

# Automatic Regularization Parameter Selection for Iterative Nonlinear MRI Reconstruction

Sathish Ramani<sup>1</sup>, Jon-Fredrik Nielsen<sup>2</sup>, Jeffrey A. Fessler<sup>1</sup>

EECS, University of Michigan, Ann Arbor, MI, United States; 2fMRI Laboratory, University of Michigan, Ann Arbor, MI, United States

**INTRODUCTION:** MRI reconstruction from undersampled  $k$ -space data requires regularization to reduce artifacts and improve image quality. Nonquadratic regularizers, e.g., edge-preserving ones or those based on the  $l_1$ -norm, have proven to be useful in MRI [1], but successful application of such criteria depends on proper selection of the regularization parameter ( $\lambda$ ) that controls the degree of *smoothness* imposed on the reconstruction. Several quantitative methods are available for automatic selection of  $\lambda$  [2] such as the *discrepancy principle* (DP) [2], the *L-curve* method (LCM) [3], *generalized cross-validation* (GCV) [4] and the estimation of *mean-squared error* (MSE) type measures [5]. DP and the LCM are able to handle a variety of nonlinear reconstruction algorithms, but can lead to over-smoothing [2] or become sensitive to changes in  $\lambda$  [3], respectively. GCV is a popular choice in reconstruction problems especially involving linear algorithms [4]. In the linear case, GCV is simple to implement and can provide *asymptotically optimal* selection of  $\lambda$  [4]. *MSE-type* estimates are attractive alternatives to GCV as such estimates can provide (near) optimal results even in the nonasymptotic regime [5]. However, both GCV [6] and estimation of MSE-type measures become nontrivial and computationally involved for nonlinear algorithms. In this work, we propose a practical means of computing GCV and an MSE-type estimate for *nonlinear* MRI reconstruction using the split-Bregman algorithm [7]. We illustrate with experiments on real MR data that they can be employed for near-optimal adjustment of  $\lambda$  for reconstruction from undersampled Cartesian  $k$ -space data using nonquadratic regularization.

## METHODS

**MRI Reconstruction:** We perform MRI reconstruction by minimizing a cost function:  $\mathbf{u}_\lambda(\mathbf{y}) = \operatorname{argmin}_{\mathbf{u}} \{\Psi(\mathbf{u}) := \|\mathbf{y} - \mathbf{A}\mathbf{u}\|^2 + \lambda\Phi(\mathbf{R}\mathbf{u})\}$ , where  $\|\cdot\|$  denotes Euclidean norm,  $\mathbf{y}$  is the  $M \times 1$  undersampled Cartesian  $k$ -space data,  $\mathbf{A}$  is the  $M \times N$  undersampled DFT matrix,  $\mathbf{R}$  is a regularization operator (e.g., finite differences),  $\Phi$  is the regularization and the  $N \times 1$  reconstruction is written as  $\mathbf{u}_\lambda(\mathbf{y})$  to indicate its dependence on  $\lambda$  and  $\mathbf{y}$ . For minimizing the cost  $\Psi$ , we employ the split-Bregman (SB) reconstruction algorithm [7] that is nonlinear and can handle several regularizers  $\Phi$  including total variation (TV) and  $l_1$ -regularization exactly. The SB algorithm is based on variable splitting that employs an auxiliary constraint variable  $\mathbf{v}$  to transform the original minimization problem (involving the cost  $\Psi$ ) in to the following equivalent constrained problem:  $\min_{\mathbf{u}, \mathbf{v}} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|^2 + \lambda\Phi(\mathbf{v})$  subject to  $\mathbf{v} = \mathbf{R}\mathbf{u}$ . This problem is then solved in an augmented Lagrangian-type framework leading to the following algorithm [7]:

$$\begin{aligned} \mathbf{u}_{(i+1)} &= \mathbf{B}^{-1}[\mathbf{A}'\mathbf{y} + \mu\mathbf{R}'(\mathbf{v}_{(i)} - \boldsymbol{\eta}_{(i)})], \\ \mathbf{v}_{(i+1)} &= \mathbf{d}_\Phi(\boldsymbol{\rho}_{(i)}), \\ \boldsymbol{\eta}_{(i+1)} &= \boldsymbol{\rho}_{(i)} - \mathbf{v}_{(i+1)}, \end{aligned}$$

where  $\mu > 0$  governs the convergence speed of the SB algorithm,  $\mathbf{B} := \mathbf{A}'\mathbf{A} + \mu\mathbf{R}'\mathbf{R}$ ,  $\boldsymbol{\rho}_{(i)} := \mathbf{R}\mathbf{u}_{(i+1)} + \boldsymbol{\eta}_{(i)}$ ,  $(\cdot)'$  denotes Hermitian-transpose, and  $\mathbf{d}_\Phi(\cdot)$  represents a denoising operator that admits explicit analytical forms for several instances of  $\Phi$  (e.g.,  $\mathbf{d}_\Phi$  represents soft-thresholding and vector-shrinkage operations for  $l_1$ -regularization and TV, respectively) [8].

**GCV & MSE-Type Measures:** The *generalized GCV* measure (that applies to linear and nonlinear algorithms [6]) is given by  $GCV(\lambda) = M^{-1}\|\mathbf{y} - \mathbf{A}\mathbf{u}_\lambda(\mathbf{y})\|^2 / (1 - M^{-1}\operatorname{Re}\{\operatorname{tr}\{\mathbf{A}\mathbf{J}_\lambda(\mathbf{y})\}\})^2$ , where  $\operatorname{Re}\{\cdot\}$  denotes the real part of a complex-valued entity,  $\operatorname{tr}\{\cdot\}$  represents the trace of a matrix and  $\mathbf{J}_\lambda(\mathbf{y})$  represents the Jacobian matrix whose rows contain gradients of the components of the reconstruction  $\mathbf{u}_\lambda(\mathbf{y})$  evaluated with respect to  $\mathbf{y}$  (treating  $\mathbf{y}^*$ , the complex conjugate of  $\mathbf{y}$ , as constant) [9].

When  $k$ -space is undersampled,  $\mathbf{y}$  has only partial information about the underlying object of interest  $\mathbf{x}$ . It is therefore possible to assess the reconstruction error only at the sample locations in  $k$ -space [10]: This corresponds to the so-called *predicted MSE* (PMSE) given by  $PMSE(\lambda) = \|\mathbf{A}(\mathbf{x} - \mathbf{u}_\lambda(\mathbf{y}))\|^2$ . Assuming  $\mathbf{y}$  is corrupted by complex-valued Gaussian noise (i.i.d. *zero-mean* with variance  $\sigma^2$ ), Stein's principle [11] can be used to estimate  $PMSE(\lambda)$  and leads to the *predicted Stein's Unbiased Risk Estimate* (PSURE) given by  $PSURE(\lambda) = M^{-1}\|\mathbf{y} - \mathbf{A}\mathbf{u}_\lambda(\mathbf{y})\|^2 - \sigma^2 + 2M^{-1}\sigma^2\operatorname{Re}\{\operatorname{tr}\{\mathbf{A}\mathbf{J}_\lambda(\mathbf{y})\}\}$  [10]. PSURE requires the knowledge of  $\sigma^2$  (that can be estimated in practice) unlike GCV, but PSURE provides relatively better selection of  $\lambda$  compared to GCV as illustrated in our experiments. Practical application of GCV and PSURE for tuning  $\lambda$  requires (apart from simple computations such as  $\|\mathbf{y} - \mathbf{A}\mathbf{u}_\lambda(\mathbf{y})\|^2$ ) the computation of  $\mathbf{J}_\lambda(\mathbf{y})$  corresponding to the algorithm used for reconstruction. In the sequel, we show how to evaluate  $\mathbf{J}_\lambda(\mathbf{y})$  analytically for the SB algorithm.

**Evaluating  $\mathbf{J}_\lambda(\mathbf{y})$ :** We recursively evaluate  $\mathbf{J}_\lambda(\mathbf{y})$  using linearity-, product- and chain-rule for Jacobian matrices [9]: As  $\mathbf{v}$  and  $\boldsymbol{\eta}$  are functions of  $\mathbf{y}$  (via  $\mathbf{u}$ ) in the SB algorithm, we get  $\mathbf{J}_{\mathbf{u}_{(i+1)}}(\mathbf{y}) = \mathbf{B}^{-1}\mathbf{A}' + \mu\mathbf{B}^{-1}\mathbf{R}'[\mathbf{J}_{\mathbf{v}_{(i)}}(\mathbf{y}) - \mathbf{J}_{\boldsymbol{\eta}_{(i)}}(\mathbf{y})]$ . Applying chain rule [9],  $\mathbf{J}_{\mathbf{v}_{(i)}}(\mathbf{y}) = \mathbf{J}_{\mathbf{d}_\Phi(\boldsymbol{\rho}_{(i)})}(\mathbf{y}) = \mathbf{J}_{\mathbf{d}_\Phi(\boldsymbol{\rho}_{(i)})}(\mathbf{y}) + \mathbf{J}_{\mathbf{d}_\Phi(\boldsymbol{\rho}_{(i)})}(\mathbf{y}^*)\mathbf{J}_{\boldsymbol{\rho}_{(i)}}(\mathbf{y})$ , where from the definition of  $\boldsymbol{\rho}_{(i)}$  above, we have that  $\mathbf{J}_{\boldsymbol{\rho}_{(i)}}(\mathbf{y}) = \mathbf{R}\mathbf{J}_{\mathbf{u}_{(i+1)}}(\mathbf{y}) + \mathbf{J}_{\boldsymbol{\eta}_{(i)}}(\mathbf{y})$ . The specific form of  $\mathbf{J}_{\mathbf{d}_\Phi(\boldsymbol{\rho}_{(i)})}$  depends on  $\Phi$  and can be analytically evaluated on a case-by-case basis for several instances of  $\Phi$  (including TV and  $l_1$ -regularization [8,10]). Finally, the update corresponding to  $\boldsymbol{\eta}_{(i+1)}$  in the SB algorithm yields  $\mathbf{J}_{\boldsymbol{\eta}_{(i+1)}}(\mathbf{y}) = \mathbf{J}_{\boldsymbol{\rho}_{(i)}}(\mathbf{y}) - \mathbf{J}_{\mathbf{u}_{(i+1)}}(\mathbf{y})$ . The Jacobian matrices  $\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})$ ,  $\mathbf{J}_{\mathbf{v}_{(i)}}(\mathbf{y})$ ,  $\mathbf{J}_{\boldsymbol{\eta}_{(i)}}(\mathbf{y})$  have enormous sizes for typical reconstruction settings, so we store and update vectors of the form  $\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\mathbf{n}$ ,  $\mathbf{J}_{\mathbf{v}_{(i)}}(\mathbf{y})\mathbf{n}$ ,  $\mathbf{J}_{\boldsymbol{\eta}_{(i)}}(\mathbf{y})\mathbf{n}$  corresponding to matrix-vector products with an i.i.d. *zero-mean unit-variance* (e.g., *Gaussian*) random vector  $\mathbf{n}$ . Then the desired trace in  $GCV(\lambda)$  and  $PSURE(\lambda)$  can be stochastically well-approximated as  $\operatorname{tr}\{\mathbf{A}\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\} \approx \mathbf{n}'\mathbf{A}\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\mathbf{n}$  [5]. Thus while running the SB algorithm for a given  $\lambda$ , we propose to simultaneously manipulate  $\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\mathbf{n}$ , stochastically estimate  $\operatorname{tr}\{\mathbf{A}\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\}$  and compute  $GCV(\lambda)$  and  $PSURE(\lambda)$  at every iteration  $i$ .

**Experimental Setup:** We acquired 10 independent sets of fully-sampled 2-D data ( $256 \times 256$ ) of a GE-phantom using a GE 3T scanner (GRE sequence with flip angle =  $35^\circ$ ,  $T_R = 200$  ms,  $T_E = 7$  ms, FOV = 15 cm). 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"  $\mathbf{x}$  (Fig. 2a) for computing  $MSE(\lambda) = \|\mathbf{x} - \mathbf{u}_\lambda(\mathbf{y})\|^2$  and  $PSNR(\lambda) = 10 \log_{10}[N \max\{\mathbf{x}\}^2 / MSE(\lambda)]$ . We separately acquired dummy-data (with the same scan setting) when no RF field was applied and used it to estimate  $\sigma^2$  (for PSURE) by the empirical variance. We retrospectively undersampled data from one of the 10 sets along the phase-encode (PE) direction in a random fashion (with 16 central fully-sampled PEs, see Fig. 2b). We ran the SB algorithm with TV regularization and tuned  $\lambda$  so as to minimize  $GCV(\lambda)$  and  $PSURE(\lambda)$  individually.

**RESULTS:** We plot  $PSNR(\lambda)$  versus  $\lambda$  in Fig. 1 where we also indicate (by vertical dashed lines)  $\lambda$ s obtained by minimizing  $GCV(\lambda)$  and  $PSURE(\lambda)$ . GCV-based selection is away, while PSURE-based selection is close to the MSE-optimal- $\lambda$ . The corresponding PSURE-based reconstruction (Fig. 2e) is therefore visually similar to the MSE-optimal one (Fig. 2d), while that due to GCV (Fig. 2f) exhibits slightly more artifacts and reduced PSNR. We obtained similar results for different undersampling rates in the above experiment and in various simulations.

**CONCLUSIONS:** We demonstrated the feasibility of using GCV and PMSE-estimation (via PSURE) for automated adjustment of the regularization parameter ( $\lambda$ ) for nonlinear MRI reconstruction using the split-Bregman (SB) algorithm. GCV and PSURE require the trace of a linear transformation of the Jacobian matrix ( $\operatorname{tr}\{\mathbf{A}\mathbf{J}_{\mathbf{u}_{(i)}}(\mathbf{y})\}$ ) that we estimated stochastically and iteratively for the SB algorithm. We illustrated with experiments on real MR (phantom) data that GCV and PSURE are able to provide near-MSE-optimal selection of  $\lambda$ . The techniques discussed here can also be extended, in principle, to other regularization criteria, reconstruction algorithms, and nonCartesian and parallel MRI.

**ACKNOWLEDGEMENTS:** This work was supported by the National Institutes of Health under Grant 1P01 CA87634.

**REFERENCES:** [1] Lustig *et al.*, 2007, MRM, 58: 1182-95; [2] Karl, 2005, *Handbook Im. Vid. Proc.*, 183-202; [3] Vogel, 1996, *Im. Prob.*, 12: 535-47; [4] Craven *et al.*, 1979, *Num. Math.*, 31: 377-403; [5] Giryes *et al.*, 2011, *App. Comp. Harm. Anal.*, 30: 407-22; [6] Deshpande *et al.*, 1991, *Curv. Surf.*, 143-8; [7] Goldstein *et al.*, 2009, *SIAM J. Img. Sci.*, 2: 323-43; [8] Ramani *et al.*, 2011, *IEEE TMI*, 30: 694-706; [9] Hjorungnes *et al.*, 2007, *IEEE TSP*, 55: 2740-6; [10] Ramani *et al.*, 2011, *IEEE TIP*, submitted; [11] Stein, 1981, *Ann. Stat.*, 9: 1135-51.

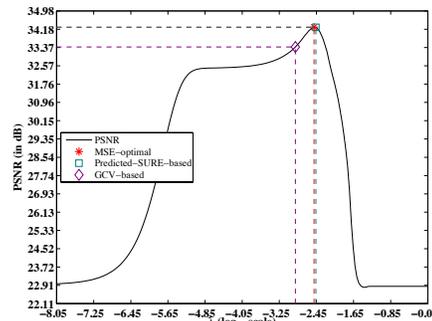


Fig. 1: Plot of  $PSNR(\lambda)$  versus  $\lambda$ .

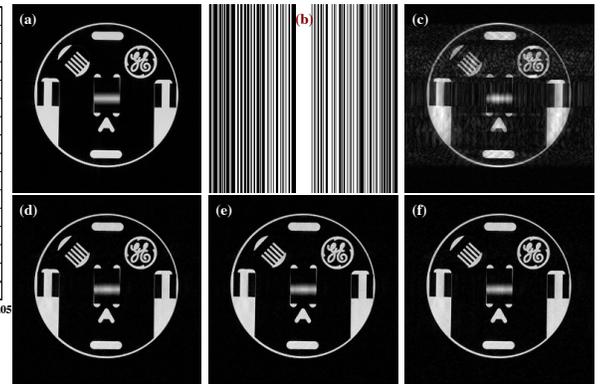


Fig. 2: (a) Reference  $\mathbf{x}$ ; (b) Retrospective phase-encode undersampling (50%); (c) Zero-filled reconstruction (22.51 dB); TV-reconstructions that minimize (d) true MSE (34.3 dB); (e) PSURE (34.3 dB); (f) GCV (33.4 dB).