

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 due to a noisy signal. 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 (TLS). However, existing TLS algorithms are 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 non-ideal imaging conditions like low SNR, high g-factors, large acceleration and misaligned sensitivity maps.

Keywords

Parallel imaging, Magnetic Resonance, Maximum Likelihood, SENSE, Total Least Squares.

I. PARALLEL IMAGING AND SENSITIVITY NOISE

THE use of multiple coils in MR imaging to substantially reduce scan time (and thus motion artifacts) has become quite popular recently [1]. These parallel imaging techniques are known as SMASH [2], [3], [4]; SENSE [5], [6], [7] [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. All these schemes use multiple coils to reconstruct the unaliased image from under-sampled data in Fourier space, also known as k-space. Each coil also has a sensitivity map depicting its spatial response over the imaging volume. Aliased data are combined using sensitivity maps to reconstruct a full, unaliased image. SENSE is a powerful method to exploit data redundancy from multiple coils and is rightly considered a major breakthrough in MR imaging. It works superbly in well-behaved situations with high SNR and low g-factors, but starts to deteriorate under non-ideal conditions. Some of these issues were highlighted in [12].

In this paper we address a major source of errors in SENSE: the situation where it is difficult to obtain artifact-free sensitivity maps. Now sensitivity maps are computed from an MR scan, typically of a phantom or after division by a body coil image. As a result, they are subject to similar noise processes that affect the data. In addition, the encoding and decoding sensitivity are not identical in practical imaging situations involving physiological motion, misalignment of coils between scans, etc. Under modest acceleration, low g-factors and high SNR, these sensitivity effects may not appear to greatly degrade SENSE performance, but our examples suggest that SENSE produces disturbing artifacts when these ideal conditions do not hold. We propose a robust reconstruction which is tolerant to unreliable sensitivity information. This may potentially open up the field of SENSE imaging to situations where it could not previously be employed. In particular, preliminary results in section V indicate strong performance under many challenging conditions like low-SNR sensitivity maps, sensitivity misalignment due to coil or physiological movement, high acceleration, and in interior regions with poor signal penetration and high g-factors.

The authors are with the Radiology Department, Weill Medical College of Cornell University. Ashish Raj is also with Radiology, University of California at San Francisco, and Ramin Zabih is also with the Cornell Computer Science Department. Address correspondence to: Ramin Zabih, 4130 Upson Hall, Cornell University, Ithaca NY 14853, rdz@cs.cornell.edu.

To some extent, the issue of poor SNR was addressed using regularization [13], [14], [15], and the issue of sensitivity noise using a Total Least Squares (TLS) approach [16]. This paper improves, extends and generalizes the TLS work via a Maximum-Likelihood (ML) formulation. We show that TLS makes assumptions about sensitivity errors which do not occur in general MR imaging practice except under very specific situations. We then obtain new algorithms from basic ML principles which overcome these problems. The parallel imaging process has a linear form

$$\bar{y} = \bar{E}\bar{x} + \bar{n}, \quad y = Ex + n \quad (1)$$

The left equation refers to k-space quantities, and the right one image-space. Coil outputs are denoted by \bar{y} and y , the desired image by \bar{x} and x , and imaging noise by \bar{n} and n . Matrix \bar{E} contains sensitivity and reduced encoding information.¹ A formalization of the imaging process is in section II-A. SENSE reconstruction assumes that n is independent and identically distributed (i.i.d.) Gaussian noise.² SENSE takes a least squares approach, which is well-known to be the maximum likelihood estimate under i.i.d. Gaussian assumption[18, Ch. 15]. We will model the imaging process as

$$y = (E + \Delta E)x + n, \quad (2)$$

where ΔE is the noise in the system matrix that results from errors in the sensitivity maps (i.e., sensitivity noise). At first glance, this appears to be an errors-in-variables problem of the kind commonly addressed with Total Least Squares (TLS) [19]. Indeed, several authors, such as [16], have suggested taking a TLS approach to sensitivity error. However, TLS algorithms assume that ΔE consists of independent elements. We will demonstrate in section II-D that this assumption is generally invalid, due to the specific structure of the system matrix E in parallel imaging.

We propose a maximum likelihood (ML) approach to solving Equation (2), which generalizes both least squares [18] and total least squares [19]. First we derive the general ML result which is applicable for arbitrary noise models and sampling trajectories, assuming only that there is no cross-coil interference. Using this as the foundation, we 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-identically distributed) noise. This models many actual imaging situations where sensitivity maps suffer from uncorrelated but spatially varying noise. The resulting algorithm, which we call Maximum Likelihood SENSE or 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.

This paper is organized as follows. The parallel MR acquisition model is detailed in section II for general and special (Cartesian) case, and our sensitivity noise model is introduced. In section III we discuss related TLS work, and show that they cannot generally handle sensitivity noise. Section IV derives our algorithm using ML principles. For more background and detailed derivations, the reader is referred to [20, Ch. 3]. We show that with Cartesian sampling the general solution reduces to a quasi-quadratic minimization problem directly in image space. We give experimental results on both simulated and clinical data in section V.

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

²See [17] for a study of noise in medical imaging.

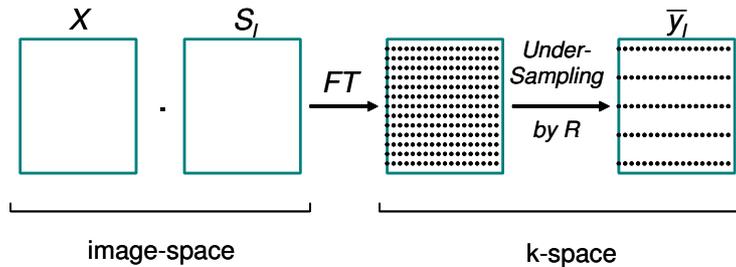


Fig. 1

SCHEMATIC DESCRIPTION OF THE PARALLEL IMAGING PROCESS FOR THE l TH COIL. THE TARGET IMAGE X IS MULTIPLIED VOXEL-BY-VOXEL WITH THE COIL SENSITIVITY S_l . AFTER FOURIER TRANSFORM, THE DATA IS IN K-SPACE. REDUCED ENCODING CORRESPONDS TO UNDERSAMPLING THIS K-SPACE R TIMES, AS SHOWN HERE FOR THE CARTESIAN CASE.

II. MR ACQUISITION MODEL

In this section we obtain the imaging model summarized in (2). Scalars, vectors and 1D objects are denoted in lower case; matrices and 2D 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 ' $x \cdot y$ ' is understood to be element-wise multiply (not the dot product which we denote by $x^T y$). The notation $diag(x)$ represents a diagonal matrix whose diagonal elements are given by those of the enclosed vector x . I is the identity matrix; boldface $\mathbf{1}$ the vector of ones. Q_{ij} denotes the (i, j) -th element of matrix Q ; q_i the i th element of vector q .

A. System model

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

$$\bar{Y}_l(k) = \int dr e^{-i2\pi rk} S_l(r) X(r). \quad (3)$$

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

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

This can be discretized by imposing a Cartesian grid on both r and k . Let vectors x , s_l and 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 2D-FT now becomes the 2-D DFT and the resampling over k may be accomplished by using a general downsampling operator in k-space. This process is depicted in Figure 1.

The input-output relationship of the l -th coil in Figure 1 is succinctly expressed as a matrix product

$$y_l = E_l x = D_{N \times M}^H \Downarrow_R D_{N \times M} S_l \cdot X = D_{N \times M}^H \Downarrow_R D_{N \times M} diag(s_l) x. \quad (5)$$

The k-space downsampling operator \Downarrow_R resamples k-space according to the specific trajectory used during the scan. Here 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, and superscript H the Hermitian operation. The operator $D_{N \times M}$ 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 G the gridding operator corresponding to a Kaiser-Bessel kernel, then the modified downsampling operator will be given by $\Downarrow'_R = \Downarrow_R G$. Henceforth we shall assume \Downarrow_R incorporates gridding, if any.

B. System 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 Equation (5) is now redefined as a *sub-sampling* operator, equivalent to removing rows of k-space.

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

$$y_l = (D_{N/R}^{col})^H \Downarrow_R D_N^{col} \text{diag}(s_l) x, \quad (6)$$

This equation can be solved separately for each column. Further, this degenerates into individual $L \times R$ aliasing equations, according to Theorem 1.

Theorem 1: Let $y_l^{(i)}$, $s_l^{(i)}$, $x^{(i)}$ be the i -th column of Y_l , S_l , X , and (6) be denoted by $y_l = E_l x^{(i)}$. Consider a partitioning of these signals into R aliasing components under Cartesian sampling:

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

whose j -th element is given by $x^{(i)}_r(j) = x^{(i)}(\frac{N}{R}(r-1) + j)$, $s_l^{(i)}_r(j) = s_l^{(i)}(\frac{N}{R}(r-1) + j)$. Then

$$y_l^{(i)} = \sum_{r=1}^R s_l^{(i)}_r \cdot x^{(i)}_r.$$

2. E has a diagonal-block structure containing $L \times R$ diagonal blocks:

$$E = \{E_l^r\}_{\substack{r=1 \dots R \\ l=1 \dots L}}$$

where each sub-block E_l^r is diagonal, with $E_l^r = \text{diag}(s_{l,r}^{(i)})$.

Proof: Proof in Appendix-A. ■

Theorem 1 is pictorially depicted in Figure 2. To maintain readability, we will henceforth drop column superscript (i) . Symbols E , x and y etc. will be used both for arbitrary and Cartesian sampling, their meaning indicated by context. Now each block is diagonal according to Theorem 1, so the

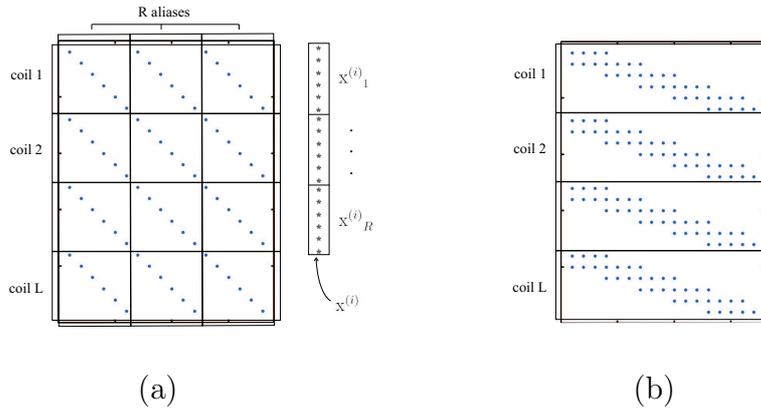


Fig. 2

STRUCTURE OF SYSTEM MATRIX UNDER REGULAR CARTESIAN SAMPLING. NON-ZERO ELEMENTS ARE INDICATED WITH AN ASTERISK. PART (A) SHOWS \bar{E} (IMAGE SPACE). AS A CONSEQUENCE OF THE PARTITIONING, IMAGE COLUMN $x^{(i)}$ SEPARATES INTO R ALIASING COMPONENTS. PART (B) SHOWS \bar{E} , THE FOURIER DUAL OF \bar{E} .

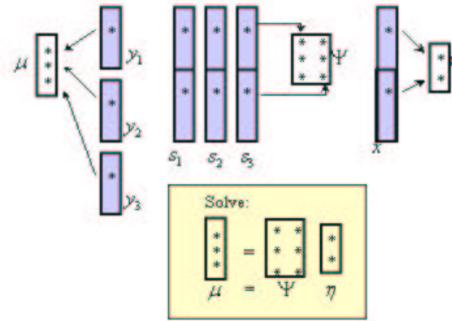


Fig. 3

CONVERTING THE FULL CARTESIAN PROBLEM FOR EACH COLUMN INTO A SET OF INDEPENDENT LINEAR SYSTEMS OF L OBSERVATIONS AND R UNKNOWN. VECTOR μ IS $L \times 1$ AND η IS $R \times 1$; MATRIX Ψ IS $L \times R$. THIS PROCESS IS REPEATED FOR EVERY SET OF R ALIASING VOXELS, AND THEN FOR EACH IMAGE COLUMN.

interactions are restricted to only R aliasing voxels at a time. Indeed, define $\mu \in \mathcal{C}^L$, $\eta \in \mathcal{C}^R$, $\Psi \in \mathcal{C}^{L \times R}$. Then for each (j, i) -th voxel in aliased images y_l , the SENSE problem becomes:

$$\begin{aligned} \text{Given} \quad & \mu(l) = y_l^{(i)}(j), \\ & \eta(r) = x^{(i)}(\text{mod}(j, R)), \\ & \Psi(l, r) = s_l^{(i)}(\text{mod}(j, R)), \\ \text{Solve} \quad & \mu = \Psi \eta. \end{aligned}$$

This process is pictorially depicted in Figure 3. The process in the figure is repeated for every set of aliasing voxels, until the entire image is reconstructed.

C. System matrix structure under arbitrary sampling

System matrices E , \bar{E} have important special forms. Individual blocks E_1, \dots, E_L are diagonal, and $\bar{E}_1, \dots, \bar{E}_L$ would have had a Toeplitz structure but for row-decimation due to under-sampling, as shown in Figure 2(b). (Recall that a Toeplitz matrix is a matrix $T = \{T_{ij}\}$ such that $T_{ij} = t_{j-i}$,

which is the $(j - i)$ th element of 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[17]. The l th coil sensitivity and output noise are modeled as

$$\begin{aligned} S_l^{noisy} &= S_l + N_l^s \\ Y_l^{noisy} &= Y_l + N_l \end{aligned}$$

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 $Var(\mathbf{n}_l) = \sigma_n \omega_l$ and $Var(\mathbf{n}_l^s) = \sigma_s \lambda_l$, where we have introduced normalized variance vectors ω_l and λ_l . Define for convenience $\Omega_l = diag(\omega_l)$, $\Lambda_l = diag(\lambda_l)$. Then the autocorrelation matrices of output and sensitivity noise are given by

$$\begin{aligned} \mathcal{E}(\mathbf{n}_l \mathbf{n}_l^H) &= \sigma_n^2 \Omega_l^2, \\ \mathcal{E}(\mathbf{n}_l^s \mathbf{n}_l^{sH}) &= \sigma_s^2 \Lambda_l^2 \end{aligned}$$

Clearly, the structure of $\Delta \mathbf{E}$ must mimic that of \mathbf{E} shown in Figure 2(a):

$$\Delta \mathbf{E} = \{ \Delta \mathbf{E}_l^r \}_{\substack{r=1 \dots R \\ l=1 \dots L}},$$

and the same holds for the k -space versions $\Delta \bar{\mathbf{E}}$ and $\bar{\mathbf{E}}$. Again, each sub-block $\Delta \mathbf{E}_l^r$ is diagonal, with entries given by the sensitivity map noise terms N_l^s . Similarly, in k -space the error matrix $\Delta \bar{\mathbf{E}}$ mimics the structure of $\bar{\mathbf{E}}$ as shown in Figure 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 non-identically 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.³ 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 non-identically distributed noise. Noise correlation across coils can be accommodated by pre-multiplying \mathbf{E} with a “whitening” matrix to remove all voxel-wise correlations among coils. Pre-whitening 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. RELATED WORK

Equation (2) appears to be an errors-in-variables problem of the kind traditionally solved with TLS. Unfortunately TLS makes the unrealistic assumption of independent matrix elements. There are variants of TLS, such as Constrained Total Least Squares, that can handle a broad class of matrix structure, including the structures that arise in parallel imaging. However, these approaches require the use of very general minimization techniques, which are very inefficient.

³This is why we explicitly allow for the noise variance of coil output and coil sensitivity to be different.

A. Total Least Squares

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 $\Delta\mathbf{E}$, as follows:

$$\hat{\mathbf{x}}_{TLS} = \arg \min_{\mathbf{x}} \|\begin{bmatrix} \Delta\mathbf{E} \\ \mathbf{n} \end{bmatrix}\|_F, \quad \text{subject to } \mathbf{n} + \Delta\mathbf{E}\mathbf{x} = \mathbf{y} - \mathbf{E}\mathbf{x} \quad (7)$$

where the indicated norm is Frobenius. This formulation assumes that the elements of $\Delta\mathbf{E}$ are independent (i.e., that $\Delta\mathbf{E}$ has no structure).

Unfortunately, TLS is ill-equipped to handle the specific system model described in §II. In Cartesian sampling $\Delta\mathbf{E}$ has a diagonal block structure as shown in Figure 2, with off-diagonal elements being zero. So even if the underlying sensitivity noise process is uncorrelated, the elements of $\Delta\mathbf{E}$ are never independent (off-diagonal entries being identically zero). Thus, the independence assumption of conventional TLS is generally violated. A similar situation occurs in k-space (1). Matrix $\Delta\bar{\mathbf{E}}$ has a Toeplitz-type structure shown in Figure 2(b). This results in the elements of $\Delta\bar{\mathbf{E}}$, being algebraically related to each other rather than being independent.

The only instance where standard TLS can actually be used in SENSE 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 reported in [16]. Unfortunately, there is no k-space equivalent of this approach; nor does it extend to non-uniform noise models like those we employ in this paper. In practice, it is frequently preferable to reconstruct entire columns or entire image together, for instance to exploit some *a priori* knowledge. Non-Cartesian data too must be reconstructed over the entire image. Direct application of TLS, while conceptually simple, is incompatible with these situations. In contrast, the proposed approach does not have these limitations, and has the added advantage of statistical optimality (in ML sense) under a large class of noise models.

B. Constrained Total Least Squares

Several generalizations of TLS, collectively known as *Constrained* TLS (CTLS), have been proposed to handle matrix structure.⁴ CTLS was proposed by [24], whose work handles *linearly structured* matrices — those matrices that can be obtained from a linear combinations of a smaller perturbation vector. For a linearly structured matrix \mathbf{E} , the CTLS approach works as follows. Define an augmented matrix $\mathbf{C} = [\mathbf{E}|\mathbf{y}]$, and a perturbation in \mathbf{C} as $\Delta\mathbf{C} = [\Delta\mathbf{E}|\mathbf{n}]$. CTLS consists of solving

$$\min_{\mathbf{v}, \mathbf{x}} \|\mathbf{v}\|, \quad \text{subject to } (\mathbf{C} + \Delta\mathbf{C}) \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} = 0 \quad \text{AND} \quad \Delta\mathbf{C} = [F_1\mathbf{v}|F_2\mathbf{v}|\cdots|F_{N+1}\mathbf{v}],$$

where the F_i 's are matrices that generate the elements of $\Delta\mathbf{C}$ from \mathbf{v} . This problem is difficult to solve for arbitrary \mathbf{E} , requiring slow general-purpose constrained minimization techniques.

Matrices described in II turn out to be linearly structured. However, by taking advantage of the particular structure of the system matrix, we can use much more efficient special-purpose unconstrained minimization methods. Further details of TLS methods in MR can be found in [20, Ch. 3].

IV. THE 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 non-linear least squares algorithm.

⁴An earlier approach, called *Structured* TLS [22], was shown to be equivalent to CTLS in [23].

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

$$\ell(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\{(\mathbf{y} - \mathbf{E}\mathbf{x})^H \mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{E}\mathbf{x})\}\right) \quad (8)$$

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

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} (\mathbf{y} - \mathbf{E}\mathbf{x})^H \mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{E}\mathbf{x}). \quad (9)$$

Under our noise model, $\mathbf{R}_{\mathbf{g}|\mathbf{x}} = \mathcal{E}(\mathbf{nn}^H + (\Delta\mathbf{E}\mathbf{x})(\Delta\mathbf{E}\mathbf{x})^H)$, and $\mathcal{E}(\mathbf{nn}^H) = \sigma_n^2 \mathbf{\Omega}^2$. 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 [25] exhibited little improvement after the log term was included, at substantial computational cost. Similar behaviour was observed during our experimentation. Consequently, we drop this term henceforth.

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

$$(\Delta\mathbf{E}_l\mathbf{x})(\Delta\mathbf{E}_{l'}\mathbf{x})^H = \mathbf{D}_{N/R \times M}^H \Downarrow_R \mathbf{D}_{N \times M} \text{diag}(\mathbf{x}) \mathcal{E}(\Delta s_l \Delta s_{l'}^H) \text{diag}(\mathbf{x}) \mathbf{D}_{N \times M}^H \Downarrow_R \mathbf{D}_{N/R \times M}, \quad (10)$$

which follows from:

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

Now $\mathcal{E}(\Delta s_l \Delta s_{l'}^H) = \sigma_s^2 \delta_{l,l'} \Lambda_l^2$ since we assume coils are decoupled, therefore

$$\mathbf{R}_{\mathbf{g}|\mathbf{x}} = \sigma_n^2 \left(\mathbf{\Omega}^2 + \beta^2 \begin{bmatrix} \mathbf{A}_1(\mathbf{x}) & & \\ & \ddots & \\ & & \mathbf{A}_L(\mathbf{x}) \end{bmatrix} \right), \quad (11)$$

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

Finally, we have

$$\mathbf{R}_{\mathbf{g}|\mathbf{x}}^{-1} = \frac{1}{\sigma_n^2} \begin{bmatrix} \mathbf{B}_1(\mathbf{x})^{-1} & & \\ & \ddots & \\ & & \mathbf{B}_L(\mathbf{x})^{-1} \end{bmatrix}, \quad (12)$$

$$\mathbf{B}_l(\mathbf{x}) = \mathbf{\Omega}_l^2 + \beta^2 \mathbf{D}_{N/R \times M}^H \Downarrow_R \mathbf{D}_{N \times M} \text{diag}(|\Lambda_l \mathbf{x}|^2) \mathbf{D}_{N \times M}^H \Downarrow_R \mathbf{D}_{N/R \times M}. \quad (13)$$

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

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \sum_l (\mathbf{y}_l - \mathbf{E}_l\mathbf{x})^H \mathbf{B}_l(\mathbf{x})^{-1} (\mathbf{y}_l - \mathbf{E}_l\mathbf{x}). \quad (14)$$

Let us summarize the significance of Equation (14): 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, (14) 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, Equation (14) is a non-quadratic 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, Ω_l and Λ_l become diagonal, and large simplifications result. Each function evaluation of (14) for general, arbitrarily sampled data involves the inversion of $NM \times NM$ generally non-sparse matrices, making direct inversion prohibitive. However, inversion may be efficiently performed iteratively since the products $B_l(\mathbf{x})\mathbf{p}$ and $B_l(\mathbf{x})^H\mathbf{p}$ 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, $B_l(\mathbf{x})$ are obviously well-conditioned due to diagonal Ω_l and Λ_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 non-linear least squares algorithms can be used to solve the problem efficiently (see [18, Ch. 10]). The ML estimate (14) will not only reduce drastically in complexity.

We do not further specify an implementation for the general case in this paper, focussing 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 $B_l(\mathbf{x}^{(i)})$. 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 $\mathbf{b}_l(\mathbf{x}^{(i)}) \triangleq 1 + \beta^2 \sum_{r=1}^R |\mathbf{x}_r^{(i)}|^2, l \in \{1, \dots, L\}$.

Then the ML estimate (14) of column (i) under Cartesian sampling is given by

$$\hat{\mathbf{x}}^{(i)} = \arg \min_{\mathbf{x}} \sum_l \|(y_l^{(i)} - \mathbf{s}_{l,r}^{(i)} \cdot \mathbf{x})\|^2 / \mathbf{b}_l(\mathbf{x}).$$

For non-i.i.d. noise with Λ and Ω : The Cartesian ML estimate is given by

$$\begin{aligned} \mathbf{x}'_l^{(i)} &\triangleq \lambda_l^{(i)} \cdot \mathbf{x}^{(i)}, \\ \mathbf{b}'_l(\mathbf{x}^{(i)}) &\triangleq \omega_l^{(i)2} + \beta^2 \sum_{r=1}^R |\mathbf{x}'_{l,r}{}^{(i)}|^2, \\ \hat{\mathbf{x}}^{(i)} &= \arg \min_{\mathbf{x}} \sum_l \|(y_l^{(i)} - \mathbf{s}_{l,r}^{(i)} \cdot \mathbf{x}')\|^2 / \mathbf{b}'_l(\mathbf{x}). \end{aligned}$$

Proof: Note that the division ‘/’ is element-by-element. Proof in Appendix-B. ■

Now for each aliasing voxel (j, i), define η, μ, Ψ , as before. Then the ML problem reduces to solving

$$\hat{\eta} = \arg \min_{\eta} \|F(\eta)\|^2, \quad F(\eta) = q(\eta)(\mu - \Psi\eta), \quad q(\eta) = 1/\sqrt{1 + \beta^2\|\eta\|^2}. \quad (15)$$

Figure 4 shows the algorithm to implement the minimization of (14) 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 non-quadratic term. But the non-quadraticity enters the equation only via a well-behaved, smooth, slowly-varying function of the norm $\|\eta\|^2$. Consequently, minimization can be achieved using a non-linear least-squares method with Newton iterations [18]. This is similar to the standard least squares method for solving the pseudoinverse Ψ^\dagger . The only difference is that the Jacobian of F is not a constant matrix any more. In Appendix-F we briefly

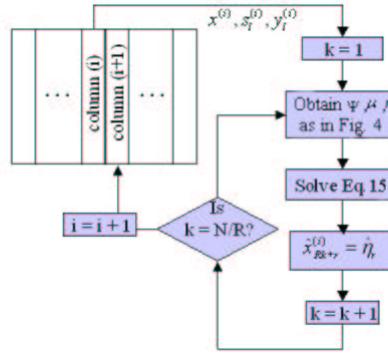


Fig. 4

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 μ , η AND Ψ . THE RESULTING $L \times R$ SYSTEM IS SOLVED, AND THE ELEMENTS OF x ARE PROGRESSIVELY FILLED IN.

describe the standard Newton algorithm for minimizing (15), 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):

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

D. Extension to Non-uniform 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 ω_l in the L -vector u , and λ_l in R -vectors r_l similar to μ and Ψ before. Consequently, the minimization proceeds exactly as before, with $q(\eta)$ now replaced by the $L \times L$ diagonal matrix $Q(\eta)$:

$$\hat{\eta} = \arg \min_{\eta} \|F(\eta)\|^2, \quad F(\eta) = Q(\eta)(\mu - \Psi\eta), \quad Q_u(\eta) = 1/\sqrt{u_l^2 + \beta^2 \|r_l \cdot \eta\|^2}. \quad (17)$$

The new algorithm, Algorithm II, is further specified in Appendix-D, and follows the same structure as shown in Figure 4 above. ML-SENSE II is slightly more challenging numerically compared to ML-SENSE I. But again, non-quadraticity is only on account of a slowly-varying functional of norm $\|r_l \cdot \eta\|^2$. The Jacobian is slightly different, but still efficiently computable. Let $\Gamma = [r_1^2, \dots, r_L^2]$. Then the new Jacobian is (see Appendix-G)

$$J(\eta) = \frac{\partial F(\eta)}{\partial \eta} = -Q(\eta)\{\Psi + \beta^2 Q^2(\eta) \text{diag}(\mu - \Psi\eta)\Gamma^T \text{diag}(\eta)\}. \quad (18)$$

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 non-uniform noise.

We propose here one method of obtaining ω_l and λ_l , used in our experiments with ML-SENSE II. Local noise variance was computed for each voxel of each coil using a weighted neighbourhood around the voxel. The weights were a monotonically decreasing function, in our work a quadratic, centered at the voxel. We chose a cut-off radius of $N/20$ for this neighbourhood. For our purposes we used unaccelerated data to obtain these estimates. For this reason our method is difficult to extend to cases

Algorithm	SENSE	ML-SENSE I	ML-SENSE II
flops per iteration	$\mathcal{O}(2MNL)$	$\mathcal{O}(3MN(L + 1))$	$\mathcal{O}(3MN(L + 2))$
avg iterations	10	30	30

TABLE I

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

where representative unaccelerated scans are not available. More sophisticated methods are currently being investigated; however, we note that in many cases accurate estimates of ω_l and λ_l may not ultimately be available. Therefore we describe Algorithms I and II separately – in absence of full noise statistics, ML-SENSE I is sub-optimal but preferable.

E. Computational Burden

The additional cost of non-quadratic 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 version R13. Typical execution times for reconstructions of size 256×256 were between three 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 non-quadratic.

V. 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.

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 and ML-SENSE is performed on data from a Shepp-Logan head phantom where both the coil data and simulated sensitivity maps are noiseless. Shown in Figure 5, this demonstrates that if there is no sensitivity error, all methods perform perfect reconstruction.

Next we simulation 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 Figure 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. In contrast our ML-SENSE algorithm is able to salvage more useful data.

A quantitative comparison is now performed for $R = 4$. For a given SNR (labeled “input SNR” in Figure 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

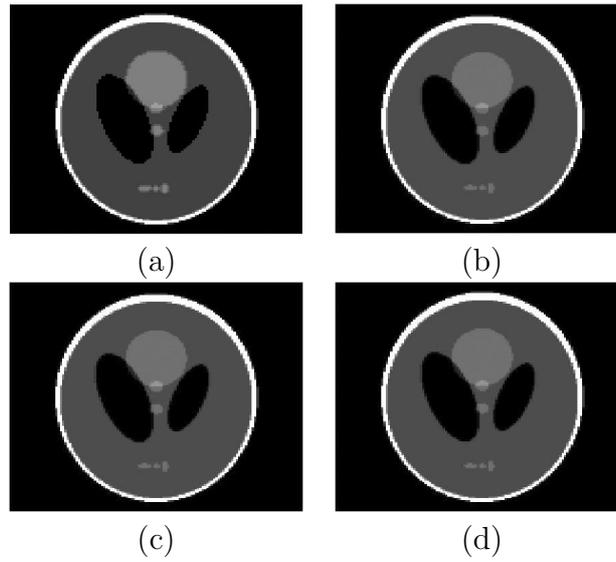


Fig. 5

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) ML-SENSE I RECONSTRUCTION, AND (D) ML-SENSE II. EACH RECONSTRUCTION IS PERFECT.

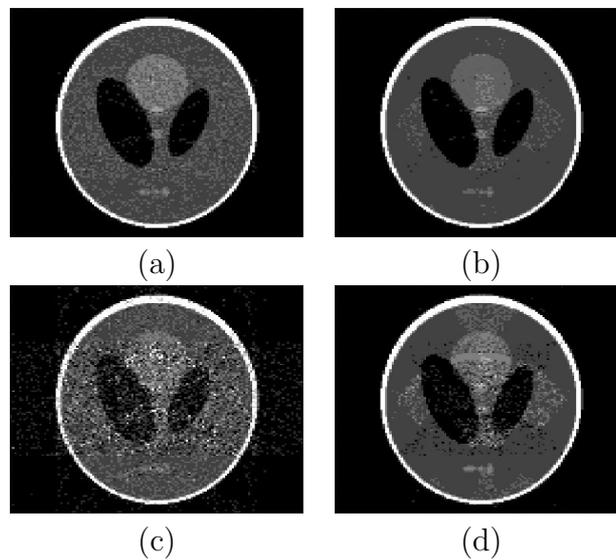


Fig. 6

RECONSTRUCTION OF NOISY SIMULATED SHEPP-LOGAN PHANTOM DATA, WITH $R = 4$, $L = 6$. (A) SENSE WITH MID-LEVEL NOISE, (B) ML-SENSE I, (C) SENSE WITH LARGE NOISE, (D) ML-SENSE I. NOTE THE EXCESSIVE NOISE AMPLIFICATION DUE TO VERY HIGH G-FACTOR IN THE MIDDLE OF FOV.

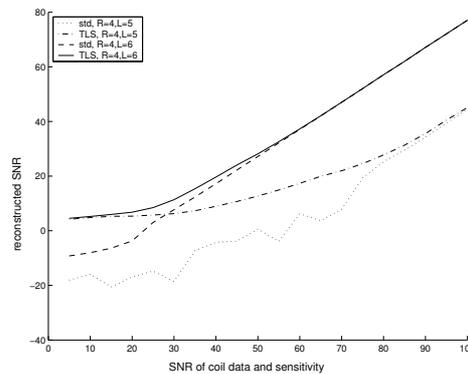


Fig. 7

THE SNR PERFORMANCE OF STANDARD AND ML-SENSE. BOTH INPUT AND OUTPUT SNR WERE OBTAINED FROM A DIRECT COMPARISON WITH THE ORIGINAL, UNACCELERATED IMAGE.

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.

B. Experiments with Sensitivity mismatch on high-resolution phantom data

Phased array 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 under the same conditions with a uniform spherical phantom (USP). However, in addition we have contributions from the tube. Unaliased data received by one coil and its measured sensitivity are shown in Figure 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 parallel imaging. A small mis-alignment of nearly 1/10 th of FOV was observed between sensitivity and data. While not standard, this experiment mimics the errors expected if sensitivity maps are mis-aligned, 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. The reconstruction results are shown in Figure 8. We also show a Tikhonov-regularized [26] version of SENSE [15], [13] to see whether these errors can be mitigated simply by regularizing the SENSE reconstruction. SENSE is unsatisfactory due to mismatch and aliasing which can not be easily improved even by regularizing. SNR improves with regularization, but at the cost of inadequate unfolding. This is expected - for large λ_{reg} matrix inverse corresponds to a copy of the observation, i.e. no unfolding. ML-SENSE results are more effective at both noise reduction and unfolding. The effect of sensitivity mismatch and aliasing are visible but considerably muted. As expected, ML-SENSE-II outperforms ML-SENSE-I, but note that ML-SENSE-I is quite effective although sensitivity noise is not i.i.d.

C. Parallel Brain Imaging With An 8-Element Head Coil

Our first in vivo results are of brain imaging under low SNR and sub-optimal coil configuration (high g-factor). Fully sampled data were obtained and under-sampling was imposed manually. 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. We first show the case of $R = 2$ times acceleration in Figure 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 (c-d) showing a small improvement over both SENSE (a) and its regularized version (b). Now consider the same data, but with $R = 4$, shown in Figure 10. SENSE reconstruction in (a) now exhibits excessive noise amplification due to a combination of high R and erroneous sensitivity maps. In order to demonstrate that this problem cannot really be addressed by regularized SENSE, (b) shows the output of a Tikhonov-regularized SENSE algorithm, with $\lambda_{reg} = 0.1$, chosen after an exhaustive L-curve analysis [13] by varying λ_{reg} in increments given by a geometric progression from 0 to 1. At each increment a data and a prior cost were computed:

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

Figure 11 plots these values and is called the L-curve. This is zero-th order regularization since it does not use a prior mean image [14]. In most interesting situations a prior mean image is not available. In our experience means computed from central k-space data produced severe boundary artifacts. Even in dynamic imaging, mean priors are useful only if the variance from mean is small.

While regularized SENSE, which was previously proposed for parallel imaging by several authors [13], [15], [16] 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. Figure 10(c)&(d) show the ML-SENSE reconstructions, which seem to suffer neither from excessive noise amplification nor ghosting. ML-SENSE II is slightly better than ML-SENSE I as expected. Comparing Figures 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. Imaging Experiments Under Non-Uniform 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 and decoding sensitivities. 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 Ω and Λ 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 Figure 12.

Next we investigate the non-uniform sensitivity noise resulting from mis-registration due to breathing, a constant problem in torso scanning. An axial torso slice with 3x speed-up was scanned using the 8-channel head coil described earlier. The central 30 k-space views were scanned densely, allowing us to obtain unaliased low-frequency sensitivity maps by the division method. In addition to mis-registration, spatially variable sensitivity error was observed due to division in areas of weak signal. ML-SENSE- II is the appropriate method in this case. We obtained Λ and Ω by the local window variance estimation described earlier. Results are shown in Figure 13(a) and (b), along with zoomed in region containing the stomach-heart interface in (c) and (d). Several artifacts contaminate the SENSE output, including loss of heart-stomach boundary definition, hyperintensity in heart reion, and stripe artifact across the liver. These artifacts are largely absent in the reconstruction using ML-SENSE II.

VI. CONCLUSIONS

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 non-quadratic objective function which may be solved efficiently via a non-linear 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. The algorithm 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 the algorithm need to be further evaluated under various clinical settings to assess its true clinical significance.

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.

APPENDICES

A. Proof of Theorem 1

Consider the 1D DFT matrix $D_N^{col} = \{e^{i\frac{2\pi}{N}(kn)}\}, k = 0 \dots N-1, n = 0, \dots, N-1$. Now the row-decimated matrix obtained by retaining every R th row in D_N^{col} can be written as $\{e^{i\frac{2\pi}{N}(Rk'n)}\}, k' = 0 \dots N/R-1, n = 0, \dots, N-1$. Expanding this in terms of $(N/R \times N/R)$ -blocks, we get, for $k' = 0 \dots N/R-1, n' = 0 \dots N/R-1$

$$\Downarrow_R D_N^{col} = \{e^{i\frac{2\pi}{N}(Rk'n')}, e^{i\frac{2\pi}{N}(Rk'(n'+\frac{N}{R}))}, \dots, e^{i\frac{2\pi}{N}(Rk'(n'+(R-1)\frac{N}{R}))}\}.$$

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

$$\Downarrow_R D_N^{col} = D_{N/R}^{col} [\mathbf{I}_{N/R} \cdots \mathbf{I}_{N/R}], \quad (19)$$

Therefore $\mathbf{y}^{(i)}_l = \mathbf{E}_l \mathbf{x}^{(i)} = [\mathbf{I}_{N/R} \cdots \mathbf{I}_{N/R}] \text{diag}(s_l^{(i)}) \mathbf{x}^{(i)} = [\text{diag}(s_1^{(i)}), \dots, \text{diag}(s_R^{(i)})] \mathbf{x}^{(i)}$.

This proves part (1) and leads immediately to the partitioning $\mathbf{E}_l = [\mathbf{E}_l^1, \dots, \mathbf{E}_l^R]$, with $\mathbf{E}_l^r = \text{diag}(s_{l,r}^{(i)})$. 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, $\Lambda = \mathbf{I}, \Omega = \mathbf{I}$. Recall that $\Downarrow_R D_{N \times M} = D_M^{row} \Downarrow_R D_N^{col}$ and $D_{N/R}^{col} \Downarrow_R D_N^{col} = [\mathbf{I}_{N/R} \cdots \mathbf{I}_{N/R}]$ from Theorem 1. Then we have for the column-wise matrix $\mathbf{B}_l(\mathbf{x}^{(i)})$ (see Eqn. (13)):

$$\begin{aligned} \mathbf{B}_l(\mathbf{x}^{(i)}) &= \mathbf{I} + \beta^2 D_{N/R}^{col} \Downarrow_R D_{N/R}^{col} \text{diag}(|\mathbf{x}^{(i)}|^2) D_{N/R}^{col} \Downarrow_R D_{N/R}^{col} \\ &= \mathbf{I} + \beta^2 [\mathbf{I}_{N/R} \cdots \mathbf{I}_{N/R}] \text{diag}(|\mathbf{x}^{(i)}|^2) [\mathbf{I}_{N/R} \cdots \mathbf{I}_{N/R}]^T \\ &= \text{diag}(1 + \beta^2 \sum_{r=1}^R |x_r^{(i)}|^2) \triangleq [\text{diag}(b_l(\mathbf{x}^{(i)}))]^{-1} \end{aligned}$$

Then the ML problem (14) for a single column becomes

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

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 Λ and Ω . ■

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

- Y_l = coil output of l th coil, in spatial domain
- S_l = sensitivity map of l th coil
- X = desired MR image of size $(N \times M)$
- L = number of coils
- R = downsampling factor.
- for $i = 1 \dots M$
 1. Define x, y_l, s_l as the i th column of X, Y_l, S_l , respectively.
 2. for $k = 1 \dots N/R$
 - (a) Define a $L \times R$ matrix Ψ , with $\Psi_{l,r} = s_{l,(r-1)N/R+k}$. Let $\mu = [y_{1,k}, \dots, y_{L,k}]^T$.
 - (b) Solve $\hat{\eta} = \arg \min_{\eta} (\frac{1}{1+\beta^2 \|\eta\|^2}) \|\mu - \Psi\eta\|^2$
 - (c) $\hat{x}_{(r-1)N/R+k} = \hat{\eta}_r$
 3. i th column of $X = \hat{x}$.

D. ML-SENSE Under Non-Uniform Sensitivity And Output Noise: ML-SENSE II

- for $i = 1 \dots M$
 1. Define x, y_l, s_l as the i th column of X, Y_l, S_l , respectively.
 2. for $k = 1 \dots N/R$
 - (a) Define $L \times R$ matrix Ψ , with $\Psi_{l,r} = s_{l,(r-1)N/R+k}$.
 - (b) For each $l = 1, \dots, L$, define R -vectors $r_l = \lambda_l^{(r-1)N/R+k,i}$ and $u_l = \omega_l^{k,i}$.
 - (c) Define $L \times R$ matrix Γ , with $\Gamma = [r_1^2, \dots, r_L^2]$.
 - (d) Define $\mu = [y_{1,k}, \dots, y_{L,k}]^T$.
 - (e) Solve $\hat{\eta} = \arg \min_{\eta} \|\mathbf{Q}(\eta)(\mu - \Psi\eta)\|^2$, with $Q_{ll} = \frac{1}{\sqrt{u_l^2 + \beta^2 \|r_l \cdot \eta\|^2}}$
 - (f) $\hat{x}_{(r-1)N/R+k} = \hat{\eta}_r$
 3. i th column of $X = \hat{x}$.

E. SENSE and Regularized SENSE Implementation

Each $L \times R$ sub-system $\mu = \Psi\eta$ is solved separately in SENSE, then the elements of the full image X are filled in from the estimates of η . Matrix Ψ is inverted through the pseudo-inverse via the popular Conjugate Gradients algorithm described previously by many authors, e.g. [5] and [6]. Thus

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

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

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

F. ML-SENSE Cost Minimization Routine

The minimization of Eqs. (15) and (17) proceeds via the well-established Gauss-Newton method [27]. In the approximate vicinity of the true solution, the Hessian is given by

$$H(\eta) \approx J(\eta)^T J(\eta).$$

The Gauss-Newton method computes, at each iteration k , a line search direction d_k starting from the current solution η_k which is the minimizer of the following least squares problem:

$$\min_{d_k} \|J(\eta_k)d_k - F(\eta_k)\|^2.$$

Since the Jacobian and function evaluations are explicitly available and cheaply computable via Eqs. (15) – (18), the above is a simple least squares problem which was solved by conventional CG algorithm. Finally, a one-dimensional line search is performed for each direction d_k using the standard method described in Section 2-6 of [27].

G. Jacobian Of $F(\eta)$

Let r_i , Γ be as defined in § IV-D, let the i th element of $F(\eta)$ be $F_i(\eta)$, and i th row of Ψ be ψ_i^T . Then

$$\frac{\partial F_i(\eta)}{\partial \eta} = -Q_{ii}(\eta)\psi_i + \frac{\partial Q_{ii}(\eta)}{\partial \eta}(\mu_i + \psi_i^T \eta)$$

Now $\frac{\partial Q_{ii}(\eta)}{\partial \eta} = -\beta^2 Q_{ii}^3(\eta)(r_i^2 \cdot \eta)$. Then

$$J_i(\eta) = \left(\frac{\partial F_i(\eta)}{\partial \eta} \right)^T = -Q_{ii}(\eta) \{ \psi_i^T + \beta^2 Q_{ii}^2(\eta) (\mu_i + \psi_i^T \eta) \eta^T \text{diag}(r_i^2) \}$$

For algorithm I, $r_i^2 = 1$, $Q_{ii}(\eta) = q(\eta)$, and we get

$$J(\eta) = \begin{bmatrix} J_1(\eta) \\ \vdots \\ J_L(\eta) \end{bmatrix} = -q(\eta) \{ \Psi + \beta^2 q^2(\eta) (\mu + \Psi \eta) \eta^T \}.$$

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

$$J(\eta) = -Q(\eta) \{ \Psi + \beta^2 Q^2(\eta) \text{diag}(\mu - \Psi \eta) \Gamma^T \text{diag}(\eta) \}.$$

REFERENCES

- [1] Florian Wiesinger, Peter Boesiger, and Klaas Pruessmann, “Electrodynamics and ultimate SNR in parallel MR imaging,” *Magnetic Resonance in Medicine*, vol. 52, no. 2, pp. 376–390, August 2004.
- [2] D. K. Sodickson and W. J. Manning, “Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays,” *Magnetic Resonance in 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,” *Magnetic Resonance in 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,” *Magnetic Resonance Materials in Biology, Physics and Medicine*, vol. 13, no. 3, pp. 158–63, 2002.
- [5] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and Peter Boesiger, “SENSE: sensitivity encoding for fast MRI,” *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 2001.
- [6] K. P. Pruessmann, M. Weiger, Peter Boernert, and Peter Boesiger, “Advances in sensitivity encoding with arbitrary k-space trajectories,” *Magnetic Resonance in Medicine*, vol. 46, pp. 638–651, 2001.
- [7] M. Weiger, K. P. Pruessmann, and P. Boesiger, “2d sense for faster 3d mri,” *Magnetic Resonance Materials in Biology, Physics and Medicine*, vol. 14, no. 1, pp. 10–19, 2002.
- [8] Johan S. van den Brink, Yuji Watanabe, Christian K. Kuhl, Taylor Chung, Raja Muthupillai, Marc Van Cauteren, Kei Yamada, Steven Dymarkowski, Jan Bogaert, Jeff H. Maki, Celso Matos, Jan W. Casselman, and Romhild M. Hoogeveen, “Implications of SENSE MR in routine clinical practice,” *European Journal of Radiology*, vol. 46, no. 1, pp. 3–27, April 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)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–10, 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," *European Journal of Radiology*, vol. 13, no. 8, pp. 1931–1936, May 2003.
- [11] Yi Wang, "Description of parallel imaging in mri using multiple coils," *Magnetic Resonance in Medicine*, vol. 44, pp. 495499, 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. Magnetic Resonance Imaging*, vol. 15, no. 4, pp. 223–36, 2004.
- [13] F. Lin, K. Kwang, J. Belliveau, and L. Wald, "Parallel imaging reconstruction using automatic regularization," *Magnetic Resonance in Medicine*, vol. 51, no. 3, pp. 559–567, 2004.
- [14] K. King and L. Angelos, "Sense image quality improvement using matrix regularization," in *Proc. ISMRM*, 2001, p. 1771.
- [15] 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," *Magnetic Resonance in Medicine*, vol. 48, pp. 128–136, 2002.
- [16] 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.
- [17] P. Gravel, G. Beaudoin, and J. A. DeGuisé, "A method for modeling noise in medical images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1221–1232, October 2004.
- [18] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery, *Numerical Recipes in C*, Cambridge, 2nd edition, 1992.
- [19] Sabine Van Huffel and Philippe Lemmerling, *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Kluwer, 2002.
- [20] Ashish Raj, *A Signal Processing And Machine Vision Approach to Problems in Magnetic Resonance Imaging*, Ph.D. thesis, Cornell University, January 2005.
- [21] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.
- [22] Vladimir Z. Mesarovic, Nikolas P. Galatsanos, and Aggelos K. Katsaggelos, "Regularized Constrained Total Least Squares image restoration," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1096–1108, August 1995.
- [23] Philippe Lemmerling, Bart de Moor, and Sabine Van Huffel, "On the equivalence of Constrained Total Least Squares and Structured Total Least Squares," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2908–2911, November 1996.
- [24] Theagenis J. Abatzoglou, Jerry M. Mendel, and Gail A. Harada, "The Constrained Total Least Squares technique and its applications to harmonic superresolution," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1070–1087, May 1991.
- [25] Vladimir Z. Mesarovic, Nikolas P. Galatsanos, and M. N. Wernick, "Iterative maximum a posteriori (map) restoration from partially-known blur for tomographic reconstruction," in *ICIP*, 1995, pp. 512–515.
- [26] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Transactions in Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 413–424, July 1986.
- [27] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, 1987.

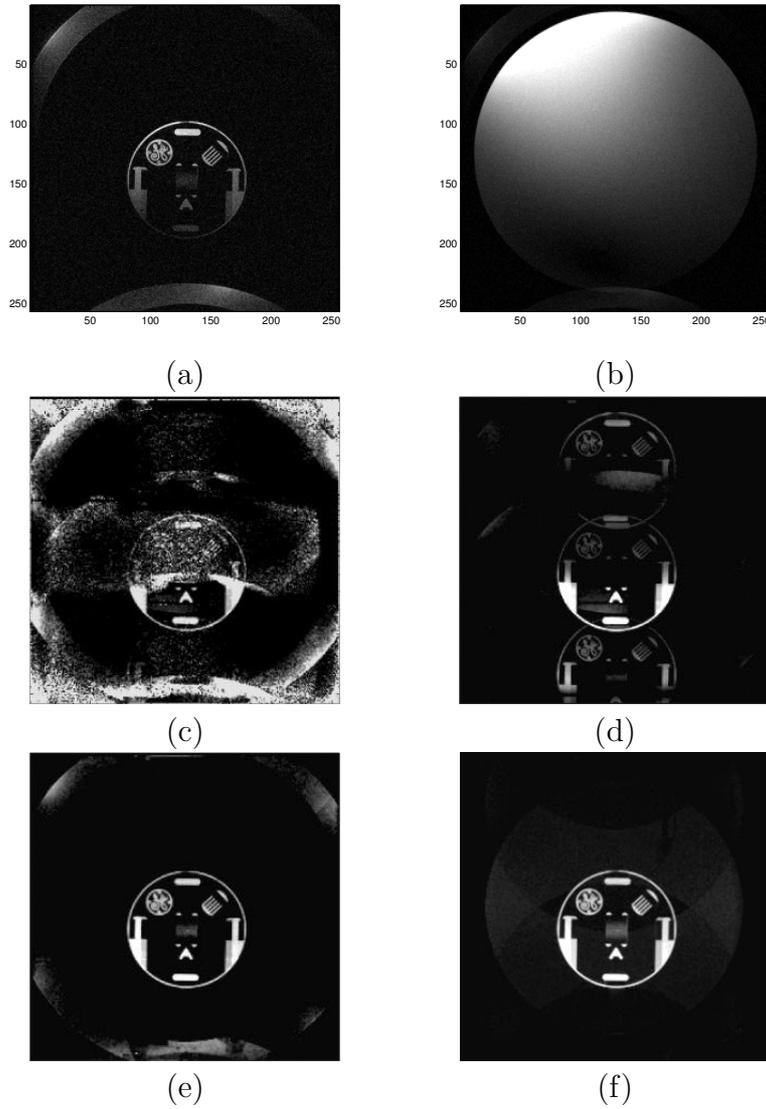


Fig. 8

RECONSTRUCTION RESULTS OF HiRES DATA, WITH $R = 3$, $L = 4$: (A) SHOWS (UNALIASED) OUTPUT OF THE HiRES PHANTOM WITHIN THE PLASTIC TUBE, AS SEEN BY ONE OF THE COILS, AND (B) SHOWS THE CORRESPONDING COIL SENSITIVITY MAP OBTAINED FROM A UNIFORM PHANTOM USING IDENTICAL GEOMETRICAL SETUPS FOR BOTH THE SENSE AND THE CALIBRATION SCAN. (C) STANDARD SENSE, (D) STANDARD SENSE WITH REGULARIZATION FACTOR OF 0.05, (E) ML-SENSE-I, AND (F) ML-SENSE-II. BOTH ML-SENSE ALGORITHMS EXHIBIT SUPERIOR PERFORMANCE COMPARED TO CONVENTIONAL METHODS.

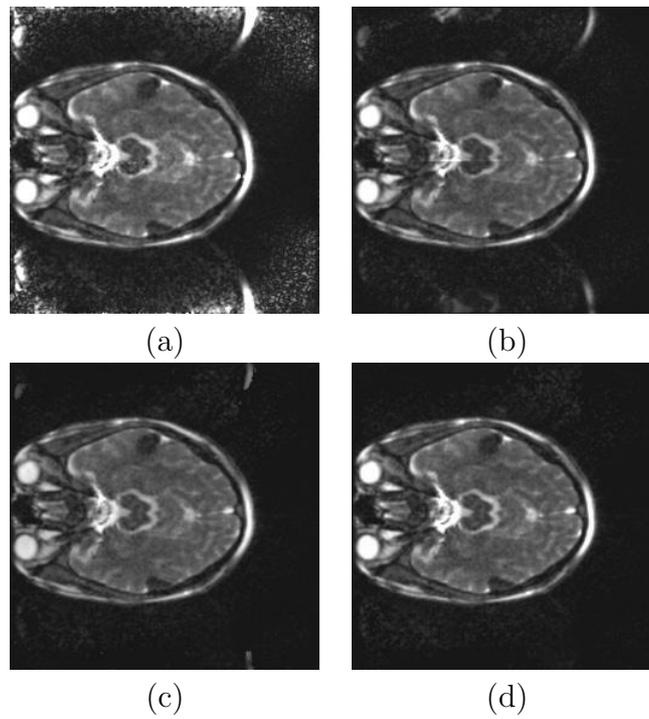


Fig. 9

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.

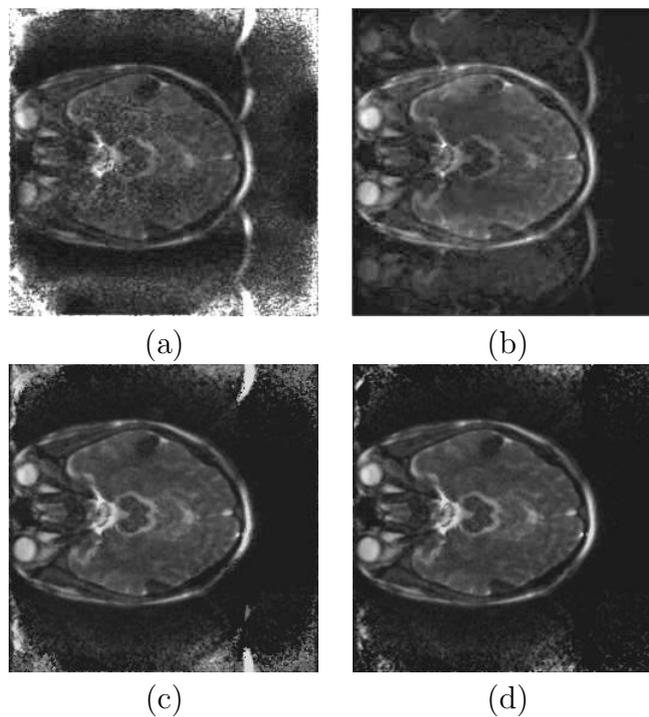


Fig. 10

BRAIN RECONSTRUCTION WITH $R = 4$, $L = 8$: (A) SENSE, (B) REGULARIZED SENSE, (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 HYPER-INTENSE FAT SIGNAL.

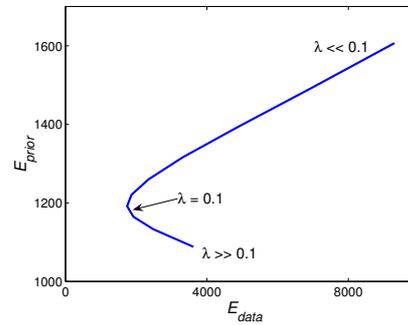


Fig. 11

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 $\lambda_{reg} = 0.1$. THIS WAS USED TO REGULARIZE SENSE IN FIGURE 10.

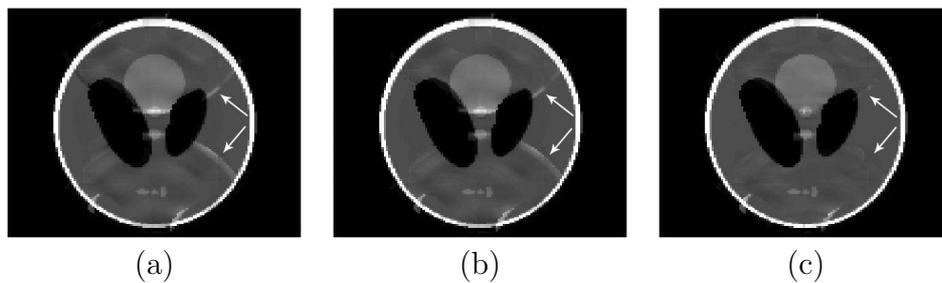


Fig. 12

RECONSTRUCTION OF SHEPP-LOGAN PHANTOM, WITH $R = 2$, $L = 4$ USING SENSITIVITY MAPS OBTAINED FROM CENTRAL (15×15) REGION OF K-SPACE AND DIVISION BY SUM OF SQUARES. SENSITIVITY WAS PERTURBED BY SMALL SHIFTS IN ORIENTATION AND POSITIONING. (A), (B), (C) ARE SENSE, ML-SENSE I AND ML-SENSE II RESPECTIVELY. NOTE THE ABSENCE OF THE ALIASING RING, POINTED OUT BY THE ARROWS, IN (C).

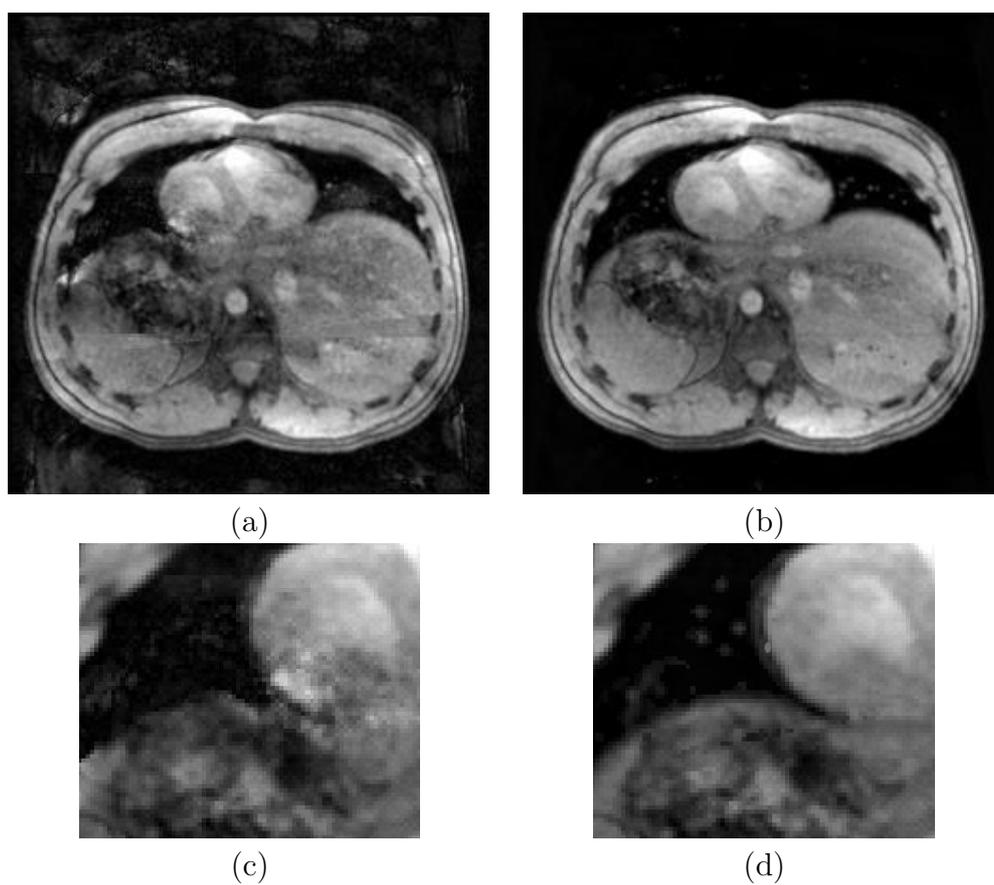


Fig. 13

RECONSTRUCTION RESULTS OF TORSO SCAN, $R = 3$, $L = 8$: (A) STANDARD SENSE WITH REGULARIZATION, (B) ML-SENSE II, (C) ZOOMING INTO (A), AND (D) ZOOMING INTO (B). NOTE THE DISTORTIONS AT THE HEART-STOMACH BOUNDARY AND STRIP ARTIFACT ACROSS THE LIVER IN SENSE OUTPUT.