An Approximate Message Passing Algorithm for Rapid Parameter-Free Compressed Sensing MRI

Charles Millard, Student Member, IEEE, Aaron T Hess, Boris Mailhé and Jared Tanner, Member, IEEE

Abstract—For certain sensing matrices, the Approximate Message Passing (AMP) algorithm and more recent Vector Approximate Message Passing (VAMP) algorithm efficiently reconstruct undersampled signals. However, in Magnetic Resonance Imaging (MRI), where Fourier coefficients of a natural image are sampled with variable density, AMP and VAMP encounter convergence problems. In response we present a new approximate message passing algorithm constructed specifically for variable density partial Fourier sensing matrices with a sparse model on wavelet coefficients. For the first time in this setting a state evolution has been observed. A practical advantage of state evolution is that Stein’s Unbiased Risk Estimate (SURE) can be effectively implemented, yielding an algorithm with no free parameters. We empirically evaluate the effectiveness of the parameter-free algorithm on simulated data and find that it converges over 5x faster and to a lower mean-squared error solution than Fast Iterative Shrinkage-Thresholding (FISTA).

Index Terms—Approximate message passing, compressed sensing, magnetic resonance imaging, Stein’s unbiased risk estimate

I. INTRODUCTION

We consider a complex linear regression problem, where complex data vector \( \mathbf{y} \in \mathbb{C}^n \) is formed of noisy linear measurements of a signal of interest \( \mathbf{x}_0 \in \mathbb{C}^N \):

\[
\mathbf{y} = \Phi \mathbf{x}_0 + \varepsilon,
\]

where \( \Phi \in \mathbb{C}^{n \times N} \) and \( \varepsilon \sim \mathcal{CN}(0, \sigma^2 \mathbb{I}_n) \), where \( \mathbb{I}_n \) is the \( n \times n \) identity matrix. Here, \( \mathcal{CN}(\mu, \Sigma) \) denotes the complex normal distribution with mean \( \mu \), covariance \( \Sigma \) and white phase. A well-studied approach is to seek a solution of

\[
\hat{x} = \arg\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \| \mathbf{y} - \Phi \mathbf{x} \|^2_2 + f(\mathbf{x})
\]

where \( f(\mathbf{x}) \) is a model-based penalty function. Compressed sensing [1, 2] concerns the reconstruction of \( \mathbf{x}_0 \) from underdetermined measurements \( n < N \). Commonly sparsity in \( \hat{x} \) is promoted by solving [2] with \( f(\mathbf{x}) = \lambda\|\Psi \mathbf{x}\|_1 \) for sparse weighting \( \lambda > 0 \) and sparsifying transform \( \Psi \).

Manuscript received January xx, xxxx; revised January xx, xxxx; accepted October xx, xxxx. Date of publication January xx, xxxx; date of current version January xx, xxxx. This work was supported by an EPSRC Industrial CASE studentship with Siemens Healthineers, voucher number 17000051, and by The Alan Turing Institute under the EPSRC grant EP/N510129/1. The concepts and information presented in this paper are based on research results that are not commercially available.

Charles Millard and Jared Tanner are with the Mathematical Institute at the University of Oxford, Oxford, OX2 6GG, UK (e-mail: millard@maths.ox.ac.uk; tanner@maths.ox.ac.uk)

Aaron T Hess is with the Oxford Centre for Clinical Magnetic Resonance at the University of Oxford, Oxford, OX3 9DU, UK (e-mail: aaron.hess@cardioox.ox.ac.uk)

Boris Mailhé is with Siemens Healthineers, 755 College Rd E, Princeton, NJ 08540, USA (e-mail: boris.mailhe@siemens-healthineers.com)

The Approximate Message Passing (AMP) algorithm [3] is an iterative method that estimates \( \mathbf{x}_0 \) in linear regression problems such as [4]. At iteration \( k \) AMP implements a denoiser \( g(r_k; \tau_k) \) on \( \mathbf{x}_0 \) estimate \( r_k \) with mean-squared error estimate \( \tau_k \). For instance, for problems of the form of [2], \( g(r_k; \tau_k) \) is the proximal operator associated with \( f(\mathbf{x}) \):

\[
g(r_k; \tau_k) = \arg\min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2\tau_k} \| r_k - \mathbf{x} \|^2_2 + f(\mathbf{x}),
\]

which is equal to soft thresholding in the case of \( f(\mathbf{x}) = \lambda \| \Psi \mathbf{x} \|_1 \) and orthogonal \( \Psi \). For certain sensing matrices and given mild conditions on \( f(\mathbf{x}) \), AMP’s state evolution guarantees that in the large system limit \( n, N \rightarrow \infty, n/N \rightarrow \delta \in (0, 1) \), vector \( r_k \) behaves as the original signal corrupted by zero-mean white Gaussian noise:

\[
r_k = \mathbf{x}_0 + \mathcal{CN}(0, \sigma^2 \mathbb{I}_N)
\]

where \( \sigma_k \) is an iteration-dependant standard deviation. The state evolution of AMP has been proven for real i.i.d. Gaussian measurements in [4] and i.i.d. sub-Gaussian measurements in [5]. It has also been empirically shown that state evolution holds for uniformly undersampled Fourier measurements of a random artificial signal [6]. When state evolution holds, AMP is known to exhibit very fast convergence. However, for generic \( \Phi \), the behavior of AMP is not well understood and it has been noted by a number of authors [6–10] that it can encounter convergence problems. The recently proposed Vector Approximate Message Passing (VAMP) [11] algorithm broadened the class of measurement matrices for which it holds, namely to those matrices that are ‘right-orthogonally invariant’, and was also found to perform very well on certain reconstruction tasks.

A. Approximate message passing for compressed sensing MRI

In compressed sensing MRI [12], measurements are formed of undersampled Fourier coefficients, so that \( \Phi = M_j \mathbf{F} \), where \( \mathbf{F} \) is a 2D or 3D discrete Fourier transform and \( M_j \in \mathbb{R}^{n \times N} \) is a undersampling mask that selects the \( j \)th row of \( \mathbf{F} \) if \( j \in \Omega \) for sampling set \( \Omega \). The signal of interest \( \mathbf{x}_0 \) is a natural image, so typically has a highly non-uniform spectral density that is concentrated at low frequencies. Accordingly, the sampling set \( \Omega \) is usually generated such that there is a higher probability of sampling lower frequencies. This work considers an \( \Omega \) with elements drawn independently from a Bernoulli distribution with non-uniform probability, such that \( \text{Prob}(j \in \Omega) = p_j \in [0, 1] \). In this variable density setting there are no theoretical guarantees for AMP or VAMP and in practice the behavior of [4] is not observed and the algorithms typically perform poorly.
This letter presents a new method for undersampled signal reconstruction which we term the Variable Density Approximate Message Passing (VDAMP) algorithm. For Fourier coefficients of a realistic image randomly sampled with variable density and orthogonal wavelet $\Psi$ we have empirical evidence that a state evolution occurs. Unlike the white effective noise of (4), the $r_k$ of VDAMP behaves as the ground truth corrupted by zero-mean Gaussian noise with a separate variance for each wavelet subband, such that

$$ r_k = w_0 + CN(0, \Sigma_k), $$

(5)

where $w_0 := \Psi x_0$ and the effective noise covariance $\Sigma_k$ is diagonal so that for a $\Psi$ with $s$ decomposition scales

$$ \Sigma_k = \begin{bmatrix} \sigma^2_{k,1} I_{N_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{k,2} I_{N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{k,1+3s} I_{N_1+3s} \end{bmatrix}, $$

(6)

where $\sigma^2_{k,j}$ and $N_j$ refer to the variance and dimension of the $j$th subband respectively. We refer to (5) as the \textit{colored} state evolution of VDAMP.

Selecting appropriate regularisation parameters such as $\lambda$ is a notable challenge in real-world compressed sensing MRI applications. We present an approach to parameter-free compressed sensing reconstruction using Stein’s Unbiased Risk Estimate (SURE) [13] in conjunction with VDAMP, building on the work on AMP introduced in [14]. A strength of the Constraints this imposes on the work on AMP introduced in [14]. A strength of automatic parameter tuning via SURE is that the it is possible to have a richer regularizer than would usually be feasible for a parameter-free denoiser [15].

For AMP and VAMP, (4) states that the effective noise $r_k - x_0$ is white, so can be fully characterised by a scalar $\tau_k$. This is appropriate for the kind of uniform measurement matrices and separable, identical sparse signal models $f(x)$ that are often encountered in abstract compressed sensing problems. However, when Fourier coefficients of a natural image are sampled the effective noise is colored, so is poorly represented by a scalar [18]. VDAMP models the richer structure of effective noise in the variable density setting with a vector $\tau_k$ that has one real number per wavelet subband.

---

**Algorithm 1 VDAMP**

**Require:** Sensing matrix $\Phi$, orthogonal wavelet transform $\Psi$, probability matrix $P$, measurements $y$, denoiser $g(r_k; \tau_k)$ and number of iterations $K_{it}$.

1. Set $\tau_0 = 0$ and compute $S = |\Phi \Phi^H|^2$
2. for $k = 0, 1, \ldots, K_{it} - 1$
3. $z_k = y - \Phi \Phi^H \tau_k$
4. $r_k = \tau_{k+1} + \Psi \Phi^H P^{-1} z_k$
5. $\tau_k = S^H P^{-1} \left( (P^{-1} - I_n) |z_k|^2 + \sigma^2 \epsilon 1_n \right)$
6. $w_k = g(r_k; \tau_k)$
7. $\alpha_k = (g'(r_k; \tau_k))_{\text{band}}$
8. $\tau_{k+1} = (w_k - \alpha_k \circ r_k) \circ (1 - \alpha_k)$
9. end for
10. return $\hat{x} = \Phi^H w_k + \Phi^H (y - \Phi \Phi^H w_k)$

---

The VDAMP algorithm is shown in Algorithm 1. Here, $P \in \mathbb{R}^{n \times n}$ is the diagonal matrix formed of sampling probabilities $p_j$ for $j \in \Omega$. The function $g(r_k; \tau_k)$ refers to some denoiser with a colored effective noise model. The notation $(g'(r_k; \tau_k))_{\text{band}}$ in line 7 refers to the (sub)-gradient of the denoiser averaged over subbands, so that for $s$ decomposition scales $\alpha_k$ has 1 + 3s unique entries, having the same structure as the $\alpha_k$ of (7).

In line 8, the notation $\circ$ is used for entry-wise multiplication and $\odot$ for entry-wise division. $| \cdot |$ refers to the entry-wise absolute magnitude of a vector or matrix.

The original AMP paper considers $\Psi = I_N$ and an i.i.d. Gaussian $\Phi$ that is normalized such that $E(\Phi \Phi^H) = I_N$. This ensures that $r_k$ is an unbiased estimate of $w_0$. For variable density sampling, the correct normalisation can be achieved by rescaling (1) by $P^{-1/2}$. In VDAMP this is manifest in the gradient step of lines 3-4, which feature a crucial weighting by $P^{-1}$. This provides the correct normalisation in expectation over $\Omega$: $E_{\Omega}(\Phi \Phi^H P^{-1} \Phi \Phi^H) = I_N$, which implies that

$$ E_{\Omega, \epsilon}(r_k) = E_{\Omega, \epsilon}(\tilde{r}_k + \Psi \Phi^H P^{-1} z_k) = w_0, $$

(8)

for any $\tilde{r}_k$. Such a rescaling is referred to as ‘density compensation’ in the MRI literature [19], [20], and was used in the original compressed sensing MRI paper with zero-filling to generate a unregularized, non-iterative baseline [12]. Line 5 of Algorithm 1 computes an estimate of the colored effective noise variance $|w_0 - r_k|^2$. It can be shown that $\tau_k$ is an unbiased estimate of the expected entry-wise squared error:

$$ E_{\Omega, \epsilon}(\tau_k) = E_{\Omega, \epsilon}(S^H P^{-1} \left( (P^{-1} - I_n) |z_k|^2 + \sigma^2 \epsilon 1_n \right)) $$

$$ = E_{\Omega, \epsilon}( |w_0 - r_k|^2), $$

(9)

for any $\tilde{r}_k$. We assume both estimators in (8) and (9) concentrate around their expectation, and leave the study of the constraints this imposes on $P$ for future works. Note that $S$ has 1 + 3s unique columns, so for fixed $s$ the complexity of VDAMP is governed by $\Psi$ and $\Phi$, whose fast implementations have complexity $O(N \log N)$.

Lines 6-8 are the model-based regularization phase from VAMP, but with a colored effective noise model, as in [21]. This phase includes the message passing Onsager correction term, which we have observed leads to the Gaussian effective noise of [5]. Line 10 implements an unweighted gradient step.
that enforces exact consistency of the image estimate with the measurements.

III. NUMERICAL EXPERIMENTS

In the experiments presented in this section the denoiser \(g(r_k; \tau_k)\) was the complex soft thresholding operator with an automatically tuned subband-dependant threshold. In other words, we used a complex adaption of SURE to approximately solve

\[
g(r_k; \tau_k) \equiv \arg\min_{w \in \mathbb{C}^N} \min_{\lambda \in \mathbb{R}^N} \frac{1}{2} ||w - r_k||^2 + \|\lambda \oplus w\|_1,
\]

where \(\sqrt{\tau_k}\) is the entry-wise square root of \(\tau_k\) and \(\lambda\) is of the form of \(\tau^v\). \((\ref{Eq:11})\) was solved using a procedure similar to the usual SureShrink but with a version of SURE adapted for effective noise that is complex and colored. Consider a ground truth subband \(v_0 \in \mathbb{R}^{N_v}\) corrupted by complex Gaussian noise with white phase: \(v = v_0 + \mathcal{N}(0, \tau, \mathbb{I}_{N_v})\). For complex soft thresholding with threshold \(t\), it can be shown that

\[
cSURE(t; v) = (t^2 + 2 \tau v) \cdot \# \{i : |v_i| > t\} - N_v \tau v + \sum_{i:|v_i| \leq t} |v_i|^2 + \sum_{i:|v_i| > t} \tau v_i/v_i,
\]

(\(\ref{Eq:12}\)) is an unbiased estimate of the risk. For each subband the optimal threshold was estimated via

\[
i = \arg\min_{t} (cSURE(t; v))
\]

by evaluating \((\ref{Eq:12})\) 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 dimensionality, there is a trade-off between the size of \(s\) and the quality of threshold selection with \(cSURE\).

We considered the reconstruction of a \(512 \times 512\) Shepp-Logan artificially corrupted with complex Gaussian noise with white phase so that \(\|F x_0\|_2^2/\sigma_x^2 = 40\)dB. We assumed that \(\sigma_x^2\) was known a priori; in practice it can be well estimated with an empty prescan. All sampling probabilities \(p_j\) were generated using the polynomial variable density sampling function from the Sparse MRI package available at https://people.eecs.berkeley.edu/~mlustig/Software.html. We considered a Haar wavelet \(\Psi\) at \(s = 4\) decomposition scales, which are referred to as scales 1-4, where scale 1 is the finest and scale 4 is the coarsest. All experiments were conducted in MATLAB 9.4 on a 1.70 GHz Intel i5-8350U processor with 8GB of RAM, and can be reproduced with code available at https://github.com/charlesmillard/VDAMP.

A. Comparison with FISTA and SURE-IT

To establish comparative reconstruction quality and convergence speed, VDAMP was compared with FISTA and SURE-IT. For FISTA we used a sparse weighting \(\lambda\) tuned with an exhaustive search so that the mean-squared error was minimised after 10 seconds. For SURE-IT the mean-squared error estimate was updated using the ground truth: \(\tau_k = \|w_0 - r_k\|_2^2/N\), and \((\ref{Eq:11})\) with \(\tau_k = \tau_k \mathbb{I}_{N_v}\) was implemented. All three algorithms were initialised with a vector of zeros.

Three sampling sets \(\Omega\) were generated at undersampling factors of approximately 4, 6 and 8, and VDAMP, FISTA and SURE-IT were run for 10 seconds. Fig. 1 shows the NMSE \(= \|\hat{x} - x_0\|_2^2/\|x_0\|_2^2\) as a function of time for each algorithm. \(\hat{x}\) at every iteration was calculated so that exact data consistency was ensured, as in line 10 of VDAMP. The mean per-iteration compute time was 0.065s for FISTA, 0.077s for SURE-IT, and 0.091s for VDAMP. Fig. 2 shows the ground truth image, the sampling set, and FISTA and VDAMP reconstruction at undersampling factor 8 after 2s. The entry-wise error \(\|\hat{x} - x_0\|\) is also shown for both algorithms.

B. State evolution and error tracking

This section empirically analyses VDAMP in greater depth, continuing to use the illustrative case of undersampling factor
Fig. 3. $|r_k - w_0|$ of VDAMP for $k = 0$, $k = 1$ and $k = 2$.

Fig. 4. Normalized quantile-quantile plots against a Gaussian of three subbands of $r_k - w_0$ for $k = 0$, $k = 1$ and $k = 2$ 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 the data comes from a Gaussian distribution. Finite dimensional effects causing small deviations from an exact Gaussian are more apparent at coarse scales, where the dimension is smaller.

8. In Fig. 3 the absolute value of the residual of $r_k$ is shown for three representative iterations: $k = 0$, $k = 1$ and $k = 2$. For the same iterations, Fig. 4 shows quantile-quantile plots of the real parts of three illustrative subbands of $r_k - w_0$: the diagonal detail at scale 1, the horizontal detail at scale 2 and the vertical detail scale 4. These figures provide empirical evidence that the effective noise of VDAMP evolves as $\mathbf{F}$.

The efficacy of automatic threshold selection with cSURE depends on how accurately the diagonal of $\Sigma_k$ from (6) is modelled by $\tau_k$. For $k = 0, 1, \ldots, 20$ Fig. 5 shows the ground truth NMSE of the wavelet coefficients at all four scales and the prediction of VDAMP, where the NMSE is per subband: $\text{NMSE}(v) = \|v - v_0\|^2_2/\|v_0\|^2_2$.

### IV. Conclusions

Like the original AMP paper [1], the claim of a state evolution has been substantiated with empirical evidence only. Theoretical work is required to establish the generality of the state evolution observed in these experiments.

In order for the algorithm to be compatible with realistic MRI acquisition protocols, practical developments are required to allow sampling with readout lines and multiple coils. The experiments conducted in this letter indicate that given such developments, VDAMP would have a number of significant advantages over algorithms currently used in the compressed sensing MRI literature. VDAMP’s state evolution provides an informed and efficient way to tune model parameters via SURE, implying that a single algorithm can be used for any image type and variable density scheme without the need for manual adjustment. More degrees of freedom are 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 such as in model-based compressed sensing [22]. The substantial reduction in required compute time observed in these experiments could be clinically valuable as real MRI reconstruction tasks typically have very high dimensionality, so it is often only viable to run a reconstruction algorithm for a limited number of iterations.

It is known that the state evolution of AMP holds for a wide range of denoisers $g(r_k; \tau_k)$ [23]. In a recent survey [24] 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 effective noise 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.
