	Size($\times 8$)	P	(α, β)
1	Cart.Sag.	data1	512×512	1	(1e-5$\sim$1e-2,0)
2	Cart.Sag.	data2	256×250	2	(1e-4,5e-5)
3	Rad.Axi.	data3	256×512	3	(1e-4,5e-5)

A. Data Acquisition

The first data set (top left in Figure 2), termed by data1, is a set of sagittal Cartesian brain images acquired on a 3T GE system (GE Healthcare, Waukesha, Wisconsin, USA). The data acquisition parameters were FOV 220mm^2, size $512\times512\times8$, TR 3060ms, TE 126ms, slice thickness 5mm, flip angle $90°$, and phase encoding direction was anterior-posterior.

The second data set (left in Figure 4) is a Cartesian brain data set acquired on a 3.0T Philips scanner (Philips, Best, Netherlands) using T2-weighted turbo spin echo (T2 TSE) sequence. The acquisition parameters were FOV 205mm^2, matrix $512 \times 500 \times 8$, TR 3000ms, TE 85ms, and the echo train length was 20. To avoid similar comparison plot due to the same data size, we reduce the image to $256 \times 250 \times 8$ and obtain full k-space data of this same size, termed by data2.

The last one (right of Figure 4), denoted by data3, is a radial brain data set acquired on a 1.5T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were FOV 220mm^2, matrix $256\times512\times 8$ (256 radial lines), slice thickness 5mm, TR 53.5ms, TE 3.4ms, and flip angle $75°$.

All three data sets were fully acquired, and then artificially down-sampled using the masks in Figure 1 for reconstruction [3]. As the overall coil sensitivities of these three data sets are fairly uniform, we set the reference image to the root of sum of squares of images which are obtained by fully acquired k-space of all channels A summary of the data information is in Table I. In Table I, "Cart.Sag." means "Cartesian sagittal brain image", and "Rad.Axi." stands for "radial axial brain image". The column P in Table I present the mask number (refer to Figure 1).

B. Test Environment

All algorithms were implemented in the Matlab programming environment (Version R2009a, MathWorks Inc., Natick,

---

[3]The pseudo random sampling can be easily set in 3D imaging. In test2 we simulated the pseudo random trajectory for 2D PPI.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
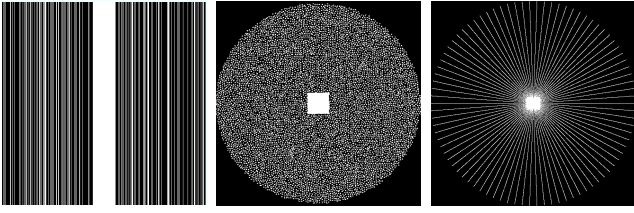
6

Fig. 1. k-space masks used (from left to right) for data1, data2 and data3, respectively. Left: Cartesian mask with net reduction factor 3. Middle: Pseudo random mask [28] with reduction factor 4. Right: Radial mask with 43 (out of 256) projections, i.e. reduction factor 6.

MA, USA). The sparsifying operator Ψ is set to Haar wavelet transform using Rice wavelet toolbox with default settings. The experiments were performed on a Dell Optiplex desktop with Intel Dual Core 2.53 GHz processors (only 1 core was used in computation), 3GB of memory and Windows operating system.

Theoretically the choice of ρ does not effect the convergence of TVL1rec. This is also demonstrated by our experiments since the results are not sensitive to ρ for a large range. Therefore in all experiments we set ρ to a moderate value 10. Algorithm 1 is automatically terminated if the relative change of u^k is less than a prescribed tolerance ϵ. In all of our experiments, we set $\epsilon = 10^{-3}$. Note that smaller ϵ leads to slightly better accuracy at the cost of more iterations and longer computational time. Other parameter settings are shown in the next section. For all algorithms tested in this paper, the sensitivity maps S_j's were estimated from the central 32×32 k-space data (which was a subset of the acquired partial data) and then fixed during the reconstructions, and the initial u^0 was set to 0.

The reconstruction results were evaluated qualitatively by zoomed-in regions of the reconstructed images, and quantitatively by relative error (to the reference image) and CPU times. Reference and reconstructed images corresponding to data1 and data3 were brightened by 3 times, and those corresponding to data2 were brightened by 2 times, to help visual justifications.

IV. COMPARISON ALGORITHMS AND RESULTS

A. Comparison with BOS

In the first experiment, we use data1 with a Cartesian sampling pattern (left in Figure 1) to undersample k-space data. We compare TVL1rec with BOS which also solves (1) via a variable splitting framework (16). To simplify comparison, we here set $\beta = 0$ in (1) and focus on the computational efficiency of two algorithms in solving (1).

The BOS algorithm solves (1) by iterating

$$
\begin{cases}
s^{k+1} = u^k - \delta^{-1} A^\top (Au^k - f) \\
w_i^{k+1} = \arg\min_{w_i} \left\{ \|w_i\| + \frac{\rho}{2} \|w_i - D_i u^k - b_i^k\|^2 \right\}, \quad \forall i \\
u^{k+1} = \arg\min_u \{\alpha\rho \|Du - w^{k+1} + b^k\|^2 + \delta \|u - s^{k+1}\|^2\} \\
b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), \quad \forall i
\end{cases}
$$

$$(39)$$

and converges if $\delta \geq \|A^\top A\|_2$, i.e. the largest eigenvalue of $A^\top A$. In SENSE applications, the magnitudes of sensitivity maps are usually normalized into $[0, 1]$. Therefore from the definition of A in (4), we have $\delta \geq \|A^\top A\|_2 = 1$ and hence set $\delta = 1$ for optimal performance of BOS. With $\beta = 0$, TVL1rec only updates w, u, b and δ in (29). As can be seen, the per iteration computational costs for BOS and TVL1rec are almost identical: the main computations consist of one shrinkage, A, $A^\top$ and two FFTs (including one inverse FFT). Therefore the computation cost for a complete reconstruction is nearly proportional to the number of iterations required by BOS and TVL1rec. In this paper, we set the stopping criterion of BOS the same as TVL1rec, namely the computation will be automatically terminated when the relative change of the iterate u^k is less than $\epsilon = 10^{-3}$.

Table II shows the performance results of TVL1rec and BOS on data1 for different values of TV regularization parameter α. In Table II, we list the following quantities: the relative error of the reconstructed images to the reference image (Err), the final objective function values (Obj), the number of iterations (Iter), and the CPU time in seconds (CPU). From Table II, we

TABLE II
RESULTS OF BOS AND TVL1REC ON DATA1.

	BOS				TVL1rec			
α	Err	Obj	Iter	CPU	Err	Obj	Iter	CPU
1e-5	8.1%	.281	33	75.1	7.2%	.252	7	18.6
1e-4	7.4%	1.01	17	38.9	7.1%	.860	11	26.7
1e-3	7.4%	6.00	39	88.2	7.3%	5.98	7	16.0
1e-2	11.5%	41.0	63	142.1	10.6%	40.7	7	15.9

can see that both BOS and TVL1 are able to stably recover the image from 34% k-space data. This is further demonstrated by Figure 2, where both method generated images very close to the reference image. Although there are still few observable aliasing artifacts due to Cartesian undersampling, the details such as edges and fine structures were well preserved in both reconstructions, as can be seen in the zoomed-ins in the right column of Figure 2. In terms of accuracy, TVL1rec gives slightly better reconstruction quality in the sense of lower relative error and objective values.

In terms of efficiency, we found that TVL1rec significantly outperforms BOS by requiring much fewer iterations (and hence less CPU time) to obtain the similar or even better image quality, as shown in Table II. Compared to BOS, TVL1rec is up to 9 times faster and hence has much higher efficiency. Although two algorithms have almost the same computational costs per iteration, TVL1rec benefits from the adaptive choice of step sizes and readily outperforms BOS which uses fixed step size $\delta = \|A^\top A\|_2$ throughout the computations. The adaptive step size selection makes TVL1rec exhibits a quasi-Newton convergence behavior in some sense because $\delta_k I$ implicitly uses partial Hessian (second order) information.

The adaptive step size selection not only leads to higher efficiency but also better stableness of TVL1rec. As shown in Table II, for a large range of α in $[10^{-5}, 10^{-2}]$, TVL1rec always requires 11 or fewer iterations to recover high quality images. In comparison, BOS appears to be quite sensitive to the choice of α: this is exemplified by the last row ($\alpha = 10^{-2}$)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
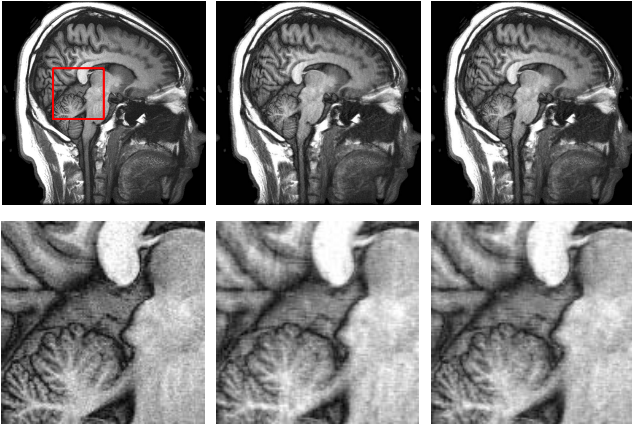
7

Fig. 2. Comparison of BOS and TVL1rec on data1 (top left). Bottom row shows the zoomed-in (square in data1) of images in the top row. From left to right: reference image, reconstructed image using BOS (Err=7.4%), and reconstructed image using TVL1rec (Err=7.1%).
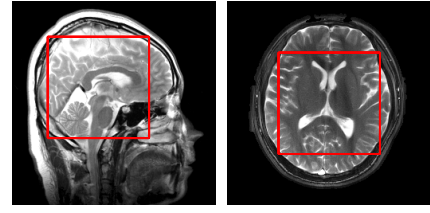

Fig. 3. Testing data used for the comparison of OS and TVL1rec: data2 (left) and data3 (right).
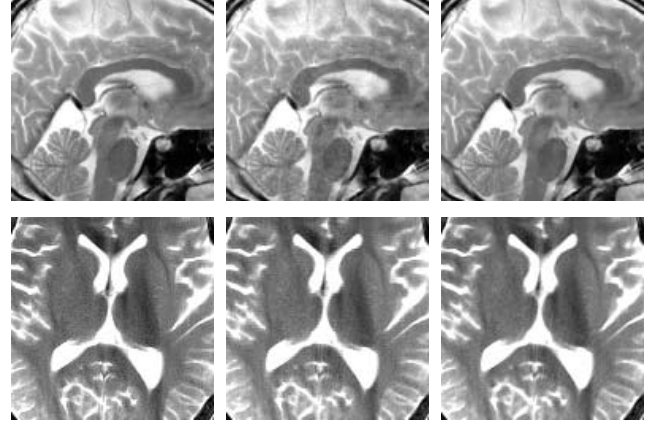

Fig. 4. Reconstructions of data2 and data3 by OS and TVL1rec. Top row (results of data2) from left to right: reference, reconstructed images by OS (Err=7.6%) and TVL1rec (Err=4.6%). Bottom row (results of data3) from left to right: reference, reconstructed images by OS (Err=6.7%) and TVL1rec (Err=6.1%).

of Table II, where BOS required much more iterations than usual; meanwhile, TVL1rec benefits from the optimal step size in each iteration and readily approximates the solution in only few iterations.

The better performance of TVL1rec over BOS relies on two phases: one is that TVL1rec imposes proximity terms not only for u but also for w and z in (29), which lead to better choices of the updates w^{k+1} and z^{k+1}; the other one is the adoption of BB method for optimal penalty parameters δ_k selection, which affects the updates of all variables as in (29) and leads much improved convergence speed.

B. Comparison with OS

For data2 and data3, we compare TVL1 with OS [24] for solving (1) with both TV and ℓ_1 terms ($\alpha = 10^{-4}, \beta = \alpha/2$). The OS scheme of [24], with a minor correction, is as follows:

$$\begin{cases} s^{k+1} = \Psi\left(u^k - (d_1/\lambda) \cdot \left(D^\top w^k + \lambda A^\top(Au^k - f)\right)\right), \\ t_i^{k+1} = w_i^k + d_2 D_i u^k, \ \forall i, \\ u^{k+1} = \Psi\left(\text{sign}(s^{k+1}) \odot \max\{|s^{k+1}| - d_1\tau/\lambda, 0\}\right), \\ w_i^{k+1} = \min(1, \|t_i^{k+1}\|) \cdot t_i^{k+1}/\|t_i^{k+1}\|_2, \ \forall i. \end{cases}$$
(40)

where $s^k \in \mathbb{C}^N$, t_i^k and $w_i^k \in \mathbb{C}^2$, $i = 1, \cdots, N$, $w^k \in \mathbb{C}^{2N}$ is formed by stacking the two columns of matrix $(w_1^k, \cdots, w_N^k)^\top$, and the "max" and "sign" operations in the computation of u^{k+1} are componentwise operations corresponding to shrinkage. The main computational cost per iteration in the OS scheme corresponds to the following operations: a 2D shrinkage during the computation of w^{k+1}, a projection during the computation of u^{k+1}, two wavelet transforms during the computation of s^{k+1} and u^{k+1}, A and $A^\top$ during the computation of s^{k+1}. In [24] it is shown that for $d_1, d_2 > 0$ in certain ranges, the OS scheme converges to a fixed point which is also a solution of (1). The iterations were stopped when either the following conditions were satisfied:

$$\|u^{k+1} - u^k\|_2/\max\{1, \|u^k\|_2\} < \epsilon_1$$
$$(f^k - f^{k+1})/\max\{1, f^k\} < \epsilon_2\sqrt{\tau_c/\tau_t},$$
(41)

where f^k is the objective value of (1) at u^k, τ_c and τ_t are the current and target values of τ respectively and ϵ_1 and ϵ_2 are prescribed stopping tolerances.

Since OS has multiple tuning parameters that affect the convergence speed and image quality: larger d_i's and ϵ_i's lead to faster convergence but result in larger relative error, whereas smaller d_i's and ϵ_i's yield monotonic decreases in objective values and better image quality at the cost of much longer computation. Based on the selection by the authors and several tries, we chose moderate values $d_1 = d_2 = 1$, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-3}$ which appear to give a best compromise between convergence speed and image quality of the OS scheme. The results on data2 and data3 are shown in Figure 4, and the comparison on relative errors and objective values are plotted in logarithmic scale in Figure 5. The horizontal label is chosen as CPU time because the per iteration computational costs for OS and TVL1rec are slightly different.

From Figures 4 and 5 we can see that TVL1rec converges much faster than OS, and achieved lower relative errors and objective values than OS overall. Therefore, it is evident that TVL1rec can outperform OS scheme in efficiency as the former requires much less computational time to reach the similar or even better image quality. It is also worth pointing out that both algorithms can further reduce the relative error slightly by setting a tighter stopping criterion at the cost of more iteration numbers. Nevertheless, the TVL1rec still can maintain lower relative error and objective value than OS during the reconstruction process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
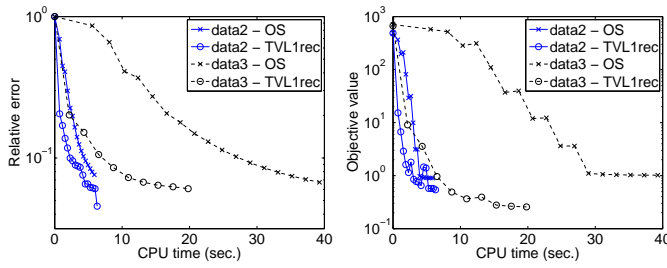
8

Fig. 5. Comparisons of OS and TVL1rec on data2 (blue solid lines) and data3 (black dashed lines). Left: relative error (in logarithm) versus CPU time and objective value. Right: objective values (in logarithm) versus CPU time.
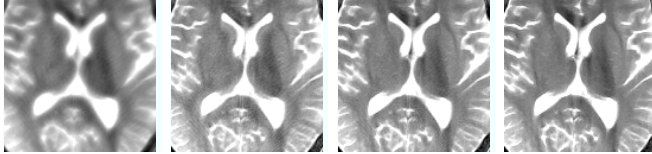

Fig. 6. From left to right: reconstructions of data3 by TVL1rec at the 1st, 4th, 10th and 20th iterations, respectively.

We checked the reconstructions of data3 by TVL1rec at the 1st, 4th, 10th and 20th iterations and depicted the corresponding zoomed-in regions in Figure 6. Recall that the main computational cost for each iteration is only $2(J + 1)$ FFTs and 2 wavelet transforms as shown in (29) and (4), we further demonstrate that TVL1rec can quickly remove artifacts and recover fine structures in CS-PPI reconstructions.

V. Conclusion

This paper presents a fast numerical algorithm, called TVL1rec, for TVL1 minimization problem (1) arising from CS reconstruction problems. The proposed algorithm incorporates the Barzilai-Borwein (BB) method into a variable splitting framework to optimize the selection of step sizes. The optimal step sizes exploit partial Hessian information and hence lead to a quasi-Newton convergence behavior of TVL1rec. Experimental results demonstrate the outstanding efficiency of the proposed algorithm in CS-PPI reconstruction.

We compared TVL1rec to another two recently developed algorithms BOS [18] and OS [24] which also solve the minimization problem (1). The common property of these algorithms is that they can deal with general sensing matrix A, and even nonlinear data fidelity term $H(u)$ other than $\|Au - f\|^2$ as long as H is convex and ∇H is computable. Meanwhile, TVL1rec significantly outperforms the other two algorithms by taking advantages of the optimal step size selection based on BB method. We hope TVL1rec can be beneficial to PPI and other medical imaging applications.

References

[1] K. Block, M. Uecker, and J. Frahm, "Undersampled radial mri with multiple coils: Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007.

[2] H. Eggers and P. Boesiger, "Gridding- and convolution-based iterative reconstruction with variable k-space resolution for sensitivity-encoded non-cartesian imaging," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2003, p. 2346.

[3] K. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, pp. 638–651, 2001.

[4] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, "Sense: Sensitivity encoding for fast mri," *Magn. Reson. Med.*, vol. 42, pp. 952–962, 1999.

[5] Y. Qian, Z. Zhang, V. Stenger, and Y. Wang, "Self-calibrated spiral sense," *Magn. Reson. Med.*, vol. 52, pp. 688–692, 2004.

[6] P. Qu, K. Zhong, B. Zhang, J. Wang, and G.-X. Shen, "Convergence behavior of iterative sense reconstruction with non-cartesian trajectories," *Magn. Reson. Med.*, vol. 54, pp. 1040–1045, 2005.

[7] A. Samsonov and C. Johnson, "Non-cartesian pocsense," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2004, p. 2648.

[8] E. Yeh, M. Stuber, C. McKenzie, R. B. RM, T. Leiner, M. Ohliger, A. Grant, J. Willig-Onwuachi, and D. Sodickson, "Inherently self-calibrating non-cartesian parallel imaging," *Magn. Reson. Med.*, vol. 54, pp. 1–8, 2005.

[9] L. Ying, B. Liu, M. Steckner, G. Wu, and S.-J. L. M. Wu, "A statistical approach to sense regularization with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 60, pp. 414–421, 2008.

[10] A. Larson, R. White, G. Laub, E. McVeigh, D. Li, and O. Simonetti, "Self-gated cardiac cine mri," *Magn. Reson. Med.*, vol. 51, pp. 93–102, 2004.

[11] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[12] J. Yang, Y. Zhang, and W. Yin, "A fast tvl1-l2 minimization algorithm for signal reconstruction from partial fourier data," CAAM Rice Univ., Tech. Rep. 08-29, 2008.

[13] T. Goldstein and S. Osher, "The split bregman method for l1 regularized problems," *SIAM J. Imag. Sci.*, vol. 2, pp. 323–343, 2009.

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

[15] J. Eckstein and D. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

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

[17] R. Glowinski and A. Marrocco, "Sur lapproximation par elements nis dordre un, et la resolution par penalisation-dualite dune classe de problemes de dirichlet nonlineaires, rev. francaise daut." *Inf. Rech. Oper.*, vol. R-2, pp. 41–76, 1975.

[18] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," CAM UCLA, Tech. Rep. 09-03, 2009.

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

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

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

[22] M. Zhu, S. Wright, and T. Chan, "Duality-based algorithms for total-variation-regularized image restoration," *Comput. Optim. Appl.*, 2008.

[23] J. Barzilai and J. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

[24] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed mr imaging using total variation and wavelets," *IEEE Proc. Conf. on Comp. Vis. Patt. Recog.*, pp. 1–8, 2008.

[25] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[26] S. J. Wright, R. D. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.

[27] H. Akaike, "On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method," *Ann. Inst. Statist. Math. Tokyo*, vol. 11, pp. 1–17, 1959.

[28] M. Lustig and J. Pauly, "Spirit: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magn. Reson. Med.*, vol. 64, no. 2, pp. 457–471, 2010.