CROSS-VALIDATION AND PREDICTED RISK ESTIMATION FOR NONLINEAR ITERATIVE REWEIGHTED LEAST-SQUARES MRI RECONSTRUCTION

Sathish Ramani∗, Jon-Fredrik Nielsen† and Jeffrey A. Fessler∗

∗EECS Department, University of Michigan, Ann Arbor, MI, USA. 
†fMRI Laboratory, University of Michigan, Ann Arbor, MI, USA.

ABSTRACT

Regularization is an effective means of reducing noise and artifacts in MR image reconstruction from undersampled k-space data. Proper application of regularization demands appropriate selection of the associated regularization parameter. Generalized cross-validation (GCV) is a popular parameter tuning technique especially for linear reconstruction methods, but its application to nonlinear iterative MRI reconstruction is more involved as it demands the evaluation of the Jacobian matrix of the reconstruction algorithm with respect to complex-valued data. We derive analytical expressions for recursively updating this Jacobian matrix for an iterative reweighted least-squares reconstruction algorithm. Our method can also be used to calculate a predicted risk estimate (PSURE) for MRI based on Stein’s principle. We demonstrate with simulations and experiments with real data that regularization parameter selection based on GCV and PSURE provides near-MSE-optimal results for nonlinear MRI reconstruction from undersampled k-space data using ℓ1-regularization.

Index Terms— MRI reconstruction, regularization, cross-validation, Stein’s unbiased risk estimate, Jacobian matrix.

1. INTRODUCTION

Magnetic resonance image (MRI) reconstruction from single-coil undersampled k-space data is an ill-posed problem and requires regularization to provide meaningful reconstruction results. Non-quadratic regularizers that promote sparsity (e.g., ℓ1-regularization) or that preserve edges (e.g., total variation) are attractive as they can effectively reduce noise and artifacts in the reconstruction [1]. However, a successful application of such regularization criteria demands proper selection of associated regularization parameters that control the balance between noise-amplification and image-smoothing. This task is generic to many regularized reconstruction problems including that of MRI and is usually performed manually.

In this paper, we focus on two quantitative techniques: generalized cross-validation (GCV) [2, 3] and the predicted form [4] of Stein’s unbiased risk estimate [5] (PSURE), for automated selection of the regularization parameter for iterative nonlinear MRI reconstruction from undersampled k-space data. Both GCV and SURE-type approaches have been used in image restoration problems, especially involving linear reconstruction algorithms [2]. Their computation for nonlinear reconstruction problems is however complicated by the nonlinearity of the associated algorithms [3, 6–9]. Specifically, they require the evaluation of the Jacobian matrix [3, 4, 6, 8, 9] of the estimator with respect to complex-valued MR data [4] for MRI reconstruction. We propose to evaluate this Jacobian matrix for a fast variant of iterative reweighted least-squares (IRLS) algorithm [1] that is gradient-based and is capable of accommodating a variety of smooth and nonsmooth regularization criteria. To the best of our knowledge, the derivation of this complex-valued Jacobian matrix has not been documented for nonlinear iterative MRI reconstruction with IRLS-type algorithms. We demonstrate using experiments with realistically simulated nonCartesian MR data [10] and with Cartesian MR data acquired with a 3T GE scanner that the use of GCV and PSURE lead to near-optimal selections for the regularization parameter for MRI reconstruction from undersampled data.

2. REGULARIZED MRI RECONSTRUCTION

2.1. Problem Formulation

We formulate regularized MRI reconstruction as a discretized optimization problem,

\[
\mathbf{u}_\text{LS}(\mathbf{y}) = \arg \min_{\mathbf{u}} \left\{ \Psi(\mathbf{u}) = \frac{1}{2} \| \mathbf{y} - \mathbf{F} \mathbf{u} \|^2 + \lambda \Psi(\mathbf{R}_l \mathbf{u}) \right\},
\]

(1)

where the \( M \times 1 \) vector \( \mathbf{y} \) represents (possibly undersampled) Cartesian / nonCartesian data from a single-channel receiver coil, \( \mathbf{F} \) is the \( M \times N \) Fourier encoding matrix associated with \( \mathbf{y} \) such that \( M \leq N \), \( \mathbf{u}_\text{LS}(\mathbf{y}) \) denotes the \( N \times 1 \) reconstruction (that is an implicit function of \( \mathbf{y} \)), \( \Psi \) is a suitable regularization, \( \mathbf{R}_l \in \mathbb{R}^{P \times N} \) is a matrix denoting the regularization operator, e.g., frames, finite differences, etc., and \( \lambda > 0 \) denotes the regularization parameter. For simplicity of exposition, we consider \( \Psi(\mathbf{R}_l \mathbf{u}) = \sum_{i=1}^{N} \Phi(|\mathbf{R}_l \mathbf{u}|_i) \), where \( \Phi \) is a suitable convex potential function, e.g., \( \Phi(x) = x \) leads to an ℓ1-regularization. The methods developed in this paper can be also extended to handle other regularizers, e.g., total variation [8].

2.2. An IRLS-MIL Algorithm

We use an iterative reweighted least-squares (IRLS) type algorithm [1] for solving (1). IRLS-type algorithms are gradient-based and provide a general framework for solving reconstruction problems like (1) for a variety of regularization criteria. In the basic form of IRLS, at iteration \( \ell \), one needs to solve a linear system of equations of the form [1]

\[
\mathbf{A}(\ell) \mathbf{u}(\ell+1) = \mathbf{F}^H \mathbf{y},
\]

(2)

where \( \mathbf{A}(\ell) = \mathbf{F}^H \mathbf{F} + \mathbf{R}_l^H \mathbf{R}_l \) and \( (.)^H \) and \( (.)^\dagger \) represent standard and Hermitian-transposes, respectively, \( \mathbf{R}_l(\ell) = \mathbf{diag}\{\gamma(\ell)\} \). The \( l \)-th element of \( \gamma(\ell) \in \mathbb{R}^N \) is \( \gamma(\ell)_l = \left( \frac{\lambda + \phi(\ell)_l}{\lambda + \phi(\ell)_l + \epsilon \| \mathbf{R}_l \mathbf{u}(\ell) \|_1} \right)^{-1} \) [1, 8]. For ℓ1-regularization (i.e., \( \Phi(x) = x \)), this leads to [8]

\[
\gamma(\ell)_l = \lambda^{-1} \| \mathbf{R}_l \mathbf{u}(\ell) \|_1.
\]

(3)

To manage nonsmooth regularization (such as the considered ℓ1-regularization) “corner-rounding” is often administered to (2), i.e., a small positive constant, \( \epsilon > 0 \), is added to \( \{\gamma(\ell)_l\} \) to ensure that
where $\lambda$ denotes the complex-valued Jacobian matrix of $u_\lambda$ with respect to $y$ that is defined in terms of its elements as [4, 12]

\[
[J(u_\lambda; y)]_{m} \triangleq \frac{1}{2} \left( \frac{\partial \langle z \rangle}{\partial z_m} \right)_{y \rightarrow y^*},
\]

where $z_m$ and $z_{m'}$ denote the $m$-th component of the real and imaginary parts of a vector $z \in \mathbb{C}^M$. When $u_\lambda$ is specified in terms of $y$ and $y^*$ (the complex conjugate of $y$), $J(u_\lambda; y)$ can be evaluated treating $y$ as a variable and $y^*$ as a constant [4, 12]. Similarly, $J(u_\lambda; y^*)$ can be evaluated treating $y$ as a constant [4, 12]. We take the real part, $\Re[\cdot]$, in the denominator of (9) to avoid spurious complex entries while computing NGCV($\lambda$) numerically.

As an alternative to NGCV($\lambda$), we also consider the predicted mean squared-error, PMSE($\lambda$) $\triangleq M^{-1} \| F(u_{\text{true}} - u_\lambda(y)) \|^2_2$, for quantifying image-quality where $u_{\text{true}}$ is the unknown (deterministic) true image. Since PMSE($\lambda$) depends on $u_{\text{true}}$, it cannot be directly used and must be estimated in practice. Assuming that noise in $y$ is zero-mean white complex-Gaussian with variance $\sigma^2$, Stein's principle [5] can be used to obtain a predicted Stein's unbiased risk
\begin{equation}
M^{-1}\|y - \mathbf{F}u_3(y)\|^2_2 - \sigma^2 + 2\sigma^2 M^{-1} \mathbb{E}\{\text{tr}(\mathbf{F}(u_3; y))\},
\end{equation}

that is unbiased in the sense that \( \mathbb{E}_y \{\text{PMSE}(\lambda)\} = \mathbb{E}_y \{\text{PSURE}(\lambda)\} \) \cite{4, 5, 7}, where \( \mathbb{E}_y \{\cdot\} \) represents expectation with respect to \( y \).

NGCV(\lambda), PSURE(\lambda) and (PMSE(\lambda)) are computed in the measurement-domain: for MRI, this corresponds to evaluating these performance-measures at sample locations in \( k \)-space. While these are not equivalent to the usual image-domain error-measure near-MSE-optimal and PSURE

\begin{equation}
J = \mathbb{E}\{\text{PSURE}(\lambda)\}
\end{equation}

end of derivation of \( J \) (Section 3), we have that

\begin{equation}
J(\mathbf{u}^{(i+1,j+1)}; \mathbf{z}) = J(\mathbf{b}^{(i+1,j)}; \mathbf{z}) - C(\mathbf{z})^T \mathbf{R}^T J(\mathbf{u}^{(i+1,j)}; \mathbf{z}),
\end{equation}

where \( J(\mathbf{b}^{(i,j)}; \cdot) \) indicates a Jacobian matrix update at a (nested) iteration indexed by \( i \) and \( j \). From the definition of complex-valued Jacobian matrices (Section 3), we have that

\begin{equation}
J(\mathbf{b}^{(i,j)}; \mathbf{y}) = C(\mathbf{z})^T \mathbf{F}^T (I_N - C(\mathbf{z})^T \mathbf{F}) J(\mathbf{u}^{(i,j)}; \mathbf{y}),
\end{equation}

Thus, evaluation of \( J(\mathbf{u}^{(i,j)}; y) \) requires \( J(\mathbf{u}^{(i,j)}; y^*) \) as mentioned earlier. It only remains to evaluate \( J(\mathbf{u}^{(i,j)}; y^*) \) for the considered \( \ell_1 \)-regularization. We have from (3) that \( \gamma(\mathbf{u}^{(i,j)}) = \{\lambda^{-1} |\mathbf{R}\mathbf{u}_{(i)}| |l| = \lambda^{-1} \sqrt{\text{Frob}(|\mathbf{R}\mathbf{u}_{(i)}|^2)} \} \), which (treating \( \mathbf{u}_{(i)} \) as a constant) leads to \(4, 12\)

\begin{equation}
J(\gamma(\mathbf{u}^{(i,j)}; \mathbf{u}_{(i)})) = \text{diag}\{2\lambda^2 \gamma(\mathbf{u}_{(i)}\} \text{diag}\{\mathbf{R}\mathbf{u}_{(i)}\} \mathbf{R}.
\end{equation}

For typical reconstruction sizes, the Jacobian matrices will be enormous and cannot be stored and manipulated directly. So we use a

\begin{equation}
\mathbf{Fj}(\mathbf{u}^{(i,j)}; \mathbf{y}) \approx \frac{\Delta}{n}\mathbf{Fj}(\mathbf{u}^{(i,j)}; \mathbf{y})\mathbf{n}.
\end{equation}

It can be shown that \( \hat{\mathbf{t}} \) is unbiased \(\mathbb{E}_t\{\cdot\} \), i.e.,

\begin{equation}
\mathbb{E}_u \{\hat{\mathbf{t}}\} = \mathbb{E}_u \{\text{tr}(\mathbf{Fj}(\mathbf{u}^{(i,j)}; y))\},
\end{equation}

and its variance can be minimized by using a binary random vector \( \mathbf{n} = n_{\pm1} \) whose entries are either +1 or -1 with probability 0.5 \cite{4} compared to using a Gaussian \( \mathbf{n} \) \cite{6}. To summarize our scheme, we run the sequence of iterations (6), (8) to obtain the reconstruction \( \mathbf{u}_{(i)} \) and simultaneously use (12)-(16) with \( \mathbf{n}_{\pm1} \) to numerically compute \( \text{NGCV}(\lambda) \) and \( \text{PSURE}(\lambda) \) at every iteration. Since \( \text{NGCV}(\lambda) \) and \( \text{PSURE}(\lambda) \) both require \( \text{tr}(\mathbf{Fj}(\mathbf{u}^{(i,j)}; y)) \), their computational complexity is very similar.

\section{EXPERIMENTAL SETUP & RESULTS}

We implemented products with \( \mathbf{F} \) using NUFFFT \(\mathbb{E}_n\{\cdot\} \) for nonCartesian data and FFT for Cartesian data. We chose \( \mathbf{C}_i(\gamma) \) \(\mathbb{E}_n\{\cdot\} \) to be a circulant matrix and implemented \( \mathbf{C}_i^{-1}(\gamma) \) using FFTs in IRLS-MIL. For nonCartesian data, we used \( \mathbf{C}_{(i)} = \mathbf{C}_{\text{Frob}} + \nu \mathbf{I}_N \) \(\mathbb{E}_n\{\cdot\} \), where
C_{Frob} is a circulant matrix that is “closest” to $F^{\dagger}F$ in Frobenius norm [14] and $\nu$ was selected such that $\nu I \succ F^{\dagger}F - C_{Frob}$. For Cartesian data, we simply set $C_{\nu}(\ell) = F^{\dagger}F + \nu I_N \forall \ell$ for some $\nu > 0$, since $F^{\dagger}F$ is already circulant in this case. We used finite differences for $R$ in (1) and performed 10 $j$-, 10 $k$- and 7 $t$-iterations of IRLS-MIL (6), (8) using $F^{\dagger}y$ as the initialization. In all experiments, the overall compute time for obtaining the reconstruction and simultaneously evaluating NGCV($\lambda$) and PSURE($\lambda$) for a given $\lambda$ under these settings was 50 seconds on a 8-core PC with 2.80-GHz Intel Xeon processors. For the purpose of illustration, we evaluated NGCV($\lambda$) and PSURE($\lambda$) for a wide range of $\lambda$ (see Figs. 1 and 2) while in practice, golden-section search can be used to optimize $\lambda$ with relatively fewer evaluations of NGCV($\lambda$) and PSURE($\lambda$).

In the first experiment, we used a spiral trajectory (with 64 leaves and 512 samples per leaf corresponding to 50% undersampling) and simulated realistic non-Cartesian MR data of 40 dB SNR using the analytical Shepp-Logan phantom of Guerquin-Kern et al. [10]. We reconstructed $256 \times 256$ images of the Shepp-Logan phantom by running the IRLS-MIL algorithm for varying $\lambda$. We assumed $\sigma^2$ was available for computing PSURE($\lambda$). Fig. 1 plots PSNR($\lambda$) $\triangleq 10 \log_{10}(\max(u_{true})^2 / \text{MSE}(\lambda))$ as a function of $\lambda$ and indicates $\lambda$-values that minimize NGCV($\lambda$), PSURE($\lambda$) and the true MSE($\lambda$). We see that both NGCV- and PSURE-based $\lambda$-selections lead to reconstructions with PSNRs almost close to that of the minimum-MSE reconstruction.

Next, we acquired 10 independent sets of fully-sampled 2-D Cartesian data ($256 \times 256$) of a GE-phantom using a 3T GE scanner (GRE sequence with flip angle $35^\circ$, $T_R = 200$ ms, $T_E = 7$ ms, FOV = 15 cm). These fully-sampled Cartesian datasets were used to reconstruct (using 3DFT) 2-D images that were then averaged to obtain a reference image that served as the “unknown” $u_{true}$ (Fig. 3a) for computing MSE($\lambda$) and PSNR($\lambda$). We estimated $\sigma^2$ for use in PSURE($\lambda$) from separately acquired dummy-data (with the same scan setting) when no RF scan setting) when no RF

6. SUMMARY, CONCLUSION & DISCUSSION

In this paper, we presented a method for computing the generalized cross-validation measure (NGCV) [3] and a predicted risk estimate (PSURE) [4] for nonlinear MRI reconstruction using an iterative reweighted least-squares-type (IRLS-MIL) [11] algorithm. Both NGCV and PSURE require the evaluation of a complex-valued Jacobian matrix [3, 7] that we carried out analytically for the IRLS-MIL algorithm. We presented numerical results using simulated (non-Cartesian) and real (Cartesian) MR data and illustrated that both NGCV and PSURE are able to provide near-optimal selections of the regularization parameter for regularized nonlinear MRI reconstruction.

Our methods can be directly applied to 3-D Cartesian and non-Cartesian MRI reconstruction (albeit with increased computation). The principles underlying our work can also be extended to other algorithms, e.g., the split-Bregman algorithm [4].

Evaluating NGCV($\lambda$) and PSURE($\lambda$) requires additional memory and computation of the same order as the IRLS-MIL algorithm used for reconstruction. Although these performance measures can be optimized using golden-section search, multiple evaluations of NGCV($\lambda$) and PSURE($\lambda$) may be necessary. Alternatively, MRI reconstruction can be formulated as a constrained minimization of a suitable (nonquadratic) regularization subject to a data-consistency constraint: this avoids the need for searching an appropriate $\lambda$, however such constrained problems are harder to tackle. We are currently investigating a comparison of the methods proposed in this paper with the constrained formulation both in terms of reconstruction quality and computation time.

7. REFERENCES