SIAM J. IMAGING SCIENCES
Vol. 5, No. 1, pp. 90–118

Fast Algorithms for Image Reconstruction with Application to Partially Parallel MR Imaging*

Yunmei Chen†, William Hager†, Feng Huang‡, Dzung Phan§, Xiaojing Ye¶, and Wotao Yin‖

Abstract. This paper presents two fast algorithms for total variation–based image reconstruction in a magnetic resonance imaging technique known as partially parallel imaging (PPI), where the inversion matrix is large and ill-conditioned. These algorithms utilize variable splitting techniques to decouple the original problem into more easily solved subproblems. The first method reduces the image reconstruction problem to an unconstrained minimization problem, which is solved by an alternating proximal minimization algorithm. One phase of the algorithm solves a total variation (TV) denoising problem, and the second phase solves an ill-conditioned linear system. Linear and sublinear convergence results are given, and an implementation based on a primal-dual hybrid gradient (PDHG) scheme for the TV problem and on a Barzilai–Borwein scheme for the linear inversion is proposed. The second algorithm exploits the special structure of the PPI reconstruction problem by decomposing it into one subproblem involving Fourier transforms and another subproblem that can be treated by the PDHG scheme. Numerical results and comparisons with recently developed methods indicate the efficiency of the proposed algorithms.

Key words. image reconstruction, variable splitting, TV denoising, nonlinear optimization

AMS subject classifications. 90C25, 68U10, 65J22

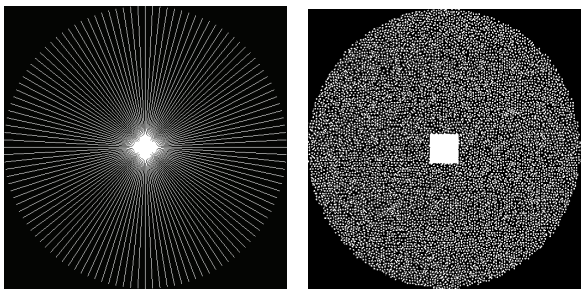
DOI. 10.1137/100792688

1. Introduction.

In this paper we provide fast numerical algorithms for image reconstruction problems that arise from an emerging magnetic resonance (MR) medical imaging technique known as *partially parallel imaging* (PPI). MR imaging is commonly used in radiology to visualize the internal structure and function of the body by noninvasive and nonionizing means. It provides better contrast between the different soft tissues than most other modalities. MR images are obtained through an inversion of Fourier data acquired by the receiver coil(s). The practical performance of inversion algorithms in terms of image quality and reconstruction speed is crucial in clinical applications.

MR images are obtained by placing an object in a strong magnetic field and then turning on and off a radio frequency electromagnetic field. Different body parts produce different signals

*Received by the editors April 19, 2010; accepted for publication (in revised form) October 24, 2011; published electronically January 24, 2012. This research was partly supported by National Science Foundation grants 0620286 and 1115568 and by Office of Naval Research grant N00014-11-1-0068.
http://www.siam.org/journals/siims/5-1/79268.html
†Department of Mathematics, University of Florida, Gainesville, FL 32611 (yun@ufl.edu, hager@ufl.edu).
‡Invivo Corporation, 3545 SW 47th Avenue, Gainesville, FL 32608 (f.huang@philips.com).
§Department of Business Analytics and Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 (phandu@us.ibm.com).
¶Corresponding author. Department of Mathematics, University of Florida, Gainesville, FL 32611. Current address: School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 (xye33@math.gatech.edu).
‖Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005 (wotao.yin@rice.edu).



(a) Radial Mask (b) Poisson Mask

Figure 1. (a) *A radial mask.* (b) *A Poisson pseudorandom mask.*

which are detected by a receiver. The resulting data is then inverted to obtain an image of the scanned object. In PPI, the image quality and reconstruction speed are improved by surrounding the scanned objects by multiple receivers and collecting in parallel part of the Fourier components at each receiver.

The undersampling patterns of the Fourier coefficients are often described by a mask. Figure 1(a) shows a radial mask, while Figure 1(b) shows a Poisson pseudorandom mask for a two-dimensional image. The white pixels correspond to the Fourier components which are measured. The white region in the center of the masks indicates that the low frequency Fourier components are all measured. The white rays in the radial mask in the surrounding darker region show the spacing between the higher frequency Fourier components that are measured. In the Poisson pseudorandom mask, about 1/4 of the Fourier components are measured.

Partial data acquisition increases the spacing between read-out lines, thereby reducing scan time; however, this reduction in the number of recorded Fourier components leads to aliasing artifacts in images which must be removed by the image reconstruction process. Image reconstruction in PPI is much different from either denoising and deblurring problems for which there are a number of algorithms. The PPI reconstruction problem leads to a large and ill-conditioned inversion matrix with much less structure than the matrices associated with denoising or deblurring problems. There are two general approaches for removing the aliasing artifacts and reconstructing high quality images: image domain–based methods and k-space–based methods. The k-space–based methods use coil sensitivity variations to reconstruct the missing k-space data, and then apply the Fourier transform to the original and reconstructed data to obtain the unaliased image [3, 22, 26]. In this paper, we employ image domain methods and coil sensitivity maps to reconstruct the underlying image [8, 15, 32, 33, 34, 35, 37, 45, 46].

Sensitivity encoding (SENSE) is the most common image domain–based parallel imaging method. It is based on the following equation which relates the partial k-space data f_j, acquired by the jth receiver, to the sensitivity map S_j and the mask M:

$$M\mathcal{F}S_j u = f_j. \tag{1}$$

Here f_j is the vector of measured Fourier coefficients at receiver j, M corresponds to the mask which is obtained by extracting from the identity those rows corresponding to the measured Fourier components, $\mathcal{F}$ is the Fourier transform, $S_j \in \mathbb{C}^{N\times N}$ is the diagonal sensitivity map for receiver j, and $u \in \mathbb{C}^N$ is the underlying image obtained by stacking all columns of the image

to form a one-dimensional vector. The sensitivity map is a diagonal matrix whose diagonal elements estimate the impact of a pixel in the image on the measured Fourier coefficients. Pixels closest to a receiver may have more impact on the signal than pixels far away from the receiver. There is one diagonal element in S_j corresponding to each pixel in the image.

Based on (1), the reconstruction of the image u could be accomplished by solving the least squares problem

$$\min_{u\in\mathbb{C}^N} \sum_{j=1}^{K} \|M\mathcal{F}S_j u - f_j\|_2^2, \tag{2}$$

where $\|\cdot\|_2$ is the 2-norm (Euclidean norm) and K is the number of channels (or receivers). Since (2) often does not have a unique solution, the minimization problem can be ill-conditioned. To alleviate the effect of the ill-conditioning, the SENSE model (2) has been improved recently by incorporating regularization terms into the energy functional to take advantage of the underlying sparsity of MR images in the finite difference domain [9, 27]. The images are recovered by solving an optimization problem of the form

$$\min_{u\in\mathbb{C}^N} \|u\|_{TV} + \lambda \sum_{j=1}^{K} \|M\mathcal{F}S_j u - f_j\|_2^2, \tag{3}$$

where $\|\cdot\|_{TV}$ is the total variation seminorm and $\lambda > 0$ is a parameter corresponding to the relative weight of the data fidelity term

$$\sum_{j=1}^{K} \|M\mathcal{F}S_j u - f_j\|_2^2.$$

The term $\|u\|_{TV}$ controls the solution sparsity. The general form of the image reconstruction problems is

$$\min_{u\in\mathbb{C}^N} J(u) + H(u), \tag{4}$$

where J is a convex and possibly nondifferentiable function and H is convex and continuously differentiable. In TV-based image reconstruction problems, J and H, respectively, have the form

$$J(u) = \|u\|_{TV} \quad \text{and} \quad H(u) = \lambda\|Au - f\|_2^2, \tag{5}$$

where f is the measured data and A is a possibly large and ill-conditioned matrix describing the imaging device or the data acquisition pattern. In the PPI problem (3),

$$A = \begin{pmatrix} M\mathcal{F}S_1 \\ \vdots \\ M\mathcal{F}S_K \end{pmatrix}, \tag{6}$$

and f is the vector formed from the data collected by the K receivers.

TV-based regularization was originally introduced in image processing by Rudin, Osher, and Fatemi in their pioneering work [36] for denoising. A significant advantage of TV regularization is that it preserves edges in the solution. The TV term in (5) leads to an underlying sparse solution of $Au = f$. The lack of smoothness in the TV term makes the solution of (4) difficult. In recent years, many algorithms were proposed to efficiently solve the TV-based image reconstruction problem (4). An overview of these algorithms will be provided in the next section. Note that the efficiency of most algorithms relies on a very special structure for the matrix A. For example, A is either the identity or diagonalizable by a discrete Fourier or cosine transform. Therefore, they do not directly apply to the PPI problem (3).

Our paper is organized as follows. In section 2 we give an overview of TV-based image reconstruction techniques. In section 3 we present two algorithms that we have used to solve the PPI problem (3). The first algorithm is based on the very general splitting $v = u$, while the second algorithm uses the PPI-based splitting $v_j = S_j u$. Section 4 studies the convergence rate of the first algorithm. Finally, section 5 uses PPI images to compare our algorithms to recently developed methods.

Notation. For a differentiable function, ∇f denotes the gradient of f, a row vector. More generally, $\partial J(x)$ denotes the subdifferential set at x, a set of row vectors. For any matrix M, $\mathcal{N}(M)$ is the null space of M. x^{T} denotes the conjugate transpose of the vector x, and $\langle x, y\rangle = x^{\mathsf{T}} y$ is the Euclidean inner product. $\|\cdot\|_p$ is the p-norm, and $\|\cdot\|_{TV}$ is the discrete total variation seminorm. A list of matrices (or vectors) separated by semicolons, such as $(A; B)$, where A and B have the same number of columns, denotes the stacked matrix with A on top of B.

2. Related work.

The image reconstruction problem (4)–(5) is equivalent to solving the problem

$$\min_{u\in\mathbb{C}^N} \|u\|_{TV} + \lambda\|Au - f\|_2^2, \tag{7}$$

where $\|\cdot\|_{TV}$ is the discrete (isotropic) TV seminorm defined by

$$\|u\|_{TV} \triangleq \sum_{i=1}^{N} \|D_i u\|_2, \tag{8}$$

where $D_i u \in \mathbb{R}^2$ contains the forward finite differences of u along its first and second dimensions, and N is the number of pixels in the image. The early work on algorithms for (7) used gradient descent methods with explicit [36] or semi-implicit schemes [24, 39] in which the TV norm was replaced by a smooth approximation

$$\|u\|_{TV,\epsilon} = \sum_{i=1}^{N} \sqrt{\|D_i u\|_2^2 + \epsilon}. \tag{9}$$

The choice of $\epsilon > 0$ was crucial to the reconstruction results and convergence speed. A large ϵ encourages a fast convergence rate but fails to preserve high quality details such as edges in the restored image; a small ϵ better preserves fine structure in the reconstruction at the expense of slow convergence.

In [40, 42], a method is developed based on the following reformulation of (7):

$$\min_{u,w} \sum_{i=1}^{N} \|w_i\|_2 + \lambda \|Au - f\|_2^2 \quad \text{subject to} \quad w_i = D_i u,\ i = 1, \ldots, N. \tag{10}$$

The linear constraint is treated with a quadratic penalty

$$\min_{u,w} \sum_{i=1}^{N} \|w_i\|_2 + \rho \|Du - w\|_2^2 + \lambda \|Au - f\|_2^2, \tag{11}$$

where $w = (w_1; \ldots; w_N) \in \mathbb{C}^{2N}$ and D is obtained by stacking the D_i matrices. For any fixed ρ, (11) can be solved by alternating minimizations, first over w and then over u. If both $D^\top D$ and $A^\top A$ can be diagonalized by the Fourier matrix, as they can if A is either the identity matrix or a blurring matrix with periodic boundary conditions, then each minimization involves shrinkage and a fast Fourier transform (FFT). A continuation method is used to deal with the slow convergence rate associated with a large value for ρ. The method, however, may not be suitable for more general A.

In [21] Goldstein and Osher develop a split Bregman method for (11). The resulting algorithm has computational complexity similar to that of the algorithm in [40]; 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) [7, 14, 19, 20] applied to the augmented Lagrangian

$$L(w, u, p) \triangleq \sum_{i=1}^{N} \|w_i\|_2 + \lambda \|Au - f\|_2^2 + \langle p, Du - w \rangle + \rho \|Du - w\|_2^2. \tag{12}$$

Nonetheless, the algorithms in [21, 40, 42] 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 [47] replaces $\|Au - f\|_2^2$ by a proximal-like term

$$\delta \|u - (u^k - \delta^{-1} A^\top (Au^k - f))\|_2^2$$

for some $\delta > 0$. BOS is an inexact Uzawa method that depends on the choice of δ. It is generally less efficient than split Bregman when A has special structure.

There are also several methods developed to solve the associated dual or primal-dual formulations of (7) based on the dual formulation of the TV norm:

$$\|u\|_{TV} = \max_{p \in X} \langle p, Du \rangle, \text{ where } X = \{p = (p_1; \ldots; p_N) \in \mathbb{C}^{2N} : p_i \in \mathbb{C}^2, \|p_i\|_2 \le 1,\ 1 \le i \le N\}. \tag{13}$$

Consequently, (7) can be written as a minimax problem,

$$\min_{u \in \mathbb{C}^N} \max_{p \in X} \langle p, Du \rangle + \lambda \|Au - f\|_2^2. \tag{14}$$

In [11], Chan, Golub, and Mulet proposed solving 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 high since the method explicitly uses second-order information and the inversion of a Hessian matrix is required. In [10], Chambolle used the dual formulation of the TV denoising problem (7) 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 [49] 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 [10, 21] for denoising and deblurring problems, but its efficiency again relies on the fact that $A^{\mathsf{T}}A$ can be diagonalized by fast transforms. More recently [50], several variations of PDHG, referred to as projected gradient descent algorithms, were applied to the dual formulation of the image denoising problem 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 [5].

Another approach that can be applied to (4) in the imaging context (5) with a general A is the forward-backward operator splitting (OS) method. In [28] the OS idea of [25] is applied to image reconstruction in compressed MR imaging. The scheme is based on the first-order optimality condition at a local minimizer u^*:

$$0 \in \partial J(u^*) + 2\lambda A^{\mathsf{T}}(Au^* - f).$$

This is rewritten in the form

$$0 \in \partial J(u^*) + \frac{1}{\delta}(u^* - s^*), \quad s^* = u^* - 2\delta\lambda A^{\mathsf{T}}(Au^* - f).$$

The iterative scheme is

$$s^k = u^k - \delta\lambda A^{\mathsf{T}}(Au^k - f),$$

$$u^{k+1} = \arg\min_u \; J(u) + \frac{1}{2\delta}\|u - s^k\|_2^2.$$

The computation of u^{k+1}, given s^k, is a TV denoising problem. If this problem is solved using a split Bregman method [21], then this is equivalent to BOS [47], which can accommodate an arbitrary matrix A. In [43], Ye, Chen, and Huang proposed a variation of BOS utilizing the Barzilai–Borwein step size to significantly improve the efficiency; however, the convergence of the algorithm is not known, although it seems to converge in numerical experiments. Numerical comparisons with the algorithm of [43] are given in section 5.

3. Proposed algorithms. In this section, we give two algorithms based on different variable splittings to solve the TV-based image reconstruction problem (4). The first algorithm is based on the general splitting $v = u$ and the alternating proximal minimization algorithm to solve a penalized problem. The convergence speed is either sublinear or linear depending on the properties of A. The practical performance of this algorithm in the context of PPI is much better than that of many recently developed methods. The second algorithm is specifically

designed for the TV-based SENSE problem in PPI (3). It employs the PPI-based splitting $v_j = S_j u$ and the alternating direction method of multipliers for which convergence is guaranteed. The numerical results in section 5 show high efficiency of these algorithms in PPI image reconstruction.

3.1. The splitting $v = u$ and the alternating minimization algorithm. To cope with the lack of smoothness in J in problem (4), we introduce an auxiliary variable v to obtain the equivalent constrained problem

$$\min_{u,v\in\mathbb{C}^N} J(v) + H(u) \quad \text{subject to} \quad u = v, \quad u, v \in \mathbb{C}^N. \tag{15}$$

The equality constrained problem is converted to an unconstrained problem using a quadratic penalty:

$$\min_{u,v\in\mathbb{C}^N} J(v) + H(u) + \alpha\|v-u\|_2^2, \tag{16}$$

where $\alpha > 0$ is a parameter. The additional variable v allows us to treat the smooth term H and the nondifferentiable term J somewhat independently. Starting from an initial guess u^0, we solve the penalized problem by first minimizing over v with u fixed, and then minimizing over u with v fixed:

$$\left.\begin{aligned} v^{k+1} &= \mathcal{T}(u^k), & \mathcal{T}(u) &\triangleq \arg\min_{v\in\mathbb{C}^N} J(v) + \alpha\|v-u\|_2^2 \\ u^{k+1} &= \mathcal{L}(v^{k+1}), & \mathcal{L}(v) &\triangleq \arg\min_{u\in\mathbb{C}^N} H(u) + \alpha\|v-u\|_2^2 \end{aligned}\right\}. \tag{17}$$

Since J and H are convex, the objective functions in both subproblems are strongly convex. Hence, for any starting guess u^0, the iteration sequence (v^k, u^k), $k \geq 1$, exists and is unique. In the imaging context (5), the iteration is

$$\left.\begin{aligned} v^{k+1} &= \arg\min_{v\in\mathbb{C}^N} \|v\|_{TV} + \alpha\|v-u\|_2^2 && \text{(TV)} \\ u^{k+1} &= \arg\min_{u\in\mathbb{C}^N} \lambda\|Au-f\|_2^2 + \alpha\|v-u\|_2^2 && \text{(LS)} \end{aligned}\right\}. \tag{18}$$

The first subproblem, denoted (TV), is a TV-based image denoising which has been extensively studied in the literature, and the second, (LS), is a least squares problem. Both subproblems can be solved quickly.

In the literature, algorithms of the form (17) are called alternating proximal minimization algorithms. References include [1, 4, 6]. Alternating proximal minimization was recently applied to the TV-based image deblurring problem in [23, 41] and to the TV-based SENSE problem in [44], with different algorithms for the subproblems. The iterates converge to a solution of (16), if a solution exists, according to [6, Cor. 4.5], for example. In general, one needs to let α tend to infinity to obtain the solution of (4). However, our numerical experience in PPI reconstruction indicates that in this application, a suitable approximation to the solution of (4) is generated using a fixed, not very large α.

We now provide implementations for the (TV) and (LS) subproblems of the alternating proximal minimization algorithm (18). One of the reasons that the splitting (15) worked well

was that each of the subproblems could be solved quickly. As discussed earlier, there are many fast algorithms for the (TV) subproblem that take advantage of the simplicity of the $\|v-u\|_2^2$ term. Recent work includes the dual approach in [10, 50], variable splitting and continuation [40, 42], split Bregman [21], and PDHG [49]. In the numerical experiments of section 5, we used a PDHG scheme which is shown to be one of the fastest methods for TV image denoising.

We now explain in detail the PDHG scheme that we use for the (TV) subproblem in (18). Based on the dual formulation of the TV norm (8), the (TV) subproblem can be written as

$$\min_v \sum_{i=1}^{N} \|D_i v\|_2 + \alpha\|v-u\|_2^2 = \min_v \max_{p\in X} \langle p, Dv\rangle + \alpha\|v-u\|_2^2,$$

where $X = \{p = (p_1; \ldots; p_N) \in \mathbb{C}^{2N} : p_i \in \mathbb{C}^2,\ \|p_i\|_2 \le 1,\ i = 1, \ldots, N\}$. The PDHG algorithm is based on the following updates for the primal and dual variables:

$$\left.\begin{aligned} p^{l+1} &= \arg\max_{p\in X}\ \Phi(v^l, p) - \tfrac{1}{2\tau_l}\|p-p^l\|_2^2, \\ v^{l+1} &= \arg\min_{v\in\mathbb{C}^N}\ \Phi(v, p^{l+1}) + \tfrac{1}{2\theta_l}\|v-v^l\|_2^2 \end{aligned}\right\}, \tag{19}$$

where $\Phi(v,p) = \langle p, Dv\rangle + \alpha\|v-u^k\|_2^2$, and θ_l and τ_l represent the primal and dual step sizes corresponding to the regularization terms in (19). Due to the simple form for the quadratic term in Φ, the iteration takes the form given in Algorithm 1.

Algorithm 1. PDHG [49] for the (TV) subproblem.

$$p^{l+1} = \Pi_X(p^l + \tau_l D v^l), \quad (\Pi_X(p))_i = p_i/\max\{\|p_i\|_2, 1\} \quad \forall\ i, \tag{20}$$

$$v^{l+1} = (1+2\alpha\theta_l)^{-1}(v^l - \theta_l D^{\mathsf{T}} p^{l+1} + 2\alpha\theta_l u^k)(1-\theta_l)v^l + \theta_l\left(u^k - (1/2\alpha)D^{\mathsf{T}}p^{l+1}\right). \tag{21}$$

In Algorithm 1, $\Pi_X : \mathbb{C}^{2N} \to \mathbb{C}^{2N}$ is the projection onto X. For the step (21), v^{l+1} is a linear combination of v^l, u^k, and $D^{\mathsf{T}}p^{l+1}$. The authors in [49] suggested that the step size be updated by the rule $\tau_l = 0.2 + 0.08l$, $\theta_l = (0.5 - \frac{5}{15+l})/\tau_l$ for improved efficiency; however, PDHG with constant step sizes already outperforms most other methods. In our experiments, we use the suggested updates for τ_l and θ_l. Note that both steps in Algorithm 1 require only pointwise operations and hence can be computed in parallel. Based on the results given in [49], Algorithm 1 is expected to be very efficient.

The (LS) subproblem in (18) is a least squares problem in u. We solve this by Nesterov's optimal gradient algorithm given in [30, 29]; however, we found that comparable or better performance was obtained using the Barzilai–Borwein (BB) method [5]. This could also be solved by a conjugate gradient method, but again, comparable or better performance was obtained using the BB method, which handles ill-conditioning much better than gradient methods with a Cauchy step [2]. The (LS) subproblem has the form

$$\min_u \lambda\|Au-f\|_2^2 + \alpha\|v-u\|_2^2. \tag{22}$$

In the standard implementation of the BB method, the Hessian of the objective function is approximated by a multiple of the identity matrix. For the LS problem, however, the Hessian of $\|v-u\|_2^2$ with respect to u is already a multiple of the identity. Hence, we only approximate the Hessian of $\|Au-f\|_2^2$ by a multiple of the identity. More precisely, if u^k is the current BB iterate, then we employ the approximation

$$\|Au-f\|_2^2 \approx \|Au^k-f\|_2^2 + 2(Au^k-f)^{\mathsf{T}}A(u-u^k) + \delta^k\|u-u^k\|_2^2, \tag{23}$$

where

$$\delta^k = \|A(u^k-u^{k-1})\|_2^2/\|u^k-u^{k-1}\|_2^2.$$

Since the $\|Au^k-f\|_2^2$ term in (23) does not depend on u, the BB method for the (LS) subproblem has the form shown in Algorithm 2.

Algorithm 2. BB method [5] for the (LS) subproblem.

$$u^{k+1} = \arg\min_{u\in\mathbb{C}^N} \lambda\left(2(Au^k-f)^{\mathsf{T}}A(u-u^k) + \delta^k\|u-u^k\|_2^2\right) + \alpha\|v-u\|_2^2. \tag{24}$$

Under suitable assumptions [12, 13, 17], the iteration (24) converges linearly to a solution of (22). Each iteration involves multiplication by A and A^{T}, where A is defined in (6). The time to multiply by M or S_j is proportional to N, while the Fourier transform $\mathcal{F}$ can be performed in time proportional to $N\log(N)$. Hence, each iteration of Algorithm 2 can be performed quickly in our target application PPI.

The scheme (18), with the (TV) subproblem solved by PDHG (Algorithm 1) and with the (LS) subproblem solved by BB (Algorithm 2), will be referred to as the alternating minimization (AM) algorithm. In theory, in order to enforce the constraint $u=v$, we must let α tend to infinity in (18). As an alternative to the penalty method for handling the equality constraint, we could apply the multiplier method. When the iteration is implemented by the alternating proximal minimization algorithm [16, 18, 48], we obtain the ADMM:

$$\left.\begin{aligned} v^{k+1} &= \arg\min_{v\in\mathbb{C}^N} \|v\|_{TV} + \langle b^k, v-u^k\rangle + \alpha\|v-u^k\|_2^2 \\ u^{k+1} &= \arg\min_{u\in\mathbb{C}^N} \lambda\|Au-f\|_2^2 + \langle b^k, v^{k+1}-u\rangle + \alpha\|v^{k+1}-u\|_2^2 \\ b^{k+1} &= b^k + 2\alpha(v^{k+1}-u^{k+1}) \end{aligned}\right\}. \tag{25}$$

ADMM converges to a solution of (15), while AM reaches a solution of (15) only in the limit, as α tends to infinity. However, we found that in our target application PPI, ADMM, and AM have almost identical performances (see Figures 7, 8, and 11 of section 5).

Another approach for treating the penalty term in (17) is the continuation method where the value of α is gradually increased. The solution for a previous α is used as a "warm start" for the next larger α. However, in our numerical tests with PPI data sets, we found it was more efficient to simply take a fixed, not very large value of α. The reason is the following: The image reconstruction techniques are designed to minimize the TV-based energy (7). The ground truth, however, is typically not a minimizer of (7). As the penalty α in the AM

algorithm tends to infinity, the iterates approach a solution of (7); however, at some point, the iterates may increase their distance to the ground truth since it does not minimize (7). We found that α does not need to be very large for a suitable image reconstruction, and that further increases in α may not improve the image quality. And in the case where α is not very large, there was no significant increase in efficiency when we implemented a continuation scheme.

3.2. The splitting $v_j = S_j u$ and ADMM. Another approach that we consider for the TV-based SENSE problem in PPI is based on the substitution $v_j = S_j u$ in (3). This leads to the problem

$$\min_{u,v_j} \ \|u\|_{TV} + \lambda \sum_{j=1}^{K} \|M\mathcal{F}v_j - f_j\|_2^2, \quad v_j = S_j u. \tag{26}$$

We employ the following augmented Lagrangian associated with (26):

$$\|u\|_{TV} + \lambda \sum_{j=1}^{K} \left(\|M\mathcal{F}v_j - f_j\|_2^2 + 2\alpha\langle b_j, v_j - S_j u\rangle + \alpha\|v_j - S_j u\|_2^2 \right). \tag{27}$$

In this context, ADMM is

$$\left.\begin{aligned} v_j^{k+1} &= \arg\min_{v_j \in \mathbb{C}^N} \ \|M\mathcal{F}v_j - f_j\|_2^2 + 2\alpha\langle b_j^k, v_j - S_j u^k\rangle + \alpha\|v_j - S_j u^k\|_2^2, \quad j = 1, \ldots, K \\ u^{k+1} &= \arg\min_{u \in \mathbb{C}^N} \ \|u\|_{TV} + \alpha\lambda \sum_{j=1}^{K} \left(2\langle b_j^k, v_j^{k+1} - S_j u\rangle + \|S_j u - v_j^{k+1}\|_2^2 \right) \\ b_j^{k+1} &= b_j^k + (v_j^{k+1} - S_j u^{k+1}), \quad j = 1, \ldots, K \end{aligned}\right\}. \tag{28}$$

In (28), v_j^{k+1} can be computed quickly since the matrix in the normal equation is

$$\mathcal{F}^{\mathsf{T}} M^{\mathsf{T}} M \mathcal{F} + \alpha I = \mathcal{F}^{\mathsf{T}} (M^{\mathsf{T}} M + \alpha I)\mathcal{F},$$

which is the product of Fourier transforms and a diagonal matrix. The solution to the u-subproblem in (28) was computed using the PDHG scheme. To put the u-subproblem into the framework for the PDHG scheme, observe that the objective function in the u-subproblem can be expressed as

$$\min_{u \in \mathbb{C}^N} \max_{p \in X} \ \Phi(u,p) \triangleq \langle p, Du\rangle + \alpha\lambda \sum_{j=1}^{K} \left(2\langle b_j^k, v_j^{k+1} - S_j u\rangle + \|S_j u - v_j^{k+1}\|_2^2 \right). \tag{29}$$

The PDHG iteration is then written as

$$\begin{aligned} p^{l+1} &= \arg\max_{p \in X} \ \Phi(u^l, p) - \frac{1}{2\tau_l}\|p - p^l\|_2^2, \\ u^{l+1} &= \arg\min_{u \in \mathbb{C}^N} \ \Phi(u, p^{l+1}) + \frac{1}{2\theta_l}\|u - u^l\|_2^2. \end{aligned} \tag{30}$$

The computation of p^{l+1} reduces to the projection given in (20). The computation of u^{l+1} is trivial since the matrix in the normal equation is $I + 2\alpha\lambda\theta_l \sum_{j=1}^{K} S_j{}^{\mathsf{T}} S_j$, a diagonal matrix. Therefore, Algorithm 3 requires only pointwise operations which can be computed in parallel. A more detailed statement of the PDHG algorithm in this context appears in Algorithm 3. The alternating direction method (28) with the u-subproblem solved by the PDHG scheme is referred to as the APD algorithm.

Algorithm 3. PDHG [49] for the (TV) subproblem in (28).

$$p^{l+1} = \Pi_X(p^l + \tau_l D u^l), \tag{31}$$

$$u^{l+1} = \left(I + 2\alpha\lambda\theta_l \sum_{j=1}^{K} S_j{}^{\mathsf{T}} S_j \right)^{-1} \left(u^l + 2\alpha\lambda\theta_l \sum_{j=1}^{K} S_j{}^{\mathsf{T}} (b_j^k + v_j^{k+1}) - \theta_l D^{\mathsf{T}} p \right). \tag{32}$$

4. Convergence analysis. In this section, we examine the convergence rate of the alternating proximal minimization scheme (17). Since H is convex, there exists a constant $\sigma \geq 0$ such that the following monotonicity condition holds for all u and $v \in \mathbb{C}^n$:

$$(\nabla H(u) - \nabla H(v))(u - v) \geq \sigma \|u - v\|_2^2. \tag{33}$$

Here, ∇H denotes the gradient, a row vector. If $\sigma > 0$, then H is strongly convex. As shown below in Corollary 4.2, strong convexity of H and convexity of J imply that the objective function in the penalized problem (16) is strongly convex, which ensures the existence of a unique minimizer.

Theorem 4.1. *If* (16) *has minimizers* v^* *and* u^**, then for each* k *we have*

$$\|v^{k+1} - v^*\|_2 \leq \frac{2\alpha}{2\alpha + \sigma} \|v^k - v^*\|_2 \quad \textit{and} \quad \|u^{k+1} - u^*\|_2 \leq \frac{2\alpha}{2\alpha + \sigma} \|u^k - u^*\|_2. \tag{34}$$

Proof. It is well known that the operators $\mathcal{T}$ and $\mathcal{L}$ in (17) are nonexpansive relative to the Euclidean norm. That is, for all u and v, we have

$$\|\mathcal{T}(v) - \mathcal{T}(u)\|_2 \leq \|v - u\|_2 \quad \text{and} \quad \|\mathcal{L}(v) - \mathcal{L}(u)\|_2 \leq \|v - u\|_2.$$

This follows from the first-order optimality conditions characterizing the minimizers in (17). For example, if $v_i = \mathcal{T}(u_i)$ for $i = 1, 2$, then $2\alpha(u_i - v_i)^{\mathsf{T}} \in \partial J(v_i)$, where ∂ denotes the subdifferential. By the convexity of J, it follows that

$$J(v_2) \geq J(v_1) + 2\alpha(u_1 - v_1)^{\mathsf{T}}(v_2 - v_1). \tag{35}$$

Likewise, interchanging v_1 and v_2 gives

$$J(v_1) \geq J(v_2) + 2\alpha(u_2 - v_2)^{\mathsf{T}}(v_1 - v_2). \tag{36}$$

We add (35) and (36) to obtain

$$\|v_2 - v_1\|_2^2 \leq (u_2 - u_1)^{\mathsf{T}}(v_2 - v_1) \leq \|u_2 - u_1\|_2 \|v_2 - v_1\|_2. \tag{37}$$

Hence, $\|v_2 - v_1\|_2 = \|\mathcal{T}(u_2) - \mathcal{T}(u_1)\|_2 \le \|u_2 - u_1\|_2$, which yields the nonexpansive property.

Since v^* and u^* achieve the minimum in (17), we have $v^* = \mathcal{T}(u^*)$. Subtracting this identity from the equation $v^{k+1} = \mathcal{T}(u^k)$ and utilizing the nonexpansive property gives

$$\|v^{k+1} - v^*\|_2 \le \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 \le \|u^k - u^*\|_2. \tag{38}$$

The first-order optimality conditions for u^k and u^* are

$$\nabla H(u^k) - 2\alpha(v^k - u^k)^{\mathsf{T}} = 0,$$
$$\nabla H(u^*) - 2\alpha(v^* - u^*)^{\mathsf{T}} = 0.$$

We subtract the second equation from the first and multiply by $(u^k - u^*)$ to obtain

$$\begin{aligned}(\nabla H(u^k) - \nabla H(u^*))(u^k - u^*) + 2\alpha\|u^k - u^*\|_2^2 &= 2\alpha(v^k - v^*)^{\mathsf{T}}(u^k - u^*) \\ &\le 2\alpha\|v^k - v^*\|_2 \, \|u^k - u^*\|_2.\end{aligned} \tag{39}$$

Utilizing the monotonicity condition (33) on the left-hand side of (39) gives

$$(\sigma + 2\alpha)\|u^k - u^*\|_2^2 \le 2\alpha\|v^k - v^*\|_2 \, \|u^k - u^*\|_2,$$

which yields

$$\|u^k - u^*\|_2 \le \left(\frac{2\alpha}{\sigma + 2\alpha}\right) \|v^k - v^*\|_2. \tag{40}$$

Combining this with (38) gives

$$\|v^{k+1} - v^*\|_2 \le \left(\frac{2\alpha}{\sigma + 2\alpha}\right) \|v^k - v^*\|_2,$$

the first inequality in (34). Combining (40), with k replaced by $k+1$, and the nonexpansive property (38) gives the second inequality in (34). ∎

Corollary 4.2. *If $\sigma > 0$, then the iterates generated by (17) converge linearly to the unique minimizer of (16).*

Proof. We first observe that when $\sigma > 0$, the objective function in (16) is strongly convex. Let $F(u, v) = H(u) + \alpha\|v - u\|_2^2$ be the part of the objective which excludes J. By the convexity inequality (33), we have

$$\begin{aligned}(\nabla F(u_1, v_1) - \nabla F(u_2, v_2)) \begin{bmatrix} \delta u \\ \delta v \end{bmatrix} &= (\nabla H(u_1) - \nabla H(u_2))(u_1 - u_2) + 2\alpha\|\delta u - \delta v\|_2^2 \\ &\ge \sigma\|\delta u\|_2^2 + 2\alpha\|\delta u - \delta v\|_2^2,\end{aligned} \tag{41}$$

where $\delta u = u_1 - u_2$ and $\delta v = v_1 - v_2$. The matrix corresponding to the quadratic in (41) is

$$2\begin{bmatrix} \alpha + \sigma/2 & -\alpha \\ -\alpha & \alpha \end{bmatrix}.$$

Since the eigenvalues of this matrix are strictly positive, F is strongly convex. The objective function in (16) is the sum $J + F$ of a convex function J and a strongly convex function F. Hence, it is strongly convex, and there exists a unique minimizer (u^*, v^*). By Theorem 4.1, the iterates generated by (17) converge to (u^*, v^*) linearly. ■

In the case $\sigma = 0$, Theorem 4.1 yields only

$$\|v^{k+1} - v^*\|_2 \le \|v^k - v^*\|_2 \quad \text{and} \quad \|u^{k+1} - u^*\|_2 \le \|u^k - u^*\|_2, \tag{42}$$

which does not imply convergence. On the other hand, by the theory for the alternating proximal minimization algorithm, we know that the iterates do converge. We now observe that the inequalities in (42) are strict except when convergence is achieved in a finite number of steps. This result is based on the following property.

Lemma 4.3. *If $\mathcal{P} : \mathbb{C}^n \to \mathbb{C}^n$ satisfies*

$$\|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2 \le \langle \mathcal{P}(u) - \mathcal{P}(v), u - v \rangle \tag{43}$$

for all u and $v \in \mathbb{C}^n$, then

$$\|\mathcal{P}(u) - \mathcal{P}(v)\|_2 \le \|u - v\|_2 \tag{44}$$

for all u and $v \in \mathbb{C}^n$ with equality only if $\mathcal{P}(u) - \mathcal{P}(v) = u - v$.

Operators satisfying (43) are called *firmly nonexpansive*. The fact that the proximal maps $\mathcal{T}$ and $\mathcal{L}$ are firmly nonexpansive is implied by (37).

Proof. The inequality (44) is a consequence of the Schwartz inequality applied to (43). Moreover, by (43) we have

$$\begin{aligned} \|(u - v) - (\mathcal{P}(u) - \mathcal{P}(v))\|_2^2 &= \|u - v\|_2^2 - 2\langle \mathcal{P}(u) - \mathcal{P}(v), u - v \rangle + \|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2 \\ &\le \|u - v\|_2^2 - \|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2. \end{aligned} \tag{45}$$

If (44) is an equality, then the right-hand side of (45) vanishes, which implies that the left-hand side vanishes:

$$(u - v) - (\mathcal{P}(u) - \mathcal{P}(v)) = 0. \quad ■$$

Theorem 4.4. *Suppose that u^* and v^* are optimal in (16). If for some k, the iterates of the alternating proximal minimization algorithm (17) satisfy $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$, then $u^j = u^k$ and $v^{j+1} = v^{k+1}$ for all $j > k$. If $\|v^{k+1} - v^*\|_2 = \|v^k - v^*\|_2$ for some k, then $v^j = v^k$ and $u^j = u^k$ for all $j > k$.*

Proof. Suppose that $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$. Since v^* and u^* are optimal in (16), we have

$$(\mathcal{L}\mathcal{T})(u^*) = \mathcal{L}(\mathcal{T}(u^*)) = \mathcal{L}(v^*) = u^*. \tag{46}$$

By (17), it follows that $u^{k+1} = (\mathcal{L}\mathcal{T})(u^k)$. Hence, the equality $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$ coupled with the nonexpansive properties of $\mathcal{L}$ and $\mathcal{T}$ yields

$$\begin{aligned} \|u^k - u^*\|_2 &= \|(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*)\|_2 = \|\mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*))\|_2 \\ &\le \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 \\ &\le \|u^k - u^*\|_2. \end{aligned} \tag{47}$$

Since the right- and left-hand sides of (47) are equal, all the inequalities in (47) are equalities. The equality $\|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 = \|u^k - u^*\|_2$ and Lemma 4.3 imply that

$$\mathcal{T}(u^k) - \mathcal{T}(u^*) = u^k - u^*. \tag{48}$$

The equality $\|\mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*))\|_2 = \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2$ and Lemma 4.3 imply that

$$(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*) = \mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*)) = \mathcal{T}(u^k) - \mathcal{T}(u^*). \tag{49}$$

Together, (48) and (49) yield

$$(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*) = u^k - u^*. \tag{50}$$

We combine (46) and (50) to obtain

$$u^{k+1} = (\mathcal{L}\mathcal{T})(u^k) = u^k.$$

Hence, u^k is a fixed point of $(\mathcal{L}\mathcal{T})$ and $u^j = u^k$ for all $j > k$. Since $v^{j+1} = \mathcal{T}(u^j)$, we conclude that $v^{j+1} = v^{k+1}$ for all $j > k$. The equality $\|v^{k+1} - v^*\|_2 = \|v^k - v^*\|_2$ is treated in the same way except that $\mathcal{L}$ and $\mathcal{T}$ are interchanged. ∎

By the convergence theory for the alternating proximal minimization algorithm, we know that the iterates converge to a solution (u^*, v^*) of (16), provided that a solution exists. Theorem 4.4 implies that

$$\|u^{k+1} - u^*\|_2 / \|u^k - u^*\|_2 < 1$$

except when $u^k = u^*$. Likewise

$$\|v^{k+1} - v^*\|_2 / \|v^k - v^*\|_2 < 1$$

except when $v^k = v^*$. This implies at least sublinear convergence of the alternating proximal minimization algorithm (17).

For any fixed α, the solution of (16) generates an approximation to a solution of (4). Let α_k, $k \geq 0$, denote an increasing sequence of values for the penalty parameter tending to infinity, and let (U^k, V^k) denote associated solutions of (16), assuming they exist. By the theory describing the convergence of the penalty scheme (see [31, Thm. 17.1]), convergent subsequences of the iterates approach a solution of (4). We now show in the context (5) of image reconstruction that the iterates (U^k, V^k) are bounded.

Theorem 4.5. *Suppose that J and H are given by (5). If $\lambda > 0$ and $\mathcal{N}(D) \cap \mathcal{N}(A) = 0$, where $\mathcal{N}$ denotes null space, then for each $\alpha_0 > 0$, there exists a compact set K which contains the solutions of (16) for all $\alpha \geq \alpha_0$. Moreover, as α tends to infinity, any convergent subsequence of the iterates approaches a solution of either (4) or the equivalent problem (15).*

Proof. In the special case (5), $J(0) = 0$ and $H(0) = \lambda\|f\|_2^2$. Let $\rho = \lambda\|f\|_2^2$ be the value of the objective function in (16) corresponding to $u = v = 0$. For any choice of α, the optimal objective function value in (16) is bounded by ρ. Hence, for any choice of α, when minimizing the objective function in (16), we should restrict our attention to those u and v satisfying

$$J(v) + H(u) + \alpha\|v - u\|_2^2 \leq \rho. \tag{51}$$

Since $J(v) = \|v\|_{TV} \geq 0$ and $H(u) = \|Au - f\|_2^2 \geq 0$, it follows from (51) that

$$\|v - u\|_2 \leq \sqrt{\rho/\alpha}, \tag{52}$$

$$\|v\|_{TV} \leq \rho, \tag{53}$$

$$\|Au - f\|_2 \leq \sqrt{\rho/\lambda}. \tag{54}$$

Decompose $u = u_n + u_p$ where $u_n \in \mathcal{N}(A)$ and u_p is orthogonal to $\mathcal{N}(A)$. By (8), (52), and (53), we have

$$\begin{aligned} \rho \geq \|v\|_{TV} &= \sum_{i=1}^{N} \|D_i v\|_2 \geq \|Dv\|_2 \geq \|Du\|_2 - \|D(v-u)\|_2 \\ &\geq \|Du_n\|_2 - \|Du_p\|_2 - \|D\|_2 \|v-u\|_2 \\ &\geq \|Du_n\|_2 - \|Du_p\|_2 - \|D\|_2 \sqrt{\rho/\alpha}. \end{aligned} \tag{55}$$

Since $\mathcal{N}(D) \cap \mathcal{N}(A) = 0$, there exists $\gamma_1 > 0$ such that

$$\|Du\|_2 \geq \gamma_1 \|u\|_2 \quad \forall\ u \in \mathcal{N}(A).$$

Hence, by (55),

$$\|u_n\|_2 \leq \left(\rho + \|Du_p\|_2 + \|D\|_2 \sqrt{\rho/\alpha}\right) / \gamma_1. \tag{56}$$

Similarly, there exists $\gamma_2 > 0$ such that

$$\|Au_p\|_2 \geq \gamma_2 \|u_p\|_2.$$

Hence, by (54), we have

$$\gamma_2 \|u_p\|_2 \leq \|Au\|_2 \leq \|f\|_2 + \|Au - f\|_2 \leq \|f\|_2 + \sqrt{\rho/\lambda}. \tag{57}$$

Combine (56) and (57) to deduce that $u = u_n + u_p$ lies in a compact set. By (52), we have

$$\|v\|_2 \leq \|u\|_2 + \sqrt{\rho/\alpha},$$

which yields a bound for $\|v\|_2$. As α increases, the level set of (16) corresponding to the objective function value ρ can only shrink. Hence, this level set is bounded for any $\alpha \geq \alpha_0$. Let α_k, $k = 0, 1, \ldots$, denote an increasing sequence of values for the penalty tending to infinity, and let (U^k, V^k) denote associated solutions of (16). By [31, Thm. 17.1], every convergent subsequence of the minimizers (U^k, V^k) approaches a solution of (15). ∎

Remark. If H is strongly convex, then (4) has a unique solution; hence, any sequence of solutions to (16) approaches the unique solution of (4) as α tends to infinity.

5. Numerical experiments. In this section we evaluate the performance of algorithms using three PPI reconstructions. In the first two tests we compare the performance of AM (18), ADMM (25), and APD (28) to that of the Bregman operator splitting (BOS) in [47] and a slightly modified version (SBB) of the algorithm proposed in [43] on pseudopartial k-space data. In the last test, we apply these algorithms to a real clinical data set.

5.1. Data acquisition and experimental setup. In our tests, all k-space data was fully acquired with an 8-channel head coil. By full acquisition we mean that each receiver coil obtains the complete k-space data and hence a high resolution image. The first data set, denoted by data1, is a collection of sagittal Cartesian brain images acquired on a 3T GE system (GE Healthcare, Waukesha, Wisconsin). The data acquisition parameters were field of view (FOV) 220mm^2, size 512×512×8, repetition time (TR) 3060ms, echo time (TE) 126ms, slice thickness 5mm, and flip angle 90°. The phase encoding direction was anterior-posterior. To make this data set less similar to the next data set, we reduced the image size to 256. The second data set, data2, is a Cartesian brain image acquired on a 3.0T Philips scanner (Philips, Best, Netherlands) using a T2-weighted turbo spin echo (T2 TSE) sequence. The acquisition parameters were FOV 205mm^2, matrix 512 × 500 × 8, TR 3000ms, and TE 85ms. The echo train length was 20. The last data set is a radial brain data set, data3. It was acquired on a 1.5T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were FOV 220mm^2, matrix 256 × 512 × 8, slice thickness 5mm, TR 53.5ms, TE 3.4ms, and flip angle 75°.

All data sets are normalized such that the intensities of reference images have range $[0, 1]$. In addition, the sensitivity maps are also normalized into the same range. The parameter setting and further discussion are based on the normalized data. In all of our experiments, a ground truth or reference image is set to the pointwise root of summed squares of the images obtained by full k-space data from all channels. Namely, the reference image $\bar{u}$ was given by the formula

$$\bar{u}_i = \left(\sum_{j=1}^{K} |(\bar{u}_j)_i|^2 \right)^{1/2}. \tag{58}$$

Here $(\bar{u}_j)_i$ is the i-component of $\bar{u}_j$, the Fourier transform of the full k-space data acquired on the jth channel.

In the first two experiments, we simulate the sensitivity maps S_j using the *central* 32×32 k-space data and generate the pseudofull k-space data by $\mathcal{F}S_j\bar{u}$, where $\bar{u}$ is obtained by full k-space data as in (58). The sensitivity map for data2 is shown in Figure 2. We add a complex-valued Gaussian noise (same level for both real and imaginary parts) with standard deviation 0.01 in magnitude to the pseudofull data, and we artificially downsample the pseudofull data using the mask shown in Figure 1(a) for data1 and the mask shown in Figure 1(b) for data2, with reduction factor (RF)—the reciprocal of undersampling ratio—3 and 4, respectively. For the last dataset, we directly use 84 out of 256 radial lines (hence the RF is 3) as in the practical radial data undersampling strategy. Then we applied the generalized autocalibration partially parallel aquisition (GRAPPA) operator gridding (GROG) technique [38] to shift the non-Cartesian radial data onto Cartesian grids such that fast Fourier transforms can be directly used in computation.

In our experiments, all algorithms are implemented in MATLAB, Version R2009b. All the experiments are performed on a Lenovo laptop with an Intel Dual Core 2 Duo 2.53 GHz processor and a Windows operating system.

5.2. Comparison algorithms. Many of the algorithms in section 2 are not very effective for PPI imaging due to the complicated structure of A. For comparison, we chose the recently

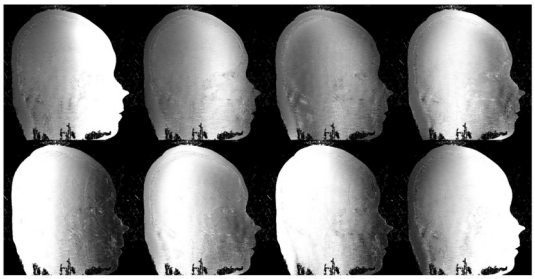


Figure 2. *The sensitivity maps for the eight channels of data*2.

proposed Bregman operator splitting (BOS) scheme from [47] and a split Bregman scheme, SBB [43], utilizing a BB step size. BOS and SBB, to the best of the authors' knowledge, currently are among the most efficient methods for solving (7) with arbitrary matrix A. The BOS scheme of [47] and the SBB scheme of [43] are iterative methods for solving (10). They are based on the ADMM applied to the Lagrangian (12), and they correspond to the following iteration:

$$\left.\begin{aligned} w_i^{k+1} &= \max\{\|D_i^k u + b_i^k\|_2 - 1/2\rho, 0\}(D_i^k u + b_i^k)/\|D_i^k u + b_i^k\|_2 \quad \forall\ i \\ u^{k+1} &= \left(\rho D^\mathsf{T} D + \lambda\delta_k I\right)^{-1}\left(\rho D^\mathsf{T}(w^{k+1} - b^k) + \lambda\delta_k u^k - \lambda A^\mathsf{T}(Au^k - f)\right) \\ b_i^{k+1} &= b_i^k - (w_i^{k+1} - D_i u^{k+1}) \quad \forall\ i \end{aligned}\right\}. \tag{59}$$

In the BOS scheme of [47], δ_k is a constant δ, and convergence to a solution of (7) holds when $\delta \in (\|A^\mathsf{T}A\|_2, \infty)$. In the SBB scheme, δ_k is given by the BB formula [5], which reduces to

$$\delta_k = \frac{\|A(u^k - u^{k-1})\|_2^2}{\|u^k - u^{k-1}\|_2^2}.$$

In either BOS or SBB, the w-subproblem represents a two-dimensional shrinkage. In the u-subproblem, $D^\mathsf{T}D$ can be diagonalized by a Fourier transform, provided that the image has periodic boundary condition; hence, $\rho D^\mathsf{T}D + \lambda\delta_k I$ can be inverted easily. Consequently, the main computational cost corresponds to multiplication by A^T and by A or, equivalently, to the evaluation of $2K$ Fourier transforms.

5.3. Experimental results. In all experiments, we set $\lambda = 0.5 \times 10^3$, for which the reconstructions of the test data by (7) have the optimal signal-to-noise ratio. Also, we set $\rho = 10$ for the BOS and SBB schemes, whereas moderate changes of ρ in $[10^0, 10^2]$ do not have much influence on the results. For BOS where δ_k is constant, we found in numerical experiments that the fastest asymptotic convergence was achieved by taking $\delta_k = 1$.

We set $\alpha = 0.1 \times \lambda = 50$ for AM, ADMM, and APD. For α in the range $[10^0, 10^2]$, the AM scheme has stable performance when the data is normalized between $[0, 1]$. The iterations in the (TV) subproblems for AM, ADMM, and APD as well as the (LS) subproblems of AM and ADMM are terminated when the relative change of the inner iterates is less than $\epsilon_{\text{inn}} = 10^{-2}$. Dynamically adjusting ϵ_{inn} as the outer iterates approach the solution can lead to improved

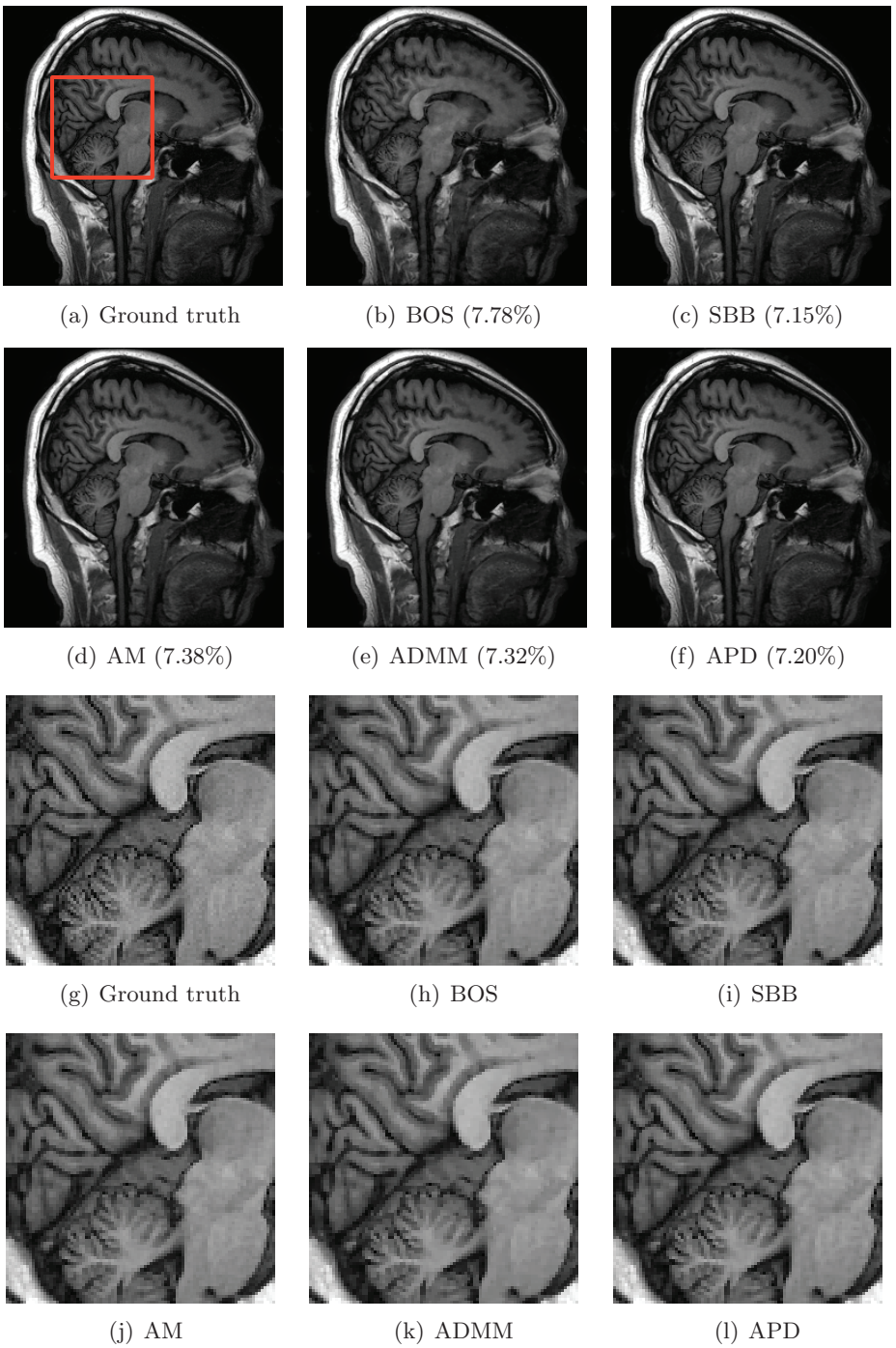
efficiency; however, in our experiments, a constant ϵ_{inn} already leads to better performance than most recently developed algorithms such as BOS.

For all algorithms tested in our experiments, we set the initial guess u^0 to zero, and we terminate the computation when the relative change $\|u^k - u^{k-1}\|_2/\|u^k\|_2$ of the (outer) iterate reaches the prescribed stopping criterion $\epsilon = 10^{-4}$. A tighter stopping criterion can lead to slightly improved accuracy for all algorithms at the expense of a much longer computational time (as shown in Table 1), so it is usually not appropriate in practice. For many real applications where A has extensive computational complexity, it is better and more practical to stop at a suboptimal solution with satisfactory quality obtained in a reasonably short time period. Through our experiments, we found that setting the relative change tolerance ϵ to 10^{-4} consistently leads to satisfactory results. Furthermore, it does not require the knowledge of the ground truth or reference image.

The reconstructed images for data1 and data2 are shown in Figures 3 and 5, respectively. The relative errors in the reconstructed image u, $\|u - \bar{u}\|_2/\|\bar{u}\|_2$, are indicated in parentheses in the figures. In Figures 3 and 5, we zoom into the square shown in the boxes of Figures 3(a) and 5(a). It is seen that all methods adequately recovered the image in the sense that most details and fine structures were accurately recovered from a small set of data samples. Furthermore, SBB, APD, AM, and ADMM appear to have higher accuracy than BOS for this stopping criterion. This can be seen in Figures 4 and 6, specifically, in terms of both the residue images $|u - \bar{u}|$ and the relative errors, which are given in parentheses.

To examine the efficiency of AM, ADMM, and APD compared to BOS and SBB, we plotted the relative error as a function of the CPU time for the two data sets. In Figures 7(a) and 8(a), we see that all algorithms converged faster for the smaller image, data1, than for data2. For both data1 and data2, SBB and APD appear to be the fastest, closely followed by AM and ADMM. AM and ADMM have almost identical performance until the prescribed criterion is met, although the latter proves to have better asymptotic behavior as it converges to an exact solution of (10) in theory. BOS has the slowest convergence speed for these data sets. This can also be seen from Figures 7(b) and 8(b). Among these five algorithms, the objective values of SBB and APD have the fastest decay. The decay rates of AM and ADMM are comparable to those of SBB and APD or slightly slower. The curve of BOS is far above, implying that it has lower efficiency, as its energy decays much slower than that of the others.

To show how accurately they can solve the minimization problem (7), we test all algorithms on data1 and data2 with a much tighter stopping criterion $\epsilon = 10^{-6}$ so that they can run until further improvements on accuracy are negligible. The CPU time in seconds (CPU), relative error to the reference image (RelErr), and final objective function value (Obj) are shown in Table 1. From Table 1, we can see that SBB and APD are very efficient in the sense that they reach lower reconstruction errors in a shorter time compared to others. AM and ADMM are comparable to or slightly less effective than SBB and APD in terms of their computational cost and relative errors, but the difference is not significant. BOS appears to be the least efficient among the tested algorithms as it takes a long time but ends up with a large error. Table 2 also suggests that a moderate stopping criterion (e.g., 10^{-4}) is more suitable in practice for the tested algorithms: A strict termination criterion 10^{-6} yields only slightly improved image qualities (in terms of reconstruction errors) in comparison to that of 10^{-4}, as shown in Figures 7(a) and 8(a), at the expense of a much longer CPU time (e.g., from



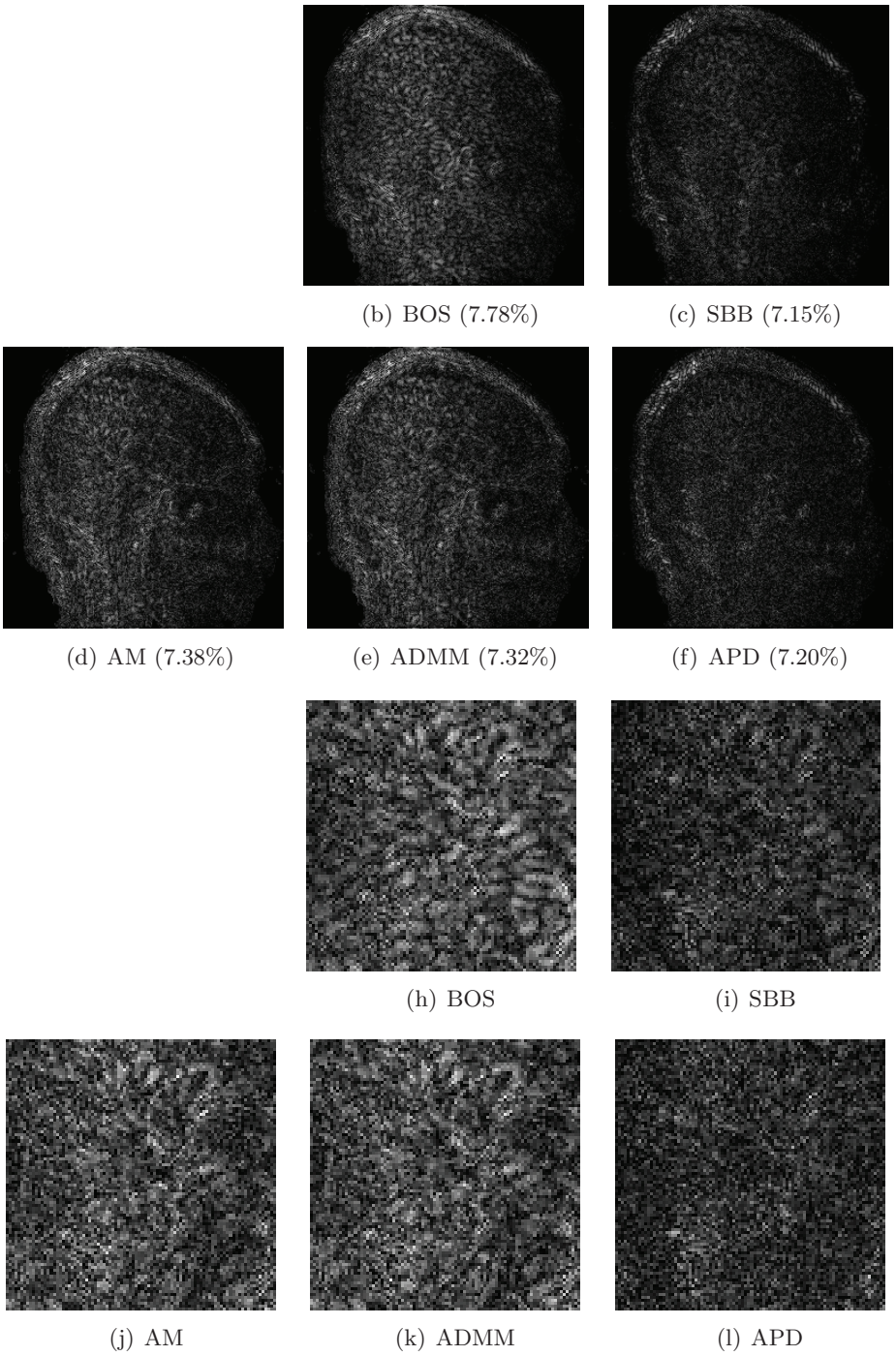
(a) Ground truth (b) BOS (7.78%) (c) SBB (7.15%)

(d) AM (7.38%) (e) ADMM (7.32%) (f) APD (7.20%)

(g) Ground truth (h) BOS (i) SBB

(j) AM (k) ADMM (l) APD

Figure 3. *Reconstructed images of data*1. (g)–(l) *zoom in on the box in* (a)–(f)*, respectively. Corresponding relative errors are indicated in parentheses.*



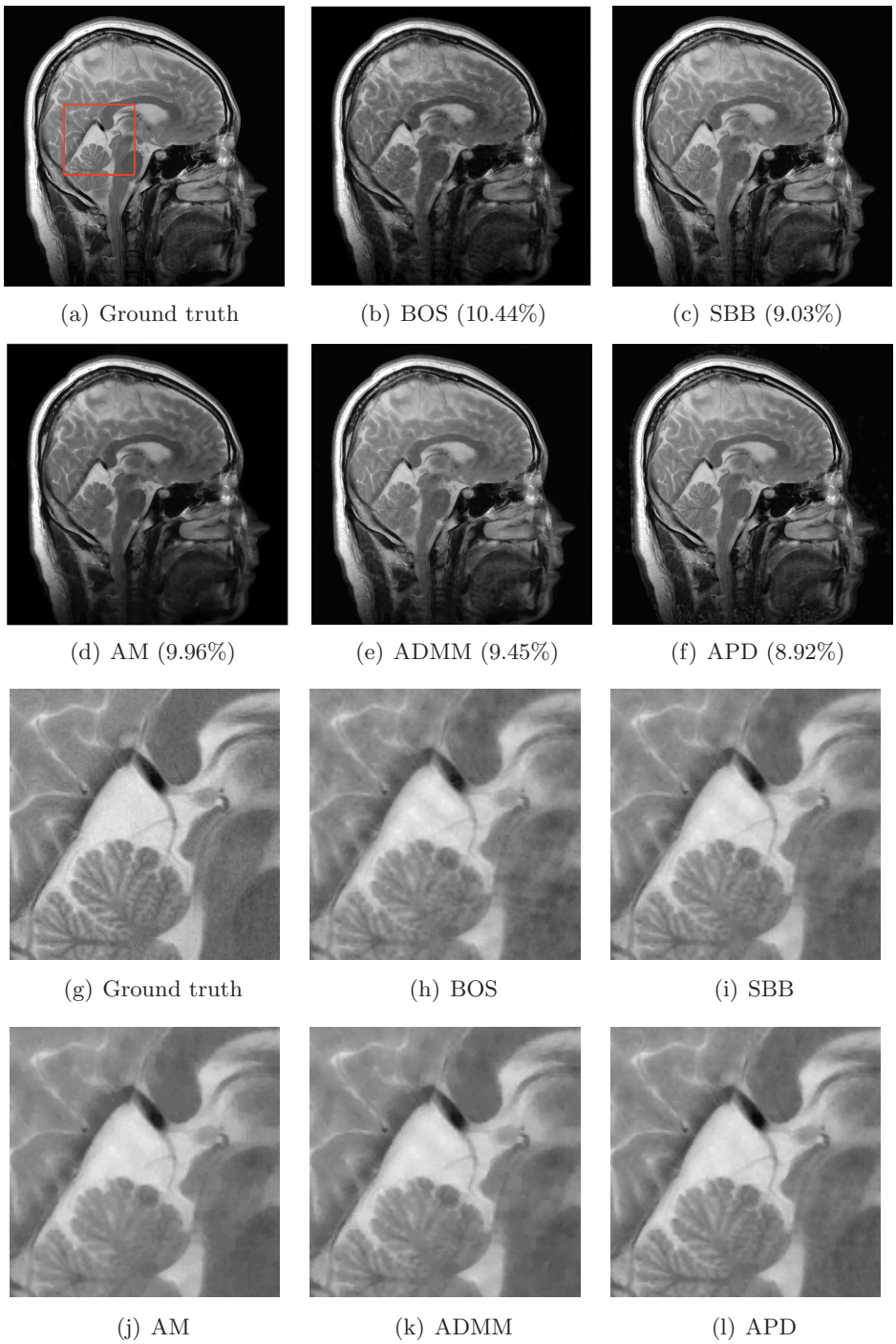
(b) BOS (7.78%) (c) SBB (7.15%)

(d) AM (7.38%) (e) ADMM (7.32%) (f) APD (7.20%)

(h) BOS (i) SBB

(j) AM (k) ADMM (l) APD

Figure 4. *Absolute errors of reconstructions u (shown in Figure 3) to the reference image $\bar{u}$, i.e., $|u-\bar{u}|$, of data*1*. All images are shown with the same brightening scale.* (h)–(l) *zoom in on the box (shown in Figure* 3(a)*) in* (b)–(f)*, respectively. Corresponding relative errors are indicated in parentheses.*



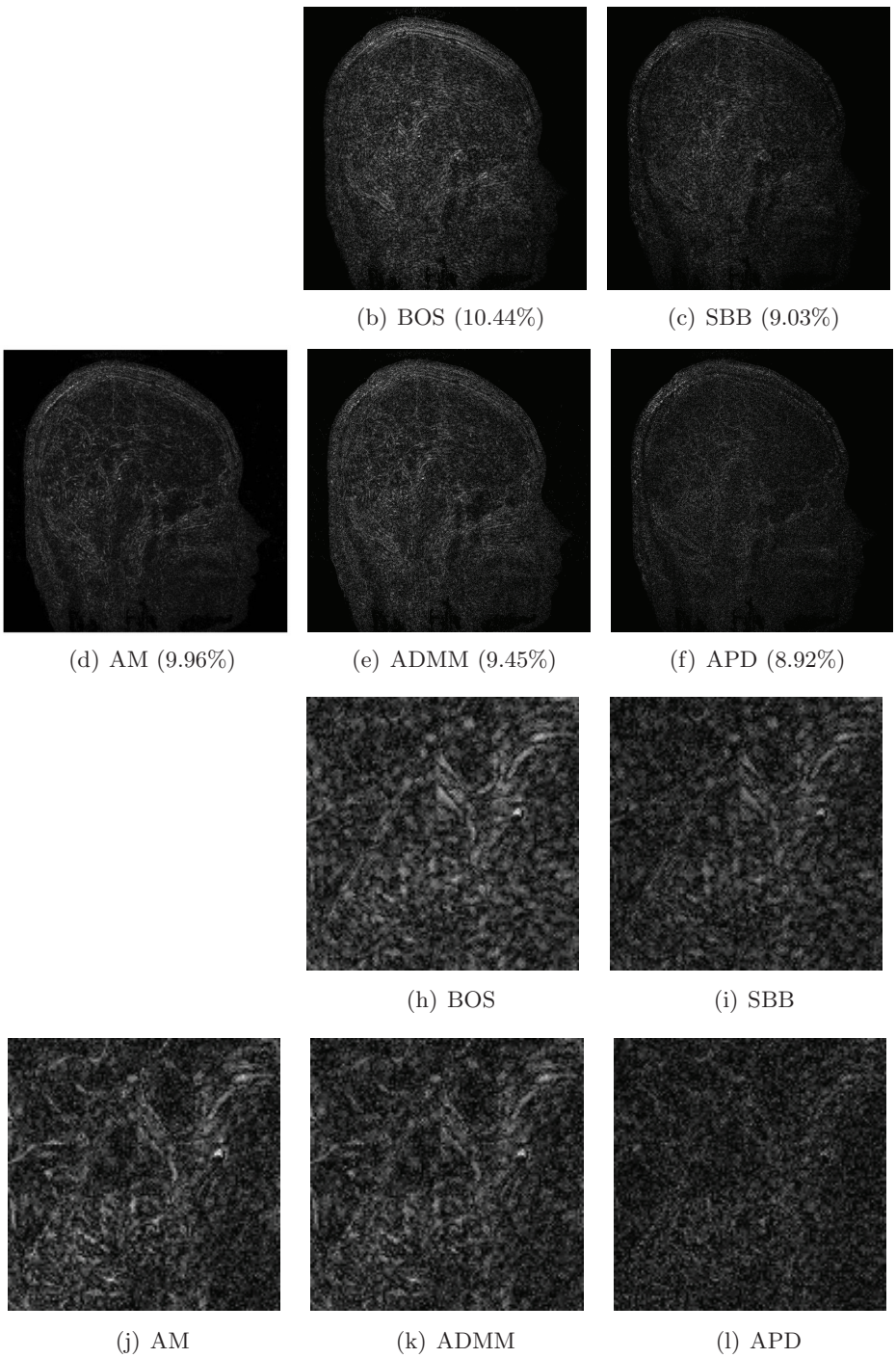
(a) Ground truth (b) BOS (10.44%) (c) SBB (9.03%)

(d) AM (9.96%) (e) ADMM (9.45%) (f) APD (8.92%)

(g) Ground truth (h) BOS (i) SBB

(j) AM (k) ADMM (l) APD

Figure 5. *Reconstructed images of data*2. (g)–(l) *zoom in on the box in* (a)–(f)*, respectively. Corresponding relative errors are indicated in parentheses.*



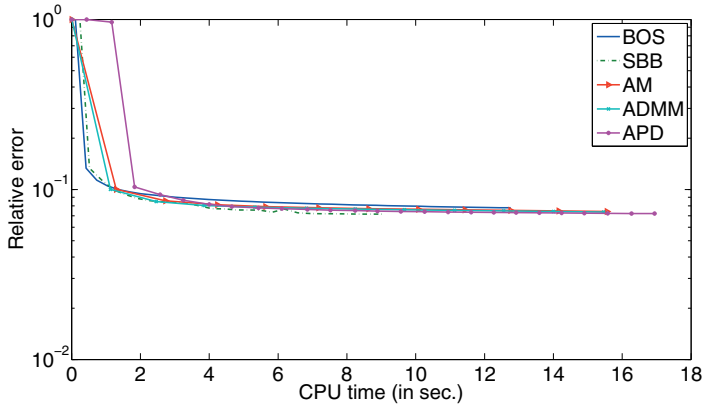
(b) BOS (10.44%) (c) SBB (9.03%)

(d) AM (9.96%) (e) ADMM (9.45%) (f) APD (8.92%)

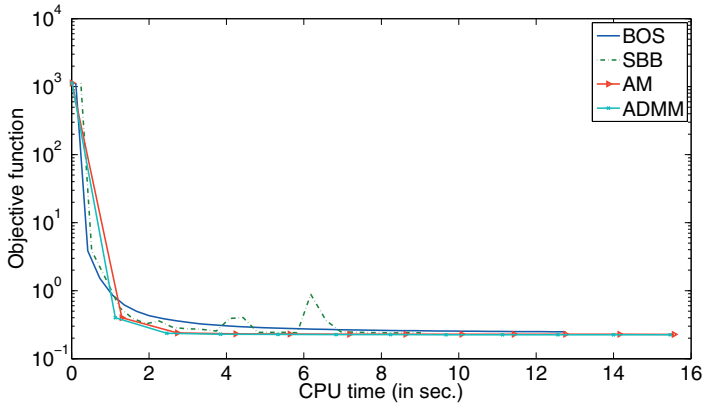
(h) BOS (i) SBB

(j) AM (k) ADMM (l) APD

Figure 6. *Absolute errors of reconstructions u (shown in Figure 5) to the reference image $\bar{u}$, i.e., $|u-\bar{u}|$, of data2. All images are shown with the same brightening scale.* (h)–(l) *zoom in on the box (shown in Figure 5(a)) in* (b)–(f)*, respectively. Corresponding relative errors are indicated in parentheses.*



(a) Relative errors versus CPU time of data1.



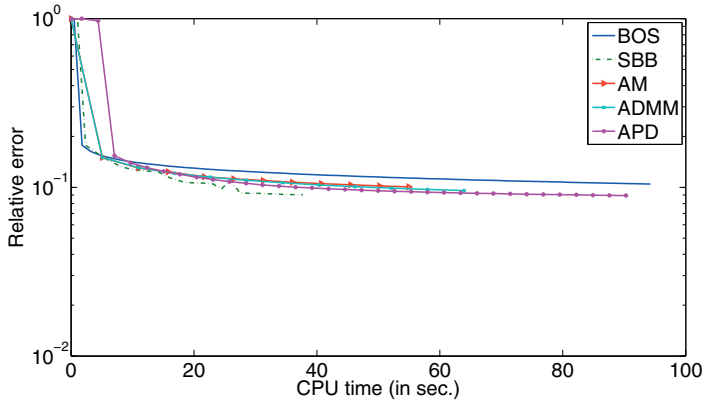
(b) Objective function versus CPU time of data1.

Figure 7. *Comparison of BOS, SBB, APD, AM, and ADMM on data*1.

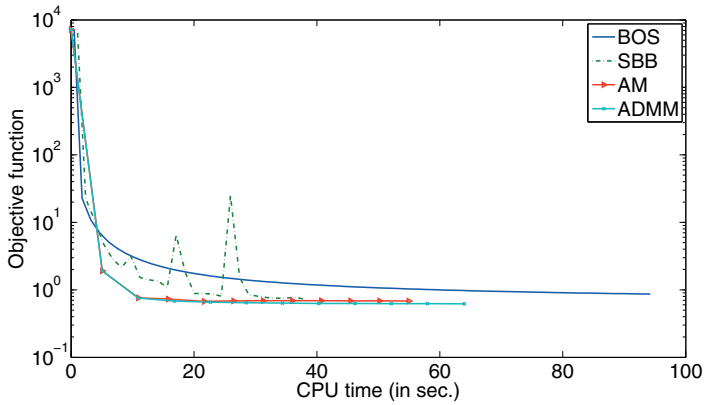
about 15 seconds to more than 60 seconds or even longer for data1). In addition to efficiency, we also summarize the characteristics of the tested algorithms in terms of their generality and convergence properties in Table 2.

The reason that BOS is slower than the other schemes is due to the total number of iterations that are required. Even though each iteration was fast, there were too many iterations to compete with the other algorithms. APD exploits the special structure of A in PPI to achieve fast convergence by solving the (LS) subproblem using FFTs. The relatively fast convergence of the BB method for the LS problem helped the performance of AM. AM, ADMM, and APD all benefit from the speed of the PDHG solver for the (TV) subproblems.

The performance of TV-based image reconstruction algorithms on real clinical data is a major concern in PPI. In [27], Lustig, Donoho, and Pauly demonstrated that minimization (7) has great potential to accurately recover images from undersampled k-space data in single channel MR imaging. This strategy also works well in parallel imaging. In this paper we test several recent algorithms for (7) on the clinical data set data3. The reconstruction results are shown in Figure 9. Due to inevitable biases in sensitivity estimation and noises



(a) Relative errors versus CPU time of data2.



(b) Objective function versus CPU time of data2.

Figure 8. *Comparison of BOS, SBB, APD, AM and ADMM on data2.*

Table 1

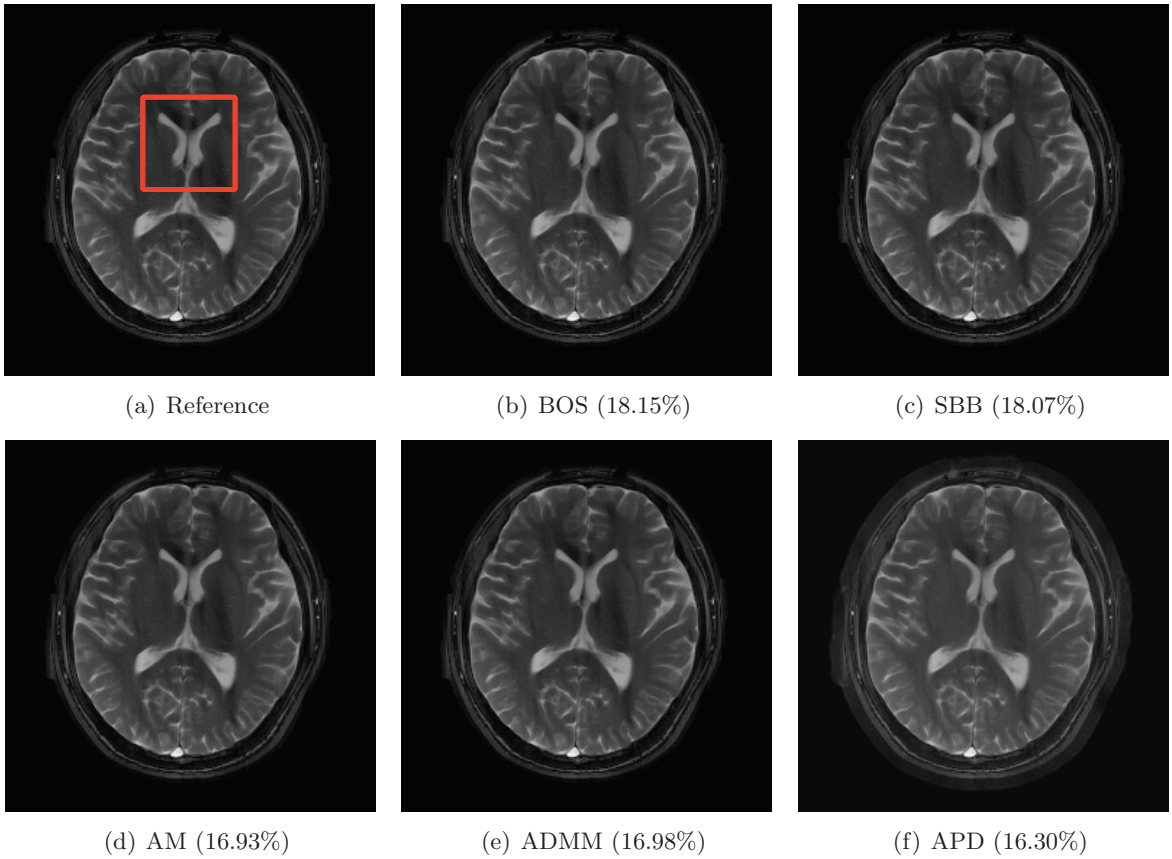
Results of the tested algorithms on data1 and data2.

		BOS	SBB	APD	AM	ADMM
data1	CPU (s)	193	67.8	59.0	65.7	66.5
	RelErr (%)	7.15	6.71	6.80	7.03	6.86
	Obj	0.2687	0.2241	0.2111	0.2172	0.2127
data2	CPU (s)	884	203	187	240	294
	RelErr (%)	10.01	8.68	8.98	9.27	9.02
	Obj	0.7854	0.6430	0.6139	0.6362	0.6146

Table 2

Properties of the tested algorithms.

	BOS	SBB	APD	AM/ADMM
Works for general A?	yes	yes	no	yes
Convergence established?	yes	no	yes	yes



(a) Reference (b) BOS (18.15%) (c) SBB (18.07%)

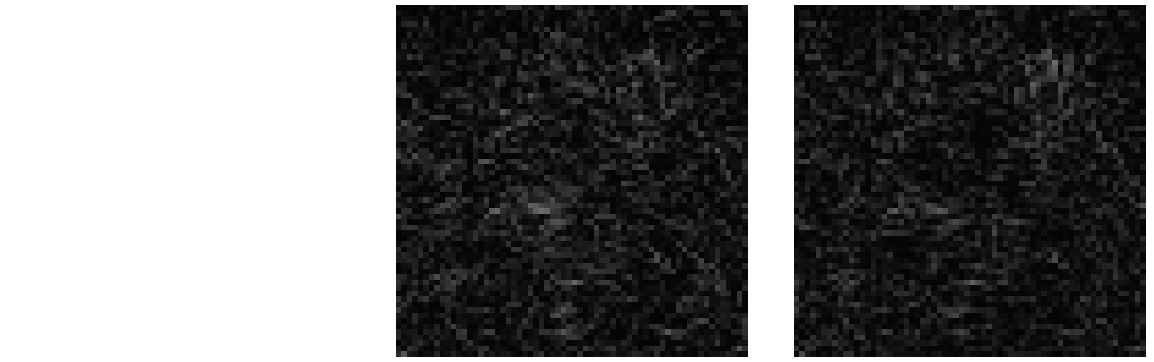
(d) AM (16.93%) (e) ADMM (16.98%) (f) APD (16.30%)

Figure 9. *Reconstructed images of clinical data3. Corresponding relative errors are indicated in parentheses.*

in data acquisition, the reconstruction errors appear larger than those in the first two tests with simulated data given the same undersampling ratio. This can be seen from the error maps shown in Figure 10. Nevertheless, model (7) still recovers good images from highly undersampled data, as shown in Figure 9. Interestingly, the AM/ADMM algorithms have superior performance in terms of objective value, as shown in Figure 11, which implies that AM/ADMM have optimal performance from the optimization point of view. It is worth noting that the APD algorithm requires extra computations to evaluate objective values, and hence its curves are not included in Figures 7(b), 8(b), and 11(b). However, a stand-alone track of the APD algorithm iterates shows that the objective values of APD are as low as those of AM/ADMM during the iterations.

Remark. In some PPI applications, the images are sparse under an orthogonal wavelet transform Ψ. In this case, we add $\|\Psi u\|_1$ to the energy function (7). To minimize the energy function, we introduce the splitting $z = \Psi u$ and again apply quadratic penalty and multiplier methods. All the algorithms (BOS, AM, ADMM, APD, and SBB) remain valid with small modifications to account for the additional term in the energy function.

6. Conclusions. Two fast algorithms for TV-based image reconstruction were introduced. The first method, AM, employs variable splitting, a quadratic penalty, and an alternating

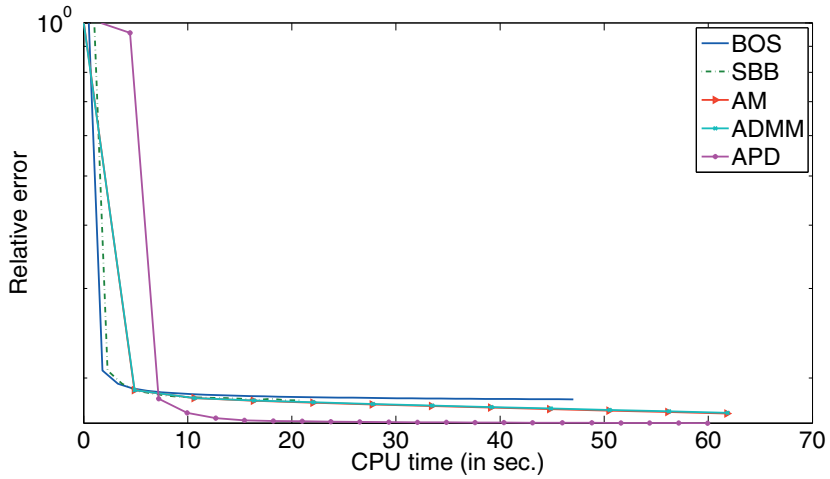


(b) BOS (18.15%) (c) SBB (18.07%)

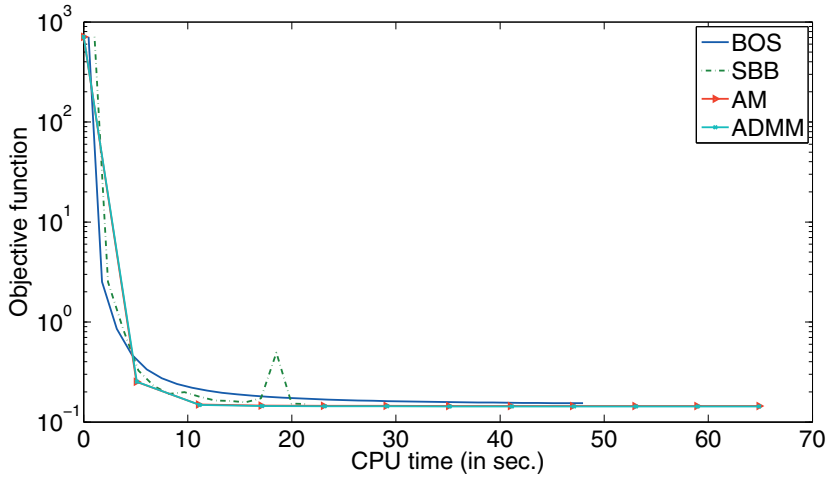
(d) AM (16.93%) (e) ADMM (16.98%) (f) APD (16.30%)

Figure 10. *Absolute errors of zoomed-in boxes (shown in Figure 9(a)) of reconstructions u (shown in Figure 9) to the reference image $\bar{u}$, i.e., $|u - \bar{u}|$, of data3. All images are shown with the same brightening scale. Corresponding relative errors are indicated in parentheses.*

proximal minimization algorithm. Linear convergence was established when the smooth part of the objective function was strongly convex, while the convergence was sublinear under a weaker convexity assumption. A modification, ADMM, based on alternating direction method of multipliers was also presented and showed performance similar to that of AM. An implementation based on a primal-dual hybrid gradient (PDHG) scheme for the TV problem and a Barzilai–Borwein (BB) method for the linear inversion is proposed. The second algorithm, APD, is based on an augmented Lagrangian and a primal-dual algorithm; it exploits the special structure of the PPI reconstruction problem by decomposing it into one subproblem involving Fourier transforms and another subproblem that can be treated by the PDHG scheme. The numerical performance of these algorithms was compared to that of a Bregman operator splitting (BOS) [47] and a modified algorithm, SBB [43], where the constant BOS step size is replaced by a variable step size based on the BB algorithm [5]. It was found that for the same stopping criteria, SBB, AM/ADMM, and APD had comparable performance and produced the highest quality images. All of them are significantly faster than BOS. Moreover, AM/ADMM and APD proposed in this paper have established convergence analysis and



(a) Relative errors versus CPU time on data3.



(b) Objective function versus CPU time on data3.

Figure 11. *Comparison of BOS, SBB, AM, ADMM, and APD on data3.*

exhibit monotone decay in terms of reconstruction error and objective function during the computation.

REFERENCES

[1] F. Acker and M.-A. Prestel, *Convergence d'un schéma de minimisation altérnée*, Ann. Fac. Sci. Toulouse Math. (5), 2 (1980), pp. 1–9.

[2] 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, 11 (1959), pp. 1–16.

[3] A. Arunachalam, A. Samsonov, and W. Block, *Self-calibrated GRAPPA method for 2D and 3D radial data*, Magn. Reson. Med., 57 (2007), pp. 931–938.

[4] H. Attouch, P. Redont, and A. Soubeyran, *A new class of alternating proximal minimization algorithms with costs-to-move*, SIAM J. Optim., 18 (2007), pp. 1061–1081.

[5] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.
[6] H. H. Bauschke, P. L. Combettes, and D. Noll, *Joint minimization with alternating Bregman proximity operators*, Pac. J. Optim., 2 (2006), pp. 401–424.
[7] D. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice–Hall, Englewood Cliffs, NJ, 1989.
[8] K. Block, M. Uecker, and J. Frahm, *Undersampled radial MRI with multiple coils: Iterative image reconstruction using a total variation constraint*, Magn. Reson. Med., 57 (2007), pp. 1086–1098.
[9] E. J. Candes, J. K. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
[10] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[11] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.
[12] Y. H. Dai, W. W. Hager, K. Schittkowski, and H. Zhang, *The cyclic Barzilai-Borwein method for unconstrained optimization*, IMA J. Numer. Anal., 26 (2006), pp. 604–627.
[13] Y. H. Dai and Y. Yuan, *Alternate minimization gradient method*, IMA J. Numer. Anal., 23 (2003), pp. 377–393.
[14] J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318.
[15] H. Eggers and P. Boesiger, *Gridding- and convolution-based iterative reconstruction with variable k-space resolution for sensitivity-encoded non-Cartesian imaging*, in Proceedings of the International Society for Magnetic Resonance in Medicine, 2003, p. 2346.
[16] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046.
[17] A. Friedlander, J. M. Martínez, B. Molina, and M. Raydan, *Gradient method with retards and generalizations*, SIAM J. Numer. Anal., 36 (1999), pp. 275–289.
[18] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange Methods: Applications to the Solution of Boundary-Valued Problems, M. Fortin and R. Glowinski, eds., North–Holland, Amsterdam, 1983, pp. 299–331.
[19] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.
[20] R. Glowinski and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un, et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, RAIRO Anal. Numér., 9 (1975), pp. 41–76.
[21] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[22] M. Griswold, P. Jakob, R. Heidemann, N. Mathias, V. Jellus, J. Wang, B. Kiefer, and A. Haase, *Generalized autocalibrating partially parallel acquisitions (GRAPPA)*, Magn. Reson. Med., 47 (2002), pp. 1202–1210.
[23] Y. Huang, M. K. Ng, and Y.-W. Wen, *A fast total variation minimization method for image restoration*, Multiscale Model. Simul., 7 (2008), pp. 774–795.
[24] D. Krishnan, P. Lin, and X. Tai, *An efficient operator splitting method for noise removal in images*, Commun. Comput. Phys., 1 (2006), pp. 847–858.
[25] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[26] C. Liu, R. Bammer, and M. Moseley, *Parallel imaging reconstruction for arbitrary trajectories using k-space sparse matrices (kSPA)*, Magn. Reson. Med., 58 (2007), pp. 1071–1081.
[27] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magn. Reson. Med., 58 (2007), pp. 1182–1195.
[28] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, *An efficient algorithm for compressed MR imaging using total variation and wavelets*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[29] Y. Nesterov, *Gradient Methods for Minimizing Composite Objective Function*, CORE Discussion Paper 2007/76, Center for Operations Research and Econometrics (CORE), Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2007.
[30] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $\mathcal{O}(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547 (in Russian).
[31] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
[32] K. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger, *Advances in sensitivity encoding with arbitrary k-space trajectories*, Magn. Reson. Med., 46 (2001), pp. 638–651.
[33] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, *SENSE: Sensitivity encoding for fast MRI*, Magn. Reson. Med., 42 (1999), pp. 952–962.
[34] Y. Qian, Z. Zhang, V. Stenger, and Y. Wang, *Self-calibrated spiral SENSE*, Magn. Reson. Med., 52 (2004), pp. 688–692.
[35] 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., 54 (2005), pp. 1040–1045.
[36] L. Rudin, S. Osher, and E. Fatemi, *Non-linear total variation noise removal algorithm*, Phys. D, 60 (1992), pp. 259–268.
[37] A. Samsonov and C. Johnson, *Non-Cartesian POCSENSE*, in Proceedings of the International Society for Magnetic Resonance in Medicine, 2004, p. 2648.
[38] N. Seiberlich, F. Breuer, M. Blaimer, K. Barkauskas, and P. J. M. Griswold, *Non-Cartesian data reconstruction using GRAPPA operator gridding (GROG)*, Magn. Reson. Med., 58 (2007), pp. 1257–1265.
[39] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.
[40] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.
[41] Y.-W. Wen, M. K. Ng, and Y.-M. Huang, *Efficient total variation minimization methods for color image restoration*, IEEE Trans. Image Process., 17 (2008), pp. 2081–2088.
[42] J. Yang, W. Yin, Y. Zhang, and Y. Wang, *A fast algorithm for edge-preserving variational multichannel image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 569–592.
[43] X. Ye, Y. Chen, and F. Huang, *Computational acceleration for MR image reconstruction in partially parallel imaging*, IEEE Trans. Med. Imag., 30 (2010), pp. 1055–1063.
[44] X. Ye, Y. Chen, W. Lin, and F. Huang, *Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories*, IEEE Trans. Med. Imag., 30 (2010), pp. 575–585.
[45] E. Yeh, M. Stuber, C. McKenzie, R. M. Botnar, T. Leiner, M. Ohliger, A. Grant, J. Willig-Onwuachi, and D. Sodickson, *Inherently self-calibrating non-Cartesian parallel imaging*, Magn. Reson. Med., 54 (2005), pp. 1–8.
[46] L. Ying, B. Liu, M. Steckner, G. Wu, M. Wu, and S.-J. Li, *A statistical approach to SENSE regularization with arbitrary k-space trajectories*, Magn. Reson. Med., 60 (2008), pp. 414–421.
[47] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.
[48] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.
[49] M. Zhu and T. Chan, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, Technical report, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.
[50] M. Zhu, S. Wright, and T. Chan, *Duality-based algorithms for total-variation-regularized image restoration*, Comput. Optim. Appl., 47 (2010), pp. 377–400.