UCSF
UC San Francisco Previously Published Works

Title
A maximum likelihood approach to parallel imaging with coil sensitivity noise.

Permalink
https://escholarship.org/uc/item/5v82k067

Journal
IEEE transactions on medical imaging, 26(8)

ISSN
0278-0062

Authors
Raj, Ashish
Wang, Yi
Zabih, Ramin

Publication Date
2007-08-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A Maximum Likelihood Approach to Parallel Imaging With Coil Sensitivity Noise

Ashish Raj, Yi Wang, and Ramin Zabih*

***Abstract*—Parallel imaging is a powerful technique to speed up magnetic resonance (MR) image acquisition via multiple coils. Both the received signal of each coil and its sensitivity map, which describes its spatial response, are needed during reconstruction. Widely used schemes such as SENSE assume that sensitivity maps of the coils are noiseless while the only errors are in coil outputs. In practice, however, sensitivity maps are subject to a wide variety of errors. At first glance, sensitivity noise appears to result in an errors-in-variables problem of the kind that is typically solved using total least squares (TLSs). However, existing TLS algorithms are in general inappropriate for the specific type of block structure that arises in parallel imaging. In this paper, we take a maximum likelihood approach to the problem of parallel imaging in the presence of independent Gaussian sensitivity noise. This results in a quasi-quadratic objective function, which can be efficiently minimized. Experimental evidence suggests substantial gains over conventional SENSE, especially in nonideal imaging conditions like low signal-to-noise ratio (SNR), high g-factors and large acceleration, using sensitivity maps suffering from misalignment, ringing, and random noise.**

***Index Terms*—Magnetic resonance (MR), maximum likelihood (ML), parallel imaging, SENSE, total least squares (TLSs).**

I. Parallel Imaging and Sensitivity Noise

PARALLEL imaging involves the use of multiple coils to substantially reduce scan time (and thus motion artifacts) by under-sampling data in k-space [1]. These techniques are known as SMASH [2]–[4]; SENSE [5]–[8]; or GRAPPA [9], [10]. They are closely related to each other [11], [12]. Mathematically, SENSE is the exact reconstruction method, and will be the focus on this work. Aliased data are combined using the coils' sensitivity maps to reconstruct a full, unaliased image. SENSE works superbly in well-behaved situations with high signal-to-noise ratio (SNR) and low g-factors, but starts to deteriorate under nonideal conditions. Some of these issues were highlighted in [12].

Manuscript received June 14, 2006; revised February 26, 2007. *Asterisk indicates corresponding author.*

A. Raj is with the Radiology Department, Weill Medical College of Cornell University, Ithaca NY 14853 USA and also with Department of Radiology, University of California, San Francisco, CA 94143 USA.

Y. Wang is with the Radiology Department, Weill Medical College of Cornell University, Ithaca, NY 14853 USA.

*R. Zabih is with the Radiology Department, Weill Medical College of Cornell University, Ithaca NY 14853 USA and also with the Computer Science Department, Cornell University, Ithaca NY 14853 USA (e-mail: rdz@cs.cornell.edu).

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

Digital Object Identifier 10.1109/TMI.2007.897364

This paper addresses a major source of errors in SENSE: the situation where it is difficult to obtain artifact-free sensitivity maps. Now coil sensitivities are subject to noise processes since they are obtained from MR data, whether from uniform phantom, after division by a body coil image, or from low-frequency calibration lines. In addition, the encoding and decoding sensitivities are not identical in practice due to physiological motion, misalignment of coils between scans, etc. Under modest acceleration, low g-factors and high SNR, SENSE does not suffer greatly from these effects, but produces disturbing artifacts when these ideal conditions do not hold. We propose a new reconstruction which is tolerant to unreliable sensitivity information due to low SNR, misalignment caused by coil or physiological movement, poor signal penetration, high g-factors in interior regions, etc. This may potentially open up SENSE method to situations where it is currently unsatisfactory.

A. Organization and Notation

A synopsis of our approach and discussion of related work follows next. The parallel MR acquisition model is detailed in Section II for general and special (Cartesian) case, and our sensitivity noise model is introduced. Section III derives our algorithm using maximum likelihood (ML) principles. Next we show that with Cartesian sampling the general solution reduces to a quasi-quadratic minimization problem. We give experimental results on both simulated and clinical data in Section IV.

Scalars, vectors, and 1-D objects are denoted in lower case; matrices and 2-D objects in upper case. Vectors and matrices are in boldface. Unitary or binary scalar operations applied on vectors or matrices are implicitly element-by-element. For instance "$\mathbf{x} \cdot \mathbf{y}$" is understood to be element-wise multiply (not the dot product which we denote by $\mathbf{x}^T\mathbf{y}$). The notation $\mathrm{diag}(\mathbf{x})$ represents a diagonal matrix whose diagonal elements are given by those of the enclosed vector $\mathbf{x}$. $\mathbf{I}$ is the identity matrix; boldface $\mathbf{1}$ the vector of ones. $\mathbf{Q}[i,j]$ denotes the (i,j)th element of matrix $\mathbf{Q}$; $\mathbf{q}[i]$ the ith element of vector $\mathbf{q}$.

B. Synopsis of Our Approach

The parallel imaging process has a linear form

$$\bar{\mathbf{y}} = \bar{\mathbf{E}}\bar{\mathbf{x}} + \bar{\mathbf{n}}, \quad \mathbf{y} = \mathbf{E}\mathbf{x} + \mathbf{n}. \tag{1}$$

The left equation refers to k-space quantities, and the right one image-space. Coil outputs are denoted by $\bar{\mathbf{y}}$ and $\mathbf{y}$, the desired image by $\bar{\mathbf{x}}$ and $\mathbf{x}$, and imaging noise by $\bar{\mathbf{n}}$ and $\mathbf{n}$. Matrix $\bar{\mathbf{E}}$ contains sensitivity and reduced encoding information.[1]

[1]We will denote k-space objects by $\bar{\mathbf{x}}$, and image-space by $\mathbf{x}$.

0278-0062/$25.00 © 2007 IEEE

SENSE takes a least squares approach which gives the ML estimate [13, Ch. 15] under independent and identically distributed (i.i.d.) Gaussian noise.[2] In this paper, we extend the noise model as follows:

$$\mathbf{y} = (\mathbf{E} + \Delta\mathbf{E})\mathbf{x} + \mathbf{n} \tag{2}$$

where $\Delta\mathbf{E}$ is the noise in the system matrix that results from errors in the sensitivity maps (i.e., sensitivity noise). We propose a ML approach to solving (2), which generalizes both least squares [13] and total least squares [15]. First we derive the general result for arbitrary noise models and sampling trajectories, then develop practical algorithms for specific situations.

While the general result is valid for any noise model, for practicality we use a natural model whereby sensitivity maps are corrupted by independent (but possibly non-i.i.d.) noise. This models many actual imaging situations where sensitivity maps suffer from uncorrelated but spatially varying noise. This algorithm, which we call ML-SENSE, was first introduced with preliminary data in [16]. ML-SENSE gives strong results, even in cases where this noise model is inaccurate. We show examples of spatially correlated sensitivity noise which are effectively mitigated by our method. This suggests that the independent noise assumption, while not completely adequate, is still much better than the conventional assumption of *zero* sensitivity errors. This is not surprising—for years workers in signal processing, radar systems, and mobile communications, for examples, have used independent Gaussian models to great effect, even in cases where they are demonstrably inaccurate. Note also that SENSE too is optimal only for i.i.d. additive Gaussian noise, but has been profitably employed in non-i.i.d. situations.

C. Relation to Prior Work

To some extent, the issue of poor SNR was addressed using regularization [17]–[19]. At first glance, (2) appears to be an errors-in-variables problem of the kind commonly addressed with total least squares (TLS) [15]. Indeed, several authors, such as [20], have suggested taking a TLS approach to sensitivity error. Classical TLS theory [21] applied on (2) attempts to find a solution that minimizes both the additive noise $\mathbf{n}$ as well as the error-in-variables $\triangle\mathbf{E}$

$$\hat{\mathbf{x}}_{\mathrm{TLS}} = \arg\min_{\mathbf{x}} \|[\triangle\mathbf{E}|\mathbf{n}]\|_F, \quad \text{where } \mathbf{n} + \triangle\mathbf{E}\mathbf{x} = \mathbf{y} - \mathbf{E}\mathbf{x}$$

where the indicated norm is Frobenius. The TLS algorithm requires the computation of singular value decomposition (SVD) of the full system matrix, which is computationally prohibitive in most cases. Further, it assumes that the elements of $\triangle\mathbf{E}$ are independent (i.e., that $\triangle\mathbf{E}$ has no structure). Unfortunately, the independence assumption is generally invalid due to the specific structure of the system matrix $\mathbf{E}$, as we show in Section II-D. In Cartesian sampling $\triangle\mathbf{E}$ has a diagonal block structure, as shown in Fig. 2. So even if the underlying sensitivity noise process is uncorrelated, the elements of $\triangle\mathbf{E}$ are never independent (off-diagonal entries being identically zero). A similar situation occurs in k-space—$\triangle\bar{\mathbf{E}}$ has a Toeplitz-type structure [Fig. 2(b)], making the elements of $\triangle\bar{\mathbf{E}}$ algebraically related rather than independent.

[2]See [14] for a study of noise in medical imaging.

The only instance where TLS can actually be used is for direct unfolding of aliasing voxels under Cartesian sampling. In this case, the problem decouples into independent $L \times R$ subproblems. This approach was suggested in [20]. Unfortunately, there is no k-space equivalent of this approach; nor does it extend to nonuniform noise models. Further, it is frequently preferable to reconstruct the entire image together, for instance to exploit some *a priori* knowledge. Non-Cartesian data too must be reconstructed over the entire image. TLS cannot be directly applied to these situations. The proposed approach removes these limitations by extending and generalizing TLS. It dos not involve expensive SVD computations and is statistically optimal (in ML sense) under a large class of noise models.

Several generalizations of TLS, collectively known as *Constrained* TLS (CTLS), have been proposed to handle *linearly structured* matrices—those matrices that can be obtained from a linear combinations of a smaller perturbation vector [22].[3] Define an augmented matrix $\mathbf{C} = [\mathbf{E}|\mathbf{y}]$, and a perturbation in $\mathbf{C}$ as $\triangle\mathbf{C} = [\triangle\mathbf{E}|\mathbf{n}]$. CTLS consists of solving

$$\min_{\mathbf{v},\mathbf{x}} \|\mathbf{v}\|, \text{ subject to } (\mathbf{C} + \triangle\mathbf{C})\begin{bmatrix}\mathbf{x}\\ -1\end{bmatrix} = 0$$
$$\text{AND } \triangle\mathbf{C} = [\mathbf{F}_1\mathbf{v}|\mathbf{F}_2\mathbf{v}|\cdots|\mathbf{F}_{N+1}\mathbf{v}]$$

where the $\mathbf{F}_i$'s are matrices that generate the elements of $\triangle\mathbf{C}$ from $\mathbf{v}$. This problem is difficult to solve for arbitrary $\mathbf{E}$, requiring slow general-purpose constrained minimization techniques. However, by taking advantage of the particular structure of the system matrix, our approach uses much more efficient special-purpose unconstrained minimization methods. Further details of TLS methods in MR can be found in [25, Ch. 3].

II. MR Acquisition Model

A. System Model

System matrices $\mathbf{E}$ and $\bar{\mathbf{E}}$ represent a concatenation over all coils of the discretized encoding operator which acts on the input image vector $\mathbf{x}$ and k-space vector $\bar{\mathbf{x}}$, respectively. The vector $\mathbf{x}$ is a discrete representation of the desired MR image $X(\mathbf{r})$, where $\mathbf{r}$ is the 2-D spatial index. The parallel imaging process for each coil $l \in \{1, \ldots, L\}$ can be summarized by Fig. 1, where Y_l is the aliased (folded) image seen by the lth coil, and S_l is its sensitivity response. Let the 2-D vectors $\mathbf{k}$ and $\mathbf{r}$ be points in k-space and image-space, respectively. The raw data from individual coils in k-space are $\bar{Y}_l(\mathbf{k})$. Then

$$\bar{Y}_l(\mathbf{k}) = \int d\mathbf{r} e^{-i2\pi\mathbf{r}^T\mathbf{k}} S_l(\mathbf{r}) X(\mathbf{r}). \tag{3}$$

Following [6], the Fourier transform above can be replaced by 2-D-FT $\mathcal{F}$ via Dirac distributions sampled at spatial index ρ

$$\bar{Y}_l(\mathbf{k}) = \mathcal{F}\left[\sum_{\rho} S_l(\mathbf{r}_\rho) X(\mathbf{r}_\rho)\delta(\mathbf{r} - \mathbf{r}_\rho)\right](\mathbf{k}). \tag{4}$$

[3]An earlier approach, called *Structured* TLS [23], was shown to be equivalent to CTLS in [24].

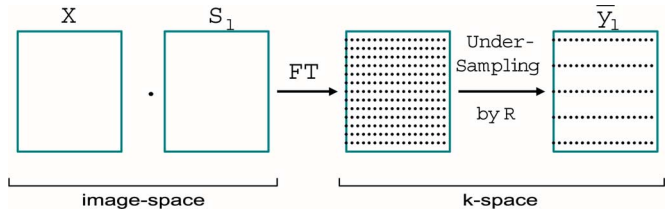

Fig. 1. Schematic of the parallel imaging process for the lth coil. Target image X is multiplied voxel-by-voxel with coil sensitivity S_l. After FT the data is in k-space. Reduced encoding corresponds to undersampling this k- space R times, as shown here for the Cartesian case.

This is discretized via a Cartesian grid on both $\mathbf{r}$ and $\mathbf{k}$. Let vectors $\mathbf{x}$, $\mathbf{s}_l$ and $\mathbf{y}_l$ be the lexicographically stacked versions of the 2-D MR image X, sensitivity responses S_l, and aliased outputs Y_l, respectively, sampled on the regular grid of size $N \times M$. The 2-D-FT now becomes the 2-D DFT and the resampling over $\mathbf{k}$ may be accomplished by using a general downsampling operator in k-space. This process is depicted in Fig. 1.

The input–output relationship of the lth coil in Fig. 1 is succinctly expressed as a matrix product

$$\begin{aligned} c\mathbf{y}_l &= \mathbf{E}_l\mathbf{x} = \mathbf{D}^H_{N\times M} \Downarrow_R \mathbf{D}_{N\times M} S_l \cdot X \\ &= \mathbf{D}^H_{N\times M} \Downarrow_R \mathbf{D}_{N\times M} \mathrm{diag}(\mathbf{s}_l)\mathbf{x}. \end{aligned} \tag{5}$$

The k-space downsampling operator $\Downarrow_R$ resamples k-space according to the specific trajectory used. It is basically an indicator function from $\mathcal{C}^{N\times M}$ to $\mathcal{C}^{N\times M}$, with zeros for every k-space point not sampled by the trajectory. The subscript R denotes the data reduction factor. The operator $\mathbf{D}_{N\times M}$ is 2-D DFT over grid $(N\times M)$. The specific form of $\Downarrow_R$ will depend on the reduction factor and the sampling method used, but it need not be explicitly computed. Note that for non-Cartesian trajectories the gridding step must always be understood to be implicit in the downsampling operator. For instance, if we denote by $\mathbf{G}$ the gridding operator corresponding to a Kaiser–Bessel kernel, then the modified downsampling operator will be given by $\Downarrow'_R = \Downarrow_R \mathbf{G}$. Henceforth, we shall assume $\Downarrow_R$ incorporates gridding, if any.

B. Matrix Structure Under Cartesian k-Space Sampling

Most MR scans are done on Cartesian grids, considerably simplifying things. The 2-D DFT reduces to two 1-D DFT's acting on rows and columns. The general-purpose sampling operator $\Downarrow_R$ in (5) is now redefined as a *subsampling* operator, equivalent to removing rows of k-space.

Writing $\mathbf{D}_{N\times M} = \mathbf{D}^{\mathrm{row}}_M \mathbf{D}^{\mathrm{col}}_N$ as the explicit row and column 1-D DFT operations, since $\Downarrow_R$ only acts on columns, we have $\Downarrow_R \mathbf{D}_{N\times M} = \mathbf{D}^{\mathrm{row}}_M \Downarrow_R \mathbf{D}^{\mathrm{col}}_N$. The output image is now $(N/R) \times M$, and (5) becomes

$$\mathbf{y}_l = \left(\mathbf{D}^{\mathrm{col}}_{N/R}\right)^H \Downarrow_R \mathbf{D}^{\mathrm{col}}_N \mathrm{diag}(\mathbf{s}_l)\mathbf{x} \tag{6}$$

which can be solved column-wise and degenerates into separate $L \times R$ equations as per Theorem 1.

Theorem 1: Let $\mathbf{y}^{(i)}_l, \mathbf{s}^{(i)}_l, \mathbf{x}^{(i)}$ be the ith column of Y_l, S_l, X, and (6) be denoted by $\mathbf{y}^{(i)}_l = \mathbf{E}_l\mathbf{x}^{(i)}$. Consider a partitioning of these signals into R aliasing components under Cartesian sampling

$$\mathbf{x}^{(i)} = \begin{bmatrix} \mathbf{x}^{(i)}{}_1 \\ \vdots \\ \mathbf{x}^{(i)}{}_R \end{bmatrix}, \quad \mathbf{s}^{(i)}_l = \begin{bmatrix} \mathbf{s}^{(i)}_l{}_1 \\ \vdots \\ \mathbf{s}^{(i)}_l{}_R \end{bmatrix}$$

such that the jth element of the rth block of the ith image column is given by $\mathbf{x}^{(i)}{}_r[j] = \mathbf{x}^{(i)}[(N/R)(r-1)+j]$, for $j = 1, \ldots, N/R$. The corresponding elements of sensitivity vectors for each coil $l \in \{1, \ldots, L\}$ are given by $\mathbf{s}^{(i)}_l{}_r[j] = \mathbf{s}^{(i)}_l[(N/R)(r-1)+j]$. An example of this partitioning is depicted in Fig. 2. Then

1)
$$\mathbf{y}^{(i)}_l = \sum_{r=1}^{R} \mathbf{s}^{(i)}_l{}_r \cdot \mathbf{x}^{(i)}{}_r.$$

2) $\mathbf{E}$ has a diagonal-block structure containing $L\times R$ diagonal blocks

$$\mathbf{E} = \{\mathbf{E}^r_l\} \quad \begin{matrix} r=1\ldots R \\ l=1\ldots L \end{matrix}$$

where each sub-block $\mathbf{E}^r_l$ is diagonal, with $\mathbf{E}^r_l = \mathrm{diag}(\mathbf{s}^{(i)}_{l,r})$.

Proof: Proof in Appendix A.

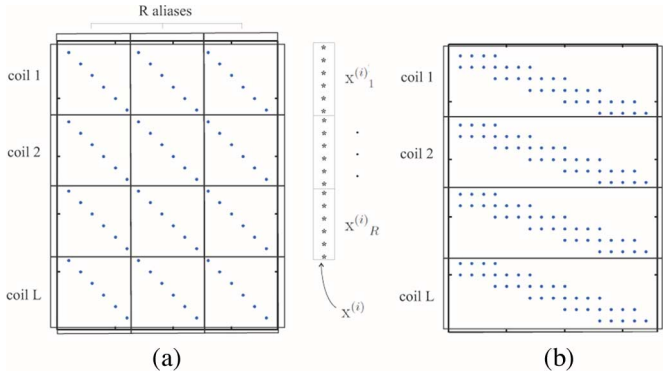

Fig. 2. Structure of system matrix under regular Cartesian sampling. Nonzero elements are indicated with an asterisk. Part (a) shows $\mathbf{E}$ (image space). As a consequence of the partitioning, image column $\mathbf{x}^{(i)}$ separates into R aliasing components. Part (b) shows $\bar{\mathbf{E}}$, the Fourier dual of $\mathbf{E}$.

Theorem 1 is pictorially depicted in Fig. 2. Now each block is diagonal according to Theorem 1, so the interactions are restricted to only R aliasing voxels at a time. Indeed, define $\boldsymbol{\mu} \in \mathcal{C}^L$, $\boldsymbol{\eta} \in \mathcal{C}^R$, $\boldsymbol{\Psi} \in \mathcal{C}^{L\times R}$. Then for each (j,i)th voxel in aliased images $\mathbf{y}_l$, the SENSE problem becomes

$$\begin{aligned} \text{Given} \quad & \boldsymbol{\mu}[l] = \mathbf{y}^{(i)}_l[j], \\ & \boldsymbol{\eta}[r] = \mathbf{x}^{(i)}\,[\mathrm{mod}(j,R)] \\ & \boldsymbol{\Psi}[l,r] = \mathbf{s}^{(i)}_l\,[\mathrm{mod}(j,R)] \\ \text{Solve} \quad & \boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\eta}. \end{aligned}$$

This process, depicted in Fig. 3, is repeated for every set of aliasing voxels, until the entire image is reconstructed. To maintain readability, we henceforth drop column index (i). Symbols

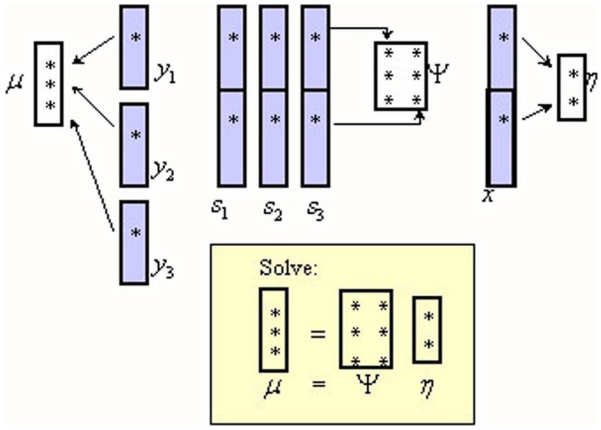

Fig. 3. Converting the full Cartesian problem for each column into a set of independent linear systems of L observations and R unknowns. Vector $\boldsymbol{\mu}$ is $L \times 1$ and $\boldsymbol{\eta}$ is $R \times 1$; matrix $\boldsymbol{\Psi}$ is $L \times R$. This process is repeated for every set of R aliasing voxels, and then for each image column.

$\mathbf{E}$, $\mathbf{x}$, and $\mathbf{y}$, etc., will be used both for arbitrary and Cartesian sampling, their meaning indicated by context.

C. System Matrix Structure Under Arbitrary Sampling

System matrices $\mathbf{E}$, $\bar{\mathbf{E}}$ have important special forms. Individual blocks $\mathbf{E}_1, \ldots, \mathbf{E}_L$ are diagonal, and $\bar{\mathbf{E}}_1, \ldots, \bar{\mathbf{E}}_L$ would have had a Toeplitz structure but for row-decimation due to under-sampling, as shown in Fig. 2(b). (Recall that $\mathbf{T}$ is Toeplitz if $\mathbf{T}[i,j] = \mathbf{t}[j-i]$, for a row vector $\mathbf{t}$.) The Toeplitz-type structure results from convolution operation in k-space. The structure of $\mathbf{E}$ for arbitrary sampling can be determined from (5), but is generally quite complicated and trajectory-dependent, unlike the simple diagonal Cartesian structure.

D. Our Noise Model

For practical implementation, we will use independent Gaussian noise model for both sensitivity and additive noise. Note that our reconstruction is complex, hence we do not have to model Rician noise which is necessary for magnitude data [14]. The lth coil sensitivity and output noise are modeled as

$$S_l^{\text{noisy}} = S_l + N_l^s$$
$$Y_l^{\text{noisy}} = Y_l + N_l$$

where both N_l and N_l^s are independent Gaussian. Let $\mathbf{n}_l$ and $\mathbf{n}_l^s$ be the vectorized representations of N_l and N_l^s, with variance given by $Var(\mathbf{n}_l) = \sigma_n \boldsymbol{\omega}_l$ and $Var(\mathbf{n}_l^s) = \sigma_s \boldsymbol{\lambda}_l$, where we have introduced normalized variance vectors $\boldsymbol{\omega}_l$ and $\boldsymbol{\lambda}_l$. Define for convenience $\boldsymbol{\Omega}_l = \text{diag}(\boldsymbol{\omega}_l)$, $\boldsymbol{\Lambda}_l = \text{diag}(\boldsymbol{\lambda}_l)$. Then the autocorrelation matrices of output and sensitivity noise are given by

$$\mathcal{E}\left(\mathbf{n}_l \mathbf{n}_l^H\right) = \sigma_n^2 \boldsymbol{\Omega}_l^2$$
$$\mathcal{E}\left(\mathbf{n}_l^s \mathbf{n}_l^{sH}\right) = \sigma_s^2 \boldsymbol{\Lambda}_l^2.$$

Clearly, the structure of $\triangle\mathbf{E}$ must mimic that of $\mathbf{E}$ shown in Fig. 2(a)

$$\triangle\mathbf{E} = \{\triangle\mathbf{E}_l^r\} \quad \begin{matrix} r=1\ldots R \\ l=1\ldots L \end{matrix}$$

and the same holds for the k-space versions $\triangle\bar{\mathbf{E}}$ and $\bar{\mathbf{E}}$. Again, each subblock $\triangle\mathbf{E}_l^r$ is diagonal, with entries given by the sensitivity map noise terms N_l^s. Similarly, in k-space the error matrix $\triangle\bar{\mathbf{E}}$ mimics the structure of $\bar{\mathbf{E}}$, as shown in Fig. 2(b).

The assumption of Gaussian noise in spatial sensitivity measurement is quite natural. A popular way to obtain sensitivity maps is through an initial scan with a uniform phantom. In this case, the effects of measurement noise clearly carry over into sensitivity maps. The effect of this noise can be exacerbated by further processing, which might introduce its own set of registration and smoothing errors. Another method is to divide the coil outputs by a body coil output [6]. This causes sensitivity errors in regions of low signal and where the body coil data itself is noisy. The sum-of-squares technique involves using densely sampled central k-space to obtain a relative sensitivity map. Both the latter methods involve voxel-wise division, which can be reasonably considered to yield nonidentically distributed but still fairly independent noise. Whenever two separate scans are used for sensitivity and data, certain other small errors such as misregistration due to motion can creep in the sensitivity map estimation.[4] All these effects add up, making the independent Gaussian assumption a reasonable one. We will demonstrate that in the absence of a detailed and exhaustive error model this model suffices.

Our noise model allows for nonidentically distributed noise. Noise correlation across coils can be accommodated by premultiplying $\mathbf{E}$ with a "whitening" matrix to remove all voxel-wise correlations among coils. Prewhitening for more complicated correlations will generally destroy diagonalization, just as it would in conventional SENSE, leading to greater computational burden. However, there is no additional burden in the non-Cartesian case since diagonalization is not available anyway.

III. ML-SENSE Algorithm

We will derive a general sensitivity-error-tolerant reconstruction which maximizes the likelihood function $\ell(\mathbf{x})$ under arbitrary sampling and general Gaussian noise. Subsequently, we obtain a specific efficient algorithm called ML-SENSE, under the independent noise model of Section II-D. Under Cartesian sampling, this involves minimizing a quasi-quadratic objective function through an efficient nonlinear least squares algorithm.

A. Deriving the Likelihood Function $\ell(\mathbf{x})$

The likelihood $\ell(\mathbf{x})$ given the observed data $\mathbf{y}$ is defined as $Pr(\mathbf{y}|\mathbf{x})$. Let the total noise be $\mathbf{g}(\mathbf{x}) = \mathbf{y} - \mathbf{E}\mathbf{x}$. Under the

[4]This is why we explicitly allow for the noise variance of coil output and coil sensitivity to be different.

Gaussian assumption, this is jointly Gaussian with zero mean. As a result we have

$$\ell(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\left\{(\mathbf{y}-\mathbf{E}\mathbf{x})^H \mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1}(\mathbf{y}(\mathbf{E}\mathbf{x})\right\}\right) \quad (7)$$

where $\mathbf{R}_{\mathbf{g}|\mathbf{x}} = \mathcal{E}((\mathbf{g}(\mathbf{x}))(\mathbf{g}(\mathbf{x}))^H)$ is the covariance matrix of the conditional noise $\mathbf{g}(\mathbf{x})|\mathbf{x}$.

The maximum likelihood estimate, which we will denote $\hat{\mathbf{x}}$, minimizes $-\log \ell(\mathbf{x})$, and is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathbf{y}-\mathbf{E}\mathbf{x})^H \mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1}(\mathbf{y}-\mathbf{E}\mathbf{x}). \quad (8)$$

Under our noise model, $\mathbf{R}_{\mathbf{g}|\mathbf{x}} = \mathcal{E}(\mathbf{n}\mathbf{n}^H + (\triangle\mathbf{E}\mathbf{x})(\triangle\mathbf{E}\mathbf{x})^H)$, and $\mathcal{E}(\mathbf{n}\mathbf{n}^H) = \sigma_n^2\mathbf{\Omega}^2$. We have omitted the $\log(\det(\mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1}))$ term for tractability, since the $\log(\cdot)$ increases slowly compared to the other terms and is safely neglected. For example, a study of Toeplitz systems in image restoration [26] exhibited little improvement after the log term was included, at substantial computational cost. Similar behavior was observed during our experimentation. Consequently, we drop this term henceforth.

The data-dependent covariance $\mathcal{E}((\triangle\mathbf{E}\mathbf{x})(\triangle\mathbf{E}\mathbf{x})^H)$ is an $L \times L$ block matrix $[(\triangle\mathbf{E}_l\mathbf{x})(\triangle\mathbf{E}_{l'}\mathbf{x})^H]_{l,l'\in\{1,\ldots,L\}}$ with the (l,l')th block given by

$$\begin{aligned}(\triangle\mathbf{E}_l\mathbf{x})(\triangle\mathbf{E}_{l'}\mathbf{x})^H &= \mathbf{D}_{N/R\times M}^H \Downarrow_R \mathbf{D}_{N\times M} \\ &\cdot\mathrm{diag}(\mathbf{x})\mathcal{E}\left(\triangle\mathbf{s}_l\triangle\mathbf{s}_{l'}^H\right)\mathrm{diag}(\mathbf{x})\mathbf{D}_{N\times M}^H \Downarrow_R^H \mathbf{D}_{N/R\times M} \quad (9)\end{aligned}$$

which follows from:

$$\begin{aligned}\triangle\mathbf{E}_l\mathbf{x} &= \mathbf{D}_{N/R\times M}^H \Downarrow_R \mathbf{D}_{N\times M}\mathrm{diag}(\triangle\mathbf{s}_l)\mathbf{x} \\ &= \mathbf{D}_{N/R\times M}^H \Downarrow_R \mathbf{D}_{N\times M}\mathrm{diag}(\mathbf{x})\triangle\mathbf{s}_l.\end{aligned}$$

Now $\mathcal{E}(\triangle\mathbf{s}_l\triangle\mathbf{s}_{l'}^H) = \sigma_s^2\delta_{l,l'}\mathbf{\Lambda}_l^2$ since we assume coils are decoupled, therefore

$$\mathbf{R}_{\mathbf{g}|\mathbf{x}} = \sigma_n^2\left(\mathbf{\Omega}^2 + \beta^2\begin{bmatrix}\mathbf{A}_1(\mathbf{x}) & & \\ & \ddots & \\ & & \mathbf{A}_L(\mathbf{x})\end{bmatrix}\right) \quad (10)$$

where $\beta = \sigma_s/\sigma_n$, and $\mathbf{A}_l(\mathbf{x}) = \mathbf{D}_{N/R\times M}^H \Downarrow_R \mathbf{D}_{N\times M}\mathrm{diag}(|\mathbf{\Lambda}_l\mathbf{x}|^2)\mathbf{D}_{N\times M}^H \Downarrow_R^H \mathbf{D}_{N/R\times M}$

Finally, we have

$$\mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1} = \frac{1}{\sigma_n^2}\begin{bmatrix}\mathbf{B}_1(\mathbf{x})^{-1} & & \\ & \ddots & \\ & & \mathbf{B}_L(\mathbf{x})^{-1}\end{bmatrix} \quad (11)$$

$$\begin{aligned}\mathbf{B}_l(\mathbf{x}) &= \mathbf{\Omega}_l^2 + \beta^2\mathbf{D}_{N/R\times M}^H \Downarrow_R \cdot\mathbf{D}_{N\times M}\mathrm{diag}\left(|\mathbf{\Lambda}_l\mathbf{x}|^2\right) \\ &\quad\times \mathbf{D}_{N\times M}^H \Downarrow_R^H \mathbf{D}_{N/R\times M}. \quad (12)\end{aligned}$$

Due to the block-diagonality, we can write the maximum likelihood estimate as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_l (\mathbf{y}_l - \mathbf{E}_l\mathbf{x})^H \mathbf{B}_l(\mathbf{x})^{-1}(\mathbf{y}_l - \mathbf{E}_l\mathbf{x}). \quad (13)$$

Let us summarize the significance of (13): it provides a general recipe for performing ML reconstruction of parallel data under the realistic assumption that noise is present in both coil outputs as well as sensitivity maps. So far we have not specified any particular model for these noise processes, other than to assume that it is Gaussian and there is no cross-coil interference. In theory, (13) can accommodate any noise model which has an adequate stochastic interpretation in terms of second order statistics captured by $\mathbf{\Omega}$ and $\mathbf{\Lambda}$. However, (13) is a nonquadratic minimization problem requiring a large number of cost function evaluations over a solution space of extremely large dimensionality. We wish to obtain practical implementations under a more specific and realistic noise model discussed in Section II-D.

B. Minimization Strategies for General Case

Introducing the independent noise model of Section II-D, $\mathbf{\Omega}_l$ and $\mathbf{\Lambda}_l$ become diagonal, and large simplifications result. Each function evaluation of (13) for general, arbitrarily sampled data involves the inversion of $NM \times NM$ generally nonsparse matrices, making direct inversion prohibitive. However, inversion may be efficiently performed iteratively since the products $\mathbf{B}_l(\mathbf{x})\mathbf{p}$ and $\mathbf{B}_l(\mathbf{x})^H\mathbf{p}$ for an arbitrary vector $\mathbf{p}$ can be computed at $\mathcal{O}(NM\log(N))$ cost due to the presence of the Fourier operator. Furthermore, $\mathbf{B}_l(\mathbf{x})$ are obviously well-conditioned due to diagonal $\mathbf{\Omega}_l$ and $\mathbf{\Lambda}_l$, which means that a fast iterative algorithm like preconditioned CG [21] can perform this inversion in relatively few steps. Since the cost function may be expressed as a data-dependent weighted least squares problem, powerful nonlinear least squares algorithms can be used to solve the problem efficiently (see [13, Ch. 10]). The ML estimate (13) will then reduce drastically in complexity. An efficient general iterative algorithm for this purpose has been developed and work is continuing to evaluate and report it. We do not further specify the general implementation in this paper, focusing instead on the special but important case of Cartesian sampling to obtain an efficient algorithm.

C. Efficient Algorithm for Cartesian Sampling

Recall that for Cartesian sampling the ML problem can be independently solved for each column (i). Further, we prove in Theorem 2 that both i.i.d. and non-i.i.d. Gaussian cases give diagonal $\mathbf{B}_l(\mathbf{x}^{(i)})$. Hence, the ML problem reduces like SENSE (Fig. 3) to NM/R subproblems, each with R variables.

Theorem 2:

For i.i.d. noise: define vectors $\mathbf{b}_l(\mathbf{x}^{(i)}) \triangleq \mathbf{1} + \beta^2\sum_{r=1}^R |\mathbf{x}_r^{(i)}|^2$, $l \in \{1,\ldots,L\}$. Then the ML estimate (13) of column (i) under Cartesian sampling is given by

$$\hat{\mathbf{x}^{(i)}} = \arg\min_{\mathbf{x}} \sum_l \left\| \left(\mathbf{y}_l^{(i)} - \sum_{r=1}^R \mathbf{s}_l^{(i)}{}_r \cdot \mathbf{x}_r\right)/\mathbf{b}_l(\mathbf{x}) \right\|^2.$$

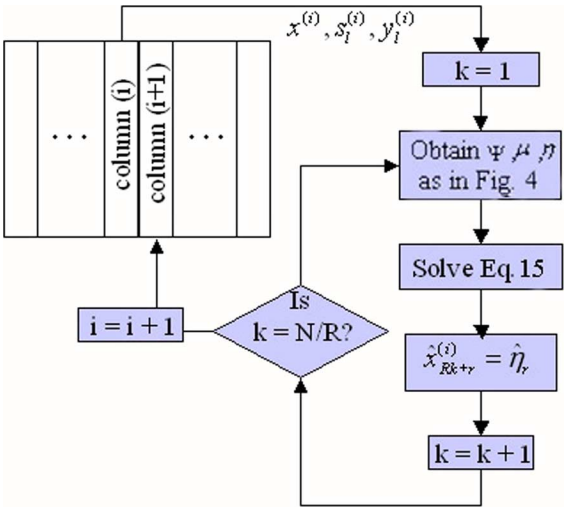

Fig. 4. ML-SENSE algorithm for Cartesian sampling. The algorithm proceeds column by column. For each aliased voxel in each column, all interacting terms are collected in $\boldsymbol{\mu}$, $\boldsymbol{\eta}$ and $\boldsymbol{\Psi}$. The resulting $L \times R$ system is solved, and the elements of x are progressively filled in.

For non-i.i.d. noise with $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$: The Cartesian ML estimate is given by

$$\mathbf{x'}_l^{(i)} \triangleq \boldsymbol{\lambda}_l^{(i)} \cdot \mathbf{x}^{(i)}$$

$$\mathbf{b}'_l\left(\mathbf{x}^{(i)}\right) \triangleq {\boldsymbol{\omega}_l}^{(i)^2} + \beta^2 \sum_{r=1}^{R} \left|\mathbf{x'}_{l_r}^{(i)}\right|^2$$

$$\hat{\mathbf{x}^{(i)}} = \arg\min_{\mathbf{x}} \sum_l \left\| \left(\mathbf{y}_l^{(i)} - \sum_{r=1}^{R} {\mathbf{s}_l^{(i)}}_r \cdot \mathbf{x}'_{lr} \right) / \mathbf{b}'_l(\mathbf{x}) \right\|^2 .$$

Proof: Note that the division '/' is element-by-element. Proof in Appendix B.

Now for each aliasing voxel (j, i), define $\boldsymbol{\eta}$, $\boldsymbol{\mu}$, $\boldsymbol{\Psi}$, as before. Then the ML problem reduces to solving

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \|\mathbf{F}(\boldsymbol{\eta})\|^2, \quad \mathbf{F}(\boldsymbol{\eta}) = q(\boldsymbol{\eta})(\boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta}) \quad (14)$$

$$q(\boldsymbol{\eta}) = 1/\sqrt{1 + \beta^2 \|\boldsymbol{\eta}\|^2}. \quad (15)$$

Fig. 4 shows the algorithm to implement the minimization of (13) under the i.i.d. assumption, one column at a time. Called Algorithm I, it is further specified in Appendix C. The minimization is challenging due to the presence of the nonquadratic term. But the nonquadraticity enters the equation only via a well-behaved, smooth, slowly-varying function of the norm $\|\boldsymbol{\eta}\|^2$. Consequently, minimization can be achieved using a nonlinear least-squares method with Newton iterations [13]. This is similar to the standard least squares method for solving the pseudoinverse $\boldsymbol{\Psi}^{\dagger}$. The only difference is that the Jacobian of $\mathbf{F}$ is not a constant matrix any more. In Appendix F, we briefly describe the standard Newton algorithm for minimizing (14), which uses the Jacobian to compute line search directions iteratively. Fortunately the Jacobian, whose knowledge speeds up Newton iterations considerably, in this case is readily available and easily computed (see Appendix G)

$$\mathbf{J}(\boldsymbol{\eta}) = \frac{\partial \mathbf{F}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = -q(\boldsymbol{\eta}) \left\{ \boldsymbol{\Psi} + \beta^2 q^2(\boldsymbol{\eta})(\boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta})\boldsymbol{\eta}^T \right\}. \quad (16)$$

D. Extension to Nonuniform Noise

The solution of extended noise model with non-i.i.d. noise is also given by Theorem 2, as a series of independent $L \times R$ subproblems. Let us capture the corresponding entries of $\boldsymbol{\omega}_l$ in the L-vector $\mathbf{u}$, and $\boldsymbol{\lambda}_l$ in R-vectors $\mathbf{r}_l$ similar to $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ before. Consequently, the minimization proceeds exactly as before, with $q(\boldsymbol{\eta})$ now replaced by the $L \times L$ diagonal matrix $\mathbf{Q}(\boldsymbol{\eta})$

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \|\mathbf{F}(\boldsymbol{\eta})\|^2, \quad \mathbf{F}(\boldsymbol{\eta}) = \mathbf{Q}(\boldsymbol{\eta})(\boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta})$$

$$\mathbf{Q}(\boldsymbol{\eta})[l,l] = 1/\sqrt{\mathbf{u}^2[l] + \beta^2 \|\mathbf{r}_l \cdot \boldsymbol{\eta}\|^2}. \quad (17)$$

The new algorithm, Algorithm II, is further specified in Appendix D, and follows the same structure, as shown in Fig. 4 above. ML-SENSE II is slightly more challenging numerically compared to ML-SENSE I. But again, nonquadraticity is only on account of a slowly-varying functional of norm $\|\mathbf{r}_l \cdot \boldsymbol{\eta}\|^2$. The Jacobian is slightly different, but still efficiently computable. Let $\boldsymbol{\Gamma} = [\mathbf{r}_1^2, \ldots, \mathbf{r}_L^2]$. Then the new Jacobian is (see Appendix G)

$$\begin{aligned} \mathbf{J}(\boldsymbol{\eta}) &= \frac{\partial \mathbf{F}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \\ &= -\mathbf{Q}(\boldsymbol{\eta}) \big\{ \boldsymbol{\Psi} + \beta^2 \mathbf{Q}^2(\boldsymbol{\eta}) \\ &\qquad \times \operatorname{diag}(\boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta}) \boldsymbol{\Gamma}^T \operatorname{diag}(\boldsymbol{\eta}) \big\}. \end{aligned} \quad (18)$$

Clearly ML-SENSE II generalizes ML-SENSE I, being especially useful under systematic rather than random sensitivity errors, for instance those caused by the division method where sensitivity information in low signal regions may be unreliable. Further, sensitivity noise propagation due to the division step will be modulated by the overall sum-of-squares, leading to nonuniform noise.

We propose here one method of obtaining $\boldsymbol{\omega}_l$ and $\boldsymbol{\lambda}_l$, used in our experiments with ML-SENSE II. Noise variance was computed for each voxel of each coil by taking the variance of a local window of size $N/20 \times N/20$ centered at the voxel. The variance calculation was weighted by a monotonically decreasing function, in our work a quadratic, centered at the voxel. We chose window size of $N/20$ after a brief trial with other window sizes. Larger windows do not produce a sufficiently local estimate while smaller windows do not have enough data points for good variance estimate. For the purpose of validating proposed method, we used unaccelerated data to obtain these estimates. This is clearly impractical for cases where representative unaccelerated scans are not available. More sophisticated methods are currently being investigated; however, we note that in many cases accurate estimates of $\boldsymbol{\omega}_l$ and $\boldsymbol{\lambda}_l$ may not ultimately be available. Therefore, we describe Algorithms I and II separately—in absence of full noise statistics, ML-SENSE I is suboptimal but preferable.

E. Computational Burden

The additional cost of nonquadratic minimization is not significantly higher than standard pseudoinverse computed through conjugate gradients, due to the easy availability of the Jacobian and its cheap evaluation from (16) and (18). The algorithms were implemented in MATLAB ver. R13. Typical execution times for reconstructions of size 256×256 were between three

TABLE I
SUMMARY OF COMPUTATIONAL BURDEN. ORDER OF FLOPS FORMULAS ARE THEORETICAL NUMBER OF MULTIPLICATION OPERATIONS PER ITERATION OF THE PCG LOOP. QUOTED AVERAGE NUMBER OF ITERATIONS ARE ROUGH ESTIMATES OBTAINED EMPIRICALLY FROM A SMALL NUMBER OF TRIALS

Algo	SENSE	ML-SENSE-I	ML-SENSE-II
flops/itn	$\mathcal{O}(2MNL)$	$\mathcal{O}(3MN(L+1))$	$\mathcal{O}(3MN(L+2))$
avg itn	10	30	30

to four times the execution time in Matlab of standard SENSE. A careful order of flops calculation, contained in Table I, indicates a roughly 50% increase in computational burden per iteration. However, ML-SENSE takes more iterations to converge than SENSE since the former is nonquadratic.

IV. RESULTS

Algorithms I and II were not considerably different even under non-i.i.d. noise. ML-SENSE II seems to perform slightly better when sensitivity errors are spatially varying AND can be properly determined; however, it is not possible in many cases to measure this variation accurately. Therefore, both ML-SENSE I and II are shown in examples below, wherever possible and appropriate. All results were compared with conventional SENSE, whose implementation details are supplied in Appendix E. The classical TLS algorithm was also evaluated on selected data sets. Our implementation of TLS follows exactly the method proposed in [21]. TLS appears to suffer from instability and noise amplification due to the final step described in [21] where a division takes place. We tried to stabilize the division step by adding a small value to the denominator. In addition, TLS displayed extremely noisy behavior for image columns with very little signal (background columns). We zeroed out these columns in the final reconstruction.

A. Simulation Results

Simulated phased-array data was obtained as follows. Sensitivity of circular coils positioned uniformly around the FOV were computed from the Biot–Savart Law. Coil data were computed by encoding a fully sampled MR image with coil sensitivities, and down-sampling by R in the PE direction. First reconstruction using SENSE, classical TLS and ML-SENSE is performed on data from a 128×128 Shepp–Logan head phantom image where both the coil data and simulated sensitivity maps are noiseless. Shown in Fig. 5, this demonstrates that if there is no sensitivity error, all methods perform perfect reconstruction.

Next, we simulate the effect of large Gaussian noise added to data and sensitivity to simulate random errors. To keep the comparison uncluttered, equal relative noise was introduced in both sensitivity and data. The performance of ML-SENSE I with $R = 4$ and $L = 6$ can be evaluated visually in Fig. 6. Since the added noise is uniform, ML-SENSE II results are the same and are not shown here. Reduced phase encoding was along the vertical direction. The standard SENSE result is almost useless in this case. The encoding matrix is badly conditioned due to large acceleration factor, causing severe noise amplification. TLS displays somewhat improved noise performance. In contrast, our ML-SENSE algorithm is able to salvage more useful data.

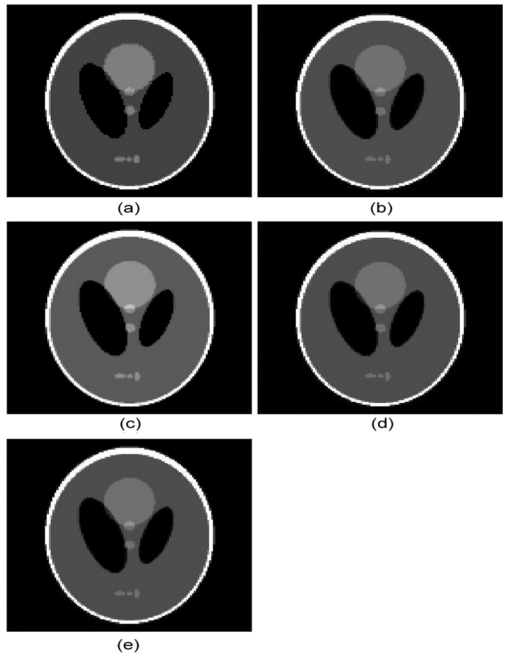

Fig. 5. Reconstruction of noise-free parallel data simulated from the Shepp–Logan phantom, with $R = 4$, $L = 6$. (a) Original phantom image. (b) Standard SENSE reconstruction. (c) TLS reconstruction. (d) ML-SENSE I reconstruction. (e) ML-SENSE II. Each reconstruction is perfect.

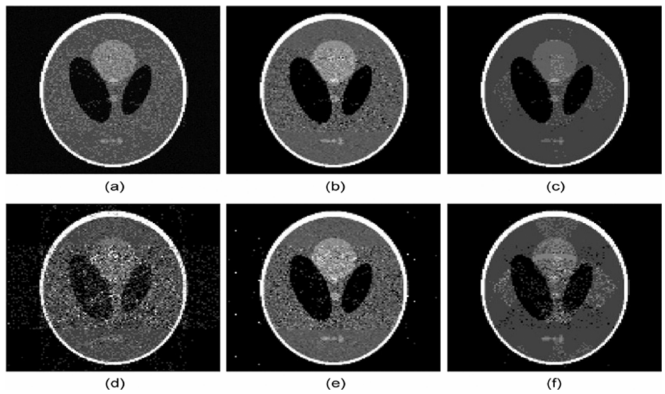

Fig. 6. Reconstruction of noisy simulated Shepp–Logan phantom data, with $R = 4$, $L = 6$. (a) SENSE with midlevel noise, (b) TLS, (c) ML-SENSE I, (d) SENSE with large noise, (e) TLS, and (f) ML-SENSE I. Note the excessive noise amplification due to very high g-factor in the middle of FOV.

data. Notice that amplification of *additive* noise occurs in areas of high g-factor in all reconstructions, but conventional SENSE suffers from additional loss of SNR due to sensitivity noise. A quantitative comparison is now performed for $R = 4$. For a given SNR (labeled "input SNR" in Fig. 7) of sensitivity and data, we determine the SNR of reconstructions using standard and ML-SENSE (labeled "reconstructed SNR"). Reconstructed SNR is available from the difference from the original unaliased image. Two sets of plots are shown for $L = 5$ and $L = 6$, respectively. Noise performance improves by around 20 dB and 14 dB, respectively, in the high noise region. The SNR performance of the algorithms converges at high input SNR, as they should. Reconstructed SNR is always lower than input SNR due to inadequate least squares averaging.

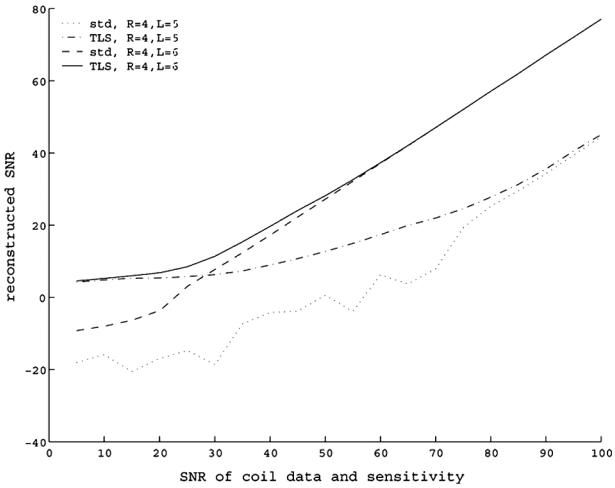

Fig. 7. The SNR performance of standard and ML-SENSE. Both input and output SNR were obtained from a direct comparison with the original, unaccelerated image.

B. Experiments With Sensitivity Mismatch on High-Resolution Phantom Data

Experiments were performed with a high-resolution phantom (HRP) to validate ML-SENSE under systematic errors like sensitivity mismatch and insufficient FOV. HRP was placed within a plastic tube around which torso coil pads were wrapped. The FOV was slightly smaller than required, resulting in controlled aliasing. Sensitivity maps were obtained with a uniform spherical phantom (USP). Measured output and sensitivity of one coil under zero acceleration are shown in Fig. 8(a). Notice that sensitivity has some contributions from the tube, of unequal intensity to USP, a kind of error normally quite disastrous for conventional SENSE. A small misalignment of nearly 1/10th of FOV was observed between sensitivity and data. While not standard, this experiment mimics the errors expected if sensitivity maps are misaligned, incorrect, or obtained by improper division with a body coil image. We wish to see if this flawed data set can be reconstructed under gross inaccuracies in sensitivity quite unlike the noise model we assumed in this work. SENSE result [Fig. 8(c)] is unsatisfactory due to mismatch and aliasing. We also performed a Tikhonov-regularized version of SENSE [19], [27]. Regularization improved SNR [Fig. 8(d)], but introduced unacceptable aliasing, a well-known regularization effect [17]. In ML-SENSE results, the effect of sensitivity mismatch and aliasing are visible but considerably muted. In implementing ML-SENSE-II, we obtained sensitivity noise statistics (vectors $\boldsymbol{\lambda}_l$) by treating all sensitivity signals outside the USP as noise. In practice, of course, it is much harder to determine what constitutes genuine sensitivity and what constitutes noise, but for this experiment we are more concerned with how well Algorithm II will perform if this kind of information were in fact available. As expected, using information about $\boldsymbol{\lambda}_l$, ML-SENSE-II outperforms ML-SENSE-I. However, it is noteworthy that ML-SENSE-I too is quite effective even though sensitivity noise is not i.i.d.

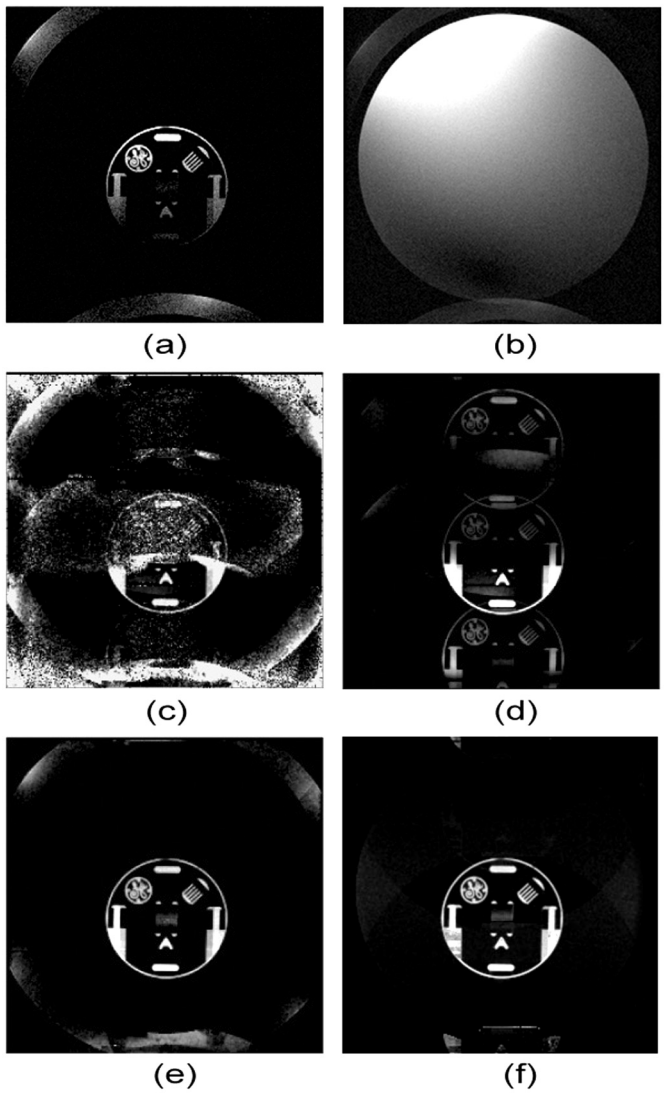

Fig. 8. Reconstruction results of HRP data, with $R = 3$, $L = 4$: (a) image of HRP from a single coil without acceleration, (b) sensitivity map obtained from USP, (c) standard SENSE, (d) regularized SENSE with $\lambda_{\text{reg}} = 0.05$, (e) ML-SENSE-I, and (f) ML-SENSE-II. Both ML-SENSE algorithms exhibit improved performance compared to conventional methods.

C. Parallel Brain Imaging With an Eight-Element Head Coil

Our first *in vivo* results are of brain imaging under low SNR and suboptimal coil configuration (high g-factor). Fully sampled data were obtained and under-sampling was performed offline by manually removing k-space views. Sensitivity was obtained by the division method from the central 30 views which were densely sampled. We tried to mitigate truncation artifacts as much as possible by using a raised-cosine window; however, some residual ringing was observed in sensitivity. These problems can be fixed by using more views for sensitivity estimation; however, doing so will reduce the effective acceleration and negate the purpose of doing parallel imaging. Further degradation of sensitivity resulted from noise amplification during the division step in regions with low signal. This example is a faithful reproduction of typical imaging errors under challenging imaging conditions. In these experiments, we chose not

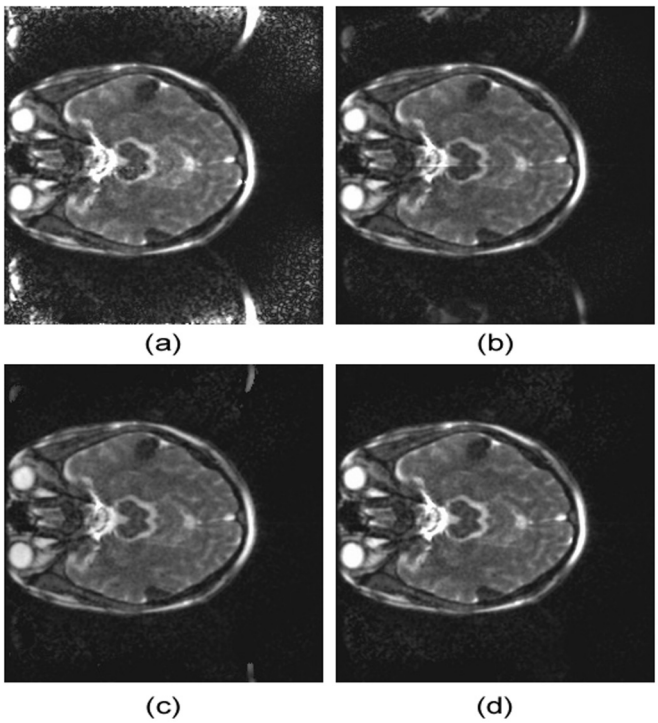

Fig. 9. Brain reconstruction results with $R = 2$, $L = 8$: (a) SENSE, (b) SENSE regularized with best parameter from Fig. 11, (c) ML-SENSE-I, and (d) ML-SENSE-II. All algorithms appear similar.

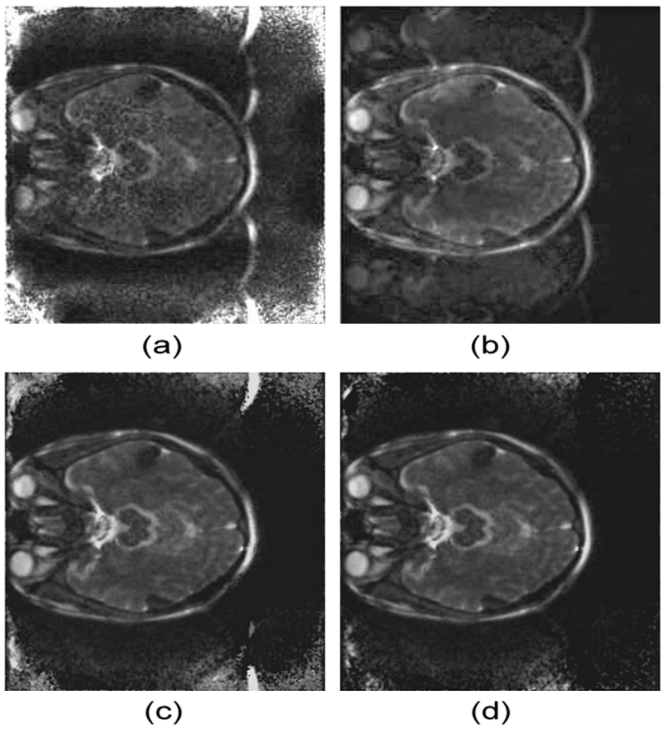

Fig. 10. Brain reconstruction with $R = 4$, $L = 8$: (a) SENSE, (b) regularized SESNE, (c) ML-SENSE-I, and (d) ML-SENSE-II. Higher acceleration has caused serious artifacts in conventional methods. ML-SENSE II seems to better suppress residual aliasing arising from the hyperintense fat signal.

to mask out the background using thresholding since its estimation from available error-prone sensitivity maps was generally challenging. If such a background mask can be reliably obtained for specific imaging situations, the results will correspondingly improve for all methods.

We first show the case of $R = 2$ times acceleration in Fig. 9. Although data and sensitivity maps are error-prone, the results from both SENSE and ML-SENSE look acceptable due to low g-factor, with ML-SENSE Fig. 9 (c)–(d) showing a small improvement over both SENSE Fig. 9(a) and its regularized version Fig. 9(b). Now consider the same data, but with $R = 4$, shown in Fig. 10. SENSE reconstruction in Fig. 10(a) now exhibits excessive noise amplification due to a combination of high R and erroneous sensitivity maps. To show that this problem cannot really be addressed by regularization, Fig. 10(b) shows the output of regularized SENSE with $\lambda_{\text{reg}} = 0.1$, chosen after an exhaustive L-curve analysis [17], as follows. We varied λ_{reg} in increments given by a geometric progression from 0 to 1. At each increment a data and a prior cost were computed

$$E_{\text{data}} = \|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2, \quad E_{\text{prior}} = \|\mathbf{x}\|^2.$$

Fig. 11 plots these values and this is called the L-curve. The optimal value of λ_{reg} is considered to be the one nearest to the "elbow" of the L-curve, since it represents the best compromise between the least squares fit to the observed data (smallest E_{data}) and the best fit to the prior (smallest E_{prior}). Note that our implementation corresponds to a zeroth-order regularization

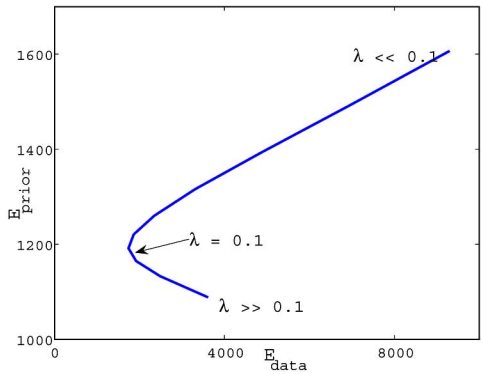

Fig. 11. L-curve for brain data obtained by varying λ_{reg} in increments of a geometric progression from 0 to 1. This is a plot of the prior cost versus data cost. The best value, given by the L-curve elbow, was around $\lambda_{\text{reg}} = 0.1$. This was used to regularize SENSE in Fig. 10.

since it does not use a prior mean image [18]. In most interesting situations, a prior mean image is not available. Low-resolution mean images computed from central k-space data were not found useful in our reconstructions since they produced severe boundary artifacts.

While regularized SENSE, which was previously proposed for parallel imaging by several authors [17] [19] [20] is less noisy, it has in fact failed to resolve the aliasing components properly. A smaller λ_{reg}, say 0.05, would have resolved the ghosting better, but with more noise amplification, as implied by the L-curve. Fig. 10(c) and (d) show the ML-SENSE reconstructions, which seem to suffer neither from excessive noise amplification nor ghosting. ML-SENSE II is slightly better than

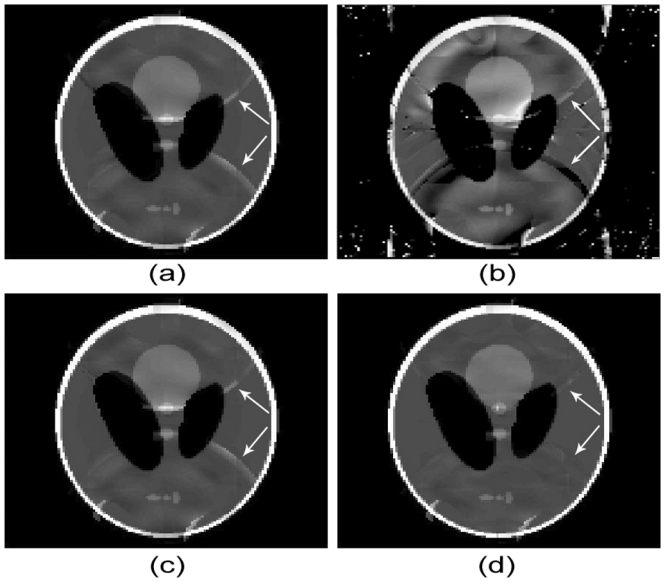

Fig. 12. Reconstruction of Shepp–Logan phantom, with $R = 2$, $L = 4$ using sensitivity maps obtained from the central 15 lines of k-space and division by sum of squares. Sensitivity was perturbed by small shifts in orientation and positioning. (a), (b), (c), (d) are SENSE, TLS, ML-SENSE I, and ML-SENSE II, respectively. Since sensitivity error is systematic and correlated, TLS performs poorly. Note the absence of the aliasing ring in (d), pointed out by arrows.

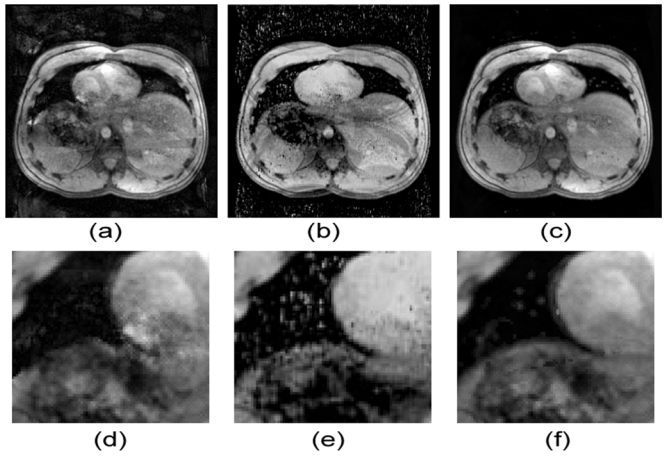

Fig. 13. Reconstruction results of torso scan, $R = 3$, $L = 8$: (a) Standard SENSE with regularization, (b) TLS, (c) ML-SENSE II, (d) Zooming into (a), (e) Zooming into (b), and (f) zooming into (c). Note the distortions at the heart–stomach boundary and strip artifact across the liver in SENSE output.

ML-SENSE I as expected. Comparing Figs. 9 and 10, we conclude that while under benign imaging conditions (low R, small g-factors) sensitivity errors may not impact quality, at higher accelerations and poorer matrix conditioning properties they can seriously degrade conventional performance. In these situations, the ML-SENSE approach appears to perform better.

D. Experiments Under Nonuniform Sensitivity Noise

Now we investigate the comparative performance of Algorithms I and II for the Shepp–Logan data set. Sensitivity was obtained from central k-space, and additional perturbations were introduced via shifts in orientation and positioning between the encoding (i.e., actual) sensitivity and decoding (estimated) sensitivity. The resulting sensitivity noise is mostly negligible except within the bright outer ring of the phantom. ML-SENSE I is not likely to produce significant improvement due to the nature of sensitivity errors in this case. However, modeling this as spatially varying noise, reliable estimates of $\mathbf{\Omega}$ and $\mathbf{\Lambda}$ are available from the simulation, and we expect ML-SENSE II to provide a better reconstruction. This was found to be the case, as shown in Fig. 12. TLS result is also shown for comparison—it is clearly unsatisfactory in this situation.

Next, we investigate the nonuniform sensitivity noise resulting from misregistration due to breathing, a constant problem in torso scanning. Motion during the acquisition of k-space views causes small amounts of mismatch between calibration lines and other lines in the data. An axial torso slice was scanned using an eight-channel torso coil (four in the anterior coil pad, four in the posterior). As before, undersampling was done manually after unaccelerated data were acquired. The central 30 k-space views were used to obtain unaliased low-frequency sensitivity maps by the division method. There were three sources of spatially variable sensitivity error in this experiment: misregistration due to motion, ringing due to truncation, and the division step in areas of weak signal. ML-SENSE-II is the appropriate method in this case. We obtained $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ by the local window variance estimation described earlier. Results are shown in Fig. 13(a) and (b), along with zoomed in region containing the stomach–heart interface in Fig. 13(c) and (d). Several artifacts contaminate the SENSE output, including loss of heart-stomach boundary definition, hyperintensity in heart region, and stripe artifact across the liver. TLS result is also unsatisfactory due to unresolved ghosting and excessive noise amplification. These artifacts are largely absent in the reconstruction using ML-SENSE II.

V. Conclusion

We addressed the problem of obtaining an optimal solution to the parallel imaging reconstruction problem in the presence of both measurement and sensitivity noise. We showed that for independent Gaussian noise the optimal solution is the minimizer of a weakly nonquadratic objective function which may be solved efficiently via a nonlinear least squares iterative technique with modest additional complexity compared to standard SENSE algorithms. We have also derived simplified expressions for the cost function as well as the Jacobian of the associated least squares problem in the case of Cartesian k-space sampling. A fast Newton algorithm with explicit Jacobian information was developed to solve the problem. Results for Cartesian k-space sampling indicate impressive improvement in performance compared to standard SENSE, amounting to almost 20 dB SNR gain in several high-noise cases. Our algorithms yields substantial improvement even in cases where the sensitivity noise is not independent. These preliminary results are promising, especially for abdominal imaging where large motion-induced sensitivity artifacts are present. But further evaluation under various clinical settings is needed to assess their true clinical significance. The issue of accurately measuring spatially varying noise directly from accelerated data must be further investigated before the utility of Algorithm II can be ascertained.

A natural extension to our work would be to handle non-Cartesian sampling schemes. The basic solution for the ML estimate remains the same, but the non-Cartesian problem can not benefit from diagonalization. Efficient implementations for arbitrary sampling as well as for more general noise models was briefly described in this paper, but detailed implementation is currently being investigated.

APPENDIX

A. Proof of Theorem 1

Consider the 1-D DFT matrix $\mathbf{D}_N^{\text{col}} = \{e^{(i2\pi/N)(kn)}\}$, $k = 0,\ldots,N-1, n = 0,\ldots,N-1$. Now the row-decimated matrix obtained by retaining every Rth row in $\mathbf{D}_N^{\text{col}}$ can be written as $\{e^{(i2\pi/N)(Rk'n)}\}$, $k' = 0,\ldots,N/R-1, n = 0,\ldots,N-1$. Expanding this in terms of $(N/R \times N/R)$-blocks, we get, for $k' = 0,\ldots,N/R-1, n' = 0,\ldots,N/R-1$

$$\Downarrow_R \mathbf{D}_N^{\text{col}} = \left\{e^{\frac{i2\pi}{N}(Rk'n')}, e^{\frac{i2\pi}{N}\left(Rk'\left(n'+\frac{N}{R}\right)\right)}, \ldots, e^{\frac{i2\pi}{N}\left(Rk'\left(n'+(R-1)\frac{N}{R}\right)\right)}\right\}.$$

Each of these terms evaluates to $exp(i(2\pi/(N/R))k'n')$, giving us

$$\Downarrow_R \mathbf{D}_N^{\text{col}} = \mathbf{D}_{N/R}^{\text{col}}[\mathbf{I}_{N/R}\ldots\mathbf{I}_{N/R}] \tag{19}$$

Therefore, $\mathbf{y}_l^{(i)} = \mathbf{E}_l\mathbf{x}^{(i)} = [\mathbf{I}_{N/R}\ldots\mathbf{I}_{N/R}]\,\text{diag}(\mathbf{s}_l^{(i)})\mathbf{x}^{(i)} = [\text{diag}(\mathbf{s}_1^{(i)}),\ldots,\text{diag}(\mathbf{s}_R^{(i)})]\mathbf{x}^{(i)}$.

This proves part (1) and leads immediately to the partitioning $\mathbf{E}_l = [\mathbf{E}_l^1,\ldots,\mathbf{E}_l^R]$, with $\mathbf{E}_l^r = \text{diag}(\mathbf{s}_{l,r}^{(i)})$. Assembling the full matrix $\mathbf{E}$ for all coils we get the result in part (2).

B. Proof of Theorem 2

Proof: For i.i.d. case, $\mathbf{\Lambda} = \mathbf{I}$, $\mathbf{\Omega} = \mathbf{I}$. Recall that $\Downarrow_R \mathbf{D}_{N\times M} = \mathbf{D}_M^{\text{row}} \Downarrow_R \mathbf{D}_N^{\text{col}}$ and ${\mathbf{D}_{N/R}^{\text{col}}}^H \Downarrow_R \mathbf{D}_N^{\text{col}} = [\mathbf{I}_{N/R}\ldots\mathbf{I}_{N/R}]$ from Theorem 1. Then we have for the column-wise matrix $\mathbf{B}_l(\mathbf{x}^{(i)})$ [see (12)]

$$\begin{aligned}\mathbf{B}_l\left(\mathbf{x}^{(i)}\right) &= \mathbf{I} + \beta^2 {\mathbf{D}_{N/R}^{\text{col}}}^H \Downarrow_R \mathbf{D}_{N/R}^{\text{col}} \text{diag} \\ &\quad \times \left(\left|\mathbf{x}^{(i)}\right|^2\right) {\mathbf{D}_{N/R}^{\text{col}}}^H \Downarrow_R^H \mathbf{D}_{N/R}^{\text{col}} \\ &= \mathbf{I} + \beta^2[\mathbf{I}_{N/R}\ldots\mathbf{I}_{N/R}]\text{diag} \\ &\quad \times \left(\left|\mathbf{x}^{(i)}\right|^2\right)[\mathbf{I}_{N/R}\ldots\mathbf{I}_{N/R}]^T \\ &= \text{diag}\left(\mathbf{1} + \beta^2\sum_{r=1}^{R}\left|\mathbf{x}_r^{(i)}\right|^2\right) \\ &\triangleq \left[\text{diag}\left(\mathbf{b}_l\left(\mathbf{x}^{(i)}\right)\right)\right]^{-1}\end{aligned}$$

Then the ML problem (13) for a single column becomes

$$\hat{\mathbf{x}^{(i)}} = \arg\min_{\mathbf{x}} \sum_l \left(\mathbf{y}_l^{(i)} - \mathbf{E}_l\mathbf{x}\right)^H \times [\text{diag}\,(\mathbf{b}_l(\mathbf{x}))]^{-1}\left(\mathbf{y}_l^{(i)} - \mathbf{E}_l\mathbf{x}\right)$$

which immediately proves part (1) of the theorem. Part (2) for non-i.i.d. case follows analogously, this time accounting for the diagonal matrices $\mathbf{\Lambda}$ and $\mathbf{\Omega}$.

C. ML-SENSE Algorithm for Cartesian Sampling: ML-SENSE I

- Y_l = coil output of lth coil, in spatial domain.
- S_l = sensitivity map of lth coil.
- X = desired MR image of size $(N \times M)$.
- L = number of coils.
- R = downsampling factor.
- *for* $i = 1,\ldots,M$:
  - define $\mathbf{x}$, $\mathbf{y}_l$, $\mathbf{s}_l$ as the ith column of X, Y_l, S_l, respectively.
  - *for* $k = 1\ldots N/R$
    Define a $L \times R$ matrix $\mathbf{\Psi}$, with $\mathbf{\Psi}[l,r] = \mathbf{s}_l[(r-1)(N/R)+k]$. Let $\boldsymbol{\mu} = [\mathbf{y}_1[k],\ldots,\mathbf{y}_L[k]]^T$.
    Solve $\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}}(1/(1+\beta^2\|\boldsymbol{\eta}\|^2))\|\boldsymbol{\mu} - \mathbf{\Psi}\boldsymbol{\eta}\|^2$
    for $r = 1\ldots R$, $\hat{\mathbf{x}}[(r-1)(N/R)+k] = \hat{\boldsymbol{\eta}}[r]$
  - ith column of $X = \hat{\mathbf{x}}$.

D. ML-SENSE Under Nonuniform Sensitivity and Output Noise: ML-SENSE II

- *for* $i = 1,\ldots,M$
  - Define $\mathbf{x}$, $\mathbf{y}_l$, $\mathbf{s}_l$ as the ith column of X, Y_l, S_l, respectively.
  - *for* $k = 1,\ldots,N/R$
    - Define $L \times R$ matrix $\mathbf{\Psi}$, with $\mathbf{\Psi}[l,r] = \mathbf{s}_l[(r-1)(N/R)+k]$. Let $\boldsymbol{\mu} = [\mathbf{y}_1[k],\ldots,\mathbf{y}_L[k]]^T$
    - For each $l = 1,\ldots,L$, define R-vectors $\mathbf{r}_l = \boldsymbol{\lambda}_l^{(i)}[(r-1)N/R+k]$ and $\mathbf{u}[l] = \boldsymbol{\omega}_l^{(i)}[(r-1)N/R+k]$
    - Define $L \times R$ matrix $\mathbf{\Gamma}$, with $\mathbf{\Gamma} = [\mathbf{r}_1^2,\cdots,\mathbf{r}_L^2]$.
    - Solve $\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}}\|\mathbf{Q}(\boldsymbol{\eta})(\boldsymbol{\mu} - \mathbf{\Psi}\boldsymbol{\eta})\|^2$, with $\mathbf{Q}(\boldsymbol{\eta})[l,l] = 1/\sqrt{u_l^2 + \beta^2\|\mathbf{r}_l\cdot\boldsymbol{\eta}\|^2}$.
    - for $r = 1,\ldots,R$, $\hat{\mathbf{x}}[(r-1)(N/R)+k] = \hat{\boldsymbol{\eta}}[r]$.
  - ith column of $X = \hat{\mathbf{x}}$.

E. SENSE and Regularized SENSE Implementation

Each $L \times R$ subsystem $\boldsymbol{\mu} = \mathbf{\Psi}\boldsymbol{\eta}$ is solved separately in SENSE, then the elements of the full image X are filled in from the estimates of $\boldsymbol{\eta}$. Matrix $\mathbf{\Psi}$ is inverted through the pseudo-inverse via the popular conjugate gradients algorithm described previously by many authors, e.g., [5] and [6]. Thus,

$$\hat{\boldsymbol{\eta}} = \mathbf{\Psi}^\dagger\boldsymbol{\mu}.$$

Regularization: Tikhonov-regularized SENSE was implemented by solving the augmented system

$$\begin{aligned}\hat{\boldsymbol{\eta}} &= {\mathbf{\Psi}'}^\dagger\boldsymbol{\mu}', \text{ where} \\ \boldsymbol{\mu}' &= [\boldsymbol{\mu}^T, \mathbf{0}^T]^T \\ \mathbf{\Psi}' &= [\mathbf{\Psi}^T, \lambda_{\text{reg}}\mathbf{I}^T]^T.\end{aligned}$$

F. ML-SENSE Cost Minimization Routine

The minimization of (14) and (17) proceeds via the well-established Gauss–Newton method [28]. In the approximate vicinity of the true solution, the Hessian is given by

$$\mathbf{H}(\boldsymbol{\eta}) \approx \mathbf{J}(\boldsymbol{\eta})^T\mathbf{J}(\boldsymbol{\eta}).$$

The Gauss–Newton method computes, at each iteration k, a line search direction $\mathbf{d}_k$ starting from the current solution $\boldsymbol{\eta}_k$ which is the minimizer of the following least squares problem:

$$\min_{\mathbf{d}_k} ||\mathbf{J}(\boldsymbol{\eta}_k)\mathbf{d}_k - \mathbf{F}(\boldsymbol{\eta}_k)||^2 .$$

Since the Jacobian and function evaluations are explicitly available and cheaply computable via (14)–(18), the above is a simple least squares problem which was solved by conventional CG algorithm. Finally, a 1-D line search is performed for each direction $\mathbf{d}_k$ using the standard method described in Section 2–6 of [28].

G. Jacobian of $\mathbf{F}(\boldsymbol{\eta})$

Let $\mathbf{r}_i, \boldsymbol{\Gamma}$ be as defined in Section III-D, let the ith element of $\mathbf{F}(\boldsymbol{\eta})$ be $F_i(\boldsymbol{\eta})$, and ith row of $\boldsymbol{\Psi}$ be $\boldsymbol{\psi}_i^T$. Then

$$\frac{\partial F_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = -\mathbf{Q}(\boldsymbol{\eta})[i,i]\boldsymbol{\psi}_i + \frac{\partial \mathbf{Q}(\boldsymbol{\eta})[i,i]}{\partial \boldsymbol{\eta}} \left(\boldsymbol{\mu}[i] + \boldsymbol{\psi}_i^T \boldsymbol{\eta}\right)$$

Now $\partial \mathbf{Q}(\boldsymbol{\eta})[i,i]/\partial \boldsymbol{\eta} = -\beta^2 \mathbf{Q}^3(\boldsymbol{\eta})[i,i](\mathbf{r}_i^2 \cdot \boldsymbol{\eta})$.. Then

$$\begin{aligned}\mathbf{J}_i(\boldsymbol{\eta}) &\triangleq \left(\frac{\partial F_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T \\ &= -\mathbf{Q}(\boldsymbol{\eta})[i,i] \cdot \Big\{\boldsymbol{\psi}_i^T + \beta^2 \mathbf{Q}^2(\boldsymbol{\eta})[i,i] \\ &\qquad \times \left(\boldsymbol{\mu}[i] + \boldsymbol{\psi}_i^T \boldsymbol{\eta}\right) \boldsymbol{\eta}^T \operatorname{diag}\left(\mathbf{r}_i^2\right)\Big\}\end{aligned} .$$

For algorithm I, $\mathbf{r}_i^2 = \mathbf{1}$, $\mathbf{Q}(\boldsymbol{\eta})[i,i] = q(\boldsymbol{\eta})$, and we get

$$\mathbf{J}(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{J}_1(\boldsymbol{\eta}) \\ \vdots \\ \mathbf{J}_L(\boldsymbol{\eta}) \end{bmatrix} = -q(\boldsymbol{\eta}) \left\{\boldsymbol{\Psi} + \beta^2 q^2(\boldsymbol{\eta})(\boldsymbol{\mu} + \boldsymbol{\Psi}\boldsymbol{\eta})\boldsymbol{\eta}^T\right\} .$$

For algorithm II, we need to stack rows $J_i(\boldsymbol{\eta})$ more carefully. It is easily verified that

$$\mathbf{J}(\boldsymbol{\eta}) = -\mathbf{Q}(\boldsymbol{\eta}) \left\{\boldsymbol{\Psi} + \beta^2 \mathbf{Q}^2(\boldsymbol{\eta}) \operatorname{diag}(\boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta}) \boldsymbol{\Gamma}^T \operatorname{diag}(\boldsymbol{\eta})\right\} .$$

REFERENCES

[1] F. Wiesinger, P. Boesiger, and K. Pruessmann, "Electrodynamics and ultimate SNR in parallel MR imaging," *Magn. Reson. Medicine*, vol. 52, no. 2, pp. 376–390, Aug. 2004.

[2] D. K. Sodickson and W. J. Manning, "Simultaneous Acquisition of Spatial Harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Medicine*, vol. 38, no. 4, pp. 591–603, 1997.

[3] C. A. McKenzie, M. A. Ohliger, E. N. Yeh, M. D. Price, and D. K. Sodickson, "Coil-by-coil image reconstruction with SMASH," *Magn. Reson. Medicine*, vol. 46, pp. 619–623, 2001.

[4] D. K. Sodickson, C. A. McKenzie, M. A. Ohliger, E. N. Yeh, and M. D. Price, "Recent advances in image reconstruction, coil sensitivity calibration, and coil array design for SMASH and generalized parallel MRI," *Magn. Reson. Materials Biol., Phys. Medicine*, vol. 13, no. 3, pp. 158–163, 2002.

[5] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[6] K. P. Pruessmann, M. Weiger, P. Boernert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Medicine*, vol. 46, pp. 638–651, 2001.

[7] M. Weiger, K. P. Pruessmann, and P. Boesiger, "2-D SENSE for faster 3D MRI," *Magn. Reson. Materials Biol., Phys. Medicine*, vol. 14, no. 1, pp. 10–19, 2002.

[8] J. S. van den Brink, Y. Watanabe, C. K. Kuhl, T. Chung, R. Muthupillai, M. Van Cauteren, K. Yamada, S. Dymarkowski, J. Bogaert, J. H. Maki, C. Matos, J. W. Casselman, and R. M. Hoogeveen, "Implications of SENSE MR in routine clinical practice," *Eur. J. Radiol.*, vol. 46, no. 1, pp. 3–27, Apr. 2003.

[9] 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)," *Magn. Reson. Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.

[10] B. J. Wintersperger, K. Nikolaou, O. Dietrich, J. Rieber, M. Nittka, M. F. Reiser, and S. O. Schoenberg, "Single breath-hold real-time cine MR imaging: Improved temporal resolution using Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA) algorithm," *Eur. J. Radiol.*, vol. 13, no. 8, pp. 1931–1936, May 2003.

[11] Y. Wang, "Description of parallel imaging in MRI using multiple coils," *Magn. Reson. Medicine*, vol. 44, pp. 495–499, 2000.

[12] M. Blaimer, F. Breuer, M. Mueller, R. M. Heidemann, M. A. Griswold, and P. M. Jakob, "SMASH, SENSE, PILS, GRAPPA: How to choose the optimal method," *Top. Magn. Reson. Imag.*, vol. 15, no. 4, pp. 223–236, 2004.

[13] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[14] P. Gravel, G. Beaudoin, and J. A. DeGuise, "A method for modeling noise in medical images," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1221–1232, Oct. 2004.

[15] S. van Huffel and P. Lemmerling, *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*. Norwell, MA: Kluwer, 2002.

[16] A. Raj and R. Zabih, "An optimal maximum likelihood algorithm for parallel imaging in presence of sensitivity noise," in *Proc. ISMRM*, 2004, p. 2411.

[17] F. Lin, K. Kwang, J. Belliveau, and L. Wald, "Parallel imaging reconstruction using automatic regularization," *Magn. Reson. Medicine*, vol. 51, no. 3, pp. 559–567, 2004.

[18] K. King and L. Angelos, "SENSE image quality improvement using matrix regularization," in *Proc. ISMRM*, 2001, p. 1771.

[19] R. Bammer, M. Auer, S. L. Keeling, M. Augustin, L. A. Stables, R. W. Prokesch, R. Stollberger, M. E. Moseley, and F. Fazekas, "Diffusion tensor imaging using single-shot SENSE-EPI," *Magn. Reson. Medicine*, vol. 48, pp. 128–136, 2002.

[20] Z. P. Liang, R. Bammer, J. Ji, N. Pelc, and G. Glover, "Making better sense: Wavelet denoising, Tikhonov regularization, and total least squares," in *Proc. ISMRM*, 2002, p. 2388.

[21] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[22] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1070–1087, May 1991.

[23] V. Z. Mesarovic, N. P. Galatsanos, and A. K. Katsaggelos, "Regularized constrained total least squares image restoration," *IEEE Trans. Image Process.*, vol. 4, no. 8, pp. 1096–1108, Aug. 1995.

[24] P. Lemmerling, B. de Moor, and S. van Huffel, "On the equivalence of constrained total least squares and structured total least squares," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2908–2911, Nov. 1996.

[25] A. Raj, "A signal processing and machine vision approach to problems in magnetic resonance imaging," Ph.D. dissertation, Cornell Univ., Ithaca, NY, Jan. 2005.

[26] V. Z. Mesarovic, N. P. Galatsanos, and M. N. Wernick, "Iterative maximum a posteriori (MAP) restoration from partially-known blur for tomographic reconstruction," in *IEEE Int. Conf. Image Process.*, Oct. 1995, vol. 2, pp. 512–515.

[27] D. Terzolpoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 4, pp. 413–424, Jul. 1986.

[28] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.