

Monte Carlo SURE-Based Parameter Selection for Parallel Magnetic Resonance Imaging Reconstruction

Daniel S. Weller,^{1*} Sathish Ramani,¹ Jon-Fredrik Nielsen,² and Jeffrey A. Fessler^{1,2}

Purpose: Regularizing parallel magnetic resonance imaging (MRI) reconstruction significantly improves image quality but requires tuning parameter selection. We propose a Monte Carlo method for automatic parameter selection based on Stein's unbiased risk estimate that minimizes the multichannel k-space mean squared error (MSE). We automatically tune parameters for image reconstruction methods that preserve the undersampled acquired data, which cannot be accomplished using existing techniques.

Theory: We derive a weighted MSE criterion appropriate for data-preserving regularized parallel imaging reconstruction and the corresponding weighted Stein's unbiased risk estimate. We describe a Monte Carlo approximation of the weighted Stein's unbiased risk estimate that uses two evaluations of the reconstruction method per candidate parameter value.

Methods: We reconstruct images using the denoising sparse images from GRAPPA using the nullspace method (DESIGN) and L_1 iterative self-consistent parallel imaging (L_1 -SPIRiT). We validate Monte Carlo Stein's unbiased risk estimate against the weighted MSE. We select the regularization parameter using these methods for various noise levels and undersampling factors and compare the results to those using MSE-optimal parameters.

Results: Our method selects nearly MSE-optimal regularization parameters for both DESIGN and L_1 -SPIRiT over a range of noise levels and undersampling factors.

Conclusion: The proposed method automatically provides nearly MSE-optimal choices of regularization parameters for data-preserving nonlinear parallel MRI reconstruction methods.

Magn Reson Med 71:1760–1770, 2014. © 2013 Wiley Periodicals, Inc.

Key words: parallel imaging reconstruction; regularization parameter selection; Stein's unbiased risk estimate; Monte Carlo methods

INTRODUCTION

Although SENSE (1), GRAPPA (2), and SPIRiT (3) successfully reconstruct full field-of-view (FOV) images from undersampled parallel MRI data, many recent developments improve on these methods using regularization

(4–7). Practical clinical use of such methods requires selecting appropriate values of the regularization parameter(s). Often, a developer hard-codes a single value that “works” over a range of tested situations, or a user tunes the parameter manually for their particular application. When the scanning protocol remains unchanged across many subjects, such reuse may be justifiable. However, in cases where image contrast varies, some users may wish to ensure that their regularization parameters are reasonable.

The problem of regularization parameter selection has a long history, leading to a variety of techniques. In an extended scan session, one can adapt regularization parameters using data acquired earlier in the scan (8). However, such adaptations require acquiring and reconstructing a high-quality data set initially without regularization. Many applications cannot provide such a ground truth acquisition. For example, accelerated diffusion-weighted imaging can use shorter echo times, so an unaccelerated “ground truth” may actually have reduced image quality (9): a longer single-shot sequence would have greater geometric distortions and reduced contrast and signal from the long echo time. A multishot acquisition would take much longer, making motion and fatigue potential issues. Dynamic studies such as abdominal or cardiac perfusion imaging are corrupted by motion and other time-varying effects that are hard to replicate to produce a ground truth suitable for regularization. The same goes for dynamic contrast-enhanced studies. Functional magnetic resonance imaging (MRI) is another area where ground truths are often inaccessible due to limited temporal resolution and nonrepeatability of the underlying dynamic processes. Although one could ignore these dynamics when they are small in amplitude, as is the case with functional activation, doing so risks selecting parameters that would not adequately preserve these small deviations in the reconstruction.

Other techniques automatically tune regularization parameters based on a statistical model of the noise and just the acquired data (no training set). Among the most well-known, the discrepancy principle (10), generalized cross-validation (11,12), and L-curve curvature maximization (4,13,14) are founded in mathematical theory derived for quadratic regularizers with closed-form solutions. Their extension to nonlinear regularization is limited (15,16). All these methods rely on the measurement residual, the difference between the acquired (noisy) data, and the measurement model applied to the reconstructed image(s), to tune parameters. Thus, these methods are ineffective when the acquired data are preserved by the reconstruction method, as the resulting residual is exactly zero, regardless of the regularization parameter values. Although noisy, the acquired data contain essential details that could be compromised by regularization, so many image reconstruction

¹Department of Electrical Engineering and Computer Science University of Michigan, Ann Arbor, Michigan, USA.

²Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant number: F32 EB015914; Grant sponsor: NIH/NCI; Grant number: P01 CA87634.

*Correspondence to: Daniel S. Weller, Ph.D., 4125 EECS, 1301 Beal Avenue, Ann Arbor, MI 48109-2122. E-mail: wellerds@umich.edu.

Received 27 December 2012; revised 24 April 2013; accepted 17 May 2013.

DOI 10.1002/mrm.24840

Published online 2 July 2013 in Wiley Online Library (wileyonlinelibrary.com).

© 2013 Wiley Periodicals, Inc.

techniques aim to preserve the acquired data, including regularized parallel imaging techniques like denoising sparse images from GRAPPA using the nullspace method (DESIGN) (6) and L_1 iterative self-consistent parallel imaging (L_1 -SPIRiT) (3,7,17,18). In this article, we propose a new error criterion that is appropriate for data-preserving regularized reconstruction methods, for which we derive a novel parameter selection scheme.

We seek the parameter that minimizes the k-space mean squared error (MSE) for reconstruction methods based on nonlinear regularizers such as the ℓ_1 norm commonly used to promote sparsity (19). Although k-space MSE may not be sensitive to local image errors in small features of diagnostic value, it is a convenient quality measure used frequently in the literature. MSE is the basis for many image reconstruction methods, including SENSE, which yields the MSE-optimal combined image consistent with the multichannel data, when given correct coil sensitivities (1). Because calculating the actual MSE would require knowing the true image, we turn to Stein's unbiased risk estimate (SURE; 20), which approximates the MSE for a gaussian noise model and accommodates nonlinear reconstruction algorithms. Using the noise model and partial derivatives of the reconstruction function, SURE yields an expression that is equal in expected value to the MSE and can be minimized to select a suitable regularization parameter value without knowing the true signal.

SURE has been exploited for parameter selection in denoising (21–23), deblurring (24–26), single-channel MRI (16,27), and image-domain noniterative SENSE reconstruction (28) for uniform (nonrandom) undersampled parallel MRI. However, in its basic formulation, evaluating SURE involves computing the Jacobian matrix of the targeted reconstruction function. Although feasible, analytically evaluating SURE is difficult for iterative methods and requires exactly differentiating nearly every computation performed in a particular implementation. A Monte Carlo SURE framework (23,27) was derived to avoid this complication, also motivating this work. The original derivation of the Monte Carlo framework, and of SURE in general, is restricted to the denoising setting, where the entire image is observed. Later work extends SURE to inverse problems such as image reconstruction, using modified error metrics such as predicted-MSE or projected-MSE (16,25–27). However, such frameworks share the limitation of discrepancy principle, cross validation, and L-curve-based methods, that the criteria are based on the measurement residual, which is insensitive to the parameter choice for data-preserving reconstructions. That limitation also extends to the Monte Carlo nonlinear generalized cross validation method described previously (16).

We extend the Monte Carlo SURE framework to regularized autocalibrating parallel MRI reconstruction methods that preserve the data. The novelty of our method consists of applying parallel imaging to express the unacquired (or full) k-space error criterion in terms of just the k-space that we measure. This approach avoids the complications that arise from data discrepancy but limits our method to parallel imaging applications. We demonstrate the generality and utility of this method using two regularized iterative algorithms on real data: DESIGN (6) and L_1 -SPIRiT (3,7). The first improves on

GRAPPA with uniform undersampling, whereas the second applies compressed sensing (19) to random Cartesian undersampling. In preliminary studies, we described the extension to parallel imaging with uniform Cartesian undersampling and depicted simulated data using SURE-optimized DESIGN (29), and we applied SURE to L_1 -SPIRiT for a range of synthetic noise levels (30). Here, we validate the Monte Carlo estimate of the weighted SURE criterion against the true weighted MSE for both parallel-imaging methods and apply the proposed method to DESIGN for different noise levels and to L_1 -SPIRiT for various undersampling factors. We conclude with a discussion of the implications of a general algorithm for automatic parameter selection and useful extensions of this work.

THEORY

Parallel imaging reconstruction methods like SENSE (1) use coil sensitivity profiles to reconstruct a single full-FOV combined image from multiple coil channels of undersampled k-space data. However, coil sensitivities may be difficult to measure accurately, motivating autocalibrating methods such as GRAPPA (2) or SPIRiT (3) that do not require explicit knowledge of those profiles. These methods typically operate in the k-space domain and fill the Cartesian k-space grid for every coil, so the MSE of the full-multichannel k-space is the cost function we aim to minimize. Methods like sum-of-squares (31) can be used to combine these full-FOV data sets afterwards.

Measurement Model and Error Criteria

The target Cartesian k-space grid for an N -voxel volume can be represented as a length- N vector; for P coils, these vectors stack to form the length- NP vector \mathbf{x} . We undersample k-space, acquiring $M < N$ points arranged in some fashion, and represented as a length- M vector. The sampling pattern for each receive coil is identical; we combine these vectors into a length- MP vector \mathbf{y} . The full k-space relates to the acquired data via the linear equation $\mathbf{y} = \mathbf{M}\mathbf{x}$. When our sampling pattern is Cartesian, \mathbf{M} is a simple $MP \times NP$ mask, where each row has a single entry equal to one and all the other elements equal to zero. To complete our model, we consider our measurements to be contaminated with additive complex Gaussian noise $\boldsymbol{\eta}$, independent across k-space locations, and correlated across channels, with zero mean and covariance $\boldsymbol{\Lambda}$:

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \boldsymbol{\eta}. \quad [1]$$

To reconstruct the full k-space \mathbf{x} from our noisy samples \mathbf{y} , we focus on improving autocalibrating linear methods such as GRAPPA and SPIRiT. The regularized version of these methods is in general nonlinear and can be written as $\hat{\mathbf{x}} = \mathbf{f}_\gamma(\mathbf{y})$, where user-specified parameter(s) $\boldsymbol{\gamma}$ control the level of regularization.

To guide selection of the regularization parameters $\boldsymbol{\gamma}$, one might aim to use a criterion based on the MSE of the full multichannel k-space:

$$\text{MSE}(\boldsymbol{\gamma}) \triangleq \|\mathbf{x} - \mathbf{f}_\gamma(\mathbf{y})\|_2^2, \quad [2]$$

which is a function of the unknown true values of the full k-space \mathbf{x} . When applied to Cartesian undersampled data,

we are interested in preserving the acquired data in the reconstruction, so $\mathbf{M}\mathbf{f}_\gamma(\mathbf{y}) = \mathbf{y}$, regardless of the value of \mathbf{y} . Thus, we attempt to minimize the MSE weighted over the set of unacquired k-space points:

$$\text{WMSE}(\mathbf{y}) \triangleq \|\tilde{\mathbf{M}}(\mathbf{x} - \mathbf{f}_\gamma(\mathbf{y}))\|_2^2, \quad [3]$$

where $\tilde{\mathbf{M}}$ is the mask complementary to \mathbf{M} for Cartesian undersampled data. One can use the weighted MSE (WMSE) criterion directly only if one postulates a Bayesian prior model for \mathbf{x} . Because such priors are rarely realistic for medical images, we take a non-Bayesian approach where \mathbf{x} is deterministic but unknown.

Applying SURE to WMSE Using Parallel Imaging

SURE is an unbiased estimate of the MSE of some unknown vector \mathbf{x} from direct measurements $\mathbf{y} = \mathbf{x} + \boldsymbol{\eta}$ corrupted by Gaussian noise $\boldsymbol{\eta}$. Thus, approximating the WMSE (or MSE) of the unacquired (or full) multichannel k-space would require directly measuring the entire k-space \mathbf{x} to apply SURE. Because we undersample k-space, we essentially need to invert the measurement matrix \mathbf{M} in Eq. 1 to minimize the WMSE in Eq. 3. Because the measurement matrix undersamples k-space, \mathbf{M} is singular, and such an inversion naturally would be ill-posed. This key problem is avoided in previous work by modifying the error criterion to be a function of the measurement error only, like projected-MSE and predicted-MSE (16,25–27), which apply the measurement mask \mathbf{M} to the reconstructed image(s). However, such efforts do not work for reconstructions that retain the measurements. Our novel contribution uses parallel imaging to perform the necessary inversion. We introduce the linear operator \mathbf{G} that estimates the full-multichannel k-space from the true k-space values at the sampled locations. For instance, unregularized parallel imaging methods like GRAPPA and SPIRiT provide approximately true values of the full-multichannel k-space when no noise is present: $\mathbf{x} \approx \mathbf{G}\mathbf{M}\mathbf{x}$. Thus, we write

$$\text{WMSE}(\mathbf{y}) \approx \|\tilde{\mathbf{M}}(\mathbf{G}\mathbf{M}\mathbf{x} - \mathbf{f}_\gamma(\mathbf{y}))\|_2^2. \quad [4]$$

Given a properly calibrated \mathbf{G} and an appropriate level of undersampling for the number of channels and geometry of the array coil, minimizing this criterion should nearly optimize the true WMSE, and minimizing the SURE for this error should yield $\boldsymbol{\gamma}$ that is also nearly WMSE-optimal. In the examples using DESIGN and L_1 -SPIRiT as the regularized reconstruction $\mathbf{f}_\gamma(\mathbf{y})$, we choose the linear mapping \mathbf{G} to be the unregularized version of the reconstruction $\mathbf{f}_\gamma(\mathbf{y})$, namely GRAPPA and SPIRiT, respectively. However, the theory does not require any such connection between \mathbf{G} and $\mathbf{f}_\gamma(\mathbf{y})$. We emphasize that we do not need to consider noise amplification in choosing \mathbf{G} , since \mathbf{G} is applied to noise-free $\mathbf{M}\mathbf{x}$ in our derivation, not to our actual observations \mathbf{y} . However, the number of available coils and the desired acceleration factor can influence the accuracy of this approximation, so we investigate the effects of violating this assumption in our experiments. This use of \mathbf{G} for regularization parameter selection is unique to this work, and the last subsection of the Methods section investigates how this approximation affects parameter selection.

Monte Carlo SURE Approximation

Expanding the approximate WMSE in Eq. 4, and using Eq. 1,

$$\begin{aligned} \text{WMSE}(\mathbf{y}) \approx & \|\tilde{\mathbf{M}}\mathbf{G}\mathbf{M}\mathbf{x}\|_2^2 + \|\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\|_2^2 \\ & - 2\mathcal{R}e\{\mathbf{y}'\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\} + 2\mathcal{R}e\{\boldsymbol{\eta}'\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\}, \end{aligned} \quad [5]$$

where $[\cdot]'$ denotes the conjugate transpose operator. The WMSE isolates \mathbf{x} to the first term of Eq. 5, which does not depend on $\boldsymbol{\gamma}$, so it can be considered a constant C and ignored when optimizing $\boldsymbol{\gamma}$. The next two terms of Eq. 5 can be computed directly using $\mathbf{f}_\gamma(\mathbf{y})$ and $\mathbf{G}\mathbf{y}$, but the last term is a function of both $\boldsymbol{\gamma}$ and the unknown noise vector $\boldsymbol{\eta}$. Stein's lemma (16,20,25) approximates this last term, so that we can find $\boldsymbol{\gamma}$ that minimizes the WMSE. Without our parallel-imaging-based approximation of the full k-space, we could not express the unacquired multichannel k-space in terms of the acquired data and the noise, and there would be no noise term in Eq. 5 to approximate using Stein's lemma.

Now, we modify previous derivations of Monte Carlo SURE (23,27) to include the matrix \mathbf{G} . When $\mathbf{f}_\gamma(\mathbf{y})$ is (sub-) differentiable, let $\mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})$ be the Jacobian matrix of all first order partial (sub-) derivatives of $\mathbf{f}_\gamma(\mathbf{y})$ with respect to \mathbf{y} . Using the chain rule for derivatives, the Jacobian of the expression $\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})$ is $\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})$. This expression in combination with Stein's lemma implies (27) that

$$\mathcal{E}\{\boldsymbol{\eta}'\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\} = \mathcal{E}\{\text{tr}\{\boldsymbol{\Lambda}\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})\}\}. \quad [6]$$

Applying Eq. 6 to the approximate WMSE in Eq. 5 yields the weighted SURE (WSURE) function, which is an unbiased estimate of the WMSE:

$$\begin{aligned} \text{WSURE}(\mathbf{y}) = & \|\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\|_2^2 - 2\mathcal{R}e\{\mathbf{y}'\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{f}_\gamma(\mathbf{y})\} \\ & + 2\mathcal{R}e\{\text{tr}\{\boldsymbol{\Lambda}\mathbf{G}'\tilde{\mathbf{M}}\tilde{\mathbf{M}}\mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})\}\} + C. \end{aligned} \quad [7]$$

However, this expression involves computing all the partial derivatives of the reconstruction function. For iterative reconstructions, the Jacobian matrix could be computed step-by-step, but such an approach requires a careful derivation customized for that particular reconstruction method (16,24,26). For an iterative reconstruction like DESIGN and L_1 -SPIRiT, expressions for the Jacobian matrices must be derived for each step in the implementation. Instead, we use Monte Carlo SURE (23,27), which does not require such in-depth knowledge of the algorithm under consideration. Under mild conditions on $\mathbf{f}_\gamma(\mathbf{y})$ (23,27,29), the directional derivative $\mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})\mathbf{b}$ can be approximated by

$$\boldsymbol{\rho}(\mathbf{y}, \mathbf{b}, \varepsilon) = (\mathbf{f}_\gamma(\mathbf{y} + (\varepsilon\|\mathbf{y}\|_2/\sqrt{MP})\mathbf{b}) - \mathbf{f}_\gamma(\mathbf{y})) / (\varepsilon\|\mathbf{y}\|_2/\sqrt{MP}), \quad [8]$$

for sufficiently small $\varepsilon > 0$. The scale factor $\|\mathbf{y}\|_2/\sqrt{MP}$ adjusts ε for the relative scale of the data; this is motivated by the fact that some algorithms (like DESIGN) use an un-normalized Fourier transform to relate k-space to image space, whereas others (like L_1 -SPIRiT) use a normalized transform. Then, choosing \mathbf{b} to be a vector of independent, identically distributed complex zero-mean, unit-variance Bernoulli-distributed random variables (i.e.,

the real and imaginary parts of each element of \mathbf{b} are drawn independently from $\{\pm 1/\sqrt{2}\}$ with equal probability, we form a Monte Carlo approximation to the trace of the Jacobian (23,27) with just one realization of \mathbf{b} :

$$\text{tr}\{\Lambda \mathbf{G}' \tilde{\mathbf{M}}' \tilde{\mathbf{M}} \mathbf{J}_{\mathbf{f}_\gamma}(\mathbf{y})\} \approx \mathbf{b}' \Lambda \mathbf{G}' \tilde{\mathbf{M}}' \tilde{\mathbf{M}} \boldsymbol{\rho}(\mathbf{y}, \mathbf{b}, \varepsilon), \quad [9]$$

Plugging this approximation into Eq. 7 yields the Monte Carlo estimate of WSURE:

$$\begin{aligned} \text{WSURE}(\boldsymbol{\gamma}) \approx & \|\tilde{\mathbf{M}} \mathbf{f}_\gamma(\mathbf{y})\|_2^2 - 2\text{Re}\{\mathbf{y}' \mathbf{G}' \tilde{\mathbf{M}}' \tilde{\mathbf{M}} \mathbf{f}_\gamma(\mathbf{y})\} \\ & + 2\text{Re}\{\mathbf{b}' \Lambda \mathbf{G}' \tilde{\mathbf{M}}' \tilde{\mathbf{M}} \boldsymbol{\rho}(\mathbf{y}, \mathbf{b}, \varepsilon)\} + C. \end{aligned} \quad [10]$$

The same \mathbf{b} can be used for evaluating WSURE with different choices of regularization parameter, so we can compute and store $\tilde{\mathbf{M}} \mathbf{G} \Lambda \mathbf{b}$ just once. When we apply WSURE to regularized GRAPPA or SPIRiT reconstructions, we can find the parameters $\boldsymbol{\gamma}$ that minimize WSURE using a derivative-free minimization algorithm. In this work, we deal with scalar γ , and we use a coarse-to-fine parameter sweep, where we evaluate WSURE for successively narrower intervals of γ .

Next, we summarize two regularized reconstruction methods that illustrate the broad applicability of Monte Carlo WSURE to regularization parameter selection: DESIGN (6) and L_1 -SPIRiT (3,7). GRAPPA is a popular and time-tested autocalibrating parallel imaging reconstruction method for uniformly spaced undersampled k-space data. The regularized method DESIGN is attractive from a computational standpoint because it denoises the GRAPPA reconstruction output, instead of iteratively reapplying GRAPPA to denoised data, which would be slower. L_1 -SPIRiT also is a useful example of an autocalibrating regularized reconstruction method because it can readily process nonuniformly or randomly undersampled k-space data.

DESIGN: Sparsity-Promoting GRAPPA

For uniformly undersampled Cartesian k-space, the GRAPPA reconstruction directly fills missing k-space in all the coil array channels with a linear combination of neighboring acquired points weighted by a shift-invariant kernel. GRAPPA reconstruction is implemented efficiently using convolution or the fast Fourier transform (FFT). For $R_x \times R_y$ two-dimensional (2D) undersampled k-space, the GRAPPA method using a set of $B_x \times B_y$ 2D kernels $\mathbf{g}_{p,q,r_x,r_y}[b_x, b_y]$ for each input/output channel pair (p, q) and target frequency shift (r_x, r_y) evaluates

$$\begin{aligned} \hat{x}[k_x + r_x, k_y + r_y, q] = & \sum_{p=1}^P \sum_{b_x=-B_x/2}^{B_x/2} \sum_{b_y=-B_y/2}^{B_y/2} \mathbf{g}_{p,q,r_x,r_y}[b_x, b_y] \\ & \cdot \mathbf{y}[k_x + b_x R_x, k_y + b_y R_y, p]. \end{aligned} \quad [11]$$

These GRAPPA kernels are calibrated the usual way, using a least-squares fit with densely sampled k-space training data, potentially with regularization. Then, we regularize the GRAPPA reconstruction using a null space formulation of the data-preserving DESIGN method (6):

$$\begin{aligned} \mathbf{f}_\gamma(\mathbf{y}) = & \tilde{\mathbf{M}} \tilde{\mathbf{x}}^* + \mathbf{M}' \mathbf{y}; \\ \tilde{\mathbf{x}}^* = & \arg \min_{\tilde{\mathbf{x}}} \frac{1}{2} \|(\Lambda_G)^{-\frac{1}{2}} (\tilde{\mathbf{x}} - \tilde{\mathbf{M}} \mathbf{G} \mathbf{y})\|_2^2 \\ & + \gamma \|\boldsymbol{\Psi} \mathcal{F}^{-1} (\tilde{\mathbf{M}} \tilde{\mathbf{x}} + \mathbf{M}' \mathbf{y})\|_{1,2}, \end{aligned} \quad [12]$$

where $\tilde{\mathbf{x}}/\tilde{\mathbf{x}}^*$ is the length- $(N-M)P$ vector of missing k-space for all the coil channels, Λ_G is the noise covariance of the GRAPPA-reconstructed missing data $\Lambda_G = \tilde{\mathbf{M}} \mathbf{G} \Lambda \mathbf{G}' \tilde{\mathbf{M}}'$, γ is the regularization parameter of interest, $\boldsymbol{\Psi}$ is the linear sparsifying transform, \mathcal{F}^{-1} is the inverse FFT, and $\|\cdot\|_{1,2}$ is the joint sparsity-promoting hybrid ℓ_1/ℓ_2 norm. To evaluate this mixed norm, for each transform coefficient, we form a vector of that coefficient's values across all the coil channels, and we compute the sum of the ℓ_2 norms of these vectors. The matrix \mathbf{G} corresponds to convolving the GRAPPA kernels with the undersampled k-space. Because GRAPPA introduces correlations among many frequencies into the noise covariance Λ_G , and such a covariance matrix is expensive to store or invert, we simplify the matrix by ignoring the correlations across frequencies, considering only the noise amplification and correlation across coils (32). The resulting optimization problem can be implemented using any number of algorithms; we use a Split-Bregman approach (33) with auxiliary variable $\mathbf{w} = \boldsymbol{\Psi} \mathcal{F}^{-1} (\tilde{\mathbf{M}} \tilde{\mathbf{x}} + \mathbf{M}' \mathbf{y})$.

Sparsity-Regularized SPIRiT

Although GRAPPA is effective for uniformly-spaced Cartesian undersampling, a more flexible method is desirable for nonuniformly undersampled k-space. The SPIRiT reconstruction method is an iterative formulation that uses kernel-weighted linear consistency equations similar to the GRAPPA convolution equation in Ref. 11 to estimate missing k-space data. However, SPIRiT consistency equations operate over the entire k-space, including unacquired frequencies. For a 2D $B_x \times B_y$ kernel $\mathbf{s}_{p,q}[b_x, b_y]$ with $s_{p,q}[0, 0] = 0$, the consistency equation for the point $\mathbf{x}[k_x, k_y, q]$ is

$$\begin{aligned} x[k_x, k_y, q] = & \sum_{p=1}^P \sum_{b_x=-B_x/2}^{B_x/2} \sum_{b_y=-B_y/2}^{B_y/2} \\ & s_{p,q}[b_x, b_y] x[k_x + b_x, k_y + b_y, p]. \end{aligned} \quad [13]$$

The data-preserving version of SPIRiT forms a linear system $\mathbf{x} = \mathbf{S} \mathbf{x}$ of these consistency equations over all of multichannel k-space and minimizes the least-squares fit constrained by the data consistency equation $\mathbf{y} = \mathbf{M} \mathbf{x}$:

$$\mathbf{G} \mathbf{y} = \arg \min_{\mathbf{x}} \|(\mathbf{S} - \mathbf{I}) \mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{y} = \mathbf{M} \mathbf{x}. \quad [14]$$

This constrained optimization problem can be regularized using an additional sparsity term on the wavelet-transformed images obtained by taking the inverse FFT of the full-multichannel k-space, yielding the regularized problem

$$\begin{aligned} \mathbf{f}_\gamma(\mathbf{y}) = & \arg \min_{\mathbf{x}} \frac{1}{2} \|(\mathbf{S} - \mathbf{I}) \mathbf{x}\|_2^2 + \gamma \|\boldsymbol{\Psi} \mathcal{F}^{-1} \mathbf{x}\|_{1,2} \\ & \text{subject to} \quad \mathbf{y} = \mathbf{M} \mathbf{x}. \end{aligned} \quad [15]$$

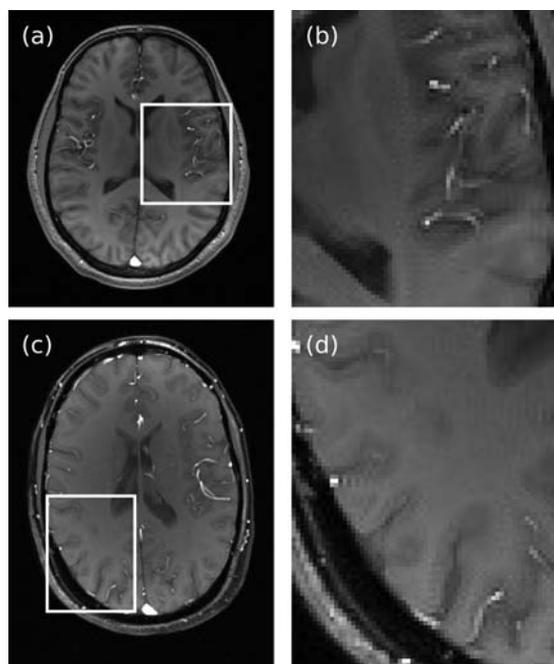


FIG. 1. The combined normalized reference images from data sets #1 (a) and #2 (c) are shown alongside inset regions enlarged to show detail (b, d).

The projection-onto-convex-sets-based implementation called L_1 -SPIRiT (7) uses a sparsity-promoting shrinkage operation on the sparse transform of \mathbf{x} , controlled by the predetermined scalar threshold γ .

METHODS

We demonstrate our parameter selection method on real data from consented subjects in accordance with institutional review board-approved protocols using a GE Discovery 3T MRI scanner with a vendor-supplied eight-channel array coil. We performed six repetitions of a T1-weighted 3D-spoiled gradient echo sequence [pulse repetition time (TR) = 10.1 ms, echo time (TE) = 4.3 ms, flip angle (FA) = 16 degrees, bandwidth (BW) = 31.25 kHz] to sample a $256 \times 256 \times 10$ Cartesian multichannel reference volume with $24 \times 24 \times 3$ -cm FOV, $0.94 \times 0.94 \times 3.0$ -mm resolution, and no acceleration. We separately saved four repetitions of a T1-weighted 2D-spoiled gradient echo sequence (TR = 20.0 ms, TE = 9.1 ms, FA = 10 degrees, BW = 15.63 kHz) to sample a 256×256 multichannel axial slice with 24×24 -cm FOV, $0.94 \times 0.94 \times 4.0$ -mm resolution, with no acceleration, similar to T1-weighted slice-by-slice acquisitions used for overlaying anatomical information. We used 3D and 2D acquisitions to add generality to the data sets for validation purposes. The averaged data are used as low-noise reference images for the experiments that follow.

We processed the acquired raw multichannel data using MATLAB. Both T1-weighted images were cropped tightly around the subject's head before undersampling and reconstructing the data. Then, we undersampled an axial slice of the reference volume using the desired sampling scheme (uniform or nonuniform Cartesian), and we reconstructed

the data using the autocalibrating method of choice (GRAPPA-based DESIGN for uniform Cartesian sampling or L_1 -SPIRiT for nonuniform). Because we are interested in reconstructing images without sensitivity maps, we formed combined images using a sum-of-squares algorithm (31). Reconstructed images are evaluated against the ground truth (when available) using difference images and WMSE. Figure 1 shows the coil-combined reference images for both T1-weighted data sets after cropping. The signal-to-noise ratio (SNR) of the acquired data (in dB) is calculated using

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\mathbf{x}\|^2}{(\# \text{ of points in } \mathbf{x}) \|\mathbf{\Lambda}\|_*^2} \right), \quad [16]$$

where $\|\cdot\|_*$ is the nuclear norm. The nuclear norm is used because each coil has a potentially different noise level, which correspond to the singular values of $\mathbf{\Lambda}$ after whitening the noise to remove correlations.

Validation of WSURE as an Estimate

Using our averaged T1-weighted acquired slices as low-noise ground truths, we added complex gaussian noise so the total noise covariance matches the covariance of a single acquisition before averaging in each instance. The synthetic noise reduces the fully sampled SNR from 30 to

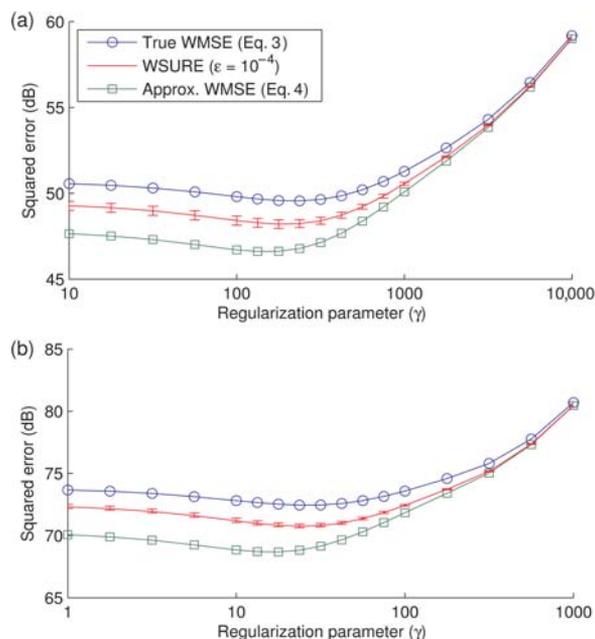


FIG. 2. WSURE for GRAPPA-based DESIGN, shown for data sets #1 (a) and #2 (b) with $\epsilon = 10^{-4}$, are similar in shape to the true WMSE curves and the GRAPPA-approximated WMSE curves and are minimized by nearly the same value of γ as the true WMSE curves in each case. The error bars on WSURE curves correspond to ± 5 standard deviations from the mean WSURE value. Only the error bars were estimated from 20 realizations of the complex Bernoulli noise vector \mathbf{b} ; the WSURE curves correspond to using just the first realization of \mathbf{b} . The curves are shown for 2×2 uniform undersampling (with a 24×24 fully sampled central autocalibration signal block), with 22.3-dB input SNR for the first data set and 19.7-dB input SNR for the second. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

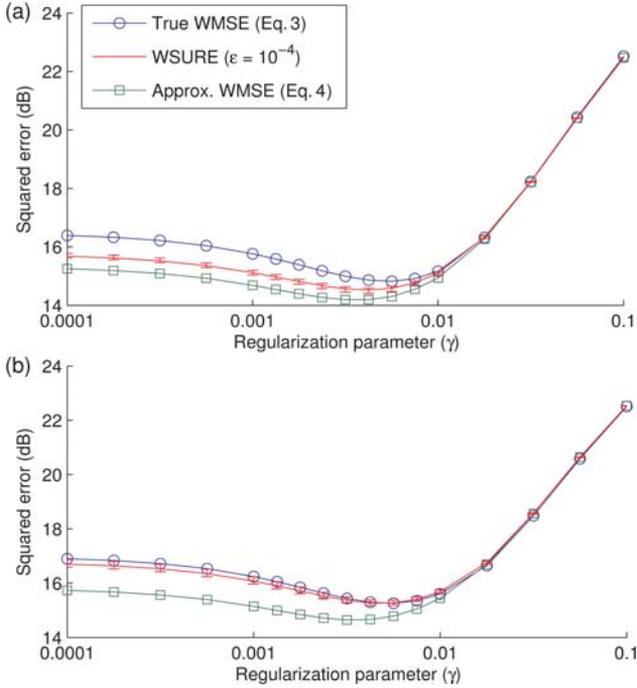


FIG. 3. WSURE for L_1 -SPIRiT, shown for data sets #1 (a) and #2 (b) with $\epsilon = 10^{-4}$, closely tracks both the true WMSE and the SPIRiT-approximated WMSE curves. WSURE curve error bars correspond to ± 5 standard deviations from the mean WSURE value. Only the error bars were estimated from 20 realizations of the complex Bernoulli noise vector \mathbf{b} ; the WSURE curves correspond to using just the first realization of \mathbf{b} . The curves are shown for $R \approx 5$ Poisson-disc undersampling (with a 24×24 fully sampled autocalibration signal region), with 22.3-dB input SNR for the first data set and 19.7-dB input SNR for the second.

22 dB for the first data set, and 26 to 20 dB for the second. We applied the GRAPPA-based DESIGN method to uniformly 2×2 undersampled k-space and the L_1 -SPIRiT method to 2D Poisson-disc nonuniformly undersampled Cartesian k-space (undersampling factor $R \approx 5$), leaving a 24×24 non-undersampled block of central k-space for calibration data in both instances to train the 4×4 -block GRAPPA and 5×5 SPIRiT kernels. In this and the experiments that follow, DESIGN was implemented using 20 Split-Bregman iterations and an isotropic (discrete) total variation sparsifying transform, and L_1 -SPIRiT was implemented using 25 projection-onto-convex-sets iterations with the four-level “db4” orthonormal DWT sparsifying transform (the code is publicly available at <http://www.eecs.berkeley.edu/~mlustig/Software.html>).

We computed WSURE values for both DESIGN and L_1 -SPIRiT using the expressions in Eqs. 7 and 10 over ranges of γ known to contain the WMSE-optimal values ($\gamma \in [10, 10^4]$ for DESIGN and $\gamma \in [10^{-4}, 0.1]$ for L_1 -SPIRiT) and several choices of ϵ ($\epsilon \in \{10^{-5}, 10^{-4}, 0.001, 0.01\}$ for both methods). The difference in ranges of γ can be ascribed to the relatively different scales of the k-space data—the implementation of L_1 -SPIRiT uses the unitary FFT operator, whereas DESIGN uses the un-normalized FFT. These WSURE estimates were generated using the same realization of the complex random vector \mathbf{b} , and we computed

$\tilde{\mathbf{M}}\mathbf{G}\mathbf{A}\mathbf{b}$ just once. Also, when evaluating WSURE for the same γ and different values of ϵ , we reused $\mathbf{f}_\gamma(\mathbf{y})$. To validate WSURE as an appropriate proxy for the MSE criterion, we compared these values to the true WMSE in Eq. 3 and approximate WMSE in Eq. 4. To gauge the precision of our estimates, we repeated our WSURE calculations for a total of 20 realizations of \mathbf{b} to compute the sample standard deviations of these Monte Carlo estimates. To generalize our validation, we repeated these experiments for both data sets, using the same values of ϵ in each.

Parameter Selection for DESIGN

Using $\epsilon = 10^{-4}$ for both T1-weighted data sets, we computed the WSURE-optimized DESIGN reconstructions for uniformly 2×2 undersampled Cartesian k-space data with various noise levels. For the first data set, we added simulated complex Gaussian noise to the averaged multi-channel reference k-space before undersampling to produce a range of 1.2 dB–13 dB SNR. We added noise to the second data set to yield a range of 4.6 dB–17 dB SNR as well. We found the WSURE-optimized reconstruction using a coarse-then-fine two-level parameter sweep of γ , with the coarse level considering γ between 0.01 and 10^6 for the first data set and between 10^{-3} and 10^5 for the second. We

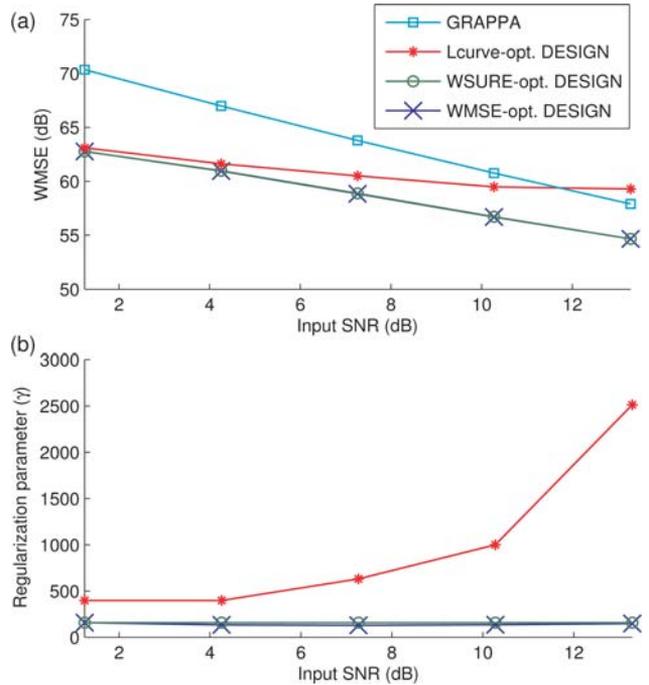


FIG. 4. Varying the noise level of 2×2 uniform undersampled data, the sparsity-promoting DESIGN method yields similarly lower WMSE (a) for both WMSE-optimal and WSURE-optimized choices of γ , relative to the un-regularized GRAPPA reconstruction, for the first data set. The WMSE of the DESIGN reconstruction using the WSURE-optimized choice is within 0.038 dB of the true WMSE-optimal DESIGN reconstruction. This behavior is consistent with the WMSE-optimal and WSURE-optimized choices for the DESIGN regularization parameter γ (b) being nearly the same for this example. The L-curve-style method, however, appears to overestimate γ , yielding images with slightly higher WMSE, and in the high-SNR case, worse than performing unregularized GRAPPA.

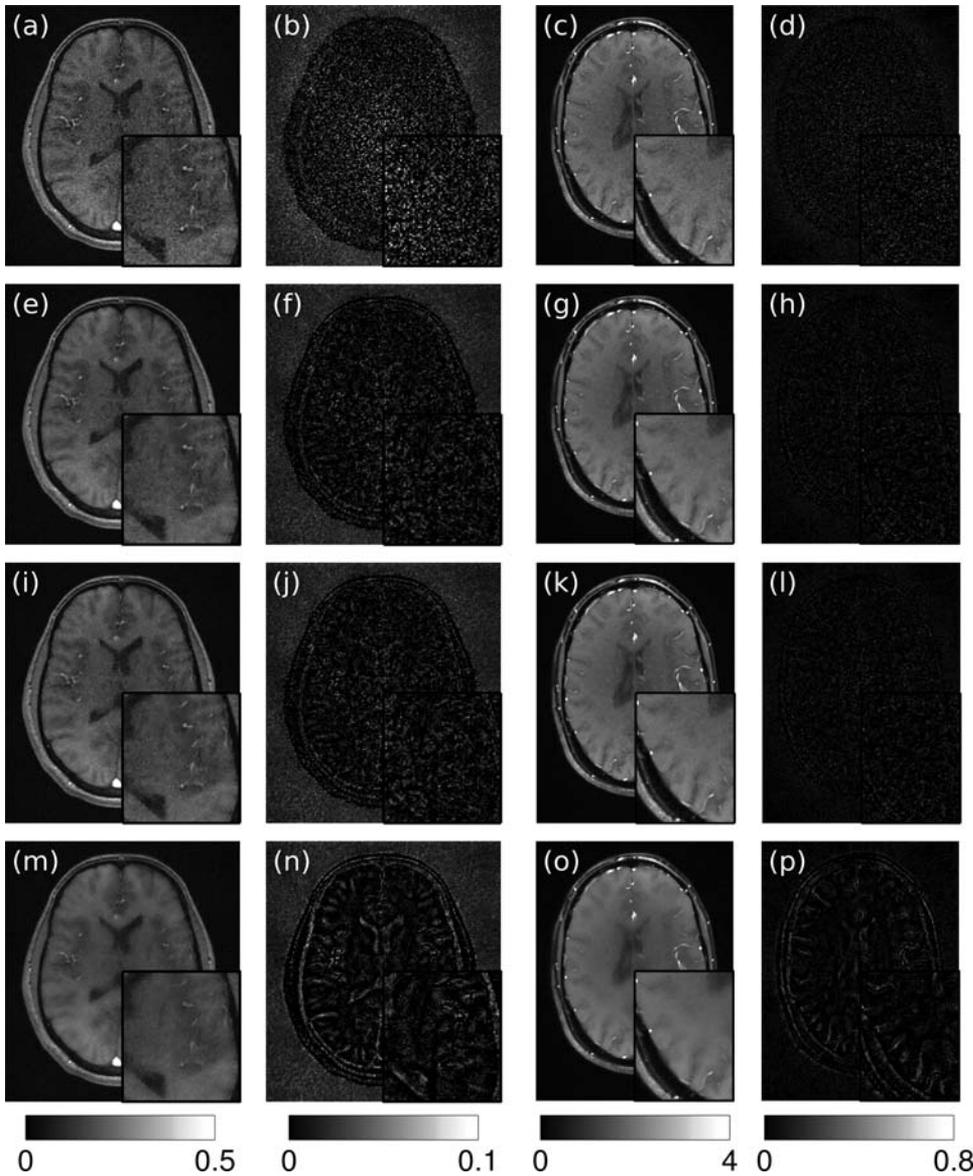


FIG. 5. The GRAPPA-reconstructed image (a) and difference image (b) for data set #1 with 10.3 dB input SNR display significant noise, while DESIGN with both WMSE-optimal (e,f) and WSURE-optimized (i,j) choices of γ effectively denoise the image while losing only a small amount of detail. The L-curve-based method (m,n) oversmooths the image, resulting in significant blurring artifacts. Similar results are observed for the second data set with 16.7-dB input SNR, where remaining noise in the GRAPPA reconstruction (c,d) is mitigated by DESIGN using the WMSE-optimal (g,h), WSURE-optimized (k,l), and L-curve-selected (o,p) choices of γ . Again, oversmoothing in the L-curve-based reconstruction is visible in the second data set. The difference images are displayed at $5\times$ scale.

compared the WSURE-optimized reconstruction against the WMSE-optimal reconstruction found using MATLAB's `fminsearch()` function from a starting point of $\gamma = 1000$ for the first data set and $\gamma = 10$ for the second. This WMSE-optimal reconstruction takes advantage of the fully sampled ground truth to avoid the $\mathbf{x} \approx \mathbf{GMx}$ approximation. We also included a comparison against an L-curve-like curvature maximization of the log-log pareto optimality curve for DESIGN, trading off $\ell_{1,2}$ norm joint sparsity for least-squares consistency with the GRAPPA reconstruction [12]. To maximize the curvature, we used a two-level parameter sweep over the same range as for WSURE. We emphasize that this method differs from L-curve methods in that the L-curve balances the data prediction residual norm with the regularization term. Because DESIGN preserves the acquired data, the data prediction residual does not depend on γ , so the standard L-curve method cannot be applied. The unregularized GRAPPA reconstruction was included as a baseline.

Performance of WSURE for L_1 -SPIRiT

Using $\varepsilon = 10^{-4}$, we studied the same performance trends for the L_1 -SPIRiT reconstruction method for a range of undersampling factors on both T1-weighted data sets. We performed Poisson-disc random Cartesian undersampling for accelerations between 2 and 6, and we repeated finding the WSURE-optimized and WMSE-optimal L_1 -SPIRiT reconstructions for each sampling pattern. The noise level was held fixed at 13-dB SNR and 14-dB SNR, respectively, for the two data sets (again adding synthetic complex Gaussian noise to the averaged reference k-space before undersampling). We used a coarse-to-fine two-level parameter sweep to find the WSURE-optimized γ , with the coarse level sweeping γ between 10^{-7} and 10 for both data sets, and we used `fminsearch()` with a starting point of $\gamma = 0.01$ (again for both data sets) to find the WMSE-optimal γ . We also computed the L-curve-like log-log pareto optimality curve of the SPIRiT-consistency and sparsity terms [15] and

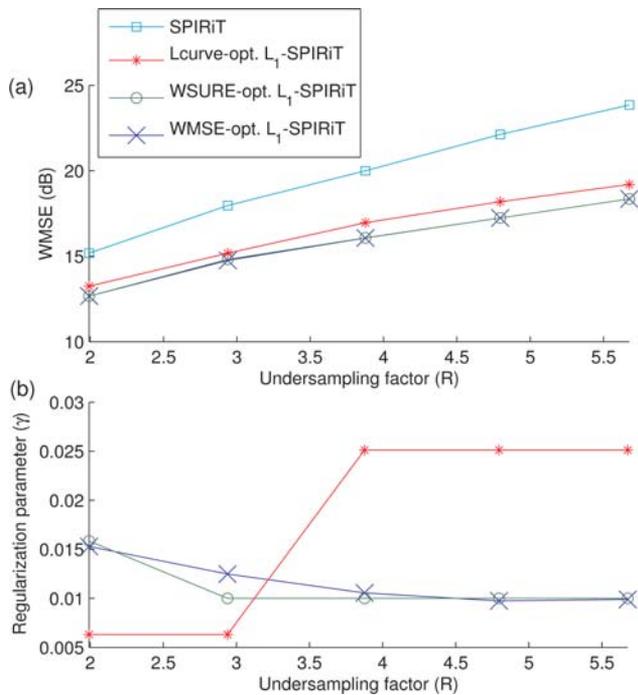


FIG. 6. The sparsity-regularized L_1 -SPIRiT reconstruction for data set #2 with 14-dB input SNR yields lower WMSE (a) for both the WMSE-optimal and WSURE-optimized choices of γ , relative to the unregularized SPIRiT reconstruction. The L_1 -SPIRiT reconstruction using the WSURE-optimized choice is within 0.056 dB of the true WMSE-optimal L_1 -SPIRiT reconstruction. The WMSE-optimal and WSURE-optimized choices for the L_1 -SPIRiT regularization parameter γ (b) tend to decrease slowly as the undersampling factor increases. The larger values of γ from the L-curve-like method increase the WMSE, but not as substantially as for DESIGN. The plotted values of R account for the central k-space calibration region. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

maximized its curvature using the same two-level parameter sweep for both data sets. This L-curve-like method is not a standard L-curve method for the same reason as for the DESIGN reconstruction. Unregularized SPIRiT was provided as a baseline.

Violating the Assumption $\mathbf{x} = \mathbf{G}\mathbf{m}$

At high levels of acceleration, or when the coil channels do not provide sufficient spatial encoding to overcome that acceleration, the linear parallel imaging method symbolized by \mathbf{G} does not satisfy $\mathbf{x} = \mathbf{G}\mathbf{m}$, even for noise-free k-space \mathbf{x} . In such cases, our WSURE metric may not accurately reflect the true WMSE, and automatic parameter tuning may suffer. To investigate the effects of such a violation, we applied L_1 -SPIRiT to the first data set at a high level of undersampling, $R \approx 6$. For this undersampling factor, we repeat the validation experiment carried out earlier on the first data set, using the same range of ϵ and γ as before.

RESULTS

The results of our validation experiments for DESIGN and L_1 -SPIRiT are illustrated for both T1-weighted data sets in

Figures 2 and 3, respectively. The value of ϵ shown is 10^{-4} for both data sets with both DESIGN and L_1 -SPIRiT. The other values of ϵ considered yielded essentially identical curves to these curves, implying that WSURE estimates are fairly insensitive to ϵ for either of these nonlinear methods. For ease of comparison, the constant term in Eq. 10, which depends on the true value \mathbf{x} , is computed from the reference k-space and added to WSURE estimates in the figures. This value is a constant offset, so it is not needed to find the WSURE-optimized γ . In both cases, the trend of WSURE estimates follows the shape of the curves for both true and approximate WMSE, although they tend to deviate more for smaller values of the regularization parameter. In addition, the approximate WMSE values are similar to the true WMSE values, validating the approximation $\mathbf{G}\mathbf{m} \approx \mathbf{x}$ made in Eq. 4. In addition to having similar shapes, these curves share similar minima, validating the accuracy of our proposed WSURE error criterion. The error bars depicting ± 5 sample standard deviations from the sample mean of our 20 WSURE estimates also suggest that a single realization of \mathbf{b} is sufficiently precise to choose γ that minimizes the WMSE effectively. The maximum observed sample standard deviations were 1.2 and 0.63 percent of the mean for DESIGN and L_1 -SPIRiT, respectively.

Figure 4 shows the optimal γ 's and WMSE values for the first T1-weighted data set of the WSURE-optimized and WMSE-optimal DESIGN reconstructions, for a range of noise levels, with 2×2 uniform Cartesian undersampling. The curves for the second data set display similar behavior and can be found online as Supporting Information. The staircase-ramp effect observed in the plot of WSURE-optimized γ is due to the coarseness of the employed parameter sweep. Additional sweep levels would not be meaningful considering that the change in WMSE is so small around the minimum as to be nearly comparable to the standard deviation in the Monte Carlo WSURE estimates. The L-curve-like method maximizing the curvature of the log-log pareto optimality frontier tends to overestimate γ for both these data sets, particularly in the cases of high-input SNR. For the second data set, the L-curve-selected γ does not increase monotonically, indicating difficulty in estimating the second derivative of the log-log pareto curve numerically for this application. The WMSEs of the reconstructions guided by L-curve selection suffer as a result. The multichannel k-space WMSE values for the WMSE-optimal and WSURE-optimized choices of γ show little difference between reconstructions optimal for these error criteria, so we examine the reconstructed sum-of-squares-combined magnitude and difference images for data set #1 with 10.3-dB input SNR and data set #2 with 16.7-dB input SNR, all in Figure 5. Both WMSE-optimal and WSURE-optimized DESIGN reconstructions display similar reduction in noise level over the GRAPPA-reconstructed image and minimal residual image structure in the difference images. Both data sets' reconstructions for DESIGN with L-curve-selected γ portray significant oversmoothing.

The results for WSURE-optimized L_1 -SPIRiT yield nearly WMSE-optimal performance over a range of undersampling factors for both T1-weighted data sets. The trends in the optimal choices of γ and the resulting WMSE's plotted in Figure 6 for the second data set with 14-dB SNR noise

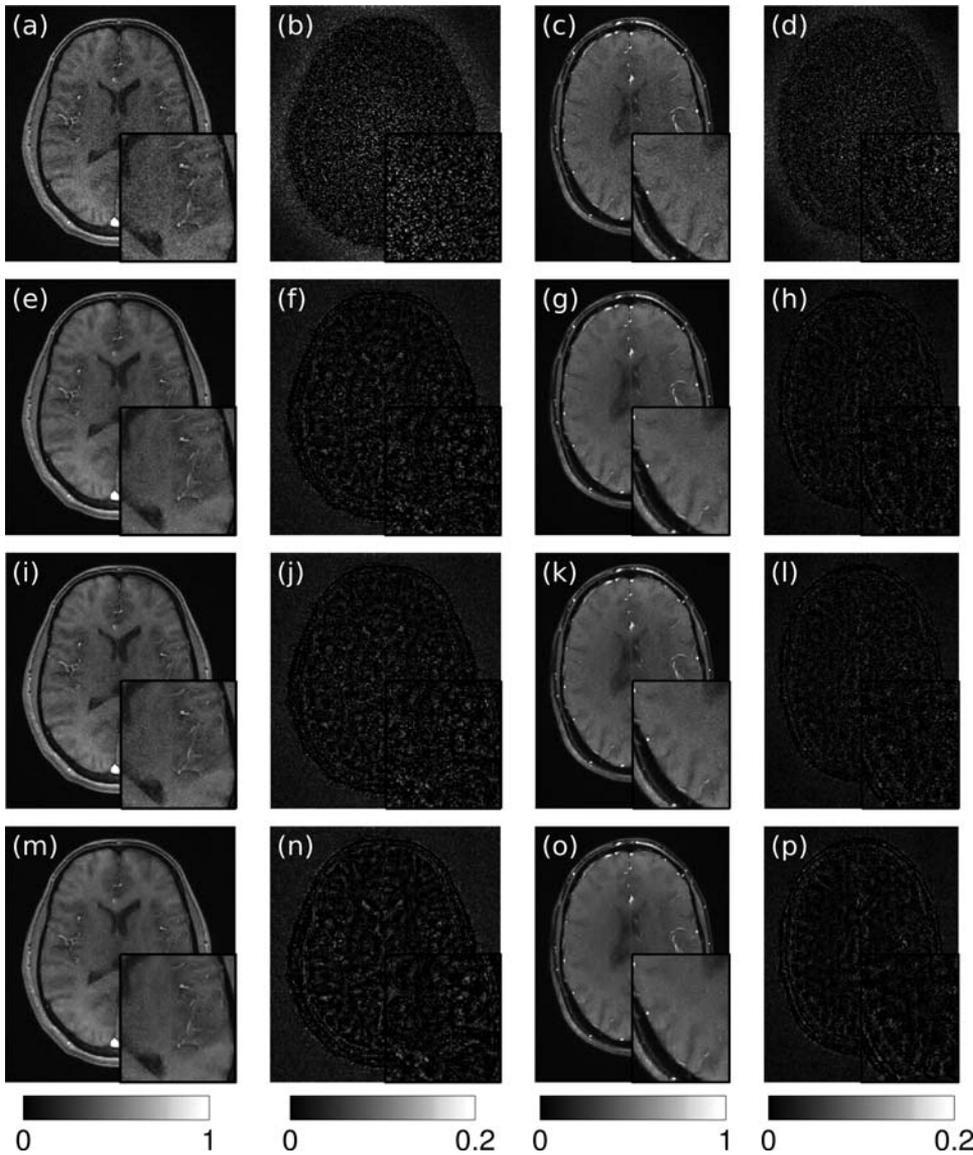


FIG. 7. The unregularized SPIRiT reconstruction (a) and difference image (b) for data set #1 with $R \approx 4$ Poisson-disc undersampling and 13-dB input SNR display significant noise, while L_1 -SPIRiT with WMSE-optimal (e,f), WSURE-optimized (i,j), and L-curve-selected (m,n) choices of γ effectively improve image quality by nearly the same amount. The L-curve-based method results in a slightly oversmoothed image, although the difference is not as apparent as for the DESIGN reconstruction. Similar conclusions can be inferred from the second data set with $R \approx 4$ Poisson-disc undersampling and 14-dB input SNR. The SPIRiT reconstruction (c,d) is relatively noisy, whereas L_1 -SPIRiT with WMSE-optimal (g,h), WSURE-optimized (k,l), and L-curve-selected (o,p) choices of γ all denoise the image while avoiding substantial oversmoothing. The difference images are displayed at $5\times$ scale.

nearly match the optimal performance over the entire range of acceleration. The curves for the first data set with 13-dB input SNR are similar to these curves; they can be found online as Supporting Information. The same staircase-ramp effect is observed for the WSURE-optimized γ curve due to the resolution of the parameter sweep. Although the L-curve-like curvature maximization of the log-log pareto optimality curve yields suboptimal γ curves for both data sets, the WMSEs appear to be much closer for L_1 -SPIRiT than DESIGN. The reconstructed images in Figure 7 confirm the numerical comparisons for the $4\times$ undersampled case for both T1-weighted data sets, as the sum-of-squares-combined magnitude and difference images are nearly indistinguishable between the WSURE-optimized, and WMSE-optimal, and L-curve-based reconstructions. Each yields noticeable noise reduction over the unregularized SPIRiT reconstruction shown for comparison.

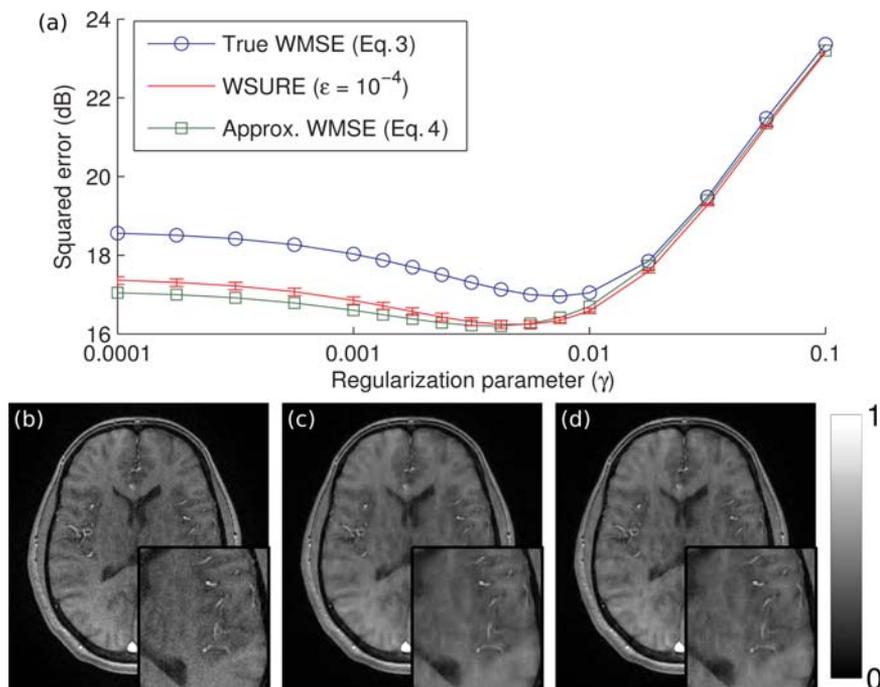
Last, we investigate the effects of violating the $\mathbf{x} = \mathbf{GMx}$ approximation used to form WSURE estimates. In the case of L_1 -SPIRiT with the first T1-weighted data set, shown in

Figure 8, WSURE validation appears to show a breakdown in the approximation starting at $R \approx 6$, but at this high level of undersampling, the image quality of even the WMSE-optimal L_1 -SPIRiT reconstruction is degraded beyond the point of diagnostic utility. The WSURE-optimized choice of γ is a bit off the WMSE-optimal choice, but the additional loss of image quality appears negligible.

DISCUSSION

The experiments validate the use of Monte Carlo WSURE for regularization parameter selection for both DESIGN with uniform Cartesian and L_1 -SPIRiT with nonuniform Cartesian undersampling, over a wide range of acquisition parameters including different undersampling factors and noise levels. Because the derivation of the Monte Carlo WSURE method is applicable to data-preserving reconstruction methods, and the implementation requires only the ability to evaluate the reconstruction function, this approach to parameter selection should readily generalize

FIG. 8. At an acceleration of $R \approx 6$, WSURE for L_1 -SPIRiT, shown for data set #1 with $\epsilon = 10^{-4}$, tracks the SPIRiT-approximated WMSE curve when the $\mathbf{x} = \mathbf{G}\mathbf{m}$ approximation is violated (a). However, both these curves do not follow the true MSE curve as closely as before. The resulting WSURE-optimized γ hence is slightly off, but the overall performance degradation in the WSURE-optimized image (d) is insignificant compared to the low quality of the WMSE-optimal L_1 -SPIRiT reconstruction (c). The unregularized SPIRiT reconstruction (b) is shown for reference. WSURE curve error bars correspond to ± 5 standard deviations from the mean WSURE value. Only the error bars were estimated from 20 realizations of the complex Bernoulli noise vector \mathbf{b} ; the WSURE curves correspond to using just the first realization of \mathbf{b} . [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



to other parallel imaging reconstruction algorithms. Also, the Monte Carlo WSURE can be applied to existing reconstruction methods such as L_1 -SPIRiT without considering the implementations of such methods. Our method requires just two evaluations of the regularized reconstruction per candidate γ . Although hard-coding regularization parameters may suffice to produce high-quality images in some instances, as suggested by the relatively narrow range of WMSE-optimal γ for the presented experiments, being able to choose such parameters automatically would reduce the burden on programmers and practitioners to ensure parameter choices are indeed reasonable. This article describes a novel way of using parallel imaging to estimate the WMSE (or MSE) with SURE for image reconstruction problems that applies to data-preserving methods. This ability to tune parameters for algorithms that preserve the acquired data sets, the proposed method apart both from previous work using SURE, and popular parameter selection methods such as discrepancy principle, cross-validation, and L-curve curvature maximization.

In this article, we consider uniform and nonuniform Cartesian undersampling, but our framework also can handle regularized non-Cartesian parallel imaging reconstructions and reconstructions that do not preserve the data with only minor modifications. However, in non-Cartesian parallel imaging, where the reconstructed k-space typically are located on a Cartesian grid, preserving the data makes less sense, and existing Monte Carlo SURE-based methods can be applied instead. More sophisticated cost functions that preserve edges and local details may also be compatible with the WMSE and WSURE framework.

Our experiments involved optimizing a single-regularization parameter, but our SURE-based method extends to tuning multiple parameters. In the multivariate case, various minimization methods can be used (34, 35). However, these methods may be more sensitive to

initialization and nonconvexity of the error criterion with respect to the regularization parameters.

A limitation of the framework is the approximation $\mathbf{G}\mathbf{m} \approx \mathbf{x}$ that is necessary to relate the measurement locations to the full k-space; at high accelerations, this approximation becomes limited by the number and geometry of the coils (or more generally, the limitations on linear reconstruction \mathbf{G}). We demonstrated that at the level of acceleration where this approximation becomes an issue, the quality of even the WMSE-optimal images is too low to be useful. Another limitation is the need to choose the noise scaling parameter ϵ . Our validation experiments indicate that choosing the absolute best ϵ is not critical, and that a wide range of ϵ would yield similar results, even for nonlinear reconstruction methods like DESIGN and L_1 -SPIRiT.

CONCLUSIONS

We described an extension to the Monte Carlo SURE method for regularized autocalibrating parallel imaging reconstruction methods such as DESIGN and L_1 -SPIRiT. We performed experiments on real data to validate WSURE estimates of the WMSE and demonstrated that the Monte Carlo WSURE estimates yield nearly WMSE-optimal reconstruction quality for both uniform undersampling with DESIGN and nonuniform Cartesian undersampling with L_1 -SPIRiT. In these experiments, we also compared our method against an L-curve-like method for parameter selection that tended to oversmooth the reconstructed images. We concluded with a discussion of useful extensions and possible limitations of our method.

REFERENCES

1. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Mag Res Med* 1999;42:952–962.

2. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Mag Res Med* 2002;47:1202–1210.
3. Lustig M, Pauly JM. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Mag Res Med* 2010;64:457–471.
4. Lin FH, Kwong KK, Belliveau JW, Wald LL. Parallel imaging reconstruction using automatic regularization. *Mag Res Med* 2004;51:559–567.
5. Liang D, Liu B, Wang J, Ying L. Accelerating SENSE using compressed sensing. *Mag Res Med* 2009;62:1574–1584.
6. Weller DS, Polimeni JR, Grady L, Wald LL, Adalsteinsson E, Goyal VK. Denoising sparse images from GRAPPA using the nullspace method (DESIGN). *Mag Res Med* 2012;68:1176–1189.
7. Murphy M, Alley M, Demmel J, Keutzer K, Vasanaawala S, Lustig M. Fast ℓ_1 -SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *IEEE Trans Med Imag* 2012;31:1250–1262.
8. Ying L, Liu B, Steckner M, Wu G, Wu M, Li SJ. A statistical approach to SENSE regularization with arbitrary k-space trajectories. *Mag Res Med* 2008;60:414–421.
9. Bernstein MA, King KF, Zhou XJ. *Handbook of MRI Pulse Sequences*. New York: Elsevier; 2004.
10. Engl HW. Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *J Optim Theory Appl* 1987;52:209–215.
11. Craven P, Wahba G. Smoothing noisy data with spline functions. *Numer Math* 1979;31:377–403.
12. Girard DA. The fast Monte-Carlo cross-validation and C_L procedures: Comments, new results and application to image recovery problems. *Comput Statist* 1995;10:205–231.
13. Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev* 1992;34:561–580.
14. Hansen PC, O’Leary DP. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J Sci Comp* 1993;14:1487–1506.
15. Girard DA. The fast Monte-Carlo cross-validation and C_L procedures: Comments, new results and application to image recovery problems—rejoinder. *Comput Statist* 1995;10:251–258.
16. Ramani S, Liu Z, Rosen J, Nielsen JF, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. *IEEE Trans Im Proc* 2012;21:3659–3672.
17. Vasanaawala SS, Alley MT, Hargreaves BA, Barth RA, Pauly JM, Lustig M. Improved pediatric MR imaging with compressed sensing. *Radiology* 2010;256:607–616.
18. Vasanaawala SS, Murphy MJ, Alley MT, Lai P, Keutzer K, Pauly JM, Lustig M. Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients. In *Proc IEEE Intl Symp. Biomed. Imag., Chicago, Illinois, USA, 2011*; pp. 1039–1043.
19. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Mag Res Med* 2007;58:1182–1195.
20. Stein C. Estimation of the mean of a multivariate normal distribution. *Ann Stat* 1981;9:1135–1151.
21. Zhang XP. Adaptive denoising based on SURE risk. *IEEE Signal Process Lett* 1998;5:265–267.
22. Blu T, Luisier F. The SURE-LET approach to image denoising. *IEEE Trans Im Proc* 2007;16:2778–2786.
23. Ramani S, Blu T, Unser M. Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans Im Proc* 2008;17:1540–1554.
24. Vonesch C, Ramani S, Unser M. Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint. In *Proc IEEE Intl Conf on Image Processing, San Diego, California, USA, 2008*. pp. 665–668.
25. Eldar YC. Generalized SURE for exponential families: applications to regularization. *IEEE Trans Sig Proc* 2009;57:471–481.
26. Giryes R, Elad M, Eldar YC. The projected GSURE for automatic parameter tuning in iterative shrinkage methods. *Appl Comput Harmon Anal* 2011;30:407–422.
27. Ramani S, Weller DS, Nielsen JF, Fessler JA. Non-Cartesian MRI reconstruction with automatic regularization via Monte-Carlo SURE. *IEEE Trans Med Imag* 2013;32:1411–1422.
28. Marin A, Chaux C, Pesquet JC, Ciuciu P. Image reconstruction from multiple sensors using Stein’s principle. Application to parallel MRI. In *Proc IEEE Intl Symp Biomed Imag., Chicago, Illinois, USA, 2011*. pp. 465–468.
29. Weller DS, Ramani S, Nielsen JF, Fessler JA. SURE-based parameter selection for parallel MRI reconstruction using GRAPPA and sparsity. In *Proc IEEE Intl Symp Biomed Imag, San Francisco, California, USA, 2013*. pp. 942–945.
30. Weller DS, Ramani S, Nielsen JF, Fessler JA. Automatic L1-SPIRiT regularization parameter selection using Monte-Carlo SURE estimation. In *Proceedings of the 21st Annual Meeting of ISMRM, Salt Lake City, Utah, USA, 2013*. p. 2602.
31. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Mag Res Med* 1990;16:192–225.
32. Weller DS, Polimeni JR, Grady L, Wald LL, Adalsteinsson E, Goyal VK. Accelerated parallel magnetic resonance imaging reconstruction using joint estimation with a sparse signal model. In *IEEE Workshop on Statistical Signal Processing, Ann Arbor, Michigan, USA, 2012*. pp. 221–224.
33. Goldstein T, Osher S. The split Bregman method for L1-regularized problems. *SIAM J Imaging Sci* 2009;2:323–343.
34. Nelder JA, Mead R. A simplex method for function minimization. *Comput J* 1965;7:308–313.
35. Kolda TG, Lewis RM, Torczon V. Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev* 2003;45:385–482.

LIST OF SUPPLEMENTARY FIGURES

Figure S1: The true and approximate WMSE validation curves for DESIGN for data set #1 (a) and data set #2 (b) are compared against WSURE repeated for various ε 's spanning several orders of magnitude. The curves (just a single trial for each point) are nearly identical, and they all provide decent approximations of the true and approximate WMSE's for both T_1 -weighted data sets.

Figure S2: The true and approximate WMSE validation curves for L_1 -SPIRiT for data set #1 (a) and data set #2 (b) are compared against WSURE repeated for various ε 's spanning several orders of magnitude. The curves (just a single trial for each point) are nearly identical, and they all provide decent approximations of the true and approximate WMSE's for both T_1 -weighted data sets.

Figure S3: The sparsity-promoting DESIGN method yields similarly lower WMSE (a) for both WMSE-optimal and WSURE-optimized choices of γ , relative to the un-regularized GRAPPA reconstruction, for the second data set. The WMSE of the DESIGN reconstruction using the WSURE-optimized choice is within 0.047 dB of the true WMSE-optimal DESIGN reconstruction. This behavior is consistent with the WMSE-optimal and WSURE-optimized choices for the DESIGN regularization parameter γ (b) being nearly the same for this example. The L-curve-style method, however, appears to overestimate γ , yielding images with slightly higher WMSE, and in the high-SNR case, again worse than performing un-regularized GRAPPA. The non-monotonic behavior of the L-curve estimates of γ confirm the difficulty of estimating the maximum curvature.

Figure S4: The sparsity-regularized L_1 -SPIRiT reconstruction for data set #1 yields similarly lower WMSE (a) for both the WMSE-optimal and WSURE-optimized choices of γ , relative to the un-regularized SPIRiT reconstruction. The L_1 -SPIRiT reconstruction using the WSURE-optimized choice is within 0.057 dB of the true WMSE-optimal L_1 -SPIRiT reconstruction. The WMSE-optimal and WSURE-optimized choices for the L_1 -SPIRiT regularization parameter γ (b) tend to decrease slowly as the undersampling factor increases. The larger values of γ from the L-curve-like method increase the WMSE, but not as substantially as for DESIGN. The plotted values of R account for the central k-space calibration region.

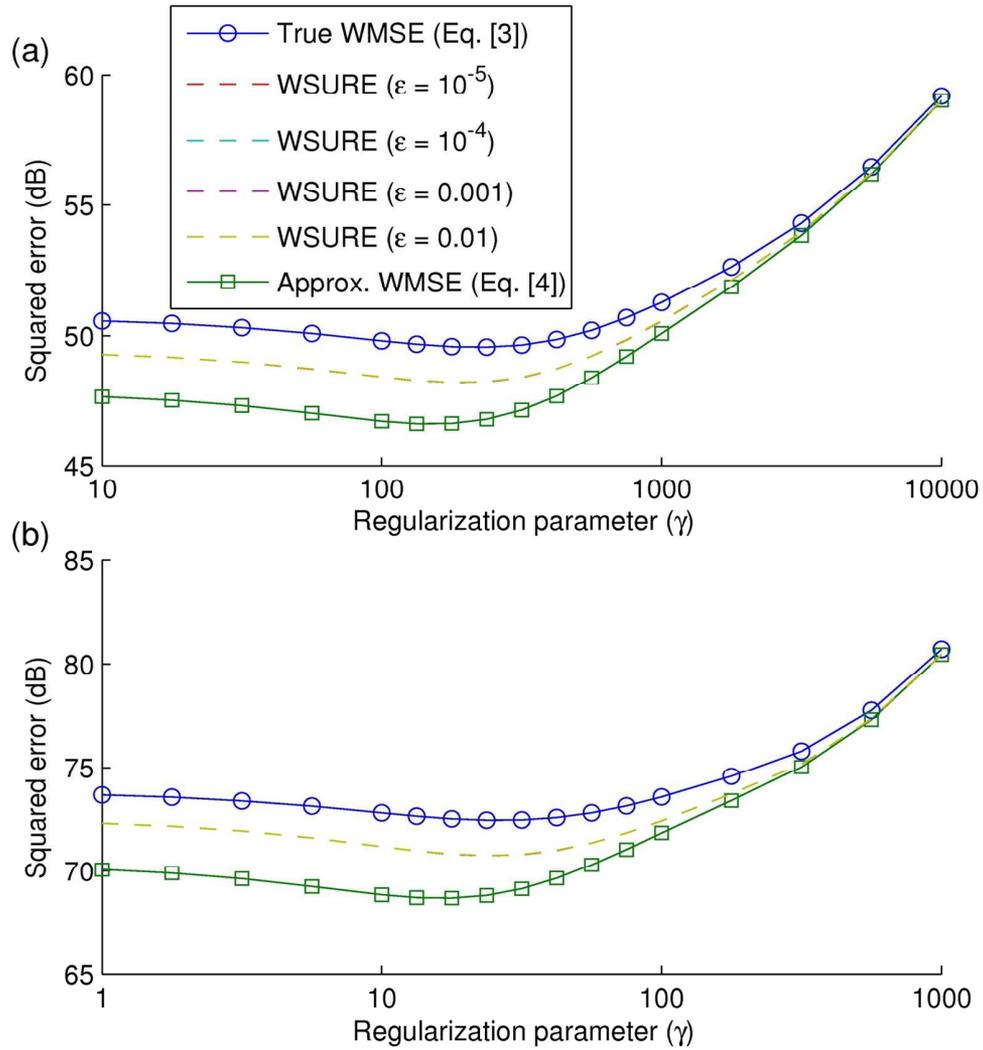


Figure S1: The true and approximate WMSE validation curves for DESIGN for data set #1 (a) and data set #2 (b) are compared against WSURE repeated for various ϵ 's spanning several orders of magnitude. The curves (just a single trial for each point) are nearly identical, and they all provide decent approximations of the true and approximate WMSE's for both T_1 -weighted data sets.
 139x153mm (300 x 300 DPI)

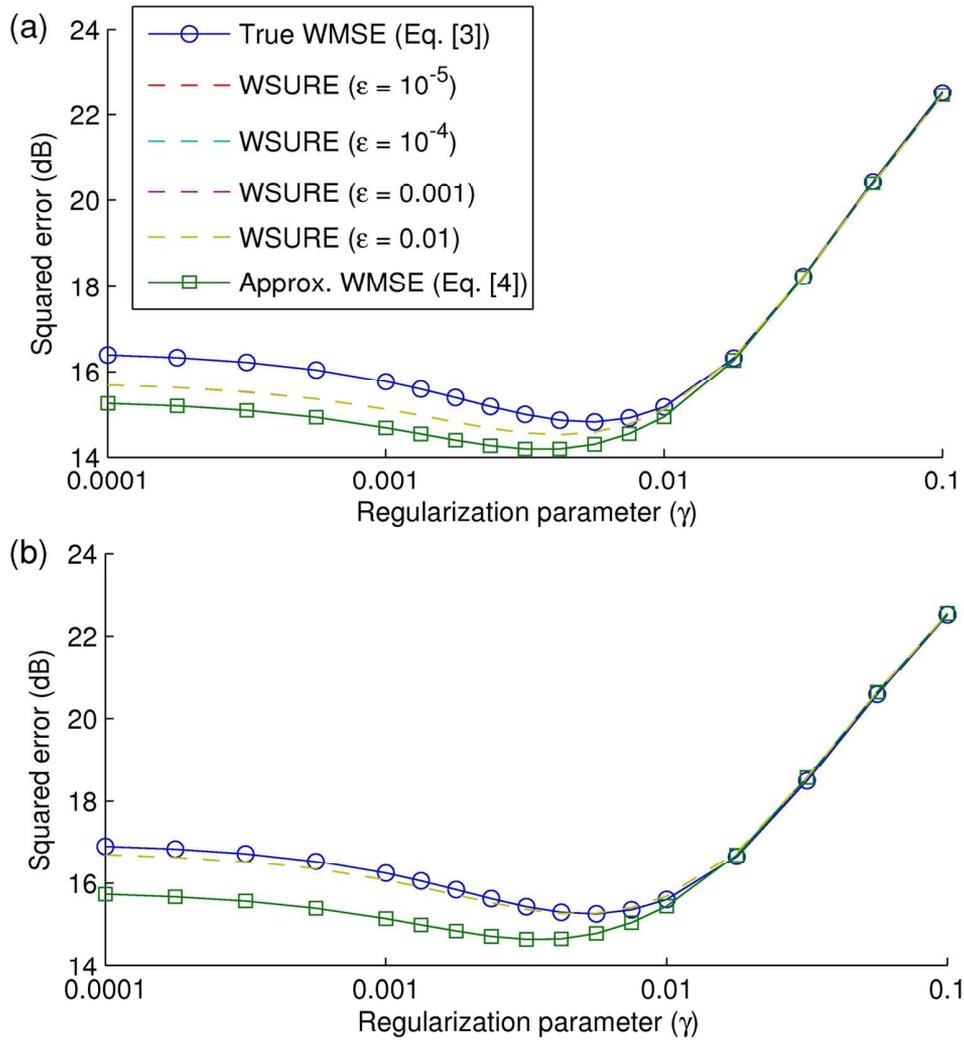


Figure S2: The true and approximate WMSE validation curves for L_1 -SPIRiT for data set #1 (a) and data set #2 (b) are compared against WSURE repeated for various ϵ 's spanning several orders of magnitude. The curves (just a single trial for each point) are nearly identical, and they all provide decent approximations of the true and approximate WMSE's for both T_1 -weighted data sets.
 139x153mm (300 x 300 DPI)

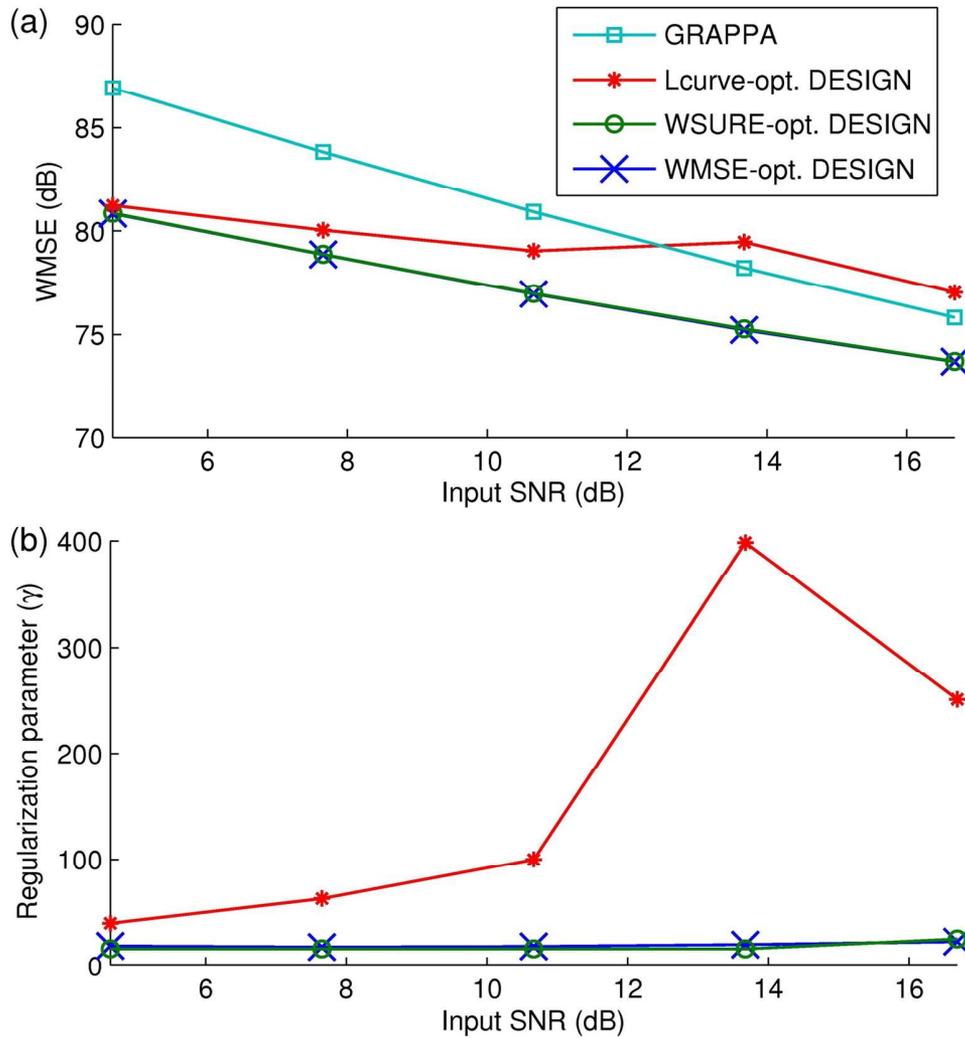


Figure S3: The sparsity-promoting DESIGN method yields similarly lower WMSE (a) for both WMSE-optimal and WSURE-optimized choices of γ , relative to the un-regularized GRAPPA reconstruction, for the second data set. The WMSE of the DESIGN reconstruction using the WSURE-optimized choice is within 0.047 dB of the true WMSE-optimal DESIGN reconstruction. This behavior is consistent with the WMSE-optimal and WSURE-optimized choices for the DESIGN regularization parameter γ (b) being nearly the same for this example. The L-curve-style method, however, appears to overestimate γ , yielding images with slightly higher WMSE, and in the high-SNR case, again worse than performing un-regularized GRAPPA. The non-monotonic behavior of the L-curve estimates of γ confirm the difficulty of estimating the maximum curvature.

139x153mm (300 x 300 DPI)

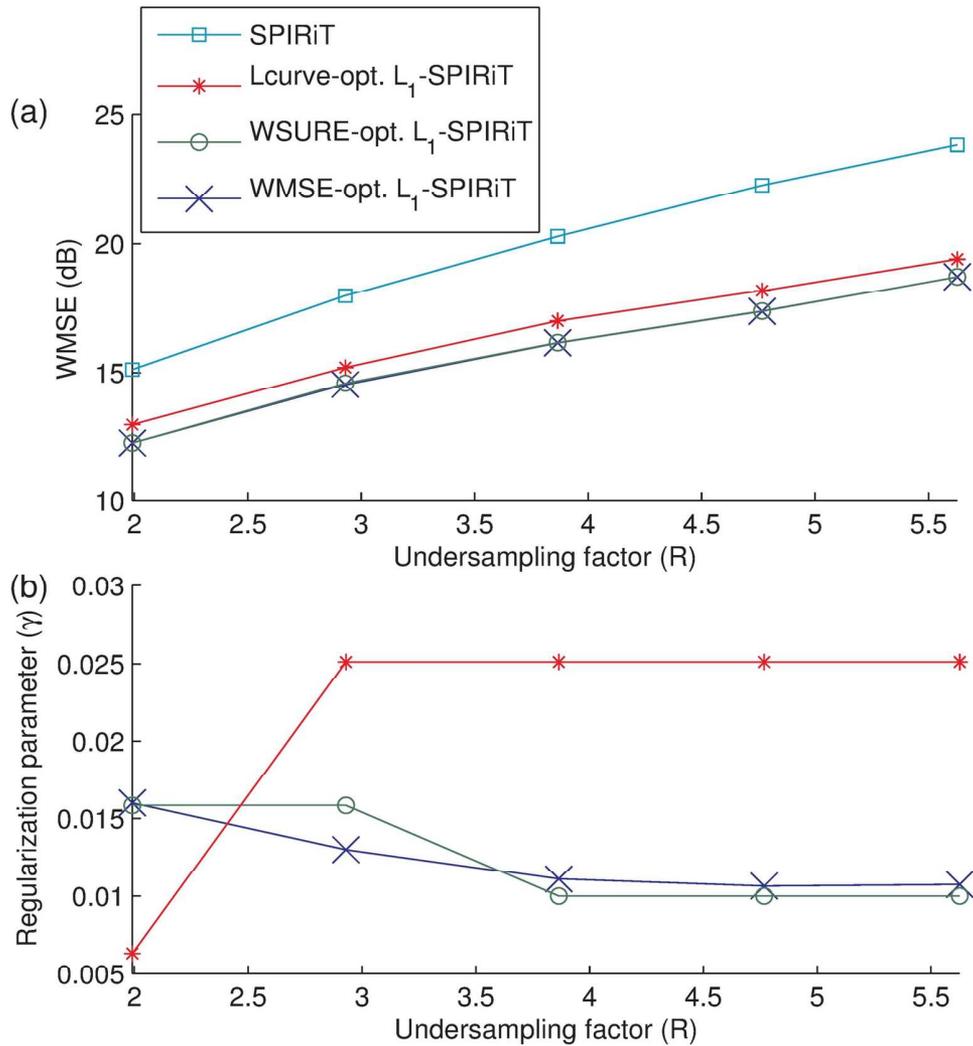


Figure S4: The sparsity-regularized L_1 -SPIRiT reconstruction for data set #1 yields similarly lower WMSE (a) for both the WMSE-optimal and WSURE-optimized choices of γ , relative to the un-regularized SPIRiT reconstruction. The L_1 -SPIRiT reconstruction using the WSURE-optimized choice is within 0.057 dB of the true WMSE-optimal L_1 -SPIRiT reconstruction. The WMSE-optimal and WSURE-optimized choices for the L_1 -SPIRiT regularization parameter γ (b) tend to decrease slowly as the undersampling factor increases. The larger values of γ from the L-curve-like method increase the WMSE, but not as substantially as for DESIGN. The plotted values of R account for the central k-space calibration region. 139x153mm (300 x 300 DPI)