

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/50225734>

# Fast MR Image Reconstruction for Partially Parallel Imaging With Arbitrary $k$ -Space Trajectories

Article · March 2011

DOI: 10.1109/TMI.2010.2088133 · Source: PubMed

CITATIONS

30

READS

162

4 authors:



**Xiaojing Ye**

Georgia State University

29 PUBLICATIONS 391 CITATIONS

[SEE PROFILE](#)



**Yunmei Chen**

University of Florida

90 PUBLICATIONS 2,510 CITATIONS

[SEE PROFILE](#)



**Wei Lin**

Philips

29 PUBLICATIONS 306 CITATIONS

[SEE PROFILE](#)



**Feng Huang**

64 PUBLICATIONS 1,441 CITATIONS

[SEE PROFILE](#)

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



fluid mechanics [View project](#)

All content following this page was uploaded by **Wei Lin** on 16 May 2014.

The user has requested enhancement of the downloaded file.

# Fast MR Image Reconstruction for Partially Parallel Imaging With Arbitrary $k$ -Space Trajectories

Xiaojing Ye\*, Yunmei Chen, Wei Lin, and Feng Huang

**Abstract**—Both acquisition and reconstruction speed are crucial for magnetic resonance (MR) imaging in clinical applications. In this paper, we present a fast reconstruction algorithm for SENSE in partially parallel MR imaging with arbitrary  $k$ -space trajectories. The proposed method is a combination of variable splitting, the classical penalty technique and the optimal gradient method. Variable splitting and the penalty technique reformulate the SENSE model with sparsity regularization as an unconstrained minimization problem, which can be solved by alternating two simple minimizations: One is the total variation and wavelet based denoising that can be quickly solved by several recent numerical methods, whereas the other one involves a linear inversion which is solved by the optimal first order gradient method in our algorithm to significantly improve the performance. Comparisons with several recent parallel imaging algorithms indicate that the proposed method significantly improves the computation efficiency and achieves state-of-the-art reconstruction quality.

**Index Terms**—Convex optimization, image reconstruction, parallel imaging, SENSE.

## I. INTRODUCTION

PARTIALLY parallel imaging (PPI) [1], [2] can efficiently reduce acquisition time, increase temporal/spatial resolution, and suppress motion related artifacts in magnetic resonance imaging (MRI) applications. Its clinical value has been demonstrated in both anatomical and functional imaging applications [3]–[6]. In PPI, a set of multichannel imaging data is acquired simultaneously from a radio-frequency (RF) coil array. The imaging is accelerated by sampling a reduced number of  $k$ -space samples. The acceleration factor (reduction factor) is defined as the reciprocal of the percentage of the acquired number of  $k$ -space samples. A special image reconstruction method is required to generate a full field-of-view (FOV) image from the undersampled data set.

Cartesian sampling trajectories were commonly used in commercial products and have achieved great success in fast PPI.

Manuscript received July 22, 2010; revised September 22, 2010; accepted September 28, 2010. Date of publication October 21, 2010; date of current version March 02, 2011. Asterisk indicates corresponding author.

\* X. Ye is with the Department of Mathematics, University of Florida, Gainesville, FL 32611 USA (e-mail: xye@ufl.edu).

Y. Chen is with the Department of Mathematics, University of Florida, Gainesville, FL 32611 USA (e-mail: yun@ufl.edu).

W. Lin and F. Huang are with the Advanced Concept Development, In-vivo Corporation, Philips HealthCare, Gainesville, FL 32608 USA (e-mail: wei.lin2@philips.com; f.huang@philips.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2010.2088133

In recent years, the advantages of non-Cartesian imaging have been demonstrated by many groups [7]–[12]. For example, radial trajectories have been successfully used for highly-accelerated dynamic imaging [13], [14]. It also has advantages in motion artifacts reduction [9], [10] and ability to achieve ultra-short echo times [15]. Acceleration of such non-Cartesian acquisitions through PPI techniques, namely non-Cartesian PPI, is currently a subject of much interests [16]–[27]. Advances on this subject will be a great benefit to many clinical applications, such as neuroimaging [28]–[31], cardiac applications [32], [33], and hyperpolarized MR acquisitions [34]. Despite of these advantages and developments of non-Cartesian PPI, there is no commercial non-Cartesian PPI package available yet. One major reason is that reconstruction takes longer than clinically acceptable time.

Similar to Cartesian PPI, the reconstruction techniques for non-Cartesian PPI can be divided into two categories: reconstruction in  $k$ -space [18], [23], [25], [27], [38] and reconstruction in image space [16], [20]–[22], [24], [26], [39], [40]. The  $k$ -space based methods use coil sensitivity variations to reconstruct the missing  $k$ -space data, then apply Fourier transform on the original and reconstructed data to obtain the unaliased image. Some  $k$ -space based methods [17], [18], [37], [38] require the symmetry of the  $k$ -space trajectory, and hence their applicability is limited. Other methods [23], [25], [27], [35], [36], which do not need the symmetry of the  $k$ -space trajectory, suffer the slow reconstruction speed which could be minutes for a single 2-D image.

Sensitivity encoding (SENSE) [1], [16] is one of the most commonly used image-space based reconstruction methods in PPI. It uses knowledge of the coil sensitivities to separate aliased pixels. SENSE involves an inversion of the encoding matrix produced by coil sensitivity modulation. For the case of sampling along a regular Cartesian  $k$ -space grid, a Fourier transform may be separated out from the inversion process as a distinct step, and the encoding matrix can be significantly simplified. However, for non-Cartesian  $k$ -space trajectories, the  $k$ -space samples are not uniformly distributed and fast Fourier transform (FFT) cannot be applied in a straightforward manner. In these cases, the inversion is much more complicated and time consuming. In [16], Pruessmann *et al.* proposed an approach that performs the reconstruction iteratively, and, combines FFT with forward and reverse gridding operations for efficient execution of the conjugate gradient (CG) iteration loops. However, gridding in each iteration consumes considerable computational time. Moreover, due to the ill-conditioning of the encoding matrix, it has been shown in [21] that the CG iteration sequence often exhibits a “semi-convergence” behavior, which

can be characterized as initially converging toward the exact solution and later diverging. Moreover, the convergence speed is low, when the acceleration factor is high.

To alleviate the ill-conditioning problem regularization techniques are often used [24], [26]. For instance, total variation (TV) based regularization has been incorporated into SENSE to improve reconstructed image quality and convergence speed over the unregularized CG method [24], [26]. In [41], the authors exploited that MR images are sparse with respect to spatial finite differences and wavelet transform, and proposed a model that employs the sparsity as regularization subjected to data consistency to reconstruct MR images from partial Fourier data. As shown in [41], the idea of reconstructing from partial data provided sparsity constraints coincides with compressed sensing (CS), which has been very successful in signal/image processing. Therefore the combination of CS and PPI (CS-PPI) becomes an emerging research topic in MR imaging for clinical applications. However, CS-PPI reconstructions suffer the non-differentiability of those  $\ell_1$ -like terms induced by sparsity.

The main purpose of this paper is to develop a fast numerical algorithm to tackle the computation problem in CS-PPI reconstruction. The proposed algorithm employs the variable splitting method and classical quadratic penalty technique to reformulate the sparsity constrained SENSE model as an unconstrained optimization problem. Then the solution to the new formulation can be obtained by alternating two simple minimizations: one is the TV and wavelet based image denoising problem which has been extensively studied in the literature and can be quickly solved by several recently developed methods; the other one is a large scale least squares problem involving an ill-conditioned inversion matrix, for which we apply an optimal gradient method for fast approximation to the inverse solution. As can be seen, our scheme requires moderate coding complexity but has high efficiency in searching for the optimal reconstruction. Moreover, the proposed method bears any  $k$ -space sampling patterns in PPI. When the  $k$ -space trajectory is non-Cartesian, we utilize the GRAPPA operator gridding (GROG) [42] to shift data onto Cartesian grids as a preprocess of data such that (inverse) gridding in each iteration can be waived. The reconstruction results show significant improvements on efficiency by using the proposed algorithm to achieve state-of-the-art reconstruction quality when compared to those by other PPI methods.

## II. PRIOR ARTS

Before going into details of CS-PPI and SENSE, we here address the notations used throughout the rest of the paper. First of all, all vectors in this paper are column vectors unless otherwise noted. Following the standard treatment, we will vectorize 2-D images or higher dimensional data into 1-D vectors. For instance, an image  $u$  consisting of  $N$  pixels is treated as an  $N$ -vector, i.e.,  $u \in \mathbb{C}^N$ . Let  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote the  $\ell_1$  norm and Euclidean norm of vectors, respectively, and  $\langle \cdot, \cdot \rangle$  denote the regular inner product. The superscript  $\top$  represents conjugate transpose of matrices. Following Matlab convention we let  $(;\cdot)$  denote the matrix formed by stacking the arguments vertically.

### A. Sensitivity Encoding (SENSE)

SENSE is an image-space based reconstruction method in PPI that uses knowledge of the coil sensitivities to separate aliased pixels. The fundamental of SENSE and the variations is as follows. In a  $J$  channels coil array, the partial  $k$ -space data  $f_j$  acquired from the  $j$ th channel relates to the true image  $\bar{u}$  through the sensitivity map  $S_j$  by

$$P\mathcal{F}(S_j\bar{u}) = f_j \quad j = 1, \dots, J \quad (1)$$

where  $\mathcal{F}$  is the Fourier transform and  $P$  is the common undersampling pattern (mask) used by all  $J$  channels. If  $f_j$ , the  $k$ -space data acquired from the  $j$ th channel, is on Cartesian grids, then the solution to (1) can be obtained by solving a least squares problem

$$\min_{u \in \mathbb{C}^N} \sum_{j=1}^J \|\mathcal{F}_p(S_j \odot u) - f_j\|_2^2 \quad (2)$$

where  $\mathcal{F}_p$  is the undersampled Fourier transform defined by  $\mathcal{F}_p \triangleq P\mathcal{F}$ , and  $\mathcal{F} \in \mathbb{C}^{N \times N}$  is the discrete Fourier transform matrix and  $P \in \mathbb{R}^{p \times N}$  is a selection matrix containing  $p$  rows of the identity matrix of order  $N$  according to the undersampling mask. In (2),  $S_j \in \mathbb{C}^N$  is the sensitivity map and  $f_j \in \mathbb{R}^p$  is the partial  $k$ -space data acquired from the  $j$ th channel. The symbol  $\odot$  is the Hadamard (or componentwise) product between two vectors. Notations in (2) can be simplified by letting  $A \in \mathbb{C}^{pJ \times N}$  and  $f \in \mathbb{C}^{pJ}$  denote the sensitivity encoding matrix and data acquired from all  $J$  channels, respectively, i.e.,

$$A \triangleq (\mathcal{F}_p S_1; \dots; \mathcal{F}_p S_J) \quad \text{and} \quad f \triangleq (f_1; \dots; f_J) \quad (3)$$

where  $S_j \triangleq \text{diag}(S_j) \in \mathbb{C}^{N \times N}$  is the diagonal matrix with  $S_j \in \mathbb{C}^N$  as the diagonal,  $j = 1, \dots, J$ . Then problem (2) can be expressed as

$$\min_{u \in \mathbb{C}^N} \|Au - f\|_2^2. \quad (4)$$

### B. SENSE With Sparsity Constraints

As shown earlier, problem (4) can yield unstable computation process due to the severe ill-conditioning of  $A$ . Moreover, to yield a well-posed problem, the reduction factor  $N/p$  can never exceed  $J$ , the number of channels in the coil array. These limitations hinder (4) in clinical applications. An appropriate remedy is by imposing constraints on the sparsity. As pointed out in [43], MR images are usually sparse in certain transform domains, such as finite difference and wavelet. For instance, consider an MR image  $\bar{u} \in \mathbb{C}^N$  and let  $\Psi = (\psi_1, \dots, \psi_N) \in \mathbb{C}^{N \times N}$  denote the (orthogonal) wavelet transform matrix, then  $\Psi^\top \bar{u} \in \mathbb{C}^N$ , the wavelet transform of  $\bar{u}$ , is sparse. Thus the  $\ell_1$  term  $\|\Psi^\top u\|_1$  can be used as the sparsity regularization in the reconstruction of  $\bar{u}$ . Hence, the SENSE model with sparsity constraints reads as

$$\min_{u \in \mathbb{C}^N} \|u\|_{TV} + \mu \|\Psi^\top u\|_1 + \frac{\lambda}{2} \|Au - f\|_2^2 \quad (5)$$

where  $\|u\|_{TV}$  is the (isotropic) TV norm defined by

$$\|u\|_{TV} \triangleq \sum_{i=1}^N \|D_i u\|_2 = \sum_{i=1}^N \sqrt{|D_i^x u|^2 + |D_i^y u|^2}. \quad (6)$$

Here  $D_i^x, D_i^y$  are the rows of  $D_i \in \mathbb{R}^{2 \times N}$  which has two nonzero entries in each row corresponding to finite difference approximations to partial derivatives of  $u$  at the  $i$ th pixel along the  $x$ - and  $y$ -coordinate axes. In (5),  $\mu, \lambda > 0$  are weights for the  $\ell_1$  and data consistency terms.

### III. PROPOSED ALGORITHM

#### A. Motivation

The main purpose of this paper is to develop a fast numerical algorithm for solving the SENSE model with sparsity constraints as in (5). A fast reconstruction algorithm can make CS-PPI more clinically applicable.

The computational challenge for solving (5) comes from the combination of two issues: one is the nondifferentiability of the TV and  $\ell_1$  terms, and the other one is the large size and severe ill-conditioning of the matrix  $A$ . To deal with nondifferentiability issue, many work used gradient decent method on (5) with the TV and  $\ell_1$  terms substituted by their smooth approximations, e.g., (17). But their results were sensitive to the perturbation parameter in the smooth approximations. Moreover, the high nonlinearity of the TV/ $\ell_1$  regularization terms and the severe ill-conditioning of  $A$  make conventional optimization approaches such as the nonlinear conjugate gradient (NLCG) method very inefficient. We will show in Section V the inferior performance of NLCG when applied to (5) with smoothed TV and  $\ell_1$  regularization (17).

Recently, there are several fast algorithms developed to solve the problem (5) with original TV norm (6) in the case that  $A$  is identity or a convolution operator in image denoising and deblurring, see, e.g., [44]–[46]. The efficiency of these methods relies on a very special condition on  $A$  such that  $A^T A$  is diagonalizable by fast transforms. However in SENSE applications,  $A$  is comprised of the undersampling Fourier operator  $\mathcal{F}_p$  and sensitivity maps  $S_j$  as shown in (3), and therefore does not satisfy this condition.

The idea of this work is to decouple the difficulties from the nonsmooth TV and  $\ell_1$  norms and the ill-conditioned  $A$ , and reduce the original problem to two minimization subproblems that can be solved efficiently. More precisely, we use the variable splitting and quadratic penalty techniques to split the original problem as a TV denoising subproblem and a least squares subproblem. The TV/ $\ell_1$  subproblem can be solved efficiently by any recently developed fast algorithm such as the split Bregman method, primal-dual method and many others, e.g., [44]–[46]. The least squares subproblem can be solved by an optimal gradient (OG) method which is very robust even with ill-conditioned  $A$ .

#### B. Variable Splitting

To tackle the computational problem of (5), we first introduce an auxiliary variable  $v$  to transform  $u$  out of the nondifferentiable terms, then apply quadratic penalty technique to reformulate (5) as an unconstrained optimization problem.

Let  $v \in \mathbb{C}^N$  be the auxiliary variable and rewrite (5) as the following constrained minimization problem:

$$\min_{u, v \in \mathbb{C}^N} \|v\|_{TV} + \mu \|\Psi^T v\|_1 + \frac{\lambda}{2} \|Au - f\|_2^2, \text{ s.t. } u = v. \quad (7)$$

Clearly minimization problem (7) is equivalent to (5) as they share the same solution  $u$ . In order to solve (7), we relax the equality constraint  $u = v$  and penalize its violation by quadratic functions and obtain

$$\min_{u, v \in \mathbb{C}^N} \|v\|_{TV} + \mu \|\Psi^T v\|_1 + \frac{\alpha}{2} \|u - v\|_2^2 + \frac{\lambda}{2} \|Au - f\|_2^2. \quad (8)$$

Then solving (8) provides an approximation to the solution (7) when  $\alpha$  is sufficiently large.

As the objective function in (8) is convex with respect to all variables, the solution can be obtained by alternately minimizing the objective function with respect to each of  $u$  and  $v$ . Namely, we solve (8) by alternating the following two minimization subproblems, termed by (TV) and (LS):

$$\begin{cases} \min_{v \in \mathbb{C}^N} \|v\|_{TV} + \mu \|\Psi^T v\|_1 + \frac{\alpha}{2} \|v - u\|_2^2 & \text{(TV)} \\ \min_{u \in \mathbb{C}^N} \frac{\alpha}{2} \|u - v\|_2^2 + \frac{\lambda}{2} \|Au - f\|_2^2 & \text{(LS).} \end{cases} \quad (9)$$

The first subproblem has the same form as TV-Wavelet (TV) based image denoising which has been extensively studied in the literature, and the second one is a least squares (LS) problem thanks to the quadratic form of both terms. The case is both subproblems can be processed in highly stable and efficient manners.

It was brought to our attention that a similar approach to image deblurring/restoration model (8) without the  $\ell_1$  term (i.e.,  $\mu = 0$ ) was introduced in [47] lately. In [47], the authors directly proposed to solve (8) with fixed parameter  $\alpha$  as an approximation to the original problem (5). They used Chambolle's dual projection method [48] and preconditioned conjugate gradient (PCG) method to solve the resulting (TV) and (LS) subproblems, respectively. However, the Chambolle's method is not as efficient as many recent methods for solving (TV) subproblem, and it cannot be applied directly to the case where both TV and  $\ell_1$  terms are presented as in (9). Moreover, PCG is also not adaptive in SENSE applications:  $A$  in (5) comprises sensitivity maps which vary in different scans/subjects, and therefore the preconditioner has to be computed case by case. In this paper, we propose to use the recently developed split Bregman method and Nesterov's optimal gradient method to solve (9) which readily tackle the complications in SENSE applications. A comprehensive analysis is also provided in Appendix B and C to exploit the convergence behavior of scheme (9). After we initially submitted this paper, we developed an algorithm with improved performance [62] but its theoretical convergence is not well established as the one presented in this paper

#### Algorithm for (TV) Subproblem

Subproblem (9)-(TV) is the TV-Wavelet based image denoising problem that solves for a "clean/smoothed" image  $v$  given a "noisy" image  $u$ . There have been numerous fast algorithms developed for this problem by taking advantages of the simplicity of the consistency term  $\|v - u\|_2^2$ . Recent

works include using dual formulation [48], variable splitting and continuation [44], [49], split Bregman [45], primal-dual hybrid gradient [46], [50]. Interested readers are referred to these papers and references therein. Since our algorithm does not depend on using specific method, we chose split Bregman which proves to be one of the fastest methods for TV-Wavelet based image denoising. Split Bregman method solves (9)-(TV) by iterating

$$\begin{cases} v^{n+1} = \arg \min_{v \in \mathbb{C}^N} \frac{\beta}{2} \|Dv - w^n + b^n\|_2^2 \\ \quad + \frac{\mu\beta}{2} \|\Psi^\top u - z^n + c^n\|_2^2 + \frac{\alpha}{2} \|v - u\|_2^2 \\ w_i^{n+1} = \arg \min_{w_i \in \mathbb{C}^2} \|w_i\|_2 + \frac{\beta}{2} \|w_i - D_i v^{n+1} - b_i^n\|_2^2, \quad \forall i \\ z^{n+1} = \arg \min_{z \in \mathbb{C}^N} \|z\|_1 + \frac{\beta}{2} \|z - \Psi^\top v^{n+1} - c^n\|_2^2 \\ b^{n+1} = b^n + (Dv^{n+1} - w^{n+1}) \\ c^{n+1} = c^n + (\Psi^\top v^{n+1} - z^{n+1}) \end{cases} \quad (10)$$

where  $D = (D^x; D^y) \in \mathbb{R}^{2N \times N}$ , where  $D^x$  and  $D^y$  are  $NN$  matrices that stack the  $D_i^x$ 's and  $D_i^y$ 's, respectively. Split Bregman (10) is equivalent to the well-known alternating direction method of multipliers (ADMM) [51]–[54] applied to augmented Lagrangian defined by

$$\begin{aligned} \mathcal{L}^\beta(v, w, z; b, c) \\ = \sum_i \left( \|w_i\|_2 + \beta \langle b_i, D_i u - w_i \rangle + \frac{\beta}{2} \|D_i u - w_i\|_2^2 \right) \\ + \mu \left( \|z\|_1 + \beta \langle c, \Psi^\top u - z \rangle + \frac{\beta}{2} \|\Psi^\top u - z\|_2^2 \right) \\ + \frac{\alpha}{2} \|v - u\|_2^2 \end{aligned}$$

and hence the convergence follows for any  $\beta > 0$ . In each iteration, split Bregman (or ADMM) minimizes  $\mathcal{L}^\beta$  with respect to  $v$ ,  $w$  and  $z$  one by one, and then updates the multipliers  $b$  and  $c$ , as shown in (10). Since the last term in  $v$ -subproblem of (10) has identity matrix (instead of  $A$ ) multiplied to  $v$ , this  $v$ -subproblem is extremely easy to solve and split Bregman can be very efficient when applied (9)-(TV). The split Bregman(10) can be summarized as in Algorithm 1. In Algorithm 1,  $\beta$  is a user-defined moderate positive number ( $= 10$  in our experiments) and  $\eta = \alpha/\beta$ , and  $L \triangleq (\mu + \eta)I + D^\top D$  can be diagonalized by discrete Fourier transform if the underlying image has periodic boundary condition. Therefore  $L^{-1}$  can be easily computed. `shrink2` and `shrinkc` refer to 2D and componentwise soft shrinkage operators, respectively [44], [45], [49]. Namely, for any fixed  $\mathbf{t} \in \mathbb{C}^2$ , the 2D soft shrinkage has the closed form of  $\mathbf{s}$  as follows:

$$\mathbf{s} = \max \left\{ \|\mathbf{t}\|_2 - \frac{1}{\beta}, 0 \right\} \cdot \frac{\mathbf{t}}{\|\mathbf{t}\|_2} \quad (11)$$

which uniquely solves the minimization

$$\min_{\mathbf{s} \in \mathbb{C}^2} \|\mathbf{s}\|_2 + \frac{\beta}{2} \|\mathbf{s} - \mathbf{t}\|_2^2 \quad (12)$$

whereas for any fixed  $t \in \mathbb{C}^N$ , the componentwise shrinkage sets each components  $s_i$  of  $\mathbf{s} = (s_1; \dots; s_N) \in \mathbb{C}^N$  by

$$s_i = \max \left\{ |t_i| - \frac{1}{\beta}, 0 \right\} \cdot \frac{t_i}{|t_i|} \quad (13)$$

such that  $\mathbf{s}$  uniquely solves the minimization

$$\min_{\mathbf{s} \in \mathbb{C}^N} \|\mathbf{s}\|_1 + \frac{\beta}{2} \|\mathbf{s} - \mathbf{t}\|_2^2 = \min_{\mathbf{s} \in \mathbb{C}^N} \sum_{i=1}^N \left( |s_i| + \frac{\beta}{2} |s_i - t_i|^2 \right) \quad (14)$$

with convention  $0/|0| \triangleq 0$ . Operations `shrink2` and `shrinkc` in Algorithm 1 perform the above simple shrinkages to each  $D_i v^{n+1} + b_i^n \in \mathbb{C}^2$  for  $i = 1, \dots, N$  and  $\Psi^\top v^{n+1} + c^n \in \mathbb{C}^N$ , and thus the computational complexity is linear to  $N$ . As shown in [45], Algorithm 1 is very efficient and computationally inexpensive. Our experiments show that (9)-(TV) can be solved within 1 s for a  $512 \times 512$  image in Matlab on a regular PC.

#### Algorithm 1 Split Bregman [45] for (TV) Subproblem

```
Initialize  $v^0 = u$ , and  $w^0 = b^0 = 0$ ,  $z^0 = c^0 = 0$ .
while  $\|v^n - v^{n-1}\|_2 / \|u^n\|_2 > \text{tol}$  do
   $v^{n+1} = L^{-1} (D^\top (w^n - b^n) + \mu \Psi (z^n - c^n) + \eta u)$ 
   $w^{n+1} = \text{shrink2}(Dv^{n+1} + b^n, 1/\beta)$ 
   $z^{n+1} = \text{shrinkc}(\Psi^\top v^{n+1} + c^n, 1/\beta)$ 
   $b^{n+1} = b^n + (Dv^{n+1} - w^{n+1})$ 
   $c^{n+1} = c^n + (\Psi^\top v^{n+1} - z^{n+1})$ 
   $n \leftarrow n + 1$ 
end while
```

As can be seen, both of the nondifferentiable TV and  $\ell_1$  terms are in the (9)-(TV) subproblem, which have been now readily tackled by Algorithm 1. Moreover, despite that the process is highly nonlinear in (9)-(TV), the computation load i.e., denoising *one* image, of (9)-(TV) subproblem is relatively lighter than the (9)-(LS) subproblem which involves data and sensitivity maps from all  $J$  channels.

#### Algorithm for (LS) Subproblem

When conventional equal-distance Cartesian acquisition scheme is used, we can solve the (9)-(LS) subproblem through a pixel-by-pixel approach based on (7) in [55]. Initially, conventional SENSE was used for reconstruction. The initial sensitivity maps can be either from the self-calibration signal or from pre-scan. The outputs of the conventional SENSE, was used as inputs of (9)-(TV) subproblem.

However, for arbitrary acquisition scheme, the (9)-(LS) subproblem is not trivial to solve. Note that (9)-(LS) is a least squares problem with respect to  $u$ . The common treatment for least squares problems is the preconditioned conjugate gradient (PCG) method. However, recall that  $A$  comprises sensitivity maps which may change due to time/subject variability, it is not always efficient to use PCG since the preconditioner has to be estimated for each scan.

In our work, we adopt the optimal gradient (OG) method which requires no conditioning and can consistently generate fast approximations to the solution of least squares problem. First recall that the regular first-order gradient method uses the time marching scheme

$$\begin{aligned} \frac{(u^{n+1} - u^n)}{\delta} &= -(\alpha(u^n - v) + \lambda A^\top (Au^n - f)) \\ \text{i.e. } u^{n+1} &= (1 - \alpha\delta)u^n + \alpha\delta v - \lambda\delta A^\top (Au^n - f) \end{aligned} \quad (15)$$

where the right hand side of the top equation in (15) is the negative gradient of the objective function in (9)-(LS). Here  $\delta = \|A^\top A\|_2$ , i.e., the largest eigenvalue of  $A^\top A$ , is the maximum step size that guarantees convergence of  $\{u^n\}_n$ . In SENSE application we have  $\delta = 1$  because of (3) and that the magnitudes of sensitivity maps are scaled in  $[0, 1]$  in practice. Regular gradient method (15) converges very slowly. To improve the speed, OG substitutes  $u^n$  in (15) by  $d^n$ , a linear combination of previous *two* iterates  $u^n$  and  $u^{n-1}$ , as given in Algorithm 2. The combination coefficients depend on  $t_n \in \mathbb{R}$ , which is also updated in iterations. OG has almost the same computational complexity as regular first order gradient method but has much improved convergence speed than the latter [56], [57]. The OG algorithm for solving (9)-(LS) can be described as in Algorithm 2. As can be seen the main computation involved in each iteration is  $A$  and  $A^\top$ , each of which comprises  $J$  component-wise matrices product and fast Fourier transforms (FFT). Hence Algorithm 2 has very low computational cost and memory requirements. Moreover, as shown in [57], with almost similar computation complexity, the convergence rate of Algorithm 2 is  $\mathcal{O}(1/n^2)$  which is much higher than  $\mathcal{O}(1/n)$  of regular gradient descent method (15).

**Algorithm 2** Optimal Gradient [56], [57] for (LS) Subproblem

```
Initialize  $u^0 = v$ ,  $d^1 = u^0$ ,  $t_1 = 1$ .
while  $\|u^n - u^{n-1}\|_2 / \|u^n\|_2 > tol2$  do
 $u^{n+1} = (1 - \alpha)d^n + \alpha v - \lambda A^\top (Ad^n - f)$ 
 $t_{n+1} = (1 + \sqrt{1 + 4t_n^2}) / 2$ 
 $d^{n+1} = u^{n+1} + (u^{n+1} - u^n) \cdot (t_n - 1) / t_{n+1}$ 
 $n \leftarrow n + 1$ 
end while
```

In practical implementations, only  $5 \sim 10$  iterations were needed to terminate Algorithm 2 with  $tol = 10^{-3}$ , which is much less than  $50 \sim 100$  for regular gradient descent method (15) to reach the same tolerance.

*Algorithm for Sense Model With Sparsity Constraints*

As both of the (TV) and (LS) subproblems in (9) can be quickly solved by Algorithms 1 and 2, respectively, we propose Algorithm 3 to solve (5), the SENSE model with sparsity constraints. Here,  $F^k$  is the objective function value of (5) when the

iterate  $u^k$  is plugged in. There are several parameters involved in Algorithm 3: the weight parameters  $\mu, \lambda$ , the penalty parameter  $\alpha$  in the objective function of (8), and three tolerance parameters  $tol1$ ,  $tol2$ , and  $tol$  for (TV) and (LS) subproblems and the main Algorithm 3, respectively. We will present specific choices of these parameters in our experiments, and discuss their selections in SENSE applications in Section VI.

**Algorithm 3** Solving SENSE Model (8)

```
Initialize  $u^0 = A^\top f$ .
while  $\|F^k - F^{k-1}\|_2 / \|F^k\|_2 > tol$  do
  Compute  $v^{k+1}$  using Algorithm 1 with  $u = u^k$ 
  Compute  $u^{k+1}$  using Algorithm 2 with  $v = v^{k+1}$ 
   $v = v^{k+1}$ ,  $k \leftarrow k + 1$ 
end while
return  $u \leftarrow u^k$ 
```

*C. Convergence Analysis of the Proposed Algorithm*

A theoretical analysis on the effectiveness of the proposed algorithm 3 is provided below. The following theorems ensure the convergence of sequence  $\{(u^k, v^k)\}_{k=1}^\infty$  generated by Algorithm 3.

*Theorem III.1:* Sequence  $\{(u^k, v^k)\}_{k=1}^\infty$  generated by Algorithm 3 converges to a solution  $(u^*, v^*)$  to (8).

The proof of Theorem III.1 is in Appendix B. Furthermore, define  $M \triangleq I + (\lambda/\alpha)A^\top A$  and then it is clear that  $\rho(M^{-1}) \leq 1$ , where  $\rho(\cdot)$  denotes the spectral radius. We can show that  $\{(u^k, v^k)\}_{k=1}^\infty$  has  $q$ -linear convergence rate as follows.

*Theorem III.2:* Let  $\{(u^k, v^k)\}_{k=1}^\infty$  be the sequence generated by Algorithm 3 and  $(u^*, v^*)$  be the limit, then

$$\begin{aligned} \|v^{k+1} - v^*\|_2 &\leq \rho(M^{-1}) \|v^k - v^*\|_2 \\ \|u^{k+1} - u^*\|_M &\leq \sqrt{\rho(M^{-1})} \|u^k - u^*\|_M. \end{aligned} \quad (16)$$

The proof of Theorem III.2 is in Appendix C.

## IV. METHOD

Experiments were designed to test the significant improvements on efficiency of the proposed method while achieving state-of-the-art reconstruction quality.

Two data sets were acquired using commercially available eight-element head coils (Invivo Corp., Gainesville, FL). The first one is a high resolution Cartesian brain data set (data1) acquired on a 3.0 T Philips scanner (Philips, Best, The Netherlands) using T2-weighted turbo spin echo (T2 TSE) sequence. The acquisition parameters were FOV 205 mm<sup>2</sup>, matrix 512 × 500 × 8, TR 3000 ms, TE 85 ms, and the echo train length was 20. The other one is a radial brain data set (data2) acquired on a 1.5 T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were FOV 220 mm<sup>2</sup>, matrix 256 × 512 × 8, slice thickness 5 mm, TR 53.5 ms, TE 3.4 ms, and flip angle 75°. Both data sets were fully acquired, and then artificially down-sampled for reconstruction.

TABLE I  
TESTS NUMBER, DATA INFORMATION, UNDERSAMPLING MASK  
 $P$  (REFER TO FIG. 1) AND PARAMETER SETTINGS

Test	Image	Abbrev.	Size( $\times 8$ )	$P$	$(\mu, \lambda)$	$\alpha/\lambda$
1	Cart.Sag.	data1	$512 \times 500$	(a)	$(0, 10^2)$	$2^{-1}$
2	Cart.Sag.	data1	$512 \times 500$	(b)	$(0.1, 10^3)$	$2^2$
3	Rad.Axi.	data2	$256 \times 512$	(c)	$(0.1, 10^3)$	$2^2$

For data2, we applied GRAPPA operator gridding (GROG) technique [42] to shift the non-Cartesian radial data onto Cartesian grids such that fast Fourier transforms can be directly used in reconstruction. All data were normalized by some constants such that the intensities of reference images have a range in  $[0, 1]$ . In all of our experiments a reference image is set to the point-wise root of summed squares of the images obtained by full  $k$ -space data from all channels. Namely the reference image  $u_0 \triangleq (\sum_j |u_j|^2)^{1/2}$  where  $u_j$  is the Fourier transform of the fully acquired data from the  $j$ -th channel. The parameter setting and further discussion are based on the normalized data.

For comparison, a nonlinear conjugate gradient (NLCG) based solver [43] for model (5) was also implemented and consistently used for comparisons in all experiments. As NLCG requires the objective function in (5) to be differentiable, we used the following common approximations:

$$\|u\|_{TV, \epsilon} = \sum_i \sqrt{\|D_i u\|_2^2 + \epsilon}, \quad \|x\|_{1, \epsilon} = \frac{x^\top x}{\sqrt{x^\top x + \epsilon}} \quad (17)$$

with  $\epsilon = 10^{-15}$  for the TV and  $\ell_1$  terms in (5). The line search parameters  $(a, b)$  were set to  $(0.01, 0.6)$  as used by default in [43].

In each experiment, the parameters  $(\mu, \lambda)$  are set the same for NLCG and the proposed method. In addition, the computation processes for both methods are terminated when the relative change in objective function [by plugging the iterate  $u^k$  into the original model (5)] is less than  $tol = 10^{-4}$ . Besides  $tol$ , the stopping criterions  $tol1$  and  $tol2$  for the (TV) and (LS) subproblems in the proposed algorithm are both set to  $10^{-3}$ . Note that such tolerances are in general small enough, such that more tight stopping criterions will not lead to significantly improved accuracy in the final reconstruction but require more iterations. The data description and parameter settings in our tests are summarized in Table I. The sparsifying transform  $\Psi$  in (5) was set to the 2D Haar wavelet transform. The sensitivity maps for data1 were estimated using the central  $32 \times 32$   $k$ -space data from all channels, which is a subset of the undersampled data. In radial sampling trajectories as in test 3, the disk of radius  $n/\pi$  (where  $n$  is the number of sampled radial lines) at the center of  $k$ -space is fully sampled and hence can be used to estimate sensitivity maps. These sensitivity maps are precomputed and then fixed during the computation.

To compare the proposed method with  $k$ -space methods, we also implemented GRAPPA [2], L1 SPIR-iT [58], and self-calibrated radial GRAPPA [59], which proved to be efficient for (a) variable density, (b) pseudo random and (c) radial  $k$ -space trajectories in Fig. 1, respectively. The convolution kernel size for L1 SPIR-iT and GRAPPA were  $3 \times 3$  and  $4 \times 5$ , respectively. The convolution kernels for self-calibrated radial GRAPPA were defined by Fig. 3 in the original paper [59]. All

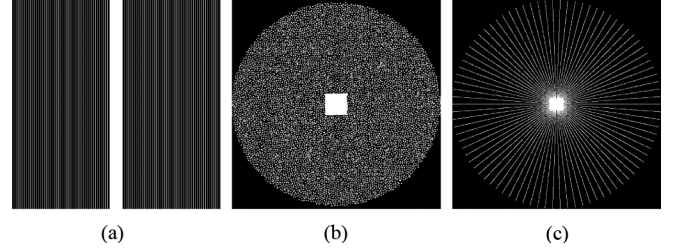


Fig. 1.  $k$ -space masks  $P$  used for reconstructions in tests 1, 2, and 3, respectively (a) Cartesian mask with net reduction factor 4 (b) Pseudo random mask with reduction factor 4 (c) Radial mask with 43 (out of 256) projections, i.e., reduction factor 6.

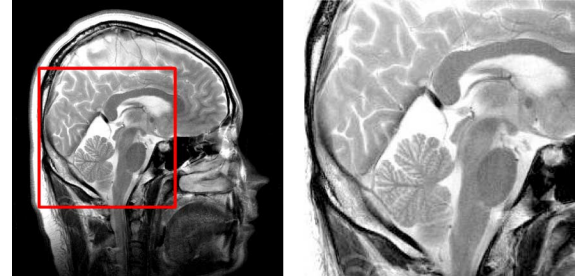


Fig. 2. data1: sagittal brain image ( $512 \times 500$ ). Left: Reference image. Right: Zoomed-in of the box in the reference image.

the convolution kernels were self-calibrated for these methods. The convolution was processed in  $k$ -space.

All algorithms were implemented in the Matlab programming environment (Version R2009a, MathWorks, Natick, MA) except few componentwise computations related to complex values in TV subproblem split Bregman solver RecPF were coded in C++. The experiments were performed on a Dell Optiplex desktop with Intel Quad Core 3.0-GHz processors (only 1 core was used in computation), 3 GB of memory and Windows operating system.

The reconstruction results were evaluated qualitatively by error maps (absolute difference of reconstruction  $u$  to the reference image  $u_0$ ) and zoomed-in regions, as shown in Figs. 3, 5, and 8. They are all at the same scale level and then brightened 5 and 2 times, respectively, to help visual justifications. To quantitatively evaluate the efficiency and accuracy of the tested methods, we also show the plots of total reconstruction time (in seconds) and relative error to the reference image ( $\|u - u_0\|_2 / \|u_0\|_2$ ) versus different reduction factors for each test.

## V. RESULTS

Figs. 3 and 4 show the results of test 1 using Cartesian trajectory [Fig. 1(a)], which is regularly used for self-calibrated PPI. Cartesian data1 (Fig. 2) was reconstructed with nominal reduction factors of 2, 3, 4, and 5 by GRAPPA, NLCG and the proposed method. In addition, the central  $32$   $k$ -space lines are also fully acquired for kernel calculation in GRAPPA and sensitivity maps estimation in NLCG and the proposed method. Then the corresponding net reduction factors are 1.9, 2.6, 3.4, and 4. For net reduction factor 4, the error maps and the zoomed-in regions of the reconstructed images are displayed on the top and bottom rows in Fig. 3, respectively. Visually one can see that



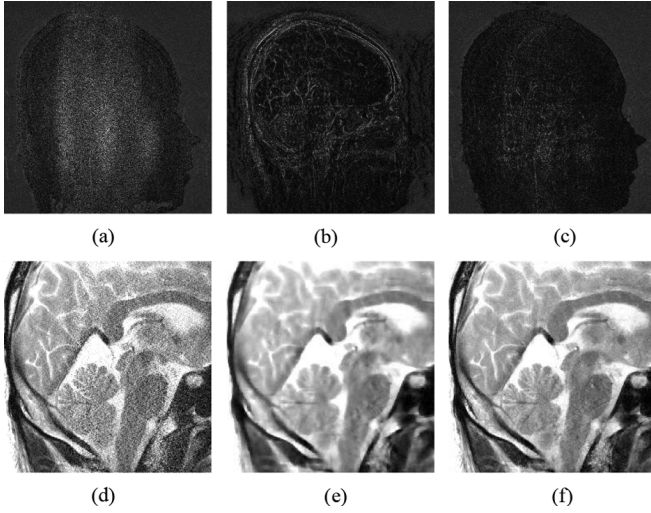


Fig. 3. Error maps (top row) and zoomed-ins (bottom row) of reconstructions in test1 with nominal reduction factor 5. Left column: Conventional GRAPPA. Middle column: NLCG. Right column: Proposed method (a) GRAPPA (b) NLCG (c) Proposed (d) GRAPPA (e) NLCG (f) Proposed.

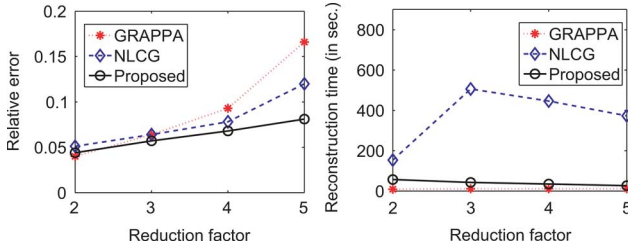


Fig. 4. Comparison of relative error and reconstruction time (sec.) versus (nominal) reduction factors of reconstructions for data1 (Fig. 7) and Cartesian trajectories using different methods.

the reconstruction by the proposed method [Fig. 3(c) and (f)] has the lowest error level and highest spatial resolution than those by GRAPPA [Fig. 3(a) and (d)] and NLCG [Fig. 3(b) and (e)]. This suggests that the proposed method has the ability to preserve spatial resolution and improve reconstruction SNR. Fig. 4 shows the comparisons on relative errors and reconstruction time by these three methods with different (nominal) reduction factors 2, 3, 4, and 5. The proposed algorithm is up to 10 times faster than NLCG with similar or lower relative errors. Although the reconstruction time of the proposed algorithm is longer than GRAPPA, the relative errors of the results by the proposed method are significantly lower than GRAPPA especially when the (nominal) reduction factor was 4 and 5.

Figs. 5 and 6 show the results of test 2 with pseudo random trajectories. Fig. 1(b) shows the acquisition trajectory, which is potentially useful for high-resolution 3D imaging [58]. Here we used it to simulate the random trajectory for 2D imaging. Cartesian data1 (Fig. 2) is reconstructed with net reduction factors of 2, 3, 4, 6, and 8 by L1 SPIR-iT [58], NLCG, and the proposed method. For reduction factor 4, the error maps (top row) and the zoomed-ins (bottom row) are shown in Fig. 5, which indicates that both L1 SPIR-iT and the proposed method provide good reconstructions with the latter one showing slightly better accuracy. Further comparisons of these three algorithms are demonstrated in Fig. 6, where the relative errors and reconstruction time versus reduction factors are presented. The left

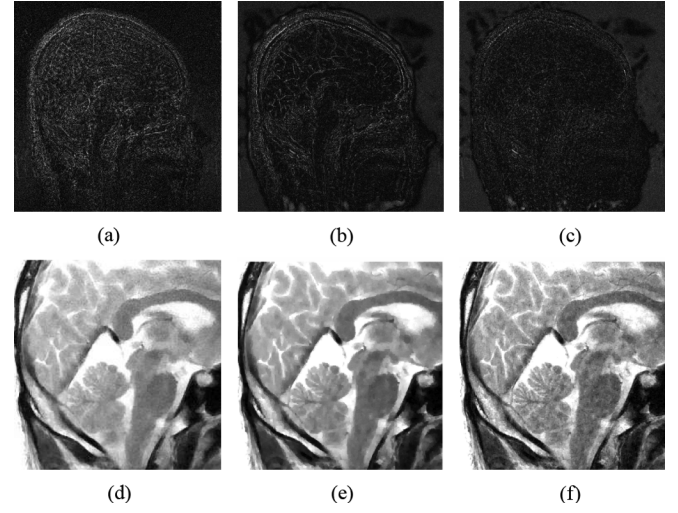


Fig. 5. Error maps (top row) and zoomed-ins (bottom row) of the reconstructions in test2 with reduction factor 4. Left column: L1 SPIR-iT. Middle column: NLCG method applied to SENSE. Right column: Proposed method (a) L1 SPIR-iT (b) NLCG (c) Proposed (d) L1 SPIR-iT (e) NLCG (f) Proposed.

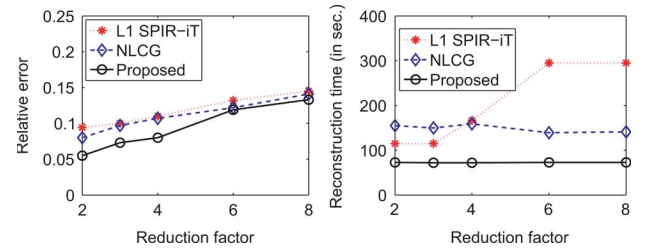


Fig. 6. Comparison of relative error and reconstruction time (in seconds) versus reduction factors for data1 (Fig. 7) and pseudo random trajectories using different methods.

shows that the proposed method consistently produces similar or even lower relative error than the other two methods. In addition, the reconstruction time used by the proposed method is much less than those by L1 SPIR-iT and NLCG, as shown on the right of Fig. 6. These two plots indicate that to generate a similar or even better reconstruction result, the proposed method cut up to 80% off the reconstruction time required by L1 SPIR-iT.

Figs. 8 and 9 show the results of test 3 with radial trajectory (Fig. 1) by self-calibrated radial GRAPPA [59], NLCG and the proposed method. Radial data2 (Fig. 7) was reconstructed with 32, 43, 64, 84, and 128 radial projections corresponding to reduction factors 8, 6, 4, 3, and 2, respectively. For 43 projections (reduction factor = 6), the error maps (top row) and zoomed-in region of the reconstructed images (bottom row) in Fig. 8 demonstrate a much better reconstruction by the proposed method than those by radial GRAPPA and NLCG in terms of relative error and SNR. As can be seen, radial GRAPPA is very fast but the reconstructions have severe aliased artifacts. On the other hand, the image reconstructed by the proposed method has both well preserved high resolution as well as much lower relative error. In the plot of reconstruction time versus number of projections (right of Fig. 9), the proposed method shows to be less effective than radial GRAPPA when the reduction factor is low (but still acceptable and is much more efficient than NLCG in all cases). As the number of projections reduces (i.e., reduction



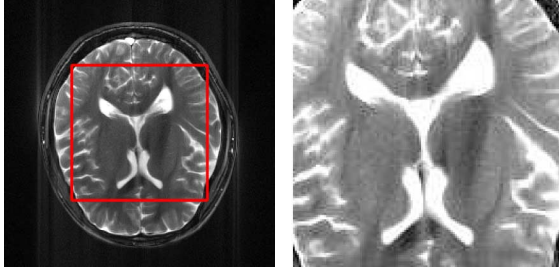


Fig. 7. Data2: radial brain image ( $256 \times 512 \times 8$ ). Left: Reference image. Right: Zoomed-in of the box in the reference image.

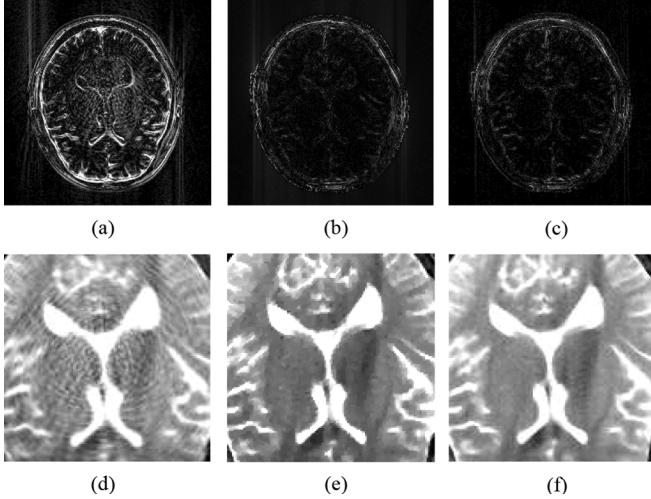


Fig. 8. Error maps (top row) and zoomed-ins (bottom row) of the reconstructions in test3 with 43 radial projections. Left column Radial GRAPPA. Middle column NLCG. Right column Proposed method (a) Radial GRAPPA (b) NLCG (c) Proposed (d) Radial GRAPPA (e) NLCG (f) Proposed.

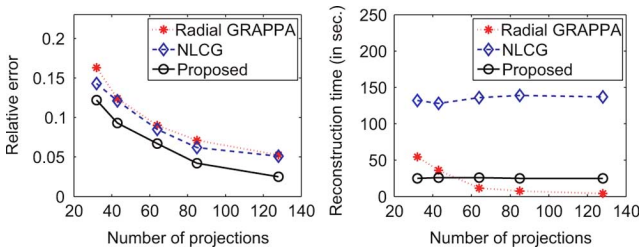


Fig. 9. Comparison of relative error and reconstruction time (seconds) versus numbers of projections for data2 (Fig. 7) with radial trajectories [Fig. 1(c)] using different methods.

factor increases), radial GRAPPA requires longer time for reconstruction but the computational cost for the proposed method remains almost the same. When the number of projections is less than 64, the proposed method turned out to be the most efficient one among the three algorithms tested in this experiment.

In all the three tests, the conventionally adopted optimization method NLCG was used for comparison. Although NLCG and the proposed algorithm are both set to solve the CS-PPI model (5), their practical performances are quite different. To test the efficiency of these two methods, we record the relative errors and objective function values [by plugging the iterates into the original objective function in (5)] in each iteration during their reconstruction processes in test 2 with reduction factor 4, and plot them (in logarithm) versus CPU time as show in Fig. 10.

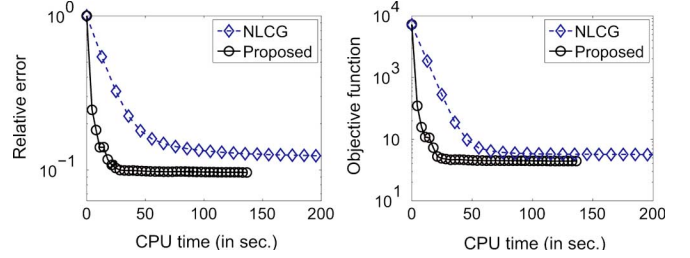


Fig. 10. Comparison of relative errors (left) and objective functions (right) in logarithm versus CPU time by NLCG and the proposed method during the reconstruction processes.

From Fig. 10, we can see that the relative error and objective function value by the proposed method decay faster and remain lower than those by NLCG. These indicate that the proposed method using alternating minimizations has better accuracy and higher efficiency than NLCG in solving (5).

In our understanding, there are two main reasons why the proposed method is more efficient and robust than NLCG. Firstly, NLCG suffers the combination of two issues in CS-PPI application greatly: the highly nonlinearity of the nonsmooth  $TV/\ell_1$  terms and the severe ill-conditioning of the matrix  $A$  in (5). Due to these two issues, NLCG has much difficulty in finding efficient gradient descent direction and also the time-consuming line searches (which requires numbers of computations of  $A$ ) consume considerably long time. On the contrary, the proposed method split the problem into two well regularized minimizations (TV) and (LS). The (TV) subproblem can be stably solved by split Bregman method by using nonlinear (but very fast) shrinkages provided that the matrix multiplied to  $v$  is identity in (9). The (LS) subproblem is smooth and can be efficiently solved by the optimal gradient method with improved conditioning. Thus the computation complexity of the original problem (5) is significantly reduced by using the proposed method and hence higher efficiency can be achieved. Secondly, NLCG appears to be very sensitive to the weight parameter  $\lambda$ : either  $\lambda$  large or small makes NLCG generate unsatisfactory results as the data fitting or the regularization becomes too significant. On the contrary,  $\lambda$  appears only in the (LS) subproblem of the proposed algorithm so it merely affects the Fourier coefficients in  $\mathcal{FS}_j(u)$  corresponding to positions where data  $f_j$  is sampled. If  $\lambda$  becomes larger, the remaining entries of  $\mathcal{FS}_j(u)$  will still update with  $v$  independent of  $\lambda$ . This splitting makes the proposed method very stable even with large  $\lambda$  and allows it to faithfully retain the effect of  $TV/\ell_1$  regularization.

## VI. FURTHER DISCUSSION ON PARAMETER SELECTION

In this section, we make further discussion on the parameter selection for the proposed algorithm. For demonstration purpose, we showed the results corresponding to the tests on data 1 with pseudo random sampling pattern [Fig. 1(b)] and reduction factor 4.

### A. Balancing the Three Terms in Model (5)

In model (5) we need two parameters  $\mu$  and  $\lambda$  to balance the three terms: the TV and  $\ell_1$  regularization terms and the data fi-

TABLE II  
RELATIVE ERRORS OF THE RECONSTRUCTIONS BY USING  
ONLY THE TV TERM AND THE  $\ell_1$  TERM IN (5)

$\lambda$	$10^2$	$10^3$	$10^4$	$10^5$
with TV only	10.67%	10.50%	12.41%	13.21%
with $\ell_1$ only	12.17%	11.55%	13.06%	13.32%

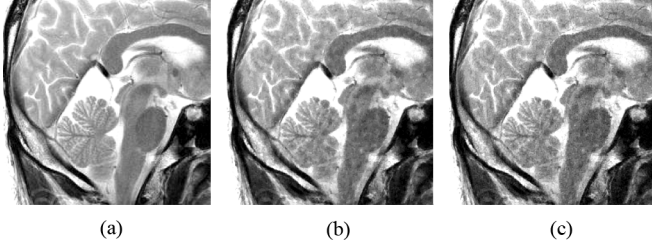


Fig. 11. Comparison of reconstructions using only TV or  $\ell_1$  term in (5), reduction factor is 4 and  $\lambda = 10^3$  (a) Reference (b) Reconstruction with TV regularization only (c) Reconstruction with  $\ell_1$  regularization only.

delity term. As will be shown soon later, the TV regularization term exhibits stronger image recovering effects than  $\ell_1$ . Therefore, we let TV term always present in the reconstruction model (5), and use the  $\mu$  and  $\lambda$  for the  $\ell_1$  term and data fidelity term respectively as weight ratios against the TV term in our experiments.

To determine the values of  $\mu$  and  $\lambda$ , we need to explore the behavior and performance of TV and  $\ell_1$  regularization terms. First, we compute the solutions to (5) with the other term turned off. Namely, we set  $\mu = 0$  and try different values of  $\lambda$  in (5) to see the effects brought by the TV term, and then turn off the TV term, set  $\mu = 1$  and try the same values of  $\lambda$  to see the effects of the  $\ell_1$  term. The parameter  $\alpha$  is set to  $2^2 \times \lambda$  in these cases. The resulting relative errors of the reconstructions are presented in Table II.

From Table II, we first observe that the model (5) can be stably solved by the proposed method with  $\lambda$  lying in a large range  $[10^2, 10^5]$ . This is demonstrated by the fact that the relative error does not change much for various  $\lambda$ . Among the values we tested in this experiment,  $\lambda$  between  $[10^2, 10^3]$  appears to be optimal as suggested by the lowest relative errors. Furthermore, we can see that the TV term appears to be more effective than the  $\ell_1$  term in reconstruction because the former consistently produces images of lower relative errors, as shown in Table II. We also show the reconstructed images when  $\lambda = 10^3$  in Fig. 11. In this case, the relative errors by using only the TV term and the  $\ell_1$  term are 10.50% and 11.55%, respectively. When compared to the reference image Fig. 11(a), the TV reconstruction [Fig. 11(b)] exhibits lower level of noises/artifacts as well as better preserved edges than  $\ell_1$  reconstruction [Fig. 11(c)].

### B. An Empirical Study of the Penalty Parameter $\alpha$

When we try to approximate the constrained problem (7) by the unconstrained minimization (8), it is usually intuitive to set the penalty parameter  $\alpha$  very large. However, our experimental results in SENSE applications show that large penalty parameter  $\alpha$  does not necessarily improve the image quality, but slows down the convergence (as shown in Theorem III.2 that  $\rho(M^{-1})$  becomes closer to 1 as  $\alpha$  gets larger). Table III lists the relative

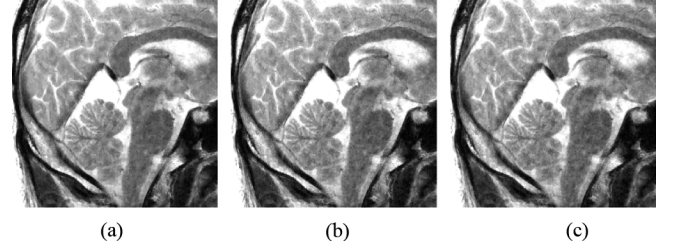


Fig. 12. Reconstructions by the proposed method in test 2 with different values of penalty parameter  $\alpha$ , reduction factor is 4 and  $\lambda = 10^3$  (a)  $\alpha = 2^{-3}$ , relative error 10.54%, (b)  $\alpha = 2^2$ , relative error 10.87%, (c)  $\alpha = 2^5$ , relative error 11.36%.

TABLE III  
RELATIVE ERRORS (RELERR) AND COMPUTATIONAL TIMES (TIME)  
IN SECONDS BY ALGORITHM 3 WITH DIFFERENT VALUES OF  $\alpha$

$\alpha/\lambda$	$2^{-3}$	$2^{-1}$	$2^0$	$2^2$	$2^5$
RelErr	10.54%	10.46%	10.47%	10.87%	11.36%
Time	30.9	35.7	37.0	39.1	65.7

errors and computational times of Algorithm 3 with different values of  $\alpha$  in test 2 with reduction factor 4 and  $\lambda = 10^3$ . The reconstruction results with  $\alpha/\lambda = 2^{-3}, 2^2, 2^5$  are also plotted in Fig. 12. The parameter  $\mu$  is set to 0.1 in these tests. From Table III, we can see those  $\alpha/\lambda \in [2^{-2}, 2^4]$  leads to fast computation as well as low reconstruction error. Tests on other data also suggest that satisfactory reconstructions in CS-PPI model (5) can be obtained when  $\alpha$  is chosen from this fairly large range. On the contrary, larger  $\alpha/\lambda$  leads to unnecessarily long computational time and sometimes even slightly larger error. One reason is that in this case the weight on the data consistency term (8) is relatively too small and hence the reconstruction error increases a bit, despite that the solution  $u$  and  $v$  are closer now than a smaller  $\alpha$  was used. Therefore, we suggest a moderate value of  $\alpha$  between  $[2^{-2}, 2^4] \times \lambda$  for optimal performance of (8).

## VII. CONCLUSION

This paper presents a fast numerical algorithm for SENSE model with sparsity constraints in CS-PPI with arbitrary  $k$ -space trajectories. The proposed algorithm adopts variable splitting, the quadratic penalty technique and an optimal gradient method. A tailored derivation of the algorithm is presented, and a comprehensive analysis showing the linear convergence rate of the algorithm is also provided. Numerical tests indicate fast, stable and accurate reconstructions obtained by the proposed algorithm. Furthermore, the choice of parameters in CS-PPI is discussed based on theoretical interpretation and experimental tests.

## APPENDIX A

### SOLVERS OF (TV)-(LS) IN (9) AND THEIR PROPERTIES

Let  $\mathcal{T}$  and  $\mathcal{L}$  denote the solvers of (TV) and (LS) subproblems in (9), respectively. Namely,  $\mathcal{T}$  and  $\mathcal{L}$  are defined by

$$\begin{aligned}\mathcal{T}(u) &\triangleq \arg \min_{v \in \mathbb{C}^N} \frac{1}{\alpha} \|v\|_{TV} + \frac{\mu}{\alpha} \|\Psi^T v\|_1 + \frac{1}{2} \|v - u\|_2^2 \\ \mathcal{L}(v) &\triangleq \arg \min_{u \in \mathbb{C}^N} \frac{\lambda}{2\alpha} \|Au - f\|_2^2 + \frac{1}{2} \|u - v\|_2^2.\end{aligned}\quad (18)$$

$\mathcal{T}, \mathcal{L} : \mathbb{C}^N \rightarrow \mathbb{C}^N$  are proximity operators as both of the objective functions above are formed by a continuous and convex function plus a proximal term  $\|u - v\|_2^2/2$ . A proximity operator  $\mathcal{P}$  is firmly nonexpansive and satisfies [61]

$$\|\mathcal{P}(x) - \mathcal{P}(y)\|_2 \leq \|x - y\|_2, \quad \forall x, y \in \mathbb{C}^N, \\ \text{and "=" holds} \iff \mathcal{P}(x) - \mathcal{P}(y) = x - y. \quad (19)$$

Hence, Algorithm 3 is simply alternating  $v^{k+1} = \mathcal{T}(u^k)$  and  $u^{k+1} = \mathcal{L}(v^{k+1})$  in each iteration.

#### APPENDIX B

##### PROOF OF THEOREM III.1

As (9) is convex and coercive and  $\mathcal{T}, \mathcal{L}$  are both continuous, it is suffice to show that  $\{u^k\}_k$  converges to a fixed point of  $\mathcal{L} \circ \mathcal{T}$ .

Let  $\hat{u}$  be any fixed point of  $\mathcal{L} \circ \mathcal{T}$ . Since  $\mathcal{L} \circ \mathcal{T}$  also satisfies (19), and there is

$$\|u^{k+1} - \hat{u}\|_2 = \|\mathcal{L} \circ \mathcal{T}(u^k) - \mathcal{L} \circ \mathcal{T}(\hat{u})\|_2 \leq \|u^k - \hat{u}\|_2.$$

This shows that the sequence of scalars  $\{\|u^k - \hat{u}\|_2\}_k$  is bounded, monotonically decreasing and hence is convergent. Boundedness also implies a convergent subsequence  $\{u^{k_j}\}_j$  of  $\{u^k\}_k$ , such that  $\lim_j u^{k_j} = u^*$  for some  $u^*$  and

$$\lim_{k \rightarrow \infty} \|u^k - \hat{u}\|_2 = \lim_{j \rightarrow \infty} \|u^{k_j} - \hat{u}\|_2 = \|u^* - \hat{u}\|_2. \quad (20)$$

Since  $\mathcal{L} \circ \mathcal{T}$  is continuous, we know  $\{u^{k_j+1}\}_j$  is also a convergent subsequence of  $\{u^k\}_k$  with limit  $\mathcal{L} \circ \mathcal{T}(u^*)$ . Hence

$$\lim_{k \rightarrow \infty} \|u^k - \hat{u}\|_2 = \lim_{j \rightarrow \infty} \|u^{k_j+1} - \hat{u}\|_2 = \|\mathcal{L} \circ \mathcal{T}(u^*) - \hat{u}\|_2.$$

Combining this with (20) gives  $\|\mathcal{L} \circ \mathcal{T}(u^*) - \hat{u}\|_2 = \|u^* - \hat{u}\|_2$ . As  $\hat{u}$  is a fixed point, and  $\mathcal{L} \circ \mathcal{T}$  satisfies (19), we have  $u^* = \mathcal{L} \circ \mathcal{T}(u^*)$ , i.e.,  $u^*$  is a fixed point of  $\mathcal{L} \circ \mathcal{T}$ .

Note that (20) holds for any fixed point of  $\mathcal{L} \circ \mathcal{T}$ , replacing  $\hat{u}$  by  $u^*$  gives  $\lim_k \|u^k - u^*\|_2 = 0$ . That is,  $\{u^k\}_k$  converges to  $u^*$ , a fixed point of  $\mathcal{L} \circ \mathcal{T}$ , which proves the theorem.

#### APPENDIX C

##### PROOF OF THEOREM III.2

As Algorithm 3 alternates  $v^{k+1} = \mathcal{T}(u^k)$  and  $u^{k+1} = \mathcal{L}(v^{k+1})$  in each iteration, there is

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

Note that the derivative of the objective function in (LS) subproblem of (9) at  $\mathcal{L}(v)$  is 0, i.e.,

$$\mathcal{L}(v) - v + \left(\frac{\lambda}{\alpha}\right) \cdot A^\top (A\mathcal{L}(v) - f) = 0$$

since  $\mathcal{L}(v)$  is the minimizer. Define  $M = I + (\lambda/\alpha)A^\top A$  then

$$\mathcal{L}(v) - \mathcal{L}(\hat{v}) = M^{-1}(v - \hat{v}), \quad \forall v, \hat{v} \in \mathbb{C}^N. \quad (21)$$

Hence we have

$$\|v^{k+1} - v^*\|_2 \leq \|u^k - u^*\|_2 = \|\mathcal{L}(v^k) - \mathcal{L}(v^*)\|_2 \\ \leq \rho(M^{-1})\|v^k - v^*\|_2.$$

Moreover, by  $u^{k+1} - u^* = \mathcal{L}(v^{k+1}) - \mathcal{L}(v^*)$  and (21), we have

$$\|u^{k+1} - u^*\|_M^2 = \langle u^k - u^*, M(u^k - u^*) \rangle \\ \leq \rho(M^{-1})\|u^k - u^*\|_2^2$$

where the last inequality is because that  $v^{k+1} = \mathcal{T}(u^k)$ ,  $v^* = \mathcal{T}(u^*)$  and  $\mathcal{T}$  satisfies (19). Also note that  $\|u^k - u^*\|_2^2 \leq \|u^k - u^*\|_M^2$ . Combining the two inequalities above, we obtain

$$\|u^{k+1} - u^*\|_M \leq \sqrt{\rho(M^{-1})}\|u^k - u^*\|_M.$$

#### REFERENCES

- [1] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, "Sense: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, pp. 952–962, 1999.
- [2] 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.*, vol. 47, pp. 1202–1210, 2002.
- [3] R. M. Heidemann, Ö. Özarlak, P. M. Parizel, J. Michiels, B. Kiefer, V. Jellus, M. Müller, F. Breuer, M. Blaimer, M. A. Griswold, and P. M. Jakob, "A brief review of parallel magnetic resonance imaging," *Eur. Radiol.*, vol. 13, pp. 2323–2337, 2003.
- [4] F. Lin, T. Huang, N. Chen, F. Wang, and S. M. Stufflebeam, "Functional mri using regularized parallel imaging acquisition," *Magn. Reson. Med.*, vol. 54, pp. 343–353, 2005.
- [5] T. Niendorf, C. Hardy, R. Giaquinto, P. Gross, H. Cline, Y. Zhu, G. Kenwood, S. Cohen, A. Grant, S. Joshi, N. Rofsky, and D. Sodickson, "Toward single breath-hold whole-heart coverage coronary mra using highly accelerated parallel imaging with a 32-channel MR system," *Magn. Reson. Med.*, vol. 56, pp. 167–176, 2006.
- [6] P. Lai, F. Huang, A. Larson, and D. Li, "Fast 4d coronary MR angiography with k-t grappa," *J. Magn. Reson. Imaging*, vol. 27, pp. 659–665, 2008.
- [7] C. B. Ahn, J. H. Kim, and Z. H. Cho, "High-speed spiral-scan echo planar NMR imaging," *IEEE Trans. Med. Imag.*, vol. 5, no. 1, pp. 2–7, Mar. 1986.
- [8] C. Meyer, B. Hu, D. Nishimura, and A. Macovski, "Fast spiral coronary artery imaging," *Magn. Reson. Med.*, vol. 28, pp. 202–213, 1992.
- [9] G. Glover and J. Pauly, "Projection reconstruction techniques for reduction of motion effects in MRI," *Magn. Reson. Med.*, vol. 28, pp. 275–289, 1992.
- [10] J. Pipe, "Motion correction with propeller MRI: Application to head motion and free-breathing cardiac imaging," *Magn. Reson. Med.*, vol. 42, pp. 963–969, 1999.
- [11] D. Noll, "Multishot rosette trajectories for spectrally selective MR imaging," *IEEE Trans. Med. Imag.*, vol. 16, no. 4, pp. 372–377, Aug. 1997.
- [12] K. Scheffler and J. Hennig, "Frequency resolved single-shot MR imaging using stochastic k-space trajectories," *Magn. Reson. Med.*, vol. 35, pp. 569–576, 1996.
- [13] D. Peters, F. Korosec, T. Grist, W. Block, J. Holden, K. Vigen, and C. Mistretta, "Undersampled projection reconstruction applied to MR angiography," *Magn. Reson. Med.*, vol. 43, pp. 91–101, 2000.
- [14] C. Mistretta, O. Wieben, J. Velikina, W. Block, J. Perry, Y. Wu, and K. Johnson, "Highly constrained backprojection for time-resolved MRI," *Magn. Reson. Med.*, vol. 55, pp. 30–40, 2006.
- [15] J. Rahmer, P. Bornert, J. Groen, and C. Bos, "Three-dimensional radial ultrashort echo-time imaging with t2 adapted sampling," *Magn. Reson. Med.*, vol. 55, pp. 1075–1082, 2006.
- [16] K. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, pp. 638–651, 2001.
- [17] M. Griswold, R. Heidemann, and P. Jakob, "Direct parallel imaging reconstruction of radially sampled data using grappa with relative shifts," *Proc. Int. Soc. Mag. Reson. Med.*, p. 2349, 2003.
- [18] R. Heidemann, M. A. Griswold, and P. Jakob, "Fast parallel image reconstruction with non-Cartesian trajectories," *Proc. Int. Soc. Mag. Reson. Med.*, p. 2347, 2003.

- [19] U. Katscher, "Underdetermined variable density sense," *Proc. Int. Soc. Mag. Reson. Med.*, p. 2342, 2003.
- [20] Y. Qian, Z. Zhang, V. Stenger, and Y. Wang, "Self-calibrated spiral sense," *Magn. Reson. Med.*, vol. 52, pp. 688–692, 2004.
- [21] P. Qu, K. Zhong, B. Zhang, J. Wang, and G.-X. Shen, "Convergence behavior of iterative sense reconstruction with non-cartesian trajectories," *Magn. Reson. Med.*, vol. 54, pp. 1040–1045, 2005.
- [22] 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.*, vol. 54, pp. 1–8, 2005.
- [23] E. Yeh, C. McKenzie, M. Ohliger, and D. Sodickson, "Parallel magnetic resonance imaging with adaptive radius in *k*-space (pars): Constrained image reconstruction using *k*-space locality in radiofrequency coil encoded data," *Magn. Reson. Med.*, vol. 53, pp. 1383–1392, 2005.
- [24] K. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils: Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007.
- [25] C. Liu, R. Bammer, and M. Moseley, "Parallel imaging reconstruction for arbitrary trajectories using *k*-space sparse matrices (kspa)," *Magn. Reson. Med.*, vol. 58, pp. 1071–1081, 2007.
- [26] L. Ying, B. Liu, M. Steckner, G. Wu, and S.-J. L. M. Wu, "A statistical approach to sense regularization with arbitrary *k*-space trajectories," *Magn. Reson. Med.*, vol. 60, pp. 414–421, 2008.
- [27] N. Seiberlich, F. Breuer, P. Ehse, H. Moriguchi, M. Blaimer, P. Jakob, and M. Griswold, "Using the grappa operator and the generalized sampling theorem to reconstruct undersampled non-Cartesian data," *Magn. Reson. Med.*, vol. 61, pp. 705–715, 2009.
- [28] T. Grotz, B. Zahneisen, A. Ella, M. Zaitsev, and J. Hennig, "Fast functional brain imaging using constrained reconstruction based on regularization using arbitrary projections," *Magn. Reson. Med.*, vol. 62, pp. 394–405, 2009.
- [29] C. Liu, M. Moseley, and R. Bammer, "Simultaneous phase correction and sense reconstruction for navigated multishot," *Magn. Reson. Med.*, vol. 54, pp. 1412–1422, 2005.
- [30] C. Liu, R. Bammer, and M. Moseley, "High resolution multi-shot sense DTI using self-navigated interleaved spirals (snails)," *Proc. Intl. Soc. Mag. Reson. Med.*, p. 10, 2005.
- [31] D. Mayer, D. Kim, D. Spielman, and R. Bammer, "Fast parallel spiral chemical shift imaging at 3t using iterative sense reconstruction," *Magn. Reson. Med.*, vol. 59, pp. 891–897, 2008.
- [32] X. Bi, J. Park, A. Larson, Q. Zhang, O. Simonetti, and D. Li, "Contrast-enhanced 4d radial coronary artery imaging at 3.0 T within a single breath-hold," *Magn. Reson. Med.*, vol. 54, no. 2, pp. 470–475, 2005.
- [33] J. Park, A. Larson, Q. Zhang, O. Simonetti, and D. Li, "4d radial coronary artery imaging within a single breath-hold: Cine angiography with phase-sensitive fat suppression (CAPS)," *Magn. Reson. Med.*, vol. 54, no. 4, pp. 833–840, 2005.
- [34] D. Blezek, A. Arunachalam, D. Whitt, K. Fish, and I. Hancu, "Sensitivity mapping for parallel imaging of hyperpolarized 13 c compounds," *Proc. Int. Soc. Mag. Reson. Med.*, p. 1751, 2008.
- [35] A. Arunachalam, A. Lu, E. Brodsky, S. Fain, and W. Block, "Parallel imaging with 3d PR (vipr)," *Proc. Int. Soc. Mag. Reson. Med.*, p. 2246, 2004.
- [36] K. Heberlein, Y. Kadah, and X. Hu, "Segmented spiral parallel imaging using grappa," *Proc. Int. Soc. Mag. Reson. Med.*, p. 328, 2004.
- [37] R. Heidemann, M. Griswold, G. Kruger, S. Kannengiesser, B. Kiefer, and P. Jakob, "Fast parallel image reconstructions for spiral trajectories," in *Proc. Int. Workshop Parallel MRI*, 2004, p. 27.
- [38] A. Arunachalam, A. Lu, E. Brodsky, and W. Block, "Grappa for the 3d radial trajectory(vipr)," *Proc. Int. Soc. Mag. Reson. Med.*, p. 2674, 2005.
- [39] H. Eggers and P. Boesiger, "Gridding- and convolution-based iterative reconstruction with variable *k*-space resolution for sensitivity-encoded non-Cartesian imaging," *Proc. Intl. Soc. Mag. Reson. Med.*, p. 2346, 2003.
- [40] A. Samsonov and C. Johnson, "Non-cartesian pocsense," *Proc. Intl. Soc. Mag. Reson. Med.*, p. 2648, 2004.
- [41] A. Larson, R. White, G. Laub, E. McVeigh, D. Li, and O. Simonetti, "Self-gated cardiac cine MRI," *Magn. Reson. Med.*, vol. 51, pp. 93–102, 2004.
- [42] N. Seiberlich, F. Breuer, M. Blaimer, K. Barkauskas, P. Jakob, and M. Griswold, "Non-cartesian data reconstruction using grappa operator gridding (grog)," *Magn. Reson. Med.*, vol. 58, pp. 1257–1265, 2007.
- [43] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [44] J. Yang, Y. Zhang, and W. Yin, "A fast TVL1-L2 minimization algorithm for signal reconstruction from partial Fourier data CAAM Rice Univ., Tech. Rep. 08-29, 2008.
- [45] T. Goldstein and S. Osher, "The split bregman method for l1 regularized problems," *SIAM J. Imag. Sci.*, vol. 2, pp. 323–343, 2009.
- [46] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration CAM UCLA, Tech. Rep. 08-34, 2008, Tech. Rep. 08-34.
- [47] Y. Huang, M. K. Ng, and Y.-W. Wen, "A fast total variation minimization method for image restoration," *Multiscale Model. Simul.*, vol. 7, no. 2, pp. 774–795, 2008.
- [48] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, pp. 89–97, 2004.
- [49] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.
- [50] M. Zhu, S. Wright, and T. Chan, "Duality-based algorithms for total-variation-regularized image restoration," *Comput. Optim. Appl.*, 2008.
- [51] D. Bertsekas, *Parallel and Distributed Computation*. Upper Saddle River, NJ: Prentice Hall, 1989.
- [52] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, 1992.
- [53] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Comput. Math. Appl.*, vol. 2, pp. 17–40, 1976.
- [54] 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, rev. française d'inf. Rech. Oper.", vol. R-2, pp. 41–76, 1975.
- [55] F. Lin, K. Kwong, J. Belliveau, and L. Wald, "Parallel imaging reconstruction using automatic regularization," *Magn. Reson. Med.*, vol. 51, pp. 559–567, 2004.
- [56] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate  $O(1/k^2)$  (in russian)," *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983.
- [57] Y. Nesterov, Gradient methods for minimizing composite objective function Center Operations Research Econometrics (CORE), Catholic Univ. Louvain, Louvain-la-Neuve, Belgium, Discussion paper 76, 2007.
- [58] M. Lustig, M. Alley, S. Vasanawala, D. Donoho, and J. Pauly, "L1 spir-it: Autocalibrating parallel imaging compressed sensing," *Proc. Int. Soc. Mag. Reson. Med.*, p. 379, 2009.
- [59] F. Huang, S. Vijayakumar, S. Reza, S. Hertel, and G. Duensing, "A self-calibration method for radial grappa/k-t grappa," *Magn. Reson. Med.*, vol. 57, pp. 1075–1085, 2007.
- [60] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbert," *C. R. Acad. Sci. Paris*, vol. 255, pp. 2897–2899, 1962.
- [61] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [62] X. Ye, Y. Chen, and F. Huang, "Computational acceleration for MR image reconstruction in partially parallel imaging," *IEEE Trans. Med. Imag.*, to be published.