Augmented Lagrangian with Variable Splitting for Faster Non-Cartesian L₁-SPIRiT MR Image Reconstruction

Daniel S. Weller, *Member, IEEE,* Sathish Ramani, *Member, IEEE,* and Jeffrey A. Fessler, *Fellow, IEEE*

Abstract—SPIRiT (iterative self-consistent parallel imaging reconstruction), and its sparsity-regularized variant L₁-SPIRiT, are compatible with both Cartesian and non-Cartesian MRI sampling trajectories. However, the non-Cartesian framework is more expensive computationally, involving a nonuniform Fourier transform with a nontrivial Gram matrix. We propose a novel implementation of the regularized reconstruction problem using variable splitting, alternating minimization of the augmented Lagrangian, and careful preconditioning. Our new method based on the alternating direction method of multipliers converges much faster than existing methods because of the preconditioners' heightened effectiveness. We demonstrate such rapid convergence substantially improves image quality for a fixed computation time. Our framework is a step forward towards rapid non-Cartesian L₁-SPIRiT reconstructions.

Index Terms—Magnetic resonance imaging, non-Cartesian reconstruction, parallel imaging reconstruction, compressed sensing, augmented Lagrangian, preconditioning.

I. INTRODUCTION

MORE sophisticated magnetic resonance image acquisition techniques have spurred development of novel image reconstruction algorithms. Applications sensitive to motion or off-resonance effects, like cardiac and abdominal imaging, can benefit from non-Cartesian sampling [1], [2], [3], [4]. However, many such developments are more computationally intensive than the simple Fourier transform-based methods used for basic Cartesian acquisitions.

Accelerated parallel imaging reconstruction with sparsity-promoting regularization is particularly expensive computationally, a consequence of both the iterative nature of most sparsity-promoting algorithms and the large amount of data acquired by multiple receivers. This limitation is true of L₁-SPIRiT [5], an iterative algorithm involving applying sparsifying transforms, convolving k-space with Iterative Self-consistent Parallel Imaging Reconstruction (SPIRiT) kernels [6], and enforcing consistency with acquired data. In addition, model-based reconstructions avoid approximation errors from gridding data by evaluating the nonuniform Fourier

transform, such as the Nonuniform Fast Fourier Transform (NUFFT) [7], during each iteration. This transform requires more computation than its uniform counterpart due to over-sampling and its nontrivial Gram matrix. These computational challenges are evident in 2D reconstructions, like the ones performed in this work, and become even more significant in 3D situations. While optimized software and parallel processing can improve overall speed, such approaches skirt the underlying difficulty that such algorithms require many iterations to resolve the problem's complicated matrix structure.

We propose a new implementation of non-Cartesian L₁-SPIRiT that accelerates reconstruction using variable splitting, alternating minimization, and preconditioning [8], [9], reducing the number of iterations required to achieve a desired reconstruction quality. Like existing SPIRiT-based methods, our implementation of L₁-SPIRiT uses autocalibration signal (ACS) data in place of explicitly measured coil sensitivity profiles. We propose a novel implementation of L₁-SPIRiT that simplifies the linear subproblem structures, enabling preconditioned conjugate gradients (PCG) [10] to more efficiently solve the linear subproblems and achieve faster convergence than existing methods. Unlike conventional gradient-based methods, such as nonlinear conjugate gradients (NLCG) [11], our approach also accommodates nonsmooth regularizers, such as the sparsity-promoting ℓ_1 norm or total variation, without corner-rounding approximations.

Non-Cartesian L₁-SPIRiT recently has been applied to myocardial perfusion imaging using a spiral trajectory [12]. To reduce the reconstruction time, previous efforts have focused on image domain implementations of the SPIRiT kernels [6] and parallel processing using graphical processing units (GPUs) [13]. Augmented Lagrangian-based techniques already have been exploited in non-Cartesian single-coil regularized MRI reconstruction [14]; our variable splitting is more involved due to the SPIRiT consistency equations [6]. Variable splitting also has been used to accelerate other non-Cartesian parallel imaging reconstructions [15], [16].

In this paper, we begin by explaining our measurement model for non-Cartesian data. Then, we provide background on the SPIRiT framework for reconstructing images from non-Cartesian multi-channel data and the regularized version called L₁-SPIRiT. To motivate our approach, we explain how to implement L₁-SPIRiT using Split-Bregman (SB) [17] iterations; then we propose our method based on the alternating direction method of multipliers (ADMM) [18], [19]. We describe how to precondition the ADMM subproblems, and we discuss

how to choose the ADMM tuning parameters to encourage fast convergence. After choosing tuning parameters based on simulated data, we demonstrate the generality of those tuning parameters on real data sets by comparing our algorithm against competitors based on NLCG and SB iterations. We conclude with a discussion of the merits of our method and possible uses and extensions beyond this paper.

II. MEASUREMENT MODEL

The parallel imaging framework called Sensitivity Encoding (SENSE) [20] assumes knowledge of the explicit sensitivity weighting relationship between the scanned object and the received data. From these weights, SENSE forms an inverse problem to reconstruct the underlying object image. Another parallel imaging method called Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) [21] uses an autocalibration signal (ACS) region of k-space in lieu of these sensitivity profiles, forming linear convolution equations that directly reconstruct missing k-space from uniformly-spaced Cartesian undersampled data. A similar approach called SPIRiT [6] also uses ACS data, establishing linear self-consistency equations between a k-space point and its immediate neighbors in all the coil channels. As described in the next section, SPIRiT uses these consistency equations to reconstruct the images or k-space recorded by all the coils. The L$_1$-SPIRiT variant regularizes the SPIRiT reconstruction problem using transform-domain joint sparsity. Whereas SENSE produces a single object image, both GRAPPA and SPIRiT reconstruct sensitivity-weighted images for each coil. One combines the reconstructed coil images into a single object image as a final post-processing step [22].

Consider a set of N_c complex-valued D-dimensional continuous coil images represented by the mapping $\mathbf{x}(\mathbf{r}) : \mathbb{R}^D \mapsto \mathbb{C}^{N_c}$, where D is usually two or three. Non-Cartesian k-space samples represent M measurements of the spatial Fourier transform of $\mathbf{x}(\mathbf{r})$ at arbitrary sample frequencies $\mathbf{\Omega}_m$, for $m = 1, \ldots, M$. The data matrix $\mathbf{D} \in \mathbb{C}^{M \times N_c}$ contains in each column the M observations for a coil channel. These measurements are typically noisy; in parallel MRI, this noise usually results from thermal fluctuations independent of the signal and modeled by zero-mean complex Gaussian ($\mathcal{CN}$) noise $\mathbf{N} \in \mathbb{C}^{M \times N_c}$, independent across frequencies but correlated among coils. The measurements at the mth k-space location are

$$[\mathbf{D}]_{m,:} = \int \mathbf{x}(\mathbf{r})e^{-j(\mathbf{\Omega}_m \cdot \mathbf{r})}\, d\mathbf{r} + [\mathbf{N}]_{m,:}, \qquad (1)$$

where $[\cdot]_{m,:}$ is the mth row of that matrix, and $[\mathbf{N}]_{m,:} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ is the zero-mean noise vector, with covariance matrix $\mathbf{\Lambda}$, for the coil array measurements at the mth k-space frequency. In (1), field inhomogeneity and relaxation effects are considered part of the image $\mathbf{x}(\mathbf{r})$ and ignored.

Consider N_v-voxel discretizations $\mathbf{X} \in \mathbb{C}^{N_v \times N_c}$ of the N_c-channel coil images represented by $\mathbf{x}(\mathbf{r})$. Assuming the continuous-domain images are appropriately bandlimited (an assumption widespread across MRI), the samples $\mathbf{D}$ are related to $\mathbf{X}$ via the discrete-space Fourier transform (DSFT). If the image has finite extent, the measurement model can be written in matrix form as

$$\mathbf{D} = \mathbf{A}\mathbf{X} + \mathbf{N}, \qquad (2)$$

where the elements of the system matrix $\mathbf{A}$ are

$$[\mathbf{A}]_{m,n} = e^{-j(\boldsymbol{\omega}_m \cdot \mathbf{r}_n)}, \qquad (3)$$

$[\boldsymbol{\omega}_m]_d = [\Delta\mathbf{r}]_d[\mathbf{\Omega}_m]_d$, $\Delta\mathbf{r}$ is the voxel size of the discretization, and $\mathbf{r}_n$ is the D-dimensional index of the nth voxel. In many such implementations of inverse problems involving this measurement model, the matrix-vector products for the DSFT matrix are approximated using a gridding/interpolating fast Fourier transform (FFT)-based algorithm. We use the NUFFT [7] with a $1.5\times$-oversampled FFT grid.

III. NON-CARTESIAN SPIRiT

As described in [6], non-Cartesian SPIRiT can be implemented in either the k-space or image domain. The unknown Cartesian k-space maps to the acquired non-Cartesian data via a gridding-like interpolation kernel, while the unknown multi-channel coil images map to the data via nonuniformly sampling the spatial Fourier transforms of the images. While the frequency-domain implementation avoids iteratively computing Fourier transforms, the image-domain formulation can become more efficient for larger images, as the SPIRiT convolutions become multiplications. Hence, we focus on the image-domain construction.

In the image-domain formulation, the data constraint objective is framed in terms of the discretized images in (2). The SPIRiT kernel consistency objective is formed by equating the underlying block convolutions in k-space with matrix multiplication in the image domain. We write the SPIRiT consistency equation $\mathbf{x} = \mathbf{G}\mathbf{x}$ in the image domain using the stacked vector of coil images $\mathbf{x} = \mathrm{col}(\mathbf{X})$ ($\mathrm{col}(\cdot)$ stacks the columns of a matrix into a single vector) and the matrix $\mathbf{G}$, which represents the k-space self-consistency convolutions in the image domain. This matrix is an $N_c \times N_c$ block matrix of diagonal submatrices of size $N_v \times N_v$, each of which represents a SPIRiT convolution from one coil channel to another.

When noisy data are acquired, denoising the data may yield favorable looking images. We allow the coil images to vary by a small amount from exact consistency with the data. The unconstrained optimization problem is developed using a tuning parameter $\lambda > 0$ in [6]:

$$\mathbf{X}^* \in \arg\min_{\mathbf{X}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X} - \mathbf{D}\|_F^2 + \frac{\lambda}{2}\|(\mathbf{G} - \mathbf{I})\,\mathrm{col}(\mathbf{X})\|_2^2, \quad (4)$$

where $\|\cdot\|_F$ is the Frobenius norm.

As demonstrated by the popular L$_1$-SPIRiT algorithm [13], regularizing the SPIRiT parallel imaging reconstruction can improve image quality significantly, such as when the image is known to have a sparse representation. In this work, we consider multi-channel regularizers $R(\mathbf{X})$ of the form

$$R(\mathbf{X}) = \gamma \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}]_{\ell,:}), \qquad (5)$$

where $\mathbf{R}_1, \ldots, \mathbf{R}_Q \in \mathbb{C}^{L \times N_v}$ are linear transform operators with circulant Gram matrices, $\psi(\cdot) : \mathbb{C}^{N_c Q} \mapsto \mathbb{R}_{\geq 0}$ is a

convex potential function, and $w_\ell \geq 0$ are weights useful for implementing shift-variant transformations. A global tuning parameter γ controls the contribution of the regularizer to the overall optimization problem.

This framework accommodates a large class of popular sparsifying regularizers used for compressed sensing in MRI [23]. The discretized total variation (TV) penalty belongs to this class, with $\mathbf{R}_1, \ldots, \mathbf{R}_Q$ implementing finite differences with Q neighboring directions, w_ℓ equaling one except on the edge of the image, where it equals zero, and $\psi(\cdot)$ being the $\ell_{1,2}$ or $\ell_{2,2}$ mixed norm (the first norm being across $\mathbf{R}$'s, and the second norm being across coil channels) for anisotropic or isotropic TV, respectively. Orthonormal and shift-invariant wavelets and frames also can be used for $\mathbf{R}$. A non-orthonormal shift-variant wavelet can be accommodated using its shift-invariant extension for $\mathbf{R}$, and downsampling by setting the weights w_ℓ to zero for the coefficients to be discarded. The extension to multiple regularizers, such as combining TV and wavelets, is straightforward.

The regularized form of (4) is:

$$\mathbf{X}^* \in \arg\min_{\mathbf{X}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X}-\mathbf{D}\|_F^2 + \frac{\lambda}{2}\|(\mathbf{G}-\mathbf{I})\operatorname{col}(\mathbf{X})\|_2^2 + R(\mathbf{X}).$$

(6)

For many common regularizers, various algorithms exist to approximately solve (6). One such method, popular for its ease of implementation, is NLCG. This algorithm is applied to SPIRiT with a "smoothed" or "corner-rounded" regularizer so that the gradient of $\psi(\cdot)$ exists. We propose approaching this problem using variable splitting, enabling closed-form solutions of the nonlinear subproblems for many common non-differentiable regularizers.

IV. REGULARIZED NON-CARTESIAN SPIRiT USING SPLIT-BREGMAN ITERATION

Split-Bregman (SB) iteration [17] minimizes (6) by introducing the auxiliary matrix $\mathbf{W} = [\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}]$. The constrained problem equivalent to (6) is

$$\{\mathbf{X}^*, \mathbf{W}^*\} \in \arg\min_{\mathbf{X}, \mathbf{W}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X}-\mathbf{D}\|_F^2 + \frac{\lambda}{2}\|(\mathbf{G}-\mathbf{I})\operatorname{col}(\mathbf{X})\|_2^2$$
$$+ \gamma \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{W}]_{\ell,:}),$$
$$\text{s.t. } \mathbf{W} = [\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}].$$

(7)

The SB iteration introduces a dual variable $\mathbf{B} \in \mathbb{C}^{L \times N_c Q}$, and an additional tuning parameter $\mu > 0$, and iterates over solving the subproblems below for $i = 1, 2, \ldots$

$$\mathbf{X}^i \leftarrow \arg\min_{\mathbf{X}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X}-\mathbf{D}\|_F^2 + \frac{\lambda}{2}\|(\mathbf{G}-\mathbf{I})\operatorname{col}(\mathbf{X})\|_2^2$$
$$+ \frac{\mu}{2}\|\mathbf{W}^{i-1} - [\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}] - \mathbf{B}^{i-1}\|_F^2, \quad (8)$$

$$\mathbf{W}^i \leftarrow \arg\min_{\mathbf{W}} \gamma \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{W}]_{\ell,:})$$
$$+ \frac{\mu}{2}\|\mathbf{W} - [\mathbf{R}_1\mathbf{X}^i, \ldots, \mathbf{R}_Q\mathbf{X}^i] - \mathbf{B}^{i-1}\|_F^2, \quad (9)$$

$$\mathbf{B}^i \leftarrow \mathbf{B}^{i-1} + ([\mathbf{R}_1\mathbf{X}^i, \ldots, \mathbf{R}_Q\mathbf{X}^i] - \mathbf{W}^i). \quad (10)$$

When using multiple regularizers, different penalty parameters μ can be used for each regularizer. This flexibility may speed convergence if one auxiliary variable converges at a different rate than another. However, varying the tuning parameters affects only the convergence rate, not the final solution, for strictly convex problems.

For the $\ell_{1,2}$ mixed norm and the TV penalty, vector soft-thresholding [24] is a low-complexity closed-form solution for the subproblem in (9). For more general regularizers, proximal gradient [25] or nonlinear conjugate gradient algorithms can approximately solve this subproblem efficiently. Either way, iterations are likely inexpensive because (9) decomposes into L smaller subproblems of size $N_c Q$ and does not include matrix-vector multiplications with $\mathbf{A}$ or $\mathbf{G}$.

However, the least squares problem in (8) remains computationally expensive, as it still involves matrix-vector products with $\mathbf{A}$, $\mathbf{G}$, and the $\mathbf{R}$'s. Iterative methods like conjugate gradients may take many steps to converge, because $\mathbf{A}$ and $\mathbf{G}$ have very different structures, the combination of which is not well-suited for preconditioning. To help alleviate this difficulty, we propose an additional variable split that separates the SPIRiT objective from the rest of the problem.

V. ADMM FOR NON-CARTESIAN SPIRiT

Now, we propose a new algorithm for (6) based on the alternating direction method of multipliers (ADMM) [26], [18], [27], [19] that offers simpler inner subproblems and leads to faster convergence than the Split-Bregman approach described in the previous section. In addition to $\mathbf{W}$ defined previously, we introduce $\mathbf{z} = \operatorname{col}(\mathbf{X})$ and solve the following constrained optimization problem that is equivalent to (6):

$$\{\mathbf{X}^*, \mathbf{z}^*, \mathbf{W}^*\} \in \arg\min_{\mathbf{X}, \mathbf{z}, \mathbf{W}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X}-\mathbf{D}\|_F^2 + \frac{\lambda}{2}\|(\mathbf{G}-\mathbf{I})\mathbf{z}\|_2^2$$
$$+ \gamma \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{W}]_{\ell,:}),$$
$$\text{s.t. } \mathbf{z} = \operatorname{col}(\mathbf{X}), \mathbf{W} = [\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}].$$

(11)

We use ADMM to solve the augmented Lagrangian form of (11). Relabeling the dual variable and penalty parameter (for $\mathbf{W}$) $\mathbf{B}_W$ and μ_W, we add a second dual variable and penalty parameter (for $\mathbf{z}$) $\mathbf{b}_z \in \mathbb{C}^{N_v N_c}$ and $\mu_z > 0$. The ith iteration consists of the following subproblems:

$$\mathbf{X}^i \leftarrow \arg\min_{\mathbf{X}} \frac{1}{2N_v}\|\mathbf{A}\mathbf{X}-\mathbf{D}\|_F^2$$
$$+ \frac{\mu_W}{2}\|[\mathbf{R}_1\mathbf{X}, \ldots, \mathbf{R}_Q\mathbf{X}] - \mathbf{W}^{i-1} + \mathbf{B}_W^{i-1}\|_F^2$$
$$+ \frac{\mu_z}{2}\|\operatorname{col}(\mathbf{X}) - \mathbf{z}^{i-1} + \mathbf{b}_z^{i-1}\|_2^2, \quad (12)$$

$$\mathbf{z}^i \leftarrow \arg\min_{\mathbf{z}} \frac{\lambda}{2}\|(\mathbf{G}-\mathbf{I})\mathbf{z}\|_2^2$$
$$+ \frac{\mu_z}{2}\|\operatorname{col}(\mathbf{X}^i) - \mathbf{z} + \mathbf{b}_z^{i-1}\|_2^2, \quad (13)$$

$$\mathbf{b}_z^i \leftarrow \mathbf{b}_z^{i-1} + (\operatorname{col}(\mathbf{X}^i) - \mathbf{z}^i), \quad (14)$$

$$\mathbf{W}^i \leftarrow \arg\min_{\mathbf{W}} \frac{\gamma}{\mu_W} \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{W}]_{\ell,:})$$
$$+ \frac{1}{2}\|[\mathbf{R}_1\mathbf{X}^i, \ldots, \mathbf{R}_Q\mathbf{X}^i] - \mathbf{W} + \mathbf{B}_W^{i-1}\|_F^2, \quad (15)$$

$$\mathbf{B}_W^i \leftarrow \mathbf{B}_W^{i-1} + ([\mathbf{R}_1\mathbf{X}^i, \ldots, \mathbf{R}_Q\mathbf{X}^i] - \mathbf{W}^i). \quad (16)$$

Given an initial set of coil images $\mathbf{X}^0$, such as the density-corrected conjugate phase reconstruction [28], we initialize $\mathbf{W}^0 \leftarrow [\mathbf{R}_1\mathbf{X}^0, \ldots, \mathbf{R}_Q\mathbf{X}^0]$, $\mathbf{z}^0 \leftarrow \mathrm{col}(\mathbf{X}^0)$, and set the scaled dual variables $\mathbf{B}_W^0$ and $\mathbf{b}_z^0$ to zero.

The subproblems in (12) and (13) are least-squares problems that we solve using preconditioned conjugate gradient methods described in Sec. VI. The subproblem for $\mathbf{W}$ remains the same as before. Besides decoupling the structures for $\mathbf{A}$ and $\mathbf{G}$, we expect to gain efficiency due to parallelism in the subproblems. The diagonal-block structure of the SPIRiT consistency operation $(\mathbf{G} - \mathbf{I})$ couples variables only across coils, meaning that we are really solving N_v subproblems of size N_c. When utilizing a relatively small array coil, or coil compression techniques on larger array coils [29], we may even solve each iteration of (13) directly. Similarly, since the measurement matrix $\mathbf{A}$ and the transforms $\mathbf{R}_1, \ldots, \mathbf{R}_Q$ only couple variables within the same coil, we are solving N_c subproblems of size N_v. From a distributed computing point of view, where memory and bandwidth requirements scale with the size of the optimization problem, natural decoupling of variables can be extremely beneficial.

VI. PRECONDITIONED CONJUGATE GRADIENTS

Without preconditioning, iterative solutions for the least-squares subproblems in (12) and (13) may take many iterations to converge, slowing the overall algorithm. Effective preconditioners take advantage of the structure of the system matrix of a problem to nearly invert that matrix efficiently, significantly reducing the number of steps needed to solve the problem. Let us examine the structures present in our subproblems.

The solution of (12) requires inverting the matrix $\frac{1}{N_v}\mathbf{A}'\mathbf{A} + \mu_W \sum_{q=1}^{Q} \mathbf{R}_q'\mathbf{R}_q + \mu_z\mathbf{I}$, where $[\cdot]'$ is the conjugate transpose. This matrix shown in Fig. 1 is a weighted sum of Gram matrices damped by $\mu_z\mathbf{I}$, which ensures the sum is strictly positive definite. When formulating the original problem, we assume that the Gram matrices for the $\mathbf{R}$'s are circulant, as is the case for total variation shift-invariant wavelets. As we shall see next, the Gram matrix $\mathbf{A}'\mathbf{A}$ is nearly circulant. Since the sum of circulant matrices is circulant, we can find an optimal circulant preconditioner for the entire matrix.

The Gram matrix $\mathbf{T}$ of the DSFT matrix $\mathbf{A}$ has entries $[\mathbf{A}'\mathbf{A}]_{n_1,n_2} = \sum_{m=1}^{M} e^{j(\boldsymbol{\omega}_m \cdot (\mathbf{r}_{n_1} - \mathbf{r}_{n_2}))}$, where $\mathbf{r}_{n_1}$ and $\mathbf{r}_{n_2}$ are the D-dimensional coordinates for sample indexes n_1 and n_2, respectively. In one dimension, this Gram matrix has Toeplitz structure. In higher dimensions, the Gram matrix inherits multi-level block-Toeplitz structure, where each block is itself a block-Toeplitz matrix [30]. A vast literature exists on preconditioning linear systems involving such matrices [31], [32], [33]. The majority of such approaches involve constructing a multi-level block-circulant matrix (composed of block-circulant blocks) that approximates or embeds the desired multi-level block-Toeplitz matrix [34], [35]. In this work, we construct the multi-level block-circulant preconditioner optimal in the Frobenius-norm sense to the Toeplitz DSFT Gram matrix, by solving

$$c(\mathbf{T}) = \underset{\mathbf{C}:\text{circulant block-circulant}}{\arg\min} \|\mathbf{C} - \mathbf{T}\|_F. \qquad (17)$$
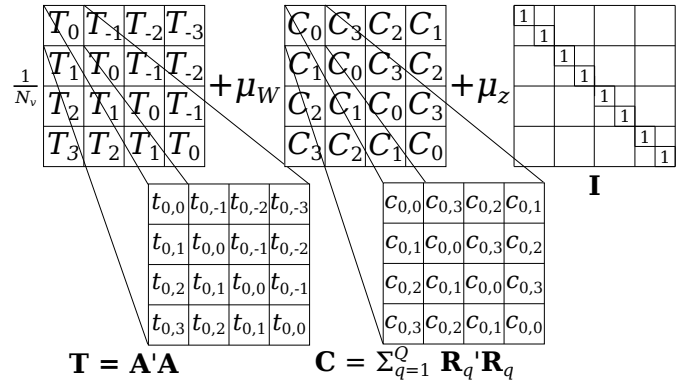

Fig. 1.   The matrix to be inverted in (12) consists of a multi-level block-Toeplitz Gram matrix $\mathbf{T} = \mathbf{A}'\mathbf{A}$, multi-level block-circulant Gram matrices $\mathbf{C} = \sum_{q=1}^{Q} \mathbf{R}_q'\mathbf{R}_q$, and a scaled identity matrix $\mu_z\mathbf{I}$. Here, we portray the first two levels of the multi-level structures, focusing on the blocks $\mathbf{T}_0$ and $\mathbf{C}_0$ for clarity. The other blocks $\mathbf{T}_n$ and $\mathbf{C}_n$ are similar.

As derived in [35], the coefficients c_n, $n = 0, \ldots, N-1$, of the optimal $N \times N$ circulant matrix are related to the coefficients t_n, $n = -(N-1), \ldots, N-1$, of the $N \times N$ Toeplitz matrix by

$$c_n = \frac{n}{N}t_{-(N-n)} + \frac{N-n}{N}t_n. \qquad (18)$$

The optimal coefficients for the multi-level block-circulant approximation can be computed recursively, starting from the lowest-level Toeplitz matrices, and averaging the higher-level coefficient blocks according to an analogous formula:

$$\mathbf{C}_n = \frac{n}{N}c(\mathbf{T}_{-(N-n)}) + \frac{N-n}{N}c(\mathbf{T}_n). \qquad (19)$$

This circulant preconditioner is positive definite if the original Toeplitz matrix is positive definite [36]; adding a scalar matrix $\mu\mathbf{I}$ (as is done in the problem in (12)) to the Gram matrix satisfies this requirement. The optimal circulant preconditioner for the linear system that solves (12) is $c(\frac{1}{N_v}\mathbf{A}'\mathbf{A} + \mu_z\mathbf{I}) + \mu_W \sum_{q=1}^{Q} \mathbf{R}_q'\mathbf{R}_q$. Our proposed method uses this preconditioner with the conjugate gradient method to solve (12).

The matrix $\boldsymbol{\Gamma} = \lambda(\mathbf{G} - \mathbf{I})'(\mathbf{G} - \mathbf{I}) + \mu_z\mathbf{I}$ for the linear subproblem in (13) has diagonal-block structure, so a simple reordering of the matrix yields a block-diagonal structure, composed of $N_c \times N_c$ blocks, as shown in Fig. 2. The inverse of this matrix, $\boldsymbol{\Gamma}^{-1}$, appears in the update of $\mathbf{z}$: $\mathbf{z}^i = \boldsymbol{\Gamma}^{-1}(\mu_z(\mathrm{col}(\mathbf{X}^i) + \mathbf{b}_z^{i-1}))$. When N_c is sufficiently small, we find $\boldsymbol{\Gamma}^{-1}$ by directly inverting each $N_c \times N_c$ block in $\boldsymbol{\Gamma}$. Otherwise, preconditioned conjugate gradients can be used for this subproblem as well. Beyond being Hermitian positive definite, these blocks do not necessarily possess additional special structure. In particular, we would not expect a circulant preconditioner to be effective here. Instead, we consider the "optimal" diagonal preconditioner (in the Frobenius sense). Analogous to the circulant preconditioner definition, the optimal diagonal preconditioner for a matrix $\boldsymbol{\Gamma}$ minimizes the norm $\|\boldsymbol{\Delta} - \boldsymbol{\Gamma}\|_F$ over the class of diagonal matrices $\boldsymbol{\Delta}$. This preconditioner simply equals the diagonal of $\boldsymbol{\Gamma}$. The computational cost of generating this preconditioner from $\mathbf{G}$ is $O(N_c^2)$, comparable to a matrix-vector product involving $\mathbf{G}$.
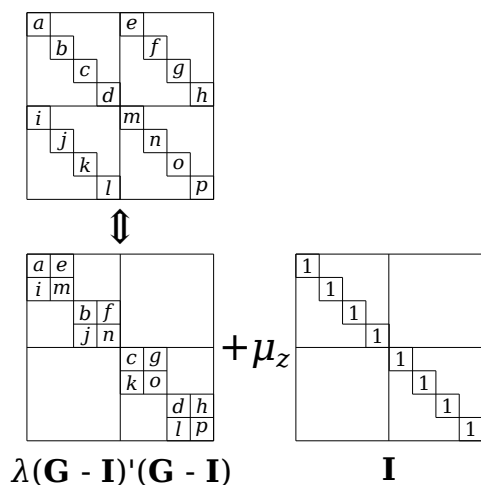
Fig. 2. The matrix to be inverted in (13) consists of a diagonal-block Gram matrix $(\mathbf{G}-\mathbf{I})'(\mathbf{G}-\mathbf{I})$ and a scaled identity matrix $\mu_z\mathbf{I}$. The diagonal-block matrix is equivalent (via a permutation) to a block-diagonal matrix, as shown, enabling fast inversion for small N_c.

One may also precondition the main Split-Bregman subproblem in (8). Unlike the matrices for the subproblems in the ADMM-based algorithm, the matrix to be inverted contains both nearly circulant and block-diagonal Gram matrices, the sum of which is neither (nearly) circulant nor diagonal. Thus, any circulant or diagonal preconditioner designed using the methods described previously will be less effective because it can model only part of the structure.

To summarize, we employ diagonal preconditioning for the Split-Bregman subproblem solving for $\mathbf{X}$ in (8), and for the ADMM-based subproblem solving for $\mathbf{z}$ in (13). We employ circulant preconditioning for the ADMM-based subproblem solving for $\mathbf{X}$ in (12). The Split-Bregman and ADMM-based subproblems solving for $\mathbf{W}$ in (9) and (15) are shrinkage steps that do not involve matrix inversion, and hence, no preconditioning. Also, (10), (14), and (16) do not involve preconditioning.

VII. ADMM PARAMETER SELECTION

As with any augmented Lagrangian-based method, the choices of tuning parameters μ_W and μ_z greatly influences the convergence behavior of the algorithm. Choosing a parameter to be too large may cause the objective function to decrease too slowly, while too small a parameter would slow the convergence of the corresponding dual variable. Additionally, since our subproblems for updating $\mathbf{X}$ (and potentially $\mathbf{z}$) are solved iteratively, the convergence properties of these subproblems depend on the conditioning of their respective system matrices. The condition number of the system matrix of the least-squares update for $\mathbf{X}$ generally increases as μ_W or μ_z decrease, while decreasing μ_z increases the condition number of the system matrix of the least-squares update for $\mathbf{z}$. Keeping these condition numbers reasonably small should ensure the inner subproblems converge quickly, promoting convergence of the overall algorithm.

To choose appropriate values of μ_W and μ_Z, we consider their effects on the various subproblems in our ADMM frame-

work. The subproblem for updating $\mathbf{W}$ is affected by μ_W. When the potential function(s) $\psi(\cdot)$ admit closed-form solutions in the form of vector shrinkage-thresholding operations, we can choose a minimum value of μ_W such that the fraction of thresholded transform coefficients in the solution of (15) for $t = 0$ is at most some value, τ_W, thus, achieving some degree of uniformity across similarly transform-sparse data sets. The condition number κ_z of the system matrix for updating $\mathbf{z}$ in (13) depends only on μ_z. By setting an upper threshold on this condition number, we can find a minimum value for μ_z. The subproblem for updating $\mathbf{X}$ involves both μ_W and μ_z. Setting a maximum threshold on the condition number of this problem's system matrix, κ_X, we can find a boundary function of values of μ_W and μ_z that ensure sufficient conditioning of this problem. Given threshold values τ_W, κ_z, and κ_X, we can compute for a given data set minimal choices of μ_W and μ_z that satisfy these constraints.

We approximate the condition number of the $\mathbf{X}$-update by finding the condition number of the circulant preconditioner of the system matrix. Computing the condition number of the $\mathbf{z}$-update system matrix for a particular choice of μ_z is relatively easy due to the diagonal-block structure of $(\mathbf{G}-\mathbf{I})'(\mathbf{G}-\mathbf{I})$, as long as N_c is not too large (we can compute κ_z in $O(N_v N_c^3)$ time). Alternatively, either condition number can be estimated using iterative methods like the Krylov subspace method used in LSQR [37] for computing a stopping criterion.

To summarize this procedure, we begin with maximum condition numbers κ_X and κ_z and thresholding fraction τ_W that optimize the convergence rate of the proposed algorithm according to a preliminary study like the experiment with simulated data described in Sec. IX. Then,

1) We compute maximum and minimum singular values for the $N_c \times N_c$ blocks of the Gram matrix $(\mathbf{G}-\mathbf{I})'(\mathbf{G}-\mathbf{I})$. Keep only the very largest (σ_1) and smallest ($\sigma_{\min}$) singular values for the entire matrix.

2) Using these singular values, find the smallest μ_z that satisfies

$$\frac{\lambda\sigma_1 + \mu_z}{\lambda\sigma_{\min} + \mu_z} \leq \kappa_z. \qquad (20)$$

Larger values of μ_z would reduce the condition number of the $\mathbf{z}$-update system matrix further.

3) Construct a matrix $\mathbf{W}^0$ of sparse transforms of the initial $\mathbf{X}$ value: $\mathbf{W}^0 = [\mathbf{R}_1\mathbf{X}^0, \cdots, \mathbf{R}_Q\mathbf{X}^0]$. Using a scalar solver or lookup table, determine the penalty parameter μ_W that thresholds no more than τ_W fraction of the total number of sparse transform coefficients. A larger value of μ_W generally will reduce the number of thresholded values further.

4) Compute the largest and smallest singular values of the circulant preconditioner for $\mathbf{A}'\mathbf{A}$, and for the circulant Gram matrix $\sum_{q=1}^{Q} \mathbf{R}_q'\mathbf{R}_q$. The singular values of these matrices are simply the multi-dimensional DFT of the matrices' first columns. Let $\sigma_{1,A}$, $\sigma_{1,R}$, $\sigma_{\min,A}$, and $\sigma_{\min,R}$ represent the largest and smallest singular values for these two Gram matrices, respectively.

5) Using a generic solver like Matlab's fmincon, minimize $\mu_z^2 + \mu_W^2$ subject to the constraints in the earlier
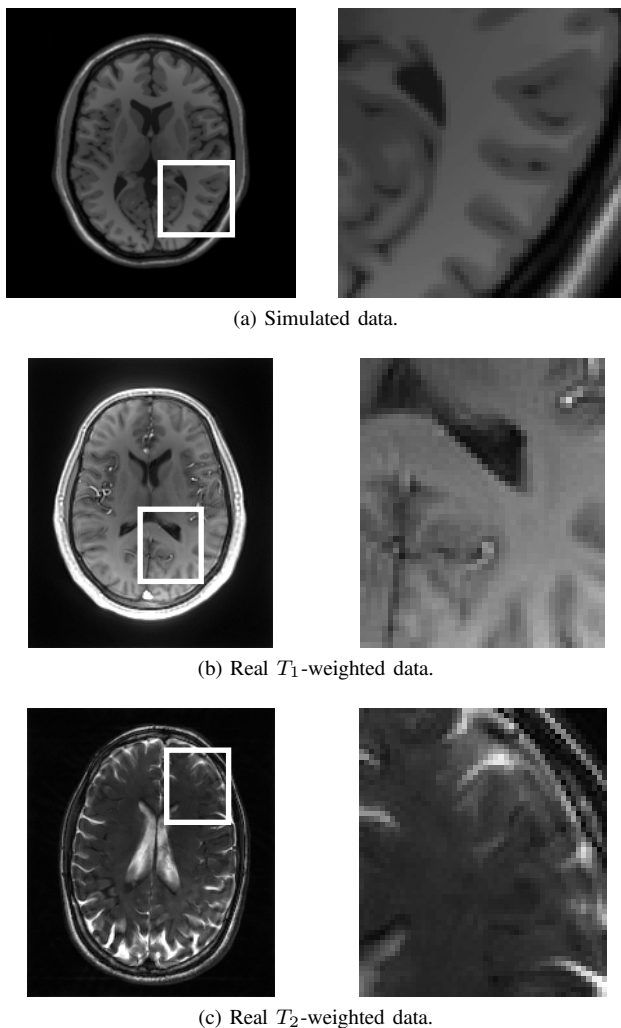
(a) Simulated data.

(b) Real T_1-weighted data.

(c) Real T_2-weighted data.

Fig. 3. Fully-sampled multi-channel (a) T_1-weighted simulated (Brainweb) data, (b) T_1-weighted real data, and (c) T_2-weighted real data magnitude images formed using sum-of-squares. The inset rectangular regions are enlarged to show fine details.

steps and the constraint

$$\frac{\sigma_{1,A}/N_v + \mu_z + \mu_W \sigma_{1,R}}{\sigma_{\min,A}/N_v + \mu_z + \mu_W \sigma_{\min,R}} \leq \kappa_X. \qquad (21)$$

VIII. DATA ACQUISITION

To test the validity and performance of our proposed implementation, we used both simulated and real data shown in Fig. 3. Our simulated data consisted of a 2D slice of 3D T_1-weighted Brainweb [38] data initially generated with no noise and isotropic 1 mm resolution. We simulated both an eight-channel and a 16-channel circular array coil at 0.5 mm resolution and applied the sensitivities to an interpolated slice to provide the illusion of continuous-space. We then reduced the multi-channel images back to 1 mm resolution; these images were retained as ground truth for our experiments. The k-space data were sampled along a multi-shot spiral trajectory (120 interleaves, 510 samples/leaf) designed for a 256×256 matrix with 1 mm isotropic resolution. For the eight-channel data, 40 leaves were retained at random ($3\times$-undersampling);

30 leaves of 16-channel data were retained at random ($4\times$-undersampling). Complex Gaussian noise correlated across channels according to the electric coupling model for linear arrays [22] was added to the multi-channel k-space, resulting in a signal-to-noise ratio (SNR) of 40 dB, where SNR is calculated using

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\mathbf{X}\|_F}{\sqrt{M \cdot \text{tr}\{\mathbf{\Lambda}\}}} \right), \qquad (22)$$

and the matrix trace $\text{tr}\{\cdot\}$ accounts for the correlations across coil channels. A Cartesian 30×30 ACS region (36×36 for the 16-channel coil) was generated (with the same $\mathbf{\Lambda}$) from the center of k-space in order to calibrate the SPIRiT kernels.

We acquired a single 2D slice of a 3D T_1-weighted custom spoiled gradient echo stack-of-spirals acquisition (120 interleaves, 832 samples/leaf) designed for a 256×256 image matrix with a 30×30 cm FOV using code [39] from http://mrsrl.stanford.edu/~brian/vdspiral/. The data was acquired on a GE 3 T Discovery scanner using a vendor-supplied head coil with eight receive channels. Because of the large FOV, we cropped the FOV in our reconstructions to yield a 208×176 matrix with the same resolution as before. The reference image was generated by averaging five time frames from the fully-sampled data set with 50 un-regularized iterations of NUFFT-based reconstruction [40]. The data was then undersampled by randomly keeping 40 interleaves ($3\times$-undersampling).

On a GE 3 T Signa scanner, using a similar 8-channel vendor-supplied head coil we acquired a single 2D slice of a 3D T_2-weighted custom steady-state free precession (SSFP) stack-of-radials acquisition (420 spokes, 257 samples/spoke) designed for a 256×256 image matrix with a 26 cm square FOV. We used a pilot scan of a gel phantom to correct analytical k-space locations for linear eddy current effects [41]. We also cropped the FOV of this data set around the head to form a 192×224 isotropic image. We reconstructed the ground truth from five averaged fully-sampled time frames using 50 un-regularized iterations of NUFFT-based reconstruction [40]. The radial data was undersampled by keeping only one fourth of the spokes (randomly selected).

In all of this paper's experiments, we used Matlab to reconstruct all undersampled data sets with non-Cartesian L_1-SPIRiT with both orthonormal wavelets (4-level 'db4') and isotropic TV regularization. We implemented these regularizations by introducing the splits $\mathbf{W}_1 = \text{DWT}(\mathbf{X})$ and $\mathbf{W}_2 = [d_x\mathbf{X}, d_y\mathbf{X}]$ for the orthonormal wavelet and isotropic TV, respectively. The same Split-Bregman μ and ADMM penalty parameters μ_W were used for both regularizers. These penalty parameters, along with the other ADMM penalty parameter μ_z, were tuned using the condition-number/thresholding rules described in the previous section. The regularization parameter $\gamma = 10^{-3}$ was used for both regularizers in all three data sets, and $\lambda = 1$ was used for all reconstructions, as these values appeared to produce consistently reasonable image quality in preliminary studies. The Matlab-based SPIRiT implementation used is publicly available at http://www.eecs.berkeley.edu/~mlustig/Software.html, and the

NLCG implementation of L$_1$-SPIRiT used for comparison is a combination of the conjugate gradient implementation of non-Cartesian SPIRiT included with that software and the single-channel NLCG implementation of compressed sensing for MRI [23] available from the same website. Visual comparisons are used to assess image quality. Magnitude images are created using a sum-of-squares combination [22], and difference images are formed from the multi-channel image data and then combined using sum-of-squares. For quantitative experiments, we used the original objective $f(\mathbf{X})$ found in (6) and subtracted the baseline objective $f^{\text{opt}} = f(\mathbf{X}^{\text{opt}})$, where $\mathbf{X}^{\text{opt}}$ is the baseline reconstruction resulting from preconditioned Split-Bregman run for 20,000 iterations. We ran so many iterations to establish a baseline because our experiments on simulated data suggest this conventional method converges relatively slowly. We also compared convergence in $\mathbf{X}$ for our ADMM method with different μ's using the normalized root mean square difference (NRMSD):

$$\text{NRMSD}(\mathbf{X}, \mathbf{X}^{\text{opt}}) = \|\mathbf{X} - \mathbf{X}^{\text{opt}}\|_2 / \|\mathbf{X}^{\text{opt}}\|_2, \qquad (23)$$

Our optimization problem may not be strictly convex, so the optimal $\mathbf{X}^{\text{opt}}$ may not be unique. When comparing across methods, we use the Split-Bregman baseline for fairness. Conjugate phase reconstructions (with density weighting) are used to initialize all the methods.

IX. SELECTING PARAMETERS FOR ADMM

Our initial experiment concerns the importance of choosing the μ's to achieve fast convergence of ADMM, and the generality of the criteria (κ_z, κ_X, and τ_W) used to identify μ for a given data set. We began by manually tuning μ_z and μ_W and running the proposed method for 500 iterations on the simulated T_1-weighted eight-channel data to establish a quality baseline for the reconstructed multi-channel images $\mathbf{X}^{\text{opt}}$ that achieve the minimum f^{opt} of the non-Cartesian L$_1$-SPIRiT objective function in (6). Then, we ran 125 iterations of ADMM (using preconditioned conjugate gradients for updating both $\mathbf{X}$ and $\mathbf{z}$) on the same undersampled data repeatedly for a range of $\mu_z \in [0.01, 1]$ and $\mu_W \in [0.001, 0.1]$, logarithmically spaced. We chose μ_W and μ_z that minimize the NRMSD of the reconstructed images (relative to the baseline $\mathbf{X}^{\text{opt}}$) and objective function values after 125 iterations. As shown in Fig. 4, the optimal μ_z and μ_W roughly agree for these two metrics (image distance and objective value), so we found roughly corresponding κ_z, κ_X, and τ_W for those values $\mu_z = 0.1$ and $\mu_W = 0.01$ that yield the fastest convergence. For these values of μ_z and μ_W, $\kappa_z = 19.5$, $\kappa_X = 3374$, and $\tau_W = 0.426$. A relatively large region of μ_z and μ_W provide fast convergence in this example, so we use rough values for these parameters ($\kappa_z = 20$, $\kappa_X = 4000$, and $\tau_W = 0.4$) for all data sets in the experiments that follow.

X. CONVERGENCE RATE COMPARISONS

Using the data and reconstruction methods described in Sec. VIII, we compare the convergence rate of our ADMM-based implementation of L$_1$-SPIRiT against both NLCG (with
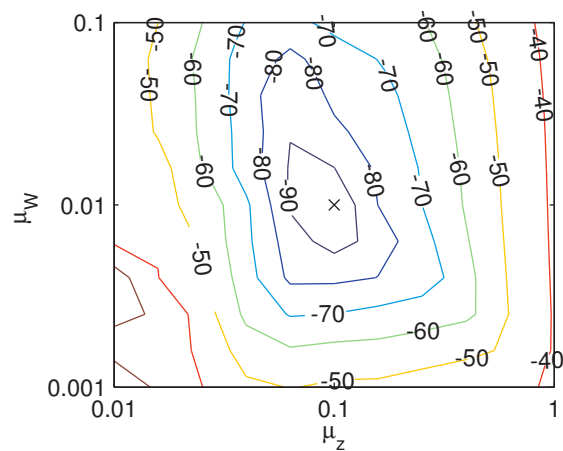
(a) NRMSD's (dB scale, relative to baseline $\mathbf{X}^{\text{opt}}$).
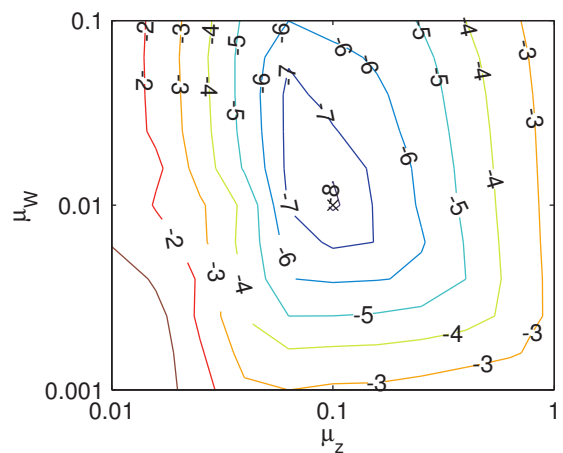
(b) Objective function values ($\log_{10}$-scale, relative to baseline f^{opt}).

Fig. 4. Contour plots show both (a) the normalized root mean square difference (NRMSD) and (b) the objective function value (plotted on a $\log_{10}$-scale relative to baseline f^{opt}) after 125 ADMM iterations are minimized around $\mu_z = 0.1$, $\mu_W = 0.01$.

adaptive backtracking line search) and Split-Bregman iteration as competing methods. For completeness, we study three versions of our method: (1) CG $\mathbf{X}$-updates (4 inner iterations) without preconditioning, and exact $\mathbf{z}$-updates, (2) with preconditioned-CG $\mathbf{X}$-updates (4 inner iterations), and exact $\mathbf{z}$-updates, and (3) with preconditioned-CG updates (4 inner iterations) for both $\mathbf{X}$ and $\mathbf{z}$. We compare our method against NLCG (without preconditioning) and Split-Bregman iteration, the latter with and without diagonal preconditioning. We examine the reconstruction quality and optimality of the objective function values (relative to the Split-Bregman baseline) after a fixed period of time (25 seconds for the real data shown in the paper, and 50 and 100 seconds for the simulated 8- and 16-channel data in the supplementary material[1]). The number of inner iterations and parameter choices for all these methods were set equal when possible to make the comparisons equitable.

Figures 5 and 6 compare the first 60 seconds of objective

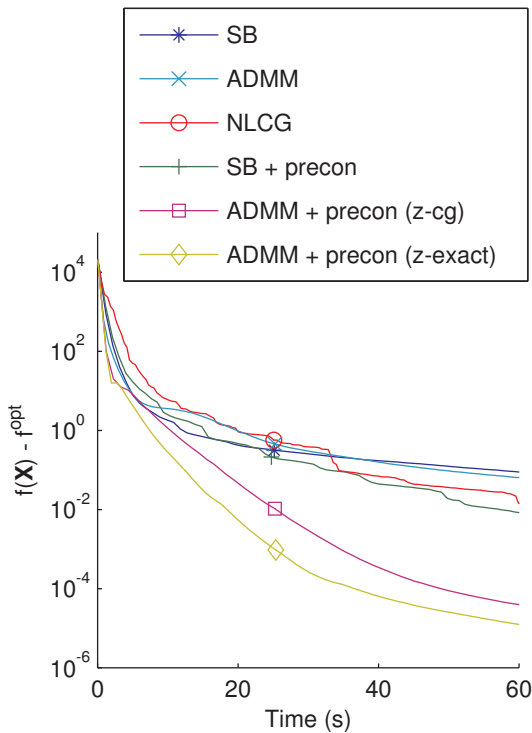[1]The supplementary material is available at http://ieeexplore.ieee.org.
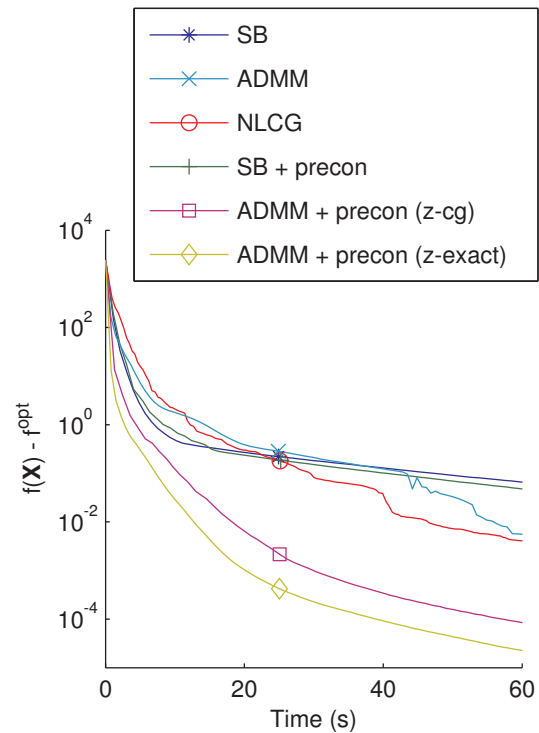
Fig. 5. The objective values $f(\mathbf{X}) - f^{\text{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for the first 60 seconds of each of the compared algorithms for the real T_1-weighted spiral data.

Fig. 6. The objective values $f(\mathbf{X}) - f^{\text{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for the first 60 seconds of each of the compared algorithms for the real T_2-weighted radial data.

function values for the proposed ADMM method against NLCG and Split-Bregman (SB) iteration for real T_1-weighted spiral and T_2-weighted radial data, respectively. The objective value $f(\mathbf{X})$ converges much more rapidly for both Split-Bregman and ADMM when using preconditioning than for their un-preconditioned counterparts. In the first sixty seconds shown in the figures, preconditioned ADMM with either iterative or exact $\mathbf{z}$-updates converges much faster than both Split-Bregman and NLCG, yielding lower objective values in a fixed amount of time. Preconditioned ADMM with exact $\mathbf{z}$-updates converges faster than ADMM with iterative $\mathbf{z}$-updates with diagonal preconditioning for both data sets, as expected, since the $\mathbf{z}$-update is no longer iterative. Convergence plots for the simulated T_1-weighted spiral data with both eight and 16 receive coil channels are provided in supplementary material. Longer-term convergence plots in terms of NRMSD and objective function values also may be found in the supplementary material for all data sets. The NRMSD curves generally follow the same trend we observe for the objective function values. These longer-term plots reveal that after 300-400 seconds of computation, the NLCG method begins to converge extremely rapidly for a limited time, after which the objective and NRMSD values remain constant. For both real data sets, NLCG even outperforms the proposed methods for a brief time, although one would rarely run so many iterations in practice.

While all the methods use common code for transforms, multiplying SPIRiT kernels, and performing the NUFFT, the number of such operations performed and other overhead affects the per-iteration computation time: (mean and standard deviation reported for 8-channel simulated data) 0.54 ± 0.095s for NLCG, 1.2 ± 0.052s for Split-Bregman without preconditioning, 1.2 ± 0.050s for preconditioned Split-Bregman, 0.85 ± 0.046s for ADMM without preconditioning (but exact $\mathbf{z}$-updates), 1.4 ± 0.053s for preconditioned ADMM with inexact $\mathbf{z}$-updates, and 0.90 ± 0.042s for preconditioned ADMM with exact $\mathbf{z}$-updates. Preconditioning requires only negligible additional computation per iteration.

Figure 7 portrays reconstructed images and difference images relative to the fully-sampled ground truth for the real T_1-weighted spiral data. After 25 seconds of reconstruction time, the competing methods' reconstructed images exhibit swirling artifacts consistent with an incomplete reconstruction, while our proposed ADMM method mitigates these artifacts with either preconditioned or exact $\mathbf{z}$-updates. Similar results are observed for the real T_2-weighted radial data set after 25 seconds of reconstruction time in Fig. 8. Additional reconstructions for the simulated T_1-weighted spiral data sets with eight and 16 receive channels can be found in supplementary material. The difference images in both figures are relative to the ground truths in Fig. 3. To save space, we did not show Split-Bregman or our proposed ADMM method without preconditioning due to the poor image quality predicted by the objective values in Figs. 5 and 6.

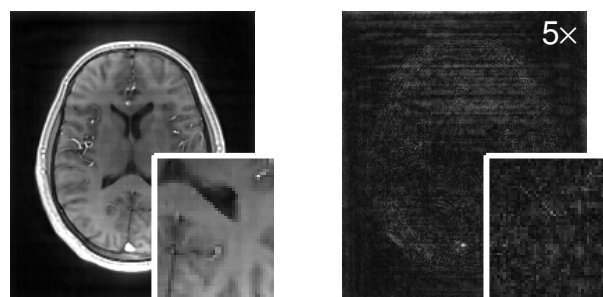## XI. DISCUSSION AND CONCLUSIONS

NLCG exhibits comparable or faster convergence in both NRMSD and objective function value than un-preconditioned
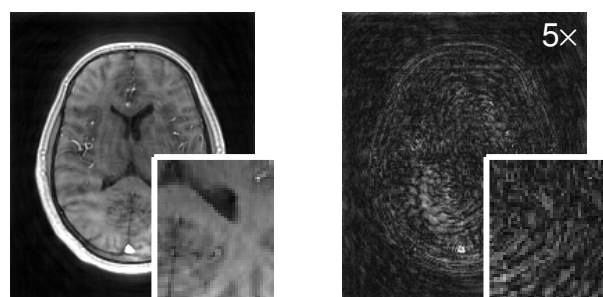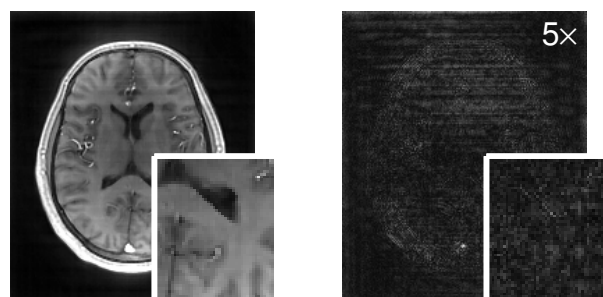
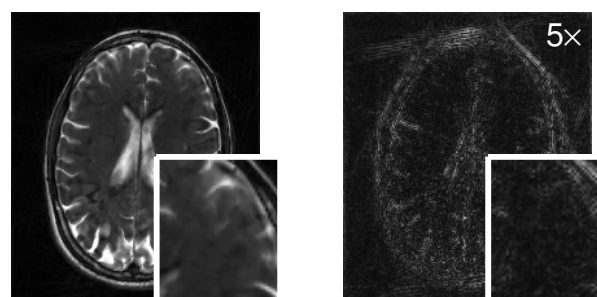(a) NLCG recon.
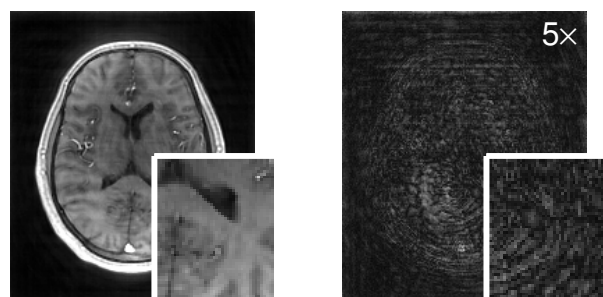
(b) Split-Bregman recon.
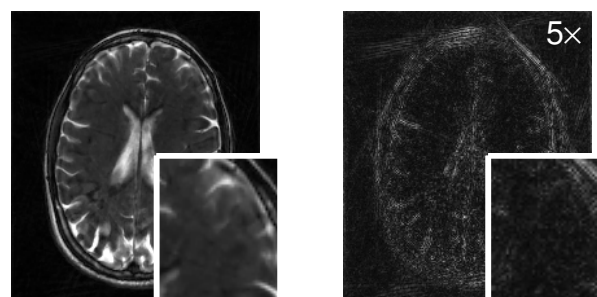

(c) ADMM with PCG for **z**-update.

(d) ADMM with exact **z**-update.

Fig. 7. Sum-of-squared combined reconstructed and difference images for the T$_1$-weighted real spiral data set after 25 seconds of reconstruction time using (a) NLCG, (b) Split-Bregman with diagonal preconditioning, (c) ADMM with preconditioning both **X** and **z**-updates, and (d) ADMM with preconditioned **X**-updates and exact **z**-updates demonstrate the advantage of rapid convergence on image quality after a fixed amount of time. Zoomed images of the inset region show differences in fine details. The difference images in the right column, relative to the ground truth, are all scaled by 5×.

(a) NLCG recon.

(b) Split-Bregman recon.

(c) ADMM with PCG for **z**-update.

(d) ADMM with exact **z**-update.

Fig. 8. Sum-of-squared combined reconstructed and difference images for the T$_2$-weighted real radial data set after 25 seconds of reconstruction time using (a) NLCG, (b) Split-Bregman with diagonal preconditioning, (c) ADMM with preconditioning both **X** and **z**-updates, and (d) ADMM with preconditioned **X**-updates and exact **z**-updates demonstrate the advantage of rapid convergence on image quality after a fixed amount of time. Zoomed images of the inset region show differences in fine details. The difference images in the right column, relative to the ground truth, are all scaled by 5×.

SB and ADMM. Generally speaking, NLCG gains from considering the (approximate) gradient of the entire objective function, while un-preconditioned SB and un-preconditioned ADMM alternate between two or more subproblems. However, in the case of preconditioned ADMM, the advantage of effective preconditioning outweighs the disadvantage of alternating
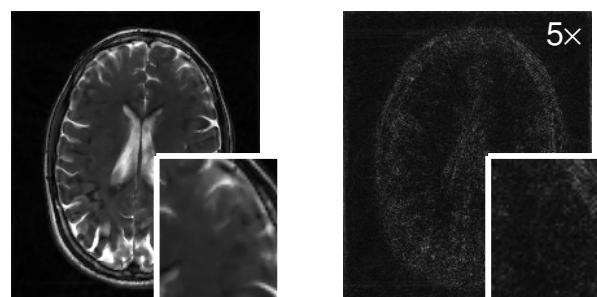
minimization.

Effective preconditioning yields demonstrable advantages for many optimization problems in terms of convergence rate, especially when appropriate preconditioners are easy to compute and invert. In designing our proposed ADMM-based method, we introduce auxiliary variables to decouple matri-

ces with dissimilar structures into separate subproblems. We employ circulant and diagonal preconditioners tailored to the structures of the matrices found in each subproblem. Applying such preconditioners to the entire problem, as is done for the Split-Bregman method, is less effective since the combination of matrices in the overall problem is neither nearly circulant nor diagonal. As a result, for both simulated and real data, our proposed ADMM-based method converges much faster than competing methods when preconditioners are used. All these methods ideally would converge to the same objective value, except for corner-rounding errors in the NLCG implementation and finite-precision effects. However, after a fixed amount of reconstruction time, both sets of reconstructions show markedly fewer artifacts than either NLCG or Split-Bregman iteration. The artifacts in these reconstructions demonstrate the importance of ensuring fast convergence in practice, when a limited number of iterations is used. Another approach, originally raised by a reviewer, would involve re-gridding the data onto an oversampled Cartesian grid, and running iterations of Cartesian L_1-SPIRiT. However, the oversampling level would have to be sufficient to approximate the reconstruction quality of the non-Cartesian reconstruction, likely resulting in DFT's large enough to outweigh any computational gains from employing Cartesian reconstruction. Alternatively, this approach may yield an improved initial image over a density-weighted conjugate phase reconstruction, reducing the work of the proposed algorithm and decreasing overall computation time.

A practical limitation of ADMM-based methods in general is the necessity of selecting tuning parameter values that lead to fast convergence. We have provided heuristics for choosing μ_z and μ_W based on the structure of the subproblems in our method, and our experiments confirm that these heuristics ensure reasonable convergence rates across multiple data sets without the need for additional tuning.

To improve upon our method and enable even faster reconstructions, we expect that our method would combine readily with the parallel computing framework described for Cartesian L_1-SPIRiT [13]. The time-consuming computations like performing the NUFFT and regularizing transforms generally reduce to operations that can be parallelized effectively.

Additionally, we expect that some modifications to our proposed method may yield further acceleration. Since a significant gap exists in our reconstructions between ADMM with exact $\mathbf{z}$-updates and ADMM with iterative $\mathbf{z}$-updates, even after diagonal preconditioning, our method may benefit from more sophisticated preconditioners for the $\mathbf{z}$-update subproblem. One candidate is the "super-optimal" [36] diagonal preconditioner. Tridiagonal preconditioners also may yield improved performance. However, computing these preconditioners from $\mathbf{G}$ require $O(N_c^3)$ operations, comparable to simply inverting the matrix.

Another approach worth further investigation concerns extending our method to regularizing transforms without circulant Gram matrices. We could accommodate such transforms in our ADMM framework via introducing an additional split, $\mathbf{Y} = \mathbf{X}$, yielding another constrained problem equivalent

to (6):

$$\{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{z}^*, \mathbf{W}^*\} \in \underset{\mathbf{X}, \mathbf{Y}, \mathbf{z}, \mathbf{W}}{\arg\min} \frac{1}{2N_v} \|\mathbf{AX} - \mathbf{D}\|_F^2$$
$$+ \frac{\lambda}{2} \|(\mathbf{G} - \mathbf{I})\mathbf{z}\|_2^2 + \gamma \sum_{\ell=1}^{L} w_\ell \psi([\mathbf{W}]_{\ell,:}),$$
$$\text{s.t. } \mathbf{z} = \text{col}(\mathbf{X}), \mathbf{Y} = \mathbf{X},$$
$$\mathbf{W} = [\mathbf{R}_1 \mathbf{Y}, \ldots, \mathbf{R}_Q \mathbf{Y}].$$
$$(24)$$

Hence, we effectively separate the subproblem containing the measurement matrix $\mathbf{A}$ from the subproblem containing the regularizer transforms. Potentially, multiple splits would help when multiple regularizer transforms have different types of structure.

The current implementation ignores main field inhomogeneity effects. Applying a spatial weighting matrix to our measurement model to account for these distortions would complicate the implementation of the $\mathbf{X}$-update step, since the Gram matrix would no longer be nearly circulant [30]. A possible extension to handle field inhomogeneity would involve an additional variable split to separate the spatial weighting matrix from the NUFFT operator. We plan to investigate this approach in future work.

Finally, we mention that our approach enables users to preserve the data in the reconstruction by introducing an additional scaled dual variable $\mathbf{B}_D$ and tuning parameter μ_D to penalize the measurement residual and enforce equality. However, such preservation typically is not done in non-Cartesian imaging, where the acquired data is not directly included in the reconstruction (unlike Cartesian imaging). Enforcing the data fit with equality would likely have undesirable effects for trajectories like spiral or radial sampling with a high sampling density near DC.

To summarize, we have proposed a novel implementation of non-Cartesian L_1-SPIRiT using variable splitting, ADMM, and preconditioned iterative methods. We provided experimental evidence that our implementation converges faster than both NLCG and Split-Bregman iteration, and our heuristic for choosing ADMM parameters generalizes across real and simulated data sets. Lastly, we discussed the merits of possible extensions to our method.

REFERENCES

[1] Y. Qian, Z. Zhang, V. A. Stenger, and Y. Wang, "Self-calibrated spiral SENSE," *Mag. Res. Med.*, vol. 52, no. 3, pp. 688–92, Sep. 2004.

[2] E. N. Yeh, M. Stuber, C. A. McKenzie, R. M. Botnar, T. Leiner, M. A. Ohliger, A. K. Grant, J. D. Willig-Onwuachi, and D. K. Sodickson, "Inherently self-calibrating non-cartesian parallel imaging," *Mag. Res. Med.*, vol. 54, no. 1, pp. 1–8, Jul. 2005.

[3] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Mag. Res. Med.*, vol. 57, no. 6, pp. 1086–98, Jun. 2007.

[4] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the golden ratio for time-resolved MRI," *IEEE Trans. Med. Imag.*, vol. 26, no. 1, pp. 68–76, Jan. 2007.

[5] M. Lustig, M. Alley, S. Vasanawala, D. L. Donoho, and J. M. Pauly, "L1 SPIR-iT: Autocalibrating parallel imaging compressed sensing," in *Proc. Intl. Soc. Mag. Res. Med.*, 2009, p. 334.

[6] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Mag. Res. Med.*, vol. 64, no. 2, pp. 457–71, Aug. 2010.

[7] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Sig. Proc.*, vol. 51, no. 2, pp. 560–74, Feb. 2003.

[8] K. Chen, *Matrix preconditioning techniques and applications*. Cambridge University Press, 2005.

[9] M. Benzi, "Preconditioning techniques for large linear systems: A survey," *J. Comp. Phys.*, vol. 182, no. 2, pp. 418–77, Nov. 2005.

[10] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Research Nat. Bur. Standards*, vol. 49, pp. 409–36, 1952.

[11] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *Comput. J*, vol. 7, no. 2, pp. 149–54, 1964.

[12] Y. Yang, X. Feng, C. H. Meyer, C. M. Kramer, and M. Salerno, "First-pass myocardial perfusion imaging with whole ventricular coverage using L1-SPIRIT accelerated spiral trajectories," *J. Cardio. Mag. Res.*, vol. 15, no. S1, p. 20, 2013.

[13] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast ℓ_1-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1250–62, Jun. 2012.

[14] J. Aelterman, H. Q. Luong, B. Goossens, A. Pizurica, and W. Philips, "Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data," *Signal Processing*, vol. 91, no. 12, pp. 2731–42, Jan. 2011.

[15] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 694–706, Mar. 2011.

[16] ——, "Accelerated non-Cartesian SENSE reconstruction using a majorize-minimize algorithm combining variable-splitting," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, 2013, pp. 700–3.

[17] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–43, 2009.

[18] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[20] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Mag. Res. Med.*, vol. 42, no. 5, pp. 952–62, Nov. 1999.

[21] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Mag. Res. Med.*, vol. 47, no. 6, pp. 1202–10, Jun. 2002.

[22] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR phased array," *Mag. Res. Med.*, vol. 16, no. 2, pp. 192–225, Nov. 1990.

[23] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Mag. Res. Med.*, vol. 58, no. 6, pp. 1182–95, Dec. 2007.

[24] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–55, Sep. 1994.

[25] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *siam-jmms*, vol. 4, no. 4, pp. 1168–200, 2005.

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

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

[28] D. C. Noll, J. A. Fessler, and B. P. Sutton, "Conjugate phase MRI reconstruction with spatially variant sample density correction," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 325–36, Mar. 2005.

[29] M. Buehrer, K. P. Pruessmann, P. Boesiger, and S. Kozerke, "Array compression for MRI with large coil arrays," *Mag. Res. Med.*, vol. 57, no. 6, pp. 1131–9, Jun. 2007.

[30] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *IEEE Trans. Sig. Proc.*, vol. 53, no. 9, pp. 3393–402, Sep. 2005.

[31] R. H. Chan and M. K. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Review*, vol. 38, no. 3, pp. 427–82, Sep. 1996.

[32] M. Van Barel, G. Heinig, and P. Kravanja, "A superfast method for solving Toeplitz linear least squares problems," *Linear Algebra and its Applications*, vol. 366, pp. 441–57, Jun. 2003.

[33] P. Favati, G. Lotti, and O. Menchi, "A divide and conquer algorithm for the superfast solution of Toeplitz-like systems," *SIAM J. Matrix. Anal. Appl.*, vol. 33, no. 4, pp. 1039–56, 2012.

[34] G. Strang, "A proposal for Toeplitz matrix calculations," *Stud. Appl. Math.*, vol. 74, pp. 171–6, 1986.

[35] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Stat. Comp.*, vol. 9, no. 4, pp. 766–71, Jul. 1988.

[36] E. Tyrtyshnikov, "Optimal and superoptimal circulant preconditioners," *SIAM J. Matrix. Anal. Appl.*, vol. 13, no. 2, pp. 459–73, 1992.

[37] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Software*, vol. 8, no. 1, pp. 43–71, Mar. 1982.

[38] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–97, Nov. 1999.

[39] J. H. Lee, B. A. Hargreaves, B. S. Hu, and D. G. Nishimura, "Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion," *Mag. Res. Med.*, vol. 50, no. 6, pp. 1276–85, Dec. 2003.

[40] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 178–88, Feb. 2003.

[41] K. T. Block and M. Uecker, "Simple method for adaptive gradient-delay compensation in radial MRI," in *Proc. Intl. Soc. Mag. Res. Med.*, 2011, p. 2816.

Augmented Lagrangian with Variable Splitting for Faster Non-Cartesian L$_1$-SPIRiT MR Image Reconstruction: Supplementary Material

Daniel S. Weller, *Member, IEEE,* Sathish Ramani, *Member, IEEE,* and Jeffrey A. Fessler, *Fellow, IEEE*

Here, we provide additional results supporting the use of our implementation of non-Cartesian L$_1$-SPIRiT, using a combination of variable splitting, an augmented Lagrangian formulation, and preconditioned conjugate gradients.

I. DATA ACQUISITION

In the paper [1], we described both simulated and real data sets. Figure 1, reprinted from [1], portrays ground truth images for all data sets. To save space, we focused on convergence plots and reconstructed images resulting from our real data experiments only. Additional convergence plots and reconstructions featuring our simulated data sets with eight and 16 channels of T_1-weighted spiral data are included in this supplement. The eight-channel simulated data was undersampled by a factor of three, while the 16-channel simulated k-space was undersampled by a factor of four to leverage the increased number of coils.

In the experiments that follow, our reconstructions use the same methods and parameter choices described in the original paper. In particular, we used the same sparsifying transforms, regularization parameters, penalty parameter selections, and conjugate phase image initializations. To evaluate the convergence rates of the different algorithms, our baseline was the same preconditioned Split-Bregman method run for 20,000 iterations.

II. CONVERGENCE RATE COMPARISONS

In the original paper, relative objective function values $f(\mathbf{X}) - f^{\text{opt}}$ are plotted for both the real T_1-weighted spiral 2D data and real T_2-weighted radial 2D data. Here, we provide the same plots over a longer time scale, so that long-term convergence behavior of the different algorithms is evident. Since the objective function may vary relatively slowly in a neighborhood around its minimum, the reconstructed image $\mathbf{X}$ may actually be farther from its optimum $\mathbf{X}^{\text{opt}}$ than a small relative objective function value would lead one to believe. Therefore, in this supplement, we also include plots of the normalized root mean squared difference (NRMSD), as introduced in the original text. To minimize the bias any

(a) Simulated data.

(b) Real data (T$_1$-weighted).
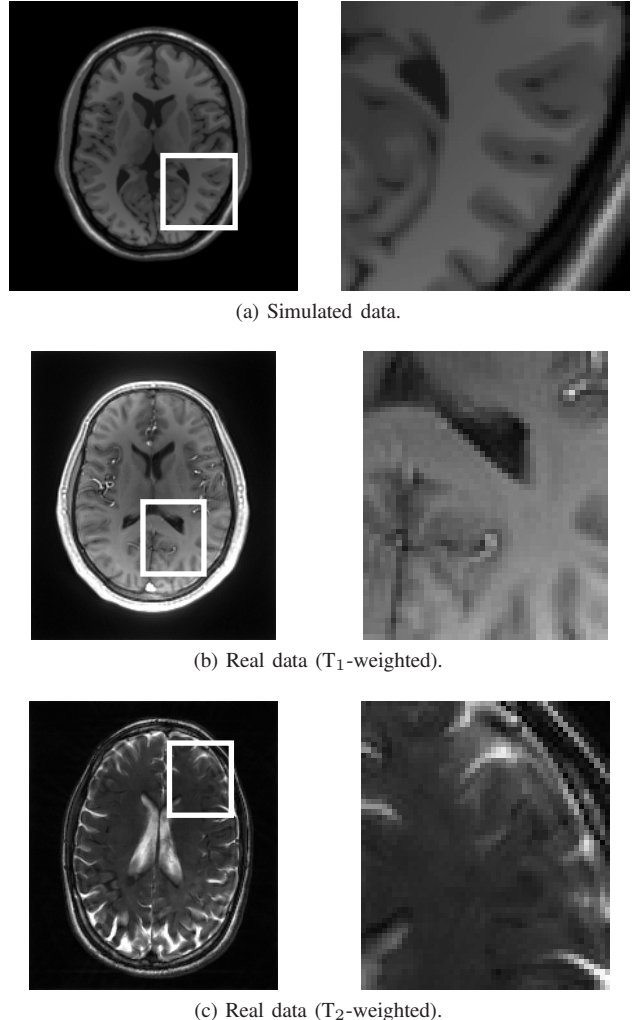

(c) Real data (T$_2$-weighted).

Fig. 1. Fully-sampled multi-channel magnitude images formed using sum-of-squares of (a) T$_1$-weighted simulated (Brainweb) data, (b) T$_1$-weighted real data, and (c) T$_2$-weighted real data. The inset rectangular regions are enlarged to show fine details. This figure is reprinted from [1].

potential non-uniqueness of the optimal reconstruction may have on our conclusions, we used the Split-Bregman baseline image.

The convergence plots for the real T$_1$-weighted spiral 2D data are shown in Fig. 2. Our preconditioned ADMM-based methods both converge more rapidly than preconditioned Split-Bregman iteration and both approaches without preconditioning, in terms of both objective function value and
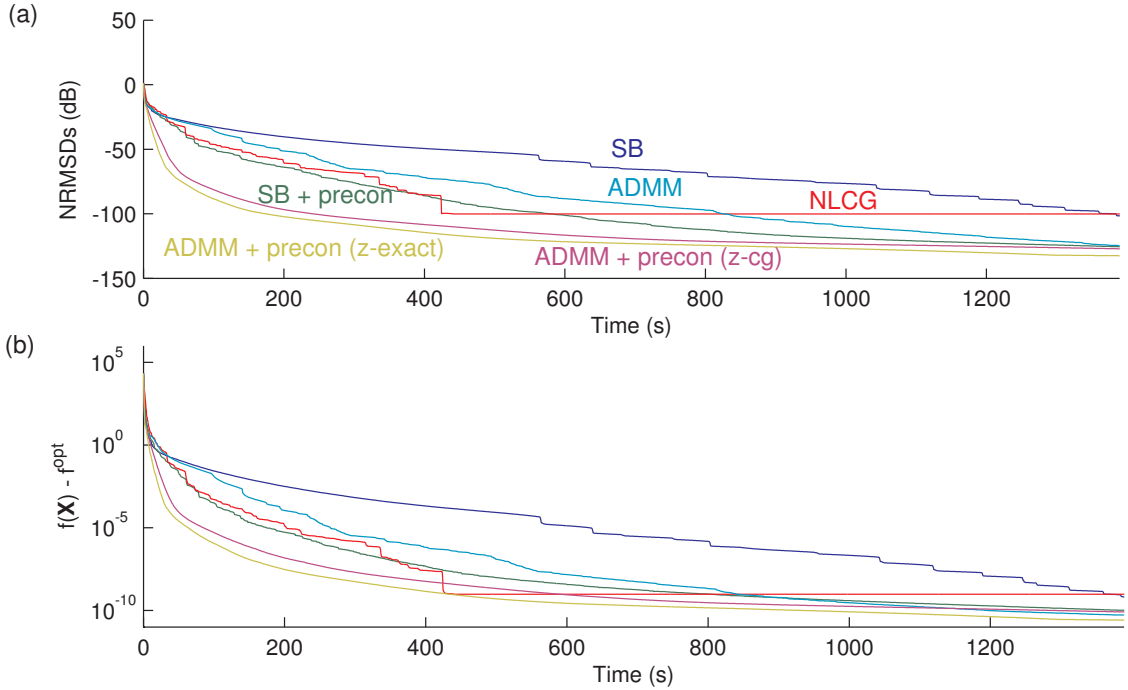
Fig. 2. The (a) NRMSDs (dB) and (b) objective values $f(\mathbf{X}) - f^{\text{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for each of the compared algorithms for the real T_1-weighted spiral data.

NRMSD. This fast convergence occurs even though we use the condition numbers calibrated with the simulated data to tune our method's penalty parameters. Our proposed method also beats the NLCG method until the NLCG method experiences rapid convergence around 400 seconds into the experiment. The NLCG method flattens out, although the difference in final objective value or NRMSD appears relatively insignificant for this data set.

The convergence rate improvements portrayed in Fig. 3 for the real T_2-weighted radially sampled 2D data echo what were observed for the real T_1-weighted spiral data. Again, our preconditioned ADMM-based methods converge faster than preconditioned Split-Bregman. The improvement over NLCG is significant in the initial minutes of run time, but the NLCG method ungoes a phase of rapid convergence around 300 seconds, before flattening out.

The convergence plots for the 8-channel simulated spiral 2D data shown in Fig. 4 and the 16-channel simulated spiral 2D data in Fig 5 both depict far more substantial advantages in both NRMSD and objective function value. The preconditioned forms of the proposed ADMM method converge much more rapidly than either competing method for the simulated data. The non-preconditioned variants have the slowest convergence rates of all. This performance holds true when measuring convergence of either the image or the objective function value. We notice that the difference in convergence rates for using exact $\mathbf{z}$-updates versus preconditioned $\mathbf{z}$-updates in our proposed algorithm appears to grow noticeably with the larger number of channels.

Reconstructed images from the 8-channel simulated spiral 2D data captured after 50 seconds are displayed in Fig. 6 for

NLCG and the preconditioned variants of the Split-Bregman and proposed ADMM-based methods. Figure 7 shows similar reconstructions from the 16-channel simulated spiral 2D data after 100 seconds of reconstruction time. Prominent spiral artifacts are visible in both the magnitude and difference images for the NLCG and Split-Bregman methods, in both sets of reconstructions. The proposed ADMM-based reconstructions contain no such visible artifacts. These images corroborate the substantial difference in reconstruction quality after a limited reconstruction time observed for the two real data sets in the original manuscript. Furthermore, we demonstrate the improvement carries over to larger receive array coils, even using the same condition numbers and shrinkage threshold fraction to tune our algorithm.

III. DISCUSSION AND CONCLUSIONS

The preconditioned variants of the proposed ADMM-based method clearly outperform conventional approaches for non-Cartesian L_1-SPIRiT reconstruction, at least for a limited reconstruction time. The uniform quality of the reconstructions of all the data sets using the same condition numbers and sparse threshold fraction demonstrates the generality of both our proposed method and our approach for tuning the penalty parameters, which are known to impact the convergence rate of such algorithms. The importance of preconditioning is also evident here, as the un-preconditioned ADMM-based method converges relatively slowly, as does the un-preconditioned Split-Bregman iteration.
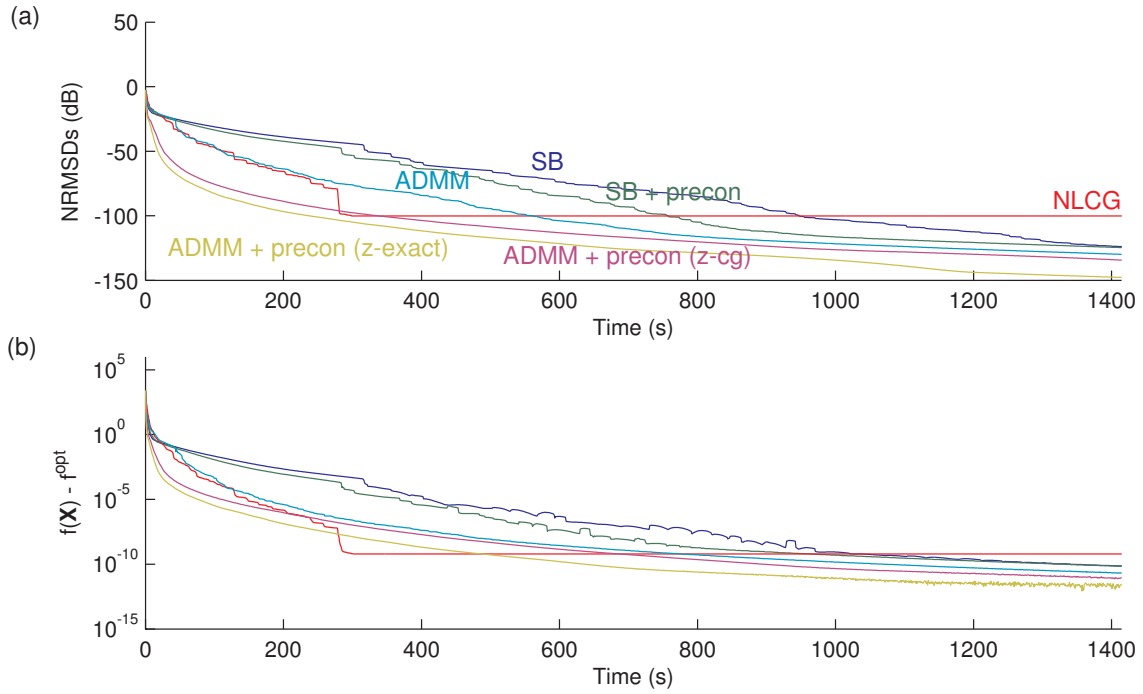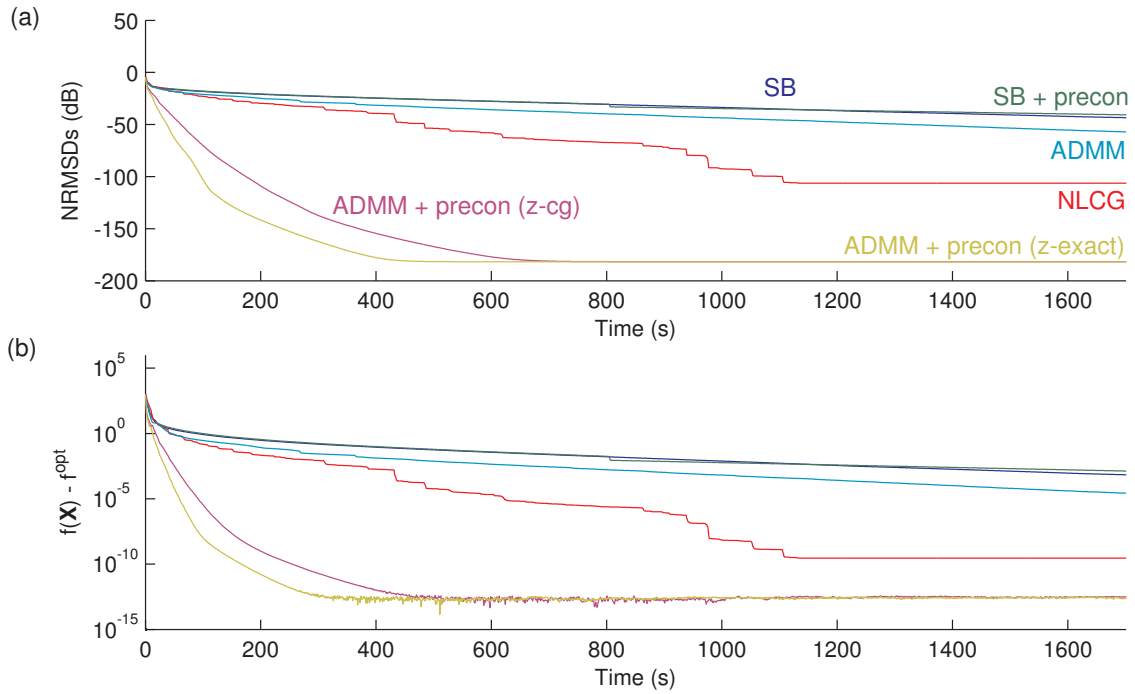
Fig. 3. The (a) NRMSDs (dB) and (b) objective values $f(\mathbf{X}) - f^{\text{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for each of the compared algorithms for the real T$_2$-weighted radial data.


Fig. 4. The (a) NRMSDs (dB) and (b) objective values $f(\mathbf{X}) - f^{\text{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for each of the compared algorithms for the simulated 8-channel spiral data.

REFERENCES

[1] D. S. Weller, S. Ramani, and J. A. Fessler, "Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction," *IEEE Trans. Med. Imag.*, to appear.
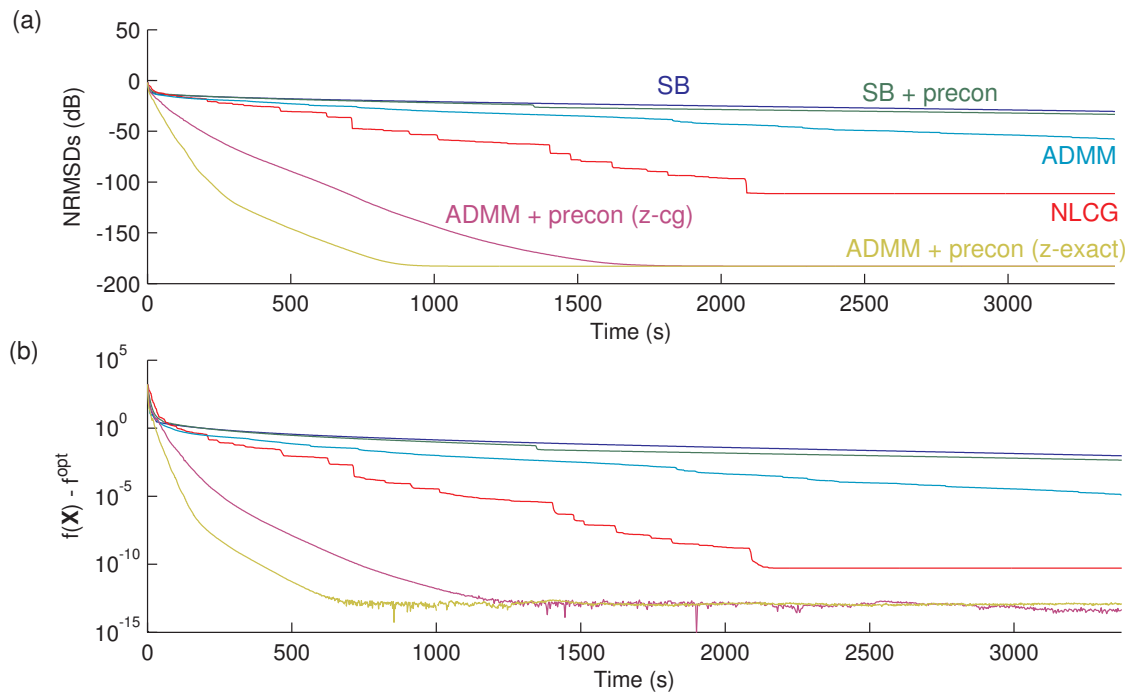
Fig. 5. The (a) NRMSDs (dB) and (b) objective values $f(\mathbf{X}) - f^{\mathrm{opt}}$ relative to the baseline image objective value f^{opt} are plotted versus time for each of the compared algorithms for the simulated 16-channel spiral data.
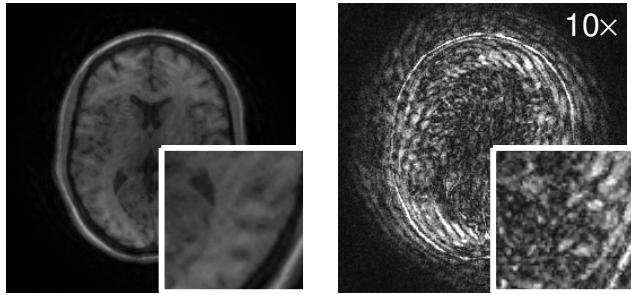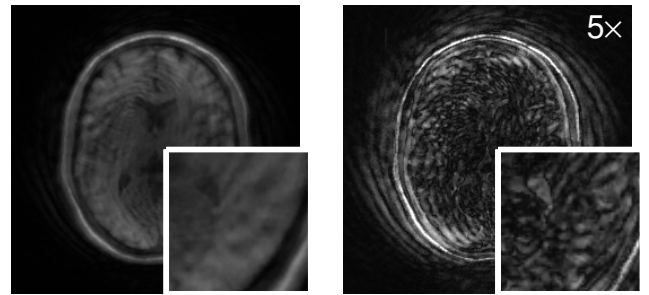
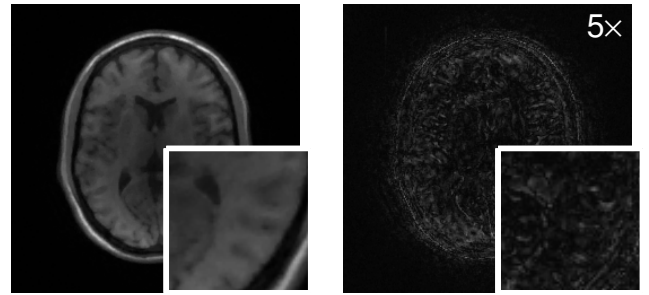(a) NLCG recon.
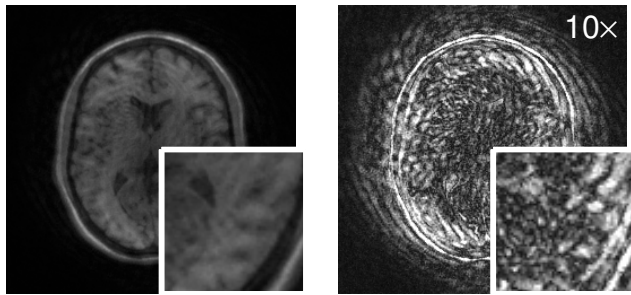
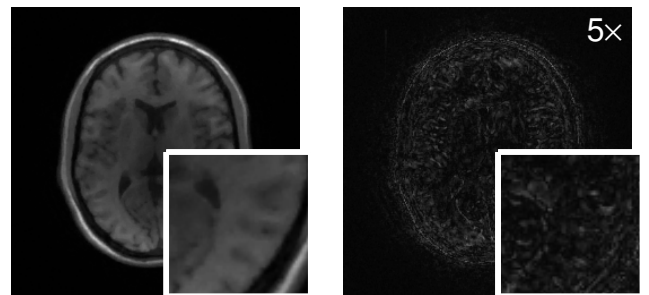(b) Split-Bregman recon.
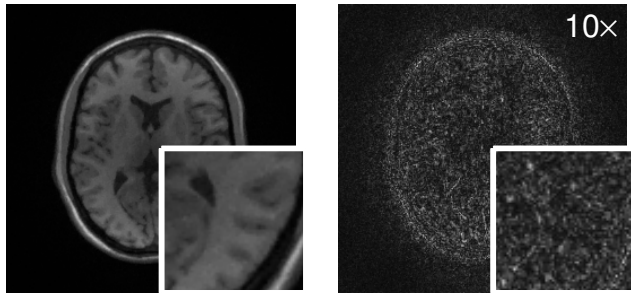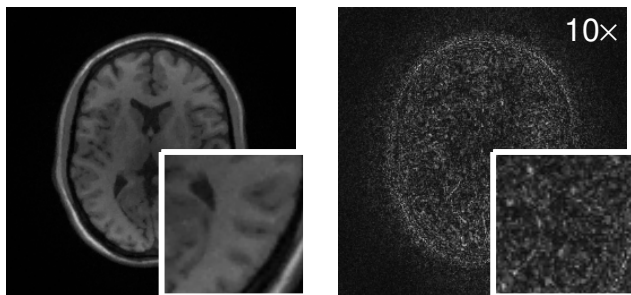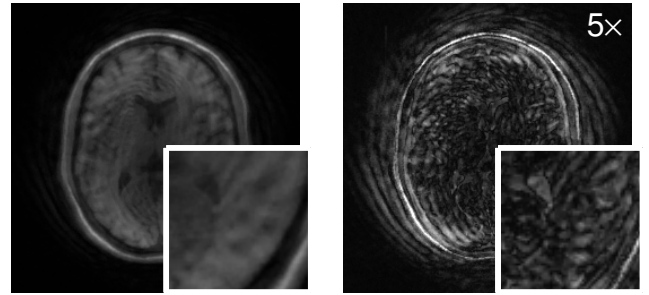
(c) ADMM with PCG for **z**-update.

(d) ADMM with exact **z**-update.

Fig. 6. Sum-of-squared combined reconstructed and difference images for the simulated 8-channel data set after 50 seconds of reconstruction time using (a) NLCG, (b) Split-Bregman with diagonal preconditioning, (c) ADMM with preconditioning both **X** and **z**-updates, and (d) ADMM with preconditioned **X**-updates and exact **z**-updates demonstrate the advantage of rapid convergence on image quality after a fixed amount of time. Zoomed images of the inset region show differences in fine details. The difference images in the right column, relative to the ground truth, are all scaled by $10\times$.

(a) NLCG recon.

(b) Split-Bregman recon.

(c) ADMM with PCG for **z**-update.

(d) ADMM with exact **z**-update.

Fig. 7. Sum-of-squared combined reconstructed and difference images for the simulated 16-channel data set after 100 seconds of reconstruction time using (a) NLCG, (b) Split-Bregman with diagonal preconditioning, (c) ADMM with preconditioning both **X** and **z**-updates, and (d) ADMM with preconditioned **X**-updates and exact **z**-updates demonstrate the advantage of rapid convergence on image quality after a fixed amount of time. Zoomed images of the inset region show differences in fine details. The difference images in the right column, relative to the ground truth, are all scaled by $5\times$.