Non-Cartesian MRI Reconstruction With Automatic Regularization Via Monte-Carlo SURE

Sathish Ramani*, *Member, IEEE*, Daniel S. Weller, *Member, IEEE*, Jon-Fredrik Nielsen, and Jeffrey A. Fessler, *Fellow, IEEE*

***Abstract*—Magnetic resonance image (MRI) reconstruction from undersampled k-space data requires regularization to reduce noise and aliasing artifacts. Proper application of regularization however requires appropriate selection of associated regularization parameters. In this work, we develop a data-driven regularization parameter adjustment scheme that minimizes an estimate [based on the principle of Stein's unbiased risk estimate (SURE)] of a suitable weighted squared-error measure in k-space. To compute this SURE-type estimate, we propose a Monte-Carlo scheme that extends our previous approach to inverse problems (e.g., MRI reconstruction) involving complex-valued images. Our approach depends only on the output of a given reconstruction algorithm and does not require knowledge of its internal workings, so it is capable of tackling a wide variety of reconstruction algorithms and nonquadratic regularizers including total variation and those based on the ℓ_1-norm. Experiments with simulated and real MR data indicate that the proposed approach is capable of providing near mean squared-error optimal regularization parameters for single-coil undersampled non-Cartesian MRI reconstruction.**

***Index Terms*—Image reconstruction, Monte-Carlo methods, non-Cartesian MRI, regularization parameter, Stein's unbiased risk estimate (SURE).**

I. Introduction

IMAGE reconstruction is a crucial task in magnetic resonance imaging (MRI). Model-based reconstruction methods [1] can improve image-quality over direct methods such as iFFT- or gridding-based reconstruction [2], especially for undersampled k-space data. The problem is usually solved by minimizing a cost function involving a model-based data-fidelity term and regularization. Regularization is often included to reduce ill-posedness of the problem for undersampled cases, to stabilize the reconstruction process and also to incorporate prior information about the object being reconstructed. Nonquadratic regularizers can better suppress noise and aliasing artifacts compared to quadratic ones [3]. Sparsity promoting regularizers such as those based on the ℓ_1-norm and edge-preserving total variation (TV) are popular nonquadratic regularizers in MRI [4]–[9]. Successful regularization requires careful selection of associated *regularization parameters* that control the strength of these regularizers during reconstruction. These parameters are often set manually (based on visual perception) for MRI reconstruction. In this paper, we focus on the problem of automatic selection of these parameters for MRI reconstruction from undersampled k-space data.

Various quantitative criteria exist for automatic selection of parameters for regularized image reconstruction in general [10], [11]. These may be broadly classified as those based on the discrepancy principle [10], [11], the L-curve [12]–[14], generalized cross-validation (GCV) [15]–[19] and estimation of (weighted) mean squared-error (MSE, also known as *risk*) using the principles underlying *Stein's unbiased risk estimate* (SURE) [20]–[27]. Unlike task-based methods [28]–[30] that focus on developing quality assessment criteria specific to a given task (e.g., detecting a lesion), the above parameter selection methods only determine a "reasonable" solution from a "feasible set" that is predetermined by the chosen cost function.

Among these methods, we focus on the weighted MSE (WMSE) based approach since WMSE is easily manipulated and estimated using the SURE-framework [23], [24], [27] and also because it is commonly used to quantify reconstruction quality [22]–[27]. Moreover, SURE-based methods can tackle noniterative nonlinear reconstruction [22], [25], [26] and iterative regularized reconstruction using nonquadratic regularizers [23], [24], [27] and also provide (near) MSE-optimal (regularization) parameter selection [22]–[27]. SURE-based parameter selection assumes that real- or complex-valued noise in the observed data follows a Gaussian distribution with known mean and covariance, so it is well-suited for MRI.

Previous applications of SURE-type parameter selection for MRI include noniterative denoising of magnitude images [25], SENSitivity Encoding [31] (SENSE) based noniterative reconstruction from *uniformly* undersampled multi-coil *Cartesian* k-space data [26] and iterative MRI reconstruction (using nonquadratic regularizers) from single-coil *Cartesian* k-space data with *arbitrary* undersampling [27]. These papers derive *analytically* a (weighted) SURE-type estimate of a (weighted) MSE for a particular (iterative) reconstruction algorithm.

In this work, we propose a SURE-based regularization parameter selection method for iterative MRI reconstruction from

Manuscript received January 08, 2013; revised March 28, 2013; accepted April 03, 2013. Date of publication April 12, 2013; date of current version July 27, 2013. This work was supported by the National Institutes of Health under Grant P01 CA87634 and Grant F32 EB015914 and by CPU donations from Intel. *Asterisk indicates corresponding author.*

*S. Ramani is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: sramani@umich.edu).

D. S. Weller and J. A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: dweller@alum.mit.edu; fessler@umich.edu).

J.-F. Nielsen is with the fMRI Laboratory, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: jfnielse@umich.edu).

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

Digital Object Identifier 10.1109/TMI.2013.2257829

undersampled data using nonquadratic regularizers. Unlike earlier work [23]–[27], we propose a Monte-Carlo scheme for computing the desired weighted SURE-type estimate. This Monte-Carlo scheme extends our previous work for real-valued denoising algorithms [32] to complex-valued reconstruction algorithms with application to MRI reconstruction. Our Monte-Carlo method depends only on the output of a given reconstruction algorithm and does not require knowledge of its internal workings beyond confirming that it satisfies certain (weak) differentiability conditions, so it is very flexible and can be applied to a wide variety of iterative/noniterative nonlinear algorithms.

We illustrate the efficacy of the proposed Monte-Carlo scheme for MRI reconstruction from single-coil undersampled *non-Cartesian* k-space data with several nonquadratic regularizers such as a smooth edge-preserving one, TV and an ℓ_1-regularizer. We present numerical results for simulations with the analytical Shepp-Logan phantom [33] and experiments with real GE phantom data and *in vivo* human brain data. These results extend those in our previous work [27] for MRI reconstruction from single-coil undersampled Cartesian data. We demonstrate that the proposed Monte-Carlo SURE-based method provides near-MSE-optimal regularization parameter selection and performs equally well or better than GCV for nonlinear algorithms [18], [27, eq. (7)]. Methods proposed in this paper can also be extended to tackle nonquadratic regularization based iterative parallel MRI reconstruction from Cartesian and non-Cartesian k-space data with arbitrary undersampling (see Section VII).

The paper is organized as follows. We introduce our data model and describe the parameter selection problem mathematically in Section II. We briefly review the principles underlying SURE in Section III and describe the proposed Monte-Carlo method in detail in Section IV. We briefly describe regularized iterative single-coil non-Cartesian MRI reconstruction in Section V. We present a variety of experimental results in Section VI and discuss implementation aspects and possible extensions to this work in Section VII. We finally conclude with Section VIII.

In the rest of the paper, $(\cdot)^\top$, $(\cdot)'$ respectively denote the non-Hermitian and Hermitian transposes, and $(\cdot)_\mathcal{R}$ and $(\cdot)_\mathcal{I}$ respectively indicate the real and imaginary components of a complex vector or matrix. The mth element of any vector $\mathbf{y}$ is denoted by either $[\mathbf{y}]_m$ or y_m and the mnth element of any matrix $\mathbf{A}$ is written as $[\mathbf{A}]_{mn}$. For any vector $\mathbf{y}$ and any matrix $\mathbf{W}$, $\|\mathbf{y}\|_\mathbf{W}^2 \triangleq \mathbf{y}'\mathbf{W}\mathbf{y}$.

II. Problem Description

A. Data Model

In MRI, noise originates in the analog domain (due to thermal fluctuations of spins) before acquisition of k-space samples but can be modeled reasonably accurately as additive Gaussian in the acquired k-space samples. So, we use the following data-model [1, eq. (12)]:

$$\mathbf{y} = \mathbf{y}_{\mathrm{true}} + \boldsymbol{\xi} \tag{1}$$

where we assume that $\mathbf{y}_{\mathrm{true}} \in \mathbb{C}^M$, containing samples of the true unknown MR signal, is a deterministic unknown, $\mathbf{y} \in \mathbb{C}^M$ contains noisy measurements, and $\boldsymbol{\xi} \in \mathbb{C}^M$ is a zero-mean complex-valued Gaussian random vector with covariance matrix $\boldsymbol{\Omega} \in \mathbb{C}^{M \times M}$.

At this point, (1) does not involve discretization of the underlying continuous-domain object χ_{true} that is being scanned. Thus, (1) can accommodate continuous-domain physical-effects representative of MR physics and imaging such as transverse relaxation, inhomogeneity of the applied magnetic field, chemical shifts and nonuniform sensitivity of receive coils [1, eq. (10)], via $\mathbf{y}_{\mathrm{true}}$. It also applies to several types of MRI including single-coil/parallel imaging, undersampled Cartesian/non-Cartesian imaging and combinations thereof.

B. Image Reconstruction

For the purpose of image reconstruction, we use the following discretized *linear* model [1, eq. (18)]

$$\mathbf{y} = \mathbf{A}\mathbf{x}_{\mathrm{true}} + \boldsymbol{\xi} \tag{2}$$

that is based on a discretization [1, eq. (14)], $\mathbf{x}_{\mathrm{true}}$, of the continuous-domain object χ_{true}. This discretization correspondingly yields [1, eqs. (14)–(17)] a system matrix, $\mathbf{A}$, that approximates continuous-domain imaging operations such as those mentioned in Section II-A. The matrix $\mathbf{A}$ depends mainly upon (among other factors such as the pulse sequence and coil geometry) the k-space trajectory used to acquire $\mathbf{y}$ and is assumed to be known. While $\mathbf{A}$ is essential for image reconstruction, we remark that $\mathbf{x}_{\mathrm{true}}$ is a hypothetical object that is not necessary for the methods proposed in this paper and is used purely for validating our simulations. For an appropriate discretization [1], $\mathbf{A}$ represents (nonuniform) discrete Fourier transform for (non-Cartesian) single-coil imaging (ignoring field inhomogeneity and relaxation effects) while for parallel MRI, it corresponds to the combined Fourier and spatial sensitivity encoding matrix [3].

Given (1) and (2), the goal of image reconstruction is to obtain a discretized estimate, $\widehat{\mathbf{x}}$, of χ_{true} from $\mathbf{y}$. This corresponds to an ill-posed inverse problem when $M < N$ and is usually tackled in a regularized-reconstruction framework where an iterative reconstruction algorithm is applied on $\mathbf{y}$ to yield $\widehat{\mathbf{x}}$. We denote the reconstruction process by

$$\widehat{\mathbf{x}} = \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y}) \tag{3}$$

where $\mathbf{u}_{\boldsymbol{\lambda}} : \mathbb{C}^M \rightarrow \mathbb{C}^N$ is a (possibly nonlinear) operator representative of the corresponding iterative reconstruction algorithm. The vector $\boldsymbol{\lambda}$ in $\mathbf{u}_{\boldsymbol{\lambda}}$ denotes one or more tunable parameters (e.g., number of iterations, regularization strength) that characterize the reconstruction method and govern the quality of $\widehat{\mathbf{x}}$. Selecting a suitable $\boldsymbol{\lambda}$ thus plays an important role in problems such as (3). Often, $\boldsymbol{\lambda}$ is adjusted manually based on visual perception of $\widehat{\mathbf{x}}$. In this work, we focus on quantitative methods for selecting $\boldsymbol{\lambda}$ automatically. Specifically, we propose to use a weighted squared-error measure in the measurement domain that can be estimated using Stein's principle [20], [21] and then minimized to yield an appropriate choice of $\boldsymbol{\lambda}$.

C. Weighted Squared-Error Measures

In imaging inverse problems, reconstruction quality is often quantified using mean squared-error, $\mathrm{MSE}(\boldsymbol{\lambda}) \triangleq N^{-1}\|\mathbf{x}_{\mathrm{true}} - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\|_2^2$, and is thus a reasonable metric for adjusting $\boldsymbol{\lambda}$. However, $\mathrm{MSE}(\boldsymbol{\lambda})$ is neither accessible in practice (due to its dependence on $\mathbf{x}_{\mathrm{true}}$) nor amenable for estimation[1] (e.g., using Stein's principle) in ill-posed inverse problems due to the ill-posedness of (2) for $M < N$ [21], [23], [27].

1) Previous Extensions to MSE: To circumvent this difficulty, some authors [21], [23] have focussed on

$$\text{Projected-MSE}(\boldsymbol{\lambda}) \triangleq M^{-1}\|\mathbf{P}[\mathbf{x}_{\mathrm{true}} - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]\|_2^2 \quad (4)$$

where $\mathbf{P} \triangleq \mathbf{A}'(\mathbf{A}\mathbf{A}')^{\dagger}\mathbf{A}$, $(\cdot)^{\dagger}$ represents pseudo-inverse. Another alternative [11], [27] is

$$\text{Predicted-MSE}(\boldsymbol{\lambda}) \triangleq M^{-1}\|\mathbf{A}[\mathbf{x}_{\mathrm{true}} - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]\|_2^2. \quad (5)$$

Both of these metrics are tractable with Stein's principle [21], [23], [27]. In our previous work [27], we considered a weighted variant

$$\mathrm{WMSE}(\boldsymbol{\lambda}) \triangleq M^{-1}\|\mathbf{A}[\mathbf{x}_{\mathrm{true}} - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]\|_{\mathbf{W}}^2 \quad (6)$$

that subsumes both $\text{Projected-MSE}(\boldsymbol{\lambda})$ and Predicted-$MSE(\boldsymbol{\lambda})$ for appropriate choices of the symmetric positive semi-definite weighting matrix $\mathbf{W} \succeq \mathbf{0}$ [27, Sec. III-B]. All of these metrics that depend on $\mathbf{x}_{\mathrm{true}}$ assume that the observed data $\mathbf{y}$ follows the discretized linear model in (2). For such a model (2), $\mathrm{WMSE}(\boldsymbol{\lambda})$ can be unbiasedly estimated using Stein's principle to yield $\mathrm{WSURE}(\boldsymbol{\lambda})$ [27, eq. (12)] when $\boldsymbol{\xi}$ in (2) is Gaussian [27, Th. 2]. Unlike $\mathrm{MSE}(\boldsymbol{\lambda})$ however, $\mathrm{WMSE}(\boldsymbol{\lambda})$ evaluates the error in the measurement-domain, i.e., the range space of $\mathbf{A}$; for MRI, $\mathrm{WMSE}(\boldsymbol{\lambda})$ corresponds to evaluating weighted squared-error in k-space. Despite this dissimilarity from $\mathrm{MSE}(\boldsymbol{\lambda})$, we found that $\mathrm{WMSE}(\boldsymbol{\lambda})$, via its estimate $\mathrm{WSURE}(\boldsymbol{\lambda})$ [27, eq. (12)], can be used to obtain near-MSE-optimal regularization parameters for iterative nonlinear image-deblurring and MRI reconstruction from undersampled Cartesian k-space data [27].

Using Stein's principle [20], [21] to estimate $\mathrm{WMSE}(\boldsymbol{\lambda})$ involves substituting $\mathbf{A}\mathbf{x}_{\mathrm{true}} = \mathbf{y} - \boldsymbol{\xi}$ from (2) in $\mathrm{WMSE}(\boldsymbol{\lambda})$ (6) and exploiting the statistics of $\boldsymbol{\xi}$ to analytically evaluate $\boldsymbol{\xi}$-related terms in the expectation sense [27, Th. 1]. The resulting *unbiased* estimate $\mathrm{WSURE}(\boldsymbol{\lambda})$ [27, eq. (12)] is independent of $\mathbf{A}\mathbf{x}_{\mathrm{true}}$ and depends only on $\mathbf{y}$, a first-order differential response of $\mathbf{u}_{\boldsymbol{\lambda}}$ and the mean and covariance of $\boldsymbol{\xi}$ thereby making it a practical proxy for $\mathrm{WMSE}(\boldsymbol{\lambda})$. However, the *unbiasedness* of $\mathrm{WSURE}(\boldsymbol{\lambda})$ to $\mathrm{WMSE}(\boldsymbol{\lambda})$ is meaningful only when the observed data follows (2). The discretized linear model (2), although crucial for image reconstruction, does not adequately describe how imaging systems work in practice: observed data $\mathbf{y}$ often involves continuous-domain imaging operations, e.g., representative of MR physics described in Section II-A, that may not be completely captured by the discretization in $\mathbf{A}\mathbf{x}_{\mathrm{true}}$. Thus, since $\mathrm{WSURE}(\boldsymbol{\lambda})$ depends on $\mathbf{y}$ and not on $\mathbf{A}\mathbf{x}_{\mathrm{true}}$, a discrepancy arises in *reasoning* that $\mathrm{WSURE}(\boldsymbol{\lambda})$ is unbiased for practical imaging inverse problems.

2) Proposed Measure: To avoid this discrepancy in reasoning, we propose to consider the following WMSE metric with respect to the True Data $\mathbf{y}_{\mathrm{true}}$ since $\mathbf{y}_{\mathrm{true}}$ accounts for continuous-domain imaging operations

$$\mathrm{WMSETD}(\boldsymbol{\lambda}) \triangleq M^{-1}\|\mathbf{y}_{\mathrm{true}} - \mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\|_{\mathbf{W}}^2. \quad (7)$$

We still require $\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ in (7) because we are reconstructing a discretized version, i.e., $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$, of the original continuous-domain object χ_{true} so that $\mathbf{A}$ maps $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ to its corresponding k-space vector. Similar to $\mathrm{WMSE}(\boldsymbol{\lambda})$, $\mathrm{WMSETD}(\boldsymbol{\lambda})$ is also a measurement-domain error metric that is not directly accessible due to its dependence on the true unknown samples $\mathbf{y}_{\mathrm{true}}$. However, since $\mathbf{y}_{\mathrm{true}}$ describes MR data-acquisition more realistically via continuous-domain operations than $\mathbf{A}\mathbf{x}_{\mathrm{true}}$, $\mathrm{WMSETD}(\boldsymbol{\lambda})$ is a more accurate representation of the k-space error than $\mathrm{WMSE}(\boldsymbol{\lambda})$. Below, we show that Stein's principle [20], [21] can be used to estimate[2] $\mathrm{WMSETD}(\boldsymbol{\lambda})$ and leads to an expression for $\mathrm{WSURE}(\boldsymbol{\lambda})$ that is very similar to that reported in our previous work [27, eq. (12)].

Due to the generality of (1) and (2), we can use $\mathrm{WMSETD}(\boldsymbol{\lambda})$ [via $\mathrm{WSURE}(\boldsymbol{\lambda})$] to tune $\boldsymbol{\lambda}$ in a variety of MRI reconstruction problems including single-coil/multi-coil MRI reconstruction (from undersampled data) with/without compensation for field-inhomogeneity and relaxation effects. However, the appropriateness of $\mathrm{WMSETD}(\boldsymbol{\lambda})$ for a given MRI technique needs to be validated using numerical experiments on a case-by-case basis. In this paper, we consider *single-coil non-Cartesian* MRI ignoring field-inhomogeneity and relaxation effects as an extension to our previous work [27] that focussed on *single-coil Cartesian*[3] MRI. We present experimental results in Section VI illustrating that $\mathrm{WSURE}(\boldsymbol{\lambda})$ can provide near-MSE-optimal regularization parameter selection for regularized MRI reconstruction from *single-coil undersampled non-Cartesian* k-space data. We also briefly discuss extensions to parallel MRI in Section VII and report results for using the proposed methods for parallel MRI reconstruction using two different algorithms in [34]–[36].

III. Estimating WMSETD Using Stein's Principle

Expanding $\mathrm{WMSETD}(\boldsymbol{\lambda})$ and using (1) to write $\mathbf{y}_{\mathrm{true}} = \mathbf{y} - \boldsymbol{\xi}$, we get that

$$\begin{aligned}\mathrm{WMSETD}(\boldsymbol{\lambda}) = M^{-1}\|\mathbf{y}_{\mathrm{true}}\|_{\mathbf{W}}^2 + M^{-1}\|\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\|_{\mathbf{W}}^2 \\ - 2M^{-1}\mathcal{R}\{\mathbf{y}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\} \\ + 2M^{-1}\mathcal{R}\{\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\}\end{aligned} \quad (8)$$

[1]In some special cases such as where $\mathbf{A}$ has full column-rank or when $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ belongs to the range-space of $\mathbf{A}'$, it is possible to estimate $\mathrm{MSE}(\boldsymbol{\lambda})$ [21], [23], [27].

[2]Since (1) and (2) are based on the same noise model, $\mathrm{WMSE}(\boldsymbol{\lambda})$ (6) and $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (7) lead to functionally similar $\mathrm{WSURE}(\boldsymbol{\lambda})$ such as [27, eq. (12)] and (12) in this paper. However, it is more apt to interpret $\mathrm{WSURE}(\boldsymbol{\lambda})$ as an unbiased estimate of $\mathrm{WMSETD}(\boldsymbol{\lambda})$ for practical imaging inverse problems.

[3]Previously [27], we assumed that the observed data followed the discretized linear model (2) for single-coil MRI reconstruction with *retrospective undersampling*, so we focussed on $\mathrm{WMSE}(\boldsymbol{\lambda})$ (6) in [27]. However, since the model in (1) is more realistic than that in (2), we prefer $\mathrm{WMSETD}(\boldsymbol{\lambda})$ over $\mathrm{WMSE}(\boldsymbol{\lambda})$ in this work.

where $\mathcal{R}\{\cdot\}$ stands for real part of a complex-number. Apart from the irrelevant constant $\|\mathbf{y}_{\mathrm{true}}\|^2_{\mathbf{W}}$ that does not depend on $\boldsymbol{\lambda}$, the only inaccessible term is $\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$. In the sequel, we use the principles underlying Stein's result [20] and generalized SURE [21] for estimating this term.

Lemma 1: Let the following be true.

1) $\boldsymbol{\xi} \in \mathbb{C}^M$ in (1) is complex Gaussian with $\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\} = \mathbf{0}$, $\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\boldsymbol{\xi}^\top\} = \mathbf{0}$, and $\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\boldsymbol{\xi}'\} = \boldsymbol{\Omega} \succ \mathbf{0}$, where $\mathsf{E}_{\boldsymbol{\xi}}$ denotes expectation with respect to $\boldsymbol{\xi}$.
2) $\mathbf{u}_{\boldsymbol{\lambda}} : \mathbb{C}^M \rightarrow \mathbb{C}^N$ is *individually analytic* [37] with respect to the real and imaginary parts of its argument (in the weak sense of distributions [38, Ch. 6]).
3) The matrix
$$\boldsymbol{\Gamma} \triangleq \boldsymbol{\Omega}\mathbf{W}\mathbf{A} \in \mathbb{C}^{M\times N} \tag{9}$$
satisfies $\mathsf{E}_{\boldsymbol{\xi}}\{|[\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_m|\} < \infty, m = 1,\dots,M$.

Then, we have that

$$\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\} = \mathsf{E}_{\boldsymbol{\xi}}\left\{\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}\right\} \tag{10}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix and $\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}) \in \mathbb{C}^{N\times M}$ is the Jacobian matrix of (weak) partial derivatives of the components of $\mathbf{u}_{\boldsymbol{\lambda}}$ with respect to the components of $\mathbf{y}$ and is defined via its elements as

$$[\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})]_{nm} \triangleq \frac{1}{2}\left(\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{R}m}} - \iota\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{I}m}}\right). \tag{11}$$

Proof: The proof is a straightforward extension of previous results [20], [21, Th. 1], [27, Lem.1] and is given in Appendix A for completeness. ∎

We now use (10) to show that

$$\mathrm{WSURE}(\boldsymbol{\lambda}) \triangleq M^{-1}\|\mathbf{y} - \mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\|^2_{\mathbf{W}} - M^{-1}\mathrm{tr}\{\boldsymbol{\Omega}\mathbf{W}\} + 2M^{-1}\mathcal{R}\{\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}\} \tag{12}$$

is an unbiased estimate of $\mathrm{WMSETD}(\boldsymbol{\lambda})$.

Theorem 1: Let $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ and $\boldsymbol{\Gamma}$ in (9) satisfy the hypotheses of Lemma 1. Then $\mathrm{WSURE}(\boldsymbol{\lambda})$ (12) is an unbiased estimate of $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (7), i.e., $\mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{WMSETD}(\boldsymbol{\lambda})\} = \mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{WSURE}(\boldsymbol{\lambda})\}$.

Proof: The proof is straightforward and uses Lemma 1 to estimate $\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ in $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (8). ∎

The estimate, $\mathrm{WSURE}(\boldsymbol{\lambda})$ (12), of $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (7) is independent of $\mathbf{y}_{\mathrm{true}}$ and depends only on $\mathbf{y}$, the noise covariance matrix $\boldsymbol{\Omega}$ and $\mathbf{u}_{\boldsymbol{\lambda}}$ via $\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}$. Thus, it is feasible to compute $\mathrm{WSURE}(\boldsymbol{\lambda})$ as a proxy for $\mathrm{WMSETD}(\boldsymbol{\lambda})$ for tuning $\boldsymbol{\lambda}$. In our previous work [27], we analytically evaluated $\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})$ recursively for some iterative reconstruction algorithms for image-deblurring and single-coil undersampled Cartesian MRI reconstruction. Although accurate, such an analytical approach demands tedious mathematical derivations that depend on the specifics of $\mathbf{u}_{\boldsymbol{\lambda}}$ and that must be repeated for different $\mathbf{u}_{\boldsymbol{\lambda}}$ individually on a case-by-case basis.

In this work, we propose a Monte-Carlo scheme for numerically estimating $\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}$ in $\mathrm{WSURE}(\boldsymbol{\lambda})$ (12). The proposed scheme does not require knowledge of the implementation details of $\mathbf{u}_{\boldsymbol{\lambda}}$ as we shall see next; this advantage makes it readily applicable to a wide variety of (weakly differentiable) estimators $\mathbf{u}_{\boldsymbol{\lambda}}$.

IV. Monte-Carlo Estimation

The proposed Monte-Carlo method for tuning $\boldsymbol{\lambda}$ extends our previous result, [32, Th. 2], that focussed on real-valued $\mathbf{u}_{\boldsymbol{\lambda}}$ for denoising applications, to handle complex-valued $\mathbf{u}_{\boldsymbol{\lambda}}$ in (3) with application to imaging inverse problems, especially MRI. Similar to [32, Th. 2], we probe $\mathbf{u}_{\boldsymbol{\lambda}}$ and analyze its response to complex-valued random perturbations in $\mathbf{y}$ to estimate $\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}$.

Theorem 2: Consider the random vector

$$\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon) \triangleq \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y} + \varepsilon\boldsymbol{\Lambda}\mathbf{b}) - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y}) \tag{13}$$

where $\mathbf{b} \in \mathbb{C}^M$ is an i.i.d. random vector independent of $\mathbf{y}$ such that $\mathsf{E}_{\mathbf{b}}\{\mathbf{b}\} = \mathbf{0}$, $\mathsf{E}_{\mathbf{b}}\{\mathbf{b}\mathbf{b}^\top\} = \mathbf{0}$, $\mathsf{E}_{\mathbf{b}}\{\mathbf{b}\mathbf{b}'\} = \mathbf{I}_M$, and $\boldsymbol{\Lambda} \in \mathbb{C}^{M\times M}$ is an invertible deterministic matrix. If $\mathbf{u}_{\boldsymbol{\lambda}}$ admits a second-order Taylor expansion in addition to satisfying the hypotheses in Lemma 1, we have that

$$\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon)\}. \tag{14}$$

Proof: When $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ admits a second-order Taylor expansion, we have that [39]

$$\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon) = \varepsilon\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\mathbf{b} + \varepsilon\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}^*)\boldsymbol{\Lambda}^*\mathbf{b}^* + o(\boldsymbol{\Lambda}\mathbf{b}, \varepsilon) \tag{15}$$

where $o(\boldsymbol{\Lambda}\mathbf{b}, \varepsilon)$ satisfies $\lim_{\varepsilon\to 0}\mathsf{E}_{\mathbf{b}}\{|b_m\, o(\boldsymbol{\Lambda}\mathbf{b}, \varepsilon)|\}/\varepsilon = \mathbf{0}$, for $m = 1,\dots,M$. Then, from (15), we have that

$$\begin{aligned}\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}&\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon)\}\\ &= \mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\mathbf{b}\}\\ &\quad + \mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}^*)\boldsymbol{\Lambda}^*\mathbf{b}^*\}\end{aligned} \tag{16}$$

where the last term in the right-hand side (rhs) of (15) vanishes due to the limit. The second term in the rhs of (16) vanishes since

$$\begin{aligned}&\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}^*)\boldsymbol{\Lambda}^*\mathbf{b}^*\}\\ &\quad= \mathsf{E}_{\mathbf{b}}\{\mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}^*)\boldsymbol{\Lambda}^*\mathbf{b}^*\mathbf{b}'\}\}\\ &\quad= \mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y}^*)\boldsymbol{\Lambda}^*(\mathsf{E}_{\mathbf{b}}\{\mathbf{b}\mathbf{b}^\top\})^*\} = \mathbf{0}\end{aligned} \tag{17}$$

while the first term can be manipulated as

$$\begin{aligned}\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\mathbf{b}\} &= \mathsf{E}_{\mathbf{b}}\{\mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\,\mathbf{b}\mathbf{b}'\}\}\\ &= \mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\,\mathsf{E}_{\mathbf{b}}\{\mathbf{b}\mathbf{b}'\}\}\\ &= \mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\boldsymbol{\Lambda}\}\\ &= \mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}\end{aligned} \tag{18}$$

which is the desired result. ∎

Theorem 2 generalizes [32, Th. 2], to complex-valued problems allowing for a correlation matrix $\boldsymbol{\Lambda}$ in (13) and (14). We briefly discuss the role of $\boldsymbol{\Lambda}$ later in this section and in Section VII. The Monte-Carlo result (14) does not explicitly rely on the functional form of $\mathbf{u}_{\boldsymbol{\lambda}}$ and is equally applicable to both linear and nonlinear $\mathbf{u}_{\boldsymbol{\lambda}}$.

A generic linear reconstruction algorithm has the form

$$\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y}) = \mathbf{H}_{\boldsymbol{\lambda}}\mathbf{y} \tag{19}$$

for some (reconstruction) matrix $\mathbf{H}_{\boldsymbol{\lambda}} \in \mathbb{C}^{N\times M}$ parametrized by $\boldsymbol{\lambda}$. Our Monte-Carlo result (14) further simplifies for linear $\mathbf{u}_{\boldsymbol{\lambda}}$ (19) as shown in the following corollary that extends our previous result [32, Prop.2] to the case of complex-valued $\mathbf{u}_{\boldsymbol{\lambda}}$.

Corollary 1: When $\mathbf{u}_{\boldsymbol{\lambda}}$ is linear, (14) holds without the limit, independent of ε leading to the following identity:

$$\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{H}_{\boldsymbol{\lambda}}\boldsymbol{\Lambda}\mathbf{b}\} = \mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{H}_{\boldsymbol{\lambda}}\}. \tag{20}$$

Proof: For linear $\mathbf{u}_{\boldsymbol{\lambda}}$ (19), the rhs of (14) reduces to $\mathsf{E}_{\mathbf{b}}\{\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\mathbf{H}_{\boldsymbol{\lambda}}\boldsymbol{\Lambda}\mathbf{b}\}$ without $\lim_{\varepsilon\to 0}$, which does not depend on ε. A manipulation similar to that in (18) leads to (20). ■

When $\boldsymbol{\Lambda} = \mathbf{I}_M$, Corollary 1 is a restatement of existing results [40]–[43] for Monte-Carlo estimation of the trace of a matrix and is useful [via $\mathrm{WSURE}(\boldsymbol{\lambda})$] for adjusting $\boldsymbol{\lambda}$ of linear MRI reconstruction algorithms [32], [40], e.g., conjugate phase reconstruction with density compensation [2], [44] where $\boldsymbol{\lambda}$ could describe some parametrization of the density compensation weights or such as those encountered when using Tikhonov-type quadratic regularizers [32], [40] where $\boldsymbol{\lambda}$ could denote regularization parameters.

For MRI reconstruction from undersampled data, it is preferable to use nonquadratic regularizers to better reduce aliasing artifacts and noise in the reconstructed image [3], [5]. The reconstruction process associated with a nonquadratic regularizer is nonlinear, so henceforth we concentrate on nonlinear $\mathbf{u}_{\boldsymbol{\lambda}}$.

In practice, for nonlinear $\mathbf{u}_{\boldsymbol{\lambda}}$, the limit in (14) cannot be applied analytically except in some special cases where $\mathbf{u}_{\boldsymbol{\lambda}}$ is analytically tractable. So we make an approximation to (14) by dropping the limit and the $\mathsf{E}_{\mathbf{b}}\{\cdot\}$ operations similar to [32, eq. (17)], and use

$$\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\} \approx \varepsilon^{-1}\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon) \tag{21}$$

for a sufficiently small ε and one realization of a complex-valued random vector $\mathbf{b}$ satisfying the hypotheses of Theorem 2. The choice of ε represents a trade-off: for too small an ε-value, $\mathbf{u}_{\boldsymbol{\lambda}}$ may be insensitive to the perturbation $\varepsilon\boldsymbol{\Lambda}\mathbf{b}$ in $\mathbf{y} + \varepsilon\boldsymbol{\Lambda}\mathbf{b}$ due to finite numerical precision of digital computers, so the Monte-Carlo estimate (21) could be unstable, i.e., it could have large variance. On the other hand, the approximation (21) may be inaccurate for large ε-values for nonlinear $\mathbf{u}_{\boldsymbol{\lambda}}$.

The robustness of (21) to the choice of ε depends on several factors such as the magnitude of the elements of $\boldsymbol{\Gamma}$ (9), the energy of $\boldsymbol{\Lambda}\mathbf{b}$, $\mathsf{E}_{\mathbf{b}}\{\|\boldsymbol{\Lambda}\mathbf{b}\|_2^2\}$, relative to that of $\mathbf{y}$, $\mathsf{E}_{\boldsymbol{\xi}}\{\|\mathbf{y}\|_2^2\}$, numerical precision of the variables used in the implementation and the sensitivity of $\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})$ to changes in $\mathbf{y}$; the approximation (21) must thus be validated for a given data model (1) and (2) and a reconstruction algorithm (3) individually. The matrix $\boldsymbol{\Lambda}$ in (21) may be chosen so as to scale the elements of $\boldsymbol{\Lambda}\mathbf{b}$ relative to those of $\mathbf{y}$, essentially allowing different amounts of perturbation for different elements of $\mathbf{y}$. This may be beneficial in some applications such as MRI where the elements of $\mathbf{y}$ span several orders of magnitude and relatively scaling the perturbation can help maintain the accuracy of the approximation (21) for a fixed, sufficiently small ε for varying $\mathbf{y}$. Although ε is a user-provided parameter, we show in Section VI-B that the choice of ε spans several decades without significantly affecting the results, so the proposed MCSURE method can be applied without having to repeatedly adjust ε.

Using (21), we thus require only two evaluations of $\mathbf{u}_{\boldsymbol{\lambda}}$ for a given $\mathbf{y}$ and $\boldsymbol{\lambda}$, i.e., the response of $\mathbf{u}_{\boldsymbol{\lambda}}$ to $\mathbf{y}$ and $\mathbf{y} + \varepsilon\boldsymbol{\Lambda}\mathbf{b}$ for estimating $\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}$ for a given $\boldsymbol{\lambda}$. Our approach does not need the knowledge of the structure of $\mathbf{u}_{\boldsymbol{\lambda}}$, so (21) is very flexible in its applicability. This is unlike the analytical development in our earlier work [27] that varied with the choice of $\mathbf{u}_{\boldsymbol{\lambda}}$ and also required storage and computation equivalent to three evaluations of $\mathbf{u}_{\boldsymbol{\lambda}}$ for a given $\boldsymbol{\lambda}$ as discussed in [27, Sec. VI-C].

Theorem 2 is somewhat restrictive in its applicability since it is based on a Taylor expansion of $\mathbf{u}_{\boldsymbol{\lambda}}$. In practice, $\mathbf{u}_{\boldsymbol{\lambda}}$ may involve *weakly differentiable* operators that do not admit (15). A typical instance is when ℓ_1-type (including total variation) regularizers are used for reconstruction; $\mathbf{u}_{\boldsymbol{\lambda}}$ for these regularizers would involve (for certain implementations) a nonsmooth shrinkage operator that satisfies Lemma 1 but not (15). In such cases, it is possible to extend the scope of Theorem 2 to *weakly differentiable* functions similar to that documented in [32, Th. 2]. However, this would require tedious derivations using measure theory and the theory of distributions [38, Ch. 6], and is beyond the scope of this paper. Instead, we investigate using (21) for $\mathbf{u}_{\boldsymbol{\lambda}}$ corresponding to ℓ_1-type regularizers based on empirical validation with numerical experiments both in the paper (see Sections VI-C and VI-D) and in a supplementary material.[4]

Finally, our Monte-Carlo result (14) precludes iterative/noniterative estimators that involve nonweakly-differentiable operators, e.g., the hard-thresholding operator [32, Sec.V-B], [45]; such operators do not satisfy the conditions of Lemma 1 and are not suitable for use with $\mathrm{WSURE}(\boldsymbol{\lambda})$.

V. Single-Coil Non-Cartesian MRI Reconstruction

The theoretical development so far has been general both in terms of the data model (1) and (2) and the reconstruction algorithm (3) due to the Monte-Carlo nature of our approach for estimating $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (7). However, numerical validation of our approach needs to be done on a case-by-case basis for different applications and reconstruction algorithms. For illustration, we henceforth focus on *single-coil non-Cartesian* MRI ignoring field-inhomogeneity and relaxation effects as an extension to our previous work [27] on *single-coil Cartesian* MRI. In this case, a good model for noise in (1) is $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M)$, so that

$$\boldsymbol{\Gamma} = \sigma^2\mathbf{W}\mathbf{A} \tag{22}$$

in (9). For the purpose of reconstruction (3), we use the discretized linear model in (2). Unlike for Cartesian MRI [27], $\mathbf{A}$ is not a simple undersampled DFT matrix for non-Cartesian MRI. But for a suitable discretization, $\mathbf{A}$ in (2) can be implemented using nonuniform FFT (NUFFT) [46] for single-coil non-Cartesian MRI. We then formulate MRI reconstruction in (3) as

$$\widehat{\mathbf{x}} = \mathbf{u}_{\lambda}(\mathbf{y}) \triangleq \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\Psi(\mathbf{R}\mathbf{x}) \tag{23}$$

[4]The supplementary material is available at http://ieeexplore.ieee.org.

where $\widehat{\mathbf{x}} \in \mathbb{C}^N$ is the reconstructed image, $\boldsymbol{\lambda} \triangleq \lambda > 0$ is the scalar regularization parameter, Ψ is a (possibly nonsmooth) convex regularizer, and $\mathbf{R} \triangleq [\mathbf{R}_1^\top \cdots \mathbf{R}_P^\top]^\top \in \mathbb{R}^{PN \times N}$ is a regularization operator, e.g., finite differences.

We used the split-Bregman (SB) scheme [47] for $\mathbf{u}_\lambda$ in (23). At each iteration, the SB algorithm requires (among other simple update steps) "inverting" a matrix $\mathbf{B} \triangleq \mathbf{A}'\mathbf{A} + \mu\mathbf{R}'\mathbf{R}$ [27, eq. (32)], for some penalty parameter[5] $\mu > 0$ [27], [47]. For Cartesian MRI, this step can be achieved via FFTs [47, Sec. 5.2], [27, Sec.IV-F]. For non-Cartesian MRI however, $\mathbf{B}$ is block-Toeplitz with Toeplitz-blocks [49] and cannot be inverted noniteratively for large image sizes, i.e., for large N, so we used a preconditioned conjugate gradient (PCG) solver with a circulant preconditioner [48] that approximately matched $\mathbf{B}^{-1}$. We implemented $\mathbf{A}'\mathbf{A}$ using the "embedding-toeplitz-in-circulant" trick, i.e., $\mathbf{A}'\mathbf{A} = \mathbf{Z}'\mathbf{Q}\mathbf{Z}$, where $\mathbf{Z}$ is a $PN \times N$ zero-padding matrix and $\mathbf{Q}$ is an appropriate $PN \times PN$ circulant matrix [50] ($P = 2$ for 1-D and $P = 4$ for 2-D images). In all our experiments, we ran five PCG iterations for this step [27, eq. (32)], and 100 iterations of the SB algorithm. These numbers ensured that the SB algorithm nearly converged in the sense that the normalized "distance" between two successive iterates $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2 / \|\mathbf{x}^{(k-1)}\|_2$ was close to zero for a large range of λ-values.

VI. Experiments

A. Setup

In all our experiments, we focussed on selecting λ in (23) by minimizing the proposed Monte-Carlo estimate, $\mathrm{WSURE}(\lambda)$ (12), of $\mathrm{WMSETD}(\lambda)$ (7). We investigated two versions of $\mathrm{WMSETD}(\lambda)$ corresponding to $\mathbf{W} = \mathbf{I}_M$ and

$$\mathbf{W} = \mathbf{W}_\mathbf{D} \triangleq \alpha\mathbf{I}_M + \mathbf{D} \tag{24}$$

where $\mathbf{D} \succeq 0$ is a diagonal matrix of suitable density compensation weights [2] for non-Cartesian trajectories and $\alpha > 0$ is chosen so that $\mathbf{W}$ has a user-provided condition number $\kappa(\mathbf{W})$; we set α such that $\kappa(\mathbf{W}) = 100$. For $\mathbf{W} = \mathbf{I}_M$, $\mathrm{WMSETD}(\lambda)$ can be interpreted as the predicted squared-error (similar to Predicted-MSE [11], [27]) that uniformly weighs the error at all sample locations in k-space. For $\mathbf{W}$ in (24), $\mathrm{WMSETD}(\lambda)$ favors errors at certain sample locations in k-space more than others depending upon $\mathbf{D}$; typically, for non-Cartesian trajectories, the central k-space is more densely sampled than outer k-space, so $\mathbf{D}$ is designed to provide higher weighting for outer k-space samples than around central k-space [2].

We implemented the SB algorithm and conducted all experiments in Matlab using double-precision variables. We used the conjugate phase (CP) reconstruction with suitable density compensation [2] (described later), $\mathbf{A}'\mathbf{D}\mathbf{y}$, to initialize the SB algorithm in all experiments.

In the proposed Monte-Carlo estimation scheme (21), we used $\mathbf{b} = \mathbf{b}_\pm \triangleq (\mathbf{b}_\mathcal{R} + \imath\mathbf{b}_\mathcal{I})/\sqrt{2}$ where $\mathbf{b}_\mathcal{R}$, $\mathbf{b}_\mathcal{I}$ are independent binary random vectors[6] whose elements are i.i.d. and assume either $+1$ or -1 with equal probability. It is easily verified that $\mathbf{b}_\pm$ satisfies the hypotheses of Theorem 2. For simplicity, we used $\boldsymbol{\Lambda} = \mathbf{I}_M$ in (21) throughout. To avoid repeated computation of $\boldsymbol{\Gamma}'\mathbf{b}$ in (21) for use in (12) with several λ-values, we precomputed and stored $\mathbf{c} \triangleq \boldsymbol{\Gamma}'\mathbf{b}$ and used $\mathbf{c}'$ in (21). In our simulations, we assumed that the noise variance σ^2 was known for computing $\mathrm{WSURE}(\lambda)$ via (12) and (22), while for experiments with real MR data, we used an estimate computed by empirical sample-variance from outer k-space data samples as those are mostly dominated by noise. We compared λ-selection using the proposed $\mathrm{WSURE}(\lambda)$ (12) against that using generalized cross-validation for nonlinear algorithms (NGCV) [18], [27, eq. (7)]

$$\mathrm{NGCV}(\lambda) \triangleq \frac{M^{-1}\|\mathbf{y} - \mathbf{A}\,\mathbf{u}_\lambda(\mathbf{y})\|_2^2}{(1 - M^{-1}\mathcal{R}\{\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_\lambda}(\mathbf{y})\}\})^2} \tag{25}$$

where we used the Monte-Carlo estimation procedure (21) for $\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_\lambda}(\mathbf{y})\}$ in the denominator of $\mathrm{NGCV}(\lambda)$. Thus, $\mathrm{NGCV}(\lambda)$ has the same computation cost as the proposed $\mathrm{WSURE}(\lambda)$.

We experimented with three types of regularizers in (23): a smooth convex regularizer with Fair potential (FP) [51], [52] given by

$$\Psi_{\mathrm{FP}}(\mathbf{R}\mathbf{x}) \triangleq \sum_{r=1}^{PN} \Phi_{\mathrm{FP}}(|\,[\mathbf{R}\mathbf{x}]_r\,|) \tag{26}$$

where $\Phi_{\mathrm{FP}}(x) \triangleq x/\delta - \log(1 + x/\delta)$, $\delta > 0$, total variation (TV) regularizer

$$\Psi_{\mathrm{TV}}(\mathbf{R}\mathbf{x}) \triangleq \sum_{r=1}^{N} \sqrt{\sum_{p=1}^{P} |\,[\mathbf{R}_p\mathbf{x}]_r\,|^2} \tag{27}$$

and an ℓ_1-regularizer

$$\Psi_{\ell_1} \triangleq \sum_{r=1}^{PN} |\,[\mathbf{R}\mathbf{x}]_r\,|. \tag{28}$$

We used finite differences for $\mathbf{R}$ in (26)–(28) with $P = 4$ (horizontal, vertical, and two diagonal) directions in all experiments.

It is possible to verify that the SB algorithm for $\mathbf{u}_\lambda$ satisfies the hypotheses of Theorem 2 for Ψ_{FP} (26) because it is differentiable everywhere. However, Theorem 2 is not directly applicable when Ψ_{TV} or Ψ_{ℓ_1} are involved in (23) as the corresponding $\mathbf{u}_\lambda$ may not satisfy the hypotheses of Theorem 2. As discussed at the end of Section IV, we demonstrate using numerical experiments in Sections VI-C–VI-D (and in the supplementary material) that the proposed Monte-Carlo approach can be used for estimating $\mathrm{WSURE}(\lambda)$ for Ψ_{TV} and Ψ_{ℓ_1} in (23). In all experiments, we minimized $\mathrm{WSURE}(\lambda)$ and $\mathrm{NGCV}(\lambda)$ as a function of λ.

B. Radial MRI Simulation

We used the analytical Shepp-Logan phantom [33] to simulate noisy data $\mathbf{y}$ of 40 dB SNR on a radial trajectory with 96 spokes each containing 512 samples (reduction factor ≈ 8). We used the approach in [53], [54] for selecting the density compensation weights $\mathbf{D}$ (24). We set $\Psi = \Psi_{\mathrm{FP}}$ (26) in (23) with $\delta = M^{-1}\|\mathbf{y}\|_2^2 \times 10^{-4}$.

[5]We chose $\mu = \mu_{\min} \times 10^{-2}$ in all experiments, where $\mu_{\min}$ minimized the condition number of $\tilde{\mathbf{A}}'\tilde{\mathbf{A}} + \mu\mathbf{R}'\mathbf{R}$ for a given $\mathbf{R}$, where $\tilde{\mathbf{A}}'\tilde{\mathbf{A}}$ is a circulant approximation to $\mathbf{A}'\mathbf{A}$ [48].

[6]Another choice is complex Gaussian $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$.

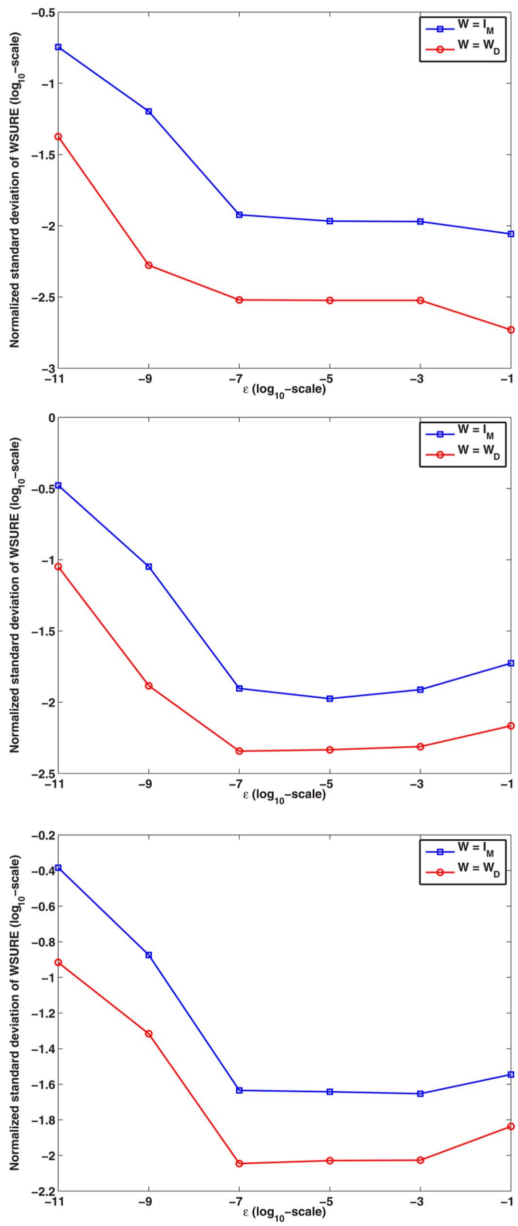


Fig. 1. Plots of standard deviation of $\mathrm{WSURE}(\lambda)$ normalized by $\mathrm{WMSETD}(\lambda)$ as a function of ε in (21) for (top) $\lambda = \lambda_{\mathrm{opt}}/10$, (middle) $\lambda = \lambda_{\mathrm{opt}}$, and (bottom) $\lambda = 10\lambda_{\mathrm{opt}}$, where λ_{opt} is the MSE-optimal value of the regularization parameter. The curves correspond to the experiment in Section VI-B1 where $\mathrm{WSURE}(\lambda)$ was obtained by averaging (21) over 25 realizations of $\mathbf{b}_{\pm}$. As expected, the variance rapidly increases for smaller ε.

1) Variance of WSURE: To analyze the accuracy of (21), we reconstructed 512×512 images of the Shepp–Logan phantom for three different values of λ, and correspondingly computed the standard deviation of Monte-Carlo $\mathrm{WSURE}(\lambda)$ by averaging it over 25 realizations of $\mathbf{b}_{\pm}$ for different ε. Fig. 1 plots the standard deviation of Monte-Carlo $\mathrm{WSURE}(\lambda)$ normalized by $\mathrm{WMSETD}(\lambda)$ as a function of ε. The plots indicate that $\varepsilon < 10^{-7}$ consistently leads to increased variance. Moreover, the variance is approximately constant for $\varepsilon \in [10^{-7}, 10^{-3}]$ indicating the robustness of the approximation in (21). We present similar results for varying SNR of data in the supplementary material.

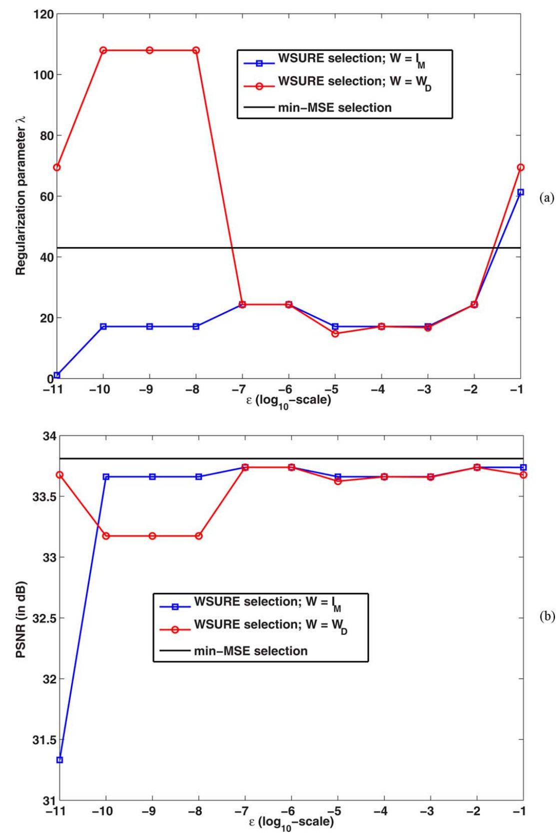


Fig. 2. Plots of (a) regularization parameter λ, and (b) $\mathrm{PSNR}(\lambda)$ as functions of ε for λ selected to minimize $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ and $\mathbf{W}_{\mathbf{D}}$ in (24) and $\mathrm{MSE}(\lambda)$ for the experiment described in Section VI-B2.

2) Selection of λ for Different ε: We used only one realization of $\mathbf{b}_{\pm}$ in (21) for computing $\mathrm{WSURE}(\lambda)$ (12). We varied ε, minimized $\mathrm{MSE}(\lambda)$ and $\mathrm{WSURE}(\lambda)$ with respect to λ for each ε. Fig. 2(a) plots the resulting λ-values, while Fig. 2(b) plots peak-SNR (PSNR) defined as

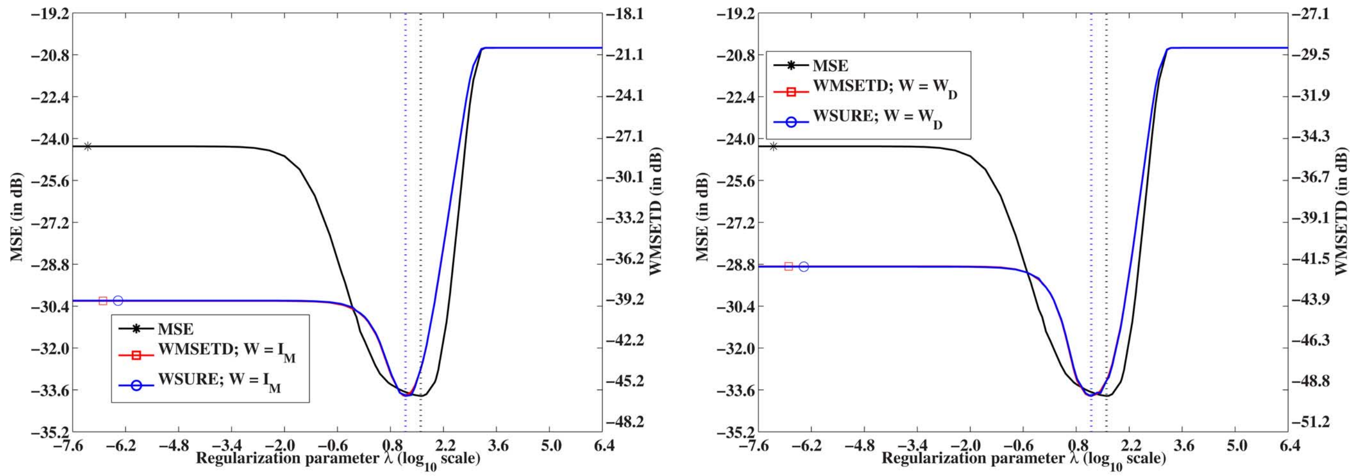
$$\mathrm{PSNR}(\lambda) \triangleq 10\log_{10}\left[\frac{\max_n \{|[\mathbf{x}_{\mathrm{true}}]_n|^2\}}{\mathrm{MSE}(\lambda)}\right]$$

as functions of ε for the various λ-selections. For $\varepsilon \in [10^{-7}, 10^{-2}]$, $\mathrm{WSURE}(\lambda)$ based λ-selection and corresponding $\mathrm{PSNR}(\lambda)$ are close to those of minimum $\mathrm{MSE}(\lambda)$ selection. We present similar results for varying SNR of data and the TV regularizer in the supplementary material.

Based on Figs. 1 and 2 and corresponding results in the supplementary material, a suitable choice of ε appears to be in the range $[10^{-7}, 10^{-2}]$. However, from our experience, it is beneficial to be conservative with ε, so we recommend choosing $\varepsilon \in [10^{-5}, 10^{-2}]$.

In the remaining experiments, we set $\varepsilon = 10^{-4}$ and used only one realization of $\mathbf{b}_{\pm}$ in (21) for computing $\mathrm{WSURE}(\lambda)$ (12) and $\mathrm{NGCV}(\lambda)$ (25).

3) Trends of WMSETD (λ) and WSURE (λ): We reconstructed 512×512 images, and computed $\mathrm{WSURE}(\lambda)$, the oracles $\mathrm{WMSETD}(\lambda)$, and $\mathrm{MSE}(\lambda)$, for a range of λ-values. Fig. 3 plots $\mathrm{WSURE}(\lambda)$, $\mathrm{WMSETD}(\lambda)$, and $\mathrm{MSE}(\lambda)$ as a function of λ. $\mathrm{WSURE}(\lambda)$ captures the trend of $\mathrm{WMSETD}(\lambda)$



(a) (b)

Fig. 3. Simulation with the analytical Shepp–Logan phantom (Section VI-B3). Plots of $\mathrm{MSE}(\lambda)$, $\mathrm{WMSETD}(\lambda)$, $\mathrm{WSURE}(\lambda)$ versus λ for $\mathbf{W} = \mathbf{I}_M$ (left) and $\mathbf{W_D}$ in (24) (right). Vertical dashed lines indicate minima of various curves. $\mathrm{WSURE}(\lambda)$ captures the trend of $\mathrm{WMSETD}(\lambda)$ in both plots and their minima are close to that of the true $\mathrm{MSE}(\lambda)$.

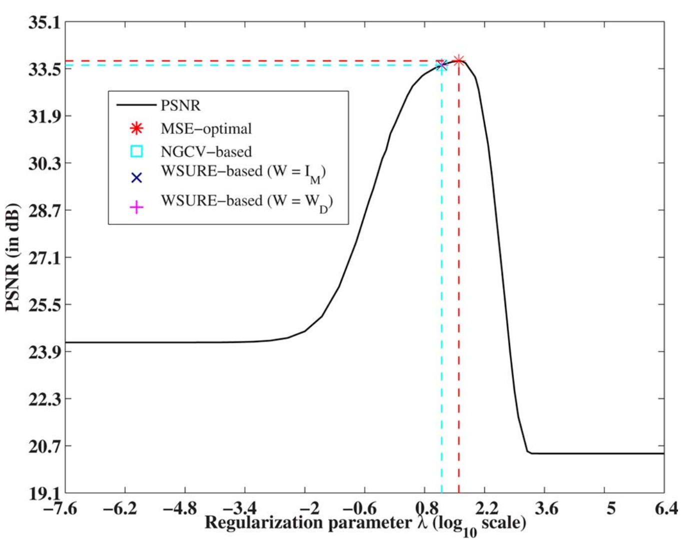


Fig. 4. Simulation with the analytical Shepp–Logan phantom (Section VI-B3). Plot of $\mathrm{PSNR}(\lambda)$ versus λ. Vertical dashed lines indicate λ-selections made by various methods. $\mathrm{WSURE}(\lambda)$ and $\mathrm{NGCV}(\lambda)$ lead to near-PSNR-optimal reconstructions.

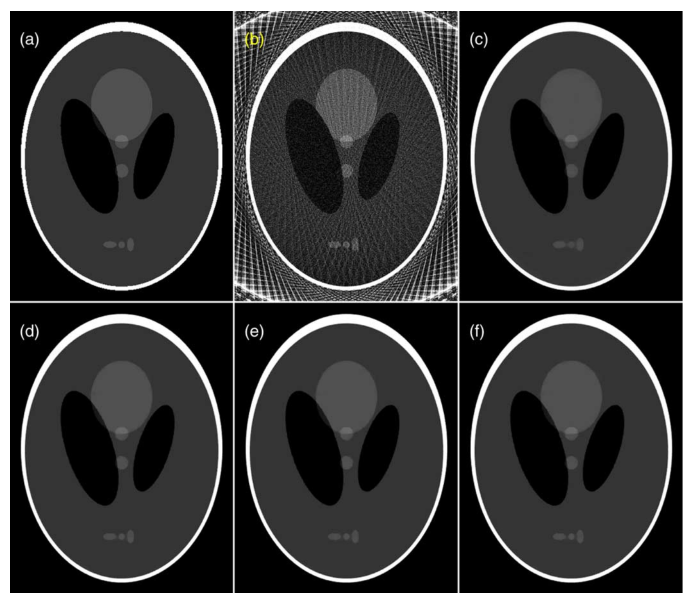


Fig. 5. Simulation with the analytical Shepp-Logan phantom (Section VI-B3). (a) Discretized noise-free 512×512 phantom; (b) CP reconstruction $(\mathrm{PSNR} = 16.57\ \mathrm{dB})$ has prominent streak artifacts and noise; Images reconstructed using Ψ_{FP} regularizer with λ selected to minimize (c) true $\mathrm{MSE}(\lambda)(\lambda = 4.3 \times 10^{-7}; \mathrm{PSNR} = 33.81\ \mathrm{dB})$; (d) $\mathrm{NGCV}(\lambda)(\lambda = 1.7 \times 10^{-7}; \mathrm{PSNR} = 33.66\ \mathrm{dB})$; (e) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M(\lambda = 1.7 \times 10^{-7}; \mathrm{PSNR} = 33.66\ \mathrm{dB})$; (f) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W_D}$ in (24) $(\lambda = 1.7 \times 10^{-7}; \mathrm{PSNR} = 33.66\ \mathrm{dB})$. In this experiment, WSURE and NGCV lead to the same λ-selections, see Fig. 4, thus resulting in similar visual quality comparable to the true $\mathrm{MSE}(\lambda)$-based reconstruction in (c).

over the entire range of λ indicating the accuracy of the proposed Monte-Carlo scheme with a single realization of $\mathbf{b}_{\pm}$. Moreover, the minima of $\mathrm{WMSETD}(\lambda)$ and $\mathrm{WSURE}(\lambda)$ are all close to that of the true $\mathrm{MSE}(\lambda)$ indicating their reliability in selecting λ. In Fig. 4, we plot $\mathrm{PSNR}(\lambda)$ for a range of λ-values indicating the λ-selections made by $\mathrm{NGCV}(\lambda)$ and $\mathrm{WSURE}(\lambda)$. Both $\mathrm{NGCV}(\lambda)$ and $\mathrm{WSURE}(\lambda)$ led to the same λ-value close to the MSE-optimal one in this experiment. Fig. 5 presents 512×512 images reconstructed using λ-values that minimized $\mathrm{NGCV}(\lambda)$ and $\mathrm{WSURE}(\lambda)$. As expected, the respective reconstructed images, Fig. 5(d) and (f), closely resemble that obtained using the true minimum-MSE-λ in Fig. 5(c). Finally, all the regularized reconstructed images, Fig. 5(c) and (f), have almost no radial-artifacts and display improved quality over CP reconstruction [Fig. 5(b)].

4) Varying Noise Level: We repeated the radial MRI simulation with varying levels of noise in the simulated data. We tabulate PSNR of reconstructed images obtained by minimizing $\mathrm{WSURE}(\lambda)$ and $\mathrm{NGCV}(\lambda)$ in Table I. $\mathrm{WSURE}(\lambda)$ was able to provide near-MSE-optimal λ-selections as indicated by the PSNR-values in Table I. NGCV also provided similar λ-selections in this experiment.

5) Varying Reduction Factor: We repeated the radial MRI simulation for varying number of spokes of the radial trajectory

TABLE I
EXPERIMENT IN SECTION VI-B4: PSNR OF IMAGES RECONSTRUCTED USING $\Psi_{\rm FP}$ WITH λ OPTIMIZED BY VARIOUS METHODS FOR DATA WITH VARYING SNR

SNR (dB)	PSNR (dB) MSE(λ)	NGCV(λ)	WSURE(λ) $\mathbf{W}=\mathbf{I}_M$	WSURE(λ) $\mathbf{W}=\mathbf{W_D}$
20	28.60	28.60	28.60	28.60
30	32.26	32.26	32.26	32.26
40	33.81	33.66	33.66	33.66

TABLE II
EXPERIMENT IN SECTION VI-B5: PSNR OF IMAGES RECONSTRUCTED USING $\Psi_{\rm TV}$ WITH λ OPTIMIZED BY VARIOUS METHODS FOR DATA WITH VARYING NUMBER OF SAMPLES (REDUCTION FACTORS)

Reduction Factor	PSNR (dB) MSE(λ)	NGCV(λ)	WSURE(λ) $\mathbf{W}=\mathbf{I}_M$	WSURE(λ) $\mathbf{W}=\mathbf{W_D}$
5	28.41	28.37	28.34	28.34
4	28.58	28.54	28.54	28.51
3	28.81	28.81	28.81	28.78
2	28.98	28.94	28.98	28.94

corresponding to reduction factors of 2, 3, 4, and 5 and for fixed data-SNR of 40 dB. We tabulate $\rm PSNR$ of reconstructed images obtained by minimizing $\rm WSURE(\lambda)$ and $\rm NGCV(\lambda)$ for $\Psi_{\rm TV}$ in Table II. $\rm WSURE(\lambda)$ was able to provide near-MSE-optimal λ-selection as indicated by the PSNR-values in Table II. $\rm NGCV$ also provides similar λ-selections. This experiment illustrates that $\rm WMSETD(\lambda)$ [via $\rm WSURE(\lambda)$] is a reasonable metric for optimizing λ for agreeable reduction factors for single-coil non-Cartesian MRI reconstruction.

C. GE Phantom MRI Scan

We scanned a GE resolution phantom using a 3T GE scanner with the following scan setting: gradient-echo sequence, $T_R = 300$ ms, $T_E \approx 2$ ms, $\rm FOV = 15$ cm, flip angle $= 40°$, slice thickness $= 5$ mm. We used a 2-D variable density (VD) spiral k-space trajectory[7] with 120 leaves each containing 841 samples. The readout duration per leaf was 3.3 ms, which is sufficiently short to make the assumption that any distortion due to field-inhomogeneity is negligible. We designed the VD spiral so that the central k-space was over-sampled by a factor of two and achieved Nyquist sampling at the periphery. We acquired three independent 2-D data-sets using the same scan-setting and averaged them to obtain a relatively less-noisy data-set. We used $\mathbf{D} = \mathrm{diag}\{\mathbf{d}\}$ in CP reconstruction $\mathbf{A}'\mathbf{D}\mathbf{y}$, where the lth element $[\mathbf{d}]_l = |k_{1l} + \imath k_{2l}|$ with k_{1l} and k_{2l} indexing the k-space sample locations in 2-D. Then, we reconstructed a 256×256 reference image, $\mathbf{x}_{\rm ref}$ in Fig. 6(a), by running the SB algorithm on this data-set using (23) with $\Psi_{\rm TV}$ and $\lambda \approx 0$ (such that $\lambda \ll \|\mathbf{y}\|^2$) in (23).

Next, we simulated undersampling of one of the three data-sets by retaining only 60 equally spaced interleaves (reduction factor $= 2$) and reconstructed 256×256 images with $\Psi_{\rm TV}$ in (23) by minimizing $\rm NGCV(\lambda)$ and $\rm WSURE(\lambda)$. The corresponding reconstructed images [Fig. 6(c)–(e)] are devoid of spiral artifacts present in CP reconstruction [Fig. 6(b)] and closely resemble $\mathbf{x}_{\rm ref}$ [Fig. 6(a)] in this experiment. These results also illustrate the reliability of the proposed Monte-Carlo scheme (21) employed in $\rm WSURE(\lambda)$ (12) and $\rm NGCV(\lambda)$ (25) for optimizing λ for $\Psi_{\rm TV}$.

[7]An illustration of the VD spiral trajectory used in this experiment is presented in the supplementary material.

D. In Vivo Human Brain Imaging

We acquired three independent 3-D VD stack-of-spiral data-sets (with the same 2-D VD spiral trajectory described in Section VI-C) of a live human brain using a 3T GE scanner with the following scan setting: spoiled gradient-echo sequence, $T_R \approx 18.5$ ms, $T_E \approx 2$ ms, $\rm FOV = 25$ cm, flip angle $= 15°$, slice thickness $= 5$ mm, number of slices $= 24$. We averaged these three data-sets and reconstructed a single 256×256 2-D reference image (corresponding to slice 14), $\mathbf{x}_{\rm ref}$ in Fig. 7(a), by running the SB algorithm with Ψ_{ℓ_1} and $\lambda \approx 0$ (such that $\lambda \ll \|\mathbf{y}\|^2$) in (23).

We again undersampled one of the three data-sets (corresponding to Slice 14) with a reduction factor of 2 and reconstructed 256×256 2-D images with Ψ_{ℓ_1} in (23) by minimizing $\rm NGCV(\lambda)$ and $\rm WSURE(\lambda)$. In this experiment, $\rm NGCV$ yielded an over-smoothed result [Fig. 7(c)] that lacks fine details in $\mathbf{x}_{\rm ref}$ [Fig. 7(a)]. However, $\rm WSURE(\lambda)$ led to images that exhibit reasonably better quality than CP reconstruction [Fig. 7(b)] and the $\rm NGCV$-result [Fig. 7(c)] and closely resemble $\mathbf{x}_{\rm xref}$. These results indicate the robustness of the proposed Monte-Carlo $\rm WSURE(\lambda)$ for λ-selection and also its applicability for Ψ_{ℓ_1} in (23). We obtained similar promising results (included in the supplementary material) for reconstructing other slices of this 3-D volume.

VII. DISCUSSION

As with other parameter tuning methods such as the discrepancy principle, L-curve, and generalized cross-validation, the proposed Monte-Carlo $\rm WSURE$-method requires multiple evaluations of the reconstruction algorithm $\mathbf{u}_{\boldsymbol\lambda}$ for optimizing $\boldsymbol\lambda$. For the purpose of illustration, we optimized $\boldsymbol\lambda = \lambda$ by searching over a range of scalar λ-values in our experiments. In practice, derivative-free optimization schemes can be used, e.g., golden-section search for optimizing the scalar λ or the Powell method [55] for optimizing the vector $\boldsymbol\lambda$.

$\rm WSURE(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ and $\mathbf{W_D}$ (24) led to similar λ-selections in all our experiments both in the paper and in supplementary material. This is probably because there is only one degree of freedom, in terms of the scalar λ, in minimizing $\rm WSURE(\lambda)$. However, minimizing $\mathrm{WSURE}(\boldsymbol\lambda)$ with respect to the vector $\boldsymbol\lambda$ may lead to different parameter selections depending upon whether $\mathbf{W} = \mathbf{I}_M$ or $\mathbf{W_D}$ (24) in $\mathrm{WMSETD}(\boldsymbol\lambda)$ (7) and $\mathrm{WSURE}(\boldsymbol\lambda)$ (12). As an illustration, we repeated the experiment in Section VI-D, but used $\Psi_{\rm FP}$ (26) and optimized λ and δ of $\Psi_{\rm FP}$ jointly by exhaustive search. Optimizing $\mathrm{WSURE}(\lambda, \delta)$ with $\mathbf{W} = \mathbf{I}_M$ led to $(\lambda, \delta) = (0.36, 0.31) \times 10^{-7}$, while $\mathrm{WSURE}(\lambda, \delta)$ with $\mathbf{W} = \mathbf{W_D}$ yielded $(\lambda, \delta) = (10, 6.7) \times 10^{-7}$. While (λ, δ)-values are different in each case, the images reconstructed with these selections [Fig. 8] appear visually similar. This is probably because the ratio λ/δ that appears in $\lambda\Psi_{\rm FP}$ (23), (26) is approximately the same for these selections.

Methods proposed in this paper can also tackle $\mathrm{WSURE}(\boldsymbol\lambda)$ with arbitrary measurement-domain symmetric positive

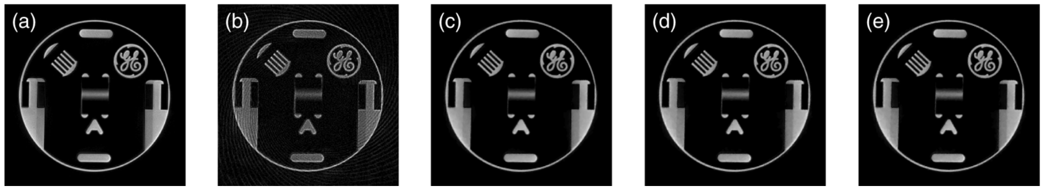


Fig. 6. Experiment with real GE phantom data (Section VI-C). (a) Very mildly Ψ_{TV}-regularized 256 × 256 reference reconstruction from "fully-sampled" data averaged over three acquisitions; (b) CP reconstruction (from 2× undersampled data from a single acquisition) is strewn with spiral artifacts; Images reconstructed from 2× undersampled data (from a single acquisition) using Ψ_{TV}-regularizer with λ selected to minimize (c) $\mathrm{NGCV}(\lambda)$ $(\lambda = 53)$; (d) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ $(\lambda = 37)$; (e) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W_D}$ in (24) $(\lambda = 37)$. The λ-value selected by NGCV is slightly higher than those selected by WSURE. The resulting image (e) is thus slightly over smoothed, although the over smoothing is not visually apparent due to the piece-wise constant nature of the GE phantom. Moreover, some fine details present in (a) are lost in (c)–(e) owing both to undersampling and regularization.

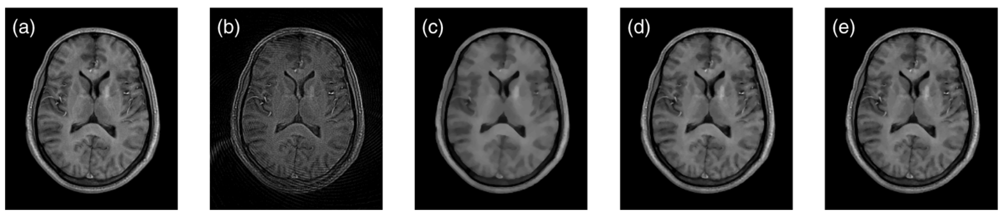


Fig. 7. Experiment with real *in vivo* human head data (Section VI-D); Slice 14. (a) Very mildly Ψ_{ℓ_1}-regularized 256 × 256 reference reconstruction from "fully-sampled" data averaged over three acquisitions; (b) CP reconstruction (from 2× undersampled data from a single acquisition) is strewn with spiral artifacts; Images reconstructed from 2× undersampled data (from a single acquisition) using Ψ_{ℓ_1}-regularizer with λ selected to minimize (c) $\mathrm{NGCV}(\lambda)$ $(\lambda = 3)$; (d) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ $(\lambda = 0.6)$; (e) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W_D}$ in (24) $(\lambda = 0.3)$. In this experiment, $\mathrm{NGCV}(\lambda)$ resulted in a noticeably over-smoothed image due to a correspondingly higher value of λ, while $\mathrm{WSURE}(\lambda)$ still yielded results comparable to the reference (a). Some fine details in (a) are lost in (d), (e) that also contain minor residual spiral artifacts; these can be attributed to undersampling of k-space data.

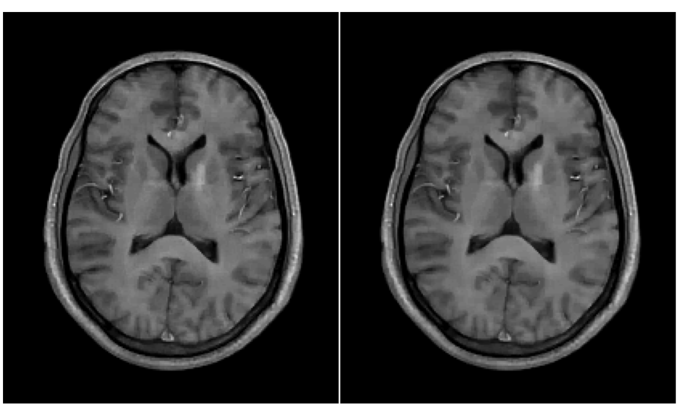


Fig. 8. Experiment with real *in vivo* human head data (Section VI-D); Slice 14. Images were reconstructed using Ψ_{FP} (26) with λ and δ chosen to minimize $\mathrm{WSURE}(\lambda, \delta)$. Left image corresponds to $\mathbf{W} = \mathbf{I}_M$, $\lambda = 0.36 \times 10^{-7}$, $\delta = 0.31 \times 10^{-7}$. Right image corresponds to $\mathbf{W} = \mathbf{W_D}$, $\lambda = 10 \times 10^{-7}$, $\delta = 6.7 \times 10^{-7}$. Although the parameter selections are different, the resulting image quality is similar in both cases and is comparable to Fig. 7(d) and (e).

semi-definite weighting matrices $\mathbf{W} \succeq \mathbf{0}$, e.g., a nondiagonal matrix such as that encountered in Projected-$\mathrm{MSE}(\boldsymbol{\lambda})$ [27, Sec. III-B] or a diagonal matrix with zeros and ones that corresponds to specifying a subset of k-space locations that contribute to $\mathrm{WMSETD}(\boldsymbol{\lambda})$ and $\mathrm{WSURE}(\boldsymbol{\lambda})$. One could also use a diagonal $\mathbf{W}$ with significantly larger weights for outer k-space samples so as to boost the error in high spatial frequencies when computing $\mathrm{WMSETD}(\boldsymbol{\lambda})$ and $\mathrm{WSURE}(\boldsymbol{\lambda})$. The proposed methods thus allow the user some freedom in choosing the type of k-space weighting $\mathbf{W}$ for the quadratic error $\mathrm{WMSETD}(\boldsymbol{\lambda})$. Finding suitable weighting matrices, $\mathbf{W_D}$, that yield "better" parameter selections than $\mathbf{W} = \mathbf{I}_M$ is interesting future work.

Theorem 2 is a key result in this work that forms the basis of our Monte-Carlo parameter selection method for single-coil MRI. While it demands strong differentiability hypotheses on $\mathbf{u}_{\boldsymbol{\lambda}}$ as presented in Section IV, numerical experiments in this paper and the accompanying supplementary material corroborate its applicability to complex-valued weakly differentiable $\mathbf{u}_{\boldsymbol{\lambda}}$ as well. Broadening the theoretical scope of Theorem 2 to such $\mathbf{u}_{\boldsymbol{\lambda}}$ along with a bias-variance analysis of the Monte-Carlo estimate (21) are interesting directions for future research. The bias-variance analysis especially is important from a practical perspective as it can help the user choose a suitable $\boldsymbol{\Lambda}$ and ε in (21) for a given reconstruction method $\mathbf{u}_{\boldsymbol{\lambda}}$.

Another interesting extension of this work is application to parameter selection for parallel MRI. A straightforward way of doing this would be to directly apply the proposed Monte-Carlo WSURE approach individually for data from each coil of a multi-coil array and combine the resulting MR images for all coils via a sum-of-squares-type method. Alternatively, one could use a SENSE-based [3], [31], [56] approach: the data model (1), proposed metric (7) and Monte-Carlo WSURE (12), (21) are directly applicable to this case with $\mathbf{A} = \mathbf{FS}$ [3], [9], where $\mathbf{F}$ represents the Fourier encoding matrix and $\mathbf{S}$ denotes the matrix of sensitivity maps for all coils. However caution must be exercised in this case: in practice, $\mathbf{S}$ is usually unknown and needs to be estimated, e.g., from low-resolution images. Since $\mathrm{WMSETD}(\boldsymbol{\lambda})$ [and $\mathrm{WSURE}(\boldsymbol{\lambda})$] involves $\mathbf{S}$ (via $\mathbf{A}$), its appropriateness as an image-quality metric depends on the quality of the estimate, $\widehat{\mathbf{S}}$, of $\mathbf{S}$, and needs to be validated for a given $\widehat{\mathbf{S}}$. One faces a similar issue with image-domain SURE-based methods for SENSE-type parallel MRI reconstruction [26].

To circumvent the dependence on $\mathbf{S}$, we recently proposed a similar Monte-Carlo WSURE-based parameter tuning scheme [34]–[36] for some existing parallel MRI reconstruction methods such as ℓ_1-SPIRiT [7] and DESIGN [8] (based on GRAPPA [57] and sparsity) that do not need explicit knowledge of coil-sensitivity maps $\mathbf{S}$. Preliminary results [34]–[36] for undersampled *Cartesian* parallel MR data indicate that our WSURE-based approach is able to provide near-MSE-optimal selection of regularization parameters for these methods. We are currently investigating extensions to undersampled *non-Cartesian* parallel MRI.

VIII. Summary and Conclusion

Selection of proper regularization parameters $\boldsymbol{\lambda}$ is a crucial task in regularized MRI reconstruction from undersampled k-space data. We proposed a weighted squared-error measure in k-space, $\mathrm{WMSETD}(\boldsymbol{\lambda})$ (7), to assess MRI reconstruction quality and thereby adjust $\boldsymbol{\lambda}$ by minimizing it. The proposed $\mathrm{WMSETD}(\boldsymbol{\lambda})$ is amenable for estimation using Stein's principle [20], [21] for Gaussian noise. The Stein-type estimate of $\mathrm{WMSETD}(\boldsymbol{\lambda})$, denoted by $\mathrm{WSURE}(\boldsymbol{\lambda})$, requires (in addition to the noise covariance matrix) computing the trace of a linear transformation of the Jacobian matrix of the MRI reconstruction algorithm $\mathbf{u}_{\boldsymbol{\lambda}}$ with respect to k-space data $\mathbf{y}$. Our major contribution in this work is a Monte-Carlo scheme that enables the estimation of this trace without requiring the knowledge of the internal working of $\mathbf{u}_{\boldsymbol{\lambda}}$. This feature thus enables its applicability for a wide-range of reconstruction algorithms involving a variety of convex nonquadratic regularizers including total variation and ℓ_1-regularization. The proposed Monte-Carlo method extends our previous result for denoising of real-valued images in [32, Th. 2] to the case of inverse problems involving complex-valued images with application to MRI reconstruction.

Although $\mathrm{WMSETD}(\boldsymbol{\lambda})$ differs from the image-domain $\mathrm{MSE}(\boldsymbol{\lambda})$ that is not amenable for estimation in practical inverse problems [21], we demonstrated using experiments with undersampled synthetic and real MR data that $\mathrm{WMSETD}(\boldsymbol{\lambda})$, via its estimate $\mathrm{WSURE}(\boldsymbol{\lambda})$, is able to provide near-MSE-optimal selection of regularization parameters for single-coil non-Cartesian MRI reconstruction. These results both extend and corroborate our previous work [27] on similar parameter-tuning methods for single-coil undersampled Cartesian MRI reconstruction. Theoretical developments in this paper are fairly general and can be readily extended to handle parameter-tuning for (iterative) linear/nonlinear parallel MRI reconstruction from undersampled Cartesian/non-Cartesian k-space data.

Appendix
Proof of Lemma 1

From the hypotheses of Lemma 1, it is clear that the probability density function of $\boldsymbol{\xi}$ is given by $g(\boldsymbol{\xi}) = K\exp(-\boldsymbol{\xi}'\boldsymbol{\Omega}^{-1}\boldsymbol{\xi})$, where $K > 0$ is some normalization constant. It is easy to verify that $g(\boldsymbol{\xi})$ satisfies

$$g(\boldsymbol{\xi})\,\boldsymbol{\xi}' = -\nabla_{\boldsymbol{\xi}} g(\boldsymbol{\xi})\boldsymbol{\Omega} \tag{29}$$

where $\nabla_{\boldsymbol{\xi}} \triangleq 1/2(\nabla_{\boldsymbol{\xi}_{\mathcal{R}}} - \iota\nabla_{\boldsymbol{\xi}_{\mathcal{I}}})$ and $\nabla_{\boldsymbol{\xi}_{\mathcal{R}}}$, $\nabla_{\boldsymbol{\xi}_{\mathcal{I}}}$ are $1 \times M$ gradient operators with respect to the real, $\boldsymbol{\xi}_{\mathcal{R}}$, and imaginary, $\boldsymbol{\xi}_{\mathcal{I}}$, parts of $\boldsymbol{\xi}$, respectively. We start from the left hand side of (10) and use (9), (29) and $d\boldsymbol{\xi} \triangleq d\boldsymbol{\xi}_{\mathcal{R}}\, d\boldsymbol{\xi}_{\mathcal{I}}$ to obtain

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\} &= \int g(\boldsymbol{\xi})\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi} \\
&= -\int \nabla_{\boldsymbol{\xi}} g(\boldsymbol{\xi})\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi} \\
&= -\frac{1}{2}\int \nabla_{\boldsymbol{\xi}_{\mathcal{R}}} g(\boldsymbol{\xi})\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi} \\
&\quad + \frac{\iota}{2}\int \nabla_{\boldsymbol{\xi}_{\mathcal{I}}} g(\boldsymbol{\xi})\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi}.
\end{aligned} \tag{30}$$

In the sequel, $m = 1, \ldots, M$ and $n = 1, \ldots, N$, respectively. We focus on the term involving $\nabla_{\boldsymbol{\xi}_{\mathcal{R}}}$ in (30) and use integration-by-parts along with the fact that $\mathsf{E}_{\boldsymbol{\xi}}\{|[\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_m|\} < \infty$, to get that [21, Th. 1]

$$\begin{aligned}
&\int \nabla_{\boldsymbol{\xi}_{\mathcal{R}}} g(\boldsymbol{\xi})\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi} \\
&\quad = -\sum_{m,n}\int g(\boldsymbol{\xi})[\boldsymbol{\Gamma}]_{mn}\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial \xi_{\mathcal{R}m}}d\boldsymbol{\xi} \\
&\quad = -\sum_{m,n}\int g(\boldsymbol{\xi})[\boldsymbol{\Gamma}]_{mn}\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{R}m}}d\boldsymbol{\xi}
\end{aligned} \tag{31}$$

where we have set $\partial/\partial\xi_{\mathcal{R}m} = \partial/\partial y_{\mathcal{R}m}$ since $\mathbf{y}_{\mathrm{true}}$ in (1) is a deterministic constant. Similarly

$$\int \nabla_{\boldsymbol{\xi}_{\mathcal{I}}} g(\boldsymbol{\xi})\boldsymbol{\Gamma}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})d\boldsymbol{\xi} = -\sum_{m,n}\int g(\boldsymbol{\xi})[\boldsymbol{\Gamma}]_{mn}\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{I}m}}d\boldsymbol{\xi}. \tag{32}$$

Combining (30)–(32) and using (11), we get that

$$\begin{aligned}
&\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}'\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})\} \\
&\quad = \mathsf{E}_{\boldsymbol{\xi}}\left\{\sum_{m,n}[\boldsymbol{\Gamma}]_{mn}\frac{1}{2}\left(\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{R}m}} - \iota\frac{\partial[\mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y})]_n}{\partial y_{\mathcal{I}m}}\right)\right\} \\
&\quad = \mathsf{E}_{\boldsymbol{\xi}}\left\{\sum_{m,n}[\boldsymbol{\Gamma}]_{mn}[\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})]_{nm}\right\} \\
&\quad = \mathsf{E}_{\boldsymbol{\xi}}\left\{\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\}\right\}
\end{aligned}$$

which is the desired result.

References

[1] J. A. Fessler, "Model-based image reconstruction for MRI," *IEEE Sig. Process. Mag.*, vol. 27, no. 4, pp. 81–89, Jul. 2010.
[2] M. Bydder, A. A. Samsonov, and J. Du, "Evaluation of optimal density weighting for regridding," *Mag. Res. Imag.*., vol. 25, no. 5, pp. 695–702, Jun. 2007.
[3] L. Ying, B. Liu, M. Steckner, G. Wu, M. Wu, and S.-J. Li, "A statistical approach to SENSE regularization with arbitrary k-space trajectories," *Mag. Res. Med.*, vol. 60, no. 2, pp. 414–421, Aug. 2008.
[4] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Mag. Res. Med.*, vol. 57, no. 6, pp. 1086–1098, Jun. 2007.
[5] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Mag. Res. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
[6] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1649–1660, Sep. 2011.

[7] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast ℓ_1-SPIRiT compressed sensing parallel imaging MRI: Scalable parallel implementation and clinically feasible runtime," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1250–1262, Jun. 2012.

[8] D. S. Weller, J. R. Polimeni, L. Grady, L. L. Wald, E. Adalsteinsson, and V. K. Goyal, "Denoising sparse images from GRAPPA using the nullspace method (DESIGN)," *Mag. Res. Med.*, vol. 68, no. 4, pp. 1176–1189, Oct. 2012.

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

[10] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image & Video Processing*, A. Bovik, Ed., 2nd ed. New York: Elsevier, 2005, pp. 183–202.

[11] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Imag. Process.*, vol. 1, no. 3, pp. 322–336, Jul. 1992.

[12] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.

[13] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.

[14] F.-H. Lin, K. K. Kwong, J. W. Belliveau, and L. L. Wald, "Parallel imaging reconstruction using automatic regularization," *Mag. Res. Med.*, vol. 51, no. 3, pp. 559–567, Mar. 2004.

[15] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 31, pp. 377–403, 1979.

[16] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Opt. Eng.*, vol. 29, no. 5, pp. 446–454, 1990.

[17] D. A. Girard, "The fast Monte-Carlo Cross-Validation and C_L procedures: Comments, new results and application to image recovery problems," *Computat. Stat.*, vol. 10, pp. 205–231, 1995.

[18] D. A. Girard, "The fast Monte-Carlo Cross-Validation and C_L procedures: Comments, new results and application to image recovery problems—Rejoinder," *Computat. Stat.*, vol. 10, pp. 251–258, 1995.

[19] J. D. Carew, G. Wahba, X. Xie, E. V. Nordheim, and M. E. Meyerandb, "Optimal spline smoothing of fMRI time series by generalized cross-validation," *NeuroImage*, vol. 18, pp. 950–961, 2003.

[20] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[21] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[22] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, Dec. 2009.

[23] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Computat. Harmonic Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.

[24] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," in *Proc. IEEE Int. Conf. Imag. Process.*, 2008, pp. 665–668.

[25] F. Luisier, T. Blu, and P. J. Wolfe, "A CURE for noisy magnetic resonance images: Chi-square unbiased risk estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3454–3466, Aug. 2012.

[26] A. Marin, C. Chaux, J.-C. Pesquet, and P. Ciuciu, "Image reconstruction from multiple sensors using Stein's principle. Application to parallel MRI," in *Proc. IEEE Int. Symp. Biomed. Imag: From Nano to Macro*, Apr. 2, 2011, pp. 465–468.

[27] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3659–3672, Aug. 2012.

[28] J. G. Brankov, Y. Yang, L. Wei, I. E. Naqa, and M. N. Wernick, "Learning a channelized observer for image quality assessment," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 991–999, Jul. 2009.

[29] L. Zhang, C. C-Ménard, P. L. Callet, and J.-Y. Tanguy, "A perceptually relevant channelized joint observer (PCJO) for the detection-localization of parametric signals," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1875–1888, Oct. 2012.

[30] H. Luong, B. Goossens, J. Aelterman, L. Platisa, and W. Philips, "Optimizing image quality in MRI: On the evaluation of k-space trajectories for under-sampled MR acquisition," in *Proc. 4th Int. Workshop Qual. Mult. Exp. (QoMEX)*, 2012, pp. 25–26.

[31] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Mag. Res. Med.*, vol. 42, no. 5, pp. 952–962, Nov. 1999.

[32] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[33] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, "Realistic analytical phantoms for parallel magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 626–636, Mar. 2012.

[34] D. S. Weller, S. Ramani, J.-F. Nielsen, and J. A. Fessler, "SURE-based parameter selection for parallel MRI reconstruction using GRAPPA and sparsity," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, 2013, pp. 942–945.

[35] D. S. Weller, S. Ramani, J.-F. Nielsen, and J. A. Fessler, "Automatic ℓ_1-SPIRiT regularization parameter selection using Monte-Carlo SURE," in *Proc. Ann. Meeting Int. Soc. Mag. Res. Med.*, 2013, p. 2602.

[36] D. S. Weller, S. Ramani, J.-F. Nielsen, and J. A. Fessler, "Monte-Carlo SURE-based parameter selection for parallel magnetic resonance imaging reconstruction," *Mag. Res. Med.*, submitted for publication.

[37] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. H: Microw., Opt. Antennas*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[38] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. Providence, RI: Am. Math. Soc., 2001.

[39] A. Van den Bos, "Complex gradient and hessian," *IEE Proc. Vis. Imag. Signal Proc.*, vol. 141, no. 6, pp. 380–382, Dec. 1994.

[40] D. A. Girard, "A fast 'Monte-Carlo cross-Validation' procedure for large least squares problems with noisy data," *Numer. Math.*, vol. 56, pp. 1–23, 1989.

[41] M. F. Hutchinson, "A stochastic estimator of the trace of the influence Matrix for Laplacian Smoothing Splines," *Commun. Stat. Simul. Comput.*, vol. 18, no. 3, pp. 1059–1076, 1989.

[42] Z. Bai, M. Fahey, and G. Golub, "Some large-scale matrix computation problems," *J. Comput. Appl. Math.*, vol. 74, pp. 71–89, 1996.

[43] S. Dong and K. Liu, "Stochastic estimation with z_2 noise," *Phys. Lett. B*, vol. 328, pp. 130–136, 1994.

[44] D. C. Noll, J. A. Fessler, and B. P. Sutton, "Conjugate phase MRI reconstruction with spatially variant sample density correction," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 325–336, Mar. 2005.

[45] C. M. Hurvich and C.-L. Tsai, "A crossvalidatory AIC for hard wavelet thresholding in spatially adaptive function estimation," *Biometrika*, vol. 85, no. 3, pp. 701–710, 1998.

[46] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, Feb. 2003.

[47] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[48] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Stat. Comp.*, vol. 9, no. 4, pp. 766–771, Jul. 1988.

[49] A. E. Yagle, "A fast algorithm for Toeplitz-block-Toeplitz linear systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2001, vol. 3, pp. 1929–1932.

[50] R. H. Chan and M. K. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Rev.*, vol. 38, no. 3, pp. 427–482, Sep. 1996.

[51] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688–699, May 1999.

[52] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.

[53] M. L. Lauzon and B. K. Rutt, "Effects of polar sampling in k-space," *Mag. Res. Med.*, vol. 36, no. 6, pp. 940–949, Dec. 1996.

[54] P. M. Joseph, "Sampling errors in projection reconstruction MRI," *Mag. Res. Med.*, vol. 40, no. 3, pp. 460–466, Sep. 1998.

[55] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Comput. J.*, vol. 7, no. 2, pp. 155–162, 1964.

[56] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Mag. Res. Med.*, vol. 46, no. 4, pp. 638–651, Oct. 2001.

[57] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Mag. Res. Med.*, vol. 47, no. 6, pp. 1202–1210, Jun. 2002.

Non-Cartesian MRI Reconstruction With Automatic Regularization Via Monte-Carlo SURE: Supplementary Material

Sathish Ramani*, *Member, IEEE,* Daniel S. Weller, *Member, IEEE,* Jon-Fredrik Nielsen, and Jeffrey A. Fessler, *Fellow, IEEE*

We provide here additional results for various experiments in [1]. Fig. 1 illustrates some of the non-Cartesian trajectories used in [1]. References to equations, tables, figures, bibliography, etc., are within this material only unless specified otherwise.

I. Robustness of Monte-Carlo Estimation

We are interested in determining the range of ε for which the Monte-Carlo estimation procedure (with only one realization of random vector $\mathbf{b}$) in [1, Sec. IV] is an adequate approximation:

$$\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\boldsymbol{\lambda}}}(\mathbf{y})\} \approx \varepsilon^{-1}\mathbf{b}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon) \qquad (1)$$

where

$$\boldsymbol{\varrho}(\mathbf{u}_{\boldsymbol{\lambda}}, \mathbf{y}, \boldsymbol{\Lambda}\mathbf{b}, \varepsilon) \triangleq \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y}+\varepsilon\boldsymbol{\Lambda}\mathbf{b}) - \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{y}). \qquad (2)$$

The Monte-Carlo estimate (1) is used in

$$\begin{aligned}\mathrm{WSURE}(\lambda) \triangleq{}& M^{-1}\|\mathbf{y}-\mathbf{A}\mathbf{u}_{\lambda}(\mathbf{y})\|^2_{\mathbf{W}} - M^{-1}\mathrm{tr}\{\boldsymbol{\Omega}\mathbf{W}\} \\ &+ 2M^{-1}\mathcal{R}\{\mathrm{tr}\{\boldsymbol{\Gamma}\mathbf{J}_{\mathbf{u}_{\lambda}}(\mathbf{y})\}\} \qquad (3)\end{aligned}$$

that is an unbiased estimate of

$$\mathrm{WMSETD}(\lambda) \triangleq M^{-1}\|\mathbf{y}_{\mathrm{true}}-\mathbf{A}\mathbf{u}_{\lambda}(\mathbf{y})\|^2_{\mathbf{W}}. \qquad (4)$$

We use the experimental setup described in [1, Sec. VI-A] throughout this material with $\mathbf{W} = \mathbf{I}_M$ and $\mathbf{W} = \mathbf{W}_{\mathrm{D}}$ in [1, Eq. (24)] and $\boldsymbol{\Lambda} = \mathbf{I}_M$ in (1)-(2). The proposed Monte-Carlo estimation scheme (1) and the hypotheses of [1, Thm. 2] are applicable to the smooth-convex regularizer Ψ_{FP} [1, Eq. (26)], but they do not directly apply to the total-variation regularizer Ψ_{TV} [1, Eq. (27)]. One of our aims in this note is to provide numerical results that further corroborate those in [1] extending the scope of (1)-(4) to nonsmooth regularizers such as Ψ_{TV}.

We repeated the radial MRI simulation in [1, Sec. VI-B.1] for varying levels of noise in the data and plotted the standard deviation of Monte-Carlo WSURE normalized by WMSETD in Figs. 2-5. The plots were generated by averaging Monte-Carlo $\mathrm{WSURE}(\lambda)$ (1)-(3) over 25 Monte-Carlo realizations of $\mathbf{b}$ in (1)-(2). These plots indicate that the variance of Monte-Carlo WSURE increases with decreasing ε consistently in all experiments and corroborate the expected behavior of (1) described in [1, Sec. IV]. From these plots, $\varepsilon = 10^{-7}$ appears to be a reasonable lower bound for ε for such experiments.

This work was supported by the National Institutes of Health under Grant P01 CA87634 and by CPU donations from Intel.

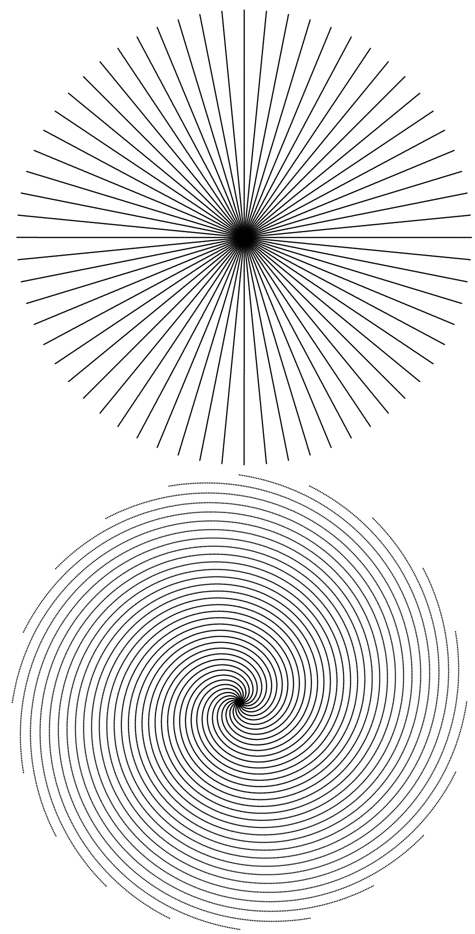


Fig. 1. Top: 32 spokes (with 512 samples each) of the non-Cartesian radial k-space trajectory used in [1, Sec. VI-B]. Bottom: 20 interleaves (with 841 samples each) of the non-Cartesian variable density spiral k-space trajectory used in [1, Secs. VI-C, VI-D].

Next, we repeated the experiment in [1, Sec. VI-B.2] for varying SNR of data using only one realization of $\mathbf{b}$ as is desirable in practice. Figs. 6-13 plot λ-values and $\mathrm{PSNR}(\lambda)$ as functions of ε where λ was chosen to minimize $\mathrm{WSURE}(\lambda)$ and the true $\mathrm{MSE}(\lambda)$. These plots indicate that a suitable choice of ε is $\varepsilon \in [10^{-5}, 10^{-2}]$; however, it should be kept in mind this range may change depending upon the type of imaging problem, the reconstruction algorithm $\mathbf{u}_{\lambda}$ in [1] and the scale of $\mathbf{y}$ relative to that of $\mathbf{b}$.

We successfully used $\varepsilon = 10^{-4}$ with the SB algorithm in all experiments in this material and also in [1] for near-MSE-optimal MRI reconstruction from single-coil undersampled

TABLE I

EXPERIMENT IN [1, SEC. VI-B]: PSNR OF IMAGES RECONSTRUCTED USING Ψ_{TV} WITH λ OPTIMIZED BY VARIOUS METHODS FOR DATA WITH VARYING SNR.

SNR (dB)	PSNR (dB) MSE(λ)	NGCV(λ)	WSURE(λ) $\mathbf{W} = \mathbf{I}_M$	WSURE(λ) $\mathbf{W} = \mathbf{W}_\mathbf{D}$
20	28.21	28.21	28.21	28.21
30	31.20	31.14	31.14	31.20
40	32.85	32.85	32.85	32.85

data (both simulated and acquired using a GE 3T MRI scanner) on different non-Cartesian (radial and variable-density spiral) k-space trajectories. These experimental results also indicate that the proposed Monte-Carlo estimation scheme (1) can be successfully used with nonsmooth regularizers such as Ψ_{TV}.

II. SIMULATION WITH VARYING NOISE LEVEL

Here, we repeated the experiment in [1, Sec. VI-C] with varying levels of noise in the simulated data, but with Ψ_{TV}. We again assumed that the noise variance σ^2 was known in each case for use in $\mathrm{WSURE}(\lambda)$ (3). We tabulate PSNR [1, Sec. VI-B] of reconstructed images obtained by minimizing $\mathrm{WSURE}(\lambda)$ and $\mathrm{NGCV}(\lambda)$ [1, Sec. VI-A] in Table I. $\mathrm{WSURE}(\lambda)$ was able to provide near-MSE-optimal λ-selections as indicated by the PSNR-values in Table I. NGCV also provides similar λ-selections in this experiment.

III. IN-VIVO HUMAN BRAIN DATA

We repeated the experiment in [1, Sec. VI-D] for different slices of the acquired 3D volume. Figs. 14-15 show images reconstructed using Ψ_{ℓ_1} [1, Sec. VI-A] as the regularizer with λ selected by minimizing $\mathrm{WSURE}(\lambda)$ and $\mathrm{NGCV}(\lambda)$ [1, Sec. VI-A]. In agreement with the results in [1, Sec. VI-D], $\mathrm{NGCV}(\lambda)$ yielded over-smoothed images for this data-set while $\mathrm{WSURE}(\lambda)$ was able to provide images that appear comparable to the corresponding references.

REFERENCES

[1] S. Ramani, D. S. Weller, J-F. Nielsen, and J. A. Fessler, "Non-Cartesian MRI reconstruction with automatic regularization via Monte-Carlo SURE," *IEEE Trans. Med. Imag.*, 2013.

[2] M. Bydder, A. A. Samsonov, and J. Du, "Evaluation of optimal density weighting for regridding," *Mag. Res. Im.*, vol. 25, no. 5, pp. 695–702, June 2007.

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

[4] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, "Realistic analytical phantoms for parallel magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 626–36, Mar. 2012.

[5] S. Ramani, Z. Liu, J. Rosen, J-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Im. Proc.*, vol. 21, no. 8, pp. 3659–72, Aug. 2012.

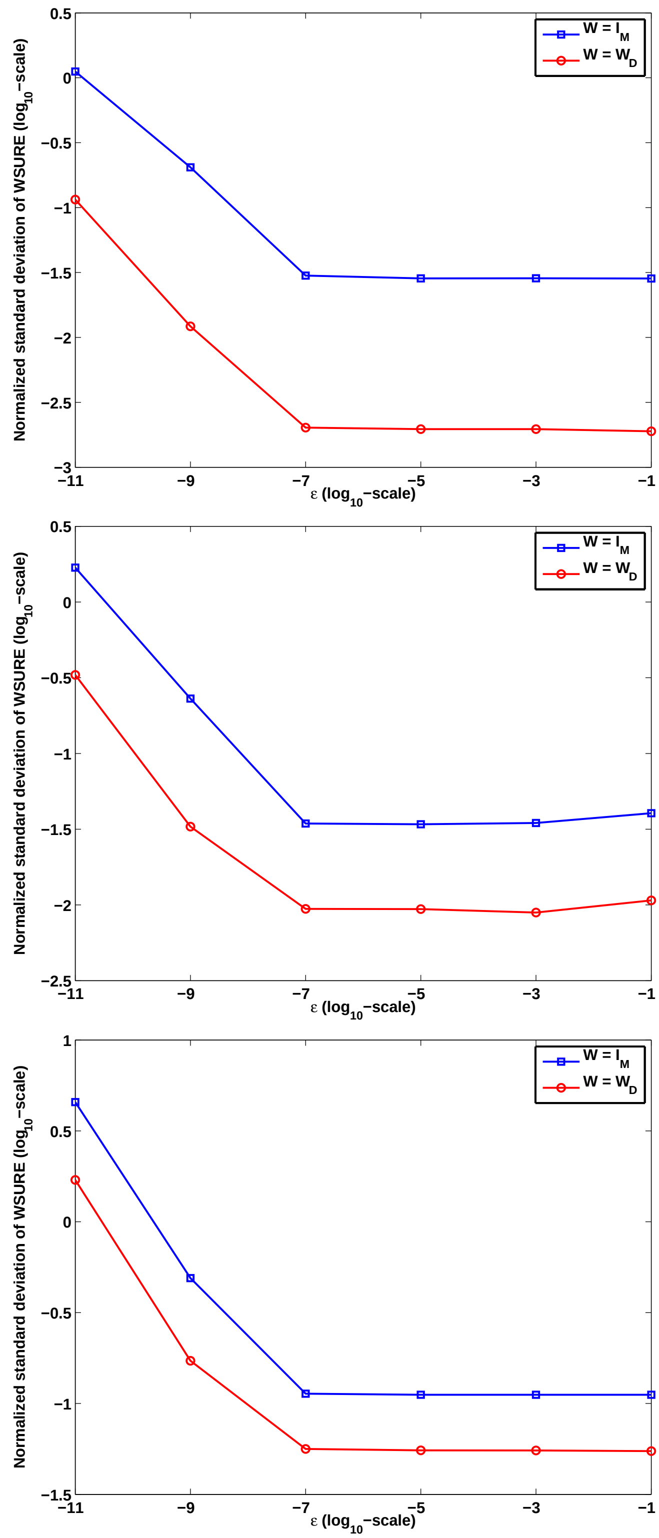


Fig. 2. Radial MRI simulation with the analytical Shepp-Logan phantom [4] in [1, Sec. VI-B.2]. Plots of standard deviation of $\mathrm{WSURE}(\lambda)$ normalized by $\mathrm{WMSETD}(\lambda)$ as a function of ε in (1) for (top) $\lambda = \lambda_{\mathrm{opt}}/10$, (middle) $\lambda = \lambda_{\mathrm{opt}}$, and (bottom) $\lambda = 10\lambda_{\mathrm{opt}}$, where λ_{opt} is the MSE-optimal value of the regularization parameter. The curves were obtained by averaging (1) over 25 realizations of $\mathbf{b}$. As expected, the variance rapidly increases for smaller ε. The SNR of data was 20 dB.

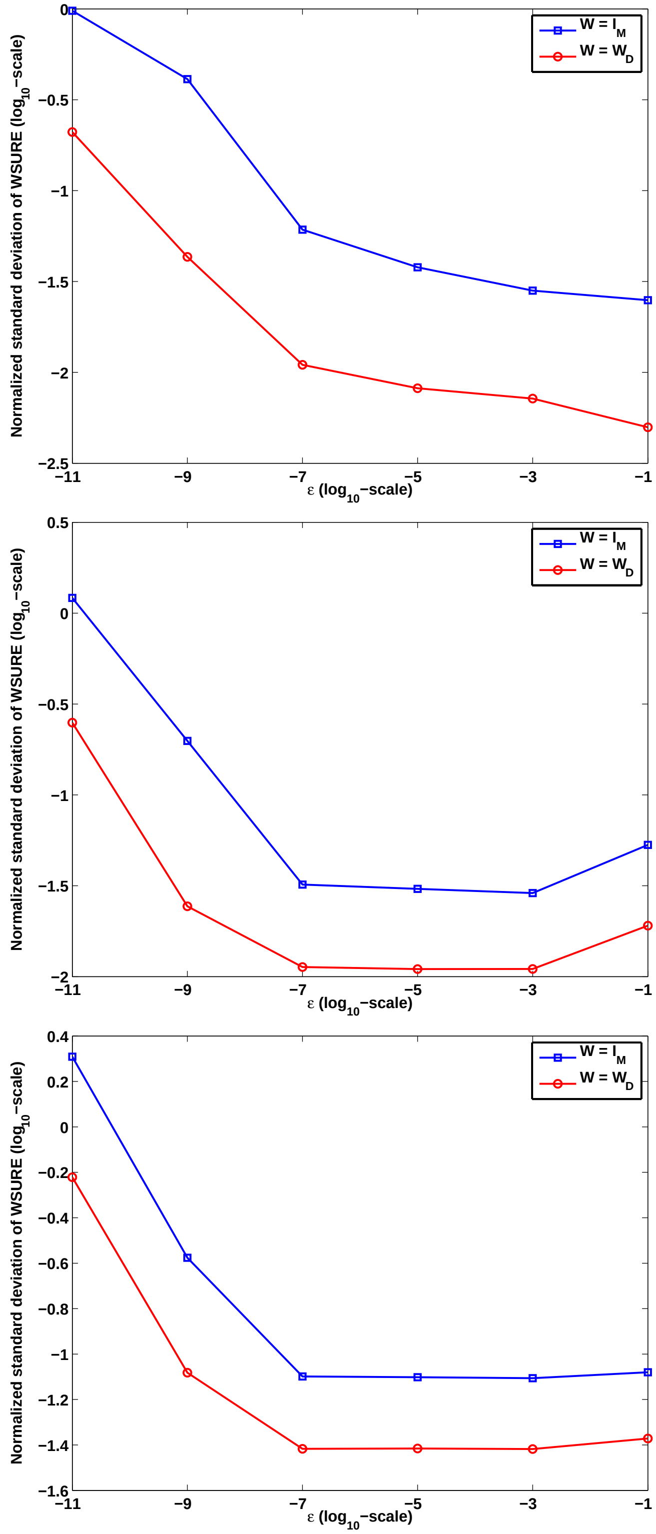


Fig. 3. Same experiment as in Fig. 2. The SNR of data was 30 dB.

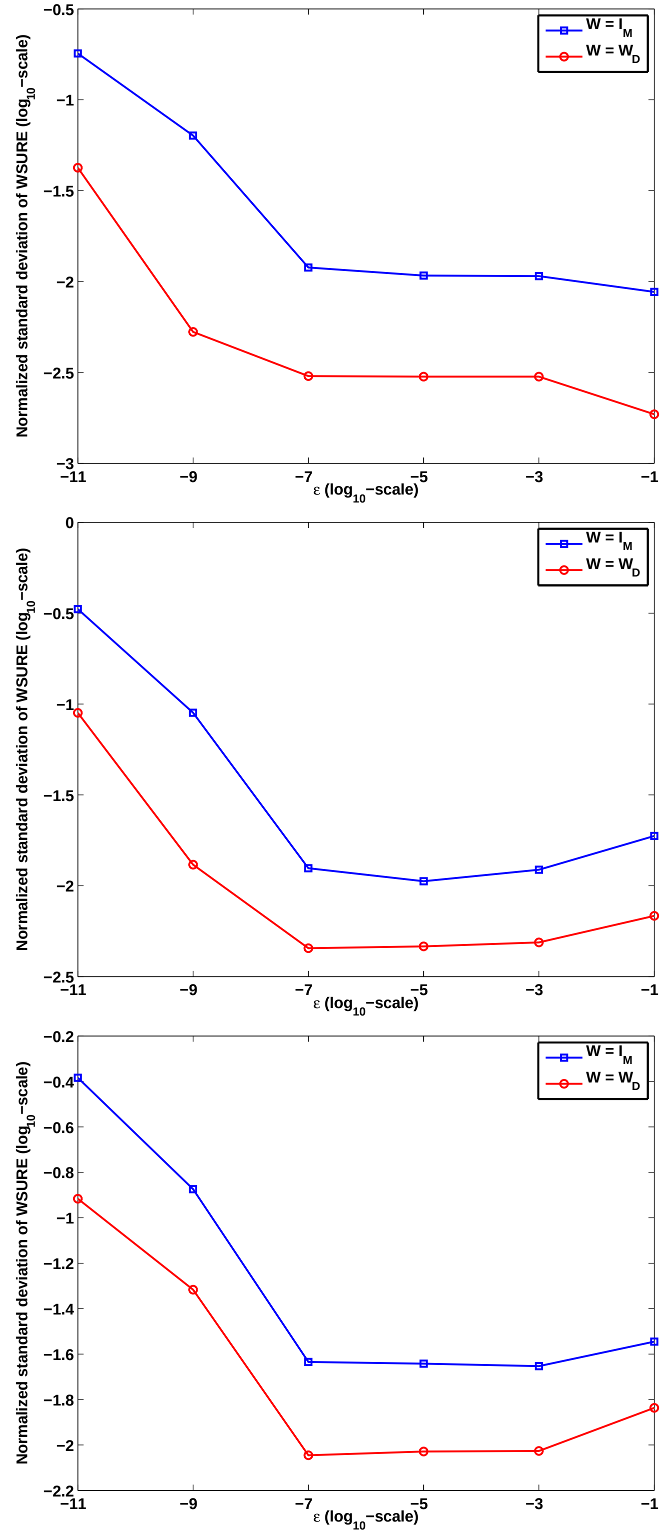


Fig. 4. Same experiment as in Fig. 2. The SNR of data was 40 dB.

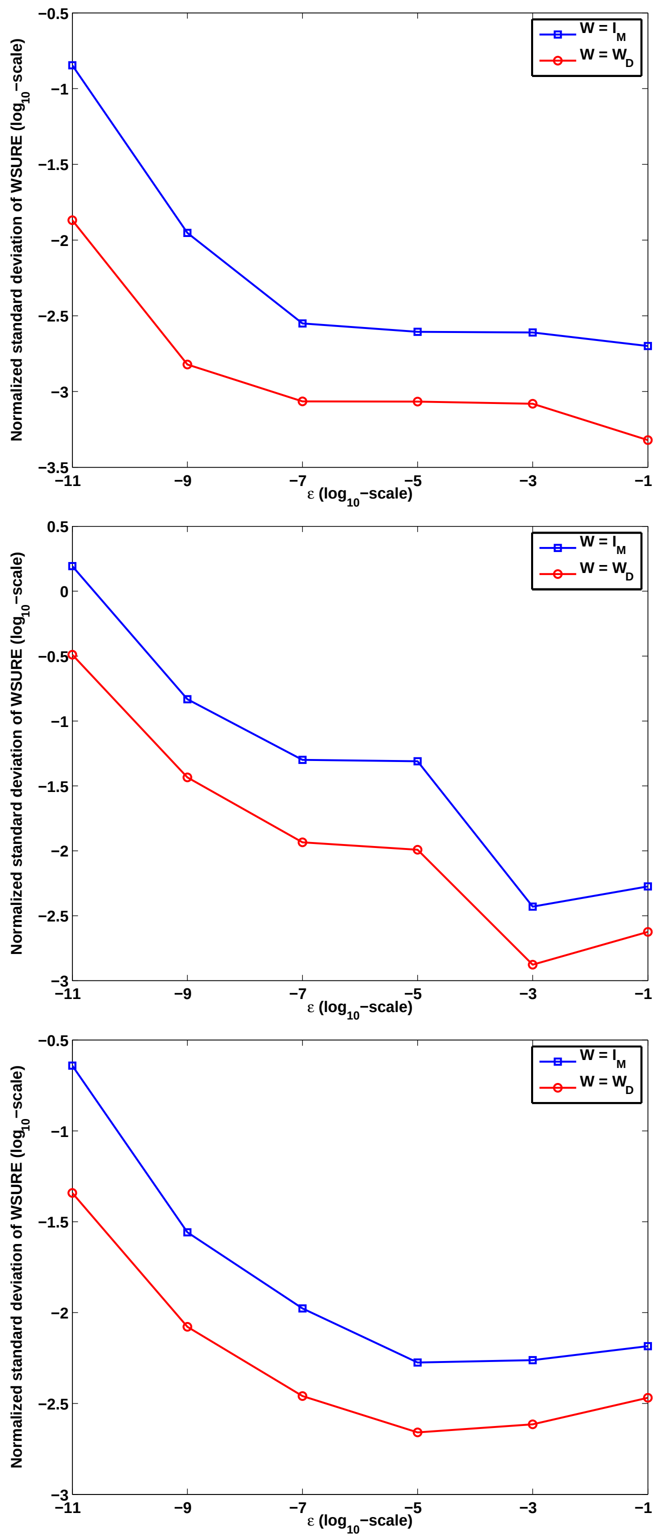


Fig. 5. Same experiment as in Fig. 2. The SNR of data was 50 dB.

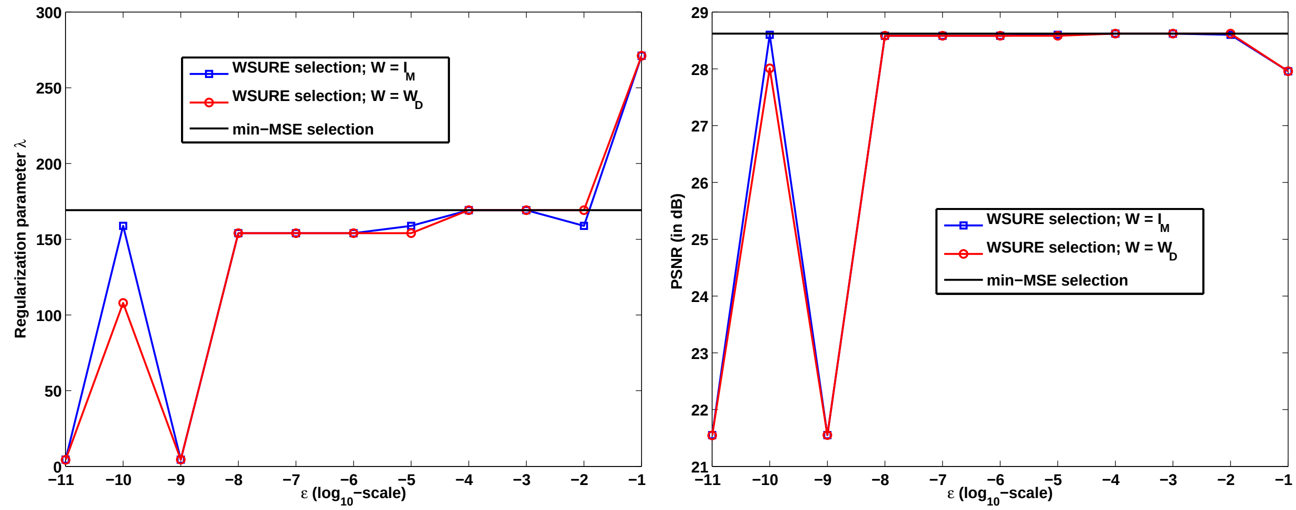


Fig. 6. Plots of (left) λ, and (right) $\mathrm{PSNR}(\lambda)$ as functions of ε for λ selected to minimize $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ and $\mathbf{W}_\mathbf{D}$ in (3) and $\mathrm{MSE}(\lambda)$ for the experiment described in [1, Sec. VI-B2] with **SNR = 20 dB and Ψ_{FP} as the regularizer.**

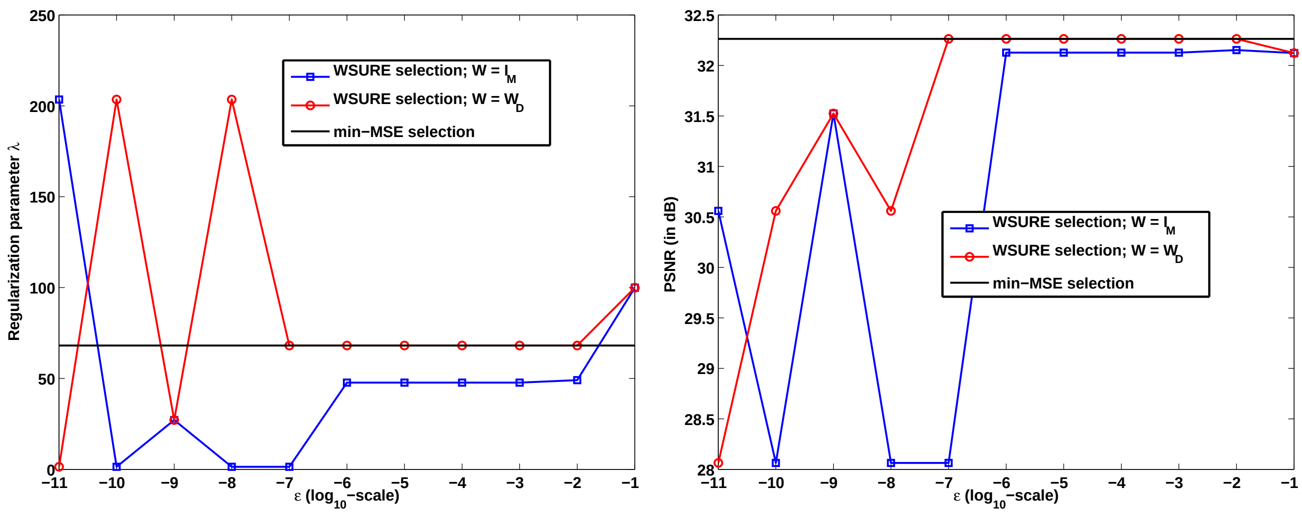


Fig. 7. Same as in Fig. 6, but SNR = 30 dB.

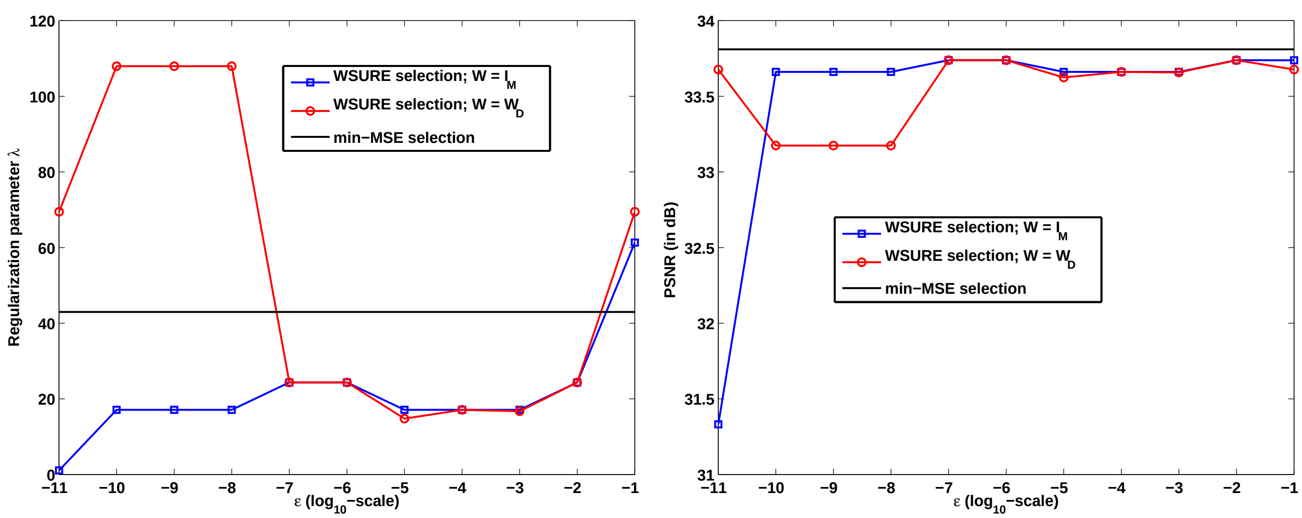


Fig. 8. Same as in Fig. 6, but SNR = 40 dB.

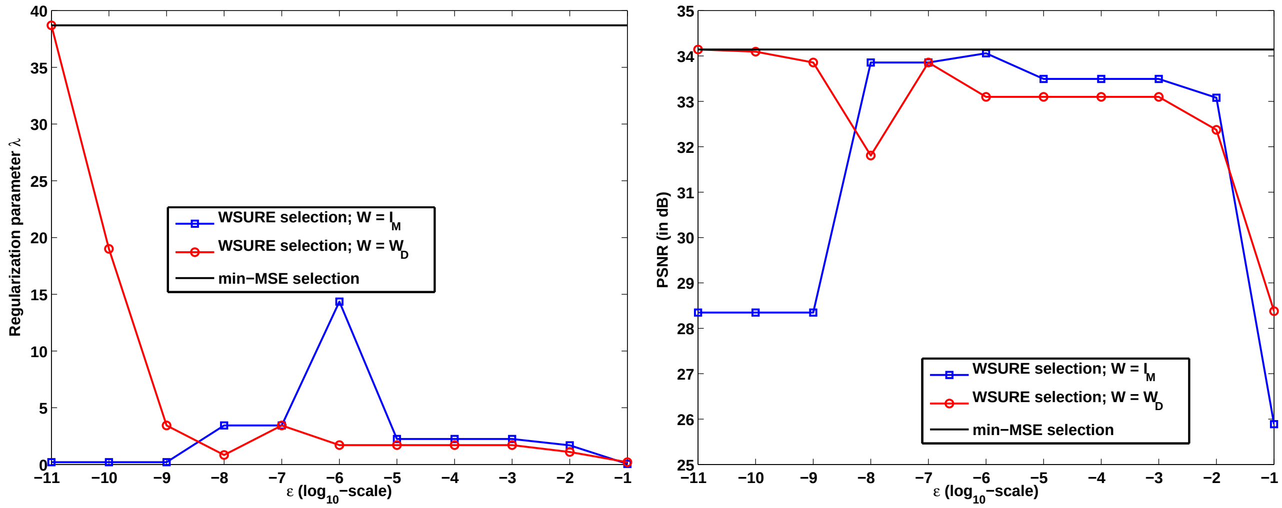


Fig. 9. Same as in Fig. 6, but SNR = 50 dB.

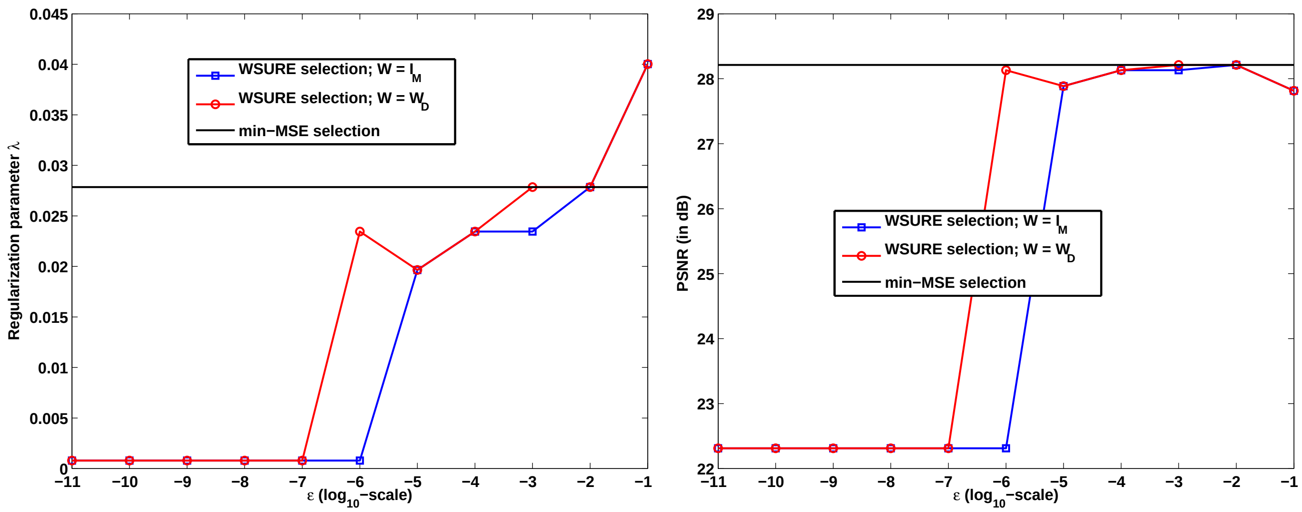


Fig. 10. Plots of (left) λ, and (right) $\mathrm{PSNR}(\lambda)$ as functions of ε for λ selected to minimize $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$ and $\mathbf{W}_\mathbf{D}$ in (3) and $\mathrm{MSE}(\lambda)$ for the experiment described in [1, Sec. VI-B2] with **SNR = 20 dB and Ψ_{TV} as the regularizer.**

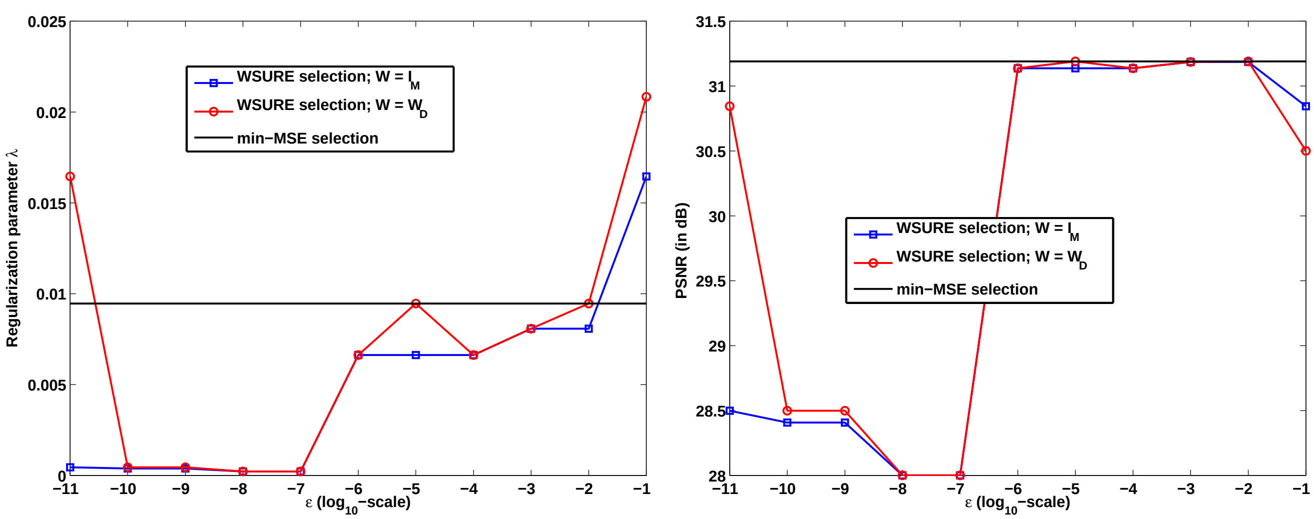


Fig. 11. Same as in Fig. 10, but SNR = 30 dB.

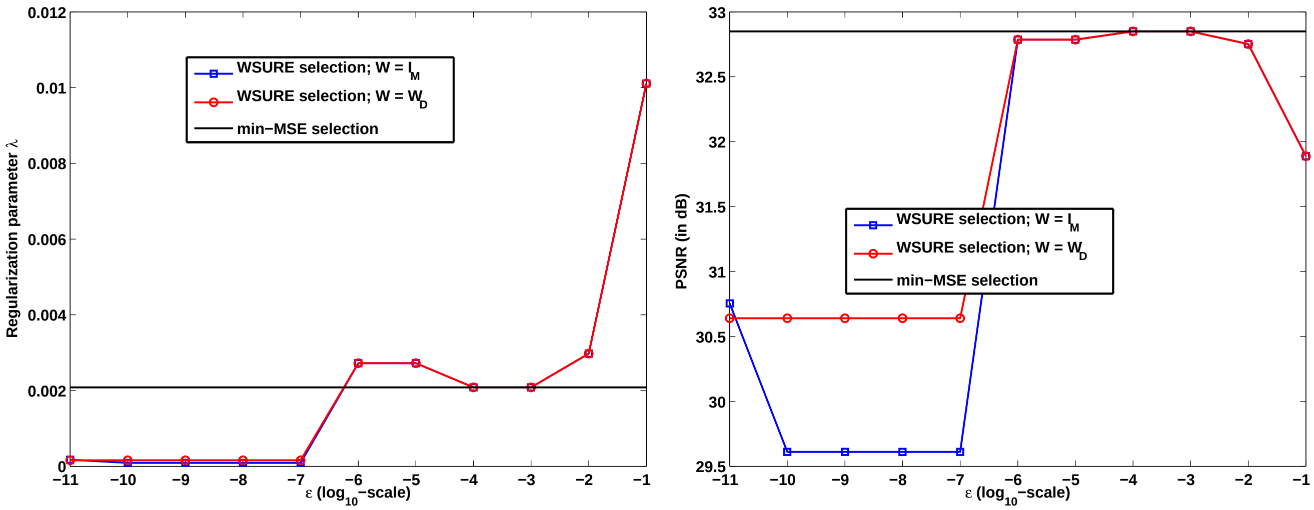


Fig. 12. Same as in Fig. 10, but SNR = 40 dB.

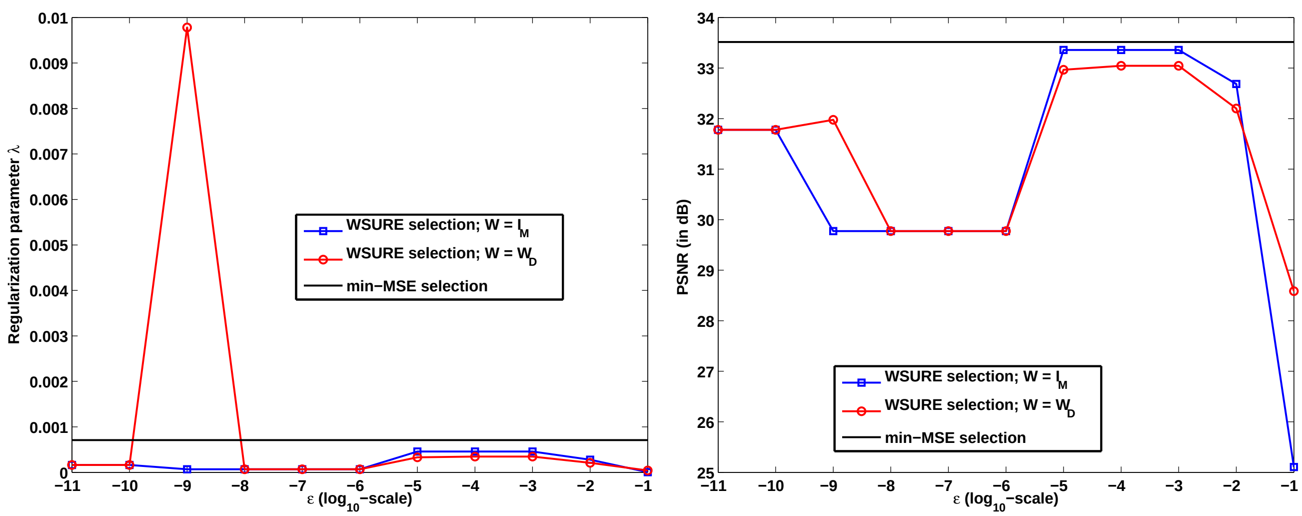


Fig. 13. Same as in Fig. 10, but SNR = 50 dB.

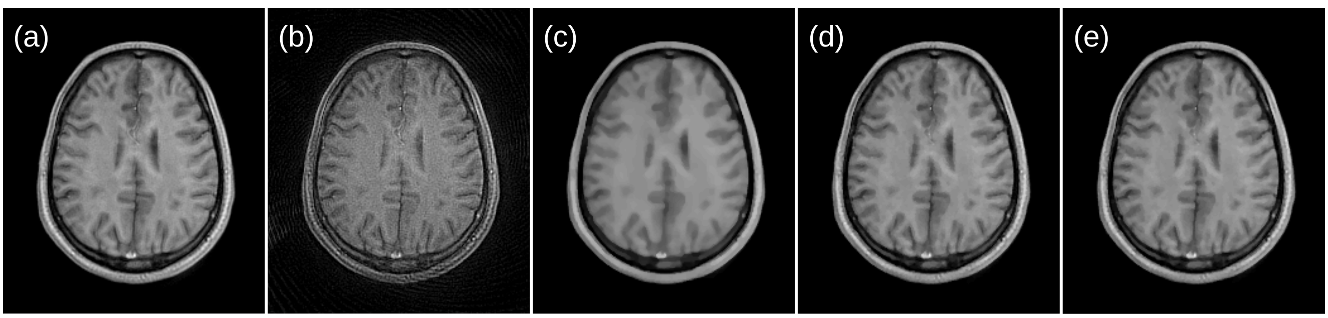


Fig. 14. Experiment with real in-vivo human head data [1, Sec. VI-D]; Slice 10. (a) Very mildly Ψ_{ℓ_1}-regularized reference reconstruction from "fully-sampled" data averaged over 3 acquisitions; (b) conjugate phase reconstruction from $2\times$ undersampled data (from a single acquisition) with density compensation; Images reconstructed from $2\times$ undersampled data (from a single acquisition) using Ψ_{ℓ_1}-regularizer with λ selected to minimize (c) $\mathrm{NGCV}(\lambda)$; (d) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W} = \mathbf{I}_M$; (e) $\mathrm{WSURE}(\lambda)$ with $\mathbf{W}_\mathbf{D}$ [1, Eq. (24)].

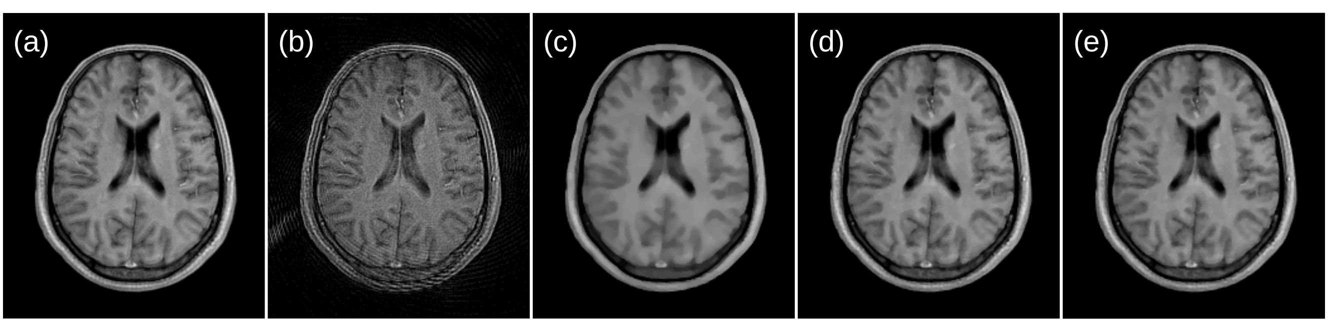


Fig. 15. Same experiment as in Fig. 14; Slice 12.