

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

Yunmei Chen\*      William W. Hager†      Feng Huang‡      Dzung T. Phan§  
Xiaojing Ye¶      Wotao Yin||

## Abstract

This paper presents two fast algorithms for total variation-based image reconstruction in partially parallel magnetic resonance 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 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 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

## 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 non-invasive and non-ionizing 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.

---

\*yun@ufl.edu, <http://www.math.ufl.edu/~yun>, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

†hager@ufl.edu, <http://www.math.ufl.edu/~hager>, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357. Partly supported by National Science Foundation Grant 0620286

‡fhuang@invivocorp.com, Invivo Corporation, 3545 SW 47th Avenue, Gainesville, FL 32608.

§phandu@us.ibm.com, Department of Business Analytics and Mathematical Sciences, IBM T.J. Watson Research Center, PO Box 218, Yorktown Heights, NY 10598. Phone (914) 945-1883. Fax (914) 945-4268.

¶xye@ufl.edu, <http://www.math.ufl.edu/~xye>, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

||wotao.yin@rice.edu, <http://www.caam.rice.edu/~wy1>, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005. Phone (713) 348-5368. Fax (713) 348-5318.

Magnetic resonance 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 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 pseudo random mask for a 2D image. The white pixels correspond to the Fourier component which are measured. The white region in

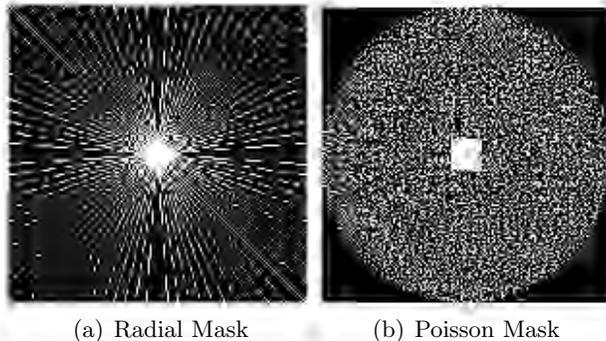


Figure 1: (a) A radial mask with 44 lines, sampling ratio 16.4%. (b) A Poisson pseudo random mask, sampling ratio 25.0%.

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 shows the spacing between the higher frequency Fourier components that are measured. In the Poisson pseudo random 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, 30, 31, 32, 33, 35, 42, 43].

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  $j$ -th receiver, to the sensitivity map  $S_j$  and the mask  $M$ :

$$M\mathcal{F}S_j u = f_j \quad (1)$$

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 gotten 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

$$\min_{u \in \mathbb{C}^N} \sum_{j=1}^K \|MFS_j u - f_j\|_2^2, \quad (2)$$

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

$$\min_{u \in \mathbb{C}^N} \|u\|_{TV} + \lambda \sum_{j=1}^K \|MFS_j u - f_j\|_2^2, \quad (3)$$

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

$$\sum_{j=1}^K \|MFS_j u - f_j\|_2^2.$$

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

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

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

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

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),

$$A = \begin{pmatrix} MFS_1 \\ \vdots \\ MFS_K \end{pmatrix}, \quad (6)$$

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 Fetami in their pioneering work [34] 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,  $\nabla f$  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  $M$ .  $x^\top$  denotes the conjugate transpose of the vector  $x$  and  $\langle x, y \rangle = x^\top y$  is the Euclidean inner product.  $\|\cdot\|_p$  is the  $p$ -norm, and  $\|\cdot\|_{TV}$  is the discrete total variation semi-norm. 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

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

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

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

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 [34] or semi-implicit schemes [24, 36] in which the TV norm was replaced by a smooth approximation

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

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

In [37, 39], a method is developed based on the following reformulation of (7):

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

The linear constraint is treated with a quadratic penalty

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

where  $w = (w_1; \dots; w_N) \in \mathbb{C}^{2N}$  and  $D$  is obtained by stacking the  $D_i$  matrices. For any fixed  $\rho$ , (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 would 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  $\rho$ . The method, however, may not be suitable to more general  $A$ .

In [21] Goldstein and Osher develop a split Bregman method for (11). The resulting algorithm has similar computational complexity to the algorithm in [37]; 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

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

Nonetheless, the algorithms in [21, 37, 39] 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 [44] 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  $\delta$ . 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:

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

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

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

In [11], 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 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 [46] 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  $A^\top A$  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 [47] 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  $u^*$ :

$$0 \in \partial J(u^*) + 2\lambda A^\top (Au^* - f).$$

This is rewritten in the form

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

The iterative scheme is

$$\begin{aligned} s^k &= u^k - \delta\lambda A^\top (Au^k - f), \\ u^{k+1} &= \arg \min_u J(u) + \frac{1}{2\delta} \|u - s^k\|_2^2. \end{aligned}$$

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 [44], which can accommodate an arbitrary matrix  $A$ . In [40], Ye *et al.* proposed a variation of BOS utilizing the Barzilai-Borwein stepsize 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 [40] 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

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

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

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

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:

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

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

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

The first subproblem, denoted (TV), is a TV-based image denoising which has been extensively studied in the literature, and second subproblem (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, 38] and to TV-based SENSE problem in [41], 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  $\alpha$  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  $\alpha$ .

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, 47], variable splitting and continuation [37, 39], split Bregman [21], primal-dual hybrid gradient [46]. In the numerical experiments of Section 5, we used a primal-dual hybrid gradient (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  $X = \{p = (p_1; \dots; p_N) \in \mathbb{C}^{2N} : p_i \in \mathbb{C}^2, \|p_i\|_2 \leq 1, i = 1, \dots, N\}$ . The PDHG algorithm is based on the following updates for the primal and dual variables:

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

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

In Algorithm 1,  $\Pi_X : \mathbb{C}^{2N} \rightarrow \mathbb{C}^{2N}$  is the projection onto  $X$ . For the step (21),  $v^{l+1}$  is a linear combination of  $v^l$ ,  $u^k$  and  $D^\top p^{l+1}$ . The authors in [46] suggested that the step size be updated by

---

**Algorithm 1** PDHG [46] for TV Subproblem
 

---

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

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


---

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 of other methods. In our experiments, we use the suggested updates for  $\tau_l$  and  $\theta_l$ . Note that both steps in Algorithm 1 only require pointwise operations and hence can be computed in parallel. Based on the results given in [46], 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 in [41], 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

$$\min_u \lambda \|Au - f\|_2^2 + \alpha \|v - u\|_2^2. \quad (22)$$

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

$$\|Au - f\|_2^2 \approx \|Au^k - f\|_2^2 + 2(Au^k - f)^\top A(u - u^k) + \delta^k \|u - u^k\|_2^2, \quad (23)$$

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 LS Subproblem
 

---

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


---

Under suitable assumptions [12, 13, 17], the iteration (24) converges linearly to a solution of (22). Each iteration involves multiplication by  $A$  and  $A^\top$  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 algorithm or (AM). In theory, in order to enforce the constraint  $u = v$ , we must let  $\alpha$  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, 45], we obtain the alternating direction method of multipliers (ADMM):

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

ADMM converges to a solution of (15), while AM reaches a solution of (15) only in the limit, as  $\alpha$  tends to infinity. However, we found that in our target application PPI, ADMM and AM have almost identical performance (see Figure 5 of Section 5).

Another approach for treating the penalty term in (17) is the continuation method where the value of  $\alpha$  is gradually increased. The solution for a previous  $\alpha$  is used as a 'warm start' for the next larger  $\alpha$ . However, in our numerical tests with PPI data sets, we found it was more efficient to simply take a fixed, not very large value  $\alpha$ . 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  $\alpha$  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  $\alpha$  does not need to be very large for a suitable image reconstruction, and that further increases in  $\alpha$  may not improve the image quality. And in the case where  $\alpha$  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

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

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

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

In this context, ADMM is

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

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

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

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

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

The PDHG iteration is then written

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

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  $I + 2\alpha\lambda\theta_l \sum_{j=1}^K S_j^\top S_j$ , a diagonal matrix. Therefore, Algorithm 3 only requires 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 [46] for TV Subproblem in (28)

---

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

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


---

## 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  $v \in \mathbb{C}^n$ :

$$(\nabla H(u) - \nabla H(v))(u - v) \geq \sigma \|u - v\|_2^2 \quad (33)$$

Here,  $\nabla H$  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*

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

*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  $2\alpha(u_i - v_i)^\top \in \partial J(v_i)$ , where  $\partial$  denotes the subdifferential. By the convexity of  $J$ , it follows that

$$J(v_2) \geq J(v_1) + 2\alpha(u_1 - v_1)^\top (v_2 - v_1). \quad (35)$$

Likewise, interchanging  $v_1$  and  $v_2$  gives

$$J(v_1) \geq J(v_2) + 2\alpha(u_2 - v_2)^\top (v_1 - v_2). \quad (36)$$

We add (35) and (36) to obtain

$$\|v_2 - v_1\|_2^2 \leq (u_2 - u_1)^\top (v_2 - v_1) \leq \|u_2 - u_1\|_2 \|v_2 - v_1\|_2. \quad (37)$$

Hence,  $\|v_2 - v_1\|_2 = \|\mathcal{T}(u_2) - \mathcal{T}(u_1)\|_2 \leq \|u_2 - u_1\|_2$ , 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

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

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

$$\begin{aligned} \nabla H(u^k) - 2\alpha(v^k - u^k)^\top &= 0, \\ \nabla H(u^*) - 2\alpha(v^* - u^*)^\top &= 0. \end{aligned}$$

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

$$\begin{aligned} (\nabla H(u^k) - \nabla H(u^*))(u^k - u^*) + 2\alpha\|u^k - u^*\|_2^2 &= 2\alpha(v^k - v^*)^\top (u^k - u^*) \\ &\leq 2\alpha\|v^k - v^*\|_2 \|u^k - u^*\|_2. \end{aligned} \quad (39)$$

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

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

which yields

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

Combining this with (38) gives

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

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

**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) 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

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

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.  $\square$

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

$$\|v^{k+1} - v^*\|_2 \leq \|v^k - v^*\|_2 \quad \text{and} \quad \|u^{k+1} - u^*\|_2 \leq \|u^k - u^*\|_2, \quad (42)$$

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  $\mathcal{P} : \mathbb{C}^n \rightarrow \mathbb{C}^n$  satisfies*

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

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

$$\|\mathcal{P}(u) - \mathcal{P}(v)\|_2 \leq \|u - v\|_2 \quad (44)$$

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  $\mathcal{T}$  or  $\mathcal{L}$  are firmly nonexpansive is implied by (37).

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

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

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

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

$\square$

**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

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

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  $\mathcal{T}$  yield

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

Since the right and left 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

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

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

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

Together, (48) and (49) yield

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

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.  $\square$

By the convergence theory for the alternating proximal minimization algorithm, we know that the iterates converge to a solution  $(u^*, v^*)$  of (16) provided 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  $\alpha$ , the solution of (16) generates an approximation to a solution of (4). Let  $\alpha_k$ ,  $k \geq 0$ , 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 [29, 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  $\mu \geq 0$ ,  $\lambda > 0$ , 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  $\alpha$  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 value in (16) corresponding to  $u = v = 0$ . For any choice of  $\alpha$ , the optimal objective function value in (16) is bounded by  $\rho$ . Hence, for any choice of  $\alpha$ , when minimizing the objective function in (16), we should restrict our attention to those  $u$  and  $v$  satisfying

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

Since  $J(v) = \|v\|_{TV} + \mu\|\Psi v\|_1 \geq 0$  and  $H(u) = \|Au - f\|_2^2 \geq 0$ , it follows from (51) that

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

$$\|v\|_{TV} \leq \rho, \quad (53)$$

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

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

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

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

$$\|Du\|_2 \geq \gamma_1 \|u\|_2 \text{ for all } u \in \mathcal{N}(A).$$

Hence, by (55)

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

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

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

Hence, by (54), we have

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

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  $\alpha$  increases, the level set of (16) corresponding to the objective function value  $\rho$  can only shrink. Hence, this level set is bounded for any  $\alpha \geq \alpha_0$ . Let  $\alpha_k$ ,  $k = 0, 1, \dots$ , denote an increasing sequence of values for the penalty tending to infinity, and let  $(U^k, V^k)$  denote associated solutions of (16). By [29, Thm. 17.1], every convergent subsequence of the minimizers  $(U^k, V^k)$  approaches a solution of (15).  $\square$

**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  $\alpha$  tends to infinity.

## 5 Numerical Experiments

In this section we evaluate the performance of algorithms using two PPI reconstructions. We compare performance of AM (18), ADMM (25), and APD (28) to that of the Bregman operator splitting (BOS) in [44] and a slightly modified version (SBB) of the algorithm proposed in [40].

### 5.1 Data Acquisition and Experimental Setup

In our tests, all  $k$ -space data were 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 data1, is a collection of sagittal Cartesian brain images acquired on a 3T GE system (GE Healthcare, Waukesha, Wisconsin, USA). The data acquisition parameters were FOV  $220\text{mm}^2$ , size  $512 \times 512 \times 8$ , TR 3060ms, TE 126ms, slice thickness 5mm, and flip angle  $90^\circ$ . 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 T2-weighted turbo spin echo (T2 TSE) sequence. The acquisition parameters were FOV  $205\text{mm}^2$ , matrix  $512 \times 500 \times 8$ , TR 3000ms, and TE 85ms. The echo train length was 20.

The ground truth or reference image  $\bar{u}$  was given by the formula

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

Here  $\bar{u}_{ij}$  is the  $i$ -component of the Fourier transform of the full  $k$ -space data acquired on the  $j$ -th channel. In all cases, we simulate the sensitivity maps  $S_j$  using the *central*  $32 \times 32$   $k$ -space data, and generate the pseudo full  $k$ -space data by  $\mathcal{F}S_j\bar{u}$ . The sensitivity map for data2 is shown in Figure 2. We add a complex valued Gaussian noise (same level for both real and imagery parts) with standard deviation 0.01 in magnitude to the pseudo full data, and downsample the pseudo full data using the mask shown in Figure 1(a) for data1 and the mask shown in Figure 1(b) for data2.

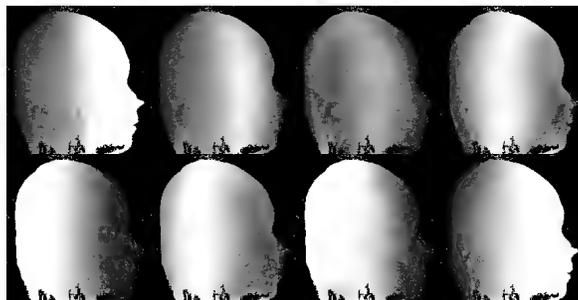


Figure 2: The sensitivity maps for the eight channels of data2.

Algorithms were implemented in MATLAB, Version R2009b. All the experiments were 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 proposed Bregman operator splitting (BOS)

scheme from [44] and a split Bregman scheme SBB [40] utilizing a BB step size. BOS and SBB are currently, to the best of the authors knowledge, among the most efficient methods for solving (7) with arbitrary matrix  $A$ . The BOS scheme of [44] and the SBB scheme of [40] are iterative methods for solving (10). They are based on the alternating direction method of multipliers applied to the Lagrangian (12), and they correspond to the following iteration:

$$\left. \begin{aligned} w_i^{k+1} &= \max\{\|D_i^k u + b_i^k\|_2 - 1/2\rho, 0\} (D_i^k u + b_i^k) / \|D_i^k u + b_i^k\|_2 \quad \forall \quad i, \\ u^{k+1} &= \left(\rho D^\top D + \lambda \delta_k I\right)^{-1} \left(\rho D^\top (w^{k+1} - b^k) + \lambda \delta_k u^k - \lambda A^\top (A u^k - f)\right), \\ b_i^{k+1} &= b_i^k - (w_i^{k+1} - D_i u^{k+1}) \quad \forall \quad i. \end{aligned} \right\} \quad (58)$$

In the BOS scheme of [44],  $\delta_k$  is a constant  $\delta$  and convergence to a solution of (7) holds when  $\delta \in (\|A^\top A\|_2, \infty)$ . In the SBB scheme,  $\delta_k$  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 2D shrinkage. In the  $u$ -subproblem,  $D^\top D$  can be diagonalized by a Fourier transform provided the image has periodic boundary condition; hence,  $\rho D^\top D + \lambda \delta_k I$  can be inverted easily. Consequently, the main computational cost corresponds to multiplication by  $A^\top$  and by  $A$ , or equivalently, to the evaluation of two fast 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  $\rho$  in  $[10^0, 10^2]$  do not have much influence on the results. For BOS where  $\delta_k$  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  $\alpha$  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 least squares subproblem of AM and ADMM are terminated when the relative change of the inner iterates is less than  $\epsilon_{\text{inn}} = 10^{-2}$ . Dynamically adjusting  $\epsilon_{\text{inn}}$  as the outer iterates approaches the solution can lead to improved efficiency, however, in our experiments, a constant  $\epsilon_{\text{inn}}$  already leads to better performance than most recently developed algorithms.

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. 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.

The reconstructed images for data1 and data2 are shown in Figures 3 and 4, respectively. The relative error in the reconstructed image  $u$ ,  $\|u - \bar{u}\|_2 / \|\bar{u}\|_2$ , is indicated in parenthesis in the Figures. In Figures 3 and 4, we zoom into the square shown in the boxes of Figure 3(a) and 4(a). It is seen that all methods adequately recovered the image while SBB, APD, AM, and ADMM have higher accuracy than BOS for this stopping criterion.

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 5(a) and 5(b), 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 fastest, closely followed by AM and ADMM. BOS has the slowest convergence speed for these data sets. In Table 1 we summarize the comparison of these algorithms for these two data sets.

Table 1: Comparison of the tested algorithms

Criterion	BOS	SBB	APD	AM/ADMM
Efficiency in PPI?	Relatively low	Very high	Very high	Relatively high
Works for general $A$ ?	Yes	Yes	No	Yes
Convergence established?	Yes	No	Yes	Yes

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.

**Remark.** In some PPI applications, the images are sparse under an orthogonal wavelet transform  $\Psi$ . 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, ADP, and SBB) remain valid with small modifications.

## 6 Conclusions

Two fast algorithm for total variation-based image reconstruction were introduced. The first method AM employs variable splitting, a quadratic penalty, and an alternating 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. An implementation based on a primal-dual hybrid gradient (PDHG) scheme for the TV problem and a Barzilai-Borwein 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) [44], a modified algorithm SBB [40] where the constant BOS stepsize is replaced by a variable stepsize based on the BB algorithm [5], and the alternating direction method of multipliers ADMM [7, 14, 19, 20]. It was found that for the same stopping criteria, SBB and APD produced the highest quality images, AM and ADMM were second in quality, and BOS was third. BOS was significantly slower than the other four methods, while both SBB and APD were slightly faster than both AM and ADMM for two Cartesian brain images.

## References

- [1] F. ACKER AND M.-A. PRESTEL, *Convergence d'un schéma de minimisation alternée*, 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. 1–17.
- [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 cost-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, *Parallel and Distributed Computation*, Prentice Hall, 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. CANDÉS, 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 Vis., 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. Optim., 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*, Mathematical 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 Proc. Intl. Soc. Mag. Reson. Med., 2003, p. 2346.
- [16] E. ESSER, X. ZHANG, AND T. CHAN, *A general framework for a class of first order primal-dual algorithms for TV minimization*, Tech. Rep. 09-67, CAM UCLA, 2009.
- [17] A. FRIEDLANDER, J. M. MARTÍNEZ, B. MOLINA, AND M. RAYDAN, *Gradient method with retracts 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., Amsterdam, 1983, North Holland, 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 nonlinéaires*, *RAIRO Analyse Numérique*, 9 (1975), pp. 41–76.
- [21] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L1 regularized problems*, *SIAM J. Imag. 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*, *IEEE Proc. Conf. on Comp. Vis. Patt. Recog.*, (2008), pp. 1–8.
- [29] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer, New York, 1999.
- [30] 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.
- [31] K. PRUESSMANN, M. WEIGER, M. SCHEIDEGGER, AND P. BOESIGER, *SENSE: Sensitivity encoding for fast MRI*, *Magn. Reson. Med.*, 42 (1999), pp. 952–962.
- [32] Y. QIAN, Z. ZHANG, V. STENGER, AND Y. WANG, *Self-calibrated spiral SENSE*, *Magn. Reson. Med.*, 52 (2004), pp. 688–692.
- [33] 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.
- [34] L. RUDIN, S. OSHER, AND E. FATEMI, *Non-linear total variation noise removal algorithm*, *Physics D.*, 60 (1992), pp. 259–268.
- [35] A. SAMSONOV AND C. JOHNSON, *Non-cartesian POCSENSE*, in *Proc. Intl. Soc. Mag. Reson. Med.*, 2004, p. 2648.

- [36] C. R. VOGEL AND M. E. OMAN, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.
- [37] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., 1 (2008), pp. 248–272.
- [38] 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.
- [39] J. YANG, Y. ZHANG, AND W. YIN, *A fast TVL1-L2 minimization algorithm for signal reconstruction from partial Fourier data*, Tech. Rep. 08-29, CAAM Rice Univ., 2008.
- [40] X. YE, Y. CHEN, AND F. HUANG, *Computational acceleration for MR image reconstruction in partially parallel imaging*, IEEE Trans. Med. Imaging, (2010, DOI: 10.1109/TMI.2010.2073717).
- [41] 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. Imaging, (2010, DOI: 10.1109/TMI.2010.2088133).
- [42] 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., 54 (2005), pp. 1–8.
- [43] 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., 60 (2008), pp. 414–421.
- [44] X. ZHANG, M. BURGER, X. BRESSON, AND S. OSHER, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, Tech. Rep. 09-03, CAM UCLA, 2009.
- [45] X. ZHANG, M. BURGER, AND S. OSHER, *A unified primal-dual algorithm framework based on Bregman iteration*, Tech. Rep. 09-99, CAM UCLA, 2009.
- [46] M. ZHU AND T. CHAN, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Tech. Rep. 08-34, CAM UCLA, 2008.
- [47] M. ZHU, S. WRIGHT, AND T. CHAN, *Duality-based algorithms for total-variation-regularized image restoration*, Comput. Optim. Appl., 47 (2010), pp. 377–400.

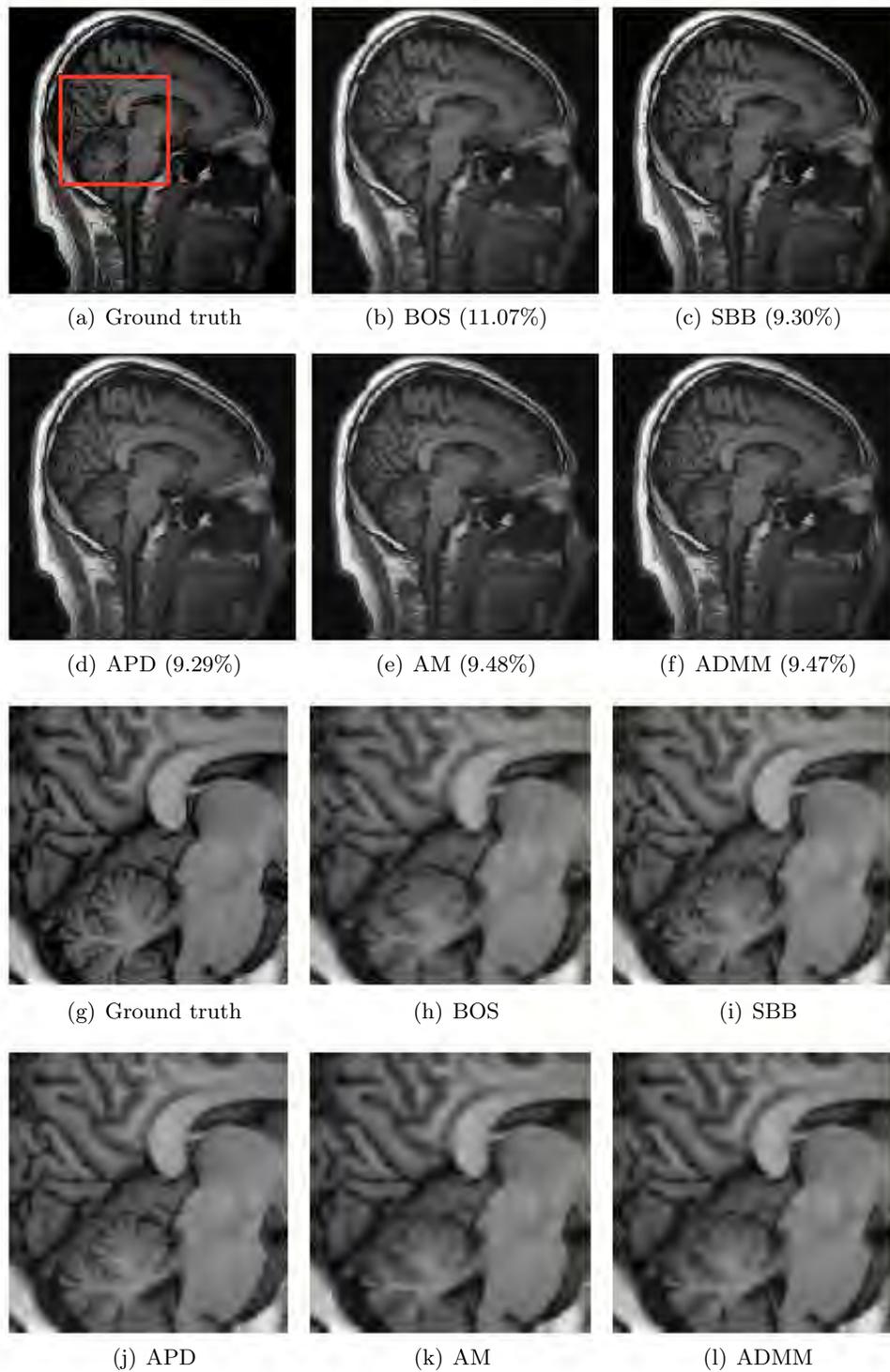


Figure 3: Reconstructed images of data1. (g)–(l) zoom in on the box in (a)–(f), respectively. Corresponding relative errors are indicated in parenthesis.

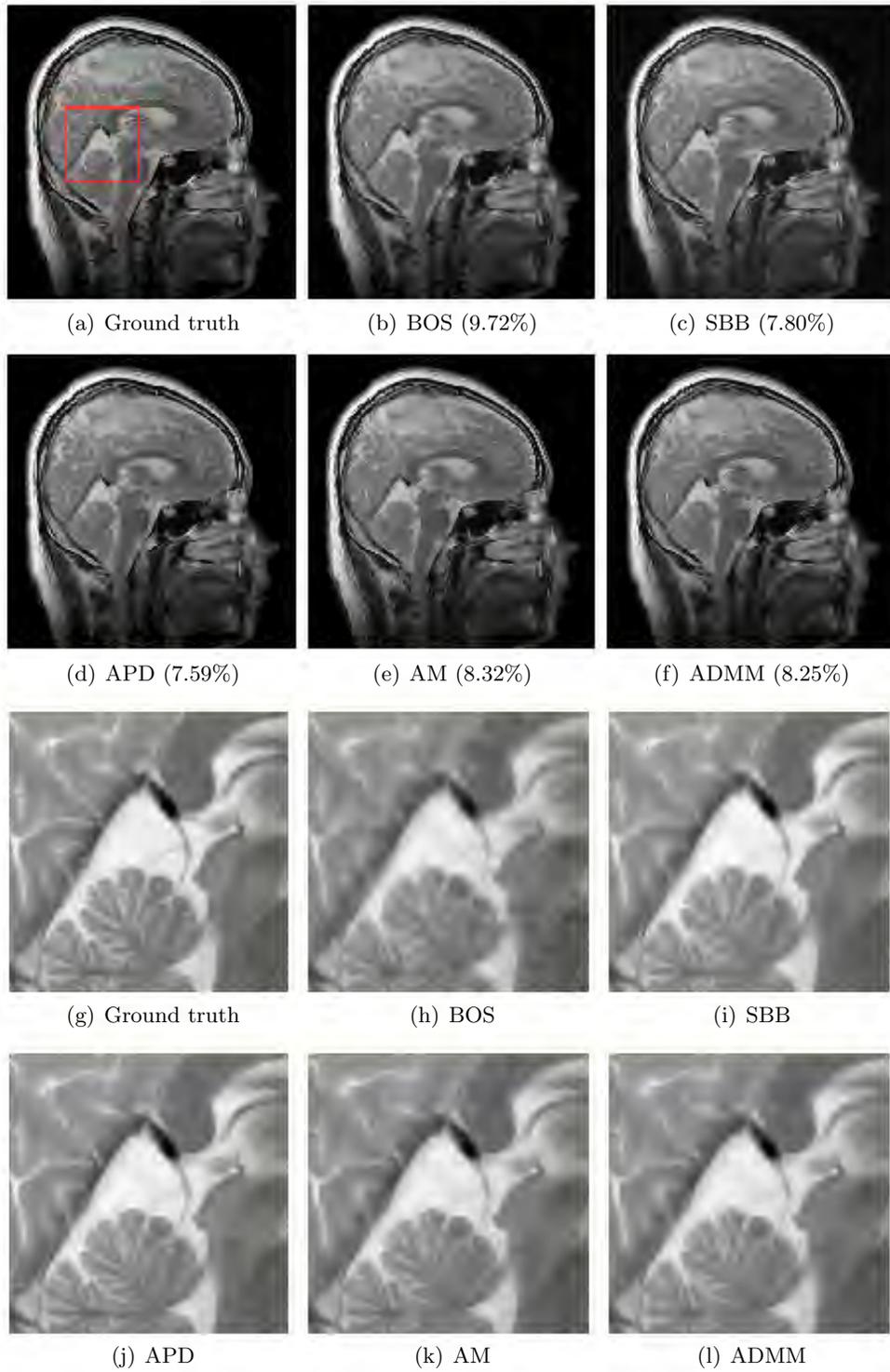
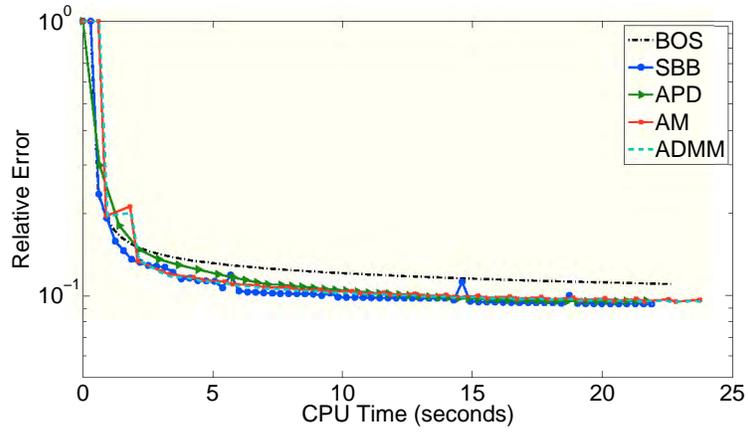
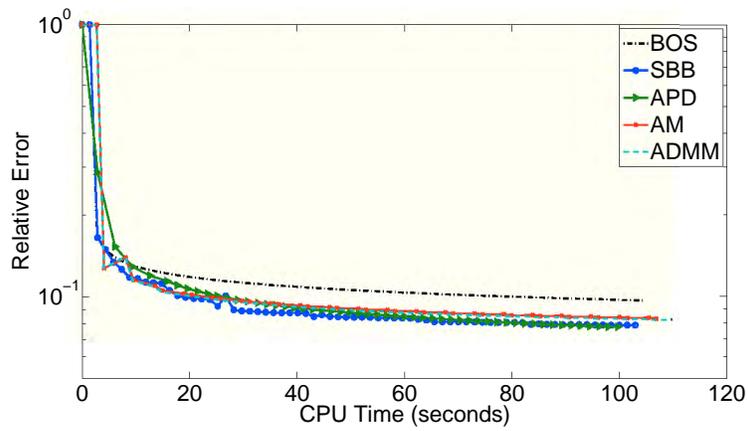


Figure 4: Reconstructed images of data2. (g)–(l) zoom in on the box in (a)–(f), respectively. Corresponding relative errors are indicated in parenthesis.



(a) Relative errors versus CPU time of data1.



(b) Relative errors versus CPU time of data2.

Figure 5: Comparison of BOS, SBB, APD, AM and ADMM on data1 and data2.