Regularization Parameter Selection for Nonlinear Iterative Image Restoration and MRI Reconstruction Using GCV and SURE-Based Methods

Sathish Ramani*, *Member, IEEE*, Zhihao Liu, Jeffrey Rosen, Jon-Fredrik Nielsen, and Jeffrey A. Fessler, *Fellow, IEEE*

***Abstract*—Regularized iterative reconstruction algorithms for imaging inverse problems require selection of appropriate regularization parameter values. We focus on the challenging problem of tuning regularization parameters for nonlinear algorithms for the case of additive (possibly complex) Gaussian noise. Generalized cross-validation (GCV) and (weighted) mean-squared error (MSE) approaches (based on Stein's Unbiased Risk Estimate—SURE) need the Jacobian matrix of the nonlinear reconstruction operator (representative of the iterative algorithm) with respect to the data. We derive the desired Jacobian matrix for two types of nonlinear iterative algorithms: a fast variant of the standard iterative reweighted least-squares method and the contemporary split-Bregman algorithm, both of which can accommodate a wide variety of analysis- and synthesis-type regularizers. The proposed approach iteratively computes two weighted SURE-type measures: Predicted-SURE and Projected-SURE (that require knowledge of noise variance σ^2), and GCV (that does not need σ^2) for these algorithms. We apply the methods to image restoration and to magnetic resonance image (MRI) reconstruction using total variation (TV) and an analysis-type ℓ_1-regularization. We demonstrate through simulations and experiments with real data that minimizing Predicted-SURE and Projected-SURE consistently lead to near-MSE-optimal reconstructions. We also observed that minimizing GCV yields reconstruction results that are near-MSE-optimal for image restoration and slightly sub-optimal for MRI. Theoretical derivations in this work related to Jacobian matrix evaluations can be extended, in principle, to other types of regularizers and reconstruction algorithms.**

EDICS
TEC-RST, COI-MRI

***Index Terms*—Regularization parameter, generalized cross-validation (GCV), Stein's unbiased risk estimate (SURE), image restoration, MRI reconstruction**

I. Introduction

Inverse problems in imaging invariably need image reconstruction algorithms to recover an underlying unknown object of interest $\mathbf{x}$ from measured data $\mathbf{y}$. Reconstruction algorithms typically depend on a set of parameters that need to be adjusted properly for obtaining good image-quality. Choosing suitable parameter values is a nontrivial, application-dependent task and has motivated research on automated parameter selection based on quantitative measures [1]–[31]. Quantitative parameter optimization methods can be broadly classified as those based on the discrepancy principle [1], the L-curve method [2]–[5], generalized cross-validation (GCV) [6]–[17] and estimation of (weighted) mean-squared error (MSE) [18]–[30]. Recently, a new measure of image-quality (different from GCV and MSE) was introduced in [31] but its applicability has been demonstrated only for denoising applications [31].

In inverse problems, typically, image reconstruction is performed by minimizing a cost function composed of data-fidelity term and (one or more) regularization terms. Image quality in such cases is governed by regularization parameters that control the bias-variance trade-off (or equivalently, the balance between image-smoothing and amplification of noise) in the reconstruction. Using discrepancy principle requires minimizing the difference between the data-fidelity term and the noise variance [1] and can lead to over-smoothing [22]. In the L-curve method, parameters are chosen so as to maximize the curvature of a (L-shaped) parametric curve (constructed from the components of the cost function) [2]–[4]. This method can be computationally expensive and sensitive to curvature evaluation [5], [25]. GCV is a popular criterion used for parameter selection in a variety of inverse problems, especially for linear reconstruction algorithms [7]–[16]. The advantage of GCV is that it does not require knowledge of noise variance and is known to yield regularization parameters for linear algorithms that asymptotically minimize the true MSE [7]. Some extensions of GCV are also available for nonlinear algorithms [15]–[17] but they are computationally more involved (see Section III-A) than for linear algorithms.

MSE-estimation-based methods can be attractive alternatives to GCV since image quality is often quantified in terms of MSE in image reconstruction problems. For Gaussian noise, Stein's unbiased risk estimate (SURE) [18] provides a practical means of unbiasedly assessing MSE for denoising problems. Unlike GCV, SURE requires knowledge of noise statistics but is optimal even in the nonasymptotic regime. SURE has been successfully employed for optimally adjusting parameters of a variety of denoising algorithms [32]–[36]. For ill-posed inverse problems, it is not possible to estimate MSE (except in some special instances [22]–[25]) since $\mathbf{y}$ may only contain partial information about $\mathbf{x}$ [24, Sec. IV]. In such

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work was supported by the National Institutes of Health under Grant P01 CA87634.

*Sathish Ramani, Zhihao Liu, Jeffrey Rosen, and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan. Jon-Fredrik Nielsen is with the fMRI Laboratory, University of Michigan, Ann Arbor, MI, U.S.A. Email: {sramani, zhihao, jayar, jfnielse, fessler}@umich.edu.

cases, the principles underlying SURE may be extended to estimate weighted variants of MSE (e.g., by evaluating the error only on components of $\mathbf{x}$ that are accessible from $\mathbf{y}$) [19], [25], [26]. Several weighted SURE-type approaches have been proposed and employed for (near) optimal parameter tuning in ill-posed inverse problems, e.g., linear restoration [19], nonlinear noniterative restoration [26], image reconstruction using sparse priors [29], [30],noniterative parallel magnetic resonance image (MRI) reconstruction [27], nonlinear restoration [25], [28] and nonlinear image upsampling [25] using iterative shrinkage-thresholding type algorithms that specifically apply to synthesis formulations [25], [28], [37], [38] of image reconstruction problems. Synthesis formulations preclude popular regularization criteria such as total variation (TV) and smooth edge-preserving regularizers (e.g., Huber [39], smoothed-Laplacian [40]) that belong to the class of analysis formulations. Bayesian methods [41]–[45] have been employed for parameter tuning in image restoration problems involving analysis-type quadratic regularizers [41], [42] and TV [43]–[45].

This paper focusses on computing the nonlinear version of GCV (denoted by NGCV) [16], [17] and weighted SURE-type measures [20], [24] for nonlinear iterative reconstruction algorithms that can tackle a variety of nonquadratic regularization criteria including synthesis- and analysis-type (e.g., TV) regularizers. Both NGCV and weighted SURE-measures require the Jacobian matrix of the reconstruction operator (representative of the iterative algorithm) evaluated with respect to the data [16], [17], [24] (see Section III). We derive the desired Jacobian matrix for two types of computationally efficient algorithms: the contemporary split-Bregman (SB) algorithm [46] and IRLS-MIL [47], [48] that uses the matrix inversion lemma (MIL) to accelerate standard iterative reweighted least squares (IRLS) [47], [48]. Our work can be interpreted as an extension to previous research [25]–[30] that focussed on applying weighted SURE-type measures to inverse problems with noniterative algorithms [26], [27] and to iterative image reconstruction based on sparsity priors [29], [30] and synthesis formulations [25], [28].

In this paper, we compute Predicted-SURE [20], [21], Projected-SURE [24] and NGCV [16], [17] for nonlinear image restoration and MRI reconstruction (from partially sampled Cartesian k-space data) using TV and an analysis-type ℓ_1-regularization. We also illustrate using simulations (for image restoration and MRI reconstruction) and experiments with real data (for MRI reconstruction) that both Predicted-SURE and Projected-SURE provide near-MSE-optimal selection of regularization parameters in these applications. We also observed that NGCV yielded near-MSE-optimal selections for image restoration and slightly sub-optimal parameter values for MRI reconstruction.

The paper is organized as follows. Section II describes the problem mathematically and presents our notation and mathematical requisites essential for theoretical derivations. Section III briefly reviews (N)GCV and weighted SURE-type measures. Section IV describes in detail the derivation of Jacobian matrices for the considered algorithms. We present experimental results for image restoration and MRI reconstruction in Section V and discuss reconstruction quality and memory / computational requirements of the considered algorithmsin Section VI. Finally, we draw our conclusions in Section VII.

II. Notation and Problem Description

We use the following linear data model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\xi}, \tag{1}$$

that is appropriate for many imaging inverse problems including image restoration and MRI reconstruction from partially sampled Cartesian k-space data. In (1), $\mathbf{y} \in \Omega^M$ is the observed data, $\mathbf{A} \in \Omega^{M\times N}$ is a known (rectangular) matrix (typically $M \leq N$), and Ω is either $\mathbb{R}$ or $\mathbb{C}$ depending on the application. We assume $\mathbf{x} \in \Omega^N$ is an unknown deterministic quantity. For image restoration, $\Omega = \mathbb{R}$, $M = N$ and we assume that $\mathbf{A}$ is circulant, while for MRI with partial Cartesian k-space sampling,[1] $\Omega = \mathbb{C}$ and $\mathbf{A} = \mathbf{M}\mathbf{Q}$, where $\mathbf{Q} \in \mathbb{C}^{N\times N}$ is the orthonormal DFT matrix, $\mathbf{M}$ is the $M \times N$ downsampling matrix that satisfies $\mathbf{M}\mathbf{M}^\top = \mathbf{I}_M$ and $\mathbf{I}_M$ is the identity matrix of size M.

Throughout the paper, $(\cdot)^\top$ denotes the transpose of a real vector or matrix, $(\cdot)^\star$ denotes the complex conjugate, $(\cdot)^{\mathrm{H}}$ is the Hermitian-transpose, $(\cdot)_{\Re}$ and $(\cdot)_{\mathcal{I}}$ indicate the real and imaginary parts, respectively, of a complex vector or matrix. The m-th element of any vector $\mathbf{y}$ is denoted by y_m and the mn-th element of any matrix $\mathbf{A}$ is written as $[\mathbf{A}]_{mn}$.

For simplicity, we model $\boldsymbol{\xi} \in \Omega^M$ as an i.i.d. zero-mean Gaussian random vector with covariance matrix $\boldsymbol{\Lambda} = \sigma^2\mathbf{I}_M$ and probability density $g_\Omega(\boldsymbol{\xi})$. For $\Omega = \mathbb{R}$, $g_{\mathbb{R}}(\boldsymbol{\xi}) = (2\pi\sigma^2)^{-\frac{M}{2}}\exp(-\boldsymbol{\xi}^\top\boldsymbol{\xi}/2\sigma^2)$, while for $\Omega = \mathbb{C}$, we assume $\boldsymbol{\xi}$ is an i.i.d. complex Gaussian random vector (which is a reasonable model for MRI applications), so $g_{\mathbb{C}}(\boldsymbol{\xi}) = (\pi\sigma^2)^{-M}\exp(-\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\xi}/\sigma^2)$. SURE-type methods discussed in this paper (see Section III-B) can be readily extended to more general cases (such as $\boldsymbol{\xi}$ with non-zero mean and covariance $\boldsymbol{\Lambda} \neq \sigma^2\mathbf{I}_M$) using the generalized SURE (GSURE) methodology developed in [24].

Given data $\mathbf{y}$, we obtain an estimate of the unknown image $\mathbf{x}$ by minimizing a cost function based on (1) composed of a data-fidelity term and some regularization that is designed using "smoothness" penalties or prior information about $\mathbf{x}$:

$$\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) \triangleq \arg\min_{\mathbf{u}} \left\{ \mathcal{J}(\mathbf{u}) \triangleq \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \Psi(\mathbf{u}) \right\}, \tag{2}$$

where $\|\cdot\|_2$ represents the Euclidean norm, Ψ represents a suitable regularizer that is (possibly nonsmooth, i.e., not differentiable everywhere) convex and $\mathbf{u}_{\boldsymbol{\theta}} : \Omega^M \to \Omega^N$ may be interpreted as a (possibly nonlinear) mapping or an algorithm, representative of the minimization in (2), that acts on $\mathbf{y}$ to yield the estimate $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$. In practice, the mapping $\mathbf{u}_{\boldsymbol{\theta}}$ depends on one or more parameters $\boldsymbol{\theta}$ that need to be set appropriately to obtain a meaningful estimate $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$. In problems such as (2), typically, $\boldsymbol{\theta} = \lambda$ is a scalar known as the regularization

[1]Partial k-space sampling on Cartesian grids is relevant for accelerating 3-D MR acquisition in practice, where undersampling is typically applied in the phase-encode plane [49].

parameter that plays a crucial role in balancing the data-fidelity and regularization terms: small λ-values can lead to noisy estimates while a large λ results in over smoothing and loss of details. Quantitative criteria such as GCV [6], [16], [17] and (weighted) SURE-type measures [24], [25] can be used for tuning $\boldsymbol{\theta}$ of a nonlinear $\mathbf{u}_{\boldsymbol{\theta}}$, but they require the evaluation of the Jacobian matrix [16], [17], [24], $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y}) \in \Omega^{N \times M}$ for $\Omega = \mathbb{R}, \mathbb{C}$ (see Sections III-A and III-B), consisting of partial derivatives of the elements $\{u_{\boldsymbol{\theta},n}(\mathbf{y})\}_{n=1}^{N}$ of $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ with respect to $\{y_m\}_{m=1}^{M}$.

Definition 1: Let $\mathbf{u}_{\boldsymbol{\theta}} : \mathbb{R}^M \to \mathbb{R}^N$ be differentiable (in the weak sense of distributions [50, Ch. 6]). The Jacobian matrix $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y}) \in \mathbb{R}^{N \times M}$ evaluated at $\mathbf{y} \in \mathbb{R}^M$ is specified using its elements as

$$[\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})]_{nm} \triangleq \left. \frac{\partial u_{\boldsymbol{\theta},n}(\mathbf{z})}{\partial z_m} \right|_{\mathbf{z}=\mathbf{y}}. \tag{3}$$

Definition 2: Let $\mathbf{u}_{\boldsymbol{\theta}} : \mathbb{C}^M \to \Omega^N$ (with $\Omega = \mathbb{R}$ or $\mathbb{C}$) be individually analytic [51] with respect to $\mathbf{y}_{\Re}$ and $\mathbf{y}_{\mathcal{I}}$ (in the weak sense of distributions [50, Ch. 6]). The Jacobian matrices $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$, $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y}^\star) \in \mathbb{C}^{N \times M}$ are specified using their respective elements as [51, Eq. 13], [52]

$$[\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})]_{nm} \triangleq \frac{1}{2} \left(\frac{\partial u_{\boldsymbol{\theta},n}(\mathbf{z})}{\partial z_{\Re m}} - \imath \frac{\partial u_{\boldsymbol{\theta},n}(\mathbf{z})}{\partial z_{\mathcal{I}m}} \right) \Bigg|_{\mathbf{z}=\mathbf{y}}, \tag{4}$$

$$[\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y}^\star)]_{nm} \triangleq \frac{1}{2} \left(\frac{\partial u_{\boldsymbol{\theta},n}(\mathbf{z})}{\partial z_{\Re m}} + \imath \frac{\partial u_{\boldsymbol{\theta},n}(\mathbf{z})}{\partial z_{\mathcal{I}m}} \right) \Bigg|_{\mathbf{z}=\mathbf{y}}. \tag{5}$$

Remark 1: When $\mathbf{u}_{\boldsymbol{\theta}} : \mathbb{C}^M \to \Omega^N$ is prescribed in terms of $\mathbf{y}$ and $\mathbf{y}^\star$, $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$ is evaluated treating $\mathbf{y}$ as a variable and $\mathbf{y}^\star$ as a constant [52], [53]. Similarly, $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y}^\star)$ is evaluated treating $\mathbf{y}$ as constant [52], [53].

For common (and some popular) instances of $\mathcal{J}$ in (2), $\mathbf{u}_{\boldsymbol{\theta}}$ satisfies the hypotheses in Definitions 1 and 2 and in turn allows the computation of GCV and weighted SURE-type measures for reliable tuning of $\boldsymbol{\theta}$ as illustrated in our experiments.

III. Generalized Cross-Validation and Weighted SURE-Type Measures

A. Generalized Cross-Validation (GCV)

GCV is based on the "leave-one-out" principle [7] that leads to a simple expression in the case of linear algorithms: for a generic linear mapping $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) = \mathbf{F}_{\boldsymbol{\theta}}\mathbf{y}$, the GCV measure (denoted by LGCV) is given by [7]

$$\mathrm{LGCV}(\boldsymbol{\theta}) \triangleq \frac{M^{-1}\|(\mathbf{I}_M - \mathbf{A}\,\mathbf{F}_{\boldsymbol{\theta}})\mathbf{y}\|_2^2}{(1 - M^{-1}\mathrm{tr}\{\mathbf{A}\,\mathbf{F}_{\boldsymbol{\theta}}\}\})^2}. \tag{6}$$

For nonlinear estimators $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$, we consider the following GCV measure (denoted by NGCV)

$$\mathrm{NGCV}(\boldsymbol{\theta}) \triangleq \frac{M^{-1}\|\mathbf{y} - \mathbf{A}\,\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\|_2^2}{(1 - M^{-1}\Re\{\mathrm{tr}\{\mathbf{A}\,\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})\}\})^2}, \tag{7}$$

adapted from [17, Sec. 3] that was originally derived using the standard "leave-one-out" principle for nonlinear algorithms [16]. We take the real part, $\Re\{\cdot\}$, in the denominator of (7) specifically for the case of $\Omega = \mathbb{C}$ to avoid spurious complex entries while evaluating $\mathrm{NGCV}(\boldsymbol{\theta})$ numerically.

LGCV has been more widely used [7]–[11], [13], [14] (for linear algorithms) than NGCV (for nonlinear algorithms), perhaps because NGCV is computationally more involved than LGCV. Recently, Liao *et al.* proposed GCV-based automatic nonlinear restoration methods using alternating minimization in [54], [55]. Although their methods are nonlinear overall, they rely on linear sub-problems arising out of alternating minimization and employ LGCV for parameter tuning. In contrast, we propose to tackle NGCV (7) directly and demonstrate its use in nonlinear image restoration and MRI reconstruction.

B. Weighted SURE-Type Measures

In the context of image reconstruction, the mean-squared error (risk) measure,

$$\mathrm{MSE}(\boldsymbol{\theta}) \triangleq N^{-1}\|\mathbf{x} - \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\|_2^2 \tag{8}$$

is often used to assess image quality and is an attractive option for optimizing $\boldsymbol{\theta}$. However, $\mathrm{MSE}(\boldsymbol{\theta})$ cannot be directly computed since the cross-term $\mathbf{x}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ depends on the unknown $\mathbf{x}$ ($\|\mathbf{x}\|_2^2$ is an irrelevant constant independent of $\boldsymbol{\theta}$) and needs to be estimated in practice. For denoising applications, i.e., $\mathbf{A} = \mathbf{I}_N$ in (1), the desired cross-term can be manipulated as $\mathbf{x}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\xi})^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ and the statistics of $\boldsymbol{\xi}$ may then be used to estimate $\boldsymbol{\xi}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$. In the Gaussian setting, $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_N)$, Stein's result [18] (for $\Omega = \mathbb{R}$) can be used for this purpose and leads to $\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}^{\top}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\} = \sigma^2\mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{tr}\{\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})\}\}$, where $\mathsf{E}_{\boldsymbol{\xi}}\{\cdot\}$ represents expectation with respect to $\boldsymbol{\xi}$. Replacing $\boldsymbol{\xi}^{\top}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ in $\mathrm{MSE}(\boldsymbol{\theta})$ with $\sigma^2\,\mathrm{tr}\{\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})\}$ thus yields the so-called Stein's unbiased risk estimate (SURE) [18],

$$\begin{aligned} \mathrm{SURE}(\boldsymbol{\theta}) \triangleq\ & N^{-1}\|\mathbf{y} - \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\|_2^2 - \sigma^2 \\ & + 2\sigma^2 N^{-1}\mathrm{tr}\{\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})\}, \end{aligned} \tag{9}$$

that is an unbiased estimator of $\mathrm{MSE}(\boldsymbol{\theta})$, i.e., $\mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{MSE}(\boldsymbol{\theta})\} = \mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{SURE}(\boldsymbol{\theta})\}$. The accuracy of $\mathrm{SURE}(\boldsymbol{\theta})$ generally increases with N (law of large numbers), so it is appealing for image-processing applications (where N is large, typically $N \geq 256^2$) [36]. Using $\mathrm{SURE}(\boldsymbol{\theta})$ as a practical alternative to $\mathrm{MSE}(\boldsymbol{\theta})$ requires (in addition to σ^2) the evaluation of $\mathrm{tr}\{\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})\}$ that can be performed analytically for some special types of *denoising* algorithms [32]–[35] or numerically using the Monte-Carlo method in [36, Th. 2] for a general (iterative / noniterative) *denoising* algorithm $\mathbf{u}_{\boldsymbol{\theta}}$.

For inverse problems modeled by (1), $\mathbf{x}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ can be manipulated in terms of $\mathbf{y}$ (and $\boldsymbol{\xi}$ and thus allows the estimation of $\mathrm{MSE}(\boldsymbol{\theta})$ using statistics of $\boldsymbol{\xi}$) only in some special instances, e.g., when $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) \in \mathsf{R}\{\mathbf{A}^{\mathrm{H}}\}$, the range space of $\mathbf{A}^{\mathrm{H}}$ [24, Sec. IV],[2] or when $\mathbf{A}$ has full column rank [25, Sec. 4].[2] In many applications, $\mathbf{A}$ has a nontrivial null-space $\mathsf{N}\{\mathbf{A}\}$: information about $\mathbf{x}$ contained in $\mathsf{N}\{\mathbf{A}\}$ is not accessible from $\mathbf{y}$ (and statistics of $\boldsymbol{\xi}$) and it is impossible to estimate $\mathrm{MSE}(\boldsymbol{\theta})$ [24] in such cases. An alternative is to compute the error using only the components of $\mathbf{x}$ that lie in the orthogonal complement of $\mathsf{N}(\mathbf{A})$: $\mathsf{N}(\mathbf{A})^{\perp} = \mathsf{R}\{\mathbf{A}^{\mathrm{H}}\}$ [24], [25]; these components

[2] If $\mathbf{u}_{\boldsymbol{\theta}} \in \mathsf{R}\{\mathbf{A}^{\mathrm{H}}\}$, we can write $\mathbf{u}_{\boldsymbol{\theta}} = \mathbf{A}^{\mathrm{H}}\mathbf{g}_{\boldsymbol{\theta}}$ for some operator $\mathbf{g}_{\boldsymbol{\theta}}$, so that $\mathbf{x}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\xi})^{\mathrm{H}}\mathbf{g}_{\boldsymbol{\theta}}(\mathbf{y})$ [23, Sec. 3.1]. Alternatively, if $\mathbf{A}$ has full column rank, then $\mathbf{x}^{\mathrm{H}}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\xi})^{\mathrm{H}}\mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$.

are in turn accessible from $\mathbf{y}$ (and $\boldsymbol{\xi}$). Such an error measure corresponds to projecting[3] the error $(\mathbf{x}-\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}))$ on to $\mathsf{R}\{\mathbf{A}^{\mathrm{H}}\}$ and is given by [24], [25]

$$\text{Projected-MSE}(\boldsymbol{\theta}) \triangleq M^{-1}\|\mathbf{P}(\mathbf{x}-\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}))\|_2^2, \tag{10}$$

where $\mathbf{P}=\mathbf{A}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}})^{\dagger}\mathbf{A}$ is the projection operator and $(\cdot)^{\dagger}$ represents pseudo-inverse.

Another quadratic error measure for inverse problems that is amenable to estimation (using statistics of $\boldsymbol{\xi}$)[3] is Predicted-MSE [20], [21] that corresponds to computing the error in the data-domain:

$$\text{Predicted-MSE}(\boldsymbol{\theta}) \triangleq M^{-1}\|\mathbf{A}(\mathbf{x}-\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}))\|_2^2. \tag{11}$$

Both (10) and (11) can be interpreted as particular instances of the following general weighted form:

$$\text{WMSE}(\boldsymbol{\theta}) \triangleq M^{-1}\|\mathbf{A}(\mathbf{x}-\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y}))\|_{\mathbf{W}}^2, \tag{12}$$

where $\|\mathbf{x}\|_{\mathbf{W}}^2 \triangleq \mathbf{x}^{\mathrm{H}}\mathbf{W}\mathbf{x}$, and $\mathbf{W}$ is a Hermitian-symmetric, $\mathbf{W}=\mathbf{W}^{\mathrm{H}}$, positive definite, $\mathbf{W}\succ\mathbf{0}$, weighting matrix. For (11), $\mathbf{W}=\mathbf{I}_M$ and the overall weighting is provided by the eigenvalues of $\mathbf{A}^{\mathrm{H}}\mathbf{A}$, while for (10) it is easy to see that $\mathbf{W}=\mathbf{W}_{\mathrm{inv}} \triangleq (\mathbf{A}\mathbf{A}^{\mathrm{H}})^{\dagger}$ since $\mathbf{P}^{\mathrm{H}}\mathbf{P}=\mathbf{P}$. For image restoration with circulant $\mathbf{A}$, $\mathbf{W}_{\mathrm{inv}}$ can be easily implemented using FFTs. For MRI with partial Cartesian k-space sampling, $\mathbf{A}=\mathbf{M}\mathbf{Q}$ (see Section II) leads to $\mathbf{W}_{\mathrm{inv}}=(\mathbf{M}\mathbf{Q}\mathbf{Q}^{\mathrm{H}}\mathbf{M}^{\mathrm{H}})^{\dagger}=\mathbf{I}_M$, so Projected-MSE and Predicted-MSE are equivalent and correspond to evaluating squared-error at the sample locations in k-space.

Similar to SURE($\boldsymbol{\theta}$), an estimator for WMSE($\boldsymbol{\theta}$) can be derived under the Gaussian assumption as summarized in the following results.

Lemma 1: Let $\mathbf{u}_{\boldsymbol{\theta}}:\Omega^M\to\Omega^N$ be differentiable (for $\Omega=\mathbb{R}$) or individually analytic (for $\Omega=\mathbb{C}$ with respect to real and imaginary parts of its argument), respectively, in the weak sense of distributions [50, Ch. 6]. Then, for any deterministic $\mathbf{T}\in\Omega^{M\times N}$ satisfying $\mathsf{E}_{\boldsymbol{\xi}}\{|[\mathbf{T}\,\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})]_m|\}<\infty$, $m=1\dots M$, we have that

$$\mathsf{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}^{\mathrm{H}}\mathbf{T}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\}=\sigma^2\mathsf{E}_{\boldsymbol{\xi}}\{\mathrm{tr}\{\mathbf{T}\,\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}\}. \tag{13}$$

The proof is very similar to those in [34, Lemma 1] [24, Th. 1] for $\Omega=\mathbb{R}$, while it constitutes a straightforward extension of [24, Th. 1] for $\Omega=\mathbb{C}$ and is presented as supplementary material (due to page limits).[4]

Theorem 1: Let $\mathbf{u}_{\boldsymbol{\theta}}$ and $\mathbf{T}=\mathbf{W}\mathbf{A}$ satisfy the hypotheses in Lemma 1 for $\mathbf{A}\in\Omega^{M\times N}$ in (1) and for a Hermitian-symmetric, positive definite matrix $\mathbf{W}\in\Omega^{M\times M}$. Then for $\Omega=\mathbb{R}$ or $\mathbb{C}$, the random variable

$$\begin{aligned}\text{WSURE}(\boldsymbol{\theta}) \;\triangleq\; & M^{-1}\|\mathbf{y}-\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})\|_{\mathbf{W}}^2-\frac{\sigma^2}{M}\mathrm{tr}\{\mathbf{W}\}\\ & +\frac{2\sigma^2}{M}\Re\{\mathrm{tr}\{\mathbf{W}\,\mathbf{A}\,\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}\}\end{aligned} \tag{14}$$

is an unbiased estimator of WMSE($\boldsymbol{\theta}$) in (12), i.e., $\mathsf{E}_{\boldsymbol{\xi}}\{\text{WMSE}(\boldsymbol{\theta})\}=\mathsf{E}_{\boldsymbol{\xi}}\{\text{WSURE}(\boldsymbol{\theta})\}$.

The proof is straightforward and uses (13) to estimate $\boldsymbol{\xi}^{\mathrm{H}}\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ in WMSE($\boldsymbol{\theta}$). Similar to SURE($\boldsymbol{\theta}$), WSURE($\boldsymbol{\theta}$) is independent of $\mathbf{x}$ and depends purely on the noise variance σ^2, the data and the reconstruction algorithm. The Monte-Carlo scheme [36, Th. 2] that uses numerical differentiation for a general nonlinear $\mathbf{u}_{\boldsymbol{\theta}}$ may be adapted to iteratively estimate $\mathrm{tr}\{\mathbf{W}\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ in (14) for the case of $\Omega=\mathbb{R}$ by considering $\mathbf{W}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}$ instead of $\mathbf{u}_{\boldsymbol{\theta}}$ in [36, Eq. 14]. In this paper, we propose to evaluate $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$ analytically for $\Omega=\mathbb{R}$ and $\mathbb{C}$. This process depends on the choice of the estimator $\mathbf{u}_{\boldsymbol{\theta}}$, the regularization Ψ in (2) and the nature of application (e.g., $\Omega=\mathbb{R}$ for restoration and $\Omega=\mathbb{C}$ for MRI), and therefore needs to be accomplished on a case-by-case basis.

IV. Evaluation of the Jacobian Matrix $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$

For nonquadratic regularizers, there is no closed-form expression for the estimator $\mathbf{u}_{\boldsymbol{\theta}}$ in (2), so it is not possible to evaluate $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$ in (14) directly. In this section, we show how to compute $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$ recursively for two types of iterative algorithms used for minimizing $\mathcal{J}$ in (2). Henceforth, we leave implicit the dependence of $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ on $\mathbf{y}$ and drop the subscript $\boldsymbol{\theta}$ when necessary, so that $\mathbf{u}$ represents either the estimator or the iteratively-reconstructed estimate depending on the context.

We focus on IRLS-MIL [47], [48] that is a fast variant of the standard iterative reweighted least-squares (IRLS) and the split-Bregman (SB) algorithm [46] that is based on variable splitting. Both algorithms are computationally efficient,[5] and can be employed for image restoration and MRI reconstruction [46]–[48], [56]. Furthermore, they can accommodatea general class of regularization criteria of the form

$$\Psi(\mathbf{R}\mathbf{u})=\lambda\sum_{l=1}^{L}\Phi_l\left(\sum_{p=1}^{P}|[\mathbf{R}_p\,\mathbf{u}]_l|^q\right), \tag{15}$$

where $\lambda>0$ is the regularization parameter, Φ_l are potential functions, $\mathbf{R}\in\mathbb{R}^{R\times N}$ with $\mathbf{R}\triangleq[\mathbf{R}_1^{\top}\cdots\mathbf{R}_P^{\top}]^{\top}$, $\mathbf{R}_p\in\mathbb{R}^{L\times N}$ are regularization operators (e.g., finite differences, frames, etc.) and $R=PL$. We consider the following convex instances of (15) that are popularly used for image restoration and MRI reconstruction:

- Analysis ℓ_1-regularization ($\Phi_l(x)=x$, $q=1$):

$$\Psi_{\ell_1}(\mathbf{R}\mathbf{u}) \triangleq \lambda\|\mathbf{R}\,\mathbf{u}\|_1=\lambda\sum_{p=1}^{P}\sum_{l=1}^{L}|[\mathbf{R}_p\,\mathbf{u}]_l|\,, \tag{16}$$

- Total variation (TV) ($\Phi_l(x)=\sqrt{x}$, $q=2$):

$$\Psi_{\mathrm{TV}}(\mathbf{R}\mathbf{u}) \triangleq \lambda\sum_{l=1}^{L}\sqrt{\sum_{p=1}^{P}|[\mathbf{R}_p\,\mathbf{u}]_l|^2}. \tag{17}$$

We derive $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$ for image restoration with the IRLS-MIL algorithm (see Sections IV-A–IV-B) and MRI reconstruction

[3]Since $\mathbf{P}^{\mathrm{H}}\mathbf{P}=\mathbf{P}$, the cross-term $\mathbf{x}^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\mathbf{P}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$ in Projected-MSE($\boldsymbol{\theta}$) (10) is nothing but $\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}})^{\dagger}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})=(\mathbf{y}-\boldsymbol{\xi})^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}})^{\dagger}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$. For Predicted-MSE($\boldsymbol{\theta}$) (11), we have that $\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})=(\mathbf{y}-\boldsymbol{\xi})^{\mathrm{H}}\mathbf{A}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y})$.

[4]Supplementary material containing a proof of Lemma 1 and additional illustrations for experimental results is available at http://tinyurl.com/supmat.

[5]IRLS-MIL has been demonstrated to converge faster than conventional methods (e.g., nonlinear conjugate gradient) [47], [48], while SB is more versatile and computationally efficient than fixed-point continuation and graph-cuts-based solvers [46].

with the SB algorithm (see Sections IV-C–IV-D). Derivations of $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$ for other combinations (i.e., image restoration with the SB algorithm and MRI reconstruction with the IRLS-MIL algorithm) can be accomplished in a similar manner and are not considered for brevity.

The derivations in Sections IV-A–IV-D can also be extended, in principle, to other instances of (15) such as smooth convex edge-preserving regularizers for $q = 1$, e.g., Huber [39], Fair or smoothed-Laplacian [48] and synthesis forms, e.g., by considering a variable $\boldsymbol{w}$ to be estimated such that $\mathbf{x} = \mathbf{S}\boldsymbol{w}$ and $\mathbf{AS}\boldsymbol{w}$ in (1) and $\Phi(\boldsymbol{w})$ in (15), for some potential function Φ and synthesis operator $\mathbf{S}$ [25], [28], [37], [38].

A. Image Restoration with IRLS-MIL Algorithm

IRLS-MIL uses matrix-splitting [57, pp. 46-50] and the matrix inversion lemma (MIL) for efficient preconditioning and fast solving of iteration-dependent linear systems arising in the standard IRLS scheme [47], [48]. We summarize the IRLS-MIL iterations below (detailed derivation of IRLS-MIL can be found in [47], [48]) for image restoration ($\Omega = \mathbb{R}$ in (1)): for any matrix $\mathbf{C}$ such that $\mathbf{C} \succ \mathbf{A}^\top\mathbf{A}$, i.e.,

$$\boldsymbol{z}^\top(\mathbf{C} - \mathbf{A}^\top\mathbf{A})\,\boldsymbol{z} > 0,\ \forall\ \boldsymbol{z} \in \mathbb{R}^N, \tag{18}$$

at outer i-iteration of IRLS-MIL, we perform inner j-iterations involving the following two-steps:

$$\mathbf{u}^{(i+1,j+1)} = \mathbf{b}^{(i+1,j)} - \mathbf{C}^{-1}\,\mathbf{R}^\top\,\mathbf{v}^{(i+1,j)}, \tag{19}$$

$$\text{solve}\{\mathbf{G}^{(i)}\mathbf{v}^{(i+1,j)} = \mathbf{R}\,\mathbf{b}^{(i+1,j)}\} \text{ for } \mathbf{v}^{(i+1,j)}, \tag{20}$$

where

$$\mathbf{b}^{(i+1,j)} \triangleq \mathbf{C}^{-1}\mathbf{A}^\top\mathbf{y} + (\mathbf{I}_N - \mathbf{C}^{-1}\mathbf{A}^\top\mathbf{A})\mathbf{u}^{(i+1,j)}, \tag{21}$$

$\mathbf{G}^{(i)} \triangleq \boldsymbol{\Gamma}^{(i)} + \mathbf{R}\ \mathbf{C}^{-1}\mathbf{R}^\top$ and $\boldsymbol{\Gamma}^{(i)} \triangleq \text{diag}\{\boldsymbol{\gamma}^{(i)}\}$ is a diagonal matrix constructed from $\boldsymbol{\gamma}^{(i)} \in \mathbb{R}^R$ with $\boldsymbol{\gamma}^{(i)} \triangleq [\boldsymbol{\gamma}_1^{(i)\top} \cdots \boldsymbol{\gamma}_P^{(i)\top}]^\top$. The l-th element of $\boldsymbol{\gamma}_p^{(i)} \in \mathbb{R}^L$ for (15) is given by $\gamma_{pl}^{(i)} =$

$$\left|[\mathbf{R}_p\,\mathbf{u}^{(i+1,0)}]_l\right|^{2-q}\left[q\lambda\,\Phi_l'\left(\sum_{s=1}^{P}\left|[\mathbf{R}_s\,\mathbf{u}^{(i+1,0)}]_l\right|^q\right)\right]^{-1}, \tag{22}$$

where Φ' denotes the first derivative of Φ. For instance,

$$\gamma_{pl}^{(i)} = \lambda^{-1}|[\mathbf{R}_p\,\mathbf{u}^{(i+1,0)}]_l|, \tag{23}$$

for the ℓ_1-regularization in (16), while for TV-regularization (17),

$$\boldsymbol{\gamma}^{(i)} = \mathbf{1}_P \otimes \breve{\boldsymbol{\gamma}}^{(i)}, \tag{24}$$

where $\mathbf{1}_P = [1 \cdots 1]^\top$ is a $P \times 1$ vector, $\otimes$ denotes the Kronecker product and the l-th element of $\breve{\boldsymbol{\gamma}}^{(i)} \in \mathbb{R}^L$ is given by $\breve{\gamma}_l^{(i)} = \lambda^{-1}\sqrt{\sum_{s=1}^{P}\left|[\mathbf{R}_s\,\mathbf{u}^{(i+1,0)}]_l\right|^2}$.

When $\mathbf{R}$ is composed of sparsifying operators (e.g., finite differences, wavelets, frames, etc.), $\mathbf{R}_p\mathbf{u}^{(i,0)}$ tends to become sparse for sufficiently large i, so for practically appealing instances of Ψ (e.g., when $1 \leq m < 2$ and Φ is an edge-preserving potential function or for nonsmooth instances such as those based on the ℓ_1-norm or TV), $\boldsymbol{\Gamma}^{(i)}$ will become sparse too. So in the standard IRLS scheme (which utilizes $[\boldsymbol{\Gamma}^{(i)}]^{-1}$ rather than $\boldsymbol{\Gamma}^{(i)}$ [47, Eqs. 3-6]), a small positive additive constant is included in $\boldsymbol{\Gamma}^{(i)}$ for maintaining numerical stability of $[\boldsymbol{\Gamma}^{(i)}]^{-1}$—this is often referred to as corner-rounding [47]. However no such corner-rounding is required for the IRLS-MIL scheme [47], [48] as it only utilizes $\boldsymbol{\Gamma}^{(i)}$.

To solve (20), we apply a matrix-splitting strategy (similar to [47, Eq. 7]) to $\mathbf{G}^{(i)}$ that leads to an iterative scheme for (20) (with iteration index k) with the following update step:

$$\mathbf{v}^{(i+1,j,k+1)} = \mathbf{D}_{\boldsymbol{\Gamma}^{(i)}}^{-1}(\mathbf{R}\,\mathbf{b}^{(i+1,j)} + \mathbf{H}_\rho\mathbf{v}^{(i+1,j,k)}), \tag{25}$$

where $\mathbf{D}_{\boldsymbol{\Gamma}^{(i)}} \triangleq \boldsymbol{\Gamma}^{(i)} + \rho\mathbf{I}_R$ is an invertible diagonal matrix, $\mathbf{H}_\rho \triangleq \rho\mathbf{I}_R - \mathbf{R}\ \mathbf{C}^{-1}\mathbf{R}^\top$ and

$$\rho > \max\ \text{eigval}\{\mathbf{R}\ \mathbf{C}^{-1}\mathbf{R}^\top\} \tag{26}$$

depends only on $\mathbf{A}$ (via $\mathbf{C}$) and $\mathbf{R}$ and can therefore be precomputed. In practice, we perform K iterations of (25) and apply the final update $\mathbf{v}^{(i+1,j,K)}$ in place of $\mathbf{v}^{(i+1,j)}$ in (19). We prefer (25) over a PCG-type solver for (20) as (25) is linear in both $\mathbf{b}^{(\cdot)}$ and $\mathbf{v}^{(\cdot)}$ and decouples the shift-variant component $\boldsymbol{\Gamma}^{(i)}$ from the rest of terms in $\mathbf{G}^{(i)}$: these features simplify the analytical derivation of $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$ for IRLS-MIL as demonstrated next.

B. Jacobian Matrix Derivation for IRLS-MIL Algorithm

Since $\mathbf{b}^{(\cdot)}$ and $\mathbf{v}^{(\cdot)}$ are functions of $\mathbf{y}$ (via $\mathbf{u}^{(\cdot)}$), using linearity of (3), at the end of K iterations of (25), we get the Jacobian matrix update corresponding to $\mathbf{u}^{(i+1,j+1)}$ from (19) as

$$\mathbf{J}(\mathbf{u}^{(i+1,j+1)}, \mathbf{y}) = \mathbf{J}(\mathbf{b}^{(i+1,j)}, \mathbf{y}) - \mathbf{C}^{-1}\mathbf{R}^\top\ \mathbf{J}(\mathbf{v}^{(i+1,j,K)}, \mathbf{y}). \tag{27}$$

From (21), we get that $\mathbf{J}(\mathbf{b}^{(i+1,j)}, \mathbf{y}) = \mathbf{C}^{-1}\mathbf{A}^\top + (\mathbf{I}_N - \mathbf{C}^{-1}\mathbf{A}^\top\mathbf{A})\ \mathbf{J}(\mathbf{u}^{(i+1,j)}, \mathbf{y})$. To obtain $\mathbf{J}(\mathbf{v}^{(i+1,j,K)}, \mathbf{y})$, we derive a recursive update for $\mathbf{J}(\mathbf{v}^{(i+1,j,k+1)}, \mathbf{y})$ from (25) using a straightforward application of product rule for Jacobian matrices and the fact that (25) involves only a diagonal matrix $\mathbf{D}_{\boldsymbol{\Gamma}^{(i)}}^{-1}$:

$$\begin{aligned}\mathbf{J}(\mathbf{v}^{(i+1,j,k+1)}, \mathbf{y}) = \mathbf{D}_{\boldsymbol{\Gamma}^{(i)}}^{-1}\Big(&\mathbf{R}\ \mathbf{J}(\mathbf{b}^{(i+1,j)}, \mathbf{y}) \\ &+\mathbf{H}_\rho\ \mathbf{J}(\mathbf{v}^{(i+1,j,k)}, \mathbf{y})\Big) \\ &-\mathbf{D}_{\boldsymbol{\Gamma}^{(i)}}^{-2}\,\mathbf{D}_{\mathbf{v}}\ \mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{y}),\end{aligned} \tag{28}$$

where $\mathbf{D}_{\mathbf{v}} \triangleq \text{diag}\{\mathbf{R}\,\mathbf{b}^{(i+1,j)} + \mathbf{H}_\rho\mathbf{v}^{(i+1,j,k)}\}$. Using chain rule for Jacobian matrices [52], we have that

$$\mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{y}) = \mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{u}^{(i+1,0)})\ \mathbf{J}(\mathbf{u}^{(i+1,0)}, \mathbf{y}), \tag{29}$$

where $\mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{u}^{(i+1,0)}) \in \mathbb{R}^{R\times N}$ constitutes derivatives of $\{\gamma_l^{(i)}\}_{l=1}^R$ with respect to $\{u_n\}_{n=1}^N$, evaluated at $\mathbf{u}^{(i+1,0)}$ and can be computed readily analytically for the regularizers in (15) using (22). For the ℓ_1-regularization[6] Ψ_{ℓ_1} in (16),

$$\mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{u}^{(i+1,0)}) = \lambda^{-1}\text{diag}\{\boldsymbol{\tau}^{(i)}\}\mathbf{R}, \tag{30}$$

[6]The derivatives are interpreted in the weak sense of distributions [50, Ch. 6] whenever conventional differentiation does not apply.

where the elements of $\boldsymbol{\tau}^{(i)} \in \mathbb{R}^R$ are $\tau_l^{(i)} \triangleq \operatorname{sign}([\mathbf{R}\mathbf{u}^{(i+1,0)}]_l)$.

For TV-regularization Ψ_{TV} in (17), we get (using tedious, but elementary calculus) that

$$\mathbf{J}(\boldsymbol{\gamma}^{(i)}, \mathbf{u}^{(i+1,0)}) = \mathbf{1}_P \otimes \left(\lambda^{-1} \sum_{p=1}^{P} \operatorname{diag}\{\boldsymbol{\omega}_p^{(i)}\} \mathbf{R}_p \right), \quad (31)$$

where the elements of $\boldsymbol{\omega}_p^{(i)} \in \mathbb{R}^L$ are given by

$$\omega_{pl}^{(i)} \triangleq \frac{[\mathbf{R}_p \mathbf{u}^{(i+1,0)}]_l}{\sqrt{\sum_{s=1}^{P} |[\mathbf{R}_s \mathbf{u}^{(i+1,0)}]_l|^2}} = \frac{[\mathbf{R}_p \mathbf{u}^{(i+1,0)}]_l}{\lambda \breve{\gamma}_l^{(i)}}, \quad (32)$$

with the assumption that $\omega_{pl}^{(i)} = 0$ whenever $[\mathbf{R}_p \mathbf{u}^{(i+1,0)}]_l = 0$, $p = 1 \ldots P$.

Thus, in addition to running (19), (25) for obtaining $\mathbf{u}^{(i,j)}$, we propose to run the sequence of iterations (27)-(28) using (29)-(32) for iteratively evaluating $\mathbf{J}(\mathbf{u}^{(i,j)}, \mathbf{y})$, (and thus, NGCV($\boldsymbol{\theta}$) and WSURE($\boldsymbol{\theta}$)) at any stage of IRLS-MIL.

C. MRI Reconstruction with Split-Bregman Algorithm

The split-Bregman (SB) algorithm [46] for solving (2) is based on a variable splitting strategy [46], [49], [56], where an auxiliary variable $\mathbf{v} \in \mathbb{C}^R$ is used to artificially introduce the constraint $\mathbf{v} = \mathbf{R}\mathbf{u}$ and separate $\mathbf{R}\mathbf{u}$ from Ψ leading to

$$\min_{\mathbf{u},\mathbf{v}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \Psi(\mathbf{v}) \text{ subject to } \mathbf{v} = \mathbf{R}\mathbf{u} \quad (33)$$

that is equivalent to (2). The above constrained problem is then solved using the so-called Bregman-iterations [46, Eqs. (2.9)-(2.10)] that consists of alternating between the minimization of an augmented Lagrangian (AL) function,[7] $\mathcal{L}(\mathbf{u}, \mathbf{v}) \triangleq \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \Psi(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v} - \mathbf{R}\mathbf{u} - \boldsymbol{\eta}\|_2^2$, jointly with respect to $(\mathbf{u}, \mathbf{v})$ [46, Eq. (3.7)] and a simple update step [46, Eq. (3.8)] for a Lagrange-multiplier-like vector, $\boldsymbol{\eta} \in \mathbb{C}^R$, for the constraint in (33). The penalty parameter $\mu > 0$ does not influence the final solution of (33) and (2), but governs the convergence speed of the Bregman-iterations [46]. In practice, the joint-minimization step is often replaced by alternating minimization [46, Sec. 3.1], i.e., $\mathcal{L}$ is minimized with respect to $\mathbf{u}$ and $\mathbf{v}$ one at a time, which decouples the minimization step and simplifies optimization. We summarize the SB algorithm below for solving (33) (and equivalently (2), for MRI reconstruction):

$$\mathbf{u}^{(i+1)} = \mathbf{B}_\mu^{-1}[\mathbf{A}^{\mathrm{H}}\mathbf{y} + \mu \mathbf{R}^\top(\mathbf{v}^{(i)} - \boldsymbol{\eta}^{(i)})], \quad (34)$$

$$\mathbf{v}^{(i+1)} = \mathbf{d}_\Psi(\boldsymbol{\varrho}^{(i)}) \triangleq \arg\min_{\mathbf{v}} \left\{ \Psi(\mathbf{v}) + \frac{\mu}{2} \|\mathbf{v} - \boldsymbol{\varrho}^{(i)}\|_2^2 \right\}, \quad (35)$$

$$\boldsymbol{\eta}^{(i+1)} = \boldsymbol{\varrho}^{(i)} - \mathbf{v}^{(i+1)}, \quad (36)$$

where $\mathbf{B}_\mu \triangleq \mathbf{A}^{\mathrm{H}}\mathbf{A} + \mu \mathbf{R}^\top \mathbf{R}$. Step (35) corresponds to a denoising problem to which we associate a denoising operator $\mathbf{d}_\Psi : \mathbb{C}^R \to \mathbb{C}^R$ that acts on

$$\boldsymbol{\varrho}^{(i)} \triangleq \mathbf{R}\mathbf{u}^{(i+1)} + \boldsymbol{\eta}^{(i)} \quad (37)$$

to yield $\mathbf{v}^{(i+1)}$. For a general Ψ such as (15), (35) can be solved iteratively in which case $\mathbf{d}_\Psi$ is representative of the iterative scheme used for (35). However, for several special instances of Ψ [58, Sec. 4] including (16)-(17), one can solve (35) exactly and $\mathbf{d}_\Psi$ admits an analytical closed-form expression. Specifically for (16)-(17), it can be shown that (35) further decouples in terms of the elements $\{v_r\}_{r=1}^R$ of $\mathbf{v}$, i.e., $\mathbf{d}_\Psi$ is a point-wise operator such that $v_r^{(i+1)} = d_\Psi(\varrho_r^{(i)})$, $r = 1 \ldots R$.

Before proceeding, it is helpful to introduce $\boldsymbol{\beta}^{(i)} \triangleq \mathbf{1}_P \otimes \breve{\boldsymbol{\beta}}^{(i)}$, where the l-th element of $\breve{\boldsymbol{\beta}}^{(i)} \in \mathbb{R}^L$ is given by $\breve{\beta}_l^{(i)} \triangleq \sqrt{\sum_{p=1}^{P} |\varrho_{(p-1)L+l}^{(i)}|^2}$.Then d_Ψ for (16)-(17) can be obtained as [46], [49]:

$$d_{\ell_1}(\varrho_r^{(i)}) = \varrho_r^{(i)} \left(1 - \lambda\, |\mu\, \varrho_r^{(i)}|^{-1}\right) \mathbb{1}(|\varrho_r^{(i)}| > \lambda/\mu) \quad (38)$$

$$d_{\mathrm{TV}}(\varrho_r^{(i)}) = \varrho_r^{(i)} \left(1 - \lambda\, (\mu\, \beta_r^{(i)})^{-1}\right) \mathbb{1}(\beta_r^{(i)} > \lambda/\mu) \quad (39)$$

where $\mathbb{1}(\cdot)$ is an indicator function that takes the value 1 when the condition in its argument is satisfied and is 0 otherwise.

D. Jacobian Matrix Derivation for Split-Bregman Algorithm

We note that $\mathbf{u}^{(\cdot)}$, $\mathbf{v}^{(\cdot)}$, $\boldsymbol{\eta}^{(\cdot)}$ and $\boldsymbol{\varrho}^{(\cdot)}$ are implicit functions of $\mathbf{y}$ and $\mathbf{y}^\star$. Therefore, we evaluate[8] $\mathbf{J}(\mathbf{u}^{(i)}, \mathbf{y})$ using (34) and the linearity of (4) to get that

$$\mathbf{J}(\mathbf{u}^{(i+1)}, \mathbf{y}) = \mathbf{B}_\mu^{-1}[\, \mathbf{A}^{\mathrm{H}} + \mu \mathbf{R}^\top (\, \mathbf{J}(\mathbf{v}^{(i)}, \mathbf{y}) - \mathbf{J}(\boldsymbol{\eta}^{(i)}, \mathbf{y})\,)\,]. \quad (40)$$

For the complex-valued case ($\Omega = \mathbb{C}$), we also need to evaluate $\mathbf{J}(\mathbf{u}^{(i+1)}, \mathbf{y}^\star)$ as explained next. For $\mathbf{J}(\mathbf{u}^{(i+1)}, \mathbf{y}^\star)$, we treat $\mathbf{y}$ as a constant in (34) (see Remark 1) so that

$$\mathbf{J}(\mathbf{u}^{(i+1)}, \mathbf{y}^\star) = \mu \mathbf{B}_\mu^{-1}\, \mathbf{R}^\top (\, \mathbf{J}(\mathbf{v}^{(i)}, \mathbf{y}^\star) - \mathbf{J}(\boldsymbol{\eta}^{(i)}, \mathbf{y}^\star)\,). \quad (41)$$

For brevity, henceforth we use $\mathbf{z}$ to represent either $\mathbf{y}$ or $\mathbf{y}^\star$ as required. From (35)-(36), we have that

$$\mathbf{J}(\mathbf{v}^{(i+1)}, \mathbf{z}) = \mathbf{J}(\mathbf{d}_\Psi(\boldsymbol{\varrho}^{(i)}), \mathbf{z}), \quad (42)$$

$$\mathbf{J}(\boldsymbol{\eta}^{(i+1)}, \mathbf{z}) = \mathbf{J}(\boldsymbol{\varrho}^{(i)}, \mathbf{z}) - \mathbf{J}(\mathbf{v}^{(i+1)}, \mathbf{z}). \quad (43)$$

Using chain rule for $\mathbf{J}(\mathbf{d}_\Psi(\boldsymbol{\varrho}^{(i)}), \mathbf{z})$ [53, Th. 1], we get that

$$\mathbf{J}(\mathbf{d}_\Psi(\boldsymbol{\varrho}^{(i)}), \mathbf{z}) = \mathbf{J}(\mathbf{d}_\Psi, \boldsymbol{\varrho}^{(i)})\, \mathbf{J}(\boldsymbol{\varrho}^{(i)}, \mathbf{z}) + \mathbf{J}(\mathbf{d}_\Psi, \boldsymbol{\varrho}^{(i)\star})\, (\mathbf{J}(\boldsymbol{\varrho}^{(i)}, \mathbf{z}^\star))^\star, \quad (44)$$

where $\mathbf{J}(\boldsymbol{\varrho}^{(i)}, \mathbf{z}) = \mathbf{R}\, \mathbf{J}(\mathbf{u}^{(i+1)}, \mathbf{z}) - \mathbf{J}(\boldsymbol{\eta}^{(i)}, \mathbf{z})$ from (37). Thus, due to (44), both $\mathbf{J}(\mathbf{u}^{(\cdot)}, \mathbf{y})$ and $\mathbf{J}(\mathbf{u}^{(\cdot)}, \mathbf{y}^\star)$ are required as mentioned earlier.

For the case of ℓ_1-regularization (16), d_{ℓ_1} in (38) depends only on $\varrho_r^{(i)}$ and $\varrho_r^{(i)\star}$, so $\mathbf{J}(\mathbf{d}_{\ell_1}, \boldsymbol{\varrho}^{(i)})$ and $\mathbf{J}(\mathbf{d}_{\ell_1}, \boldsymbol{\varrho}^{(i)\star})$ become diagonal matrices:

$$\mathbf{J}(\mathbf{d}_{\ell_1}, \boldsymbol{\varrho}^{(i)}) = \mathbf{D}_{\mathbb{1},\boldsymbol{\varrho}^{(i)}} \left(\mathbf{I}_R - \frac{\lambda}{2\mu} \mathbf{D}_{|\boldsymbol{\varrho}^{(i)}|}^{-1} \right), \quad (45)$$

$$\mathbf{J}(\mathbf{d}_{\ell_1}, \boldsymbol{\varrho}^{(i)\star}) = \frac{\lambda}{2\mu}\, \mathbf{D}_{\mathbb{1},\boldsymbol{\varrho}^{(i)}}\, \mathbf{D}_{\boldsymbol{\varrho}^{(i)}}^2 \mathbf{D}_{|\boldsymbol{\varrho}^{(i)}|}^{-3}, \quad (46)$$

[7] For (33), $\mathcal{L}$ is equivalent to the sum of the Bregman-distance [46] and a quadratic penalty term for the constraint in (33) up to irrelevant constants.

[8] The derivatives are interpreted in the weak sense of distributions [50, Ch. 6] whenever conventional differentiation does not apply.

where

$$\mathbf{D}_{\mathbb{1},\boldsymbol{\varrho}^{(i)}} \triangleq \operatorname{diag}\left\{\left\{\mathbb{1}(|\varrho_r^{(i)}|>\lambda/\mu)\right\}_{r=1}^{R}\right\}, \quad (47)$$

$$\mathbf{D}_{\boldsymbol{\varrho}^{(i)}} \triangleq \operatorname{diag}\{\boldsymbol{\varrho}^{(i)}\}, \quad (48)$$

$$\mathbf{D}_{|\boldsymbol{\varrho}^{(i)}|} \triangleq \operatorname{diag}\left\{\left\{|\varrho_r^{(i)}|\right\}_{r=1}^{R}\right\}. \quad (49)$$

For the case of TV-regularization (17), we apply (tedious, but elementary) product rule to obtain

$$\mathbf{J}(\mathbf{d}_{\mathrm{TV}},\boldsymbol{\varrho}^{(i)}) = \mathbf{D}_{\mathbb{1},\boldsymbol{\beta}^{(i)}}\left[\mathbf{I}_R-\frac{\lambda}{\mu}\mathbf{D}^{-1}_{\boldsymbol{\beta}^{(i)}} + \frac{\lambda}{\mu}\mathbf{D}_{\boldsymbol{\varrho}^{(i)}}\mathbf{D}^{-2}_{\boldsymbol{\beta}^{(i)}}\left(\mathbf{1}_P\otimes\mathbf{J}(\breve{\boldsymbol{\beta}}^{(i)},\boldsymbol{\varrho}^{(i)})\right)\right], \quad (50)$$

$$\mathbf{J}(\mathbf{d}_{\mathrm{TV}},\boldsymbol{\varrho}^{(i)\star}) = \frac{\lambda}{\mu}\mathbf{D}_{\mathbb{1},\boldsymbol{\beta}^{(i)}}\mathbf{D}_{\boldsymbol{\varrho}^{(i)}}\mathbf{D}^{-2}_{\boldsymbol{\beta}^{(i)}} \times\left(\mathbf{1}_P\otimes(\mathbf{J}(\breve{\boldsymbol{\beta}}^{(i)},\boldsymbol{\varrho}^{(i)}))^{\star}\right), \quad (51)$$

where

$$\mathbf{J}(\breve{\boldsymbol{\beta}}^{(i)},\boldsymbol{\varrho}^{(i)}) = \frac{1}{2}(\mathbf{1}_P^{\top}\otimes\mathbf{D}^{-1}_{\breve{\boldsymbol{\beta}}^{(i)}})\,\mathbf{D}^{\star}_{\boldsymbol{\varrho}^{(i)}}, \quad (52)$$

$$\mathbf{D}_{\boldsymbol{\beta}^{(i)}} \triangleq \operatorname{diag}\{\boldsymbol{\beta}^{(i)}\}, \quad (53)$$

$$\mathbf{D}_{\mathbb{1},\boldsymbol{\beta}^{(i)}} \triangleq \operatorname{diag}\left\{\left\{\mathbb{1}(\beta_r^{(i)}>\lambda/\mu)\right\}_{r=1}^{R}\right\}, \quad (54)$$

$$\mathbf{D}_{\breve{\boldsymbol{\beta}}^{(i)}} \triangleq \operatorname{diag}\{\breve{\boldsymbol{\beta}}^{(i)}\}. \quad (55)$$

Thus, similar to the case of IRLS-MIL, we propose to run (34)-(36) for obtaining $\mathbf{u}^{(i)}$ and (40)-(55) for iteratively evaluating $\mathbf{J}(\mathbf{u}^{(i)},\mathbf{y})$ (and thus, NGCV($\boldsymbol{\theta}$) and WSURE($\boldsymbol{\theta}$)) at any stage of the SB algorithm.

E. Monte-Carlo Trace Estimation

The Jacobian matrices $\mathbf{J}(\cdot,\cdot)$ in Sections IV-B and IV-D have enormous sizes for typical reconstruction settings and cannot be stored and manipulated directly to compute the desired traces, $\mathrm{tr}\{\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ in (7) and $\mathrm{tr}\{\mathbf{W}\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ in (14). So we use a Monte-Carlo method to estimate $\mathrm{tr}\{\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ and $\mathrm{tr}\{\mathbf{W}\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ that is based on the following well-established identity [12], [59]–[61].

Proposition 1: Let $\mathbf{n}\in\mathbb{R}^M$ be an i.i.d. zero-mean random vector with unit variance and $\mathbf{T}\in\Omega^{M\times M}$ (for $\Omega=\mathbb{R}$ or $\mathbb{C}$) be deterministic with respect to $\mathbf{n}$. Then

$$\mathsf{E}_{\mathbf{n}}\{\mathbf{n}^{\top}\mathbf{T}\,\mathbf{n}\} = \mathrm{tr}\{\mathbf{T}\}. \quad (56)$$

For practical applications, $\mathsf{E}_{\mathbf{n}}\{\cdot\}$ in (56) can be replaced by sample mean, $1/N_r\sum_{i=1}^{N_r}\mathbf{n}_i^{\top}\mathbf{T}\,\mathbf{n}_i$, with N_r independent realizations $\{\mathbf{n}_i\}_{i=1}^{N_r}$. In image-processing applications where, typically, M is large and $\mathbf{T}$ has a sparse off-diagonal structure, $\hat{t}_{\mathbf{T}} \triangleq \mathbf{n}^{\top}\mathbf{T}\,\mathbf{n}$ (corresponding to $N_r=1$) provides a reliable estimate of $\mathrm{tr}\{\mathbf{T}\}$ [11], [12], [25], [28]. To use this type of stochastic estimation for $\mathrm{tr}\{\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$ and $\mathrm{tr}\{\mathbf{W}\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\}$, we adopt the procedure applied in [25], [28]: we take products with $\mathbf{n}$ in (27)-(29), (40)-(44) and store and update vectors of the form $\mathbf{J}(\mathbf{u}^{(\cdot)},\cdot)\mathbf{n}$, $\mathbf{J}(\mathbf{v}^{(\cdot)},\cdot)\mathbf{n}$, $\mathbf{J}(\boldsymbol{\eta}^{(\cdot)},\cdot)\mathbf{n}$ in IRLS-MIL and SB algorithms, respectively. At any point during the course of the algorithms, the desired traces in (7) and (14) are stochastically approximated, respectively, as

$$\mathrm{tr}\{\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\} \approx \hat{t}_{\mathrm{NGCV}} \triangleq \mathbf{n}^{\top}\mathbf{A}\,\mathbf{J}(\mathbf{u}^{(\cdot)},\mathbf{y})\mathbf{n}, \quad (57)$$

$$\mathrm{tr}\{\mathbf{W}\mathbf{A}\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})\} \approx \hat{t}_{\mathrm{WSURE}} \triangleq \mathbf{n}^{\top}\mathbf{W}\,\mathbf{A}\,\mathbf{J}(\mathbf{u}^{(\cdot)},\mathbf{y})\mathbf{n}. \quad (58)$$

To improve accuracy of (57)-(58), $\mathbf{n}$ can be designed to decrease the variance of $\hat{t}_{\mathrm{NGCV}}$ and $\hat{t}_{\mathrm{WSURE}}$: it has been shown [59], [61] that variance of a Monte-Carlo trace estimate (such as $\hat{t}_{\mathrm{NGCV}}$ or $\hat{t}_{\mathrm{WSURE}}$) is lower for a binary random vector $\mathbf{n}_{\pm1}$ whose elements are either $+1$ or -1 with probability 0.5 than for a Gaussian random vector $\mathbf{n}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_M)$ employed in [25], [28]. So in our experiments, we used one realization of $\mathbf{n}_{\pm1}$ in (57)-(58). Figs. 1, 2 present outlines for implementing IRLS-MIL and SB algorithms with recursions for $\mathbf{J}(\cdot,\cdot)\mathbf{n}_{\pm1}$ to compute and monitor NGCV($\boldsymbol{\theta}$) and WSURE($\boldsymbol{\theta}$) as these algorithms evolve.

F. Implementation of IRLS-MIL and Split-Bregman Algorithms

The convergence speed of IRLS-MIL (19), (25) depends primarily on the "proximity" of $\mathbf{C}$ to $\mathbf{A}^{\top}\mathbf{A}$ while ensuring (18) [47], [48]. Ideally, we would like to choose the circulant matrix $\mathbf{C}_{\mathrm{opt}}=\mathbf{Q}^{\mathrm{H}}\mathrm{diag}\{\boldsymbol{\alpha}_{\mathrm{opt}}\}\mathbf{Q}$, where $\mathbf{Q}$ is the DFT matrix and $\boldsymbol{\alpha}_{\mathrm{opt}} \triangleq \arg\min_{\mathrm{diag}\{\boldsymbol{\alpha}\}\succeq\mathbf{Q}\mathbf{A}^{\top}\mathbf{A}\mathbf{Q}^{\mathrm{H}}}|||\mathrm{diag}\{\boldsymbol{\alpha}\}-\mathbf{Q}\mathbf{A}^{\top}\mathbf{A}\mathbf{Q}^{\mathrm{H}}|||$ for some matrix norm $|||\cdot|||$, e.g., the Frobenius norm. However, $\boldsymbol{\alpha}_{\mathrm{opt}}$ can be both challenging and computationally expensive to obtain for a general $\mathbf{A}$. For image restoration, typically, $\mathbf{A}^{\top}\mathbf{A}\in\mathbb{R}^{N\times N}$ is circulant, so $\boldsymbol{\alpha}_{\mathrm{opt}}$ is simply the eigenvalues of $\mathbf{A}^{\top}\mathbf{A}$. In our experiments, we used $\mathbf{C}=\mathbf{C}_{\nu}\triangleq\mathbf{A}^{\top}\mathbf{A}+\nu\mathbf{I}_N$ and implemented $\mathbf{C}_{\nu}^{-1}$ using FFTs. The parameter $\nu>0$ was chosen to achieve a prescribed condition number of $\mathbf{C}_{\nu}$, $\kappa(\mathbf{C}_{\nu})$, that can be easily computed as a function of ν. In general, setting $\kappa(\mathbf{C}_{\nu})$ to a large value can lead to numerical instabilities in $\mathbf{C}_{\nu}^{-1}$ and IRLS-MIL, while a small $\kappa(\mathbf{C}_{\nu})$ reduces convergence speed of IRLS-MIL [47], [48]. In our experiments, we found that ν leading to $\kappa(\mathbf{C}_{\nu})\in[20,100]$ yielded good convergence speeds for a fixed number of outer (i.e., index by i) iterations of IRLS-MIL, so we simply set ν such that $\kappa(\mathbf{C}_{\nu})=100$.

For MRI reconstruction from partially sampled Cartesian k-space data, $\mathbf{A}^{\mathrm{H}}\mathbf{A}\in\mathbb{C}^{N\times N}$ is circulant [49]. We chose $\{\mathbf{R}_p\}_{p=1}^{P}$ in (16)-(17) to be shift-invariant with periodic boundary extensions so that $\mathbf{R}^{\top}\mathbf{R}$, and thus $\mathbf{B}_{\mu}$ in (34), are circulant as well. Then we implemented $\mathbf{B}_{\mu}^{-1}$ in (34) using FFTs. One way to select the penalty parameter μ for the SB algorithm is to minimize the condition number $\kappa(\mathbf{B}_{\mu})$ of $\mathbf{B}_{\mu}$: $\mu=\mu_{\min}\triangleq\arg\min_{\mu}\kappa(\mathbf{B}_{\mu})$ [46]. We found in our experiments that the empirical selection $\mu=\mu_{\min}\times\mu_{\mathrm{factor}}$ with $\mu_{\mathrm{factor}}\in[10^{-5},10^{-2}]$ yielded favorable convergence speeds of the SB algorithm for a fixed number of iterations compared to using $\mu_{\min}$, so we set $\mu_{\mathrm{factor}}=10^{-4}$ throughout.

V. Experimental Results

A. Experimental Setup

In all our experiments, we focussed on tuning the regularization parameter λ (16)-(17) for a fixed number of (outer)

1) Initialization: $\mathbf{u}^{(0,0)} \triangleq \mathbf{A}^\top \mathbf{y}$, $\mathbf{J}(\mathbf{u}^{(0,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{A}^\top \mathbf{n}$, $i = 0$
2) **Repeat Steps 3-12 until Stop Criterion is met**
3) If $i = 0$
   $\mathbf{u}^{(i+1,0)} = \mathbf{u}^{(i,0)}$, $\mathbf{v}^{(i+1,0,0)} = \mathbf{R}\mathbf{u}^{(i,0)}$, $\mathbf{J}(\mathbf{u}^{(i+1,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{J}(\mathbf{u}^{(i,0)},\mathbf{y})\mathbf{n}$, $\mathbf{J}(\mathbf{v}^{(i+1,0,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{R}\mathbf{J}(\mathbf{u}^{(i,0)},\mathbf{y})\mathbf{n}$
   Else
   $\mathbf{u}^{(i+1,0)} = \mathbf{u}^{(i,J)}$, $\mathbf{v}^{(i+1,0,0)} = \mathbf{v}^{(i,J-1,K)}$, $\mathbf{J}(\mathbf{u}^{(i+1,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{J}(\mathbf{u}^{(i,J)},\mathbf{y})\mathbf{n}$, $\mathbf{J}(\mathbf{v}^{(i+1,0,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{J}(\mathbf{v}^{(i,J-1,K)},\mathbf{y})\mathbf{n}$
4) Compute $\mathbf{\Gamma}^{(i)}$ using (22); set $j = 0$
5) **Run J iterations of Steps 6-10**
6) Compute $\mathbf{b}^{(i+1,j)}$ (21) and $\mathbf{J}(\mathbf{b}^{(i+1,j)},\mathbf{y})\mathbf{n}$
7) If $j > 0$ set $\mathbf{v}^{(i+1,j,0)} = \mathbf{v}^{(i+1,j-1,K)}$ and $\mathbf{J}(\mathbf{v}^{(i+1,j,0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{J}(\mathbf{v}^{(i+1,j-1,K)},\mathbf{y})\mathbf{n}$
8) Run K iterations of (25) and (28) to get $\mathbf{v}^{(i+1,j,K)}$ and $\mathbf{J}(\mathbf{v}^{(i+1,j,K)},\mathbf{y})\mathbf{n}$
9) Compute $\mathbf{u}^{(i+1,j+1)}$ (19) and $\mathbf{J}(\mathbf{u}^{(i+1,j+1)},\mathbf{y})\mathbf{n}$ (27)
10) Set j = j+1 and return to Step 5
11) Compute NGCV($\boldsymbol{\theta}$) and / or WSURE($\boldsymbol{\theta}$) at iteration i using (57)-(58), (7) and (14), respectively
12) Set $i = i + 1$ and return to Step 2

Fig. 1. Iterative computation of WSURE($\boldsymbol{\theta}$) and NGCV($\boldsymbol{\theta}$) for image deblurring using IRLS-MIL algorithm (with J iterations of (19)-(25) and K iterations of (25)). We use a pregenerated binary random vector $\mathbf{n} = \mathbf{n}_{\pm 1}$ for Monte-Carlo computation (57)-(58) of the required traces in (7) and (14), respectively. Vectors of the form $\mathbf{J}(\cdot^{(\cdot)},\cdot)\mathbf{n}$ are stored and manipulated in place of actual matrices $\mathbf{J}(\cdot^{(\cdot)},\cdot)$.

1) Initialization: $\mathbf{u}^{(0)} \triangleq \mathbf{A}^{\mathrm{H}}\mathbf{y}$, $\mathbf{v}^{(0)} \triangleq \mathbf{R}\mathbf{u}^{(0)}$, $\boldsymbol{\eta}^{(0)} \triangleq \mathbf{0}$, $i = 0$
   $\mathbf{J}(\mathbf{u}^{(0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{A}^{\mathrm{H}}\mathbf{n}$, $\mathbf{J}(\mathbf{v}^{(0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{R}\mathbf{A}^{\mathrm{H}}\mathbf{n}$, $\mathbf{J}(\boldsymbol{\eta}^{(0)},\mathbf{y})\mathbf{n} \triangleq \mathbf{0}$, $\mathbf{J}(\mathbf{u}^{(0)},\mathbf{y}^\star)\mathbf{n} \triangleq \mathbf{0}$, $\mathbf{J}(\mathbf{v}^{(0)},\mathbf{y}^\star)\mathbf{n} \triangleq \mathbf{0}$, $\mathbf{J}(\boldsymbol{\eta}^{(0)},\mathbf{y}^\star)\mathbf{n} \triangleq \mathbf{0}$,
2) **Repeat** Steps 3-7 until Stop Criterion is met
3) Compute $\mathbf{u}^{(i+1)}$, $\mathbf{J}(\mathbf{u}^{(i+1)},\mathbf{y})\mathbf{n}$, $\mathbf{J}(\mathbf{u}^{(i+1)},\mathbf{y}^\star)\mathbf{n}$, respectively, using (34), (40), (41)
4) Compute $\mathbf{v}^{(i+1)}$ using (35), (38)-(39) and $\mathbf{J}(\mathbf{v}^{(i+1)},\mathbf{y})\mathbf{n}$, $\mathbf{J}(\mathbf{v}^{(i+1)},\mathbf{y}^\star)\mathbf{n}$, respectively, using (42), (44)-(55)
5) Compute $\boldsymbol{\eta}^{(i+1)}$, $\mathbf{J}(\boldsymbol{\eta}^{(i+1)},\mathbf{y})\mathbf{n}$, $\mathbf{J}(\boldsymbol{\eta}^{(i+1)},\mathbf{y}^\star)\mathbf{n}$, respectively, using (36) and (43)
6) Compute NGCV($\boldsymbol{\theta}$) and / or WSURE($\boldsymbol{\theta}$) at iteration i using (57)-(58), (7) and (14), respectively
7) Set $i = i + 1$ and return to Step 2

Fig. 2. Iterative computation of WSURE($\boldsymbol{\theta}$) and NGCV($\boldsymbol{\theta}$) for MRI reconstruction with split-Bregman algorithm. We use a pregenerated binary random vector $\mathbf{n} = \mathbf{n}_{\pm 1}$ for Monte-Carlo computation (57)-(58) of the required traces in (7), (14), respectively. Vectors of the form $\mathbf{J}(\cdot^{(\cdot)},\cdot)\mathbf{n}$ are stored and manipulated in place of actual matrices $\mathbf{J}(\cdot^{(\cdot)},\cdot)$.

iterations for both IRLS-MIL and SB algorithms, although in principle, we can apply the greedy method[9] of Giryes *et al.* [25, Sec. 5.2] to minimize WSURE and NGCV as functions of both the number of iterations and λ. For IRLS-MIL, we used $J = K = 1$ (J iterations of (19)-(25) and K iterations of (25), see Fig. 1) and set the maximum number of iterations (indexed by i) to 100 for both algorithms. We used 2 levels of the undecimated Haar wavelet transform (excluding the approximation level) for $\mathbf{R}$ in Ψ_{ℓ_1} (16) and horizontal and vertical finite differences for $\{\mathbf{R}_p\}_{p=1}^2$ in Ψ_{TV} (17), all with periodic boundary extensions.

Both NGCV (7) and WSURE (14) require the evaluation of $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}},\mathbf{y})$, therefore their computation costs are similar for a given reconstruction algorithm.We evaluated NGCV(λ) in (7) for image restoration and MRI reconstruction and the (oracle) MSE using (8). We assumed that σ^2 (the variance of noise in $\mathbf{y}$) was known[10] in all simulations to compute the following WSURE-based measures: Predicted-SURE(λ) with $\mathbf{W} = \mathbf{I}_M$ in (14) and Projected-SURE(λ) with $\mathbf{W} = (\mathbf{A}\mathbf{A}^{\mathrm{H}})^\dagger$ in (14) that correspond to Predicted-MSE(λ) (11) and Projected-MSE(λ) (10), respectively. For image restoration, we computed $\mathbf{W} = (\mathbf{A}\mathbf{A}^{\mathrm{H}})^\dagger$ for Projected-SURE using FFTs.[11] For MRI reconstruction from partially sampled Cartesian k-space data, Predicted-SURE and Projected-SURE are equivalent (since $\mathbf{W} = \mathbf{I}_M$, see Section III-B) and correspond to evaluating the error at sample locations in the k-space.

[9]The Jacobian matrix $\mathbf{J}(\mathbf{u}^{(\cdot)},\mathbf{y})$ is updated at every (outer) iteration of IRLS-MIL and SB algorithms (see Figs. 1, 2), so (57)-(58) can be used to monitor NGCV($\boldsymbol{\theta}$) and WSURE($\boldsymbol{\theta}$), respectively, during the course of the algorithms as elucidated in Figs. 1, 2.

[10]In practice, σ can be estimated fairly reliably using, e.g., the techniques proposed in [22, Sec. V].

[11]We set the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$ below a threshold of 10^{-5} to zero for numerical stability of $(\mathbf{A}\mathbf{A}^{\mathrm{H}})^\dagger$.

B. Results for Image Restoration

We performed three sets of experiments with simulated data corresponding to the setups (with standard blur kernels [37]) summarized in Table I. In each simulation, data was generated corresponding to a blur kernel and a prescribed BSNR (SNR of blurred and noisy data) [43]. IRLS-MIL was then applied for varying λ and the quality of the deblurred images was assessed by computing Projected-SURE(λ), Predicted-SURE(λ) and NGCV(λ). We also included the following GCV-measure adapted from [25, Eq. 11] in our tests:

$$\mathrm{RLGCV}(\lambda) \triangleq \frac{M^{-1}\|(\mathbf{y} - \mathbf{A}\mathbf{u}_\lambda(\mathbf{y})\|_2^2}{(1 - M^{-1}\mathbf{n}_\pm^\top \mathbf{A}\mathbf{u}_\lambda(\mathbf{n}_\pm))^2}. \tag{59}$$

where $\mathbf{n}_\pm$ is the binary random vector specified in Section IV-E. RLGCV in (59) is a randomized version of LGCV in (6) that applies to linear algorithms but has been suggested for use with nonlinear algorithms as well in [25]. We minimized these measures over λ using golden-section search and calculated the improvement in SNR (ISNR) [43] of corresponding deblurred images (after minimizing the various measures).

Tables II and III summarize the ISNR-results for Experiments IR-A and IR-B, respectively, for varying BSNR. Minimization of Projected-SURE(λ) yields deblurred images with ISNR (reasonably) close to the corresponding minimum-MSE (oracle) result in all cases. Surprisingly, data-domain predicted-type measures Predicted-SURE and NGCV, which

TABLE I
SETUP FOR IMAGE RESTORATION (IR) EXPERIMENTS

Experiment	Test image (256×256)	Blur	Regularization
IR-A	Cameraman	Uniform 9×9	Ψ_{TV}
IR-B	House	$(1 + x_1 + x_2)^{-1}$, $-7 \leq x_1, x_2 \leq 7$	Ψ_{ℓ_1}
IR-C	Cameraman	Uniform (with varying sizes)	Ψ_{TV}

TABLE II
ISNR† (IN dB) OF DEBLURRED IMAGES FOR EXPERIMENT IR-A AND VARYING BSNR

BSNR	σ^2	MSE (oracle)	Projected-SURE	Predicted-SURE	NGCV	RLGCV in (59)
20	3.08e+01	3.85	3.73	**3.84**	**3.84**	2.45
30	3.08e+00	5.85	**5.84**	**5.85**	**5.85**	2.40
40	3.08e-01	8.50	**8.50**	**8.49**	**8.49**	2.41
50	3.08e-02	11.02	**10.97**	**11.00**	**11.01**	2.38

TABLE III
ISNR† (IN dB) OF DEBLURRED IMAGES FOR EXPERIMENT IR-B AND VARYING BSNR

BSNR	σ^2	MSE (oracle)	Projected-SURE	Predicted-SURE	NGCV	RLGCV in (59)
20	1.65e+01	5.85	**5.80**	**5.83**	5.72	3.48
30	1.65e+00	8.49	**8.49**	**8.49**	**8.49**	2.94
40	1.65e-01	11.68	**11.68**	**11.67**	**11.63**	2.85
50	1.65e-02	16.00	15.76	15.76	15.76	2.85

TABLE IV
ISNR† (IN dB) OF DEBLURRED IMAGES FOR EXPERIMENT IR-C: UNIFORM BLUR OF VARYING SIZES AND BSNR $= 40$ dB

Blur size	σ^2	MSE (oracle)	Projected-SURE	Predicted-SURE	NGCV	RLGCV in (59)
5×5	3.36e-01	9.82	**9.82**	**9.74**	**9.74**	2.67
9×9	3.08e-01	8.50	**8.50**	**8.48**	**8.48**	2.41
15×15	2.78e-01	7.42	**7.38**	**7.42**	**7.42**	2.22
21×21	2.57e-01	6.86	**6.78**	**6.82**	**6.82**	2.33

† ISNR values within 0.1 dB of the oracle are indicated in bold in Tables II-IV.

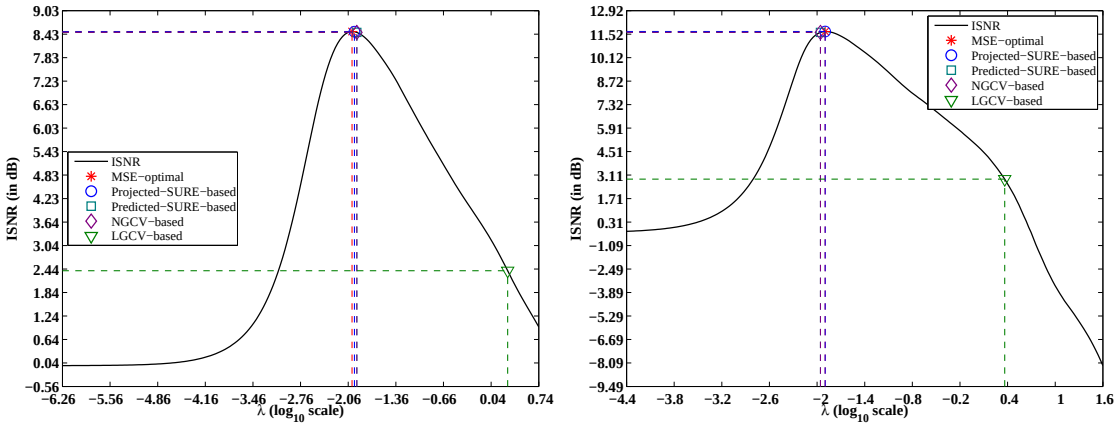


Fig. 3. Plot of ISNR(λ) as a function of regularization parameter λ. Left: Experiment IR-A corresponding to third row of Table II; Right: Experiment IR-B corresponding to third row of Table III. The plots indicate that λ's that minimize Projected-SURE, Predicted-SURE, NGCV and the (oracle) MSE are very close to each other. RLGCV-based selection (59) is far away from that of oracle MSE-based selection and leads to over-smoothing and loss of details, e.g., see Fig. 4g.

are known to undersmooth linear deblurring algorithms [22], [24], also consistently yield ISNRs that are remarkably near the corresponding oracle-ISNRs. These observations are also substantiated by Fig. 3 where we plot ISNR(λ) versus λ for specific instances of Experiments IR-A and IR-B: ISNRs corresponding to the optima of Projected-SURE(λ), Predicted-SURE(λ) and NGCV(λ) are close to the oracle-ISNR. Accordingly, the deblurred images (corresponding to an instance of Experiment IR-A) obtained by minimizing Projected-SURE(λ) (Fig. 4d), Predicted-SURE(λ) (Fig. 4e) and NGCV(λ) (Fig. 4f) closely resemble the corresponding minimum-MSE result (Fig. 4c) in terms of visual appearance. We present additional illustrations (for Experiments IR-A and IR-B) that corroborate these inferences as supplementary material.[4]

To further investigate the potential of Predicted-SURE and NGCV, we generated $\mathbf{y}$ corresponding to uniform blur of varying sizes (for a fixed BSNR of 40 dB: Experiment IR-C) and minimized the various measures (using golden-section search) in each case. The ISNR-results summarized in Table IV for this experiment indicate that minimization of Predicted-SURE(λ) and NGCV(λ) (and also Projected-SURE(λ)) lead to deblurred images with ISNRs close to that of the corresponding MSE-optimal ones. We obtained similar promising results at

Fig. 4. Experiment IR-A corresponding to third row of Table II: Zoomed images of (a) Noisefree Cameraman; (b) Blurred and noisy data; and TV-deblurred images with regularization parameter λ selected to minimize (c) (oracle) MSE (8.50 dB); (d) Projected-SURE (8.50 dB); (e) Predicted-SURE (8.49 dB); (f) NGCV (8.49 dB); (g) RLGCV in (59) (2.41 dB). Projected-SURE-, Predicted-SURE- and NGCV-based results (d)-(f) visually resemble the oracle MSE-based result (c) very closely, while the RLGCV-based (59) result is considerably over-smoothed.

varying (BSNR = 20, 30 dB) levels of noise (results not shown). These observations suggest that Predicted-SURE and NGCV may be reasonable alternatives to Projected-SURE for tuning λ for nonlinear restoration.

In all image-restoration experiments, the RLGCV-measure (59) yielded λ-values that were larger (by at least an order of magnitude, e.g., see Fig. 3) than corresponding oracle-optimum-λ, leading to over-smoothing and loss of details (e.g., see Fig. 4g), and thus reduced ISNR (see RLGCV-column in Tables II-IV). These results are perhaps due to the fact that RLGCV (based on LGCV in (6)) is primarily designed for linear algorithms and is therefore unable to cope with nonlinearity of (2) for the strongly nonquadratic regularizers in (16)-(17). On the contrary, NGCV (7), which is specifically designed to handle nonlinear algorithms [16], [17], provides a reliable means of selecting λ for nonlinear restoration. We do not show results for RLGCV hereafter.

C. Results for MRI Reconstruction

We conducted experiments with both synthetic and real MR data (setups summarized in Table V) for MRI reconstruction. In the synthetic case, we considered two test images (of size 256×256): the Shepp-Logan phantom (Experiments MRI-A) and a noisefree T_2-weighted MR image (Experiment MRI-B, see Fig. 6a) from the Brainweb database [62]. Partial sampling of k-space was simulated by applying a sampling mask (confined to a Cartesian grid)[12] on the Fourier transform (FFT) of test images. We considered two types of masks corresponding to a near uniform (less than Nyquist rate) but random sampling of k-space with a 8×8 fully sampled[13] central portion (see Fig. 6b) and radial patterns that densely sample the center[13] but sparsely sample the outer k-space (e.g., see Fig. 7b). Complex (i.i.d., zero-mean) Gaussian noise of appropriate variance was added at sample locations to simulate noisy data of prescribed SNR in Experiments MRI-A and MRI-B.

For experiments with real MR data, we acquired 10 in-

[12]Cartesian undersampling is more appropriate for 3-D MRI in practice and is applied here retrospectively for 2-D MRI for illustration purposes.

[13]Partial k-space sampling schemes typically involve dense sampling of the central portion (as that contains most of the signal energy) and undersampling of outer portions of k-space [49], respectively.

TABLE V
SETUP FOR EXPERIMENTS WITH SIMULATED AND REAL MR DATA

Experiment	Test image / Real MR data (256×256)	Retrospective (Cartesian) undersampling	Regularization
MRI-A	Shepp-Logan phantom	Radial (30 lines, 89% undersampling)	Ψ_{TV}
MRI-B	Noisefree T_2-weighted MR image	Random (60% undersampling)	Ψ_{ℓ_1}
MRI-C	Real GE-phantom dataset	Radial (with varying number of lines)	Ψ_{TV}

TABLE VI
PSNR† (IN dB) OF MRI RECONSTRUCTIONS FOR EXPERIMENT MRI-A AND VARYING DATA SNR

Data SNR (in dB)	σ^2	MSE (oracle)	Predicted-SURE	NGCV
30	2.69e+01	13.69	**13.66**	12.72
40	2.69e+00	22.28	**22.21**	21.68
50	2.69e-01	31.90	**31.86**	30.74
60	2.69e-02	42.33	**42.33**	42.12

TABLE VII
PSNR† (IN dB) OF MRI RECONSTRUCTIONS FOR EXPERIMENT MRI-B AND VARYING DATA SNR

Data SNR (in dB)	σ^2	MSE (oracle)	Predicted-SURE	NGCV
30	1.33e+01	7.77	7.33	7.33
40	1.33e+00	10.58	**10.53**	10.38
50	1.33e-01	11.62	**11.62**	**11.58**
60	1.33e-02	11.83	**11.81**	**11.83**

TABLE VIII
PSNR† (IN dB) OF MRI RECONSTRUCTIONS FOR EXPERIMENT MRI-C AND VARYING UNDERSAMPLING RATES

Number of radial lines	% undersampling	MSE (oracle)	Predicted-SURE	NGCV
20	93	26.19	**26.16**	25.82
30	89	30.08	**30.06**	29.42
40	85	31.71	**31.69**	31.09
50	82	33.03	**33.03**	32.16
60	78	33.65	**33.56**	32.85

† PSNR values within 0.1 dB of the oracle are indicated in bold in Tables VI-VIII.

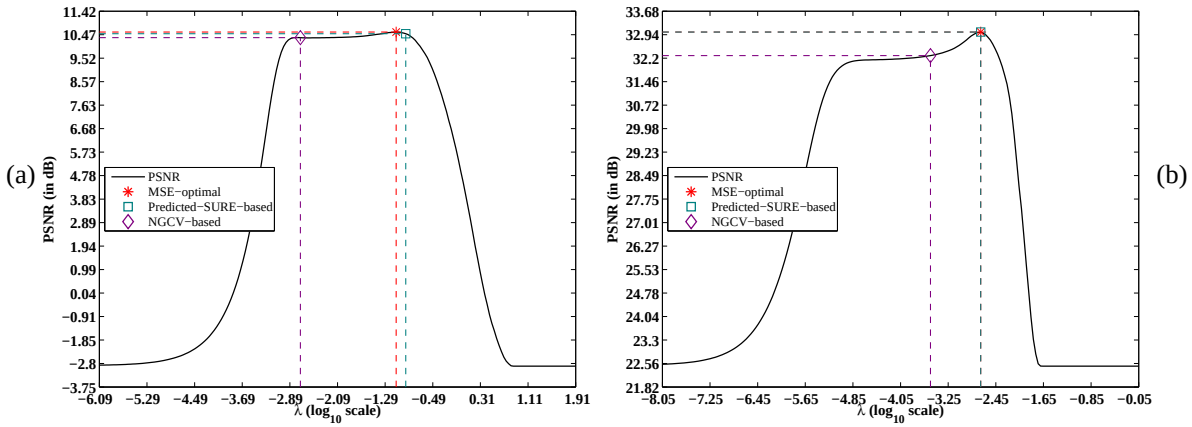


Fig. 5. Plot of PSNR(λ) as a function of regularization parameter λ: (a) Experiment MRI-B corresponding to second row of Table VII; (b) Experiment MRI-C corresponding to fourth row of Table VIII. The plots indicate that λ's that minimize Predicted-SURE and the (oracle) MSE are very close to each other and lead to almost identical PSNRs. NGCV-based selection is away from the MSE-based selection in both plots: in case of (a) it still yields a reconstruction Fig. 6f that is agreeably close to the oracle in terms of PSNR and visual quality Fig. 6d, but in (b) it leads to a slight reduction in PSNR and correspondingly the reconstruction Fig. 7f exhibits slightly more artifacts at the center and around the object.

dependent sets of fully-sampled 2-D data (256×256) of a GE-phantom using a GE 3T scanner (gradient-echo sequence with flip angle = 35°, repetition time = 200 ms, echo time = 7 ms, field of view (FOV) = 15 cm and voxel-size = 0.6×0.6 mm^2). These fully-sampled datasets were used to reconstruct (using iFFT) 2-D images that were then averaged to obtain a reference image that served as the true "unknown" $\mathbf{x}$ (see Fig. 7a) for computing the oracle MSE (8). We separately acquired 2-D data from a dummy scan (with the same scan setting) where no RF field was applied. We used this dummy-data to estimate σ^2 by the empirical variance. We retrospectively undersampled data from one of the 10 sets by applying radial sampling patterns (confined to a Cartesian grid)[13] with varying number of spokes in Experiment MRI-C.

We ran the SB algorithm and minimized Predicted-SURE(λ) and NGCV(λ) using golden-section search for each instance of Experiments MRI-A, MRI-B, and MRI-C, respectively. Tables VI-VIII present PSNR (computed as $20\log_{10}(\sqrt{N}\max\{\mathbf{x}\}/\|\mathbf{x}-\mathbf{u}_\lambda(\mathbf{y})\|_2)$ of reconstructions obtained after minimization of Predicted-SURE(λ) and

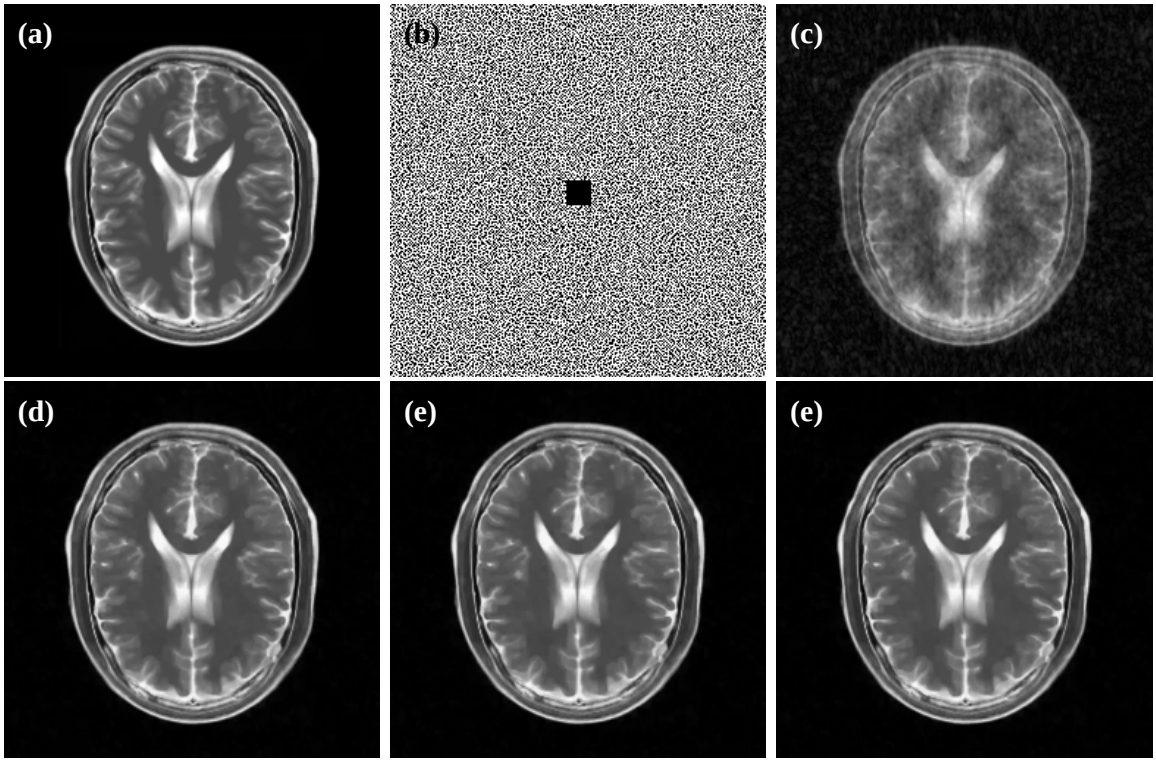


Fig. 6. Simulations corresponding to Experiment MRI-B and second row of Table VII: (a) Noisefree T_2-weighted MR test image; (b) Retrospective random undersampling (black dots indicate sample locations on a Cartesian grid, 60% undersampling); (c) Magnitude of zero-filled iFFT reconstruction from undersampled data (-2.88 dB db); and magnitude of reconstructions obtained (using analysis ℓ_1-regularization with 2 levels of undecimated Haar wavelet) with regularization parameter λ selected to minimize (d) (oracle) MSE (10.58 dB); (e) Predicted-SURE (10.53 dB); (f) NGCV (10.38 dB). Regularized reconstructions (d)-(f) have reduced noise and artifacts compared to the zero-filled iFFT reconstruction (c). Both Predicted-SURE-based and NGCV-based results (e) and (f) closely resemble the oracle MSE-based result (d) in this experiment.

NGCV(λ). In almost all experiments, NGCV-based selections resulted in worse PSNRs than those corresponding to Predicted-SURE-selections. This is also corroborated by Fig. 5 where we plot PSNR(λ) as a function of λ for specific instances of Experiment MRI-B and MRI-C. NGCV-based selections are away (approximately, by an order of magnitude) from both Predicted-SURE- and oracle-selections. As the PSNR-profile in Fig. 5a exhibits a plateau[14] over a large range of λ-values, NGCV-based reconstruction in Fig. 6d is visually similar to the corresponding minimum-MSE reconstruction in Fig. 6f. However, this is not the case with Fig. 5b and correspondingly, the NGCV-based reconstruction in Fig. 7f exhibits slightly more artifacts at the center and around the object's periphery compared to Predicted-SURE-based, Fig. 7e, and minimum-MSE, Fig. 7d, reconstructions. These results indicate that NGCV may not be as consistently robust for MRI reconstruction from partially sampled Cartesian data (which is a severely ill-posed problem where $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ has a lot of zero-eigenvalues) as for image restoration (where only fewer eigenvalues of $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ are zero, especially for the blurs considered in Section V-B).

On the other hand, Predicted-SURE-based tuning consistently yields PSNRs close to the corresponding (minimum-MSE) oracle-PSNRs as seen from Tables VII-VIII and Fig. 5. Predicted-SURE also leads to reconstructions (see Figs. 6e, 7e) that are visually similar to the respective minimum-MSE reconstructions (see Figs. 6d, 7d). These results demonstrate the potential of Predicted-SURE for selection of λ for MRI reconstruction.

[14]This is perhaps because the problem is less ill-posed in Experiment MRI-B as the k-space is sampled in a nearly uniform (but random) fashion (see Fig. 6b) compared to other setups, MRI-A and MRI-C, respectively, that use radial sampling (e.g., see Fig. 7b) where the corners of k-space are sparsely sampled.

VI. Discussion

A. Reconstruction Quality

Reconstruction quality in inverse problems of the form (1)-(2) is mainly governed by (*a*) the cost criterion $\mathcal{J}$ in (2), and (*b*) the choice of associated regularization parameter(s). In this work, we have only addressed the latter aspect, i.e., (*b*), for specific (but popular) regularizers such as TV and those based on the ℓ_1-norm. As we achieve near-MSE-optimal tuning of the regularization parameter for these regularizers, our TV-based image restoration results are comparable to those in [43], [45], [54]. It should be noted that this optimality (achieved by considering (*b*) alone) however is only over the set of solutions prescribed by the minimization problem in (2) for a given regularizer. It is possible to further improve quality by considering more sophisticated regularizers, e.g., higher-degree total variation [63], Hessian-based [64] and nonlocal regularization [65]. Extending the applicability of our current parameter selection techniques to these advanced regularizers requires more investigation and is a possible direction for future research.

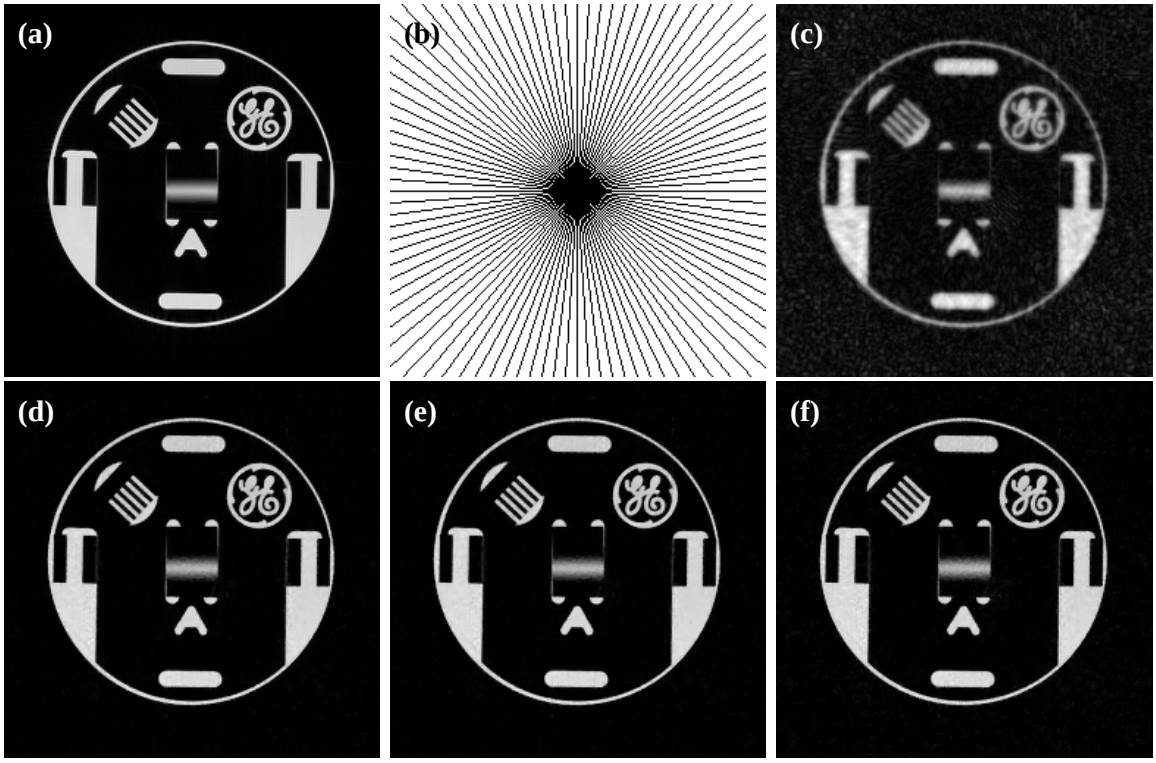


Fig. 7. Experiment MRI-C with real MR GE-phantom data (corresponding to fourth row of Table VIII): (a) Magnitude of reference image reconstructed (using iFFT) and averaged over 10 fully-sampled acquisitions; (b) Retrospective sampling along radial lines (50 lines on a Cartesian grid with 82% undersampling, black lines indicate sample locations); (c) Magnitude of zero-filled iFFT reconstruction from undersampled data (22.51 dB); and magnitude of TV-regularized reconstructions with regularization parameter λ selected to minimize (d) (oracle) MSE (33.03 dB); (e) Predicted-SURE (33.03 dB); (f) NGCV (32.30 dB). Regularized reconstructions (d)-(f) have reduced artifacts compared to the zero-filled iFFT reconstruction (c). Predicted-SURE-based result (e) closely resembles the oracle MSE-based result (d), while NGCV-based result (f) exhibits slightly more artifacts at the center and around the object's periphery.

B. IRLS-MIL and Split-Bregman Algorithms

Both IRLS-MIL and split-Bregman (SB) algorithms can tackle general minimization problems of the form (2) with arbitrary convex regularizers. However, the inner-steps of the SB algorithm (34)-(35) may not admit exact updates for a general $\mathbf{A}$ and / or general regularizer Ψ such as (15) or those in [63], [64].In such cases, iterative schemes may be needed for the updates in (34)-(35), and correspondingly, evaluation of $\mathbf{J}(\mathbf{u}^{(\cdot)}, \mathbf{y})$ has to be performed on a case-by-case basis depending on the type of iterative schemes used for (34)-(35). In this respect, IRLS-MIL is slightly more general: as it is based on the standard gradient-descent IRLS scheme [47], [48], it may be more amenable to tackle sophisticated regularizers [63], [64] and / or a data model involving a more general[15] $\mathbf{A}$.

C. Memory and Computation Requirements

Evaluating reconstruction quality through quantitative measures generally involves additional memory and computational requirements [31]. In our case, it is clear from (27)-(32) that storing and manipulating $\mathbf{J}(\mathbf{u}^{(\cdot)}, \mathbf{y})\mathfrak{n}$, $\mathbf{J}(\mathbf{v}^{(\cdot)}, \mathbf{y})\mathfrak{n}$, $\mathbf{J}(\boldsymbol{\gamma}^{(\cdot)}, \mathbf{y})\mathfrak{n}$ and evaluating NGCV(λ), Predicted-SURE(λ) and Projected-SURE(λ) for one instance of the parameter vector λ demand similar memory and computational load as the IRLS-MIL iterations (19), (25) themselves. These requirements are also comparable to those of the iterative risk estimation techniques in [25], [28] and the Monte-Carlo divergence estimator in [36, Th. 2] (that needs two algorithm evaluations for one instance of λ).The complex-valued case ($\Omega = \mathbb{C}$) demands even more memory and computations (compared to the real-valued case $\Omega = \mathbb{R}$) as one has to tackle Jacobian matrices evaluated with respect to $\mathbf{y}$ and $\mathbf{y}^\star$. This additional requirement is purely a consequence of complex-calculus. In general, the exact amount of storage and computation necessary for evaluating NGCV and WSURE depends on how the reconstruction algorithm is implemented.

Furthermore, in our experiments, we optimize NGCV(λ), Predicted-SURE(λ) and Projected-SURE(λ) using golden-section search that necessitates multiple evaluations of these performance measures for several instances of λ. To save computation time, it is desirable to optimize λ simultaneously during reconstruction. Designing such a scheme is not straightforward when the reconstruction problem is posed as (2) since intermittently changing λ affects the cost function $\mathcal{J}$ and alters the original problem (2).

To avoid this difficulty, image reconstruction can be formulated as a *penalty problem* using variable splitting and penalty techniques [66]. Alternating minimization can then be employed to decouple the original *penalty problem* into simpler linear and nonlinear subproblems [66]. The advantage of this approach is that it provides the option for optimizing parameters based on the subproblems, which can be achieved relatively easily. Liao *et al.* [54], [55] demonstrated the prac-

[15]The matrix $\mathbf{C}$ needs to be chosen in accordance with (18), but since it depends only on $\mathbf{A}$ (and $\mathbf{A}^{\mathrm{H}}$), it can be predetermined for a given problem.

ticability of this approach for TV-based image restoration, but they optimized regularization parameters only based on linear subproblems (using LGCV) and used continuation techniques to adjust other parameters associated with nonlinear subproblems. Since the techniques developed in this paper can handle nonlinear algorithms, they may be adapted to optimize parameters (e.g., using NGCV) associated with the nonlinear subproblems in the *penalty* formulation. As part of future work, we plan to investigate the *penalty* approach for biomedical image reconstruction with simultaneous optimization of penalty parameters.

VII. Summary & Conclusions

Proper selection of the regularization parameter (λ) is an important part of regularized methods for inverse problems. GCV and (weighted) MSE-estimation based on the principle of Stein's Unbiased Risk Estimate [18]—SURE (in the Gaussian setting) can be used for selecting λ, but they require the trace of a linear transformation of the Jacobian matrix, $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$, associated with the nonlinear (possibly iterative) reconstruction algorithm represented by the mapping $\mathbf{u}_{\boldsymbol{\theta}}$. We derived recursions for $\mathbf{J}(\mathbf{u}_{\boldsymbol{\theta}}, \mathbf{y})$ for two types of nonlinear iterative algorithms: the iterative reweighted least squares method with matrix inversion lemma (IRLS-MIL) [47] and the variable splitting-based split-Bregman (SB) algorithm [46], both of which are capable of handling (synthesis-type and) a variety of analysis-type regularizers.

We estimated the desired trace for nonlinear image restoration and MRI reconstruction (from partially sampled Cartesian k-space data) by applying a Monte-Carlo procedure similar to that in [25], [28]. We implemented IRLS-MIL and SB along with computation of NGCV(λ), Predicted-SURE(λ), and Projected-SURE(λ) for total variation and analysis ℓ_1-regularization. Through simulations, we showed for image restoration that selecting λ by minimizing NGCV(λ), Predicted-SURE(λ), and Projected-SURE(λ) consistently yielded reconstructions that were close to corresponding minimum-MSE reconstructions both in terms of visual quality and SNR improvement. For MRI (with partial Cartesian k-space sampling), we conducted experiments with both synthetic and real phantom data and found that NGCV-based reconstructions were slightly sub-optimal in terms of SNR improvement, while minimizing Predicted-SURE(λ) (equivalent to Projected-SURE(λ) in this case) consistently yielded near-MSE-optimal reconstructions both in terms of SNR improvement and visual quality. These results indicate the feasibility of applying GCV- and weighted SURE-based selection of λ for iterative nonlinear reconstruction using analysis-type regularizers. The philosophy underlying theoretical developments in this work can also be extended, in principle, to handle other regularizers, reconstruction algorithms and inverse problems involving Gaussian noise.

References

[1] W. C. Karl, "Regularization in Image restoration and reconstruction," in *Handbook of Image & Video Processing*, A. Bovik, Ed., pp. 183–202. ELSEVIER, 2nd edition, 2005.

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

[3] 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.

[4] T. Regińska, "A regularization parameter in discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 17, no. 3, pp. 740–749, 1996.

[5] C. R. Vogel, "Non-convergence of the L-curve regularization parameter selection method," *Inverse Prob.*, vol. 12, no. 4, pp. 535–47, Aug. 1996.

[6] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

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

[8] G. H. Golub, M. Heath, and G. Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.

[9] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Opt. Engg.*, vol. 29, no. 5, pp. 446–54, 1990.

[10] S. J. Reeves, "Optimal space-varying regularization in iterative image resotration," *IEEE Trans. Im. Proc.*, vol. 3, no. 3, pp. 319–24, 1994.

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

[12] D. A. Girard, "A fast Monte-Carlo cross-validation procedure for large least squares problems with noisy data," *Numerische Mathematik*, vol. 56, no. 1, pp. 1–23, Jan. 1989.

[13] 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. 950–61, 2003.

[14] S. Sourbron, R. Luypaert, P. V. Schuerbeek, M. Dujardin, and T. Stadnik, "Choice of the regularization parameter for perfusion quantification with MRI," *Phys. Med. Biol.*, vol. 49, pp. 3307–24, 2004.

[15] F. O'Sullivan and G. Wahba, "A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments," *J. Comp. Phys.*, vol. 59, no. 3, pp. 441–55, July 1985.

[16] L. N. Deshpande and D. A. Girard, "Fast computation of cross-validated robust splines and other non-linear smoothing splines," in *Curves and Surfaces*, Le Méhauté Laurent and Schumaker, Eds., pp. 143–8. Academic, Boston, 1991.

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

[18] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–51, Nov. 1981.

[19] L. Desbat and D. A. Girard, "The "Minimum Reconstruction Error" choice of regularization parameters: some more efficient methods and their applicaiton to deconvolution problems," *SIAM J. Sci. Comput.*, vol. 16, no. 6, pp. 1387–1403, Nov. 1995.

[20] J. A. Rice, "Choice of smoothing parameter in deconvolution problems," *Contemporary Mathematics*, vol. 59, pp. 137–51, 1986.

[21] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A Study of Methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 4, pp. 326–339, 1991.

[22] 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. Im. Proc.*, vol. 1, no. 3, pp. 322–336, 1992.

[23] S. Ramani, C. Vonesch, and M. Unser, "Deconvolution of 3D fluorescence micrographs with automatic risk minimization," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, pp. 732–5, 2008.

[24] Y. C. Eldar, "Generalized SURE for exponential families: applications to regularization," *IEEE Trans. Sig. Proc.*, vol. 57, no. 2, pp. 471–81, Feb. 2009.

[25] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–22, May 2011.

[26] J-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Sig. Proc.*, vol. 57, no. 12, pp. 4616–32, Dec. 2009.

[27] A. Marin, C. Chaux, J-C. Pesquet, and P. Ciuciu, "Image reconstruction from multiple sensors using stein's principle. Application to parallel MRI," in *Biomedical Imaging: From Nano to Macro, IEEE International Symposium on*, pp. 465–468, 2011.

[28] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," *Proc. IEEE Intl. Conf. Img. Proc.*, pp. 665–8, 2008.

[29] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," *Proc. IEEE Intl. Conf. Img. Proc.*, pp. 1261–4, October 2006.

[30] O. Batu and M. Çetin, "Parameter selection in sparsity-driven SAR imaging," *IEEE Trans. Aero. Elec. Sys.*, vol. 47, no. 4, pp. 3040–50, 2011.

[31] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Im. Proc.*, vol. 19, no. 12, pp. 3116–32, Dec. 2010.

[32] X. -P. Zhang and M. D. Desai, "Adaptive denoising based on SURE risk," *IEEE Sig. Proc. Lett.*, vol. 5, no. 10, pp. 265–267, 1998.

[33] F. Luisier, T. Blu, and M. Unser, "A New SURE approach to image denoising: Interscale Orthonormal Wavelet Thresholding," *IEEE Trans. Im. Proc.*, vol. 16, no. 3, pp. 593–606, 2007.

[34] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Im. Proc.*, vol. 16, no. 11, pp. 2778–86, 2007.

[35] D. Van De Ville and M. Kocher, "Nonlocal means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. Im. Proc.*, vol. 20, no. 9, pp. 2683–90, 2011.

[36] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Im. Proc.*, vol. 17, no. 9, pp. 1540–54, Sept. 2008.

[37] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Im. Proc.*, vol. 12, no. 8, pp. 906–16, Aug. 2003.

[38] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inv. Prob.*, vol. 23, pp. 947–968, 2007.

[39] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.

[40] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.

[41] G. Archer and D. M. Titterington, "On some Bayesian / regularization methods for image restoration," *IEEE Trans. Im. Proc.*, vol. 4, no. 7, pp. 989–95, 1995.

[42] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Im. Proc.*, vol. 8, no. 2, pp. 231–46, 1999.

[43] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Processing*, vol. 89, pp. 1683–93, 2009.

[44] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior," *IEEE Trans. Im. Proc.*, vol. 18, no. 1, pp. 12–26, 2009.

[45] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Im. Proc.*, vol. 17, no. 3, pp. 326–39, 2008.

[46] 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.

[47] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, pp. 257–60, 2010.

[48] S. Ramani, J. Rosen, Z. Liu, and J. A. Fessler, "Iterative weighted risk estimation for nonlinear image restoration with analysis priors," *Proc. SPIE Elec. Img.*, vol. 8296, pp. 82960N1–12, 2012.

[49] 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, 2011.

[50] E. H. Lieb and M. Loss, *Analysis*, American Mathematical Society, Providence, RI, USA, 2nd, revised edition, 2001.

[51] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proceedings H: Microwaves, Optics and Antennas*, vol. 130, no. 1, pp. 11–6, Feb. 1983.

[52] M. Brookes, "The Matrix Reference Manual," [online], http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/calculus.html, 2011.

[53] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2740–6, 2007.

[54] H. Liao, F. Li, and M. K. Ng, "Selection of regularization parameter in total variation image restoration," *J. Opt. Soc. Am. A*, vol. 26, no. 11, pp. 2311–20, 2009.

[55] H. Liao and M. K. Ng, "Blind deconvolution using generalized cross-validation approach to regularization parameter estimation," *IEEE Trans. Im. Proc.*, vol. 20, no. 3, pp. 670–80, 2011.

[56] M. V. Afonso, José M Bioucas-Dias, and Mário A T Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Im. Proc.*, vol. 19, no. 9, pp. 2345–56, Sept. 2010.

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

[58] C. Chaux, P. L. Combettes, J-C. Pesquet, and Valérie R Wajs, "A variational formulation for frame-based inverse problems," *Inverse Prob.*, vol. 23, no. 4, pp. 1495–518, Aug. 2007.

[59] M. F. Hutchinson, "A stochastic estimator for the trace of the influence matrix for Laplacian smoothing splines," *Comm. in Statistics - Simulation and Computation*, vol. 19, no. 2, pp. 433–50, 1990.

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

[61] S. Dong and K. Liu, "Stochastic estimation with z_2 noise," *Phys. Lett. B*, vol. 328, pp. 130–6, 1994.

[62] "Brainweb: Simulated MRI Volumes for Normal Brain," McConnell Brain Imaging Centre, http://www.bic.mni.mcgill.ca/brainweb/.

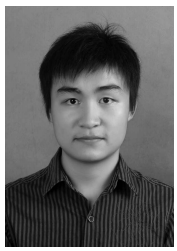
[63] Y. Hu and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery," *IEEE Trans. Im. Proc.*, DOI: 10.1109/TIP.2012.2183143, 2012, in press.

[64] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Trans. Im. Proc.*, vol. 21, no. 3, pp. 983–95, 2012.

[65] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing," *IEEE Trans. Im. Proc.*, vol. 17, no. 7, pp. 1047–60, 2008.

[66] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Img. Sci.*, vol. 1, no. 3, pp. 248–72, 2008.

Sathish Ramani (S'08, M'09) received the M.Sc. degree in Physics from Sri Sathya Sai Institute of Higher Learning (SSSIHL), Puttaparthy, Andhra Pradesh, India in 2000, the M.Sc. degree in Electrical Engineering from Indian Institute of Science (IISc), Bangalore, India in 2004 and Ph.D. in Electrical Engineering from Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland in 2009. He received a Best Student Paper Award at the IEEE International Conference on Acoustics, Speech and Signal Processing in 2006. He was a Swiss National Science Foundation Post-Doctoral Fellow in Systems Group, Electrical Engineering and Computer Science Department, University of Michigan, USA from 2009 to 2010 and continues to work there as a post-doctoral research scholar. His research interests are in splines, interpolation, variational methods in image processing and optimization algorithms for biomedical imaging applications.



Zhihao Liu was born in Hunan, China, on July 7, 1990. He is currently a senior year undergraduate student in Electrical Engineering, EECS department of the University of Michigan, Ann Arbor. Prior to that, he spent two years of undergraduate studies in the Science and Integrated Class, Hongshen Honors School, Chongqing University, China, from 2008 to 2010. He joined the University of Michigan in September 2010 and plans to continue with graduate studies after completing his B.S degree. His research interests are generally in signal and image processing, especially with respect to medical applications.

Jeffrey Rosen is an undergraduate student at the University of Michigan (UM), Ann Arbor pursuing a B.S.E. degree in electrical engineering (expected graduation April 2012). He will join the Computer Science Engineering graduate program, also at UM, Ann Arbor, in Fall 2012. His research interests include computer architecture, particularly multicore and multiprocessor programmability

Jon-Fredrik Nielsen received a PhD in physics from Ohio State University in 2001, and was a post-doctoral fellow at University of Southern California before joining the functional MRI laboratory at University of Michigan as research faculty. His research interests include rapid steady-state MRI sequences, functional MRI, and MRI image reconstruction.

Jeffrey A. Fessler (F'06) received the BSEE degree from Purdue University in 1985, the MSEE degree from Stanford University in 1986, and the M.S. degree in Statistics from Stanford University in 1989. From 1985 to 1988 he was a National Science Foundation Graduate Fellow at Stanford, where he earned a Ph.D. in electrical engineering in 1990. He has worked at the University of Michigan since then. From 1991 to 1992 he was a Department of Energy Alexander Hollaender Post-Doctoral Fellow in the Division of Nuclear Medicine. From 1993 to 1995 he was an Assistant Professor in Nuclear Medicine and the Bioengineering Program. He is now a Professor in the Departments of Electrical Engineering and Computer Science, Radiology, and Biomedical Engineering. He became a Fellow of the IEEE in 2006, for contributions to the theory and practice of image reconstruction. He received the Francois Erbsmann award for his IPMI93 presentation. He serves as an associate editor for IEEE Transactions on Medical Imaging and is a past associate editor for the IEEE Transactions on Image Processing and the IEEE Signal Processing Letters. He was co-chair of the 1997 SPIE conference on Image Reconstruction and Restoration, technical program co-chair of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI), and general chair of ISBI 2007. His research interests are in statistical aspects of imaging problems, and he has supervised doctoral research in PET, SPECT, X-ray CT, MRI, and optical imaging problems.

Regularization Parameter Selection for Nonlinear Iterative Image Restoration and MRI Reconstruction Using GCV and SURE-Based Methods—Supplementary Material

Sathish Ramani*, *Member, IEEE,* Zhihao Liu, Jeffrey Rosen, Jon-Fredrik Nielsen, and Jeffrey A. Fessler, *Fellow, IEEE*

University of Michigan, Ann Arbor, MI, USA.

We present a proof for Lemma 1 and show additional illustrations that we could not accommodate in the paper due to space constraints. References to equations, figures, etc. are within this material only unless specified otherwise.

I. Lemma 1 in the Paper

Lemma 1: Let $\mathbf{f}_{\boldsymbol{\lambda}} : \Omega^M \to \Omega^N$ be differentiable (for $\Omega = \mathbb{R}$) or individually analytic (for $\Omega = \mathbb{C}$ with respect to real and imaginary parts of its argument), respectively, in the weak sense. Then, for any deterministic $\mathbf{T} \in \Omega^{M \times N}$ satisfying $\mathcal{E}_{\boldsymbol{\xi}}\{|[\mathbf{T}\,\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})]_m|\} < \infty$, $m = 1 \dots M$, we have for $\Omega = \mathbb{R}$ or $\mathbb{C}$ that

$$\mathcal{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}^{\mathrm{H}}\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\} = \sigma^2 \mathcal{E}_{\boldsymbol{\xi}}\{\mathrm{tr}\{\mathbf{T}\,\mathbf{J}^{\Omega}_{\mathbf{f}_{\boldsymbol{\lambda}}}(\mathbf{y})\}\}. \quad \blacksquare \tag{1}$$

Proof: The proof for $\Omega = \mathbb{R}$ is very similar to those in [1, Th. 1], [2, Lemma 1], so we focus on the case of $\Omega = \mathbb{C}$. The density function $g_{\mathbb{C}}(\boldsymbol{\xi}) = (\pi\sigma^2)^{-M} \exp(-\boldsymbol{\xi}^{\mathrm{H}}\boldsymbol{\xi}/\sigma^2)$ is analytic with respect to $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\star}$ individually and satisfies the identity

$$\boldsymbol{\xi}^{\mathrm{H}} g_{\mathbb{C}}(\boldsymbol{\xi}) = -\frac{\sigma^2}{2}[\nabla_{\boldsymbol{\xi}_{\Re}} - \imath \nabla_{\boldsymbol{\xi}_{\Im}}] g_{\mathbb{C}}(\boldsymbol{\xi}), \tag{2}$$

where $\nabla_{\boldsymbol{\xi}_{\Re}}$ and $\nabla_{\boldsymbol{\xi}_{\Im}}$ denote $1 \times M$ gradient operators consisting partial derivatives with respect to

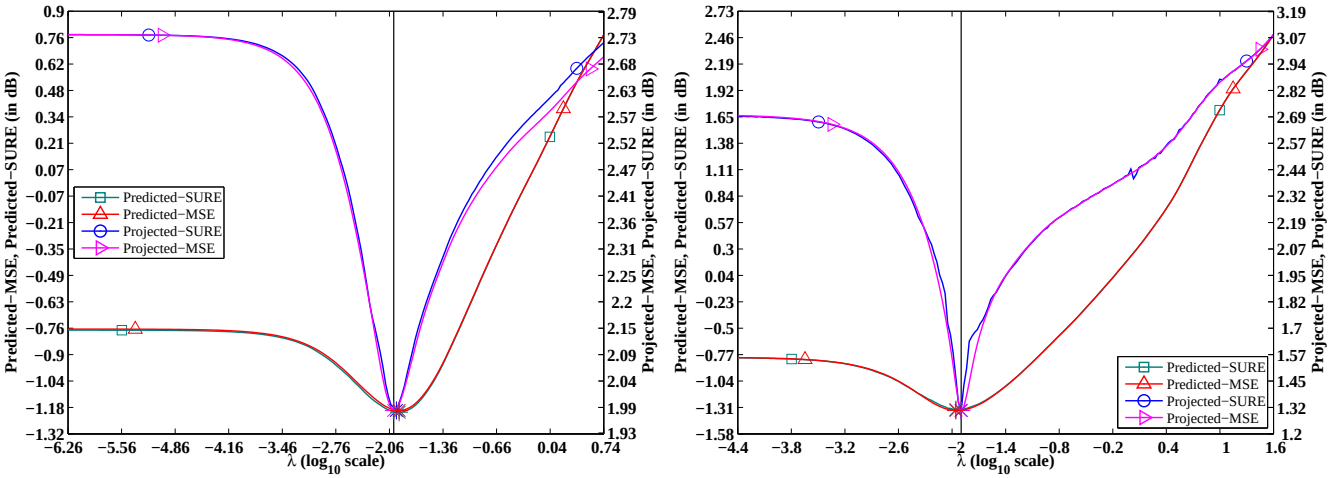


Fig. 1. Plot of Predicted-MSE(λ), Predicted-SURE(λ), Projected-MSE(λ) and Projected-SURE(λ) as functions of regularization parameter λ. Left: Experiment IR-A corresponding to third row of Table II in the paper; Right: Experiment IR-B corresponding to third row of Table III in the paper. The plots indicate that the SURE-curves closely capture the trends of the respective MSE-curves and their minima (indicated by '*') are close to that of the (oracle) MSE indicated by the solid vertical line.

$\{\xi_{\mathfrak{R}m}\}_{m=1}^{M}$, $\{\xi_{\mathcal{I}m}\}_{m=1}^{M}$. Expanding the LHS of (1) and using (2), we get that

$$\mathcal{E}_{\boldsymbol{\xi}}\{\boldsymbol{\xi}^{\mathrm{H}}\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\} = \int g_{\mathbb{C}}(\boldsymbol{\xi})\,\boldsymbol{\xi}^{\mathrm{H}}\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}} = -\frac{\sigma^2}{2}\int \left[\left(\nabla_{\boldsymbol{\xi}_{\mathfrak{R}}} - \iota\nabla_{\boldsymbol{\xi}_{\mathcal{I}}}\right) g_{\mathbb{C}}(\boldsymbol{\xi})\right]\,\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}}. \quad (3)$$

Integrating-by-parts the term involving $\nabla_{\boldsymbol{\xi}_{\mathfrak{R}}}$ in (3) and using the fact that $\lim_{|\xi_{\mathfrak{R}m}|\to\infty} g_{\Omega}(\boldsymbol{\xi})\,[\mathbf{T}\,\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})]_m = 0$, when $\mathcal{E}_{\boldsymbol{\xi}}\{|[\mathbf{T}\,\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})]_m|\} < +\infty$ [1], [2], we get that

$$\begin{aligned}\int \nabla_{\boldsymbol{\xi}_{\mathfrak{R}}} g_{\mathbb{C}}(\boldsymbol{\xi})\,\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}} &= \sum_{m=1}^{M}\sum_{n=1}^{N}\int \frac{\partial g_{\mathbb{C}}(\boldsymbol{\xi})}{\partial \xi_{\mathfrak{R}m}}\,T_{mn}f_{\boldsymbol{\lambda},n}(\mathbf{y})\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}} \\ &= -\sum_{m=1}^{M}\sum_{n=1}^{N}\int g_{\mathbb{C}}(\boldsymbol{\xi})T_{mn}\frac{\partial f_{\boldsymbol{\lambda},n}(\mathbf{y})}{\partial \xi_{\mathfrak{R}m}}\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}} \\ &= -\sum_{m=1}^{M}\sum_{n=1}^{N}\int g_{\mathbb{C}}(\boldsymbol{\xi})T_{mn}\frac{\partial f_{\boldsymbol{\lambda},n}(\mathbf{y})}{\partial y_{\mathfrak{R}m}}\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}}, \quad (4)\end{aligned}$$

where we have set $\partial/\partial\xi_{\mathfrak{R}m} = \partial/\partial y_{\mathfrak{R}m}$ since $\mathbf{A}\mathbf{x}$ (Eq. (1) in the paper) is a deterministic constant. Going through a similar derivation for the integral involving $-\iota\nabla_{\boldsymbol{\xi}_{\mathcal{I}}}$ and using $\partial/\partial\xi_{\mathcal{I}m} = \partial/\partial y_{\mathcal{I}m}$, we get that

$$-\iota\int \nabla_{\boldsymbol{\xi}_{\mathcal{I}}} g_{\mathbb{C}}(\boldsymbol{\xi})\,\mathbf{T}\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{y})\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}} = \iota\sum_{m=1}^{M}\sum_{n=1}^{N}\int g_{\mathbb{C}}(\boldsymbol{\xi})T_{mn}\frac{\partial f_{\boldsymbol{\lambda},n}(\mathbf{y})}{\partial y_{\mathcal{I}m}}\,d\boldsymbol{\xi}_{\mathfrak{R}}\,d\boldsymbol{\xi}_{\mathcal{I}}. \quad (5)$$

Combining (3)-(5) with the definition of $\mathbf{J}_{\mathbf{f}_{\boldsymbol{\lambda}}}^{\Omega}(\mathbf{y})$ (Eq. (4) in the paper) yields the desired result (1). ■

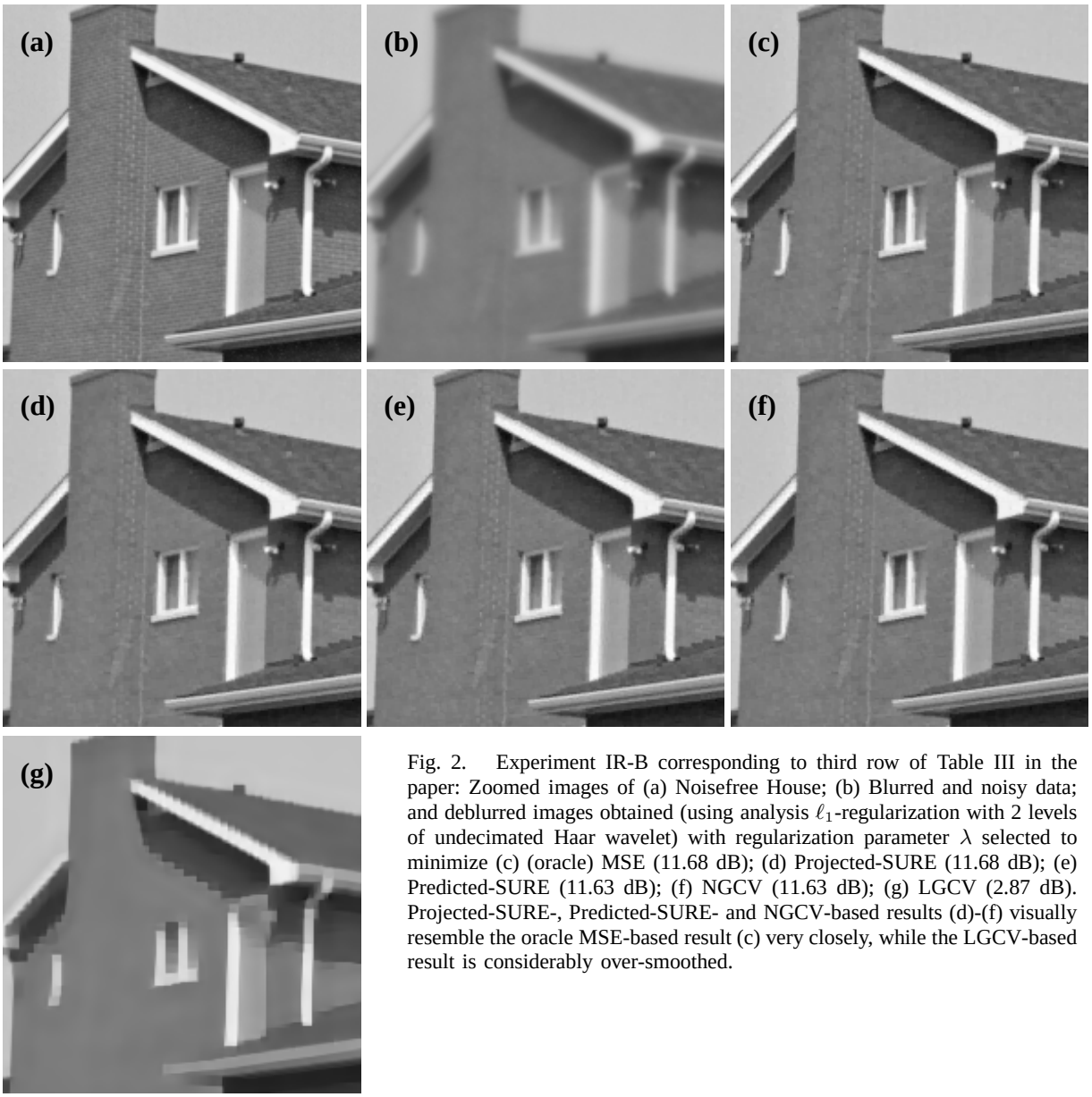
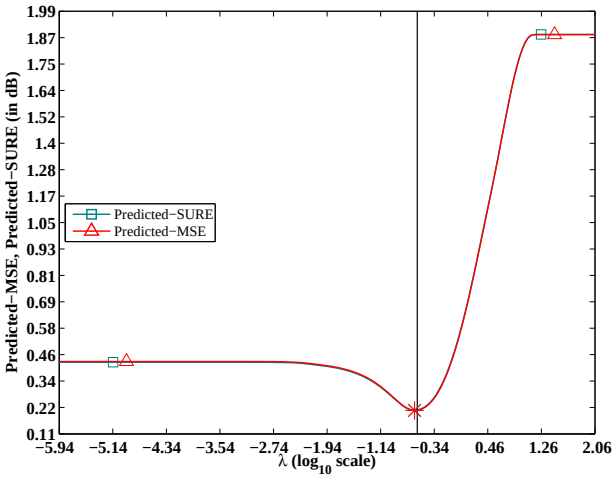


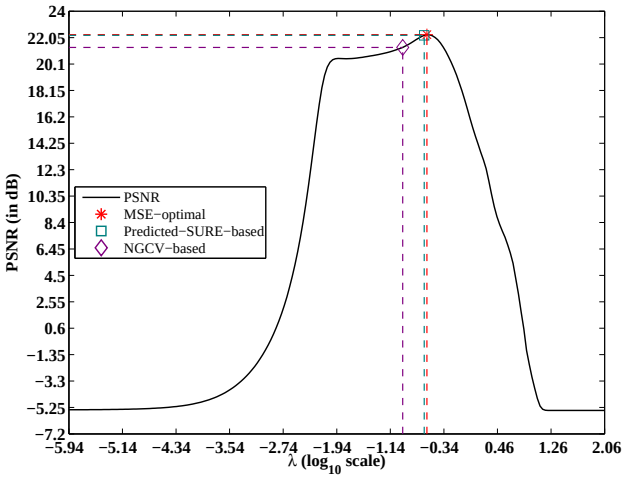
Fig. 2. Experiment IR-B corresponding to third row of Table III in the paper: Zoomed images of (a) Noisefree House; (b) Blurred and noisy data; and deblurred images obtained (using analysis ℓ_1-regularization with 2 levels of undecimated Haar wavelet) with regularization parameter λ selected to minimize (c) (oracle) MSE (11.68 dB); (d) Projected-SURE (11.68 dB); (e) Predicted-SURE (11.63 dB); (f) NGCV (11.63 dB); (g) LGCV (2.87 dB). Projected-SURE-, Predicted-SURE- and NGCV-based results (d)-(f) visually resemble the oracle MSE-based result (c) very closely, while the LGCV-based result is considerably over-smoothed.

II. Additional Illustrations for Image Restoration

Fig. 1 plots Projected-MSE(λ), Projected-SURE(λ), Predicted-MSE(λ) and Predicted-SURE(λ) versus λ for specific instances of Experiments IR-A and IR-B (corresponding to Cameraman and House test images, respectively, see Table I in the paper). Projected-SURE- and Predicted-SURE-curves are accurate in capturing the trends of Projected-MSE- and Predicted-MSE-curves and also exhibit minima (indicated by *) close to that of MSE (solid vertical line). The deblurred images (corresponding to an instance of Experiment IR-B) obtained by minimizing Projected-SURE(λ) (Fig. 2d), Predicted-SURE(λ) (Fig. 2e) and NGCV(λ) (Fig. 2f) closely resemble the corresponding minimum-MSE result (Fig. 2c) in terms of

(a)

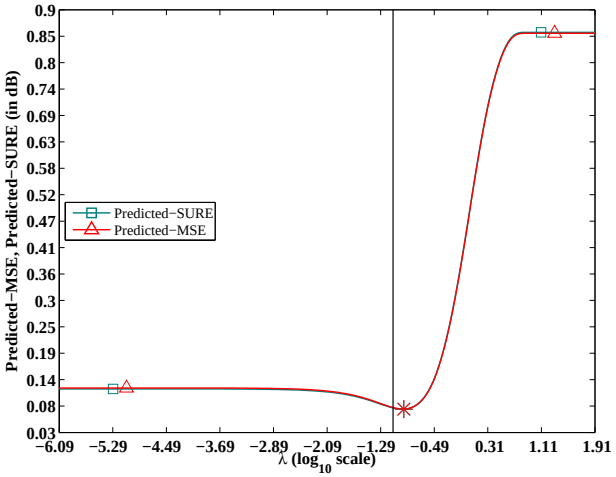


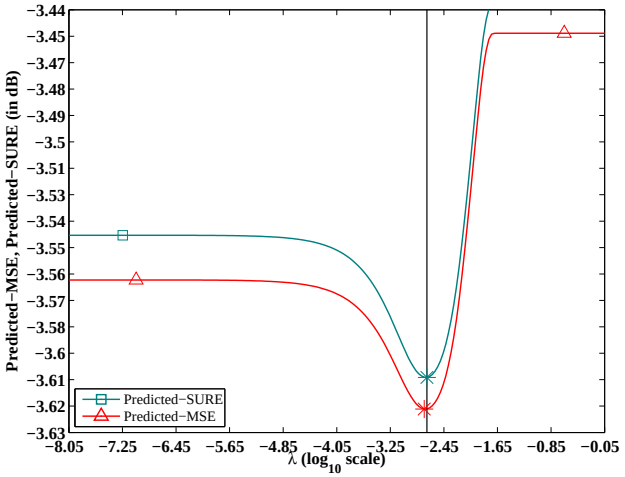


(b)

Fig. 3. Experiment MRI-A with (Shepp-Logan phantom) corresponding to second row of Table VI in the paper: (a) Plot of Predicted-MSE and Predicted-SURE as functions of regularization parameter λ indicates that Predicted-SURE closely captures the trend of Predicted-MSE and their minima (indicated by '*') are close to that of the (oracle) MSE indicated by the solid vertical line; (b) Plot of PSNR(λ) versus λ indicate that λ's that minimize Predicted-SURE and the (oracle) MSE are very close to each other and lead to almost identical PSNRs and reconstructions (see Fig. 5). NGCV-based selection is away from the MSE-based selection but it still yields a reconstruction Fig. 5f that is agreeably close to the oracle Fig. 5d in terms of visual quality in this instance of Experiment MRI-A.

(a)





(b)

Fig. 4. Plot of Predicted-MSE and Predicted-SURE as functions of regularization parameter λ: (a) Experiment MRI-B with synthetic data corresponding to second row of Table VII in the paper; (b) Experiment MRI-C with real MR data corresponding to fourth row of Table VIII in the paper. The plots indicate that Predicted-SURE closely captures the trend of Predicted-MSE and their minima (indicated by '*') are close to that of the (oracle) MSE indicated by the solid vertical line.

visual appearance, while that obtained by minimizing LGCV(λ) is over-smoothed (see Fig. 2g). These results are also consistent with what is reported in the paper (e.g., see Figs. 3 and 4 in the paper).

III. Additional Illustrations for MRI Reconstruction

We evaluated Predicted-MSE(λ), Predicted-SURE(λ) as functions of λ and plotted them in Figs. 3a and 4 for specific instances of Experiments MRI-A (with synthetic data corresponding to noisefree Shepp-Logan phantom), MRI-B (with synthetic data corresponding to a noisefree T_2-weighted MR image)

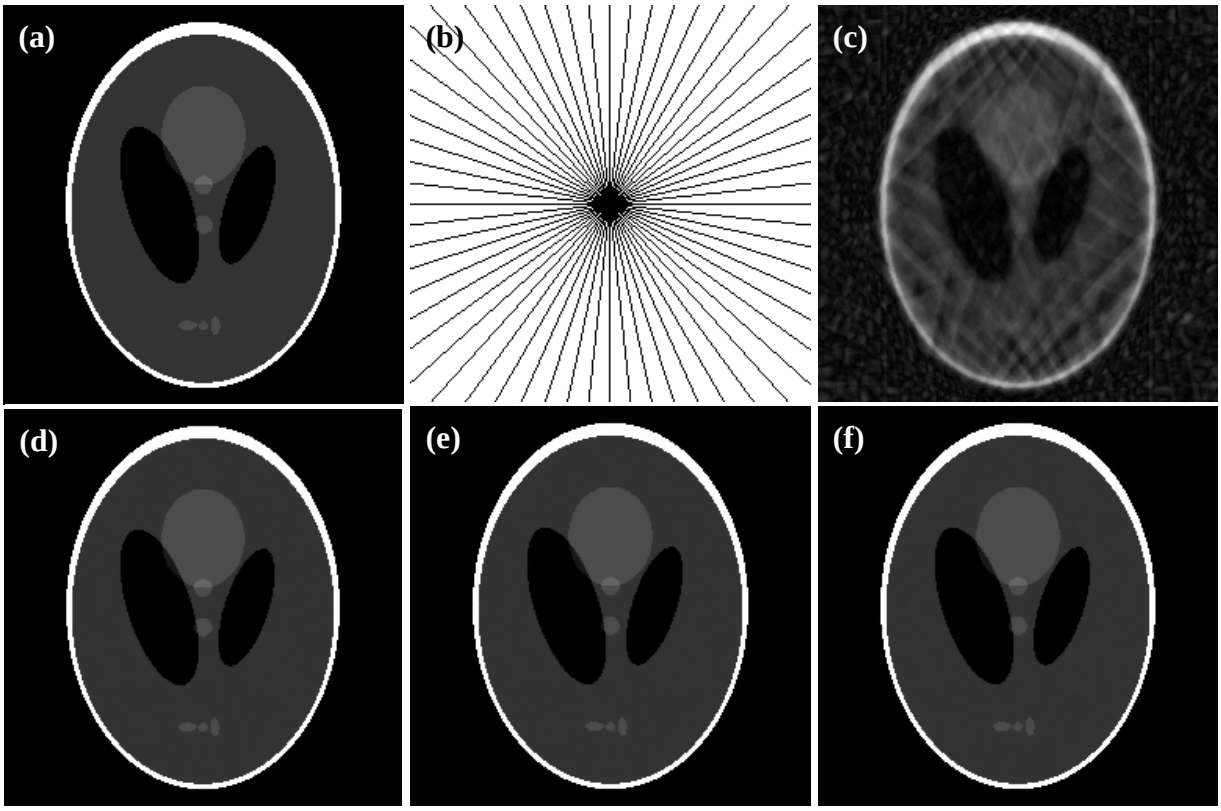


Fig. 5. Experiment MRI-A with Shepp-Logan phantom (corresponding to second row of Table VI: (a) Noisefree Shepp-Logan (256×256) phantom; (b) Retrospective sampling along radial lines (30 lines on a Cartesian grid with 89% undersampling, black lines indicate sample locations); (c) Magnitude of zero-filled iFFT reconstruction from undersampled data; and magnitude of TV-regularized reconstructions with regularization parameter λ selected to minimize (d) (oracle) MSE (22.28 dB); (e) Predicted-SURE (22.21 dB); (f) NGCV (21.68 dB). Regularized reconstructions (d)-(f) have reduced artifacts compared to the zero-filled iFFT reconstruction (c). Predicted-SURE-based and NGCV-based results (e), (f), respectively, closely resemble the oracle MSE-based result (d) in this experiment.

and MRI-C (with real GE phantom data) in the paper. Predicted-SURE not only captures the trend of Predicted-MSE in Figs. 3a and 4, but also exhibits minima (indicated by *) close to that of MSE (solid vertical line).

We also plot PSNR(λ) versus λ in Fig. 3b and present reconstructions in Fig. 5 for the instance of Experiment MRI-A considered in Figs. 3a: Although NGCV-selection is slightly away from the (oracle) MSE-selection in Fig. 3b, the corresponding reconstruction Fig. 5f is visually similar to the MSE-based one in Fig. 5d in this case. Predicted-SURE-selection is close to the (oracle) MSE-selection in Fig. 3b and therefore naturally leads a reconstruction Fig. 5e that resembles the MSE-based one in Fig. 5d. These illustrations point to (the sub-optimality of NGCV and) the accuracy of Predicted-SURE(λ) for MRI reconstruction from partially sampled Cartesian k-space data and are consistent with our results portrayed in the paper.

REFERENCES

[1] Y. C. Eldar, "Generalized SURE for exponential families: applications to regularization," *IEEE Trans. Sig. Proc.*, vol. 57, no. 2, pp. 471–81, Feb. 2009.

[2] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Im. Proc.*, vol. 16, no. 11, pp. 2778–86, 2007.