AN ACCELERATED ITERATIVE REWEIGHTED LEAST SQUARES ALGORITHM FOR COMPRESSED SENSING MRI

Sathish Ramani and Jeffrey A. Fessler

EECS Department, University of Michigan, Ann Arbor, MI 48109-2122, USA.

ABSTRACT

Compressed sensing for MRI (CS-MRI) attempts to recover an object from undersampled k-space data by minimizing sparsity-promoting regularization criteria. The iterative reweighted least squares (IRLS) algorithm can perform the minimization task by solving iteration-dependent linear systems, recursively. However, this process can be slow as the associated linear system is often poorly conditioned for ill-posed problems. We propose a new scheme based on the matrix inversion lemma (MIL) to accelerate the solving process. We demonstrate numerically for CS-MRI that our method provides significant speed-up compared to linear and nonlinear conjugate gradient algorithms, thus making it a promising alternative for such applications.

Index Terms— Compressed sensing, MRI, iterative reweighted least squares, matrix inversion lemma, nonlinear conjugate gradient.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a popular technique that offers great flexibility in imaging soft tissues. However, a fundamental drawback of MRI is its low data-acquisition speed. Emerging trends to reduce MR scan time focus on undersampling k-space and attempting to estimate the underlying object by mathematical modeling. The theory of compressed sensing for MRI (CS-MRI) [1] provides insight into how this can be achieved: The basic assumption is that many MR images are inherently sparse in some transform domain and can be reconstructed with high accuracy from significantly undersampled k-space data by minimizing transform-domain sparsity-promoting regularization criteria subject to data-consistency. This problem is usually handled in a numerical optimization framework using iterative algorithms.

In this work, we concentrate on iterative reweighted least squares (IRLS) algorithms as they are versatile in accommodating multiple convex/nonconvex regularization criteria simultaneously. The IRLS algorithm is a simple technique that performs the minimization task by repetitively solving many linear systems: The key point is, at each iteration i, the associated Hessian matrix, $\mathbf{A}_{(i)}$, depends on the previous iterate, thus making it a nonlinear algorithm on the whole. However, for ill-posed problems, $\mathbf{A}_{(i)}$ may not be well-conditioned due to the "nonsmooth" nature of sparsity-promoting regularization criteria. In such cases, iterative solvers like conjugate gradient (CG) may have slow overall convergence rates. Preconditioning the linear system can alleviate this problem. Since the ideal preconditioner is $\mathbf{A}_{(i)}^{-1}$, we propose to investigate the use of the matrix inversion lemma (MIL) for this problem. We devise a fixed-point solver for the original linear system, which asymptotically becomes equivalent to having a close approximation to $\mathbf{A}_{(i)}^{-1}$ as the preconditioner. We perform various synthetic experiments for CS-MRI and demonstrate numerically that the proposed MIL-based method significantly improves the convergence speed compared to solvers that use conventional preconditioners. We also found that our method outperforms the nonlinear conjugate gradient algorithm. Our approach is fairly general and can be readily extended to treat other inverse problems such as deconvolution.

This work was supported by the Swiss National Science Foundation under fellowship PBELP2-125446 and in part by the National Institutes of Health under grant P01 CA87634.

2. PROBLEM FORMULATION

We consider the standard CS-MRI problem where the measurements are taken in the Fourier domain (k-space) but are undersampled to reduce the scan time. Reconstruction from such partially sampled data is an ill-posed problem and it invariably requires the use of regularization for obtaining meaningful results. The problem is formulated as [1]:

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \mathcal{R}\{\mathbf{u}\} \;\text{ such that }\; \|\mathbf{y} - \mathbf{F}\,\mathbf{u}\|_{\ell_2}^2 < \epsilon^2, \qquad (1)$$

where $\hat{\mathbf{u}}$ and $\mathbf{y}$ are vectors that correspond to the reconstruction and the noisy k-space data, respectively, $\mathbf{F}$ denotes the matrix corresponding to the partially-sampled Fourier transform (after suitable discretization), and $\mathcal{R}$ is a regularization operator. The constraint in (1) ensures that the reconstruction is robust to measurement errors such as noise in the data. The constrained optimization problem (1) can be converted to the following unconstrained form

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \{\|\mathbf{y} - \mathbf{F}\,\mathbf{u}\|_{\ell_2}^2 + \lambda\,\mathcal{R}\{\mathbf{u}\}\},$$

where the regularization parameter λ decides the trade-off between the fidelity of the solution $\mathbf{u}$ to the data $\mathbf{y}$ and the amount of imposed regularization. In practice, λ can be chosen so as to meet the constraint in (1).

The reconstruction quality depends on the choice of $\mathcal{R}$. A classical choice is $\mathcal{R}\{\mathbf{u}\} = \|\mathbf{u}\|_{\ell_2}^2$ which, in the limit $\epsilon \to 0$, leads to the "minimum-energy" reconstruction obtained by first filling the missing k-space samples with zeros and then taking the inverse Fourier transform [2]. However, this approach is highly suboptimal as the resulting output would be

strewn with blurring and aliasing artifacts [2]. Instead, several authors apply the theory of compressed sensing and select $\mathcal{R}$ so as to exploit the assumed sparsity in MR images. A popular choice is $\mathcal{R}\{\mathbf{u}\} = \|\Psi\{\mathbf{u}\}\|_{\ell_1}$ [1, 2], where Ψ is a sparsifying operator such as wavelets. It has also been noted in [1] that including a total-variation (TV) type edge-preserving regularization is beneficial in such problems. We will therefore consider the formulation

$$\begin{aligned} \hat{\mathbf{u}} &= \arg\min_{\mathbf{u}} J_{\lambda_{1,2}}(\mathbf{u}), \qquad (2) \\ J_{\lambda_{1,2}}(\mathbf{u}) &= \|\mathbf{y} - \mathbf{F}\,\mathbf{u}\|_{\ell_2}^2 + \lambda_1 \|\mathbf{W}\mathbf{u}\|_{\ell_1} + \lambda_2\, \mathrm{TV}\{\mathbf{u}\}, \end{aligned}$$

where the matrix $\mathbf{W}$ corresponds to a wavelet transform, and $\mathrm{TV}\{\cdot\}$ is the discretized version of the total-variation norm given by $\mathrm{TV}\{\mathbf{u}\} = \sum_k \sqrt{\sum_{n=1}^{d} |\mathbf{D}_n\mathbf{u}|_k^2}$, where $(\mathbf{u})_k$ represents k-th element of $\mathbf{u}$, d is the number of spatial dimensions of the image $\mathbf{u}$ (d = 2 or 3 in imaging applications), and $\{\mathbf{D}_n\}_{n=1}^d$ are matrices that correspond to finite-differences along the n-th dimension. In this work, we focus on shift-invariant wavelet transforms for $\mathbf{W}$ which are known to exhibit less block-artifacts than orthonormal wavelet transforms. The methods also apply to the orthonormal case.

3. METHODS

We propose to handle (2) in the analysis-form (i.e., where the solution $\mathbf{u}$ is obtained by "analyzing", or penalizing, the coefficients $\mathbf{W}\mathbf{u}$ and $\mathbf{D}_n\mathbf{u}$) since $\{\mathbf{D}_n\}_{n=1}^d$ are not invertible. The associated criterion $J_{\lambda_{1,2}}$, which is convex, can be effectively minimized using iterative gradient-descent-based methods. A prominent method in this category is the nonlinear conjugate gradient (NCG) algorithm which iteratively minimizes $J_{\lambda_{1,2}}$ by traversing along conjugate directions to the minimizer. A line-search is necessary at each iteration to find the optimal step-size along a minimizing direction. A potential alternative to NCG is the iterative reweighted least squares (IRLS) scheme. In the sequel, we give details of the IRLS algorithm applied to (2) and then present our approach for improving its convergence speed.

Iterative Reweighted Least Squares (IRLS) Algorithm

We use the majorize-minimize formulation [3] to describe the IRLS algorithm although it can be shown that the Euler-Lagrange equations or the half-quadratic approaches also lead to an identical method for problem (2).

Consider the auxiliary criterion $J_{\mathrm{AUX}}(\mathbf{u}, \mathbf{u}_{(i)})$ defined at a given iterate $\mathbf{u}_{(i)}$ obtained using [3, Equation (11)]:

$$J_{\mathrm{AUX}}(\mathbf{u}, \mathbf{u}_{(i)}) = \|\mathbf{y} - \mathbf{F}\,\mathbf{u}\|_{\ell_2}^2 + \mathbf{u}^{\mathrm{H}}\,\mathbf{R}^{\mathrm{H}}\boldsymbol{\Theta}_{(i)}\mathbf{R}\,\mathbf{u} + C,$$

where $\mathbf{R} = [\sqrt{\lambda_1}\mathbf{W}^{\mathrm{H}}\ \sqrt{\lambda_2}\mathbf{D}_1^{\mathrm{H}}\ \cdots\ \sqrt{\lambda_2}\mathbf{D}_d^{\mathrm{H}}]^{\mathrm{H}}$, $(\cdot)^{\mathrm{H}}$ is the Hermitian-transpose, C is a constant with respect to $\mathbf{u}$,

$$\begin{aligned} \boldsymbol{\Theta}_{(i)} &= \mathrm{diag}\{\boldsymbol{\Theta}_{\mathbf{W}_{(i)}}\ \boldsymbol{\Theta}_{\mathrm{TV}_{(i)}}\ \cdots\ \boldsymbol{\Theta}_{\mathrm{TV}_{(i)}}\}, \qquad (3) \\ \boldsymbol{\Theta}_{\mathbf{W}_{(i)}} &= \mathrm{diag}\left\{\left(2\sqrt{|\mathbf{W}\,\mathbf{u}_{(i)}|_k^2 + \mu}\right)^{-1}\right\}, \qquad (4) \\ \boldsymbol{\Theta}_{\mathrm{TV}_{(i)}} &= \mathrm{diag}\left\{\left(2\sqrt{\textstyle\sum_{n=1}^{d} |\mathbf{D}_n\,\mathbf{u}_{(i)}|_k^2 + \mu}\right)^{-1}\right\}, \qquad (5) \end{aligned}$$

and where μ is a small positive constant that is commonly added to ℓ_1-type analysis-priors in $J_{\lambda_{1,2}}$ to increase numerical stability of corresponding minimization algorithms. It can be shown that $J_{\mathrm{AUX}}(\mathbf{u}, \mathbf{u}_{(i)})$ majorizes $J_{\lambda_{1,2}}(\mathbf{u})$, i.e., $J_{\lambda_{1,2}}(\mathbf{u}) < J_{\mathrm{AUX}}(\mathbf{u}, \mathbf{u}_{(i)})\ \forall\, \mathbf{u} \neq \mathbf{u}_{(i)}$ and $J_{\lambda_{1,2}}(\mathbf{u}_{(i)}) = J_{\mathrm{AUX}}(\mathbf{u}_{(i)}, \mathbf{u}_{(i)})$, so that minimizing J_{AUX} decreases $J_{\lambda_{1,2}}$ (this however holds strictly only when $\mu > 0$ is accounted for in $J_{\lambda_{1,2}}$). Since J_{AUX} is quadratic in $\mathbf{u}$, it can be minimized by setting its gradient with respect to $\mathbf{u}$ to zero which leads to the following linear system of equations

$$\mathbf{A}_{(i)}\,\mathbf{u}_{(i+1)} = \mathbf{F}^{\mathrm{H}}\mathbf{y} \qquad (6)$$

that must be solved to obtain the next iterate $\mathbf{u}_{(i+1)}$, where $\mathbf{A}_{(i)} = \mathbf{F}^{\mathrm{H}}\mathbf{F} + \mathbf{R}^{\mathrm{H}}\,\boldsymbol{\Theta}_{(i)}\,\mathbf{R}$.

Standard Approach for Solving (6)

In imaging applications, (6) cannot be solved exactly due to the large dimensionality of $\mathbf{u}$. A fairly standard practice therefore is to run a few conjugate gradient (CG) iterations on (6) to partially solve for $\mathbf{u}_{(i+1)}$. However, because $\mathbf{W}$ and $\{\mathbf{D}_n\}_{n=1}^d$ are sparsifying operators, some values $(\mathbf{W}\mathbf{u}_{(i)})_k$ and $(\mathbf{D}_n\mathbf{u}_{(i)})_{n,k}$ may be small, and so $\mathbf{A}_{(i)}$ may be poorly conditioned due to (4) and (5). A preconditioned CG (PCG) algorithm is frequently employed to accelerate the convergence. Some common preconditioners include the Jacobi diagonal preconditioner $\mathbf{M}_{\mathrm{diag}} = \mathrm{diag}\{1/(\mathbf{A}_{(i)})_{kk}\}$, and the preconditioner $\mathbf{M}_{\mathrm{circ}} = (\mathbf{F}^{\mathrm{H}}\mathbf{F} + \mathbf{R}^{\mathrm{H}}\,\mathbf{R})^{-1}$ (obtained by discarding $\boldsymbol{\Theta}_{(i)}$ in $\mathbf{A}_{(i)}$) which turns out to be a circulant matrix when sample locations are specified on a cartesian grid in k-space and when $\mathbf{R}^{\mathrm{H}}\mathbf{R}$ is circulant (e.g., the case where $\mathbf{W}$ is shift-invariant). The resulting IRLS-PCG algorithm is

Algorithm 1: IRLS-PCG algorithm for minimizing $J_{\lambda_{1,2}}$
Step 1: Initial estimate $= \mathbf{u}_{(0)}$; $i = 0$
Repeat Steps 2 - 3 **until Stop Criterion is met**
Step 2: Update $\boldsymbol{\Theta}_{(i)}$ using $\mathbf{u}_{(i)}$ in (3) - (5)
Step 3: Run few PCG iterations on (6) to get $\mathbf{u}_{(i+1)}$; Set $i = i+1$

Proposed Scheme for Solving (6)

Since the major task in IRLS is to solve (6), we propose to apply matrix inversion lemma (MIL) to $\mathbf{A}_{(i)}^{-1}$. Because $\mathbf{F}$ is not invertible, we propose the following matrix-splitting scheme

$$\mathbf{u}_{(i+1,j+1)} = \mathbf{B}_{(i)}^{-1}(\mathbf{F}^{\mathrm{H}}\mathbf{y} + (\mathbf{C}_{(i)} - \mathbf{F}^{\mathrm{H}}\mathbf{F})\mathbf{u}_{(i+1,j)}), \qquad (7)$$

where $\mathbf{B}_{(i)} = (\mathbf{C}_{(i)} + \mathbf{R}^{\mathrm{H}}\,\boldsymbol{\Theta}_{(i)}\,\mathbf{R})$, and $\mathbf{C}_{(i)}$ is chosen to be an invertible matrix such that $\mathbf{C}_{(i)} \succ \mathbf{F}^{\mathrm{H}}\mathbf{F}$. The above iteration on j is bound to converge to the solution of (6) since the spectral radius $\varrho(\mathbf{B}_{(i)}^{-1}(\mathbf{C}_{(i)} - \mathbf{F}^{\mathrm{H}}\mathbf{F})) < 1$ whenever $\mathbf{C}_{(i)} \succ \mathbf{F}^{\mathrm{H}}\mathbf{F}$ [4, Theorem 3.2.7, p. 48]. We now apply MIL to $\mathbf{B}_{(i)}^{-1}$ to get $\mathbf{B}_{(i)}^{-1} = \mathbf{C}_{(i)}^{-1} - \mathbf{C}_{(i)}^{-1}\mathbf{R}^{\mathrm{H}}\,\mathbf{G}_{(i)}^{-1}\mathbf{R}\ \mathbf{C}_{(i)}^{-1}$, where $\mathbf{G}_{(i)} = \boldsymbol{\Theta}_{(i)}^{-1} + \mathbf{R}\,\mathbf{C}_{(i)}^{-1}\mathbf{R}^{\mathrm{H}}$. Therefore, (7) becomes

$$\begin{aligned} \mathbf{u}_{(i+1,j+1)} &= \mathbf{b}_{(i+1,j)} - \mathbf{C}_{(i)}^{-1}\,\mathbf{R}^{\mathrm{H}}\,\mathbf{v}_{(i+1,j)}, \qquad (8) \\ \mathbf{G}_{(i)}\,\mathbf{v}_{(i+1,j)} &= \mathbf{R}\,\mathbf{b}_{(i+1,j)}, \qquad (9) \end{aligned}$$

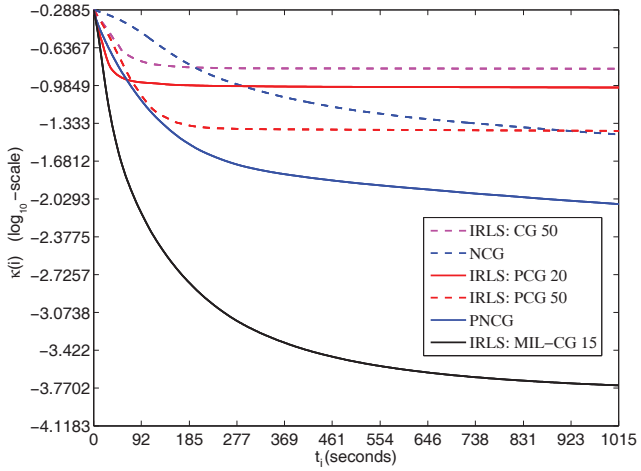


Fig. 1. Experiment 1 (SL phantom): Plot of $\kappa(i)$ versus time t_i.

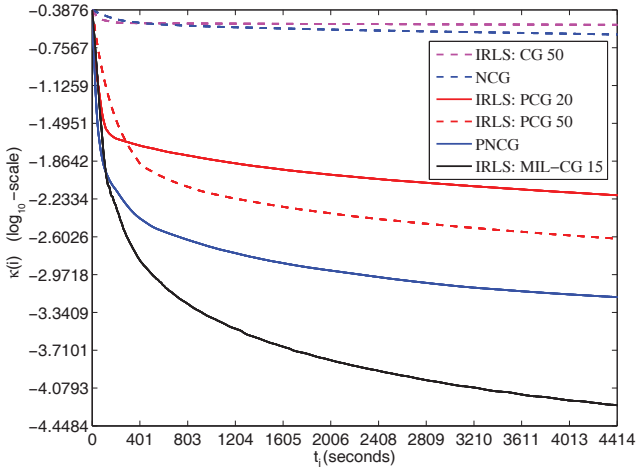


Fig. 2. Experiment 2 (T1-MR image): Plot of $\kappa(i)$ versus time t_i.

where $\mathbf{b}_{(i+1,j)} = \mathbf{C}_{(i)}^{-1}\mathbf{F}^{\mathrm{H}}\mathbf{y} + (\mathbf{I} - \mathbf{C}_{(i)}^{-1}\mathbf{F}^{\mathrm{H}}\mathbf{F})\mathbf{u}_{(i+1,j)}$. An advantage of this approach is that $\mathbf{G}_{(i)}$ depends on $\mathbf{\Theta}_{(i)}^{-1}$ rather than $\mathbf{\Theta}_{(i)}$, which provides the option to do away with the undesirable constant μ in (3)-(5). Based on (7)-(9), we present our IRLS-MIL algorithm below where we have assumed that $\mathbf{C}_{(i)}$ is chosen *a priori* for all iterations i.

Algorithm 2: IRLS-MIL algorithm for minimizing $J_{\lambda_{1,2}}$
Step 1: Initial estimate $= \mathbf{u}_{(0)}$; $i = 0$
While Stop Criterion for i-loop **is not met**
Step 2: Update $\mathbf{\Theta}_{(i)}^{-1}$ using $\mathbf{u}_{(i)}$; Set $j = 0$, $\mathbf{u}_{(i+1,0)} = \mathbf{u}_{(i)}$
While Stop Criterion for j-loop **is not met**
Step 3: Compute $\mathbf{b}_{(i+1,j)}$
Step 4: Run few CG iterations on (9) to get $\mathbf{v}_{(i+1,j)}$
Step 5: Update $\mathbf{u}_{(i+1,j+1)}$ using (8); Set $j = j+1$
end of j-loop
Step 6: Set $i = i + 1$
end of i-loop

IRLS-MIL is certainly more involved than IRLS-PCG both in terms of computation (products with $\mathbf{C}_{(i)}^{-1}$) and storage requirements (for handling $\mathbf{v}_{(i+1,j)}$). However, we hope that using (7) will provide faster convergence than IRLS-(P)CG since $\varrho(\mathbf{B}_{(i)}^{-1}(\mathbf{C}_{(i)} - \mathbf{F}^{\mathrm{H}}\mathbf{F})) \to 0$ when $\mathbf{C}_{(i)} \to \mathbf{F}^{\mathrm{H}}\mathbf{F}$. Alternatively, the fixed point of (7) satisfies $\mathbf{B}_{(i)}^{-1}\mathbf{A}_{(i)}\mathbf{u}_{(i+1)} = \mathbf{B}_{(i)}^{-1}\mathbf{F}^{\mathrm{H}}\mathbf{y}$. Thus, using (7) is equivalent to having $\mathbf{B}_{(i)}^{-1}$ as the preconditioner for solving (6), which approaches the ideal preconditioner $\mathbf{A}_{(i)}^{-1}$ as $\mathbf{C}_{(i)} \to \mathbf{F}^{\mathrm{H}}\mathbf{F}$.

Choice of C

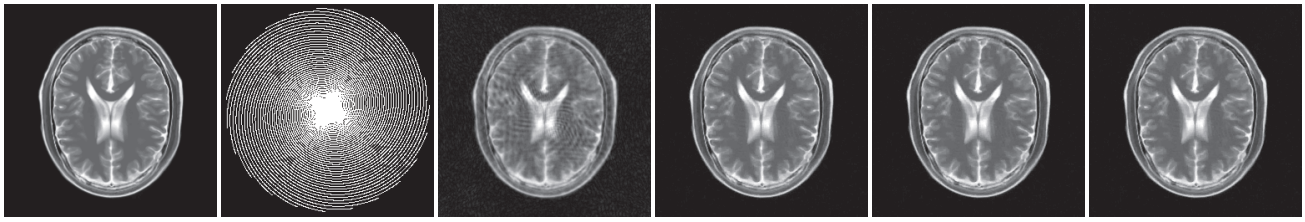
In CS-MRI, $\mathbf{F}^{\mathrm{H}}\mathbf{F}$ has a block-Toeplitz structure which can be exploited to write $\mathbf{F}^{\mathrm{H}}\mathbf{F}\mathbf{u} = \mathbf{Z}^{\mathrm{H}}\mathbf{Q}\mathbf{Z}\mathbf{u}$, where the action of $\mathbf{Z}$ on $\mathbf{u}$ is to pad $\mathbf{u}$ with zeros to obtain a vector that is twice long as $\mathbf{u}$ and $\mathbf{Q}$ is a circulant matrix [5]. Then, choosing $\mathbf{C}_{(i)} = \mathbf{Z}^{\mathrm{H}}\mathbf{Q}\mathbf{Z} + \alpha\mathbf{I}$ ensures $\mathbf{C}_{(i)} \succ \mathbf{F}^{\mathrm{H}}\mathbf{F}$ for $\alpha > 0$. This form of $\mathbf{C}_{(i)}$ can also be efficiently inverted using the technique proposed in [6]. However, caution must be exercised in selecting α because, as $\alpha \to 0$, $\mathbf{C}_{(i)}^{-1}$ will become numerically unstable. On the other hand, when $\alpha \gg 0$, $\varrho(\mathbf{B}_{(i)}^{-1}(\mathbf{C}_{(i)} - \mathbf{F}^{\mathrm{H}}\mathbf{F})) \to 1$, resulting in slow convergence of IRLS-MIL. Therefore, the key to the successful application of IRLS-MIL relies in selecting α such that $\mathbf{C}_{(i)} \approx \mathbf{F}^{\mathrm{H}}\mathbf{F}$ while ensuring a numerically stable $\mathbf{C}_{(i)}^{-1}$.

In this paper, we restrict our focus to sampling patterns that are confined to a cartesian grid. Then, $\mathbf{F}^{\mathrm{H}}\mathbf{F}$ becomes circulant, in which case we select $\mathbf{C}_{\mathrm{circ}_{(i)}} = \mathbf{F}^{\mathrm{H}}\mathbf{F} + \alpha\mathbf{R}^{\mathrm{H}}\mathbf{R}$ since $\mathbf{C}_{\mathrm{circ}_{(i)}}$ becomes a circulant approximation to $\mathbf{A}_{(i)}$ (when $\mathbf{W}$ is shift-invariant) and can be inverted efficiently using FFTs. We found that $\mathbf{C}_{\mathrm{circ}_{(i)}}$ with $\alpha \in [0.001,\ 0.02]$ provided excellent results in our experiments.

4. EXPERIMENTS

We performed three experiments, one with each of the following 256×256 synthetic images: Shepp-Logan (SL) phantom, T1- and T2-MR images obtained from the Brainweb database [7]. We considered three different k-space sampling schemes (confined to a cartesian grid): radial sampling (18 radial lines, 93% undersampling), variable-density random sampling (80% undersampling), and variable-density multi-shot spiral (70% undersampling), one for each test image, respectively. In each experiment, complex additive white Gaussian noise was added to simulate noisy data with 30 dB SNR. For the experiment with SL phantom, we only considered the TV-regularization (i.e., we set $\lambda_1 = 0$ in $J_{\lambda_{1,2}}$) since the phantom already has a sparse representation in terms of its image-gradient. However, for experiments with MR images, we used both regularization terms taking care to set $\lambda_{1,2}$ to obtain visually appealing reconstructions. For $\mathbf{W}$, we used 2 levels of the undecimated Haar wavelet transform excluding the 'scaling' coefficients.

In all experiments, we compared the following variations of NCG and IRLS algorithms: (i) **NCG** (no preconditioning), (ii) **PNCG** (preconditioned NCG), (iii) **IRLS-CG** $\boldsymbol{N}$ (N CG iterations at Step 3 of Algorithm 1), (iv) **IRLS-PCG** $\boldsymbol{N}$ (N preconditioned CG iterations at Step 3 of Algorithm 1), (v) **IRLS-MIL-CG** $\boldsymbol{N}$ (N CG iterations at Step 5 of Algorithm 2 with only one iteration of the j-loop). We used $\mathbf{M}_{\mathrm{circ}}$ as the preconditioner since $\mathbf{M}_{\mathrm{diag}}$ did not provide any significant improvement. NCG algorithms used the line-search (with 3 iterations) described in [8] that guarantees monotonic de-



(a) (b) (c) (d) (e) (f)

Fig. 3. Reconstruction results for Experiment 3: (a) Original noise-free T2-MR image, (b) multi-shot spiral scheme (70% undersampling), (c) minimum-energy reconstruction, (d) $\mathbf{u}_{(\infty)}$ the actual solution of (2), (e) IRLS-MIL output ($\kappa(\cdot) = 10^{-3}$ after $t_i = 253$ seconds), (f) NCG output ($\kappa(\cdot) = 10^{-3}$ after $t_i = 1928$ seconds). As expected, (e) and (f) are virtually identical to (d) indicating that the outputs of IRLS-MIL and NCG have converged to $\mathbf{u}_{(\infty)}$, but IRLS-MIL has done so nearly 7.5 times faster than NCG.

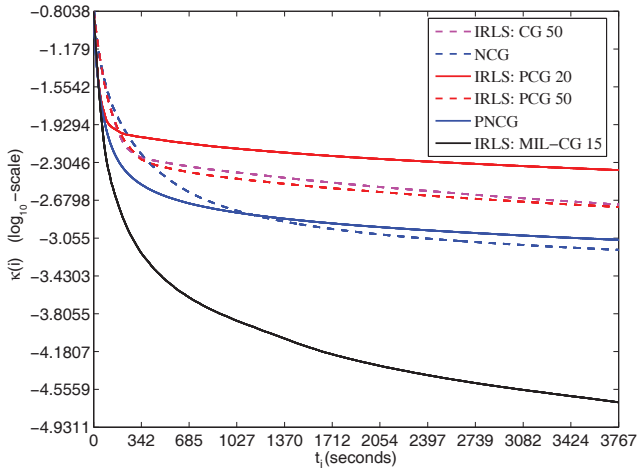


Fig. 4. Experiment 3 (T2-MR image): Plot of $\kappa(i)$ versus time t_i.

crease of $J_{\lambda_{1,2}}$. The parameter μ in (4)-(5) was set to 10^{-10} for the algorithms based on (P)NCG and IRLS-(P)CG while it was set to zero for IRLS-MIL. The algorithms were implemented in MATLAB and the experiments were performed on a PC with 3.4 GHz Intel Xeon processor. Since all algorithms minimized the same $J_{\lambda_{1,2}}$ (which decides the reconstruction quality), we focused on their convergence speed rather than their SNR improvement. Therefore, for all algorithms, we examined the normalized ℓ_2-distance between $\mathbf{u}_{(i)}$ and $\mathbf{u}_{(\infty)}$ given by $\kappa(i) = \|\mathbf{u}_{(i)} - \mathbf{u}_{(\infty)}\|_{\ell_2} / \|\mathbf{u}_{(\infty)}\|_{\ell_2}$, where $\mathbf{u}_{(\infty)}$ (corresponds to the limiting case $i \to \infty$) represents the actual solution of (2); it was obtained in each experiment by running hundreds of iterations of the IRLS-PCG 500 algorithm. The $\mathbf{u}_{(\infty)}$ solution was numerically verified by separately running several thousands of iterations of the PNCG algorithm.

Since the algorithms have different computational loads per iteration, we evaluated $\kappa(i)$ as a function of algorithm run-time t_i (time elapsed from start till iteration i). Figures 1, 2 and 4 plot $\kappa(i)$ for various algorithms as a function of t_i for the three experiments, respectively. The preconditioned algorithms, PNCG and IRLS-PCG, generally exhibit some improvement over their unpreconditioned counterparts. However, the proposed IRLS-MIL scheme demonstrates a clear lead in convergence speed over all IRLS and (P)NCG algorithms considered in this work. Figure 3 displays $\mathbf{u}_{(\infty)}$ and the outputs of IRLS-MIL and NCG (which is slightly faster than PNCG) for Experiment 3. As expected, $\mathbf{u}_{(\infty)}$ and the IRLS-MIL and NCG images are visually indistinguishable because these algorithms converge to $\mathbf{u}_{(\infty)}$. However, the proposed IRLS-MIL does so much sooner than NCG.

5. CONCLUSIONS AND DISCUSSION

Iterative reweighted least squares (IRLS) algorithms find the minimizer of a nonquadratic cost criterion by attempting to solve a weighted-linear system at each iteration. We proposed the use of matrix inversion lemma (MIL) for this task and demonstrated based on numerical experiments that our scheme—IRLS-MIL—converges to the true minimizer significantly faster than some popular methods like nonlinear conjugate gradient (NCG) for the problem of compressed sensing MRI. The proposed method can also be directly applied to other inverse problems that are based on shift-invariant data-acquisition models, e.g., deconvolution. Expanding the applicability of IRLS-MIL to nonconvex priors [2] and shift-variant imaging systems is a research direction that we hope to pursue in the future.

6. REFERENCES

[1] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[2] J. Trzasko and A. Manduca, "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic ℓ_0-Minimization," *IEEE Trans. Medical Imaging*, vol. 28, no. 1, pp. 106–121, January 2009.

[3] S. Ramani, P. Thévenaz, and M. Unser, "Regularized Interpolation for Noisy Data," *Proc. ISBI*, pp. 612–615, 2007.

[4] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer-Verlag, New York, 1994.

[5] F. Wajer and K. P. Pruessmann, "Major speedup of reconstruction for sensitivity encoding with arbitrary trajectories," *Proc. ISMRM*, p. 767, 2001.

[6] A. E. Yagle, "New fast preconditioners for toeplitz-like linear systems," *Proc. ICASSP*, pp. 1365–1368, 2002.

[7] "Brainweb: Simulated MRI volumes for normal brain," McConnell Brain Imaging Centre, http://www.bic.mni.mcgill.ca/brainweb/.

[8] J. A. Fessler and S. D. Booth, "Conjugate-Gradient Preconditioning Methods for Shift-Variant PET Image Reconstruction," *IEEE Trans. Image Processing*, vol. 8, no. 5, pp. 688–699, May 1999.