Approximate Message Passing With a Colored Aliasing Model for Variable Density Fourier Sampled Images

CHARLES MILLARD [1,2] (Student Member, IEEE), **AARON T. HESS**[2], **BORIS MAILHÉ**[3] (Member, IEEE),
AND JARED TANNER [1] (Member, IEEE)

[1] Mathematical Institute, University of Oxford, Oxford OX2 6GG, U.K.
[2] Oxford Centre for Clinical Magnetic Resonance, University of Oxford, Oxford OX3 9DU, U.K.
[3] Siemens Healthineers, Princeton, NJ 08540 USA

CORRESPONDING AUTHOR: CHARLES MILLARD

ABSTRACT The Approximate Message Passing (AMP) algorithm efficiently reconstructs signals which have been sampled with large i.i.d. sub-Gaussian sensing matrices. Central to AMP is its "state evolution," which guarantees that the difference between the current estimate and ground truth (the "aliasing") at every iteration obeys a Gaussian distribution that can be fully characterized by a scalar. However, when Fourier coefficients of a signal with non-uniform spectral density are sampled, such as in Magnetic Resonance Imaging (MRI), the aliasing is intrinsically colored, AMP's scalar state evolution is no longer accurate and the algorithm encounters convergence problems. In response, we propose the Variable Density Approximate Message Passing (VDAMP) algorithm, which uses the wavelet domain to model the colored aliasing. We present empirical evidence that VDAMP obeys a "colored state evolution," where the aliasing obeys a Gaussian distribution that can be fully characterized with one scalar per wavelet subband. A benefit of state evolution is that Stein's Unbiased Risk Estimate (SURE) can be effectively implemented, yielding an algorithm with subband-dependent thresholding that has no free parameters. We empirically evaluate the effectiveness of VDAMP on three variations of Fast Iterative Shrinkage-Thresholding (FISTA) and find that it converges in around 10 times fewer iterations on average than the next-fastest method, and to a comparable mean-squared-error.

INDEX TERMS Approximate message passing, compressed sensing, magnetic resonance imaging (MRI), Stein's unbiased risk estimate, variable density sampling.

I. INTRODUCTION

We consider a complex data vector $\boldsymbol{y} \in \mathbb{C}^N$ formed of noisy Fourier coefficients of a deterministic signal of interest $\boldsymbol{x}_0 \in \mathbb{C}^N$:

$$\boldsymbol{y} = \boldsymbol{M}_\Omega (\boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{\varepsilon}), \tag{1}$$

where $\boldsymbol{F}$ is a multi-dimensional discrete Fourier transform and $\boldsymbol{M}_\Omega \in \mathbb{R}^{N \times N}$ is a diagonal undersampling mask with 1 on the jth diagonal entry if $j \in \Omega$ and 0 otherwise, where Ω is a sampling set with $|\Omega| = n$ for $n < N$. Here, $\boldsymbol{\varepsilon} \frown \mathcal{CN}(\boldsymbol{0}, \sigma_\varepsilon^2 \mathbb{1}_N)$ where $\mathbb{1}_N$ is the $N \times N$ identity matrix and

$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$ denotes the distribution with independent real and imaginary parts that are normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}^2/2$. A well-studied approach is to seek a solution of

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{C}^N}{\text{argmin}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x}) \tag{2}$$

where $f(\boldsymbol{x})$ is a model-based penalty function. Compressed sensing [2], [3] concerns the reconstruction of signals of interest from underdetermined measurements, where sparsity

in $\hat{x}$ is promoted by solving (2) with $f(x) = \lambda\|\Psi x\|_1$ for sparse weighting $\lambda > 0$ and sparsifying transform Ψ.

A prominent success of compressed sensing with Fourier measurements is accelerated Magnetic Resonance Imaging (MRI) [4]–[8]. Images of interest typically have a highly non-uniform spectral density that is concentrated at low frequencies. Accordingly, it is well-known that better image restoration is possible if the sampling set Ω is generated with variable density, so that there is a higher probability of sampling low frequencies [9]–[13]. This work considers an Ω with elements drawn independently from a Bernoulli distribution with generic non-uniform probability, so that $\text{Prob}(j \in \Omega) = p_j \in [0, 1]$.

A. APPROXIMATE MESSAGE PASSING

The Approximate Message Passing (AMP) algorithm [14] is an iterative method that, for certain sensing matrices $\Phi \in \mathbb{R}^{n \times N}$, efficiently estimates x_0 in problems of the form $y = \Phi x_0 + \varepsilon$. At iteration k, AMP implements a denoiser $g(r_k; \tau_k)$ on r_k with mean-squared error estimate τ_k, which can be, for instance, the proximal operator associated with penalty function $f(x)$:

$$g(r_k; \tau_k) = \operatorname*{argmin}_{x \in \mathbb{C}^N} \frac{1}{2\tau_k}\|r_k - x\|_2^2 + f(x), \tag{3}$$

which is equal to soft thresholding in the case of $f(x) = \lambda\|\Psi x\|_1$ and orthogonal Ψ. Under certain circumstances [14], AMP with proximal denoising shares a fixed point with optimization problems of the form of (2). Further, for certain sensing matrices and given mild conditions on $f(x)$, AMP obeys a *state evolution*, which guarantees that in the large system limit $n, N \to \infty$, $n/N \to \delta \in (0, 1)$, vector r_k is the original signal corrupted by zero-mean Gaussian noise with a covariance matrix that is proportional to the identity:

$$r_k = x_0 + \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbb{1}_N), \tag{4}$$

where σ_k is a scalar iteration-dependent standard deviation. In this work, the term *aliasing* is used to refer to the difference between a given estimate and the ground truth. For instance, the aliasing of r_k is $r_k - x_0$. Also, when the covariance matrix is proportional to the identity, as in (4), the aliasing and state evolution are referred to as *white*.

AMP was originally constructed for real, zero-mean, i.i.d. Gaussian measurements, and its white state evolution was proven for this case in [15] and subsequently proven for i.i.d. sub-Gaussian measurements in [16]. It has also been shown empirically that it holds for uniformly undersampled Fourier measurements of an artificial i.i.d. signal [14]. When state evolution holds, AMP is known to exhibit very fast convergence. However, for generic Φ, the behavior of AMP is not well understood and it has been noted by a number of authors [17]–[20] that it can encounter convergence problems. The recent Orthogonal AMP (OAMP) [21] and related Vector Approximate Message Passing (VAMP) [22] algorithm obey a white state evolution for a broader class of measurement matrices Φ, and were found to perform very well on certain

---

Algorithm 1: OAMP [21].

Require: Matrix Φ, measurements y, denoiser $g(r_k; \tau_k)$, number of iterations K_{it}.
1: Set $\tilde{r}_0 = \mathbf{0}$
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3:      $z_k = y - \Phi\tilde{r}_k$
4:      $r_k = \tilde{r}_k + \Phi^H z_k$
5:      Update τ_k
6:      $\hat{x}_k = g(r_k; \tau_k)$
7:      $\alpha_k = \langle g'(r_k; \tau_k)\rangle$
8:      Update c_k, e.g. $c_k = 1, 2$ or 3
9:      $\tilde{r}_{k+1} = c_k \cdot (\hat{x}_k - \alpha_k r_k)$
10: **end for**
11: **return** $\hat{x}_k$

---

reconstruction tasks. For VAMP, white state evolution was proven for sensing matrices that are "right-orthogonally invariant": see [22] for details.

The matched filter variation of the OAMP algorithm [21], which forms the basis of the algorithm presented in this work, is stated in Algorithm 1. Here, $\langle \cdot \rangle$ is the empirical averaging operator and $g'(r_k; \tau_k)$ is the diagonal of the Jacobian of $g(r_k; \tau_k)$ with respect to r_k. The scalar τ_k estimates the variance σ_k^2 from (4): see Eqn. (31) of [21] for details of the update formula. The relationship between OAMP and the well-known Iterative Shrinkage-Thresholding Algorithm (ISTA) [23] can be seen by considering lines 6–9 as a single function:

$$\tilde{g}(r_k; \tau_k) = c_k \cdot (g(r_k; \tau_k) - \alpha_k r_k). \tag{5}$$

Then OAMP is equivalent to ISTA with $\tilde{g}(r_k; \tau_k)$ in place of the usual shrinkage step. The $\alpha_k r_k$ subtraction, known as the *Onsager correction* [14], causes the function $\tilde{g}(r_k; \tau_k)$ for large N to be approximately *divergence-free* [21], defined as

$$\langle \tilde{g}'(r_k; \tau_k)\rangle \approx 0. \tag{6}$$

The divergence-free property of $\tilde{g}(r_k; \tau_k)$ is the vital aspect of OAMP that leads to the white state evolution of (4) [21]. Although any divergence-free function can be employed in place of lines 6–9 of Algorithm 1, and is not required to take the form of (5), this work focuses on a $\tilde{g}(r_k; \tau_k)$ of the form of (5) with $g(r_k; \tau_k)$ as the soft thresholding operator.

Any choice of c_k update in line 8 of Algorithm 1 is consistent with the divergence-free requirement of (6). For soft thresholding, OAMP [21] gives no explicit suggestions for c_k in practice. Instead, OAMP demonstrates the generality of its state evolution using three arbitrary chosen values $c_k = 1, 2, 3$. For the algorithm presented in this work, two c_k updates are suggested, stated in (24) and (26), which are motivated by VAMP [22] and computed by Stein's Unbiased Risk Estimate (SURE) [24], [25] respectively.

B. COLORED ALIASING

AMP, OAMP and VAMP assume that the sensing matrix is sufficiently random to ensure that the aliasing is white.
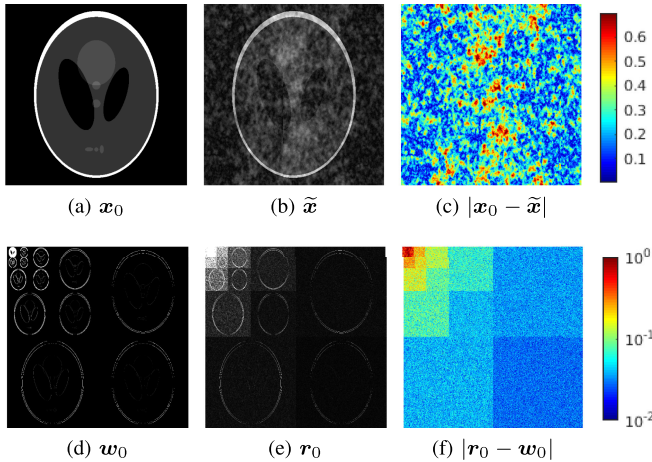
FIGURE 1. The ground truth, unbiased estimate and entry-wise absolute error for a uniformly sampled Shepp-Logan with $p_j = 1/2$, where the colorbars are given as a proportion of the maximum of x_0. The top row shows the image domain and the bottom shows the wavelet domain. The colored aliasing evident in Fig. 1(c) illustrates the infeasibility of white state evolution for Fourier sampling of signals with non-uniform spectral density. The anisotropy of the spectral density of x_0 causes the horizontal, vertical and diagonal variances to differ, even at the same scale.

However, when an image is sampled in the Fourier domain, the aliasing is innately colored. To see this, consider the natural initialization for an approximate message passing algorithm: the unbiased estimator $\widetilde{x} = F^H P^{-1} y$, where P is the diagonal matrix formed from sampling probabilities p_j. Denoting $y_0 = Fx_0$, the power spectrum of the aliasing of $\widetilde{x}$ is shown in Appendix A to be

$$\mathbb{E}_{\Omega,\varepsilon}\{|y_0 - P^{-1}y|^2\} = (P^{-1} - \mathbb{1}_N)|y_0|^2 + \sigma_\varepsilon^2 P^{-1}\mathbb{1}_N \quad (7)$$

where $|\cdot|$ is the entry-wise absolute value and $\mathbb{1}_N$ is the N-dimensional vector of ones. Equation (7) depends on P and $|y_0|^2$, which are non-uniform and anisotropic in general. Note that although the specific case $p_j = (\sigma_\varepsilon^2 + |y_{0,j}|^2)/(\alpha + |y_{0,j}|^2)$ for constant α does lead to white aliasing, it requires knowledge of the ground truth spectral density $|y_0|^2$ so is not a feasible sampling scheme in practice.

A visual example of colored aliasing is shown in the top row of Fig. 1. Here, the unbiased estimate of a 512×512 synthetic Shepp-Logan uniformly sampled with $p_j = 1/2$ for all j is shown, with $\sigma_\varepsilon = 0$. By (7), the power spectrum of the aliasing of $\tilde{x}$ in this case is $\mathbb{E}_{\Omega,\varepsilon}\{|y_0 - P^{-1}y|^2\} = |y_0|^2$. Since $|y_0|^2$ is tightly concentrated at low frequencies, the aliasing has strong low-frequency components, which is manifest in Fig. 1(c) as local correlations. In this example, uniform sampling was chosen to exaggerate the colored aliasing property. In practice, when variable density sampling is used, the sampling distribution partially compensates for the power spectrum of the signal, so the aliasing is still colored, but less strongly.

The intrinsically colored aliasing of variable density sampling from a non-uniform spectral density implies that the white state evolution of AMP, OAMP, and VAMP, (4), cannot be relied upon. The primary development of this work is based

on the use of the Discrete Wavelet Transform (DWT) to compute a multiresolution decomposition of the power spectrum of the aliasing, as used for colored noise analysis in [26]–[28]. In the wavelet domain, colored aliasing has a structure that resembles a state evolution. To illustrate this, consider again the unbiased initialization $\widetilde{x}$. The corresponding estimator in the wavelet domain is

$$r_0 = \Psi\widetilde{x}. \quad (8)$$

The bottom row of Fig. 1 shows $w_0 = \Psi x_0$, r_0 and the entry-wise absolute difference $|w_0 - r_0|$ for the same image and sampling set as the top row, where Ψ is a Haar DWT with 4 decomposition scales. Qualitatively, Fig. 1(f) suggests that the power spectrum given by (7) has a simple structure in the wavelet domain; in particular, it suggests that the per-subband aliasing within each subband is approximately uniform. In Section III we show that, in fact, *the aliasing within each subband is quantitatively consistent with a white Gaussian distribution*, see Fig. 6. Further, Table 3 presents evidence that the Gaussianity holds for a variety of image types and undersampling factors.

C. COLORED STATE EVOLUTION

Herein we present a new method for undersampled signal reconstruction that we term the Variable Density Approximate Message Passing (VDAMP) algorithm, see Algorithm 2. We present empirical evidence that VDAMP preserves the subband-dependent noise structure illustrated in Fig. 1(f) for all iterations. Explicitly, the r_k of VDAMP behaves as

$$r_k = w_0 + \mathcal{CN}(0, \Sigma_k^2), \quad (9)$$

where Ψ is an orthogonal DWT and the covariance matrix Σ_k^2 is diagonal so that for a Ψ with s decomposition scales,

$$\Sigma_k^2 = \begin{bmatrix} \sigma_{k,1}^2\mathbb{1}_{N_1} & 0 & \dots & 0 \\ 0 & \sigma_{k,2}^2\mathbb{1}_{N_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{k,1+3s}^2\mathbb{1}_{N_{1+3s}} \end{bmatrix}, \quad (10)$$

where $\sigma_{k,b}^2$ and N_b refer to the variance and dimension of the bth subband respectively. We refer to (9) and (10) as the *colored state evolution* of VDAMP and the aliasing of r_k as the *effective noise* of VDAMP.

The joint space-frequency localization provided by the wavelet transform decomposes the color of the effective noise while retaining incoherence. The algorithm presented in this work considers a sparse model on wavelet coefficients, however, we emphasize that the wavelet transform is primarily used as a tool for decomposing the aliasing, and is not necessarily constrained to model wavelet coefficients directly [25], [29].

VAMP for Image Recovery (VAMPire) [30] is an adaption of VAMP for variable density Fourier sampled images that uses wavelets to decompose the effective noise in to frequency

Algorithm 2: VDAMP.

Require: Sampling set Ω, wavelet transform $\boldsymbol{\Psi}$, probability matrix $\boldsymbol{P}$, measurements $\boldsymbol{y}$, denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$, number of iterations K_{it}.

1:   Set $\widetilde{\boldsymbol{r}}_0 = \boldsymbol{0}$ and compute $\boldsymbol{S} = |\boldsymbol{F}\boldsymbol{\Psi}^H|^2$
2:   **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3:       $\boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H \widetilde{\boldsymbol{r}}_k$
4:       $\boldsymbol{r}_k = \widetilde{\boldsymbol{r}}_k + \boldsymbol{\Psi}\boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{z}_k$
5:       $\boldsymbol{\tau}_k = \boldsymbol{S}^H \boldsymbol{M}_\Omega \boldsymbol{P}^{-1}[(\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{z}_k|^2 + \sigma_\varepsilon^2 \mathbb{1}_N]$
6:       $\hat{\boldsymbol{w}}_k = \boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$
7:       $\boldsymbol{\alpha}_k = \langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k))\rangle_{\text{sband}}$
8:       Update $\boldsymbol{c}_k$
9:       $\widetilde{\boldsymbol{r}}_{k+1} = \boldsymbol{c}_k \odot (\hat{\boldsymbol{w}}_k - \boldsymbol{\alpha}_k \odot \boldsymbol{r}_k)$
10: **end for**
11: **return** $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k + \boldsymbol{F}^H(\boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k)$

bands that are subsequently "whitened" by a hand-tuned prediction of the per-subband energy. As in OAMP, Algorithm 1, the effective noise model of VAMPire is represented by a scalar τ_k. In contrast, we propose making the necessary algorithmic adaptions to *allow the aliasing to be colored*, and to model the color with a *vector* $\boldsymbol{\tau}_k$. To our knowledge, VDAMP is the first algorithm for variable density Fourier sampling of images where a state evolution has been observed.

II. DESCRIPTION OF ALGORITHM
The VDAMP algorithm, Algorithm 2, adapts OAMP, Algorithm 1, to a colored aliasing model. VDAMP's colored aliasing model is updated in line 5, where the scalar $\tau_k \in \mathbb{R}$ of OAMP is replaced by a vector $\boldsymbol{\tau}_k \in \mathbb{R}^N$ that models the diagonal of $\boldsymbol{\Sigma}_k^2$. In line 6, OAMP's denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ is replaced by the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$, which takes the vector $\boldsymbol{\tau}_k$ as its input. Lines 7–9 of Algorithm 2 is the Onsager correction from lines 7–9 of Algorithm 1 adapted to a subband-wise aliasing model. Here, the notation $\boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k))$ in line 7 replaces OAMP's $\boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)$, defined as the function with jth entry

$$\partial_j(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) = \frac{1}{2}\left(\frac{\partial \Re[g_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Re[r_j]} + \frac{\partial \Im[g_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Im[r_j]}\right) \tag{11}$$

where $\Re[\cdot]$ and $\Im[\cdot]$ are the real and imaginary parts respectively. The form of (11) is justified in section II-D. Also in line 7, the notation $\langle\cdot\rangle_{\text{sband}}$ is an operator that empirically averages entries within subbands, so that $\boldsymbol{\alpha}_k$ has the structure

$$\boldsymbol{\alpha}_k = \begin{bmatrix} \alpha_{k,1}\mathbf{1}_{N_1} \\ \alpha_{k,2}\mathbf{1}_{N_2} \\ \vdots \\ \alpha_{k,1+3s}\mathbf{1}_{N_{1+3s}} \end{bmatrix} \tag{12}$$

with

$$\alpha_{k,b} = \frac{1}{N_b}\sum_{j\in J_b} \partial_j(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)), \tag{13}$$

where J_b is the set of indices associated with subband b and $N_b = |J_b|$. In line 8 of Algorithm 2, $\boldsymbol{c}_k$ is a vector with the

piecewise-constant structure of (12), and in line 9 the notation $\odot$ refers to entry-wise multiplication. The remainder of this section works through Algorithm 2 in detail, line-by-line.

A. DENSITY COMPENSATED GRADIENT DESCENT, LINES 3–4
To ensure that $\boldsymbol{r}_k$ is an unbiased estimate of $\boldsymbol{x}_0$, the sensing matrix must be correctly normalized. In VDAMP this is manifest in the gradient step of lines 3–4, which features a crucial weighting by $\boldsymbol{P}^{-1}$ that is absent in previous applications of AMP to variable density sampling [30]–[32], where a state evolution was not observed. This provides the correct normalization in expectation over Ω: $\mathbb{E}_\Omega\{\boldsymbol{\Psi}\boldsymbol{F}^H \boldsymbol{P}^{-1}\boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H\} = \mathbb{1}_N$. Note that VDAMP's $\boldsymbol{r}_0$ is the unbiased estimator from (8). Such a rescaling is referred to as *density compensation* in the MRI literature [33], [34], and was used in the original compressed sensing MRI paper with zero-filling to generate a unregularized, non-iterative baseline [4]. However, to our knowledge, VDAMP is the first *iterative* method that employs density compensated gradient descent. Density compensation also arises in recovery guarantees for variable density Fourier measurements in [9], [11], although it was considered an artifact of the proof and was not used in the numerical evaluations of these works. The connection of VDAMP to these theoretical results is beyond the scope of this paper and is left as future work.

Density compensation increases the variance of the measurement noise at frequencies sampled with low probability, and its inclusion in the gradient step will lead to a $\boldsymbol{r}_k$ with higher mean-squared error than an unweighted gradient step. However, as shown in Section III, a careful choice of denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ that leverages VDAMP's state evolution can cause lines 6–9 to be very effective, leading to faster overall convergence than competing methods.

The final step of VDAMP, line 11, is a gradient step without a $\boldsymbol{P}^{-1}$ weighting, which generates a biased image estimate $\hat{\boldsymbol{x}}$ with high data fidelity.

B. COLORED EFFECTIVE NOISE MODEL, LINE 5
Line 5 of Algorithm 2 computes an estimate of the colored effective noise covariance matrix $\boldsymbol{\Sigma}_k^2$ from (9). Through a similar derivation to that for (7), shown in Appendix A, the power spectrum of the aliasing of $\boldsymbol{r}_k$ is

$$\mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{r}_k - \boldsymbol{y}_0|^2\} = (\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{F}\boldsymbol{\Psi}^H\widetilde{\boldsymbol{r}}_k - \boldsymbol{y}_0|^2 + \sigma_\varepsilon^2 \boldsymbol{P}^{-1}\mathbf{1}_N. \tag{14}$$

Eqn. (14) depends on the ground truth $\boldsymbol{y}_0$, so is of limited practical use. An estimate of (14) that does not require knowledge of $\boldsymbol{y}_0$ is

$$\boldsymbol{\tau}_k^y = \boldsymbol{M}_\Omega \boldsymbol{P}^{-1}[(\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{z}_k|^2 + \sigma_\varepsilon^2 \mathbf{1}_N]. \tag{15}$$

Estimating properties of a distribution using samples from another is known as *importance sampling* in the statistics literature [35], [36]. Eqn. (15) uses importance sampling with $\boldsymbol{P}$ as the importance distribution which, as proven in Appendix

B, is an unbiased estimator of (14). We assume the estimator τ_k^y concentrates around its expectation, and leave the study of how this depends on the importance distribution $\boldsymbol{P}$ for future works.

The computation of τ_k in line 5 of Algorithm 2 is a linear transform of τ_k^y to the wavelet domain: $\tau_k = |\boldsymbol{\Psi}\boldsymbol{F}^H|^2 \tau_k^y$, which, as shown in Appendix C, is a unbiased estimate of $|\boldsymbol{r}_k - \boldsymbol{w}_0|^2$ when $\boldsymbol{r}_k$ has unbiased independent entries, as expected by state evolution. $|\boldsymbol{\Psi}\boldsymbol{F}^H|^2$ is the power spectrum of $\boldsymbol{\Psi}$, so has $1 + 3s$ unique rows; line 5 therefore requires $1 + 3s$ inner products. For fixed s the complexity of VDAMP is therefore governed by $\boldsymbol{\Psi}$ and $\boldsymbol{F}$, whose fast implementations have complexity $O(N)$ and $O(N \log N)$ respectively.

C. COMPLEX SOFT THRESHOLD TUNING WITH SURE, LINE 6

Selecting appropriate regularisation parameters such as λ is a notable challenge in real-world compressed sensing applications. We present an approach to parameter-free compressed sensing reconstruction that leverages VDAMP's state evolution by applying Stein's Unbiased Risk Estimate (SURE) [24], building on work on AMP in [19], [37], [38].

A strength of automatic parameter tuning via SURE is that it is possible to have a richer regularizer than would usually be feasible for a hand-tuned $f(\boldsymbol{x})$. In this work, the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ was the complex soft thresholding operator with a subband-dependent threshold that is tuned automatically with SURE. In other words, SURE with an effective noise model given by $\boldsymbol{r}_k = \boldsymbol{w}_0 + \mathcal{CN}(\boldsymbol{0}, \text{Diag}(\boldsymbol{\tau}_k))$ was used to approximately solve

$$\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k) \approx \underset{\boldsymbol{w} \in \mathbb{C}^N}{\text{argmin}} \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \frac{1}{2} \|(\boldsymbol{w} - \boldsymbol{r}_k) \oslash \sqrt{\boldsymbol{\tau}_k}\|_2^2 + \|\boldsymbol{\lambda} \odot \boldsymbol{w}\|_1,$$
(16)

where $\oslash$ denotes entry-wise division, $\sqrt{\boldsymbol{\tau}_k}$ is the entry-wise square root of $\boldsymbol{\tau}_k$ and $\boldsymbol{\lambda}$ has the piecewise-constant structure of (12). The possibility of using a scale-dependent thresholds is well known, such as in [23]. We emphasize that (16) is *subband*-dependent rather than scale-dependent, enabling higher order, anisotropic signal modeling [39], [40].

Equation (16) was solved using a procedure related to SureShrink [41] but for Gaussian noise that is complex and colored [28]. Consider a vector $\boldsymbol{v}_0 \in \mathbb{C}^{N_v}$ corrupted by white complex Gaussian noise: $\boldsymbol{v} = \boldsymbol{v}_0 + \mathcal{CN}(\boldsymbol{0}, \tau_v \mathbb{1}_{N_v})$. Let $\boldsymbol{d}(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{h}(\boldsymbol{v})$ be an estimator of $\boldsymbol{v}_0$. SURE [24] adapted to complex variables is

$$\text{cSURE}(\boldsymbol{d}(\boldsymbol{v})) = \|\boldsymbol{h}(\boldsymbol{v})\|_2^2 + N_v \tau_v [2 \langle \boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v})) \rangle - 1].$$
(17)

cSURE is of interest to denoising problems because, as shown in Appendix D, it is an unbiased estimate of the risk $\mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v}) - \boldsymbol{v}_0\|_2^2\}$. The optimal parameters of the denoiser $\boldsymbol{d}(\boldsymbol{v})$ can therefore be estimated by minimizing cSURE as a proxy for the true risk. Consider the case where $\boldsymbol{d}(\boldsymbol{v})$ is the complex soft thresholding operator $\boldsymbol{\eta}(\boldsymbol{v}; t)$ with threshold t,

which acts component-wise as

$$\eta_j(v_j; t) := v_j \left(1 - \min\left\{\frac{t}{|v_j|}, 1\right\}\right).$$
(18)

The jth entry of $\boldsymbol{\partial}(\boldsymbol{\eta}(\boldsymbol{v}; t))$ is

$$\partial_j(\eta_j(v_j; t)) = \begin{cases} 0, & \text{if } |v_j| \leq t \\ 1 - \frac{t}{2|v_j|}, & \text{otherwise.} \end{cases}$$
(19)

By (17), an unbiased estimate of the risk of soft thresholding is therefore [41]

$$\text{cSURE}(\boldsymbol{\eta}(\boldsymbol{v}; t)) = (t^2 + 2\tau_v) \cdot \#\{j : |v_j| > t\} - N_v \tau_v$$

$$+ \sum_{j : |v_j| \leq t}^{N_v} |v_j|^2 - \sum_{j : |v_j| > t}^{N_v} t\tau_v / |v_j|. \quad (20)$$

The optimal threshold for each subband can be estimated with

$$\hat{t} = \underset{t}{\text{argmin}}(\text{cSURE}(\boldsymbol{\eta}(\boldsymbol{v}; t)))$$
(21)

by evaluating (20) for trial thresholds $t = |v_1|, |v_2|, \ldots, |v_{N_v}|$. For large dimension N_v one would expect by the law of large numbers that cSURE is close to the true risk, and for the threshold to be almost optimal. Since a larger number of decomposition scales s give subbands with lower dimension, there is a trade-off between the size of s and the quality of threshold selection with cSURE.

SURE has previously been employed for parameter-free compressed sensing MRI in [42], where FISTA was used with (16) in place of the usual shrinkage step. This algorithm is herein referred to as SURE-IT, and is discussed in detail in Section III-A.

D. COMPLEX, COLORED ONSAGER CORRECTION, LINES 7–9

Lines 7–9 of Algorithm 2 can be understood intuitively as follows: since lines 7–9 of Algorithm 1 apply to white noise, and the effective noise of VDAMP is white within each subband, the Onsager correction must be applied subband-by-subband. For soft thresholding, the use of $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) \rangle_{\text{sband}}$ in place of OAMP's $\langle \boldsymbol{g}'(\boldsymbol{r}_k; \boldsymbol{\tau}_k) \rangle$ leads the function formed by merging lines 7–9,

$$\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k) = \boldsymbol{c}_k \odot (\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k) - \boldsymbol{\alpha}_k \odot \boldsymbol{r}_k),$$
(22)

to obey, for all subbands b,

$$\frac{1}{N_b} \sum_{j \in J_b} \frac{\partial \Re[\tilde{g}_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Re[r_j]} \approx \frac{1}{N_b} \sum_{j \in J_b} \frac{\partial \Im[\tilde{g}_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Im[r_j]} \approx 0, \quad (23)$$

which is a colored aliasing version of OAMP's divergence-free condition, (6), applied to both real and imaginary parts. The Onsager correction employed here is not the only choice that leads to a $\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ that satisfies (23) [43], however, we have found that this particular choice performs well.

Like the scalar c_k in Algorithm 1, the $\boldsymbol{c}_k$ updated in line 8 of Algorithm 2 is not constrained by (23), except to have the

piecewise-constant structure of (12). In the experiments in this work, two possibilities for the c_k update are considered. First,

$$c_k^\alpha = \mathbf{1}_N \oslash (\mathbf{1}_N - \boldsymbol{\alpha}_k). \qquad (24)$$

In this case, lines 6–9 of Algorithm 2 are a colored version of the 'denoising' phase of VAMP when written in LMMSE form (see [22], Algorithm 3). VDAMP with c_k updated with (24) is herein referred to as VDAMP-α.

Secondly, as in [25], we suggest using cSURE for a second time to estimate the c_k that minimizes the mean squared error of $\widetilde{r}_k$:

$$c_k^{SURE} \approx \arg\min_c \|\widetilde{g}(r_k, \tau_k) - w_0\|_2^2. \qquad (25)$$

Since by (23) $\langle \partial(\widetilde{g}(r_k; \tau_k)) \rangle_{\text{sband}} \approx 0$, optimizing (17) reduces to a series of ℓ_2 minimization problems with a closed-form solution, so that for the bth subband

$$c_{k,b}^{SURE} = \arg\min_c \|c(g_b(r_k; \tau_k) - \alpha_{k,b} r_{k,b}) - r_{k,b}\|_2^2 \quad (26a)$$

$$= \frac{r_{k,b}^H (g_b(r_k; \tau_k) - \alpha_{k,b} r_{k,b})}{\|g_b(r_k; \tau_k) - \alpha_{k,b} r_{k,b}\|_2^2}. \qquad (26b)$$

Vector c_k^{SURE} is formed from the $c_{k,b}^{SURE}$ so that it has the structure of (12). VDAMP with c_k updated with c_k^{SURE} is herein referred to as VDAMP-S.

As for OAMP [21], we do not claim that the either of the two c_k updates suggested here are necessarily optimal. Instead, we show in Section III that both updates lead to aliasing consistent with state evolution, and empirically evaluate their performance, which we observe to converge to a similar or lower NMSE as other algorithms, but with approximately a tenth the time to convergence.

III. NUMERICAL EXPERIMENTS

This section illustrates the performance of VDAMP-α and VDAMP-S compared with the Fast Iterative Shrinking-Thresholding Algorithm (FISTA) algorithm [23], [44], and two FISTA-based methods wth subband-dependent thresholding [40], [42]. We also present empirical evidence for VDAMP's state evolution. All experiments were conducted in MATLAB 9.4 and can be reproduced with code online, available at https://github.com/charlesmillard/VDAMP.

A. DESCRIPTION OF COMPARATIVE FISTA-BASED ALGORITHMS

The three versions of FISTA [40], [42], [44] considered in this work are summarized in Algorithm 3. In line 7, $g(r_k; \tau_k \mathbf{1})$ refers to the subband-dependent thresholding function with thresholds tuned by cSURE, stated in (16), with a scalar aliasing model τ_k. In line 9, $\eta(r_k; \tau_k \lambda \Lambda^{-1} \mathbf{1})$ is the soft thresholding function from (18) with entry-wise threshold $\tau_k \lambda \Lambda^{-1} \mathbf{1}$. As in (3), the threshold employed here is weighted by the variance estimate τ_k, causing the threshold to decrease over iterations, which, as in [45], was found to significantly reduce the time to convergence. Line 14 is an unweighted gradient descent step, as in line 11 of Algorithm 2, which outputs an estimate $\hat{x}$ with

---

Algorithm 3: FISTA [44], S-FISTA [40] and SURE-IT [42] for Fourier Sampled Images.

Require: Sampling set Ω, wavelet transform Ψ, per-subband weighting W set to $W = \mathbb{1}_N$ for FISTA and SURE-IT or calculated with (27) for S-FISTA, sparse weighting λ for FISTA and S-FISTA, measurements y, number of iterations K_{it}.

1: Set $\widetilde{r}_0 = \mathbf{0}$, $\hat{w}_{-1} = \mathbf{0}$ and $h_{-1} = 1$
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3:      $z_k = y - F\Psi^H \widetilde{r}_k$
4:      $r_k = \widetilde{r}_k + W^{-1} \Psi F^H M_\Omega z_k$
5:      Update τ_k
6:      **if** SURE-IT **then**
7:          $\hat{w}_k = g(r_k; \tau_k \mathbf{1})$
8:      **else**
9:          $\hat{w}_k = \eta(r_k; \tau_k \lambda W^{-1} \mathbf{1})$
10:     **end if**
11:     $h_k = (1 + \sqrt{1 + 4h_{k-1}^2})/2$
12:     $\widetilde{r}_{k+1} = \hat{w}_k + (h_{k-1} - 1)(\hat{w}_k - \hat{w}_{k-1})/h_k$
13: **end for**
14: **return** $\hat{x} = \Psi^H \hat{w}_k + F^H(y - M_\Omega F\Psi^H \hat{w}_k)$

---

high data fidelity. This was suggested in [45], and we found that this output had a lower reconstruction error than $\Psi^H \hat{w}_k$. The variations described in Algorithm 3, discussed in detail below, are referred to in this work as *FISTA*, *S-FISTA* [40] and *SURE-IT* [42].

FISTA refers to Algorithm 3 with $W = \mathbb{1}_N$, so that a global threshold $\lambda \tau_k$ is applied in line 9. Despite not discriminating between subbands, this version of FISTA was considered because it is widely-used in MRI, and we found that it performed well compared with the subband-dependent algorithms, so is an instructive baseline. FISTA with a hand-tuned subband-dependent λ was not considered as it is not feasible in practice.

S-FISTA refers to Algorithm 3 with a diagonal weight matrix W that has one unique entry per wavelet subband, so that $\text{diag}(W)$ has the piecewise-constant structure of (12). In [40], a method was suggested for selecting the weight for the bth subband, which we denote as w_b. Let Φ_b be the block of $M_\Omega F\Psi^H$ corresponding to the bth subband. In [40], it is shown that a choice of w_b that satisfies

$$\frac{1}{w_b} > \sum_{b'=1}^{3s+1} \sqrt{\lambda_{\max}(\Phi_{b'}^H \Phi_b \Phi_b^H \Phi_{b'})}, \qquad (27)$$

where $\lambda_{\max}(\cdot)$ is the largest eigenvalue, ensures that $W - \Phi^H \Phi$ is a positive operator, and therefore guarantees that the algorithm converges: see [40] for details. Following [46], we computed the w_b by calculating $\lambda_{\max}$ once per Ω for all b and b' using the power iteration method. In contrast with VDAMP, S-FISTA's per-subband weighting is fixed across all iterations, and depends only on the sensing matrix and wavelet family, and not on the per-iteration signal estimate.

TABLE 1. Convergence Time in Seconds. The Shortest Time is Highlighted in Bold

Image	SHEPP-LOGAN			BRAIN			CARDIAC			BARBARA		
n/N	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	57.99	69.55	76.64	1.40	1.53	1.68	3.37	3.45	3.20	1.01	1.65	3.18
S-FISTA	79.91	80.62	81.13	4.22	4.93	4.79	10.41	8.53	9.66	6.00	6.78	10.10
SURE-IT	8.60	4.45	5.25	0.91	0.90	1.08	1.84	20.77	1.67	1.05	1.15	4.50
VDAMP-α	2.99	4.41	**0.96**	**0.16**	**0.13**	0.14	**0.39**	0.38	**0.24**	**0.18**	**0.19**	0.43
VDAMP-S	**2.21**	**2.37**	4.70	0.28	**0.13**	**0.11**	0.43	**0.26**	1.28	0.56	0.33	**0.31**

Image	BOAT			CAMERAMAN			HOUSE			PEPPERS		
n/N	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	2.80	4.06	3.08	1.56	1.58	1.22	1.52	1.97	1.94	1.43	1.66	1.94
S-FISTA	10.44	11.71	8.13	5.15	4.94	4.20	4.44	5.08	4.44	5.24	5.29	5.57
SURE-IT	2.31	2.73	2.36	1.15	0.94	0.81	2.27	2.09	3.03	0.73	0.19	0.15
VDAMP-α	**0.44**	**0.42**	**0.31**	**0.19**	**0.13**	0.73	0.17	0.17	0.14	0.14	**0.09**	**0.09**
VDAMP-S	0.70	3.62	2.23	0.99	0.68	**0.06**	**0.14**	**0.09**	**0.09**	**0.10**	0.10	**0.09**

(a) 256×256 Brain (b) 208×416 Cardiac
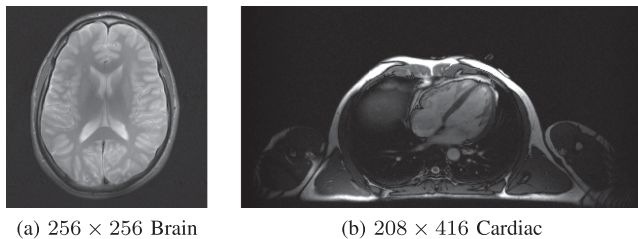
FIGURE 2. MRI images used to evaluate the performance of VDAMP.

SURE-IT refers to Algorithm 3 with $\boldsymbol{W} = \mathbb{1}_N$ and a subband-dependent, automatically tuned denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k \boldsymbol{1})$, as in [42]. Although SURE-IT's thresholding is iteration-dependent, the aliasing of its $\boldsymbol{r}_k$ is highly non-Gaussian, so deviates from a proper theoretical basis for using SURE and does not give near-optimal threshold selection. Further, unlike VDAMP, the aliasing of SURE-IT is modeled by a scalar τ_k.

In the following experiments, the scalar sparse weighting λ of FISTA and S-FISTA was tuned with an exhaustive search so that the mean-squared error was minimized at $k = 100$. Since the threshold was weighted by τ_k, we tuned λ separately for FISTA and S-FISTA. The variance estimate τ_k on line 5 of Algorithm 3 was updated using the ground truth: $\tau_k = \|\boldsymbol{r}_k - \boldsymbol{w}_0\|_2^2/N$.

B. EXPERIMENTAL METHOD

We considered the reconstruction of 8 test images: the Shepp-Logan shown in Fig. 1(a), a brain and a cardiac MRI image, shown in Fig. 2, and 5 standard test images: Barbara, Boat, Cameraman, House and Peppers, downloaded from [47]. In all cases the undersampled data was artificially corrupted with complex Gaussian noise so that $\|\boldsymbol{x}_0\|_2^2/N\sigma_\varepsilon^2 = 40$ dB. We assumed that σ_ε^2 was known a priori. For simplicity $\boldsymbol{\Psi}$ was chosen as the Haar wavelet with s = 4 scales.

A variety of variable density sampling schemes for MRI have been suggested [4], [10], [12], [45], [48], [49], including some with recovery guarantees [11], [13]. In the experiments presented in this section, we generated Ω using the variable

density sampling function from the Sparse MRI package.[1] The focus of this work is on the reconstruction algorithm, not on the sampling scheme, and we do not claim that this scheme is necessarily the best choice, only that it an instructive example because it common in MRI and known to perform well in practice [4]. The Ω generated with this package for the $N/n = 4$ cardiac image is shown in Fig. 4 f. We chose variable density schemes $\boldsymbol{P}$ with $\mathbb{E}_\Omega\{|\Omega|\} = n$ so that the acceleration factors $N/n = 4, 6, 8$ were used, except for the Shepp-Logan, where we used an increased acceleration of $N/n = 8, 10, 12$. All algorithms were initialized with a vector of zeros and run for $K_{it} = 500$ iterations, except for the Shepp-Logan, which was found to require more iterations, so was run until $K_{it} = 1000$.

C. TIME TO CONVERGE

Table 1 shows the time to converge for each algorithm, defined as the time taken until the NMSE is within 0.1 dB of its value at K_{it}. In all test cases, both variations of VDAMP converge considerably more rapidly than the competing FISTA-based methods. Across all experiments, convergence time compared with FISTA was 14.0 times shorter for VDAMP-α and 11.8 times shorter for VDAMP-S on average, corresponding to a 16.5 and 15.2 times reduction in the required number of iterations respectively. Of the three FISTA-based algorithms, SURE-IT had the shortest time to convergence, but still required 10.7 and 10.0 times more iterations on average than VDAMP-α and VDAMP-S respectively. Note that the times listed in Table 1 do not include the time required to tune λ for FISTA and S-FISTA, nor the time to calculate S-FISTA's subband-weighting w_b.

The NMSE vs iteration for the Shepp-Logan undersampled at $N/n = 10$ and the cardiac image undersampled at $N/n = 4$ are shown up to K_{it} and $k = 200$ respectively in Fig. 3, which visualizes VDAMP's comparative rapidity of convergence. In Fig. 4, the cardiac image is shown at $k = 10$, where VDAMP-α and VDAMP-S had converged, demonstrating a visible reduction in blocking artifacts for VDAMP.

---

[1][Online]. Available: https://people.eecs.berkeley.edu/~mlustig/Software.html

TABLE 2. Reconstruction NMSE in dB at K_{it} for All 8 Test Images at Different Undersampling Factors. The Lowest NMSE is Highlighted in Bold

Image	SHEPP-LOGAN			BRAIN			CARDIAC			BARBARA		
n/N	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	-36.2	-31.9	-25.7	-20.4	-18.2	-17.1	-17.9	**-14.5**	**-13.0**	**-17.5**	**-16.3**	-15.5
S-FISTA	-33.2	-28.1	-23.5	-20.3	-18.1	-17.1	-17.7	-14.3	-12.7	**-17.5**	**-16.3**	-15.5
SURE-IT	-18.7	-13.3	-13.1	-18.9	-17.3	-16.6	-14.9	-12.9	-10.5	-17.0	-16.1	-15.5
VDAMP-α	-35.1	-29.4	-20.5	**-20.7**	**-18.5**	**-17.5**	-17.9	**-14.5**	-12.9	**-17.5**	**-16.3**	**-15.6**
VDAMP-S	**-38.1**	**-34.9**	**-30.8**	-20.6	-18.4	-17.4	**-18.0**	-14.3	-12.4	-16.8	-15.4	-14.1

Image	BOAT			CAMERAMAN			HOUSE			PEPPERS		
n/N	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	-21.4	-19.6	-18.6	**-20.9**	**-18.4**	**-17.1**	**-25.4**	-22.8	**-21.6**	**-20.0**	-17.5	-16.2
S-FISTA	-21.4	-19.5	-18.5	-20.7	-18.3	-17.0	-25.3	-22.7	-21.4	-19.9	-17.5	-16.1
SURE-IT	-20.3	-18.9	-18.3	-18.7	-16.7	-15.9	-23.3	-21.4	-20.6	-18.5	-16.0	-15.3
VDAMP-α	**-21.6**	**-19.9**	**-18.9**	-20.8	-18.2	-16.0	**-25.4**	**-22.9**	-21.5	-19.9	**-17.7**	**-16.7**
VDAMP-S	-19.5	-16.7	-15.6	-18.7	-16.1	-15.5	-21.5	-18.8	-17.8	-18.0	-15.7	-14.3

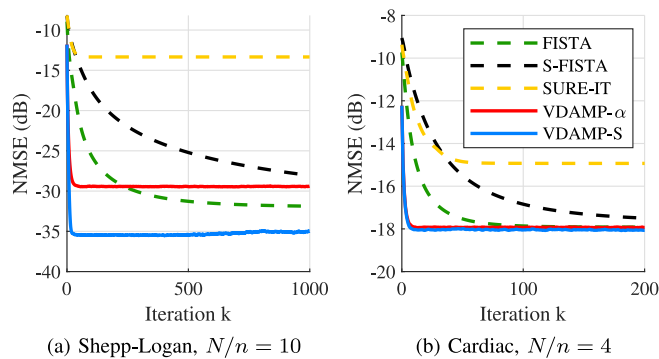(a) Shepp-Logan, $N/n = 10$

(b) Cardiac, $N/n = 4$

FIGURE 3. NMSE of FISTA, S-FISTA, SURE-IT, VDAMP-α and VDAMP-S of a Shepp-Logan undersampled at $N/n = 10$ and the cardiac image undersampled at $N/n = 4$. The NMSE at $k = 0$ differs between algorithms as the image estimate at $k = 0$ is defined to be the after the first thresholding is applied. The cardiac example is shown up to $k = 200$, not $K_{it} = 500$, so that the behavior of VDAMP can clearly be seen.

D. NMSE COMPARISON

Table 2 shows the normalized mean-squared error (NMSE) $\|\hat{x} - x_0\|_2^2 / \|x_0\|_2^2$ of the reconstructed image for each algorithm. The NMSE of FISTA, S-FISTA and VDAMP are generally comparable. Given that VDAMP has 13 model parameters while FISTA has one, one might expect the NMSE of VDAMP would consistently be lower. However, for the Cameraman, the cardiac image at $N/n = 8$, House at $N/n = 4$ and Peppers at $N/n = 4$ FISTA's NMSE was found to be lower. This is due to density compensation in the gradient step, line 4 of Algorithm 2, which effectively increases the measurement noise for coefficients sampled with low probability. Note that FISTA's NMSE advantage in these instances may not necessarily arise in realistic, prospectively undersampled reconstruction tasks, as the ground truth cannot be used to hand-tune FISTA's sparse weighting λ to a near optimal value, as in the experiments here. Also note that VDAMP-S and VDAMP-α both perform comparatively well on the MRI images, which are of primary importance for the algorithm's intended application.

Despite employing subband-dependent thesholding, S-FISTA's NMSE at K_{it} was often slightly higher than FISTA.

Fig. 5 shows $k = 1, 2, 3$ of VDAMP's wavelet-domain aliasing for the $N/n = 12$ Shepp-Logan, which demonstrates that the ratio of the aliasing between subbands may not be constant over iterations. For instance, at $k = 1$, the coarse level has greater variance than the fine levels, but at $k = 3$ the coarse variance is visibly lower than the fine levels. This is poorly reflected by S-FISTA's threshold weighting, which is fixed over iterations and not dependent on the current estimate. Further, [40] notes that while (27) is sufficient to ensure convergence, the inequality is not tight so may lead to weights that are smaller than necessary, which slows convergence. In [46], which uses different sampling schemes to that employed here, it was found that S-FISTA performed slightly better than FISTA, suggesting that S-FISTA's relative performance may be particularly dependent on the sampling scheme employed. The comparatively poor performance of SURE-IT highlights the need for zero-mean Gaussian aliasing for effective automatic parameter tuning with SURE.

E. EMPIRICAL EVIDENCE OF STATE EVOLUTION

This section presents empirical evidence that VDAMP obeys the colored state evolution given by (9) using kurtosis and quantile-quantile plots. The excess kurtosis of a real random variable X is defined as $\text{Kurt}\{X\} = \mu_4/\sigma^4 - 3$, where μ_4 is the fourth central moment and σ is the standard deviation. The Gaussianity of the aliasing of r_k was tested by calculating the mean of per-subband empirical kurtosis of the real part,

$$\overline{\text{Kurt}}\{\Re[r_k - w_0]\} = \frac{1}{3s+1} \sum_{b=1}^{3s+1} \text{Kurt}\{\Re[r_{k,b} - w_{0,b}]\},$$

and comparing to zero, which is the kurtosis of a white Gaussian distribution. Table 3 shows the mean kurtosis for all images and sampling factors at $k = K_{it}$. The proximity to zero is consistent with a colored state evolution for all image types and undersampling factors, and for both VDAMP-α and VDAMP-S, confirming that the difference in performance between the algorithms is not due to a breakdown in state evolution. The imaginary part, which is not included here for conciseness, was found to have a similarly small mean excess kurtosis.
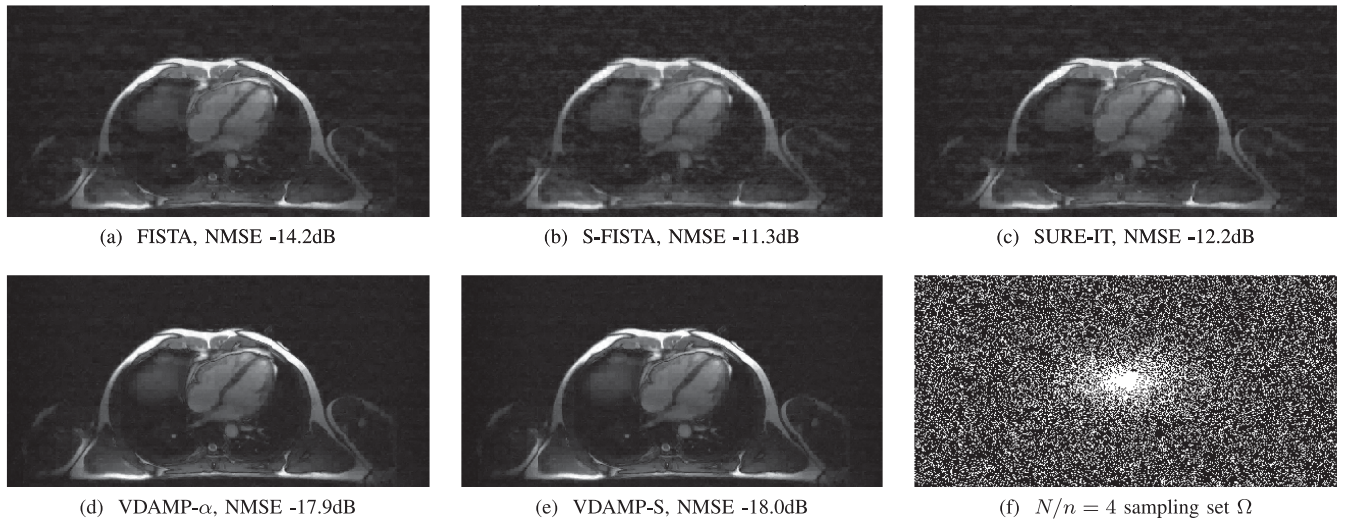
(a) FISTA, NMSE -14.2dB

(b) S-FISTA, NMSE -11.3dB

(c) SURE-IT, NMSE -12.2dB

(d) VDAMP-α, NMSE -17.9dB

(e) VDAMP-S, NMSE -18.0dB

(f) $N/n = 4$ sampling set Ω

FIGURE 4. Reconstructions of the Cardiac image undersampled at $N/n = 4$, shown at $k = 10$.

TABLE 3. Test for Gaussianity Using the Mean Excess Kurtosis $\overline{\text{Kurt}}\{\Re[r_k - w_0]\}$ at K_{it}. An Exact Gaussian has Zero Mean Excess Kurtosis. The Smallest Absolute Values are Highlighted in Bold

Image	SHEPP-LOGAN			BRAIN			CARDIAC			BARBARA		
n/N	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	27.99	32.57	38.01	2.86	4.43	5.76	3.94	6.35	9.55	2.10	2.75	3.09
S-FISTA	43.69	45.63	46.83	9.58	11.35	11.97	20.46	23.56	25.42	3.80	4.35	4.55
SURE-IT	39.27	49.64	51.08	4.57	7.32	8.76	6.82	13.92	21.99	2.38	2.88	3.21
VDAMP-α	**0.06**	**0.05**	0.13	-0.06	-0.02	**-0.03**	**-0.06**	**-0.05**	**-0.09**	0.04	**0.00**	**0.00**
VDAMP-S	0.10	0.07	**0.01**	**-0.03**	**-0.01**	0.06	0.20	0.07	0.11	**0.03**	0.10	0.04

Image	BOAT			CAMERAMAN			HOUSE			PEPPERS		
n/N	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	2.17	3.35	4.20	3.70	5.33	7.25	3.01	5.23	6.96	2.90	5.13	7.19
S-FISTA	4.79	5.53	5.96	7.26	8.24	8.92	6.14	7.54	8.45	10.04	11.49	12.17
SURE-IT	3.30	4.74	5.27	5.68	8.09	8.92	5.26	7.55	8.45	5.21	10.01	11.30
VDAMP-α	0.02	0.09	0.05	**-0.06**	**-0.04**	0.07	**-0.05**	**-0.01**	**-0.01**	0.00	0.07	**0.03**
VDAMP-S	**0.01**	**0.02**	**0.01**	-0.08	0.13	-0.09	-0.06	-0.04	-0.07	**-0.07**	**0.00**	-0.06

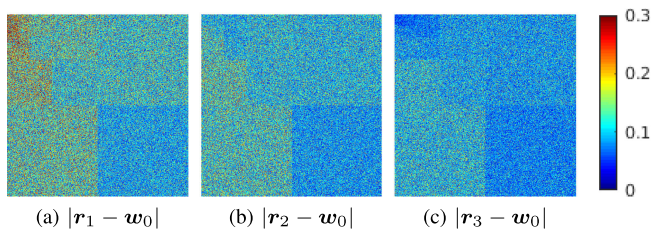(a) $|r_1 - w_0|$

(b) $|r_2 - w_0|$

(c) $|r_3 - w_0|$

FIGURE 5. The magnitude of the effective noise for VDAMP-S for a Shepp-Logan undersampled with $N/n = 12$ for iterations $k = 1, 2, 3$, where the colorbar shows the proportion of the maximum of x_0.

Using the example of the Shepp-Logan undersampled with $N/n = 12$, Fig. 5 shows VDAMP-S's $|r_k - w_0|$ for $k = 1, 2, 3$, visualizing the preservation of the unbiased subband-dependent aliasing structure shown for uniform sampling in Fig. 1. For $k = 0, 5, 20$, Fig. 6 shows quantile-quantile plots against a Gaussian of the three illustrative subbands of $r_k - w_0$: the diagonal detail at scale 1, the horizontal detail at scale 2 and the vertical detail at scale 4, where scale 1 is the finest and scale 4 is the coarsest. The linearity of the blue points

provides strong evidence that the per-subband effective noise is Gaussian.

The efficacy of automatic threshold selection with cSURE depends on how accurately the diagonal of Σ_k^2 is modeled by τ_k. For $k = 0, 1, \ldots, 20$, Fig. 7 shows the ground truth sub-band NMSE $\|r_{k,b} - w_{0,b}\|_2^2 / \|w_{0,b}\|_2^2$ at all four scales and the prediction of VDAMP, where the NMSE is per subband. The true NMSE is closely tracked by τ_k at all scales, which implies that parameter selection with SURE is truly near-optimal. Since, by Appendix C, τ_k is unbiased when the aliasing of r_k is zero mean and i.i.d., Fig. 7 also provides further verification of colored state evolution.

IV. CONCLUSION

Based on the observation that Fourier sampling from a non-uniform spectral density leads to colored aliasing, we propose VDAMP, an algorithm based on OAMP that obeys a colored state evolution. State evolution provides an effective way to tune model parameters via cSURE, implying that a single algorithm can be used for arbitrary variable density scheme and image type without the need for manual adjustment.
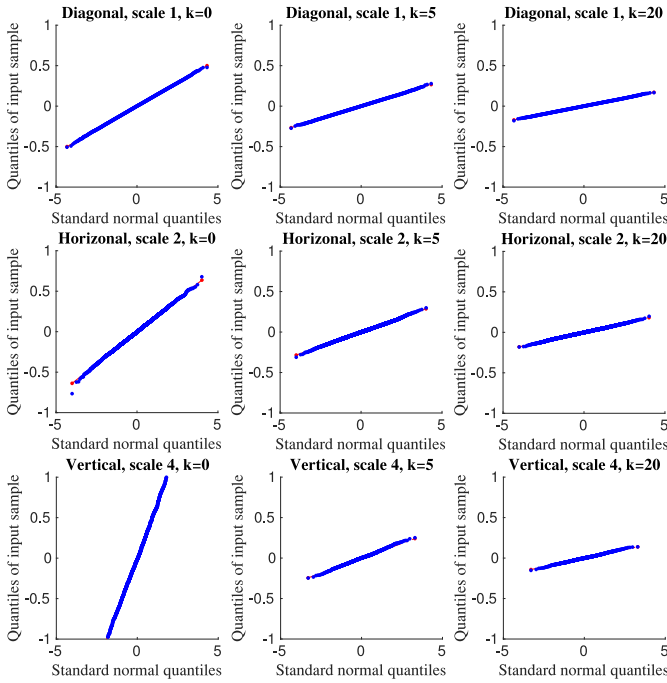
FIGURE 6. Normalized quantile-quantile plots against a Gaussian for three subbands of VDAMP-S's $r_k - w_0$ at $k = 0, 5, 20$ for the Shepp-Logan sampled with $N/n = 12$ in blue, and points along a straight line in red. The real part is plotted in the top and bottom rows and the imaginary is plotted in the middle row. Linearity of the blue points indicates that that the data comes from a Gaussian distribution, and the decreasing gradient shows that the variance decreases with increasing k. Finite dimensional effects causing small deviations from an exact Gaussian are more apparent at coarse scales, where the dimension is smaller.

More degrees of freedom are feasibly allowed in the model, enabling higher order prior information such as anisotropy, variability across scales and structured sparsity, without the need to estimate the structure a priori [50].

VDAMP was motivated by the application of compressed sensing to accelerated MRI. Developments are required for VDAMP to be applicable to MRI data acquired across multiple coils that possess a sensitivity profile [34], [51], and for VDAMP to be applicable to Fourier sampling with 1D readout curves, where elements of Ω are generated dependently.

It is known that the state evolution of OAMP holds for a wide range of denoisers $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ [25], [29]. In [52], a number of standard compressed sensing algorithms that leverage image denoisers designed for Gaussian noise were shown to perform well on MRI reconstruction tasks, despite the mismatch between the aliasing and its model. A sophisticated denoiser equipped to deal with wavelet coefficients corrupted with known colored Gaussian noise would be expected to perform well in conjunction with VDAMP. There has also been interest in algorithms that leverage the statistical modeling capabilities of neural networks [52]–[55]. VDAMP with a neural network denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ could accommodate ground-truth free training by using cSURE as the loss, as shown for AMP in [56], [57].

APPENDIX A
THE POWER SPECTRUM OF THE ALIASING OF $\tilde{x}$
This appendix proves that the magnitude of the entry-wise difference between $\boldsymbol{y}_0$ and the unbiased estimate $\boldsymbol{P}^{-1}\boldsymbol{y}$ is

$$\mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{y}_0 - \boldsymbol{P}^{-1}\boldsymbol{y}|^2\} = (\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{y}_0|^2 + \sigma_\varepsilon^2 \boldsymbol{P}^{-1}\mathbb{1}_N \quad (28)$$

Proof: Since the entries of $\boldsymbol{y}$ are independent and $\boldsymbol{P}^{-1}$ is diagonal, we can consider (28) as N distinct one-dimensional expressions. The ith entry of (28) is

$$\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|y_{0,i} - \frac{y_i}{p_i}\right|^2\right\}$$

$$= \mathbb{E}_{m_i,\varepsilon_i}\left\{\left|y_{0,i} - \frac{m_i}{p_i}(y_{0,i} + \varepsilon_i)\right|^2\right\}$$

$$= \mathbb{E}_{m_i,\varepsilon_i}\left\{\left|\left(1 - \frac{m_i}{p_i}\right)y_{0,i} - \frac{m_i}{p_i}\varepsilon_i\right|^2\right\}, \quad (29)$$

where m_i is the ith diagonal of $\boldsymbol{M}_\Omega$. By assumption, m_i is distributed according to a Bernoulli distribution with $\mathbb{E}_{m_i}\{m_i\} = p_i$. The expectation over m_i can therefore be found by resolving (29) at $m_i = 1$ and $m_i = 0$ and summing with weights p_i and $1 - p_i$ respectively:

$$\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|\left(1 - \frac{m_i}{p_i}\right)y_{0,i} - \frac{m_i}{p_i}\varepsilon_i\right|^2\right\}$$

$$= \mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1 - p_i}{p_i}\right)y_{0,i} - \frac{\varepsilon_i}{p_i}\right|^2 + (1 - p_i)|y_{0,i}|^2\right\}.$$

Expanding the first term, and noting that $\mathbb{E}_{\varepsilon_i}\{\varepsilon_i\} = 0$ and $\mathbb{E}_{\varepsilon_i}\{|\varepsilon_i|^2\} = \sigma_\varepsilon^2$,

$$\mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1 - p_i}{p_i}\right)y_{0,i} - \frac{\varepsilon_i}{p_i}\right|^2 + (1 - p_i)|y_{0,i}|^2\right\}$$

$$= [t]\mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1 - p_i}{p_i}\right)y_{0,i}\right|^2 + p_i\left|\frac{\varepsilon_i}{p_i}\right|^2\right.$$
$$\left. - \left(\frac{1 - p_i}{p_i}\right)(y_{0,i}^*\varepsilon_i + y_{0,i}\varepsilon_i^*) + (1 - p_i)|y_{0,i}|^2\right\}$$

$$= \frac{(1 - p_i)^2}{p_i}|y_{0,i}|^2 + \frac{\sigma_\varepsilon^2}{p_i} + (1 - p_i)|y_{0,i}|^2$$

$$= \left(\frac{1 - p_i}{p_i}\right)|y_{0,i}|^2 + \frac{\sigma_\varepsilon^2}{p_i}$$

which is the ith entry of the right-hand-side of (28). This completes the proof.

APPENDIX B
PROOF THAT τ_k^y IS UNBIASED
This appendix shows that the τ_k^y update, (15), is an unbiased estimate of the power spectrum of the aliasing of $\boldsymbol{r}_k$ (14):

$$\mathbb{E}\{\boldsymbol{M}_\Omega \boldsymbol{P}^{-1}[(\boldsymbol{P}^{-1} - \mathbb{1}_N)|z_k|^2 + \sigma_\varepsilon^2\mathbb{1}_N]\}$$

$$= (\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{F}\boldsymbol{\Psi}^H\tilde{\boldsymbol{r}}_k - \boldsymbol{y}_0|^2 + \sigma_\varepsilon^2 \boldsymbol{P}^{-1}\mathbb{1}_N \quad (30)$$
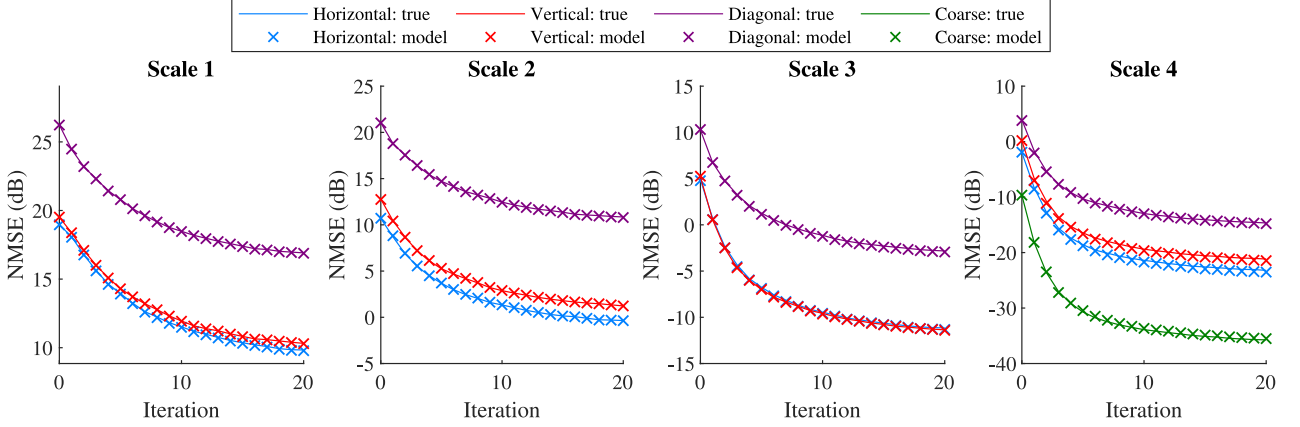
FIGURE 7. Per subband NMSE $\|r_{k,b} - w_{0,b}\|_2^2 / \|w_{0,b}\|_2^2$ versus iteration index k for a $N/n = 12$ undersampled Shepp-Logan reconstructed with VDAMP-S. Lines show the actual NMSE and crosses show the predictions from τ_k.

Proof: Defining $\widetilde{y}_k = F\Psi^H \widetilde{r}_k$, the ith entry of the left-hand-side of (30) in terms of y_0 is

$$\mathbb{E}_{m_i, \varepsilon_i} \left\{ \frac{m_i}{p_i} \left[\left(\frac{1 - p_i}{p_i} \right) |y_i - m_i \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2 \right] \right\}$$

$$= \mathbb{E}_{m_i, \varepsilon_i} \left\{ \frac{m_i}{p_i} \left[\left(\frac{1 - p_i}{p_i} \right) |m_i(y_{0,i} + \varepsilon_i) - m_i \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2 \right] \right\}$$

The expectation over m_i is

$$\mathbb{E}_{m_i, \varepsilon_i} \left\{ \frac{m_i}{p_i} \left[\left(\frac{1 - p_i}{p_i} \right) |m_i(y_{0,i} + \varepsilon_i) - m_i \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2 \right] \right\}$$

$$= \mathbb{E}_{\varepsilon_i} \left\{ \left(\frac{1 - p_i}{p_i} \right) |y_{0,i} + \varepsilon_i - \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2 \right\}$$

Since $\mathbb{E}_{\varepsilon_i}\{\varepsilon_i\} = 0$ and $\mathbb{E}_{\varepsilon_i}\{|\varepsilon_i|^2\} = \sigma_\varepsilon^2$,

$$\mathbb{E}_{\varepsilon_i} \left\{ \left(\frac{1 - p_i}{p_i} \right) |y_{0,i} + \varepsilon_i - \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2 \right\}$$

$$= \left(\frac{1 - p_i}{p_i} \right) (|y_{0,i} - \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2) + \sigma_\varepsilon^2$$

$$= \left(\frac{1 - p_i}{p_i} \right) |y_{0,i} - \widetilde{y}_{k,i}|^2 + \frac{\sigma_\varepsilon^2}{p_i}$$

which is the ith entry of the right-hand-side of (30). This completes the proof.

APPENDIX C
TRANSFORMING τ_k^y TO THE WAVELET DOMAIN
This appendix proves that when $\mathbb{E}\{r_k\} = w_0$, and the entries of r_k are independent,

$$\mathbb{E}\{|r_k - w_0|^2\} = |\Psi F^H|^2 \mathbb{E}\{\tau_k^y\} \qquad (31)$$

where τ_k^y is defined in (15).

Proof: Let the wavelet-domain error be $r_k - w_0 = Au$, where $u = F\Psi^H r_k - y_0$ is the Fourier-domain residual, and $A = \Psi F^H$, where the iteration index k has been removed to simplify notation. The ith entry of $|Au|^2$ is

$$\left| \sum_j A_{ij} u_j \right|^2 = \left(\sum_j A_{ij} u_j \right) \left(\sum_l A_{il}^* u_l^* \right)$$

$$= \sum_j \left(A_{ij} A_{ij}^* u_j u_j^* + \sum_{l \neq j} A_{ij} A_{il}^* u_j u_l^* \right)$$

Since, by assumption, r_k is unbiased and independent, the $l \neq j$ terms are zero in expectation. Therefore

$$\mathbb{E} \left\{ \left| \sum_j A_{ij} u_j \right|^2 \right\} = \mathbb{E} \left\{ \sum_j A_{ij} A_{ij}^* u_j u_j^* \right\}$$

$$= \sum_j |A_{ij}|^2 \mathbb{E}\{|u_j|^2\}.$$

Since, by Appendix B, $\mathbb{E}\{|u|^2\} = \mathbb{E}\{\tau_k^y\}$, we have

$$\mathbb{E}\{|Au|^2\} = |A|^2 \mathbb{E}\{|u|^2\} = |\Psi F^H|^2 \mathbb{E}\{\tau_k^y\}$$

This completes the proof.

APPENDIX D
SURE FOR COMPLEX VARIABLES
This appendix proves that cSURE, defined in (17), is an unbiased estimate of the risk, so that

$$\mathbb{E}\{\|d(v) - v_0\|_2^2\} = \mathbb{E}\{\text{cSURE}(d(v))\} \qquad (32)$$

where the expectation is over $v = v_0 + \mathcal{CN}(0, \tau_v \mathbb{1}_{N_v})$ and $d(v) = v + h(v)$ is a denoiser. The proof in this appendix is a complex noise variation on the standard proof of SURE, as found in [24], [58].

Proof: The Euclidean distance between the ground truth v_0 and the denoised vector $d(v)$ can be expanded as

$$\mathbb{E}\{\|d(v) - v_0\|_2^2\} = \mathbb{E}\{\|h(v)\|_2^2 + \|v - v_0\|_2^2$$
$$- 2(\Re[h(v)]^H \Re[v - v_0] + \Im[h(v)]^H \Im[v - v_0])\} \qquad (33)$$

By the noise model of $\boldsymbol{v}$, the second term on the right hand side is

$$\mathbb{E}\{\|\boldsymbol{v} - \boldsymbol{v}_0\|_2^2\} = N_v \tau_v \tag{34}$$

By Stein's lemma [24] (see also (6) of [58]), and recalling that $\mathcal{CN}(\boldsymbol{0}, \tau_v \mathbb{1}_{N_v})$ is defined such that the variance of the real and imaginary parts is $\tau_v/2$, the final term of the right-hand-side of (33) is

$$\mathbb{E}\{\Re[\boldsymbol{h}(\boldsymbol{v})]^H \Re[\boldsymbol{v} - \boldsymbol{v}_0]\} = \frac{\tau_v}{2} \mathbb{E}\left\{ \sum_j \frac{\partial \Re[h_j(\boldsymbol{v})]}{\partial \Re[v_j]} \right\}$$

$$= \frac{\tau_v}{2}\left(N_v - \mathbb{E}\left\{ \sum_j \frac{\partial \Re[d_j(\boldsymbol{v})]}{\partial \Re[v_j]} \right\} \right)$$

and similarly for the imaginary part. Overall, (33) is therefore

$$\mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v}) - \boldsymbol{v}_0\|_2^2\}$$

$$= \mathbb{E}\left\{ \|\boldsymbol{h}(\boldsymbol{v})\|_2^2 - N_v \tau_v + \tau_v \sum_j \left(\frac{\partial \Re[d_j(\boldsymbol{v})]}{\partial \Re[v_j]} + \frac{\partial \Im[d_j(\boldsymbol{v})]}{\partial \Im[v_j]} \right) \right\}$$

$$= \mathbb{E}\{\|\boldsymbol{h}(\boldsymbol{v})\|_2^2 + N_v \tau_v [2 \langle \boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v})) \rangle - 1]\}$$

which is the expectation of cSURE, given in (17). This completes the proof.

REFERENCES

[1] C. Millard, A. T. Hess, B. Mailhé, and J. Tanner, "An approximate message passing algorithm for rapid parameter-free compressed sensing MRI," no. 4, 2019. [Online]. Available: http://arxiv.org/abs/1911.01234

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1614066/

[3] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1580791/

[4] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Mag. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007. [Online]. Available: http://doi.wiley.com/10.1002/mrm.21391

[5] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Mag. Reson. Med.*, vol. 64, no. 3, pp. 767–776, Sep. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20535813

[6] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: A review of the clinical literature," *Brit. J. Radiol.*, vol. 88, no. 1056, Dec. 2015, Art. no. 20150487. [Online]. Available: http://www.birpublications.org/doi/10.1259/bjr.20150487

[7] J. C. Ye, "Compressed sensing MRI: A review from signal processing perspective," *BMC Biomed. Eng.*, vol. 1, no. 1, p. 8, Dec. 2019. [Online]. Available: https://bmcbiomedeng.biomedcentral.com/articles/10.1186/s42490-019-0006-z

[8] D. Donoho, "How high-dimensional geometry is transforming the MRI Industry," 2017. [Online]. Available: https://vimeo.com/225634059

[9] G. Puy, P. Vandergheynst, and Y. Wiaux, "On variable density compressive sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 595–598, Oct. 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5976374/

[10] Z. Wang and G. R. Arce, "Variable density compressed image sampling," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 264–270, Jan. 2010.

[11] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 612–622, Feb. 2014.

[12] N. Chauffert, P. Ciuciu, and P. Weiss, "Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies," *Proc. - Int. Symp. Biomed. Imag.*, pp. 298–301, 2013.

[13] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," *Forum Math., Sigma*, vol. 5, p. e4, Feb. 2017. [Online]. Available: https://www.cambridge.org/core/product/identifier/S2050509416000323/type/journal_article

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," Jul. 2009. [Online]. Available: http://dx.doi.org/10.1073/pnas.0909892106

[15] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5695122/

[16] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *The Ann. Appl. Probability*, vol. 25, no. 2, pp. 753–822, Apr. 2015. [Online]. Available: http://projecteuclid.org/euclid.aoap/1424355130

[17] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inf. Theory.*, Jun. 2014, pp. 236–240. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6874830

[18] F. Caltagirone, L. Zdeborova, and F. Krzakala, "On convergence of approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory.*, Jun. 2014, pp. 1812–1816. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6875146

[19] C. Guo and M. E. Davies, "Near optimal compressed sensing without priors: parametric SURE approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2130–2141, Apr. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7054509/

[20] S. Rangan, P. Schniter, E. Riegler, A. K. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," *IEEE Trans. Inform. Theory*, vol. 62, no. 12, pp. 7464–7474, Dec. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7600404/

[21] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7817805/

[22] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inform. Theory*, pp. 1–1, Oct. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8713501/

[23] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004. [Online]. Available: http://doi.wiley.com/10.1002/cpa.20042

[24] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981. [Online]. Available: http://projecteuclid.org/euclid.aos/1176345632

[25] Z. Xue, J. Ma, and X. Yuan, "D-OAMP: A denoising-based signal recovery algorithm for compressed sensing," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Dec. 2016, pp. 267–271. [Online]. Available: http://ieeexplore.ieee.org/document/7905845/

[26] B. Goossens, J. Aelterman, H. Luong, A. Pizurica, and W. Philips, "Wavelet-based analysis and estimation of colored noise," in *Proc. Discrete Wavelet Transforms - Algorithms Appl.*, H. Olkkonen, Ed. Intech, 2011, ch. 15, pp. 255–280.

[27] T. Li, M. Wang, and W. Xiong, "A novel method for filtering of Gaussian colored noise in images with wavelet transform," in *Proc. Int. Conf. Electron. Inf. Eng.*, vol. 1, no. ICEIE, pp. 184–189, 2010.

[28] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Stat. Soc. Series B: Statist. Methodol.*, vol. 59, no. 2, pp. 319–351, 1997.

[29] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.

[30] P. Schniter, S. Rangan, and A. Fletcher, "Plug-and-play image recovery using vector AMP." BASP Frontiers Workshop 2017, 2017. [Online]. Available: http://www2.ece.ohio-state.edu/ schniter/pdf/basp17_poster.pdf

[31] K. Sung, B. L. Daniel, and B. A. Hargreaves, "Location constrained approximate message passing for compressed sensing MRI," *Magnetic Resonance Med.*, vol. 70, pp. 370–381, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.24468

[32] E. M. Eksioglu and A. K. Tanc, "Denoising AMP for MRI reconstruction: BM3D-AMP-MRI," *SIAM J. Imag. Sci.*, vol. 11, no. 3, pp. 2090–2109, Jan. 2018. [Online]. Available: https://epubs.siam.org/doi/10.1137/18M1169655

[33] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: Rationale and an iterative numerical solution," *Mag. Reson. Med.*, vol. 41, no. 1, pp. 179–186, Jan. 1999. [Online]. Available: http://doi.wiley.com/10.1002/percent28SICIpercent291522-2594percent28199901percent2941percent3A1percent3C179percent3Apercent3AAID-MRM25percent3E3.0.COpercent3B2-V

[34] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Mag. Reson. Med.*, vol. 46, no. 4, pp. 638–651, Oct. 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11590639

[35] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press, 2006.

[36] S. Chatterjee, P. Diaconis *et al.*, "The sample size required in importance sampling," *The Ann. Appl. Probability*, vol. 28, no. 2, pp. 1099–1135, 2018.

[37] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Parameterless optimal approximate message passing," Oct. 2013. [Online]. Available: http://arxiv.org/abs/1311.0035

[38] M. Bayati, M. A. Erdogdu, and A. Montanari, "Estimating LASSO risk and noise level," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2013.

[39] C. Vonesch and M. Unser, "A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Apr. 2008.

[40] I. Bayram and I. W. Selesnick, "A subband adaptive iterative shrinkage/thresholding algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 3 PART 1, pp. 1131–1143, Mar. 2010.

[41] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, p. 1200, Dec. 1995. [Online]. Available: https://www.jstor.org/stable/2291512?origin=crossref

[42] K. Khare, C. J. Hardy, K. F. King, P. A. Turski, and L. Marinelli, "Accelerated MR imaging using compressive sensing with no free parameters," *Magn. Resonance Med.*, vol. 68, no. 5, pp. 1450–1457, Nov. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22266597

[43] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing," *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.

[44] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009. [Online]. Available: http://epubs.siam.org/doi/10.1137/080716542

[45] S. S. Vasanawala *et al.*, "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," *Proc. - Int. Symp. Biomed. Imag.*, pp. 1039–1043, 2011.

[46] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1649–1660, Sep. 2011.

[47] Y. Mäkinen *et al.*, "Image and video denoising by sparse 3D transform-domain collaborative filtering," 2020. [Online]. Available: http://www.cs.tut.fi/ foi/GCF-BM3D/index.html#ref_results

[48] C. M. Tsai and D. G. Nishimura, "Reduced aliasing artifacts using variable-density k-space sampling trajectories," *Magn. Resonance Med.*, vol. 43, no. 3, pp. 452–458, 2000.

[49] C. Lazarus *et al.*, "SPARKLING: Variable-density k-space filling curves for accelerated T2*-weighted MRI," *Magn. Resonance Med.*, vol. 81, no. 6, pp. 3643–3661, 2019.

[50] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5437428/

[51] D. Liang, B. Liu, J. Wang, and L. Ying, "Accelerating SENSE using compressed sensing," *Magn. Resonance Med.*, vol. 62, no. 6, pp. 1574–1584, Dec. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19785017

[52] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: using denoisers for image recovery," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 105–116, Jan. 2020.

[53] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, pp. 399–406.

[54] C. Metzler, A. Mousavi, and R. Baraniuk, "Learned D-AMP: Principled neural network based compressive image recovery," in *Proc. Adv. Neural Inf. Process. Syst. 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., pp. 1772–1783.

[55] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3113–3125, Jun. 2019.

[56] C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsupervised learning with stein's unbiased risk estimator," 2018. [Online]. Available: http://arxiv.org/abs/1805.10531

[57] M. Zhussip, S. Soltanayev, and S. Y. Chun, "Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 10 247–10 256, 2019.

[58] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

CHARLES MILLARD (Student Member, IEEE) received the M.Sci. degree in physics with theoretical physics from Imperial College London, London, U.K. He is currently working toward the D.Phil. degree with the University of Oxford, Oxford, U.K., working jointly with the Mathematical Institute and the Oxford Centre for Clinical Magnetic Resonance Research. His research focuses on the application of compressed sensing to magnetic resonance imaging.

AARON T. HESS received the Ph.D. degree from the University of Cape Town, Cape Town, South Africa. He is currently a Research Scientist with the Oxford Centre for Clinical Magnetic Resonance Research, University of Oxford, Oxford, U.K. His research investigates methods for high resolution MRI of the heart, including motion correction and use of ultra-high magnetic fields.

BORIS MAILHÉ (Member, IEEE) received the master's degree in computer science and the Ph.D. degree in signal processing from Rennes 1 University, Rennes, France. He is currently a Data Scientist with the Digital Technology and Innovation of Siemens Healthineers, Princeton, NJ, USA. His research interests include MR compressed sensing, image processing for medical imaging, and generative modeling.

JARED TANNER (Member, IEEE) received the B.S. degree in physics from The University of Utah, Salt Lake City, UT, USA, and the Ph.D. degree in applied mathematics from UCLA, Los Angeles, CA, USA, where his Ph.D. adviser was Eitan Tadmor. He is currently a Professor of the mathematics of information with the University of Oxford, Oxford, U.K., a Fellow of Exeter College, and Oxford Turing University Lead for the Alan Turing Institute. His research interests include signal processing with emphasis in compressed sensing, sparse approximation, applied Fourier analysis, and spectral methods.