Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Plug in estimation in high dimensional linear inverse problems a rigorous analysis*

Alyson K Fletcher[1], **Parthe Pandit**[2], **Sundeep Rangan**[3], **Subrata Sarkar**[4] **and Philip Schniter**[4,5]

[1] Department of Statistics, UC Los Angeles, CA, United States of America
[2] Department of ECE, UC Los Angeles, CA, United States of America
[3] Department of ECE, NYU, New York, NY, United States of America
[4] Department of ECE, The Ohio State University, Columbus, OH, United States of America

E-mail: akfletcher@ucla.edu, parthepandit@ucla.edu, srangan@nyu.edu, sarkar.51@osu.edu and schniter.1@osu.edu

Received 15 May 2019
Accepted for publication 6 June 2019
Published 20 December 2019

Online at stacks.iop.org/JSTAT/2019/124021
https://doi.org/10.1088/1742-5468/ab321a

Abstract. Estimating a vector $\mathbf{x}$ from noisy linear measurements $\mathbf{Ax}+\mathbf{w}$ often requires use of prior knowledge or structural constraints on $\mathbf{x}$ for accurate reconstruction. Several recent works have considered combining linear least-squares estimation with a generic or 'plug-in' denoiser function that can be designed in a modular manner based on the prior knowledge about $\mathbf{x}$. While these methods have shown excellent performance, it has been difficult to obtain rigorous performance guarantees. This work considers plug-in denoising combined with the recently-developed vector approximate message passing (VAMP) algorithm, which is itself derived via expectation propagation techniques. It shown that the mean squared error of this 'plug-and-play' VAMP can be exactly predicted for high-dimensional right-rotationally invariant random $\mathbf{A}$ and Lipschitz denoisers. The method is demonstrated on applications in image recovery and parametric bilinear estimation.

Keywords: machine learning

S Supplementary material for this article is available online

* This article is an updated version of: Fletcher A K, Pandit P, Rangan S, Sarkar S and Schniter P 2018 Plug-in estimation in high-dimensional linear inverse problems: a rigorous analysis *Advances in Neural Information Processing Systems 31* ed S Bengio, H Wallach, H Larochelle, K Grauman, N Cesa-Bianchi and R Garnett (Red Hook, NY: Curran Associates, Inc) pp 7440–7449.
[5] Author to whom any correspondence should be addressed.

1742-5468/19/124021+15$33.00

Contents

1. Introduction

The estimation of an unknown vector $\mathbf{x}^0 \in \mathbb{R}^N$ from noisy linear measurements $\mathbf{y}$ of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \mathbf{w} \in \mathbb{R}^M, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a known transform and $\mathbf{w}$ is disturbance, arises in a wide-range of learning and inverse problems. In many high-dimensional situations, such as when the measurements are fewer than the unknown parameters (i.e. $M \ll N$), it is essential to incorporate known structure on $\mathbf{x}^0$ in the estimation process. A fundamental challenge is how to perform structured estimation of $\mathbf{x}^0$ while maintaining computational efficiency and a tractable analysis.

Approximate message passing (AMP), originally proposed in [1], refers to a powerful class of algorithms that can be applied to reconstruction of $\mathbf{x}^0$ from (1) that can easily incorporate a wide class of statistical priors. In this work, we restrict our attention to $\mathbf{w} \sim \mathcal{N}(0, \gamma_w^{-1}\mathbf{I})$, noting that AMP was extended to non-Gaussian measurements in

[2–4]. AMP is computationally efficient, in that it generates a sequence of estimates $\{\widehat{\mathbf{x}}_k\}_{k=0}^{\infty}$ by iterating the steps

$$\widehat{\mathbf{x}}_k = \mathbf{g}(\mathbf{r}_k, \gamma_k) \tag{2a}$$

$$\mathbf{v}_k = \mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}_k + \frac{N}{M}\langle \nabla \mathbf{g}(\mathbf{r}_k, \gamma_k)\rangle \mathbf{v}_{k-1} \tag{2b}$$

$$\mathbf{r}_{k+1} = \widehat{\mathbf{x}}_k + \mathbf{A}^{\mathsf{T}}\mathbf{v}_k, \quad \gamma_{k+1} = M/\|\mathbf{v}_k\|^2, \tag{2c}$$

initialized with $\mathbf{r}_0 = \mathbf{A}^{\mathsf{T}}\mathbf{y}$, $\gamma_0 = M/\|\mathbf{y}\|^2$, $\mathbf{v}_{-1} = 0$, and assuming $\mathbf{A}$ is scaled so that $\|\mathbf{A}\|_F^2 \approx N$. In (2), $\mathbf{g} : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ is an estimation function chosen based on prior knowledge about $\mathbf{x}^0$, and $\langle \nabla \mathbf{g}(\mathbf{r}, \gamma)\rangle := \frac{1}{N}\sum_{n=1}^{N} \frac{\partial g_n(\mathbf{r},\gamma)}{\partial r_n}$ denotes the divergence of $\mathbf{g}(\mathbf{r}, \gamma)$. For example, if $\mathbf{x}^0$ is known to be sparse, then it is common to choose $\mathbf{g}(\cdot)$ to be the componentwise soft-thresholding function, in which case AMP iteratively solves the LASSO [5] problem.

Importantly, for large, i.i.d., sub-Gaussian random matrices $\mathbf{A}$ and Lipschitz denoisers $\mathbf{g}(\cdot)$, the performance of AMP can be exactly predicted by a scalar *state evolution* (SE), which also provides testable conditions for optimality [6–8]. The initial work [6, 7] focused on the case where $\mathbf{g}(\cdot)$ is a separable function with identical components (i.e. $[\mathbf{g}(\mathbf{r}, \gamma)]_n = g(r_n, \gamma)\ \forall n$), while the later work [8] allowed non-separable $\mathbf{g}(\cdot)$. Interestingly, these SE analyses establish the fact that

$$\mathbf{r}_k = \mathbf{x}^0 + \mathcal{N}(0, \mathbf{I}/\gamma_k), \tag{3}$$

leading to the important interpretation that $\mathbf{g}(\cdot)$ acts as a *denoiser*. This interpretation provides guidance on how to choose $\mathbf{g}(\cdot)$. For example, if $\mathbf{x}$ is i.i.d. with a known prior, then (3) suggests to choose a separable $\mathbf{g}(\cdot)$ composed of minimum mean-squared error (MMSE) scalar denoisers $g(r_n, \gamma) = \mathbb{E}(x_n | r_n = x_n + \mathcal{N}(0, 1/\gamma))$. In this case, [6, 7] established that, whenever the SE has a unique fixed point, the estimates $\widehat{\mathbf{x}}_k$ generated by AMP converge to the Bayes optimal estimate of $\mathbf{x}^0$ from $\mathbf{y}$. As another example, if $\mathbf{x}$ is a natural image, for which an analytical prior is lacking, then (3) suggests to choose $\mathbf{g}(\cdot)$ as a sophisticated image-denoising algorithm like BM3D [9] or DnCNN [10], as proposed in [11]. Many other examples of structured estimators $\mathbf{g}(\cdot)$ can be considered; we refer the reader to [8] and section 5. Prior to [8], AMP SE results were established for special cases of $\mathbf{g}(\cdot)$ in [12, 13]. Plug-in denoisers have been combined in related algorithms [14–16].

An important limitation of AMP's SE is that it holds only for large, i.i.d., sub-Gaussian $\mathbf{A}$. AMP itself often fails to converge with small deviations from i.i.d. sub-Gaussian $\mathbf{A}$, such as when $\mathbf{A}$ is mildly ill-conditioned or non-zero-mean [4, 17, 18]. Recently, a robust alternative to AMP called *vector AMP* (VAMP) was proposed and analyzed in [19], based closely on expectation propagation [20]—see also [21–23]. There it was established that, if $\mathbf{A}$ is a large right-rotationally invariant random matrix and $\mathbf{g}(\cdot)$ is a separable Lipschitz denoiser, then VAMP's performance can be exactly predicted by a scalar SE, which also provides testable conditions for optimality. Importantly, VAMP applies to arbitrarily conditioned matrices $\mathbf{A}$, which is a significant benefit over AMP, since it is known that ill-conditioning is one of AMP's main failure mechanisms [4, 17, 18].

J. Stat. Mech. (2019) 124021

Algorithm 1. Vector AMP (LMMSE form).

Require: LMMSE estimator $\mathbf{g}_2(\cdot, \gamma_{2k})$ from (4), denoiser $\mathbf{g}_1(\cdot, \gamma_{1k})$, and number of iterations K_{it}.

Select initial $\mathbf{r}_{10}$ and $\gamma_{10} \geqslant 0$.
for $k = 0, 1, \ldots, K_{\text{it}}$ **do**
    // Denoising
    $\widehat{\mathbf{x}}_{1k} = \mathbf{g}_1(\mathbf{r}_{1k}, \gamma_{1k})$
    $\alpha_{1k} = \langle \nabla \mathbf{g}_1(\mathbf{r}_{1k}, \gamma_{1k}) \rangle$
    $\eta_{1k} = \gamma_{1k}/\alpha_{1k}$, $\gamma_{2k} = \eta_{1k} - \gamma_{1k}$
    $\mathbf{r}_{2k} = (\eta_{1k}\widehat{\mathbf{x}}_{1k} - \gamma_{1k}\mathbf{r}_{1k})/\gamma_{2k}$

    // LMMSE estimation
    $\widehat{\mathbf{x}}_{2k} = \mathbf{g}_2(\mathbf{r}_{2k}, \gamma_{2k})$
    $\alpha_{2k} = \langle \nabla \mathbf{g}_2(\mathbf{r}_{2k}, \gamma_{2k}) \rangle$
    $\eta_{2k} = \gamma_{2k}/\alpha_{2k}$, $\gamma_{1,k+1} = \eta_{2k} - \gamma_{2k}$
    $\mathbf{r}_{1,k+1} = (\eta_{2k}\widehat{\mathbf{x}}_{2k} - \gamma_{2k}\mathbf{r}_{2k})/\gamma_{1,k+1}$
end for
Return $\widehat{\mathbf{x}}_{1K_{\text{it}}}$.

Unfortunately, the SE analyses of VAMP in [24] and its extension in [25] are limited to separable denoisers. This limitation prevents a full understanding of VAMP's behavior when used with non-separable denoisers, such as state-of-the-art image-denoising methods as recently suggested in [26]. The main contribution of this work is to show that the SE analysis of VAMP can be extended to a large class of non-separable denoisers that are Lipschitz continuous and satisfy a certain convergence property. The conditions are similar to those used in the analysis of AMP with non-separable denoisers in [8]. We show that there are several interesting non-separable denoisers that satisfy these conditions, including group-structured and convolutional neural network based denoisers.

An extended version with all proofs and other details are provided in [27].

2. Review of vector AMP

The steps of VAMP algorithm of [19] are shown in algorithm 1. Each iteration has two parts: a denoiser step and a linear MMSE (LMMSE) step. These are characterized by *estimation functions* $\mathbf{g}_1(\cdot)$ and $\mathbf{g}_2(\cdot)$ producing estimates $\widehat{\mathbf{x}}_{1k}$ and $\widehat{\mathbf{x}}_{2k}$. The estimation functions take inputs $\mathbf{r}_{1k}$ and $\mathbf{r}_{2k}$ that we call *partial estimates*. The LMMSE estimation function is given by,

$$\mathbf{g}_2(\mathbf{r}_{2k}, \gamma_{2k}) := \left(\gamma_w \mathbf{A}^{\top}\mathbf{A} + \gamma_{2k}\mathbf{I}\right)^{-1}\left(\gamma_w \mathbf{A}^{\top}\mathbf{y} + \gamma_{2k}\mathbf{r}_{2k}\right), \tag{4}$$

where $\gamma_w > 0$ is a parameter representing an estimate of the precision (inverse variance) of the noise $\mathbf{w}$ in (1). The estimate $\widehat{\mathbf{x}}_{2k}$ is thus an MMSE estimator, treating the $\mathbf{x}$ as having a Gaussian prior with mean given by the partial estimate $\mathbf{r}_{2k}$. The estimation function $\mathbf{g}_1(\cdot)$ is called the *denoiser* and can be designed identically to the denoiser $\mathbf{g}(\cdot)$ in the AMP iterations (2). In particular, the denoiser is used to incorporate the

J. Stat. Mech. (2019) 124021

structural or prior information on $\mathbf{x}$. As in AMP, in lines 5 and 11, $\langle \nabla \mathbf{g}_i \rangle$ denotes the normalized divergence.

The main result of [24] is that, under suitable conditions, VAMP admits a state evolution (SE) analysis that precisely describes the mean squared error (MSE) of the estimates $\widehat{\mathbf{x}}_{1k}$ and $\widehat{\mathbf{x}}_{2k}$ in a certain large system limit (LSL). Importantly, VAMP's SE analysis applies to arbitrary right rotationally invariant $\mathbf{A}$. This class is considerably larger than the set of sub-Gaussian i.i.d. matrices for which AMP applies. However, the SE analysis in [24] is restricted separable Lipschitz denoisers that can be described as follows: let $g_{1n}(\mathbf{r}_1, \gamma_1)$ be the nth component of the output of $\mathbf{g}_1(\mathbf{r}_1, \gamma_1)$. Then, it is assumed that,

$$\hat{x}_{1n} = g_{1n}(\mathbf{r}_1, \gamma_1) = \phi(r_{1n}, \gamma_1), \tag{5}$$

for some function scalar-output function $\phi(\cdot)$ that does not depend on the component index n. Thus, the estimator is separable in the sense that the nth component of the estimate, $\hat{x}_{1n}$ depends only on the nth component of the input r_{1n} as well as the precision level γ_1. In addition, it is assumed that $\phi(r_1, \gamma_1)$ satisfies a certain Lipschitz condition. The separability assumption precludes the analysis of more general denoisers mentioned in the introduction.

J. Stat. Mech. (2019) 124021

3. Extending the analysis to non-separable denoisers

The main contribution of the paper is to extend the state evolution analysis of VAMP to a class of denoisers that we call *uniformly Lipschitz* and *convergent under Gaussian noise*. This class is significantly larger than separable Lipschitz denoisers used in [24]. To state these conditions precisely, consider a sequence of estimation problems, indexed by a vector dimension N. For each N, suppose there is some 'true' vector $\mathbf{u} = \mathbf{u}(N) \in \mathbb{R}^N$ that we wish to estimate from noisy measurements of the form, $\mathbf{r} = \mathbf{u} + \mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^N$ is Gaussian noise. Let $\hat{\mathbf{u}} = \mathbf{g}(\mathbf{r}, \gamma)$ be some estimator, parameterized by γ.

Definition 1. The sequence of estimators $\mathbf{g}(\cdot)$ are said to be *uniformly Lipschitz continuous* if there exists constants A, B and $C > 0$, such that

$$\|\mathbf{g}(\mathbf{r}_2, \gamma_2) - \mathbf{g}(\mathbf{r}_1, \gamma_1)\| \leqslant (A + B|\gamma_2 - \gamma_1|)\|\mathbf{r}_2 - \mathbf{r}_1\| + C\sqrt{N}|\gamma_2 - \gamma_1|, \tag{6}$$

for any $\mathbf{r}_1, \mathbf{r}_2, \gamma_1, \gamma_2$ and N.

Definition 2. The sequence of random vectors $\mathbf{u}$ and estimators $\mathbf{g}(\cdot)$ are said to be *convergent under Gaussian noise* if the following condition holds: let $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^N$ be two sequences where (z_{1n}, z_{2n}) are i.i.d. with $(z_{1n}, z_{2n}) = \mathcal{N}(0, \mathbf{S})$ for some positive definite covariance $\mathbf{S} \in \mathbb{R}^{2\times 2}$. Then, all the following limits exist almost surely:

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{g}(\mathbf{u}+\mathbf{z}_1, \gamma_1)^{\mathsf{T}}\mathbf{g}(\mathbf{u}+\mathbf{z}_2, \gamma_2), \quad \lim_{N\to\infty} \frac{1}{N}\mathbf{g}(\mathbf{u}+\mathbf{z}_1, \gamma_1)^{\mathsf{T}}\mathbf{u}, \tag{7a}$$

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{u}^{\mathsf{T}}\mathbf{z}_1, \quad \lim_{N\to\infty} \frac{1}{N}\|\mathbf{u}\|^2 \tag{7b}$$

$$\lim_{N\to\infty} \langle \nabla \mathbf{g}(\mathbf{u}+\mathbf{z}_1,\gamma_1)\rangle = \frac{1}{NS_{12}}\mathbf{g}(\mathbf{u}+\mathbf{z}_1,\gamma_1)^{\mathsf{T}}\mathbf{z}_2, \tag{7c}$$

for all γ_1, γ_2 and covariance matrices $\mathbf{S}$. Moreover, the values of the limits are continuous in $\mathbf{S}$, γ_1 and γ_2.

With these definitions, we make the following key assumption on the denoiser.

Assumption 1. *For each N, suppose that we have a 'true' random vector $\mathbf{x}^0 \in \mathbb{R}^N$ and a denoiser $\mathbf{g}_1(\mathbf{r}_1, \gamma_1)$ acting on signals $\mathbf{r}_1 \in \mathbb{R}^N$. Following definition 1, we assume the sequence of denoiser functions indexed by N, is uniformly Lipschitz continuous. In addition, the sequence of true vectors $\mathbf{x}^0$ and denoiser functions are convergent under Gaussian noise following definition 2.*

The first part of assumption 1 is relatively standard: Lipschitz and uniform Lipschitz continuity of the denoiser is assumed several AMP-type analyses including [6, 24, 28] What is new is the assumption in definition 2. This assumption relates to the behavior of the denoiser $\mathbf{g}_1(\mathbf{r}_1, \gamma_1)$ in the case when the input is of the form, $\mathbf{r}_1 = \mathbf{x}^0 + \mathbf{z}$. That is, the input is the true signal with a Gaussian noise perturbation. In this setting, we will be requiring that certain correlations converge. Before continuing our analysis, we briefly show that separable denoisers as well as several interesting non-separable denoisers satisfy these conditions.

3.1. Separable denoisers

We first show that the class of denoisers satisfying assumption 1 includes the separable Lipschitz denoisers studied in most AMP analyses such as [6]. Specifically, suppose that the true vector $\mathbf{x}^0$ has i.i.d. components with bounded second moments and the denoiser $\mathbf{g}_1(\cdot)$ is separable in that it is of the form (5). Under a certain uniform Lipschitz condition, it is shown in the extended version of this paper [27] that the denoiser satisfies assumption 1.

3.2. Group-based denoisers

As a first non-separable example, let us suppose that the vector $\mathbf{x}^0$ can be represented as an $L \times K$ matrix. Let $\mathbf{x}^0_\ell \in \mathbb{R}^K$ denote the ℓth row and assume that the rows are i.i.d. Each row can represent a *group*. Suppose that the denoiser $\mathbf{g}_1(\cdot)$ is *groupwise separable*. That is, if we denote by $\mathbf{g}_{1\ell}(\mathbf{r}, \ell)$ the ℓth row of the output of the denoiser, we assume that

$$\mathbf{g}_{1\ell}(\mathbf{r}, \gamma) = \phi(\mathbf{r}_\ell, \gamma) \in \mathbb{R}^K, \tag{8}$$

for a vector-valued function $\phi(\cdot)$ that is the same for all rows. Thus, the ℓth row output $\mathbf{g}_\ell(\cdot)$ depends only on the ℓth row input. Such groupwise denoisers have been used in AMP and EP-type methods for group LASSO and other structured estimation problems [29–31]. Now, consider the limit where the group size K is fixed, and the number of groups $L \to \infty$. Then, under suitable Lipschitz continuity conditions, the extended version of this paper [27] shows that groupwise separable denoiser also satisfies assumption 1.

3.3. Convolutional denoisers

As another non-separable denoiser, suppose that, for each N, $\mathbf{x}^0$ is an N sample segment of a stationary, ergodic process with bounded second moments. Suppose that the denoiser is given by a linear convolution,

$$\mathbf{g}_1(\mathbf{r}_1) := T_N(\mathbf{h} * \mathbf{r}_1), \tag{9}$$

where $\mathbf{h}$ is a finite length filter and $T_N(\cdot)$ truncates the signal to its first N samples. For simplicity, we assume there is no dependence on γ_1. Convolutional denoising arises in many standard linear estimation operations on wide sense stationary processes such as Weiner filtering and smoothing [32]. If we assume that $\mathbf{h}$ remains constant and $N \to \infty$, the extended version of this paper [27] shows that the sequence of random vectors $\mathbf{x}^0$ and convolutional denoisers $\mathbf{g}_1(\cdot)$ satisfies assumption 1.

3.4. Convolutional neural networks

In recent years, there has been considerable interest in using trained deep convolutional neural networks for image denoising [33, 34]. As a simple model for such a denoiser, suppose that the denoiser is a composition of maps,

$$\mathbf{g}_1(\mathbf{r}_1) = (F_L \circ F_{L-1} \circ \cdots \circ F_1)(\mathbf{r}_1), \tag{10}$$

where $F_\ell(\cdot)$ is a sequence of layer maps where each layer is either a multi-channel convolutional operator or Lipschitz separable activation function, such as sigmoid or ReLU. Under mild assumptions on the maps, it is shown in the extended version of this paper [27] that the estimator sequence $\mathbf{g}_1(\cdot)$ can also satisfy assumption 1.

3.5. Singular-value thresholding (SVT) denoiser

Consider the estimation of a low-rank matrix $\mathbf{X}^0$ from linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{X}^0)$, where $\mathcal{A}$ is some linear operator [35]. Writing the SVD of $\mathbf{R}$ as $\mathbf{R} = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$, the SVT denoiser is defined as

$$\mathbf{g}_1(\mathbf{R}, \gamma) := \sum_i (\sigma_i - \gamma)_+ \mathbf{u}_i \mathbf{v}_i^\top, \tag{11}$$

where $(x)_+ := \max\{0, x\}$. In the extended version of this paper [27], we show that $\mathbf{g}_1(\cdot)$ satisfies assumption 1.

4. Large system limit analysis

4.1. System model

Our main theoretical contribution is to show that the SE analysis of VAMP in [19] can be extended to the non-separable case. We consider a sequence of problems indexed by the vector dimension N. For each N, we assume that there is a 'true' random vector $\mathbf{x}^0 \in \mathbb{R}^N$ observed through measurements $\mathbf{y} \in \mathbb{R}^M$ of the form in (1) where

J. Stat. Mech. (2019) 124021

$\mathbf{w} \sim \mathcal{N}(0, \gamma_{w0}^{-1}\mathbf{I})$. We use γ_{w0} to denote the 'true' noise precision to distinguish this from the postulated precision, γ_w, used in the LMMSE estimator (4). Without loss of generality (see below), we assume that $M = N$. We assume that $\mathbf{A}$ has an SVD,

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}, \quad \mathbf{S} = \mathrm{diag}(\mathbf{s}), \quad \mathbf{s} = (s_1, \ldots, s_N), \tag{12}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and $\mathbf{S}$ is non-negative and diagonal. The matrix $\mathbf{U}$ is arbitrary, $\mathbf{s}$ is an i.i.d. random vector with components $s_i \in [0, s_{\max}]$ almost surely. Importantly, we assume that $\mathbf{V}$ is Haar distributed, meaning that it is uniform on the $N \times N$ orthogonal matrices. This implies that $\mathbf{A}$ is *right rotationally invariant* meaning that $\mathbf{A} \stackrel{d}{=} \mathbf{A}\mathbf{V}_0$ for any orthogonal matrix $\mathbf{V}_0$. We also assume that $\mathbf{w}$, $\mathbf{x}^0$, $\mathbf{s}$ and $\mathbf{V}$ are all independent. As in [19], we can handle the case of rectangular $\mathbf{V}$ by zero padding $\mathbf{s}$.

These assumptions are similar to those in [19]. The key new assumption is assumption 1. Given such a denoiser and postulated variance γ_w, we run the VAMP algorithm, algorithm 1. We assume that the initial condition is given by,

$$\mathbf{r} = \mathbf{x}^0 + \mathcal{N}(0, \tau_{10}\mathbf{I}), \tag{13}$$

for some initial error variance τ_{10}. In addition, we assume

$$\lim_{N\to\infty} \gamma_{10} = \overline{\gamma}_{10}, \tag{14}$$

almost surely for some $\overline{\gamma}_{10} \geqslant 0$.

Analogous to [24], we define two key functions: *error functions* and *sensitivity functions*. The error functions characterize the MSEs of the denoiser and LMMSE estimator under AWGN measurements. For the denoiser $\mathbf{g}_1(\cdot, \gamma_1)$, we define the error function as

$$\mathcal{E}_1(\gamma_1, \tau_1) := \lim_{N\to\infty} \frac{1}{N}\|\mathbf{g}_1(\mathbf{x}^0 + \mathbf{z}, \gamma_1) - \mathbf{x}^0\|^2, \quad \mathbf{z} \sim \mathcal{N}(0, \tau_1\mathbf{I}), \tag{15}$$

and, for the LMMSE estimator, as

$$\begin{gathered}\mathcal{E}_2(\gamma_2, \tau_2) := \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\|\mathbf{g}_2(\mathbf{r}_2, \gamma_2) - \mathbf{x}^0\|^2, \\ \mathbf{r}_2 = \mathbf{x}^0 + \mathcal{N}(0, \tau_2\mathbf{I}), \quad \mathbf{y} = \mathbf{A}\mathbf{x}^0 + \mathcal{N}(0, \gamma_{w0}^{-1}\mathbf{I}).\end{gathered} \tag{16}$$

The limit (15) exists almost surely due to the assumption of $\mathbf{g}_1(\cdot)$ being convergent under Gaussian noise. Although $\mathcal{E}_2(\gamma_2, \tau_2)$ implicitly depends on the precisions γ_{w0} and γ_w, we omit this dependence to simplify the notation. We also define the *sensitivity functions* as

$$\mathcal{A}_i(\gamma_i, \tau_i) := \lim_{N\to\infty} \langle \nabla\mathbf{g}_i(\mathbf{x}^0 + \mathbf{z}_i, \gamma_i)\rangle, \quad \mathbf{z}_i \sim \mathcal{N}(0, \tau_i\mathbf{I}). \tag{17}$$

The LMMSE error function (16) and sensitivity functions (17) are identical to those in the VAMP analysis [19]. The denoiser error function (15) generalizes the error function in [19] for non-separable denoisers.

4.2. State evolution of VAMP

We now show that the VAMP algorithm with a non-separable denoiser follows the identical state evolution equations as the separable case given in [19]. Define the error vectors,

J. Stat. Mech. (2019) 124021

$$\mathbf{p}_k := \mathbf{r}_{1k} - \mathbf{x}^0, \quad \mathbf{q}_k := \mathbf{V}^{\mathsf{T}}(\mathbf{r}_{2k} - \mathbf{x}^0). \tag{18}$$

Thus, $\mathbf{p}_k$ represents the error between the partial estimate $\mathbf{r}_{1k}$ and the true vector $\mathbf{x}^0$. The error vector $\mathbf{q}_k$ represents the transformed error $\mathbf{r}_{2k} - \mathbf{x}^0$. The SE analysis will show that these errors are asymptotically Gaussian. In addition, the analysis will exactly predict the variance on the partial estimate errors (18) and estimate errors, $\widehat{\mathbf{x}}_i - \mathbf{x}^0$. These variances are computed recursively through what we will call the *state evolution* equations:

$$\overline{\alpha}_{1k} = \mathcal{A}_1(\overline{\gamma}_{1k}, \tau_{1k}), \quad \overline{\eta}_{1k} = \frac{\overline{\gamma}_{1k}}{\overline{\alpha}_{1k}}, \quad \overline{\gamma}_{2k} = \overline{\eta}_{1k} - \overline{\gamma}_{1k} \tag{19a}$$

$$\tau_{2k} = \frac{1}{(1-\overline{\alpha}_{1k})^2}\left[\mathcal{E}_1(\overline{\gamma}_{1k}, \tau_{1k}) - \overline{\alpha}_{1k}^2\tau_{1k}\right], \tag{19b}$$

$$\overline{\alpha}_{2k} = \mathcal{A}_2(\overline{\gamma}_{2k}, \tau_{2k}), \quad \overline{\eta}_{2k} = \frac{\overline{\gamma}_{2k}}{\overline{\alpha}_{2k}}, \quad \overline{\gamma}_{1,k+1} = \overline{\eta}_{2k} - \overline{\gamma}_{2k} \tag{19c}$$

$$\tau_{1,k+1} = \frac{1}{(1-\overline{\alpha}_{2k})^2}\left[\mathcal{E}_2(\overline{\gamma}_{2k}, \tau_{2k}) - \overline{\alpha}_{2k}^2\tau_{2k}\right], \tag{19d}$$

which are initialized with $k=0$, τ_{10} in (13) and $\overline{\gamma}_{10}$ defined from the limit (14). The SE equations in (19) are identical to those in [19] with the new error and sensitivity functions for the non-separable denoisers. We can now state our main result, which is proven in the extended version of this paper [27].

Theorem 1. *Under the above assumptions and definitions, assume that the sequence of true random vectors* $\mathbf{x}^0$ *and denoisers* $\mathbf{g}_1(\mathbf{r}_1, \gamma_1)$ *satisfy assumption 1. Assume additionally that, for all iterations k, the solution* $\overline{\alpha}_{1k}$ *from the SE equations (19) satisfies* $\overline{\alpha}_{1k} \in (0,1)$ *and* $\overline{\gamma}_{ik} > 0$*. Then,*

(a) For any k, the error vectors on the partial estimates, $\mathbf{p}_k$ *and* $\mathbf{q}_k$ *in (18) can be written as,*

$$\mathbf{p}_k = \tilde{\mathbf{p}}_k + O(\frac{1}{\sqrt{N}}), \quad \mathbf{q}_k = \tilde{\mathbf{q}}_k + O(\frac{1}{\sqrt{N}}), \tag{20}$$

where, $\tilde{\mathbf{p}}_k$ *and* $\tilde{\mathbf{q}}_k \in \mathbb{R}^N$ *are each i.i.d. Gaussian random vectors with zero mean and per component variance* τ_{1k} *and* τ_{2k}*, respectively.*

(b) For any fixed iteration $k \geqslant 0$*, and* $i = 1, 2$*, we have, almost surely*

$$\lim_{N\to\infty} \frac{1}{N}\|\widehat{\mathbf{x}}_i - \mathbf{x}^0\|^2 = \frac{1}{\overline{\eta}_{ik}}, \quad \lim_{N\to\infty}(\alpha_{ik}, \eta_{ik}, \gamma_{ik}) = (\overline{\alpha}_{ik}, \overline{\eta}_{ik}, \overline{\gamma}_{ik}). \tag{21}$$

In (20), we have used the notation, that when $\mathbf{u}, \tilde{\mathbf{u}} \in \mathbb{R}^N$ are sequences of random vectors, $\mathbf{u} = \tilde{\mathbf{u}} + O(\frac{1}{\sqrt{N}})$ means $\lim_{N\to\infty}\frac{1}{N}\|\mathbf{u} - \tilde{\mathbf{u}}\|^2 = 0$ almost surely. Part (a) of theorem 1 thus shows that the error vectors $\mathbf{p}_k$ and $\mathbf{q}_k$ in (18) are approximately i.i.d. Gaussian.

J. Stat. Mech. (2019) 124021

The result is a natural extension to the main result on separable denoisers in [19]. Moreover, the variance on the variance on the errors, along with the mean squared error (MSE) of the estimates $\widehat{\mathbf{x}}_{ik}$ can be exactly predicted by the same SE equations as the separable case. The result thus provides an asymptotically exact analysis of VAMP extended to non-separable denoisers.

5. Numerical experiments

5.1. Compressive image recovery

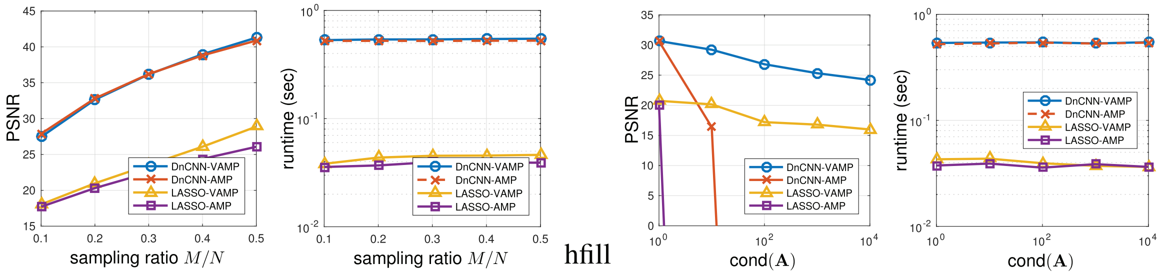
We first consider the problem of compressive image recovery, where the goal is to recover an image $\mathbf{x}^0 \in \mathbb{R}^N$ from measurements $\mathbf{y} \in \mathbb{R}^M$ of the form (1) with $M \ll N$. This problem arises in many imaging applications, such as magnetic resonance imaging, radar imaging, computed tomography, etc, although the details of $\mathbf{A}$ and $\mathbf{x}^0$ change in each case.

One of the most popular approaches to image recovery is to exploit sparsity in the wavelet transform coefficients $\mathbf{c} := \mathbf{\Psi}\mathbf{x}^0$, where $\mathbf{\Psi}$ is a suitable orthonormal wavelet transform. Rewriting (1) as $\mathbf{y} = \mathbf{A}\mathbf{\Psi}\mathbf{c} + \mathbf{w}$, the idea is to first estimate $\mathbf{c}$ from $\mathbf{y}$ (e.g. using LASSO) and then form the image estimate via $\widehat{\mathbf{x}} = \mathbf{\Psi}^\top \hat{\mathbf{c}}$. Although many algorithms exist to solve the LASSO problem, the AMP algorithms are among the fastest (see, e.g. [36, figure 1]). As an alternative to the sparsity-based approach, it was recently suggested in [11] to recover $\mathbf{x}^0$ directly using AMP (2) by choosing the estimation function $\mathbf{g}$ as a sophisticated image-denoising algorithm like BM3D [9] or DnCNN [10].

Figure 1(a) compares the LASSO- and DnCNN-based versions of AMP and VAMP for 128×128 image recovery under well-conditioned $\mathbf{A}$ and no noise. Here, $\mathbf{A} = \mathbf{JPHD}$, where $\mathbf{D}$ is a diagonal matrix with random ± 1 entries, $\mathbf{H}$ is a discrete Hadamard transform (DHT), $\mathbf{P}$ is a random permutation matrix, and $\mathbf{J}$ contains the first M rows of $\mathbf{I}_N$. The results average over the well-known *lena*, *barbara*, *boat*, *house*, and *peppers* images using ten random draws of $\mathbf{A}$ for each. The figure shows that AMP and VAMP have very similar runtimes and PSNRs when $\mathbf{A}$ is well-conditioned, and that the DnCNN approach is about 10 dB more accurate, but 10× as slow, as the LASSO approach. Figure 2 shows the state-evolution prediction of VAMP's PSNR on the *barbara* image at $M/N = 0.5$, averaged over 50 draws of $\mathbf{A}$. The state-evolution accurately predicts the PSNR of VAMP.

To test the robustness to the condition number of $\mathbf{A}$, we repeated the experiment from figure 1(a) using $\mathbf{A} = \mathbf{J}\mathrm{Diag}(\mathbf{s})\mathbf{PHD}$, where $\mathrm{Diag}(\mathbf{s})$ is a diagonal matrix of singular values. The singular values were geometrically spaced, i.e. $s_m/s_{m-1} = \rho\ \forall m$, with ρ chosen to achieve a desired $\mathrm{cond}(\mathbf{A}) := s_1/s_M$. The sampling rate was fixed at $M/N = 0.2$, and the measurements were noiseless, as before. The results, shown in figure 1(b), show that AMP diverged when $\mathrm{cond}(\mathbf{A}) \geqslant 10$, while VAMP exhibited only a mild PSNR degradation due to ill-conditioned $\mathbf{A}$. The original images and example image recoveries are included in the extended version of this paper.

J. Stat. Mech. (2019) 124021



hfill

(a) (b)

Figure 1. Compressive image recovery: PSNR and runtime versus rate M/N and cond($\mathbf{A}$). (a) Average PSNR and runtime with versus M/N with well-conditioned $\mathbf{A}$ and no noise after 12 iterations (b) Average PSNR and runtime versus cond($\mathbf{A}$) at $M/N = 0.2$ and no noise after ten iterations.

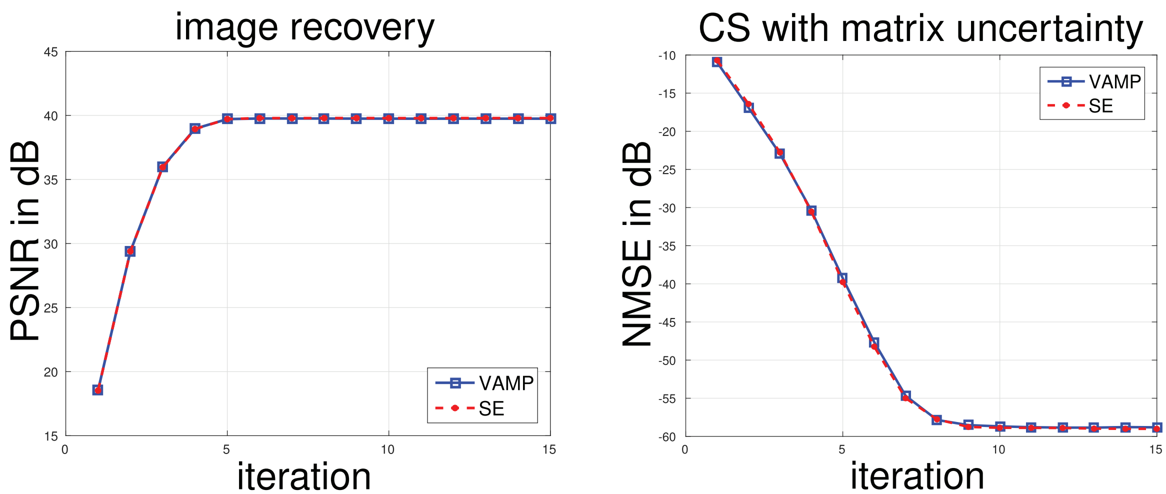


Figure 2. SE prediction & VAMP for image recovery and CS with matrix uncertainty.

5.2. Bilinear estimation via lifting

We now use the structured linear estimation model (1) to tackle problems in *bilinear* estimation through a technique known as 'lifting' [37–40]. In doing so, we are motivated by applications like blind deconvolution [41], self-calibration [39], compressed sensing (CS) with matrix uncertainty [42], and joint channel-symbol estimation [43]. All cases yield measurements $\mathbf{y}$ of the form

$$\mathbf{y} = \left(\textstyle\sum_{l=1}^{L} b_l \mathbf{\Phi}_l\right)\mathbf{c} + \mathbf{w} \in \mathbb{R}^M, \tag{22}$$

where $\{\mathbf{\Phi}_l\}_{l=1}^{L}$ are known, $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}/\gamma_w)$, and the objective is to recover both $\mathbf{b} := [b_1, \ldots, b_L]^\mathsf{T}$ and $\mathbf{c} \in \mathbb{R}^P$. This bilinear problem can be 'lifted' into a linear problem of the form (1) by setting

$$\mathbf{A} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_L \end{bmatrix} \in \mathbb{R}^{M \times LP} \text{ and } \mathbf{x} = \mathrm{vec}(\mathbf{c}\mathbf{b}^\mathsf{T}) \in \mathbb{R}^{LP}, \tag{23}$$

where vec($\mathbf{X}$) vectorizes $\mathbf{X}$ by concatenating its columns. When $\mathbf{b}$ and $\mathbf{c}$ are i.i.d. with known priors, the MMSE denoiser $\mathbf{g}(\mathbf{r}, \gamma) = \mathbb{E}(\mathbf{x}|\mathbf{r} = \mathbf{x} + \mathcal{N}(0, \mathbf{I}/\gamma))$ can be implemented

J. Stat. Mech. (2019) 124021

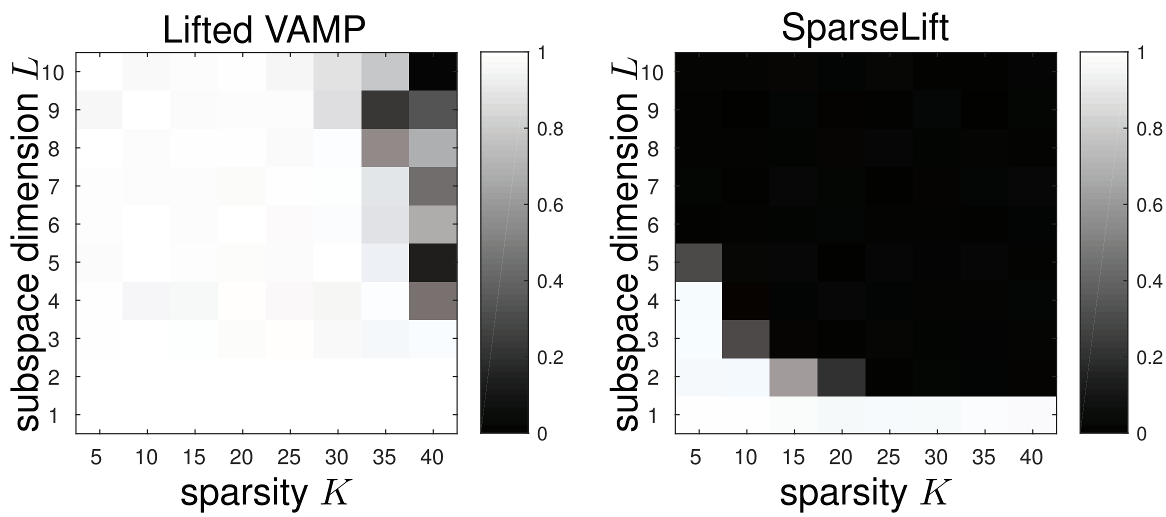


Figure 3. Self-calibration: success rate versus sparsity K and subspace dimension L.

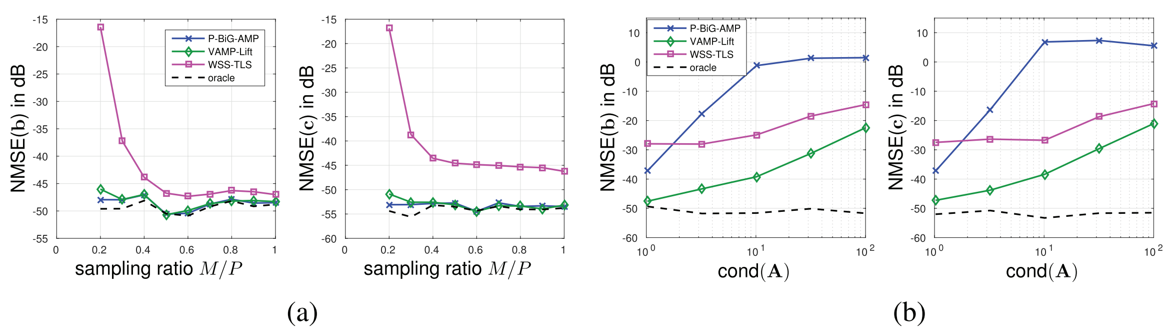


Figure 4. Compressive sensing with matrix uncertainty. (a) NMSE versus M/P with i.i.d. $\mathcal{N}(0,1)$ $\mathbf{A}$. (b) NMSE versus cond($\mathbf{A}$) at $M/P = 0.6$.

near-optimally by the rank-one AMP algorithm from [44] (see also [45–47]), with divergence estimated as in [11].

We first consider *CS with matrix uncertainty* [42], where b_1 is known. For these experiments, we generated the unknown $\{b_l\}_{l=2}^{L}$ as i.i.d. $\mathcal{N}(0,1)$ and the unknown $\mathbf{c} \in \mathbb{R}^P$ as K-sparse with $\mathcal{N}(0,1)$ nonzero entries. Figure 2 shows that the MSE on $\mathbf{x}$ of lifted VAMP is very close to its SE prediction when $K = 12$. We then compared lifted VAMP to PBiGAMP from [48], which applies AMP directly to the (non-lifted) bilinear problem, and to WSS-TLS from [42], which uses non-convex optimization. We also compared to MMSE estimation of $\mathbf{b}$ under oracle knowledge of $\mathbf{c}$, and MMSE estimation of $\mathbf{c}$ under oracle knowledge of support($\mathbf{c}$) and $\mathbf{b}$. For $b_1 = \sqrt{20}$, $L = 11$, $P = 256$, $K = 10$, i.i.d. $\mathcal{N}(0,1)$ matrix $\mathbf{A}$, and SNR $= 40$ dB, figure 4(a) shows the normalized MSE on $\mathbf{b}$ (i.e. $\mathsf{NMSE}(\mathbf{b}) := \mathbb{E}\|\hat{\mathbf{b}} - \mathbf{b}^0\|^2/\mathbb{E}\|\mathbf{b}^0\|^2$) and $\mathbf{c}$ versus sampling ratio M/P. This figure demonstrates that lifted VAMP and PBiGAMP perform close to the oracles and much better than WSS-TLS.

Although lifted VAMP performs similarly to PBiGAMP in figure 4(a), its advantage over PBiGAMP becomes apparent with non-i.i.d. $\mathbf{A}$. For illustration, we repeated the previous experiment, but with $\mathbf{A}$ constructed using the SVD $\mathbf{A} = \mathbf{U}\mathrm{Diag}(\mathbf{s})\mathbf{V}^{\top}$

J. Stat. Mech. (2019) 124021

with Haar distributed $\mathbf{U}$ and $\mathbf{V}$ and geometrically spaced $\mathbf{s}$. Also, to make the problem more difficult, we set $b_1 = 1$. Figure 4(b) shows the normalized MSE on $\mathbf{b}$ and $\mathbf{c}$ versus cond($\mathbf{A}$) at $M/P = 0.6$. There it can be seen that lifted VAMP is much more robust than PBiGAMP to the conditioning of $\mathbf{A}$.

We next consider the *self-calibration* problem [39], where the measurements take the form

$$\mathbf{y} = \mathrm{Diag}(\mathbf{Hb})\boldsymbol{\Psi}\mathbf{c} + \mathbf{w} \in \mathbb{R}^M. \tag{24}$$

Here the matrices $\mathbf{H} \in \mathbb{R}^{M\times L}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{M\times P}$ are known and the objective is to recover the unknown vectors $\mathbf{b}$ and $\mathbf{c}$. Physically, the vector $\mathbf{Hb}$ represents unknown calibration gains that lie in a known subspace, specified by $\mathbf{H}$. Note that (24) is an instance of (22) with $\boldsymbol{\Phi}_l = \mathrm{Diag}(\mathbf{h}_l)\boldsymbol{\Psi}$, where $\mathbf{h}_l$ denotes the lth column of $\mathbf{H}$. Different from 'CS with matrix uncertainty,' all elements in $\mathbf{b}$ are now unknown, and so WSS-TLS [42] cannot be applied. Instead, we compare lifted VAMP to the SparseLift approach from [39], which is based on convex relaxation and has provable guarantees. For our experiment, we generated $\boldsymbol{\Psi}$ and $\mathbf{b} \in \mathbb{R}^L$ as i.i.d. $\mathcal{N}(0, 1)$; $\mathbf{c}$ as K-sparse with $\mathcal{N}(0, 1)$ nonzero entries; $\mathbf{H}$ as randomly chosen columns of a Hadamard matrix; and $\mathbf{w} = 0$. Figure 3 plots the success rate versus L and K, where 'success' is defined as $\mathbb{E}\|\hat{\mathbf{c}}\hat{\mathbf{b}}^\top - \mathbf{c}^0(\mathbf{b}^0)^\top\|_F^2/\mathbb{E}\|\mathbf{c}^0(\mathbf{b}^0)^\top\|_F^2 < -60$ dB. The figure shows that, relative to SparseLift, lifted VAMP gives successful recoveries for a wider range of L and K.

6. Conclusions

We have extended the analysis of the method in [24] to a class of non-separable denoisers. The method provides a computational efficient method for reconstruction where structural information and constraints on the unknown vector can be incorporated in a modular manner. Importantly, the method admits a rigorous analysis that can provide precise predictions on the performance in high-dimensional random settings.

Acknowledgments

A K Fletcher and P Pandit were supported in part by the National Science Foundation under Grants 1738285 and 1738286 and the Office of Naval Research under Grant N00014-15-1-2677. S Rangan was supported in part by the National Science Foundation under Grants 1116589, 1302336, and 1547332, and the industrial affiliates of NYU WIRELESS. P Schniter and S Sarkar were supported in part by the National Science Foundation under Grant CCF-1716388.

References

[1] Donoho D L, Maleki A and Montanari A 2009 Message-passing algorithms for compressed sensing *Proc. Natl Acad. Sci.* **106** 18914–9

[2] Rangan S 2011 Generalized approximate message passing for estimation with random linear mixing *Proc. IEEE ISIT* pp 2174–8

[3] Rangan S, Schniter P, Riegler E, Fletcher A and Cevher V 2013 Fixed points of generalized approximate message passing with arbitrary matrices *Proc. IEEE ISIT* pp 664–8

[4] Rangan S, Schniter P and Fletcher A K 2014 On the convergence of approximate message passing with arbitrary matrices *Proc. IEEE ISIT* pp 236–40

[5] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc.* B **58** 267–88

[6] Bayati M and Montanari A 1996 The dynamics of message passing on dense graphs, with applications to compressed sensing *IEEE Trans. Inf. Theory* **57** 764–85

[7] Javanmard A and Montanari A 2013 State evolution for general approximate message passing algorithms, with applications to spatial coupling *Inf. Inference* **2** 115–44

[8] Berthier R, Montanari A and Nguyen P-M 2017 State evolution for approximate message passing with non-separable functions (arXiv:1708.03950)

[9] Dabov K, Foi A, Katkovnik V and Egiazarian K 2007 Image denoising by sparse 3-D transform-domain collaborative filtering *IEEE Trans. Image Process.* **16** 2080–95

[10] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising *IEEE Trans. Image Process.* **26** 3142–55

[11] Metzler C A, Maleki A and Baraniuk R G 2016 From denoising to compressed sensing *IEEE Trans. Inf. Theory* **62** 5117–44

[12] Donoho D, Johnstone I and Montanari A 2013 Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising *IEEE Trans. Inf. Theory* **59** 3396–433

[13] Ma Y, Rush C and Baron D 2017 Analysis of approximate message passing with a class of non-separable denoisers *Proc. ISIT* pp 231–5

[14] Venkatakrishnan S V, Bouman C A and Wohlberg B 2013 Plug-and-play priors for model based reconstruction *Proc. IEEE Global Conf. on Signal and Inf. Processing* pp 945–8

[15] Chen S, Luo C, Deng B, Qin Y, Wang H and Zhuang Z 2017 BM3D vector approximate message passing for radar coded-aperture imaging *PIERS-FALL* pp 2035–8

[16] Wang X and Chan S H 2017 Parameter-free plug-and-play ADMM for image restoration *Proc. IEEE Acoustics, Speech and Signal Processing* (IEEE) pp 1323–7

[17] Caltagirone F, Zdeborová L and Krzakala F 2014 On convergence of approximate message passing *Proc. IEEE ISIT* pp 1812–6

[18] Vila J, Schniter P, Rangan S, Krzakala F and Zdeborová L 2015 Adaptive damping and mean removal for the generalized approximate message passing algorithm *Proc. IEEE ICASSP* pp 2021–5

[19] Rangan S, Schniter P and Fletcher A K 2017 Vector approximate message passing *Proc. IEEE ISIT* pp 1588–92

[20] Opper M and Winther O 2005 Expectation consistent approximate inference *J. Mach. Learn. Res.* **1** 2177–204

[21] Fletcher A K, Sahraee-Ardakan M, Rangan S and Schniter P 2016 Expectation consistent approximate inference: generalizations and convergence *Proc. IEEE ISIT* pp 190–4

[22] Ma J and Ping L 2017 Orthogonal AMP *IEEE Access* **5** 2020–33

[23] Takeuchi K 2017 Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements *Proc. ISIT* pp 501–5

[24] Rangan S, Schniter P and Fletcher A K 2016 Vector approximate message passing (arXiv:1610.03082)

[25] Fletcher A K, Sahraee-Ardakan M, Rangan S and Schniter P 2017 Rigorous dynamics and consistent estimation in arbitrarily conditioned linear systems *Proc. NIPS* pp 2542–51

[26] Schniter P, Fletcher A K and Rangan S 2017 Denoising-based vector AMP *Proc. Int. Biomedical and Astronomical Signal Process. Workshop* pp 77

[27] Fletcher A K, Pandit P, Rangan S, Sarkar S and Schniter P 2018 Plug-in estimation in high-dimensional linear inverse problems: a rigorous analysis (arXiv:1806.10466)

[28] Kamilov U S, Rangan S, Fletcher A K and Unser M 2014 Approximate message passing with consistent parameter estimation and applications to sparse learning *IEEE Trans. Inf. Theory* **60** 2969–85

[29] Taeb A, Maleki A, Studer C and Baraniuk R 2013 Maximin analysis of message passing algorithms for recovering block sparse signals (arXiv:1303.2389)

[30] Andersen M R, Winther O and Hansen L K 2014 Bayesian inference for structured spike and slab priors *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran Associates, Inc.) pp 1745–53

[31] Rangan S, Fletcher A K, Goyal V K, Byrne E and Schniter P 2017 Hybrid approximate message passing *IEEE Trans. Signal Process.* **65** 4577–92

[32] Scharf L L and Demeure C 1991 *Statistical signal Processing: Detection, Estimation, and Time Series Analysis* vol 63 (Reading, MA: Addison-Wesley)

[33] Xie J, Xu L and Chen E 2012 Image denoising and inpainting with deep neural networks *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran Associates, Inc.) pp 341–9

[34] Xu L, Ren J S, Liu C and Jia J 2014 Deep convolutional neural network for image deconvolution *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran Associates, Inc.) pp 1790–8

[35] Cai J-F, Candès E J and Shen Z 2010 A singular value thresholding algorithm for matrix completion *SIAM J. Optim.* **20** 1956–82

[36] Borgerding M, Schniter P and Rangan S 2017 AMP-inspired deep networks for sparse linear inverse problems *IEEE Trans. Signal Process.* **65** 4293–308

[37] Candès E J, Strohmer T and Voroninski V 2013 PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming *Commun. Pure Appl. Math.* **66** 1241–74

[38] Ahmed A, Recht B and Romberg J 2014 Blind deconvolution using convex programming *IEEE Trans. Inf. Theory* **60** 1711–32

[39] Ling S and Strohmer T 2015 Self-calibration and biconvex compressive sensing *Inverse Problems* **31** 115002

[40] Davenport M A and Romberg J 2016 An overview of low-rank matrix recovery from incomplete observations *IEEE J. Sel. Top. Signal Process.* **10** 608–22

[41] Haykin S S (ed) 1994 *Blind Deconvolution* (Upper Saddle River, NJ: Prentice-Hall)

[42] Zhu H, Leus G and Giannakis G B 2011 Sparsity-cognizant total least-squares for perturbed compressive sampling *IEEE Trans. Signal Process.* **59** 2002–16

[43] Sun P, Wang Z and Schniter P 2018 Joint channel-estimation and equalization of single-carrier systems via bilinear AMP *IEEE Trans. Signal Process.* **66** 2772–85

[44] Rangan S and Fletcher A K 2012 Iterative estimation of constrained rank-one matrices in noise *Proc. IEEE ISIT* (Cambridge, MA) pp 1246–50

[45] Deshpande Y and Montanari A 2014 Information-theoretically optimal sparse PCA *Proc. ISIT* pp 2197–201

[46] Matsushita R and Tanaka T 2013 Low-rank matrix reconstruction, clustering via approximate message passing and *Proc. NIPS* pp 917–25

[47] Lesieur T, Krzakala F and Zdeborova L 2015 Phase transitions in sparse PCA *Proc. IEEE ISIT* pp 1635–9

[48] Parker J and Schniter P 2016 Parametric bilinear generalized approximate message passing *IEEE J. Sel. Top. Signal Process.* **10** 795–808

J. Stat. Mech. (2019) 124021