SURE-based Automatic Parameter Selection For ESPIRiT Calibration

Siddharth Iyer¹  |  Frank Ong¹  |  Mariya Doneva²  |  Michael Lustig¹

¹Department of Electrical Engineering and Computer Science, University of California, Berkeley, California, 94720, United States of America
²Philips Research Laboratories, Hamburg, Germany

Purpose
Parallel imaging methods in MRI have resulted in faster acquisition times and improved noise performance. ESPIRiT is one such technique that estimates coil sensitivity maps from the auto-calibration region using an eigenvalue-based method. This method requires choosing several parameters for the map estimation. Even though ESPIRiT is fairly robust to these parameter choices, occasionally, poor selection can result in reduced performance. The purpose of this work is to automatically select parameters in ESPIRiT for more robust and consistent performance across a variety of exams.

Theory and Methods
Stein’s unbiased risk estimate (SURE) is a method of calculating an unbiased estimate of the mean squared error of an estimator under certain assumptions. We show that this can be used to estimate the performance of ESPIRiT. We derive and demonstrate the use of SURE to optimize ESPIRiT parameter selection.
Modern MRI leverages coil arrays and parallel imaging for faster acquisition-time and improved signal to noise ratio (SNR). The two most common parallel imaging techniques are SENSE [2], which uses a calibration scan to derive explicit spatial coil sensitivity information, and GRAPPA [3], which derives data consistency operators from either auto-calibration signal (ACS) lines or a separate auto calibration scan. ESPIRIT [1] is a method that bridges SENSE and GRAPPA by using ACS to derive so called ESPIRiT maps that can be used in a SENSE-like reconstruction.

The high level mechanics of all parallel imaging methods can be broken up into two steps: the calibration step and the reconstruction step. The calibration step typically involves exploiting data redundancy to derive data consistency operators that define the subspace the signal lives in; and the reconstruction step then enforces data consistency by using these derived operators along with other priors to reconstruct the image. For example, GRAPPA derives k-space consistency kernels from ACS that are used to interpolate missing data in k-space; and ESPIRiT maps define a linear subspace that the desired image is expected to lie in.

In this work, the calibration step of ESPIRiT is studied. There are several user-set parameters in the calibration that determine the effectiveness of the resulting ESPIRiT maps. These parameters are described in the theory section. While ESPIRiT is in general robust to variation in these parameter, occasionally, poor parameter choice can result in reduced performance, and image artifacts. This is exemplified in Fig. 1, where the resulting ESPIRiT maps exhibit significant variations with different parameter choices. The purpose of this work is automatic, data driven parameter selection in ESPIRiT for more consistent performance across a variety of scans for robust clinical use.

Much of the discussion associated to parameter choices in parallel imaging techniques have been on regularizers that enforce priors in the reconstruction step. There have been numerous works done on automatically optimizing for these reconstruction parameters using Stein's unbiased risk estimate (SURE) [4]. Ramani et al. presented in [5] a framework for tuning non-linear reconstructions based on their respective Jacobians evaluated on acquired data; Marin et al. presented in [6] a parameterized wavelet-based estimator that uses SURE to determine the optimal parameters for reconstruction; Weller et al. presented in [7] a method of selecting regularization for GRAPPA to near-optimally balance the contributions of GRAPPA and sparsity; and Weller et al. presents in [8] a Monte Carlo SURE method of
FIGURE 1  Variability in ESPRIT maps is exemplified by varying the signal subspace size \( w \) in WNSVN and eigenvalue crop threshold \( c \). These parameters are explained in the theory section. ESPRIT maps within the blue highlights are desirable, as they capture the anatomy without cropping any signal. The green maps are considered good, as they capture anatomy but also allow some signal outside of the field of view of the object in. The red maps are undesirable because they attenuate signal.

In this work, we focus on using SURE to automatically select parameters in the calibration step of ESPRIT. We propose a SURE-based approach to quantify the performance of ESPRIT that uses the self-consistency aspect of parallel imaging techniques under the assumption that noise is additive, complex normal noise. We formulate the calibration step of ESPRIT to be the construction of a linear, self-consistency operator from particular parameter choices and we quantify the performance of this operator using the SURE principle. We then verify the accuracy of SURE as an estimator with simulation experiments and demonstrate the reliability of this method with in-vivo experiments.

2  |  THEORY

2.1  |  ESPRIT

ESPRIT is a technique that combines the data-based calibration advantages of GRAPPA to derive SENSE-like relative coil-sensitivity maps. It exploits the data redundancy introduced through the use of multiple receiver channels by characterizing the null space of a matrix (called the auto-calibration matrix) constructed by sweeping a kernel through the calibration region (as depicted in Figure 2). This information is then used to describe a linear subspace that the desired signal is expected to reside in. In the following, we will provide a brief overview of ESPRIT, with emphasis on the
This summarizes the technique of generating ESPIRiT maps that are used as SENSE-like projection operators. (a) A kernel is swept through the calibration region to construct the calibration matrix (which consequently has block Hankel structure). (b) The SVD of the calibration matrix is taken and (c) the right singular vectors corresponding to the largest singular values are reshaped into k-space kernels. (d) The inverse Fourier transform of these kernels are taken, followed by (e) the eigenvalue decomposition of each pixel along the coil dimension in the image domain. The eigenvectors corresponding to eigenvalues \( \approx 1 \) are used to construct the ESPIRiT operator. (f) depicts how ESPIRiT maps can be used as a projection operator to denoise data in the ideal case. (g) Consequently, the null space of the ideal projection operator would contain only noise.

Let \( A \) be the auto-calibration matrix, \( V_\parallel \) be a matrix consisting of the right singular vectors of \( A \) corresponding to the dominant singular values of \( A \) and \( V_\perp \) be a matrix consisting of the remaining singular vectors that span the null space of \( A \). Then,

\[
A = U \Sigma V^* \quad \text{with} \quad V = \begin{bmatrix} V_\parallel & V_\perp \end{bmatrix}
\]

Let \( y \) be the true, underlying multi-channel k-space data. Let \( R_r \) be an operator that extracts a block from \( y \) around the k-space position \( r \) (including \( r \) itself) across all the channels. Since the signal \( y \) should be orthogonal to the null-space
of $A$, we have the following normal equations:

$$\left( \sum_r R^*_r V_V^* R_r \right) y = 0 \quad (2a)$$

$$\left( \sum_r R^*_r \left( I - V_V^* \right) R_r \right) y = 0 \quad (2b)$$

$$\left( \sum_r R^*_r R_r \right)^{-1} \left( \sum_r R^*_r \left( V_V^* \right) R_r \right) y = y \quad (2c)$$

$(\sum_r R^*_r R_r)^{-1}$ effectively scales each channel by a scalar that is the inverse of the number of k-space elements selected by $R_r$. Thus, ‘$W$’ is a convolution with a matrix-valued kernel where the matrix operates on the channel dimension. This convolution operator is decoupled into pixel-wise operations along the channel dimension in the image domain. In other words, let $x$ be the true multi-channel image data such that $y = F x$ where $F$ is the unitary Fourier transform operator. Equation (2c) becomes,

$$(F^* W F) x = x, \quad (3)$$

where $G$ is defined to be $F^* W F$. Particularly, since $G$ is decoupled into pixel-wise operators, we can look at the effect of $G$ on a particular image pixel index $q$.

$$G(q) x(q) = x(q) \quad (4)$$

$x(q)$ is a vector of dimension equal to the number of coil-sensitivity maps. Let the source image be $m$ and let $S$ be the vector constructed from stacking the coil sensitivities of the different channels. The SENSE model tells us that,

$$x(q) = S(q) m(q) \quad (5)$$

$m(q)$ is a scalar and $S(q)$ is a vector of dimension equal to the number of coil-sensitivity maps. Applying this to (4), we get the following:

$$G(q) S(q) m(q) = S(q) m(q) \quad (6)$$
If $m(q) \neq 0$, we get the following eigenvalue-eigenvector condition.

$$G(q)S(q) = 1 \times S(q)$$  \hspace{1cm} (7)

This tells us that sensitivity maps, in the ideal case, are eigenvectors of $G$ with eigenvalues one, with the other eigenvectors of $G$ having eigenvalues much smaller than one. (This comes from observing that $W$ is an average of projections and is consequently positive semi-definite with eigenvalues smaller or equal to one.) In practice, due to data-inconsistencies (like noise), the observed eigenvalues of the eigenvector maps are very close to (but not exactly) one. This motivates defining an approximate “$\approx 1$” condition where the eigenvalues that would be 1 in the ideal case but are instead close to one. Thus, we can take the eigenvalue decomposition of $G$ and consider the eigenvector corresponding to eigenvalue “$\approx 1$” to be ESPIRiT maps that function as sensitivity maps.

In some cases, multiple eigenvalues “$\approx 1$” appear such as when the calibration region supports a smaller field-of-view than the object which results in multiple sensitivity values at a pixel location due to aliasing [1]. This motivates using more that one set of these eigenvector maps to better capture the desired signal.

2.2 Parameter choices in ESPIRiT

There are three parameters in the ESPIRiT calibration. The first is the size of the window, or kernel size, that is swept through ACS to construct the auto-calibration matrix $A$, the second is the size of the signal subspace used to partition $V_{||}$ from $V_{\perp}$, and the third is the threshold condition used to recognize eigenvalues that would ideally be one. The parameters are denoted the kernel size ($k$), the subspace size ($w$) and the eigenvalue crop threshold ($c$) respectively.

Let $n_c$ be the number of channels. For 2D data, a kernel size ($k$) would imply that window of dimensions ($k \times k \times n_c$) is swept through the calibration region to construct the rows of the auto-calibration matrix $A$. This, in turn, implies that $V_{||}$ and $V_{\perp}$ together span a linear space of dimension ($k \times k \times n_c$), and that the rank of $V_{||}$ can vary from a minimum of 0 to a maximum of ($k \times k \times n_c$). The chosen rank of $V_{||}$ is the subspace size ($w$) and is measured in Window Normalized Singular Values Number (WNSVN). This normalizes the rank by the kernel spatial dimensions and thus ($w$) is in the range from 0 to $n_c$. Succinctly, a subspace size of ($w$) would imply that $V_{||}$ consists of $w \times k^2$ orthogonal vectors. Finally, the eigenvalue crop threshold ($c$) determines the pixel positions ($q$) where the eigenvectors of operator $G(q)$ are well defined, which corresponds to pixels positions ($q$) within the object’s FOV. Too high a threshold would result in direct attenuation of the signal, and too small a threshold would allow eigenvectors of operator $G(q)$ from positions that are not well defined in terms of [7], which in turn allows in signal that is not necessarily from the object, such as noise. The feasible values of the eigenvalue crop threshold ($c$) are from 0 to 1, with realistic values residing in the range from 0.8 to 0.95.

While ESPIRiT is fairly robust to these parameter choices, there is variability in map performance (such as how well the maps capture the field of view of the object) and choosing parameters that result in optimal performance is desirable. Figure 1 exemplifies the variability in ESPIRiT maps when varying the subspace size ($w$) and eigenvalue crop threshold ($c$) for a fixed kernel size ($k$).

In order to develop a robust, data-driven method of automatically picking parameters, we will explore using Stein’s unbiased risk estimate (SURE) as a metric to select parameters that are optimal in an expected mean squared error sense.
2.3 | Stein’s unbiased risk estimate

Stein’s unbiased risk estimate is a data-driven method of calculating the expected mean squared error of an estimator in the presence of zero-mean, additive, normal noise, as long as the estimator is differentiable with respect to the data almost everywhere [4]. We will present the SURE expression for a self-adjoint linear operator and then extend it to ESPRIT. For the definition and properties of self-adjoint operators, we refer the reader to [9].

Let \( P_\theta \in \mathbb{C}^{m \times m} \) be the self-adjoint linear operator parameterized by \( \theta \). Let \( x \in \mathbb{C}^m \) be the ground truth to be estimated; \( n \) be zero-mean, additive, Gaussian complex noise with standard deviation \( \sigma \); and \( x_{acq} = x + n \) be the acquired data. Let \( \mathbb{E} \{ \cdot \} \) denote the expected value operation. Partitioning our complex vector space into real and imaginary parts and noting that the divergence of a linear operator is the trace of the linear operator (steps described in the appendix), Stein’s first theorem [4] tells us the following:

\[
\mathbb{E} \left\{ \| P_\theta x_{acq} - x \|_2^2 \right\} = \mathbb{E} \left\{ \text{SURE} P_\theta (x_{acq}) \right\} \quad (8a)
\]

\[
\text{SURE} P_\theta (x_{acq}) = -m\sigma^2 + \| (P_\theta - I) x_{acq} \|_2^2 + 2\sigma^2 \left\{ \text{div} x_{acq} (P_\theta) \right\} (x_{acq})
\]

\[
= -m\sigma^2 + \| (P_\theta - I) x_{acq} \|_2^2 + 2\sigma^2 \text{trace} (P_\theta) \quad (8b)
\]

Particularly, note that Equation (8b) is independent of \( x \), and only depends on the acquired data \( x_{acq} \). We can thus use SURE as a surrogate for the expected mean squared error to find the optimal parameters \( \theta^* \).

2.4 | SURE with ESPRiT

We first illustrate the main concepts through a non-accelerated, densely-sampled, high-resolution case. In this case, we can apply a Discrete Fourier Transform to the multi-channel data and consider our measurements \( x_{acq} \) to be multi-channel images. The requirement of zero-mean additive normal noise is satisfied. We will later extend to the case when we are restricted to the densely-sampled, low-resolution calibration region.

In order to use SURE to tune ESPRiT parameters, we must quantify the quality of the maps through denoising. To do this, we will define an ESPRiT projection operator, which denoises the acquired data using ESPRiT maps. An illustration is provided in Fig. 2.

Concretely, let \( n_c \) be the number of coils. Let \( S^i(q) \) be the \( i^{th} \) eigenvector of \( \mathcal{G}(q) \) with eigenvalue \( \lambda_j(q) \). \( S^i(q) \) is a vector of dimension \( n_c \) and has unit norm. The ESPRiT projection operator at image position \( q \), which will be denoted as \( P(q) \), is defined as:

\[
P(q) = \begin{bmatrix}
S^1(q) & \cdots & S^{n_c}(q) \\
| & \cdots & | \\
| & \cdots & | \\
S^1(q) & \cdots & S^{n_c}(q)
\end{bmatrix}^*
\quad (9)
\]
The aggregate ESPIRiT projection operator $P$ can be represented by stacking the pixel-wise operators diagonally:

$$
P = \begin{bmatrix}
P(1) & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & P(N)
\end{bmatrix}
$$

where $N$ is the number of image pixels. (10)

Succinctly, the ESPIRiT projection operator is a linear projection operator whose range space describes the linear subspace the desired signal is expected to reside in. Consequently, the projection onto this subspace is expected to get rid of undesirable signal such as additive, white, Gaussian noise that contaminates MRI data. This interpretation of the ESPIRiT operator as a denoiser allows us to apply SURE to ESPIRiT.

Let $P_\theta$ be the projection operator derived from a particular ESPIRiT parameter set $\theta = (k, w, c)$. Define $x$ be the densely-sampled, high-resolution, noise-less, multi-channel coil images; $n$ be additive, complex normal noise of standard deviation $\sigma$; $x_{\text{acq}} = x + n$ be the acquired coil images; and $I$ be the identity operator. In the densely-sampled, high-resolution case, finding the optimal projection operator is reduced to finding the optimum $\theta^*$ that results in a projection operator $P_\theta^*$ that best denoises the input data $x_{\text{acq}}$. That is to say,

$$
\theta^* = \arg \min_\theta \| x - P_\theta x_{\text{acq}} \|_2^2
$$

Since ESPIRiT is a pixel-wise linear projection operator in the image domain, the divergence contributed by a single pixel $x(q)$ is the trace of the linear operator affecting that pixel (denoted $P_\phi(q)$) (described in the appendix). We can thus calculate the SURE value of $P_\theta$ by summing over all pixel positions $q$.

$$
\text{SURE}_\theta(x_{\text{acq}}) = \sum_q \left[ -n_c \sigma^2 + \| (P_\theta(q) - I) x_{\text{acq}}(q) \|_2^2 + 2\sigma^2 \text{trace} P_\theta(q) \right]
$$

Equation (12) can then be used as a surrogate for (11).

$$
\theta^* = \arg \min_\theta \| P_\theta x_{\text{acq}} - x \|_2^2 \approx \arg \min_\theta \text{SURE}_\theta(x_{\text{acq}})
$$

Thus, in the above case, we can sweep through different values of $\theta$ to determine the optimal projection operator $P_\theta$ for calibration. We use (13) as a basis to to optimize for the ESPIRiT projection operator. We will present a variant of (13) for the case in which we are limited to only the ACS data and noise information.

### 2.5 Accelerated case with the auto-calibration signal

In order to accelerate parallel imaging scans while still being auto-calibrating, we have to estimate the ESPIRiT operator given only the low-resolution, densely sampled ACS data. To do this, we incorporate a Fourier sampling operator into the projection operator $P$ and enforce data consistency within the ACS data in k-space.

Define $R$ to be the operator that outputs an ACS region from densely-sampled k-space. We define a new, augmented
projection operator $\mathbf{P}_\theta^R$ as follows:

$$\mathbf{P}_\theta^R = R F S S^* F^* R^*$$  \hspace{1cm} (14)$$

Let $y$ be the noise-less, densely-sampled, low-resolution auto-calibration region in k-space; $n$ be additive, zero-mean, complex normal noise of standard deviation $\sigma$; and $y_{\text{acq}} = y + n$ be the acquired auto-calibration region. Then, interpreting (14) as a denoiser of ACS data yields the following SURE expression:

$$\text{SURE}_\theta(y) = \left\| \left( \mathbf{P}_\theta^R - I \right) y_{\text{acq}} \right\|_2^2 + 2\sigma^2 \text{trace} \mathbf{P}_\theta^R + C$$  \hspace{1cm} (15)$$

Here, $C$ is some constant term that we ignore because it does not affect the minimum. Enforcing consistency of ACS data and the corresponding SURE expression is seen to be a good representative of the performance of ESPRIT maps. In particular, the parameters ($\theta$) obtained from minimizing (12) and (15) often correspond in practice. This allows us to sweep through the different parameters in ESPRIT to obtain the ESPRIT projection operator that results in near-optimal performance in the expected mean squared error sense even when restricted to ACS data.

2.6 Soft-threshold based weighting for subspace estimation

In practice, sweeping through different kernel sizes and signal subspace sizes (or the rank of $V_j$) is computationally intensive whereas sweeping through different thresholds to determine the eigenvalue crop threshold ($c$) is relatively quick. To aid viability, we present a heuristic that appropriately weights the right singular vectors based on singular value soft-thresholding as an alternative to sweeping through different rank values.

Since we are observing the same data through multiple channels, the auto-calibration matrix $A$ is expected to be low rank. However, due to noise and other data inconsistencies present in the data, the observed auto-calibration matrix often has full rank. We can construct a low-rank matrix estimate of $A$ by hard thresholding the singular values of $A$. Ideally, we would like to use SURE to determine the optimal low-rank matrix estimate in the sense of “denoising” the matrix. However, this is difficult to do since a hard threshold is not weakly-differentiable. A common alternative is to soft-threshold the singular values.

Consider the singular value decomposition of $A$ in its dyadic form.

$$A = \sum_j s_j u_j v_j^*$$  \hspace{1cm} (16)$$

Here, $u_j$ are the left singular vectors, $v_j$ are the right singular vectors and $s_j$ are the singular values. A soft-threshold low rank matrix estimate of $A$, which will be denoted as $\hat{A}$, is constructed by soft-thresholding the singular values by threshold $\lambda$.

$$\hat{A} = \sum_j (s_j - \lambda)_+ u_j v_j^*$$  \hspace{1cm} (17)$$

The advantage of this formulation is that we can utilize the computationally efficient SURE-based method presented in [10] to find the optimal $\lambda$ in (17). We denote the output threshold as $\lambda^*$. We refer the reader to [10] for the derivation.
FIGURE 3  This illustrates the motivation behind weighting the signal subspace using weights derived from the soft-threshold. A hard threshold on the singular values can be modeled as a multiplication of the right singular vectors with a diagonal matrix whose diagonal entries represent binary decisions which in turn determines the subspace size as in Equation \((18)\). A soft-threshold based weighted subspace can be described as the matrix multiplication of the right singular vectors with a diagonal matrix whose diagonal entries represent our confidence in that vector being a part of the signal subspace as in Equation \((19)\).

To motivate weighting the singular vectors, consider the following. The subspace selection problem can be modeled by a hard threshold on the singular values using a threshold \(\lambda\).

\[ V_{||} = VW \text{ where } W \text{ is a diagonal weight matrix with } W_{ii} = \begin{cases} 1, & (s_i > \lambda) \\ 0, & \text{otherwise} \end{cases} \]  

\((18)\)

Instead, we weight the singular vectors instead with its soft-threshold variant that is used in \((17)\). We will denote this weighted subspace as \(V_{||}^w\).

\[ V_{||}^w = VW \text{ where } W \text{ is a diagonal weight matrix with } W_{ii} = \frac{(s_i - \lambda)^+}{s_i} \]  

\((19)\)

We use \(\lambda^*\) as calculated by Candès’ SURE method in \((19)\) to get a weighted subspace estimate \(V_{||}^w\) and plug it into \((20)\) instead of \(V_{||}\). This is illustrated in Figure 3.
3 METHODS

In the spirit of reproducible research, we provide the source code used to conduct the following simulation experiments. MATLAB (MathWorks, Natick, MA) code can be downloaded from: [https://github.com/mikgroup/auto-espirit](https://github.com/mikgroup/auto-espirit) (doi:10.5281/zenodo.1401470). Furthermore, the soft-threshold based weighting subspace heuristic and automated eigenvalue crop threshold selection for ESPIRiT has been written in the C language as a part of the BART Toolbox [11] from version 0.4 onwards (doi:10.5281/zenodo.216718). The latest version of BART can be obtained from: [https://mrirecon.github.io/bart/](https://mrirecon.github.io/bart/). To verify the efficacy of our technique, we conduct simulation experiments to compare the true squared error and the expected risk as calculated by SURE for ESPIRiT. Further more, we demonstrate feasibility by conducting in-vivo experiments using ESPIRiT.

3.1 Simulation Experiments

Fully-sampled, high-resolution data of the human brain was acquired on a 1.5T scanner (GE, Waukesha, WI) using an eight-channel coil for multiple subjects. It was obtained using inversion-recovery prepared 3D RF-spoiled gradient-echo sequence with the following parameters: \( T_R/T_E = 12.2/5.2 \) ms, \( T_I = 450 \) ms, \( FA = 20^\circ \), \( BW = 15 \) kHz, and a matrix size of \( 256 \times 180 \times 230 \) with \( 1 \) mm isotropic resolution. This was done to have a ground-truth to compare against and verify the accuracy of SURE as an estimator of the mean squared error. The 3D dataset was Fourier transformed along the readout dimension and a slice along the readout dimension was taken. ESPIRiT maps were calculated from this slice and the projection of this slice onto the ESPIRiT maps was considered to be the true, underlying ground truth. The ground truth has dimensions \( 230 \times 180 \times 8 \), where \( 8 \) is the number of channels, and an \( l_2 \) - norm of 8369.46. Additive complex k-space noise of standard deviation 10 was retrospectively added to the ground-truth and the result was considered to be the acquired data.

Parameters \((\theta)\) were varied to get the corresponding projection operator \((P_\theta)\). The mean squared error between the estimated k-space and ground truth was compared to the SURE value. Furthermore, the ability to determine optimal parameters while being limited to ACS data was verified.

For different parameter values \((\theta)\), we generated ESPIRiT maps and calculated the true squared error between the projection and the ground truth. We compared this to the SURE value calculated. We also tested the soft-threshold based weighting heuristic by similarly varying kernel sizes and eigenvalue crop thresholds and comparing the true squared error to SURE. For all the previous cases, we also tested the case in which we are restricted to low-resolution, densely-sampled ACS data.

In Experiment (a), a fixed kernel size of 6 was used and the subspace size and eigenvalue crop thresholds were varied. For each subspace size and eigenvalue crop threshold, the true squared error, SURE value given high-resolution, densely sampled data, and the SURE value given low-resolution, densely-sampled ACS data was calculated. The latter curve was normalized by a constant so as to better compare the minimums of each of the curves.

In Experiment (b), a fixed kernel size of 6 was used along with the soft-threshold based subspace weighting heuristic, and the eigenvalue crop threshold was varied. For each eigenvalue crop threshold, the true squared error, SURE value given high-resolution, densely sampled data, and the SURE value given low-resolution, densely-sampled ACS data was calculated. Once again, the latter curve was normalized by a constant to better compare the minimums of each of the curves.

In Experiment (c), the same three curves were calculated from varying the kernel size, subspace size and eigenvalue crop threshold parameters. The minimums given a particular kernel size across the subspace thresholds and eigenvalue crop thresholds for that kernel size were taken. This experiment was conducted to test the dependence of ESPIRiT maps...
on the kernel size \((k)\) assuming optimal subspace size and eigenvalue crop threshold for that particular kernel size \((k)\).

In Experiment (d), the same three curves were calculated from varying the kernel size and eigenvalue crop threshold parameters while using the soft-threshold based subspace weighting heuristic. The minimums across the eigenvalue crop thresholds given a particular kernel size were taken. This experiment was conducted to test the dependence of ESPRiT maps on the kernel size \((k)\) assuming an optimal eigenvalue crop threshold for that particular kernel size \((k)\).

### 3.2 Calibration With Aliasing Due To FOV Smaller Than Object

We applied our SURE-based parameter selection to the same data used in [1] that demonstrated ESPRiT’s robustness to aliasing due to the calibration region supporting a FOV smaller than the object. We used the same retrospectively
2-fold under-sampled 2D spin-echo dataset \((T R / T E = 550/14 ms, F A = 90^\circ, B W = 19 k H z, m a t r i x: 320 \times 168, s l i c e t h i c k n e s s: 3 m m, 24 r e f e r e n c e l i n e s)\) with an FOV of \((200 \times 150) m m^2\), acquired at 1.5T using an 8-channel head coil. The data was Fourier transformed to the image domain and noise variance was estimated from a corner of the image data that did not contain any desired signal. To determine the ESPRiT maps, SURE values were calculated for eigenvalue crop thresholds varied from 0.7 to 0.99, with a fixed kernel size of 6 and the soft-threshold based subspace weighting heuristic. The ESPRiT map corresponding to the lowest SURE value was then calculated to verify robustness to aliasing.

### 3.3 In-Vivo Experiments

Two pre-whitened 3D accelerated datasets were acquired on a 3T Achieva scanner (Philips, Best, The Netherlands). The acquisitions were \(T_1\)-weighted, TFE datasets acquired using poisson disk under-sampling with \(18 \times 18\) ACS lines. These datasets are noted as Dataset A and Dataset B. A unitary inverse Fourier transform was taken along the readout direction and a slice was extracted. ESPRiT calibration with parallel imaging and compressed sensing reconstructions (with an \(l_1\) regularization of 0.01) were executed on the datasets using ESPRiT’s default parameters (described below). ESPRiT calibration was then done with parameters selected by the SURE-based method with the soft-threshold based weighting heuristic. An identical reconstruction was performed with the SURE-calibrated ESPRiT maps to compare the results.

The reconstructions used BART [11]. The manual parameters used were a kernel size of 5, a subspace size that corresponded to the number of singular values greater that 0.001 times the largest singular value, and an eigenvalue crop threshold of 0.8. For the SURE-based method, a fixed kernel size of 5 is used along with the soft-threshold based weighted subspace estimate. The eigenvalue crop threshold was varied from 0.7 to 0.99 with a step size of 0.01, and the one corresponding to the minimum SURE value as calculated from the ACS data was used.

### 4 RESULTS

#### 4.1 Simulation Results

The simulation results are illustrated in Figure[4] in Experiment (a), we see that the true squared error calculated with the densely-sampled, high-resolution data and SURE correspond well. Further more, it is seen that SURE as calculated from ACS data has a minimum close to the minimum of the true squared error. An identical result is seen in Experiment (b), where the true squared error and SURE given densely-sampled, high-resolution data line well with the SURE as estimated from ACS data having a minimum that corresponds to the minimum of the true squared error. Experiments (c) and (d) show identical results. It is interesting to note that by auto-tuning the other parameters, ESPRiT does not seem
FIGURE 4  ESPIRIT simulation experiment results. In (a), a fixed kernel size of 6 is used and the subspace size and eigenvalue crop thresholds are varied. The figure shows the true squared error, SURE given densely-sampled, high-resolution data and normalized SURE calculated from ACS data as the subspace size and the eigenvalue crop threshold are varied. In (b), the eigenvalue crop threshold is varied with a fixed kernel size of 6 along with the soft-threshold based subspace weighting heuristic. The figure shows the true squared error, SURE given densely-sampled, high-resolution data and normalized SURE calculated from ACS data as the eigenvalue crop threshold is varied. In (c), kernel size, subspace size and eigenvalue crop threshold parameters are varied, and the minimum given a particular kernel size is taken. The figure shows the true squared error, SURE given densely-sampled, high-resolution data and normalized SURE calculated from ACS data as functions of the kernel size assuming optimal subspace size and eigenvalue crop threshold given the particular kernel size. The purple crosses in (a), (b) and (c) show the locations of the respective minimum values. (d) is similar to (c), except that instead of calculating the optimal subspace size given the kernel size, the soft-threshold based subspace weighting heuristic is used. (b), (c) and (d) share the same legend.

4.2  |  Calibration With Aliasing Due To FOV Smaller Than Object Result

The simulation results are illustrated in Figure 5. It is seen that the SURE-based parameter selection results in parameters that retain ESPIRIT's robustness to image aliasing. Particularly, the second set of ESPIRIT maps capturing the aliased signal tightly about its FOV allows us to more confidently use a second ESPIRIT map.

4.3  |  In-Vivo Results

The results are illustrated in Figures 6 and 7. They compare the ESPIRIT calibration maps obtained from using the manual parameters (described in the methods section) and the SURE-based method. The manual parameter calibration maps are denoted "Manual", and the SURE-calibrated maps are denoted "SURE". Note that the manual parameters to be significantly dependent on the kernel size.

On simulated data that fits the model, it is seen that SURE is an accurate estimator of the squared error. Furthermore, it is seen that restricting ourselves to ACS data results in near-optimal parameter choice.
results in ESPRiT maps that capture more than the desired signal, particularly in the second map. The SURE-calibrated maps better fit the desired signal without any support in the second map, which is desirable in this case. Since the SURE calibrated maps better capture the desired signal than the manual parameter calibrated maps, the resulting parallel imaging and compressed sensing reconstruction using the SURE-calibrated maps does not result in a second map image.

5 | DISCUSSION

By using SURE as a metric to determine parameters, we obtain parameter choices that result in consistent performance across different datasets. Further more, since SURE is used as a proxy to the squared error, the resulting parameter choices are optimal in this expected error sense. In practice, with the parameter ranges mentioned in the in-vivo experiment, the resulting parameter choices tend to optimize for SNR performance while causing no signal attenuation.

One restriction of the presented technique is that we have to test each element from a finite set of parameter choices in order to make the optimal choice. This adds overhead as it is computationally expensive. To overcome this, we presented the soft-threshold based weighted subspace estimate. We also made design choices in setting the range of eigenvalue crop threshold choices and using a fixed kernel size. This is because Figure suggests that optimizing over kernel sizes would give us diminishing returns. While we did choose a kernel size of 6, an argument can be made
This figure compares the ESPIRiT calibration maps obtained from using the manual parameters described in the methods section and the SURE-based method. The manual parameter calibration maps are denoted "Manual" and the SURE-calibrated maps are denoted "SURE". Note that the manual parameters results in ESPIRiT maps that capture more than the desired signal, particularly in the second map. The SURE-calibrated maps better fit the desired signal without any support in the second map, which is desirable in this case.

about using a smaller kernel size (say, 3) when auto-picking parameters which would improve computation performance with respects to calculating the SVD of the calibration matrix. Having made these design choices, in our experience, the computational overhead of our method on in-vivo data was not significant.

This method offers a robust, data-consistent metric. By adapting to the noise level, this technique optimizes maps according to the expected mean squared error while avoiding signal attenuation.

For practical usage, we recommend a kernel size of 4 – 6 with eigenvalue crop thresholds varying from 0.7 to 0.99 with a step size of 0.01 along with the soft-threshold based weighted subspace heuristic.

6 | CONCLUSIONS

Using SURE as a metric to determine parameter in ESPIRiT allows for automatic parameter selections that are optimal in an expected mean squared error sense. The efficacy of SURE as an estimator for the mean squared error and the resulting optimal performance of parallel imaging methods are verified. This allows for data-driven, consistent parameter selection. The validity of this method has also been verified using in-vivo experiments.
FIGURE 7  This figure demonstrates the resulting parallel imaging and compressed sensing reconstruction using the manually calibrated ESPIRiT maps and the SURE-calibrated ESPIRiT maps. The SURE calibrated maps better capture the desired signal while the manual parameter calibrated maps result in some noise being captured in the second map.

REFERENCES

[1] Uecker M, Lai P, Murphy MJ, Virtue P, Elad M, Pauly JM, et al. ESPIRiT - an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. Magnetic resonance in medicine 2014;71(3):990–1001.

[2] Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P, et al. SENSE: sensitivity encoding for fast MRI. Magnetic resonance in medicine 1999;42(5):952–962.

[3] Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic resonance in medicine 2002;47(6):1202–1210.

[4] Stein CM. Estimation of the mean of a multivariate normal distribution. The annals of Statistics 1981:p. 1135–1151.

[5] Ramani S, Liu Z, Rosen J, Nielsen JF, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. IEEE Transactions on Image Processing 2012;21(8):3659–3672.

[6] Marin A, Chaux C, Pesquet JC, Ciuciu P. Image reconstruction from multiple sensors using Stein's principle. Application to parallel MRI. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro IEEE, 2011. p. 465–468.

[7] Weller DS, Ramani S, Nielsen JF, Fessler JA. SURE-based parameter selection for parallel MRI reconstruction using GRAPPA and sparsity. In: 2013 IEEE 10th International Symposium on Biomedical Imaging IEEE; 2013. p. 954–957.

[8] Weller DS, Ramani S, Nielsen JF, Fessler JA. Monte Carlo SURE-based parameter selection for parallel magnetic resonance imaging reconstruction. Magnetic resonance in medicine 2014;71(5):1760–1770.

[9] Axler S. Linear Algebra Done Right (Undergraduate Texts in Mathematics). Springer; 2016.

[10] Candes EJ, Sing-Long CA, Trzasko JD. Unbiased risk estimates for singular value thresholding and spectral estimators. IEEE transactions on signal processing 2013;61(19):4643–4657.

[11] Uecker M, Ong F, Tamir JI, Bahri D, Virtue P, Cheng JY, et al. Berkeley advanced reconstruction toolbox. In: Proc. Intl. Soc. Mag. Reson. Med. vol. 23; 2015. p. 2486.
7 | APPENDIX

7.0.1 | Divergence of a Linear Operator (Used in Equations (8b) and (12))

Consider arbitrary $x \in \mathbb{C}^m$, $A \in \mathbb{C}^{m \times m}$ and let $f(x) = Ax$.

\[
    f(x) = \begin{bmatrix}
    A_{11} & \ldots & A_{1m} \\
    \vdots & \ddots & \vdots \\
    A_{m1} & \ldots & A_{mm}
    \end{bmatrix}
    \begin{bmatrix}
    x_1 \\
    \vdots \\
    x_m
    \end{bmatrix}
\]

(Succinctly,

\[
    f_i(x) = \sum_{j=1}^{m} A_{ij} x_j
\]

Thus,

\[
    [\text{div}_x f](x) = \sum_{i=1}^{m} \frac{\partial}{\partial x_i} f_i(x) \\
    = \sum_{i=1}^{m} [A_{ii}] (x) \\
    = \text{trace}(A)
\]

7.0.2 | Derivation of Equation (8)

Let $x \in \mathbb{C}^m$ be the true, underlying data; $n$ be zero-mean, additive, Gaussian, complex noise with the real and imaginary standard deviations of $\frac{\sigma}{\sqrt{2}}$ each; and $x_{\text{acq}} = x + n$ be acquired data. Let $P$ be the self-adjoint projection operator.

We can partition our complex vector space into real and imaginary parts to reduce this problem to the real case. Let $x_R$ denote the partitioned form of $x$, $P_R$ denote the partitioned form of $P$ and so on. Then,

\[
    x_R = [\text{Re}(x) \ldots \text{Im}(x)]^T, \quad x_{\text{acq}R} = [\text{Re}(x_{\text{acq}}) \ldots \text{Im}(x_{\text{acq}})]^T, \quad P_R = \begin{bmatrix}
    \text{Re}(P) & -\text{Im}(P) \\
    \text{Im}(P) & \text{Re}(P)
    \end{bmatrix}
\]

This partitioning allows us to use [4] directly. Observe that the noise variance in the partitioned case is now $\frac{\sigma^2}{2}$.

\[
    \text{SURE}_P(x_{\text{acq}}) = -2m \frac{\sigma^2}{2} + \|(P - I)x_{\text{acq}}\|_2^2 + 2 \frac{\sigma^2}{2} \text{trace}(P_R) \\
    = -m\sigma^2 + \|(P - I)x_{\text{acq}}\|_2^2 + 2\sigma^2\text{trace}(P)
\]

Using the fact that the eigenvalues of $P$ are real (since $P$ is self-adjoint) and that the trace of a matrix is independent from basis of representation [2], we can conclude that,

\[
    \text{trace}(P_R) = 2\text{trace}(P)
\]