EXPECTATION CONSISTENT PLUG-AND-PLAY FOR MRI

Saurav K. Shastri\textsuperscript{1}, Rizwan Ahmad\textsuperscript{\ast}, Christopher A. Metzler\textsuperscript{\ast}, and Philip Schniter\textsuperscript{\dagger}

\textsuperscript{1}Dept. ECE, The Ohio State Univ., Columbus, OH, 43210, \{shastri.19, schniter.1\}@osu.edu
\textsuperscript{\ast}Dept. BME, The Ohio State Univ., Columbus, OH, 43210, rizwan.ahmad@osumc.edu
\textsuperscript{\dagger}Dept. CS, The Univ. of Maryland, College Park, MD, 20742, metzler@umd.edu

ABSTRACT

For image recovery problems, plug-and-play (PnP) methods have been developed that replace the proximal step in an optimization algorithm with a call to an application-specific denoiser, often implemented using a deep neural network. Although such methods have been successful, they can be improved. For example, the denoiser is often trained using white Gaussian noise, while PnP’s denoiser input error is often far from white and Gaussian, with statistics that are difficult to predict from iteration to iteration. PnP methods based on approximate message passing (AMP) are an exception, but only when the forward operator behaves like a large random matrix. In this work, we design a PnP method using the expectation consistent (EC) approximation algorithm, a generalization of AMP, that offers predictable error statistics at each iteration, from which a deep-net denoiser can be effectively trained.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a medical imaging approach that uses magnetic fields to create detailed anatomical images. Although MRI provides excellent soft-tissue contrast without the use of ionizing radiation, it takes a long time to fully sample the measurement space. Thus, it is common to take relatively few measurements and apply sophisticated post-processing to reconstruct an accurate image. Although our paper focuses on MRI, the methods we propose apply to any application where the goal is to recover a signal from undersampled Fourier measurements.

The measurements $y \in \mathbb{C}^{CM}$ collected in $C$-coil MRI, known as “k-space” measurements, can be modeled as

$$y = Ax_0 + w,$$

where $x_0 \in \mathbb{C}^N$ is a vectorized version of the $N$-pixel image we wish to recover, $F \in \mathbb{C}^{N \times N}$ is a unitary 2D discrete Fourier transform (DFT), $M \in \mathbb{R}^{M \times N}$ is a sampling mask formed from $M$ rows of the identity matrix $I \in \mathbb{R}^{N \times N}$, $s_c \in \mathbb{C}^N$ is the $c$th coil-sensitivity map, and $w \sim N(0, I/\gamma_w)$ is additive white Gaussian noise (AWGN) with precision $1/\gamma_w$. In the special case of single-coil MRI, $C = 1$ and $s_1 = 1$, the all-ones vector. In this paper, we use the variable-density sampling masks like that shown in Fig. 1. In MRI, the pixels-to-measurement ratio, $R \triangleq N/M$, is known as the “acceleration rate.” When $R > 1$, one cannot uniquely determine $x_0$ from $y$ due to the nullspace of $A$, and so prior information about $x_0$ is needed for its recovery.

To recover images from undersampled MRI measurements, many methods have been proposed. Some are based on iterative optimization [1, 2], as described in the sequel. More recently, deep neural networks (DNNs) that directly map MRI measurements $y$ to an image $\hat{x}$ have been proposed, e.g., [3, 4]. Although such DNNs work well, training them requires huge fully-sampled k-space datasets (which may be difficult or impossible to obtain) and changes in the acquisition parameters (e.g., sampling mask) from training to testing can degrade performance [5].

In this work, we focus on the “plug and play” approach that iteratively calls a DNN for image denoising, which brings several advantages. First, DNN denoisers can be trained using image patches, implying the need for relatively few images and no k-space data. Second, the denoiser is trained independently of the acquisition parameters, so that it generalizes to any acquisition scenario. Our approach is based on the generalized expectation consistent (GEC) approximation algorithm from [6], which lives in the family of approximate message passing (AMP) algorithms like [7, 8].

2. BACKGROUND

2.1. Compressed-sensing-based methods

The conventional approach to MRI image recovery [1, 2] is to pose and solve an optimization problem of the form

$$\hat{x} = \text{arg min}_x \{ g_1(x) + g_2(x) \},$$

where $g_1(x)$ promotes measurement fidelity and $g_2(x)$ is an image-based regularizer. Typical choices are

$$g_1(x) = \frac{\gamma_w}{2} \| y - Ax \|^2$$

for (1) and $g_2(x) = \lambda \| \Psi x \|_1$ with a suitable transform $\Psi$ (e.g., wavelet or total-variation) and carefully chosen $\lambda > 0$. Such $g_2$ encourage sparsity in the transform coefficients $\Psi x$. 
Many algorithms have been proposed to solve (2) with convex
$g_1$ and $g_2$ [2]. For example, the alternating directions method of
multipliers (ADMM) [9] iterates
\begin{align}
x_t &\leftarrow \text{prox}_{g_1}(x_t - u) \\
y_t &\leftarrow \text{prox}_{g_2}(y_t + u) \\
u_t &\leftarrow u + (x_t - x_{t-1}),
\end{align}
where the proximal operator is defined as
\[
\text{prox}_g(r) \triangleq \arg \min_x \{ \rho(x) + \frac{1}{2} \| x - r \|^2 \}.
\]
In (4), $\gamma > 0$ is a tunable stepsize that affects the speed of ADMM’s
convergence but not its fixed point.

2.2. Plug-and-play methods
The prox operator (5) can be interpreted as a denoiser, in particu-
lar, the maximum a posteriori (MAP) estimator of $x_0$ with prior
$p(x_0) \propto e^{-\frac{1}{2}g(x_0)}$ from an observation $r = x_0 + w$ with $\gamma$
precision AWGN $e$. Leveraging this fact, Bouman et al. [10] pro-
posed to replace ADMM line (4b) with a call to a high-performance
image denoiser $f_2(x)$ like BM3D [11] or DnCNN [12], giving
rise to “plug-and-play” (PnP) ADMM. PnP extensions of other al-
gorithms, such as primal-dual splitting (PDS) [13] and proximal
gradient (PG) [14], have also been proposed. As shown in the recent
overview paper, PnP methods have been shown to significantly out-
perform compressed-sensing-based approaches in MRI [5]. Note,
however, that when $\gamma$ is increased with a denoising step of the form
“$x_t \leftarrow f_2(x_t + u)$,” the stepsize $\gamma$ does affect the fixed-point [5]
and thus must be tuned.

Although PnP algorithms work well in MRI, there is room for
improvement. For example, while image denoisers are typically de-
signed/trained to remove the effects of AWGN, PnP algorithms do
not provide the denoiser with an AWGN-corrupted input at each
iteration. Rather, the denoiser’s input error has iteration-dependent
statistics that are difficult to analyze or predict.

2.3. Approximate message passing
In (2), if we interpret $g_1(x)$ as a log-likelihood and $g_2(x)$ as a log
prior, then $x$ can be interpreted as the MAP estimate of $x$ from $y$.
However, because image recovery results are often judged by mean-
squared error (MSE), one may be more interested in the minimum
MSE (MMSE) estimate of $x$ from $y$. Interestingly, both MMSE and
MAP estimation are facilitated by approximate message passing
(APM) methods like [7,8].

The AMP algorithm from [7] iterates
\begin{align}
v_{t+1} &\leftarrow \beta \cdot \left( y - A x_t + \frac{1}{\sqrt{T}} \left\{ \nabla f_2^0(x_t) + A^H v_t \right\} \right) \\
\epsilon_{t+1} &\leftarrow \frac{1}{\beta} \| v_{t+1} \|^2 \\
x_{t+1} &\leftarrow f_2^{t+1}(x_{t+1} + A^H \epsilon_{t+1})
\end{align}
over $t = 0, 1, 2, \ldots$, starting from $v_0 = 0 = x_0$, where $f_2^0(\cdot)$ is a
Lipschitz denoising function, $\left\{ \nabla f_2^0(r) \right\}$ is the trace of the Jaco-
bian of $f_2^0$ at $r$, and $\beta = N / \| A \|^2$. When configured for MAP esti-
mation, AMP uses the MAP denoiser $f_2^0(r) = \text{prox}_{g_2}(r)$. When
configured for MMSE estimation, AMP instead uses the MMSE
denoiser $f_3^0(r) = \mathbb{E} [ x | r ]$ for $r = x + w$ with $w \sim \mathcal{N}(0, \sigma^2 I)$.

Importantly, when the forward operator $A$ is large and i.i.d. sub-
Gaussian, and when $f_2^0$ is Lipschitz, the macroscopic behavior of
AMP is rigorously characterized by a scalar state-evolution [15,16].

When $f_2^0$ is the MMSE denoiser and the state-evolution has a unique
fixed point, AMP provably converges to the MMSE-optimal esti-
mate $x_{\text{MMSE}}$ [15,16]. For images, the MMSE denoiser can be ap-
proximated by BM3D or a DNN, as proposed in [17], leading to
“denoising-AMP” (D-AMP). There, the trace-Jacobian in (6a) is ap-
proximated using the Monte-Carlo approach [18]
\[
\text{tr} ( \nabla f_2^0(r) ) \approx \delta^{-1} q^H ( f_2^0(r + \delta q) - f_2^0(r) ),
\]
with random $q \sim \mathcal{N}(0, I)$ and small $\delta > 0$.

More recently, the vector AMP (VAMP) algorithm [8] was pro-
posed, with similar properties as AMP (e.g., rigorous state evolution
and provable MMSE estimation) but applicability to a wider class of
random matrices: right orthogonally invariant (ROI) ones. Inspired
by D-AMP, a denoising VAMP (D-VAMP) was proposed in [19] and
analyzed in [20].

2.4. AMP for MRI
Neither AMP nor VAMP works as intended in MRI because $A$ in (1)
lacks sufficient randomness. In fact, these algorithms tend to diverge
in MRI if applied without modification.

The failure of AMP and VAMP can be understood from their
error recursions. For AMP, the error recursion is [21]
\[
e^{t+1} = (I - A^T A) e^t + A^T ( w + \frac{1}{\sqrt{T}} v^t \{ \nabla f_2^0(x_0 + e^t) \} )
\]
\[
e^{t+1} = f_2^{t+1}(x_0 + e^t) - x_0.
\]
It is important to keep in mind that images $x_0$ have much more en-
ergy at low Fourier frequencies than at high ones. The same tends
to be true of the output error $e^t$ of an image denoiser. Even so, if
$A \in \mathbb{R}^{M \times N}$ was large and i.i.d. (with zero mean and element-
wise variance $\frac{1}{T}$), then the $I - A^T A$ term in (8) would randomize $e^t$
such that the denoiser input error vector $e^{t+1}$ looks like AWGN. In
MRI, however, both $A$ and $e^t$ have Fourier structure, this random-
ization does not happen, and AMP behaves unpredictably. A similar
behavior plagues VAMP.

Several MRI-specific variations of AMP and VAMP have been
proposed to counter these deficiencies. For example, [22] proposed
D-AMP with a very small $\beta$, which helps the algorithm converge, but
at the cost of degrading its fixed points. [23] proposed a damped D-
VAMP that, combined with a novel initialization, showed improved
performance and runtime over Pn-ADMM for MRI.

Several other VAMP-based algorithms for MRI have been de-
signed to recover the wavelet coefficients of the image rather than
the image itself. The motivation is that, in this case, $A$ is a Fourier-
Wavelet matrix, which is approximately block diagonal [24], where
the blocks correspond to the wavelet subbands. With an appro-
priate modification of VAMP, the subband error vectors can be made
to behave more like AWGN, albeit with different variances. The
first incarnation of this idea appeared in [25], where a fixed band-
wise normalization procedure was used. Later, for single-coil MRI
with variable-density sampling masks, a “variable density AMP”
(VDAMP) algorithm with band-specific adaptive wavelet thresh-
olding was proposed in [26], which was able to successfully predict
the noise variance in each subband at each iteration. More recently, the
D-VDAMP algorithm [27] extended VDAMP to DNN denoising in
each subband.

Although D-VDAMP is the state-of-the-art AMP algorithm for
MRI, it is based on a non-standard modification of VAMP with de-
graded fixed points, which makes early stopping critical for good
performance. Also, it is not clear how to extend D-VDAMP to multi-
core MRI. These issues motivate our approach, which is described
next.
3. PROPOSED APPROACH

3.1. Denoising GEC

Our approach uses the GEC framework from [6], which is summarized in Alg. 1. When solving a convex optimization problem of the form (2), the functions $f_i$ in Alg. 1 take the form

$$f_i(r, \gamma) = \text{gprox}_{\rho, \gamma}(r), \quad i = 1, 2,$$

for the generalized proximal operator

$$\text{gprox}_{\rho, \gamma}(r) \triangleq \arg \min_x \left\{ \rho(x) + \frac{1}{2} \|x - r\|^2 \right\},$$

where $\|q\|_\gamma \triangleq \sqrt{q^T \text{Diag}(\gamma) q}$ and $\text{Diag}(\cdot)$ creates a diagonal matrix from its vector argument. Note that if $\gamma = \gamma 1$, then $\text{gprox}_{\rho, \gamma} = \text{gprox}_{\rho, 1}$. Furthermore, if the $\gamma_i$ vectors were held fixed over the iterations and took the form $\gamma_i = \gamma_1$, then Alg. 1 reduces to a variant of ADMM (4) with two dual updates: (4c) and a similar step between (4a) and (4b). So, GEC can be interpreted as an ADMM-like algorithm with two adaptive vector-valued stepsizes, $\gamma_1$ and $\gamma_2$.

**Algorithm 1 Generalized EC (GEC)**

**Require:** $f_1(\cdot, \cdot), f_2(\cdot, \cdot)$, and $\text{gdiag}(\cdot)$.

1: Select initial $r_1, \gamma_1$.
2: repeat
3: // Measurement fidelity
4: $\tilde{x}_1 \leftarrow f_1(r_1, \gamma_1)$
5: $\eta_1 \leftarrow \text{Diag}(\text{gdiag}(\nabla f_1(r_1, \gamma_1)))^{-1} \gamma_1$
6: $\gamma_2 \leftarrow \eta_1 - \gamma_1$
7: $r_2 \leftarrow \text{Diag}(\gamma_2)^{-1} (\text{Diag}(\eta_1) \tilde{x}_1 - \text{Diag}(\gamma_1)r_1)$
8: // Denoising
9: $\tilde{x}_2 \leftarrow f_2(r_2, \gamma_2)$
10: $\eta_2 \leftarrow \text{Diag}(\text{gdiag}(\nabla f_2(r_2, \gamma_2)))^{-1} \gamma_2$
11: $\gamma_1 \leftarrow \eta_2 - \gamma_2$
12: $r_1 \leftarrow \text{Diag}(\gamma_1)^{-1} (\text{Diag}(\eta_2) \tilde{x}_2 - \text{Diag}(\gamma_2)r_2)$
13: until Terminated

In lines 5 and 10, GEC averages the diagonal of the Jacobian separately over $L$ coefficient subsets using $\text{gdiag}: \mathbb{R}^{N \times N} \to \mathbb{R}^N$:

$$\text{gdiag}(Q) \triangleq [d_11_{N_1}, \ldots, d_L1_{N_L}]^T, \quad d_i = \frac{\text{tr}(Q_{i,i})}{N_i}. \quad (12)$$

In (12), $N_i$ is the size of the $i$th subset and $Q_{i,i} \in \mathbb{R}^{N_i \times N_i}$ is the $i$th diagonal subblock of the matrix input $Q$. When $L = 1$, GEC reduces to VAMP.

We focus on the quadratic loss (3), which yields

$$f_i(r, \gamma) = (\gamma w A^H A + \text{Diag}(\gamma))^{-1} (\gamma w A^H y + \text{Diag}(\gamma)r). \quad (13)$$

For $f_2$, we propose to “plug in” a DNN denoiser. For both $f_1$ and $f_2$, we approximate the $\text{tr}(Q_{i,i})$ term in (12) using

$$\text{tr}(Q_{i,i}) \approx \delta_i^{-1} q_i^H [f_i(r + \delta q_i, \gamma) - f_i(r, \gamma)], \quad (14)$$

where the $i$th coefficient subset in $q_i$ is i.i.d. unit-variance Gaussian and the others are zero. Inspired by D-AMP and D-VAMP, we call this approach “denoising GEC” (D-GEC).

3.2. D-GEC for Image Recovery

Like [25–27], we recover the wavelet coefficients $c_0$ rather than the image pixels $x_0$. For orthogonal wavelet transform $\Psi$, we have $c_0 = \Psi x_0$ and $x_0 = \Psi^T c_0$, so that we can rewrite (1) as

$$y = Bc_0 + w$$

with $B \triangleq A \Psi^T$. \quad (15)

To apply D-GEC to $c_0$-recovery, we choose $f_1$ as in (13), but with $B$ in place of $A$, and for the diagonalization subsets we choose the $L = 3D + 1$ subbands of a depth-$D$ 2D wavelet transform. As in [27], we perform denoising in the wavelet domain using a denoiser $f_2$ that can exploit knowledge of the noise variance in each wavelet subband, as provided by the precision vector $\gamma_2$.

4. NUMERICAL EXPERIMENTS

We now compare the proposed D-GEC algorithm to the existing D-VDAMP [27] and PnP-PDS [13] algorithms. Based on the extensive experiments in [27], D-VDAMP is state-of-the-art among PnP algorithms. PnP-PDS is a useful baseline, since it has the same fixed points as PnP-ADMM and PnP-PG.

**Denoisers:** For a fair comparison to D-VDAMP [27], we use the DNN denoiser proposed in [27], which is modification of DnCNN [12] that accepts the noise standard deviation (SD) in each wavelet subband. The denoiser was trained using noise that was white in each subband but with SD that varies across subbands. In particular, 5 copies of the denoiser were trained using subband noise SDs uniformly distributed in the ranges 0-10, 10-20, 20-50, 50-120, and 120-500, respectively. The DNNs were trained using patches from 70 MRI images of the Stanford 2D FSE dataset available at http://mridata.org. For PnP-PDS, we used a standard DnCNN denoiser trained on the same data with white noise of SD uniformly distributed in 20-50, as in [27]. Because we used real-valued images, the denoisers use only the real part of the input and generate a real-valued output.

**Test data:** For evaluation, we used the ten 352×352 MRI images in Fig. 2, which were not in the training dataset. The measurements $y$ were constructed using (1) with complex AWGN $\omega$ whose variance was adjusted to give a pre-masking SNR of 40 dB. For the multicoil experiments, we used coil sensitivities $s_i$, simulated using the Biot-Savart law, while in the single-coil case, we used $s_1 = 1$.

**Algorithm parameters:** For D-GEC and D-VDAMP, we used a 2D Haar wavelet transform with $D = 4$ levels, giving $L = 13$ wavelet subbands. D-GEC used the auto-tuning scheme from [28] and the damping scheme from [23] with parameter 0.4. D-VDAMP code was obtained from the authors and run under default settings, which are detailed in [27]. PnP-PDS was run for 200 iterations using the stepsize that maximized PSNR on the training set.

**Single-coil results:** Table 1 shows that D-GEC outperformed D-VDAMP in all single-coil experiments and outperformed PnP-PDS in all but SSIM at $R = 8$. Figure 5 shows an example of the wavelet

Fig. 2: Test images from http://mridata.org.
coefficients input to D-GEC’s denoiser at the 10th iteration, and their error relative to the true coefficients. Figure 3 shows the evolution of the standard deviation at the input to D-GEC’s denoiser in each subband; there is a good agreement between true and predicted values. Figure 6 suggests that the subband errors are Gaussian. Figure 4 shows image recoveries and error maps for one test image at \( R = 4 \).

| method     | \( R = 4 \) PSNR | SSIM | \( R = 8 \) PSNR | SSIM |
|------------|------------------|------|------------------|------|
| PnP-PDS    | 40.66            | 0.968| 37.38            | 0.951|
| D-VDAMP    | 42.36            | 0.972| 35.92            | 0.918|
| D-GEC      | 42.97            | 0.977| 37.65            | 0.946|

Table 1: Recovery results averaged over the 10 test images.

5. CONCLUSION

We designed a GEC-based plug-and-play algorithm for MRI that renders the subband errors white and Gaussian with predictable variance, and used it with a denoiser trained to handle subband errors that are white and Gaussian with known variance. Experiments show good performance relative to previous approaches in single- and multi-coil settings.
6. REFERENCES

[1] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 72–82, Mar. 2008.

[2] J. A. Fessler, “Optimization methods for magnetic resonance image reconstruction,” IEEE Signal Process. Mag., vol. 37, no. 1, pp. 33–40, 2020.

[3] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” IEEE Trans. Image Process., vol. 26, no. 9, pp. 4509–4522, 2017.

[4] K. Hammernik, T. Klatzer, E. Kobler, M. P. Rec ht, D. K. Sod-ickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated MRI data,” Magnetic Resonance Med., vol. 79, no. 6, pp. 3055–3071, 2018.

[5] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reechorst, and P. Schniter, “Plug and play methods for magnetic resonance imaging;” IEEE Signal Process. Mag., vol. 37, no. 1, pp. 105–116, 2020.

[6] A. K. Fletcher, M. Sahraee-Ardakan, S. Rangan, and P. Schniter, “Expectation consistent approximate inference: Generalizations and convergence,” in Proc. IEEE Int. Symp. Inform. Thy., 2016, pp. 190–194.

[7] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing,” Proc. Nat. Acad. Sci., vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[8] S. Rangan, P. Schniter, and A. K. Fletcher, “Vector approximate message passing,” IEEE Trans. Inform. Theory, pp. 6664–6684, 2019.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

[10] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction;” in Proc. IEEE Global Conf. Signal Info. Process., 2013, pp. 945–948.

[11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080–2095, 2007.

[12] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, 2017.

[13] S. Ono, “Primal-dual plug-and-play image restoration,” IEEE Signal Process. Lett., vol. 24, no. 8, pp. 1108–1112, 2017.

[14] U. Kamilov, H. Mansour, and B. Wohlberg, “A plug-and-play priors approach for solving nonlinear imaging inverse problems,” IEEE Signal Process. Lett., vol. 24, no. 12, pp. 1872–1876, May 2017.

[15] R. Berthier, A. Montanari, and P.-M. Nguyen, “State evolution for approximate message passing with non-separable functions,” Inform. Inference, 2019.

[16] 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.