Bayes-Optimal Estimation in Generalized Linear Models via Spatial Coupling

Pablo Pascual Cobo
University of Cambridge, UK
pp423@cam.ac.uk

Kuan Hsieh
Luminance Technologies Ltd, UK
kuanhsieh95@gmail.com

Ramji Venkataramanan
University of Cambridge, UK
rv285@cam.ac.uk

***Abstract*—We consider the problem of signal estimation in a generalized linear model (GLM). GLMs cover many canonical problems in statistical estimation including linear regression and phase retrieval. Recent work has precisely characterized the asymptotic minimum mean-squared error (MMSE) for GLMs with i.i.d. Gaussian sensing matrices. However, in many models there is a significant gap between the MMSE and the performance of the best known feasible estimators. In this work we address this gap by considering GLMs defined via *spatially coupled* sensing matrices. We propose an efficient approximate message passing (AMP) algorithm for estimation and prove that with a simple choice of spatially coupled design, the MSE of a carefully tuned AMP estimator approaches the asymptotic MMSE. Numerical results show that for finite signal dimensions, spatially coupled designs can yield substantially lower MSE than i.i.d. Gaussian designs when used with AMP algorithms.**

I. Introduction

Consider a generalized linear model (GLM), where the goal is to estimate a signal $\boldsymbol{x} \in \mathbb{R}^n$ from an observation $\boldsymbol{y} \in \mathbb{R}^m$ obtained as:

$$\boldsymbol{y} = \varphi(\boldsymbol{z}, \boldsymbol{\varepsilon}), \quad \text{where } \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}. \tag{1}$$

Here $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a known sensing matrix, $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ is an unknown noise vector and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a known function applied row-wise to the input. The model (1) covers many widely studied problems in statistical estimation and signal processing, including linear regression [1], [2] ($\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$), phase retrieval [3], [4] ($\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \boldsymbol{\varepsilon}$), and 1-bit compressed sensing [5] ($\boldsymbol{y} = \mathrm{sign}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})$).

In this paper, we consider the high-dimensional setting where m, n are both large with $\lim_{n\to\infty} \frac{m}{n} = \delta$, a positive constant that does not scale with n. The components of the signal $\boldsymbol{x}$ are assumed to be distributed according to a prior P_X, and the components of the noise $\boldsymbol{\varepsilon}$ according to a distribution $P_{\bar{\varepsilon}}$. If the entries of $\boldsymbol{A}$ are also assumed to be drawn from a given distribution, the minimum mean-squared error is:

$$\mathsf{MMSE}_n := \frac{1}{n}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}\|^2\}. \tag{2}$$

Two key questions are: i) What is the limiting behavior of MMSE_n as $m, n \to \infty$ with their ratio fixed at a given δ? ii) Noting that the MMSE estimator $\mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}$ is computationally infeasible for large n, what is the smallest mean-squared error (MSE) achievable by efficient estimators?

The first question has been precisely answered for i.i.d. Gaussian sensing matrices: Barbier et al. [6] proved that under mild technical conditions, MMSE_n converges to a well-defined limit that can be numerically computed. Regarding computationally efficient estimators, a variety of estimators based on convex relaxations, spectral methods, and non-convex methods has been proposed for specific GLMs, such as sparse linear regression [7], [8], phase retrieval [9]–[14] and one-bit compressed sensing [15], [16]. Most of these techniques are generic and are not well-equipped to exploit the signal prior, so their estimation error (MSE) is generally much higher than the MMSE.

Approximate message passing (AMP) is a family of iterative algorithms which can be tailored to take advantage of the signal prior. AMP algorithms were first proposed for estimation in linear models [17]–[21] and have since been applied to a range of statistical estimation problems, including GLMs [22]–[26] and low-rank matrix estimation [27]–[30]. An attractive feature of AMP is that under suitable model assumptions, its estimation performance in the high-dimensional limit is precisely characterized by a succinct deterministic recursion called *state evolution*. Using the state evolution analysis, it has been proved that the AMP is Bayes-optimal for certain problems such as symmetric rank-1 matrix estimation [27], [30], i.e., its MSE converges to the limiting MMSE. However, for many canonical GLMs such as phase retrieval, the MSE of AMP can be substantially higher than the limiting MMSE [6].

In this work, we show that for any GLM that satisfies certain mild conditions, the limiting MMSE can be achieved using a *spatially coupled* sensing matrix and an AMP algorithm for estimation. A spatially coupled Gaussian sensing matrix is composed of blocks with different variances; it consists of zero-mean Gaussian entries that are independent but not identically distributed across different blocks. To ensure a fair comparison with i.i.d. Gaussian matrices and their limiting MMSE, the variances in the spatially coupled sensing matrix $\boldsymbol{A}$ are chosen so that the signal strength $\frac{1}{m}\mathbb{E}\{\|\boldsymbol{A}\boldsymbol{x}\|^2\}$ is the same as for the i.i.d. Gaussian design.

Spatially coupled sensing matrices for compressed sensing were introduced in [31], and Donoho et al. [32] showed that they achieve the Bayes-optimal estimation error in Gaussian linear models. They have also been used for constructing capacity-achieving sparse superposition codes for the AWGN channel, see e.g., [33]–[36]. Our work significantly expands the scope of spatial coupling beyond linear models, and shows how to efficiently achieve the Bayes-optimal error for canonical GLMs such as phase retrieval. We must mention that spatial coupling assumes that there is some flexibility in constructing the sensing matrix, which is a reasonable assumption in many imaging and signal processing applications.

Structure of the paper and main contributions: Section II-B describes how a spatially coupled Gaussian sensing matrix is constructed using a base matrix that specifies the variances in the different blocks of the sensing matrix. In Section II-C, we present the AMP algorithm for estimation in a GLM with a spatially coupled matrix. Our first main result, Theorem 1, gives the state evolution performance guarantees of AMP with a generic spatially coupled matrix. It shows that in each AMP iteration, the joint empirical distribution of the signal and AMP estimate converges to the law of a pair of random variables defined via the signal prior and deterministic state evolution parameters. The state evolution parameters depend on a pair of 'denoising' functions used to define the AMP algorithm. Our second result, Theorem 3, shows that with a simple spatially coupled design and specific choices for denoising functions, AMP achieves the Bayes-optimal error for any GLM. We present numerical results which show that spatially coupled designs give significantly smaller estimation error than i.i.d. Gaussian designs, with AMP used for estimation in both settings.

Notation: We write $[n]$ for the set $\{1, \dots, n\}$. We use boldface for matrices and vectors, and plain font for scalars. Thus, x_i denotes the ith component of the vector $\boldsymbol{x}$.

II. Spatially Coupled GLMs

A. Model assumptions

As $n, m \to \infty$, we assume that $\frac{m}{n} \to \delta$, for some constant $\delta > 0$. Both the signal $\boldsymbol{x}$ and the noise vector $\boldsymbol{\varepsilon}$ are independent of the sensing matrix $\boldsymbol{A}$. As $n \to \infty$, the empirical distributions of the signal and the noise vectors are assumed to converge in Wasserstein distance to well-defined limits. More precisely, write $\nu_n(\boldsymbol{x})$ and $\nu_m(\boldsymbol{\varepsilon})$ for the empirical distributions of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ respectively. Then for some $k \in [2, \infty)$, there exist scalar random variables $X \sim P_X$ and $\bar{\varepsilon} \sim P_{\bar{\varepsilon}}$ with $\mathbb{E}\{|X|^k\}, \mathbb{E}\{|\bar{\varepsilon}|^k\} < \infty$, such that $d_k(\nu_n(\boldsymbol{x}), P_X) \to 0$ and $d_k(\nu_m(\boldsymbol{\varepsilon}), P_{\bar{\varepsilon}}) \to 0$ almost surely, where $d_k(P, Q)$ is the k-Wasserstein distance between distributions P, Q defined on the same Euclidean probability space. We note that this assumption on the empirical distributions of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ is more general than assuming that their entries are i.i.d. according to P_X and $P_{\bar{\varepsilon}}$, respectively (which it includes as a special case).

B. Spatially Coupled Sensing Matrix

Consider the GLM (1) with sensing matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. A spatially coupled (Gaussian) sensing matrix $\boldsymbol{A}$ consists of independent zero-mean normally distributed entries whose variances are specified by a *base matrix* $\boldsymbol{W}$ of dimension $\mathsf{R} \times \mathsf{C}$. The sensing matrix $\boldsymbol{A}$ is obtained by replacing each entry of the base matrix W_{rc} by an $\frac{m}{\mathsf{R}} \times \frac{n}{\mathsf{C}}$ matrix with entries drawn i.i.d. $\sim \mathcal{N}(0, \frac{W_{\mathsf{rc}}}{m/\mathsf{R}})$, for $\mathsf{r} \in [\mathsf{R}]$, $\mathsf{c} \in [\mathsf{C}]$. This is similar to the "graph lifting" procedure for constructing spatially coupled LDPC codes from protographs [37]. See Fig. 1 for an example.

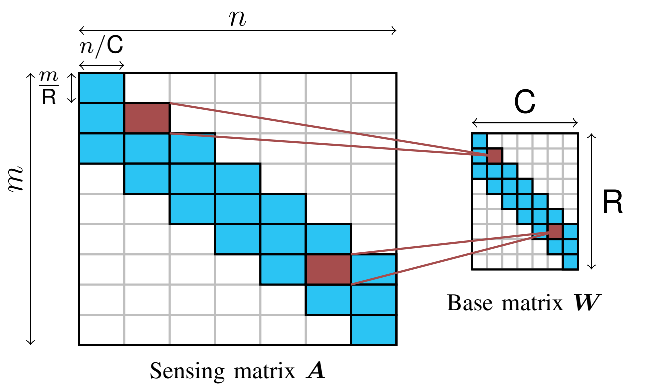

Fig. 1. The entries of $\boldsymbol{A}$ are independent with $A_{ij} \sim \mathcal{N}(0, \frac{1}{m/\mathsf{R}} W_{\mathsf{r}(i),\mathsf{c}(j)})$, where $\boldsymbol{W}$ is the base matrix. Here $\boldsymbol{W}$ is an (ω, Λ) base matrix with $\omega = 3, \Lambda = 7$. The white parts of $\boldsymbol{A}$ and $\boldsymbol{W}$ correspond to zeros.

From the construction, $\boldsymbol{A}$ has independent Gaussian entries

$$A_{ij} \sim \mathcal{N}\left(0, \frac{1}{m/\mathsf{R}} W_{\mathsf{r}(i),\mathsf{c}(j)}\right), \quad \text{for } i \in [m], \; j \in [n]. \tag{3}$$

Here the operators $\mathsf{r}(\cdot) : [m] \to [\mathsf{R}]$ and $\mathsf{c}(\cdot) : [n] \to [\mathsf{C}]$ in (3) map a particular row or column index to its corresponding *row block* or *column block* index in $\boldsymbol{W}$. Each entry of the base matrix corresponds to an $\frac{m}{\mathsf{R}} \times \frac{n}{\mathsf{C}}$ block of the sensing matrix $\boldsymbol{A}$, and each block can be viewed as an (uncoupled) i.i.d. Gaussian sensing matrix with sampling ratio

$$\delta_{\text{in}} := \frac{m/\mathsf{R}}{n/\mathsf{C}} = \frac{\mathsf{C}}{\mathsf{R}} \delta. \tag{4}$$

As in [32], we will assume that entries of the base matrix $\boldsymbol{W}$ are scaled to satisfy:

$$\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}} = 1 \quad \text{for } \mathsf{c} \in [\mathsf{C}], \qquad \kappa_1 \leq \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \leq \kappa_2, \tag{5}$$

for some $\kappa_1, \kappa_2 > 0$. The first assumption ensures that the columns of the sensing matrix $\boldsymbol{A}$ have expected squared norm equal to one, to match the standard assumption for AMP with i.i.d. Gaussian matrices. The second assumption ensures that the variance of each entry of $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$ is bounded above and below for all $i \in [m]$. Indeed, it can be verified that the variance of z_i for i with row block index $\mathsf{r} \in [\mathsf{R}]$ is

$$\mathbb{E}\{Z_{\mathsf{r}}^2\} = \frac{\mathbb{E}\{X^2\}}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \in \left[\frac{\mathbb{E}\{X^2\}}{\delta_{\text{in}}}\kappa_1, \frac{\mathbb{E}\{X^2\}}{\delta_{\text{in}}}\kappa_2\right]. \tag{6}$$

The trivial base matrix with $\mathsf{R} = \mathsf{C} = 1$ (single entry equal to 1) corresponds to the sensing matrix with i.i.d. $\mathcal{N}(0, \frac{1}{m})$ entries. The quantity, $\mathbb{E}\{\|\boldsymbol{z}\|^2\}/m$, which can be interpreted as the measurement strength, is equal to $\mathbb{E}\{X^2\}/\delta$ for both the i.i.d. Gaussian and spatially coupled matrices. Indeed, from (6), we have that with spatial coupling,

$$\frac{\mathbb{E}\{\|\boldsymbol{z}\|^2\}}{m} = \frac{\mathbb{E}\{X^2\}}{\mathsf{R}\,\delta_{\text{in}}} \sum_{\mathsf{r}=1}^{\mathsf{R}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} = \frac{\mathbb{E}\{X^2\}}{\delta}. \tag{7}$$

This ensures that for a given δ, the i.i.d. Gaussian and the spatially coupled setups can be fairly compared.

To prove the Bayes optimality of spatial coupling with AMP, we will use the following base matrix, first used in the context of sparse superposition codes for communication over Gaussian channels [35], [36], and inspired by the construction of protograph based spatially coupled LDPC codes [37].

Definition 2.1: An (ω, Λ) base matrix $\boldsymbol{W}$ is described by two parameters: the coupling width $\omega \geq 1$ and the coupling length $\Lambda \geq 2\omega - 1$. The matrix has $\mathsf{R} = \Lambda + \omega - 1$ rows and $\mathsf{C} = \Lambda$ columns, with the (r, c)th entry of the base matrix, for $\mathsf{r} \in [\mathsf{R}], \mathsf{c} \in [\mathsf{C}]$, given by

$$W_{\mathsf{rc}} = \begin{cases} \frac{1}{\omega} & \text{if } \mathsf{c} \leq \mathsf{r} \leq \mathsf{c} + \omega - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The base matrix in Fig. 1 has parameters $(\omega = 3, \Lambda = 7)$.

With an (ω, Λ) base matrix, from (4) we have $\delta_{\text{in}} = \frac{\Lambda}{\Lambda+\omega-1}\delta$. As $\omega > 1$ in spatially coupled systems, the difference $\delta - \delta_{\text{in}}$ is often referred to as a "rate loss" in the literature of spatially coupled error correcting codes [38]–[40], and becomes negligible when Λ is much larger than ω.

C. Spatially Coupled Generalized AMP

We now describe the algorithm for estimating $\boldsymbol{x}$ from $(\boldsymbol{y}, \boldsymbol{A})$, which we call spatially coupled Generalized AMP (SC-GAMP). For iteration $t \geq 1$, SC-GAMP iteratively computes $\boldsymbol{p}^t \in \mathbb{R}^m$ and $\boldsymbol{q}^t \in \mathbb{R}^n$ via functions $\bar{g}_{\text{in}}(\cdot\,;t) : \mathbb{R}^n \to \mathbb{R}^n$ and $\bar{g}_{\text{out}}(\cdot,\cdot\,;t) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$. Starting from an initialization $\boldsymbol{x}^0 \equiv \bar{g}_{\text{in}}(\boldsymbol{q}(0)\,;0) \in \mathbb{R}^n$ and $\boldsymbol{p}(0) = \boldsymbol{A}\boldsymbol{x}^0$, for $t \geq 0$ we compute:

$$\boldsymbol{q}(t+1) = \bar{g}_{\text{in}}(\boldsymbol{q}(t)\,;t) + \boldsymbol{\alpha^q}(t+1) \odot \boldsymbol{A}^\top \bar{g}_{\text{out}}(\boldsymbol{p}(t), \boldsymbol{y}\,;t), \tag{9}$$

$$\begin{aligned}\boldsymbol{p}(t+1) = \boldsymbol{A}\bar{g}_{\text{in}}(&\boldsymbol{q}(t+1);t+1) \\ &- \boldsymbol{\alpha^p}(t+1) \odot \bar{g}_{\text{out}}(\boldsymbol{p}(t), \boldsymbol{y}\,;t).\end{aligned} \tag{10}$$

Here the vectors $\boldsymbol{\alpha^p}(t+1) \in \mathbb{R}^n$ and $\boldsymbol{\alpha^q}(t+1) \in \mathbb{R}^m$ are defined below via state evolution, and $\odot$ denotes the Hadamard (element-wise) product.

The functions $\bar{g}_{\text{in}}$ and $\bar{g}_{\text{out}}$ are assumed to have a separable block-wise structure. Partition $[m]$ and $[n]$ into R and C equal-sized blocks, respectively, denoted by

$$[n] = \cup_{\mathsf{c}=1}^{\mathsf{C}} \mathcal{J}_{\mathsf{c}}, \quad [m] = \cup_{\mathsf{r}=1}^{\mathsf{R}} \mathcal{I}_{\mathsf{r}}, \tag{11}$$

where

$$\begin{aligned}\mathcal{J}_{\mathsf{c}} &= \{(\mathsf{c}-1)\frac{n}{\mathsf{C}} + 1, \ldots, \mathsf{c}\,\frac{n}{\mathsf{C}}\}, \quad \text{for } \mathsf{c} \in [\mathsf{C}], \\ \mathcal{I}_{\mathsf{r}} &= \{(\mathsf{r}-1)\frac{m}{\mathsf{R}} + 1, \ldots, \mathsf{r}\,\frac{m}{\mathsf{R}}\}, \quad \text{for } \mathsf{r} \in [\mathsf{R}].\end{aligned} \tag{12}$$

Then, for each $t \geq 0$, we assume that there exist functions $g_{\text{in}}(\cdot\,;t+1) : \mathbb{R} \times [\mathsf{C}] \to \mathbb{R}$ and $g_{\text{out}}(\cdot\,;t) : \mathbb{R}^2 \times [\mathsf{R}] \to \mathbb{R}$ such that

$$\begin{aligned}&\bar{g}_{\text{in},j}(\boldsymbol{q}(t+1)\,;t+1) = g_{\text{in}}(q_j(t+1), \mathsf{c}\,;t+1), \quad \text{for } j \in \mathcal{J}_{\mathsf{c}}, \\ &\bar{g}_{\text{out},i}(\boldsymbol{p}(t), \boldsymbol{y}\,;t) = g_{\text{out}}(p_i(t), y_i, \mathsf{r}\,;t), \quad \text{for } i \in \mathcal{I}_{\mathsf{r}}.\end{aligned} \tag{13}$$

State Evolution: We will show in Theorem 1 that under suitable assumptions, for each $t \geq 0$, the empirical distribution of $(\boldsymbol{q}_{\mathsf{c}}(t) - \mu^q_{\mathsf{c}}(t)\,\boldsymbol{x}_{\mathsf{c}})$ converges to a Gaussian $\mathcal{N}(0, \tau^q_{\mathsf{c}}(t))$, for each $\mathsf{c} \in [\mathsf{C}]$. Here $\boldsymbol{q}_{\mathsf{c}}(t), \boldsymbol{x}_{\mathsf{c}} \in \mathbb{R}^{n/\mathsf{C}}$ denote the cth block of $\boldsymbol{q}(t), \boldsymbol{x} \in \mathbb{R}^n$, respectively. Thus the cth block of the function $\bar{g}_{\text{in}}(;t)$ can be viewed as estimating $\boldsymbol{x}_{\mathsf{c}}$ from an observation of the form $\mu^q_{\mathsf{c}}(t)\,\boldsymbol{x}_{\mathsf{c}}$ plus Gaussian noise of variance $\tau^q_{\mathsf{c}}(t)$.

Analogously, the joint empirical distribution of the rows of $(\boldsymbol{z}_{\mathsf{r}}(t), \boldsymbol{p}_{\mathsf{r}}(t)) \in \mathbb{R}^{(m/\mathsf{R}) \times 2}$ converges to a bivariate Gaussian distribution $\mathcal{N}(0, \Lambda_{\mathsf{r}}(t))$, for $\mathsf{r} \in [\mathsf{R}]$. The constants $\mu^q_{\mathsf{c}}(t), \tau^q_{\mathsf{c}}(t) \in \mathbb{R}$, and $\Lambda_{\mathsf{r}}(t) \in \mathbb{R}^{2\times 2}$ are defined via the state evolution recursion below. For $t \geq 1$, given the coefficients $\mu^q_{\mathsf{c}}(t), \tau^q_{\mathsf{c}}(t)$ for $\mathsf{c} \in [\mathsf{C}]$, and $\Lambda_{\mathsf{r}}(t)$ for $\mathsf{r} \in [\mathsf{R}]$, define the random variables $Q_{\mathsf{c}}(t)$ and $P_{\mathsf{r}}(t)$ as follows:

$$\begin{aligned}&Q_{\mathsf{c}}(t) = \mu^q_{\mathsf{c}}(t) X + G^q_{\mathsf{c}}(t), \\ &\qquad\qquad \text{where } X \sim P_X \perp G^q_{\mathsf{c}}(t) \sim \mathcal{N}(0, \tau^q_{\mathsf{c}}(t)), \\ &(P_{\mathsf{r}}(t),\, Z_{\mathsf{r}}) \sim \mathcal{N}(0, \Lambda_{\mathsf{r}}(t)).\end{aligned} \tag{14}$$

Recalling that $\boldsymbol{y} = \varphi(\boldsymbol{z}, \boldsymbol{\varepsilon})$ where $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$, we write $g_{\text{out}}(p, y, \mathsf{r}\,;t) = g_{\text{out}}(p, \varphi(z,\varepsilon), \mathsf{r}\,;t)$ and let $g'_{\text{out}}(p, \varphi(z,\varepsilon), \mathsf{r}\,;t)$ denote the derivative with respect to the first argument and $\partial_z\, g_{\text{out}}(p, \varphi(z,\varepsilon), \mathsf{r}\,;t)$ the derivative with respect to z. Then the coefficients for $(t+1)$ are computed as follows, for $\mathsf{c} \in [\mathsf{C}]$ and $\mathsf{r} \in [\mathsf{R}]$:

$$\tau^q_{\mathsf{c}}(t+1) = \frac{\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}}\, \mathbb{E}\{g_{\text{out}}(P_{\mathsf{r}}(t),\, \varphi(Z_{\mathsf{r}}, W), \mathsf{r}\,;t)^2\}}{\left[\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}}\, \mathbb{E}\{g'_{\text{out}}(P_{\mathsf{r}}(t),\, \varphi(Z_{\mathsf{r}}, W), \mathsf{r}\,;t)\}\right]^2}, \tag{15}$$

$$\mu^q_{\mathsf{c}}(t+1) = \frac{\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}}\, \mathbb{E}\{\partial_z g_{\text{out}}(P_{\mathsf{r}}(t),\, \varphi(Z_{\mathsf{r}}, W), \mathsf{r}\,;t)\}}{-\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}}\, \mathbb{E}\{g'_{\text{out}}(P_{\mathsf{r}}(t),\, \varphi(Z_{\mathsf{r}}, W), \mathsf{r}\,;t)\}}, \tag{16}$$

$$[\Lambda_{\mathsf{r}}(t+1)]_{11} = \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \mathbb{E}\{g_{\text{in}}(Q_{\mathsf{c}}(t+1), \mathsf{c}\,;(t+1))^2\}, \tag{17}$$

$$\begin{aligned}[\Lambda_{\mathsf{r}}(t+1)]_{12} &= [\Lambda_{\mathsf{r}}(t+1)]_{21} \\ &= \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} \mathbb{E}\{X\, g_{\text{in}}(Q_{\mathsf{c}}(t+1), \mathsf{c}\,;(t+1))\},\end{aligned} \tag{18}$$

$$[\Lambda_{\mathsf{r}}(t+1)]_{22} = \mathbb{E}\{Z_{\mathsf{r}}^2\} = \frac{\mathbb{E}\{X^2\}}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}}\,. \tag{19}$$

To initialize the state evolution, we make the following assumption on the SC-GAMP initialization $\boldsymbol{x}^0 \in \mathbb{R}^n$.

(A0) Denoting by $\boldsymbol{x}^0_{\mathsf{c}} \in \mathbb{R}^{n/\mathsf{C}}$ the cth block of $\boldsymbol{x}^0 \in \mathbb{R}^n$, we assume that there exists a symmetric non-negative definite $\Xi_{\mathsf{c}} \in \mathbb{R}^{2\times 2}$ for each $\mathsf{c} \in [\mathsf{C}]$ such that we almost surely have

$$\lim_{n\to\infty} \frac{1}{n/\mathsf{C}} \begin{bmatrix} \langle \boldsymbol{x}^0_{\mathsf{c}}, \boldsymbol{x}^0_{\mathsf{c}} \rangle & \langle \boldsymbol{x}_{\mathsf{c}}, \boldsymbol{x}^0_{\mathsf{c}} \rangle \\ \langle \boldsymbol{x}_{\mathsf{c}}, \boldsymbol{x}^0_{\mathsf{c}} \rangle & \langle \boldsymbol{x}_{\mathsf{c}}, \boldsymbol{x}_{\mathsf{c}} \rangle \end{bmatrix} = \Xi_{\mathsf{c}}. \tag{20}$$

From the assumptions on signal in Section II-A, we have that $[\Xi_{\mathsf{c}}]_{22} = \mathbb{E}\{X^2\}$.

The state evolution iteration is initialized with $\Lambda_{\mathsf{r}}(0)$ for $\mathsf{r} \in [\mathsf{R}]$, whose entries are given by:

$$[\Lambda_{\mathsf{r}}(0)]_{11} = \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} [\Xi_{\mathsf{c}}]_{11},$$

$$[\Lambda_{\mathsf{r}}(0)]_{22} = \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} [\Xi_{\mathsf{c}}]_{22} = \mathbb{E}\{Z_{\mathsf{r}}^2\},$$

$$[\Lambda_{\mathsf{r}}(0)]_{12} = [\Lambda_{\mathsf{r}}(0)]_{21} = \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}} [\Xi_{\mathsf{c}}]_{12}. \quad (21)$$

The entries of coefficient vectors $\boldsymbol{\alpha^p}(t+1) \in \mathbb{R}^m$ and $\boldsymbol{\alpha^q}(t+1) \in \mathbb{R}^n$ in (9)-(10) are defined block-wise as follows. Starting with $\alpha^p_{\mathsf{r}}(0) = 0$ for all r, for $t \geq 0$, we recursively compute the following for $\mathsf{c} \in [\mathsf{C}]$ and $\mathsf{r} \in [\mathsf{R}]$:

$$\alpha^q_{\mathsf{c}}(t+1) = -\left(\sum_{\mathsf{r}=1}^{\mathsf{R}} W_{\mathsf{rc}} \mathbb{E}\{ g'_{\text{out}}(P_{\mathsf{r}}(t), Y_{\mathsf{r}}, \mathsf{r}\,;\, t)\} \right)^{-1}, \quad (22)$$

$$\alpha^p_{\mathsf{r}}(t+1) = \frac{1}{\delta_{\text{in}}} \sum_{\mathsf{c}=1}^{\mathsf{C}} W_{\mathsf{rc}}\, \alpha^q_{\mathsf{c}}(t+1)\, \mathbb{E}\{ g'_{\text{in}}(Q_{\mathsf{c}}(t+1), \mathsf{c}\,;\, t+1)\}. \quad (23)$$

We then set

$$\alpha^p_i(t+1) = \alpha^p_{\mathsf{r}}(t+1), \quad \text{for } i \in \mathcal{I}_{\mathsf{r}},$$
$$\alpha^q_j(t+1) = \alpha^q_{\mathsf{c}}(t+1), \quad \text{for } j \in \mathcal{J}_{\mathsf{c}}. \quad (24)$$

We note that for $(Z_{\mathsf{r}},\, P_{\mathsf{r}}(t)) \sim \mathcal{N}(0, \Lambda_{\mathsf{r}}(t))$, by standard properties of Gaussian random vectors we have

$$(P_{\mathsf{r}}(t),\, Z_{\mathsf{r}}) \stackrel{\text{d}}{=} (P_{\mathsf{r}}(t),\, \mu^p_{\mathsf{r}}(t) P_{\mathsf{r}}(t) + G^p_{\mathsf{r}}(t)), \quad (25)$$

where $G^p_{\mathsf{r}}(t) \sim \mathcal{N}(0, \tau^p_{\mathsf{r}}(t))$ is independent of $P_{\mathsf{r}}(t)$ with

$$\mu^p_{\mathsf{r}}(t) = \frac{[\Lambda_{\mathsf{r}}(t)]_{12}}{[\Lambda_{\mathsf{r}}(t)]_{11}}, \qquad \tau^p_{\mathsf{r}}(t) = [\Lambda_{\mathsf{r}}(t)]_{22} - \frac{[\Lambda_{\mathsf{r}}(t)]_{12}^2}{[\Lambda_{\mathsf{r}}(t)]_{11}}. \quad (26)$$

Performance Characterization of SC-GAMP: For the state evolution result (Theorem 1), we make the following assumption on the functions $g_{\text{in}}, g_{\text{out}}$ in (13) used to define SC-GAMP.

(A1) For $t \geq 0$, the function $g_{\text{in}}(\cdot, \mathsf{c}\,; t+1)$ is Lipschitz on $\mathbb{R}$ for $\mathsf{c} \in [\mathsf{C}]$, and $\tilde{g}_{\mathsf{r},t} : (p, z, \varepsilon) \mapsto g_{\text{out}}(p,\, \varphi(z, \varepsilon),\, \mathsf{r};\, t)$ is Lipschitz on $\mathbb{R}^3$ for $\mathsf{r} \in [\mathsf{R}]$.

The state evolution result is stated in terms of *pseudo-Lipschitz* test functions. A function $\Psi : \mathbb{R}^d \to \mathbb{R}$ is called pseudo-Lipschitz of order k if $|\Psi(x) - \Psi(y)| \leq C(1 + \|x\|_2^{k-1} + \|y\|_2^{k-1})\|x - y\|_2$ for all $x, y \in \mathbb{R}^d$. Pseudo-Lipschitz functions of order $k = 2$ include $\Psi(u) = u^2$ and $\Psi(u, v) = uv$.

Theorem 1: Consider a GLM with a spatially coupled sensing matrix defined via a base matrix $\boldsymbol{W}$ satisfying (5), and the SC-GAMP algorithm in (9)-(10). Assume that the assumptions in II-A and **(A0)**,**(A1)** hold. Let $\Psi : \mathbb{R}^2 \to \mathbb{R}$ and $\Phi : \mathbb{R}^3 \to \mathbb{R}$ be any pseudo-Lipschitz functions of order k, where k is specified in Section II-A. Then, for all $t \geq 1$ and each $\mathsf{r} \in [\mathsf{R}]$ and $\mathsf{c} \in [\mathsf{C}]$, we almost surely have:

$$\lim_{n \to \infty} \frac{1}{n/\mathsf{C}} \sum_{j \in \mathcal{J}_{\mathsf{c}}} \Psi(q_j(t)\,, x_j) = \mathbb{E}\{\Psi(Q_{\mathsf{c}}(t)\,, X)\}, \quad (27)$$

$$\lim_{m \to \infty} \frac{1}{m/\mathsf{R}} \sum_{i \in \mathcal{I}_{\mathsf{r}}} \Phi(p_i(t)\,, z_i\,, \varepsilon_i) = \mathbb{E}\{\Phi(P_{\mathsf{r}}(t)\,, Z_{\mathsf{r}},\, \bar{\varepsilon})\}. \quad (28)$$

An outline of the proof is given in the long version of this paper [41]. The main idea is to show, via a suitable change of variables, that the spatially coupled GAMP in (9)-(10) is an instance of an abstract AMP recursion with matrix-valued iterates for i.i.d. Gaussian matrices. A state evolution result for this abstract AMP recursion was established in [42]. This result is translated via the change of variables to establish the state evolution result in Theorem 1.

Theorem 1 allows us to compute the performance measures such as asymptotic MSE of SC-GAMP. Indeed, in (27) taking $\Psi(q, x) = (x - g_{\text{in}}(q, \mathsf{c}; t))^2$ for $\mathsf{c} \in [\mathsf{C}]$, we obtain:

$$\lim_{n \to \infty} \frac{\|\boldsymbol{x} - \bar{g}_{\text{in}}(\boldsymbol{q}(t))\|^2}{n} = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} \mathbb{E}\{(X - g_{\text{in}}(Q_{\mathsf{c}}(t), \mathsf{c}; t))^2\} \quad \text{a.s.} \quad (29)$$

III. MMSE and Fixed Points of Bayes AMP

The limiting value of MMSE_n in (2) for an i.i.d. Gaussian sensing matrix is characterized in terms of a *potential function* [6].

Definition 3.1 (Potential function): For $x \in [0, \text{Var}(X)]$, let

$$\upsilon(x; \delta) := 1 - \frac{\delta}{x} \mathbb{E}_{P,Y} \Big\{ \text{Var}\Big(Z \Big| P, Y;\, \frac{x}{\delta} \Big) \Big\}, \quad (30)$$

where $Y = \varphi(Z, \bar{\varepsilon})$ with $\varepsilon \sim P_{\bar{\varepsilon}} \perp Z$, and $\text{Var}(Z|P, Y;\, \frac{x}{\delta})$ denotes the conditional variance computed with $(P, Z) \sim \mathcal{N}(0, \Lambda)$ where

$$\Lambda = \frac{1}{\delta} \begin{bmatrix} \mathbb{E}\{X^2\} - x & \mathbb{E}\{X^2\} - x \\ \mathbb{E}\{X^2\} - x & \mathbb{E}\{X^2\} \end{bmatrix}. \quad (31)$$

Then the scalar potential function $U(x; \delta)$ is defined as

$$U(x; \delta) := -\delta \upsilon(x; \delta) + \int_0^x \frac{\delta}{z} \upsilon(z; \delta)\, dz + 2I\Big(X\,;\, \sqrt{(\delta/x)\upsilon(x;\delta)}\, X + G \Big), \quad (32)$$

for $x \in [0, \text{Var}(X)]$, where the mutual information is computed with $X \sim P_X \perp G \sim \mathcal{N}(0, 1)$.

Theorem 2 (MMSE for i.i.d. Gaussian design [6]): Consider a GLM with sensing matrix entries i.i.d. $\sim \mathcal{N}(0, \frac{1}{m})$. Suppose that the components of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ are i.i.d. according to P_X and $P_{\bar{\varepsilon}}$, respectively, and $\frac{m}{n} \to \delta$ as $n \to \infty$. Furthermore, assume that $U(x; \delta)$ has a unique minimum. Then:

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A},\, \boldsymbol{y}\}\|^2\} = \underset{x \in [0, \text{Var}(X)]}{\arg\min}\; U(x; \delta). \quad (33)$$

We mention that there are a few additional technical conditions required for Theorem 2; these are omitted due to space constraints and can be found in [6, Sec. 4].

Bayes SC-GAMP: Consider SC-GAMP with the choice

$$g^*_{\text{in}}(q, \mathsf{c}\,;\, t) = \mathbb{E}\{X \mid Q_{\mathsf{c}}(t) = q\}, \quad \text{for } \mathsf{c} \in [\mathsf{C}], \tag{34}$$

$$g^*_{\text{out}}(p, y, \mathsf{r}\,;\, t) = \frac{1}{\tau^p_{\mathsf{r}}(t)}(\mathbb{E}\{Z_{\mathsf{r}} \mid P_{\mathsf{r}}(t) = p,\, Y_{\mathsf{r}} = y\} \,-\, p), \quad \text{for } \mathsf{r} \in [\mathsf{R}]. \tag{35}$$

For the uncoupled case, i.e., $\mathsf{R} = \mathsf{C} = 1$, these choices maximize the correlations $\frac{(\mathbb{E}\{P(t)Z\})^2}{\mathbb{E}\{P(t)^2\}\mathbb{E}\{Z^2\}}$ and $\frac{(\mathbb{E}\{XQ(t)\})^2}{\mathbb{E}\{X^2\}\mathbb{E}\{Q(t)^2\}}$ for each $t \geq 1$ [43, Sec. 4.2]. We therefore refer to the algorithm with this choice of functions as Bayes SC-GAMP.

Our second result shows that with an (ω, Λ) base matrix (defined in Definition 2.1), the MSE of SC-GAMP after a large number of iterations is bounded in terms of the minimizer of $U(x; \delta_{\text{in}})$, where $\delta_{\text{in}} = \delta\frac{\Lambda}{\Lambda+\omega-1}$. Moreover, with an i.i.d. Gaussian sensing matrix, the MSE of GAMP is characterized by the largest stationary point of $U(x; \delta)$ in $x \in [0, \text{Var}(X)]$.

Theorem 3: Consider a GLM with the assumptions in Section II-A, and estimation via Bayes SC-GAMP initialized with $\boldsymbol{x}^0 = \mathbb{E}\{X\}\mathbf{1}_n$. Also assume that **(A1)** is satisfied by the functions g^*_{in} and g^*_{out}. Then:

1) With an (ω, Λ) base matrix, for any $\gamma > 0$ there exist $\omega_0 < \infty$ and $t_0 < \infty$ such that for all $\omega > \omega_0$ and $t > t_0$, the asymptotic MSE of Bayes SC-GAMP almost surely satisfies:

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} \,-\, \bar{g}^*_{\text{in}}(\boldsymbol{q}(t);\, t)\|^2 \leq \left(\max\left(\underset{x\in[0,\text{Var}(X)]}{\arg\min}\; U(x; \delta_{\text{in}})\right) + \gamma\right)\frac{\Lambda+\omega}{\Lambda}. \tag{36}$$

2) With an i.i.d. Gaussian sensing matrix (i.e., 1×1 base matrix with $W_{11} = 1$), the asymptotic MSE of Bayes GAMP converges almost surely as:

$$\lim_{t\to\infty}\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} \,-\, \bar{g}^*_{\text{in}}(\boldsymbol{q}(t);\, t)\|^2 = \max\left\{x \in [0, \text{Var}(X)]\} \,:\, \frac{\partial U(x;\delta)}{\partial x} = 0\right\}. \tag{37}$$

By combining Part 1 of the theorem above with the asymptotic MMSE formula in Theorem 2, and noting that $\delta_{\text{in}} = \delta\frac{\Lambda}{\Lambda+\omega-1}$, we immediately obtain the following corollary.

Corollary 3.1 (Bayes-optimality of SC-GAMP): Consider the setup of Theorem 1, part 1, and suppose that there exists some $\delta_0 > \delta$ such that $U(x; \delta_{\text{in}})$ has a unique minimizer in x for $\delta_{\text{in}} \in [\delta, \delta_0]$. Then for any $\epsilon > 0$, there exist finite ω_0, t_0 such that for all $\omega > \omega_0$, $t > t_0$, and Λ sufficiently large, the asymptotic MSE of Bayes SC-GAMP almost surely satisfies:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} \,-\, \bar{g}^*_{\text{in}}(\boldsymbol{q}(t);\, t)\|^2 \\
&\leq \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A},\, \boldsymbol{y}\}\|^2\} + \epsilon \\
&= \underset{x\in[0,\text{Var}(X)]}{\arg\min}\; U(x; \delta) + \epsilon.
\end{aligned} \tag{38}$$

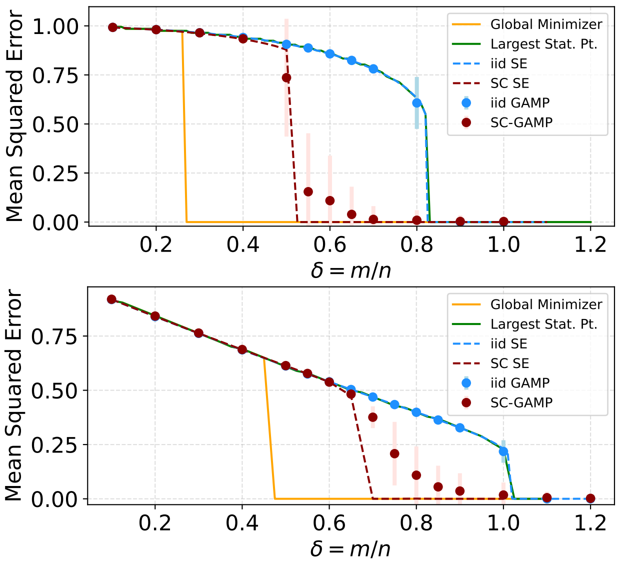

Fig. 2. MSE of Bayes-optimal GAMP with i.i.d. Gaussian and spatially coupled designs, for noiseless phase retrieval (top) and ReLU model (bottom). For phase retrieval the signal entries are drawn i.i.d. from $\{-a, a\}$ with probabilities $\{0.4, 0.6\}$, and a chosen such that the variance is 1. For ReLU, the signal entries are drawn i.i.d. from $\{-b, 0, b\}$ with probabilities $\{0.25, 0.5, 0.25\}$, and b chosen such that the variance is 1. In both cases, signal dimension $n = 20000$, the base matrix parameters are $(\omega = 6, \Lambda = 40)$, and the empirical performance is obtained by averaging over 100 trials.

The Bayes-optimality of SC-GAMP is analogous to the *threshold saturation* phenomenon in coding theory where the threshold of SC-LDPC codes with iterative decoding approaches the MAP decoding threshold in the large system limit. If $U(x; \delta)$ has a unique stationary point in $x \in [0, \text{Var}(X)]$ at which the minimum is attained, then Theorem 3 and Corollary 3.1 imply that the MMSE can be achieved by an i.i.d. Gaussian sensing matrix with estimation via standard GAMP, i.e., spatial coupling is not required in this case. However, as illustrated by the numerical results in Fig. 2 , when $U(x; \delta)$ has multiple stationary points the MSE of a spatially coupled design can be substantially lower than that of an i.i.d. design.

Numerical Results: Bayes SC-GAMP was evaluated for two different GLMs: noiseless phase retrieval ($\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$) and ReLU ($\boldsymbol{y} = \max(\boldsymbol{A}\boldsymbol{x}, \mathbf{0})$). Fig. 2 plots the MSE of Bayes SC-GAMP and Bayes GAMP (with i.i.d. Gaussian matrix) at convergence for a range of δ values. The orange solid line is the global minimizer of $U(x; \delta_{\text{in}})$ and the green solid line the largest stationary point of $U(x; \delta)$. The dashed lines are the state evolution MSE predictions for GAMP, given by (29). For both phase retrieval and ReLU, the curves show that the minimum required δ for near-perfect reconstruction is significantly lower for the spatially coupled design compared to the i.i.d. Gaussian one. The difference between the empirical performance of SC-GAMP and its state evolution curve is due to finite length effects. The difference between the state evolution curve (dashed red) and the global minimizer (orange) is because of the small values ω, Λ. In particular, we do not expect that the $\omega > \omega_0$ condition of Theorem 3 is satisfied with the relatively small value of $\omega = 6$.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, April 2006.

[2] Y. C. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*. Cambridge University Press, 2012.

[3] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[4] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," *Acta Numerica*, vol. 29, p. 125–228, 2020.

[5] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Conf. on Inf. Sciences and Systems (CISS)*, 2008, pp. 16–21.

[6] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.

[7] R. Tibshirani, "Regression shrinkage and selection with the Lasso," *J. Royal. Statist. Soc B*, vol. 58, pp. 267–288, 1996.

[8] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the Lasso and generalizations*. Chapman and Hall/CRC, 2019.

[9] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Neural Information Processing Systems*, 2013.

[10] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[11] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[12] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *Foundations of Computational Mathematics*, vol. 19, pp. 703–773, 2019.

[13] W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2347–2356, 2019.

[14] Y. M. Lu and G. Li, "Phase transitions of spectral initialization for high-dimensional non-convex estimation," *Information and Inference: A Journal of the IMA*, vol. 9, no. 3, pp. 507–541, 2020.

[15] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1275–1297, 2013.

[16] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.

[17] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A*, vol. 36, pp. 11 111–11 121, 2003.

[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[19] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[20] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.

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

[22] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2011, pp. 2168–2172.

[23] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1043–1055, 2014.

[24] J. Ma, J. Xu, and A. Maleki, "Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3600–3629, 2019.

[25] A. Maillard, B. Loureiro, F. Krzakala, and L. Zdeborová, "Phase retrieval in high dimensions: Statistical and computational phase transitions," in *Neural Information Processing Systems (NeurIPS)*, 2020.

[26] M. Mondelli and R. Venkataramanan, "Approximate message passing with spectral initialization for generalized linear models," in *Int. Conf. Artificial Intelligence and Stat. (AISTATS)*, 2021, pp. 397–405.

[27] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse pca," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 2197–2201.

[28] A. K. Fletcher and S. Rangan, "Iterative reconstruction of rank-one matrices in noise," *Information and Inference: A Journal of the IMA*, vol. 7, no. 3, pp. 531–562, 2018.

[29] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 7, p. 073403, 2017.

[30] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Annals of Statistics*, vol. 45, no. 1, pp. 321–345, 2021.

[31] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," *CoRR*, vol. abs/1010.6020, 2010. [Online]. Available: http://arxiv.org/abs/1010.6020

[32] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

[33] J. Barbier, C. Schülke, and F. Krzakala, "Approximate message-passing with spatially coupled structured operators, with applications to compressed sensing and sparse superposition codes," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, p. P05013, 2015.

[34] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity achieving sparse superposition codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 4894–4927, Aug. 2017.

[35] C. Rush, K. Hsieh, and R. Venkataramanan, "Capacity-achieving spatially coupled sparse superposition codes with AMP decoding," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4446–4484, 2021.

[36] K. Hsieh, C. Rush, and R. Venkataramanan, "Near-optimal coding for many-user multiple access channels," *IEEE Journal on Special Areas in Inf. Theory*, vol. 3, no. 1, pp. 21–36, March 2022.

[37] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, "Spatially coupled LDPC codes constructed from protographs," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4866–4889, Sept. 2015.

[38] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb 2011.

[39] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec 2013.

[40] D. J. Costello, L. Dolecek, T. E. Fuja, J. Kliewer, D. G. M. Mitchell, and R. Smarandache, "Spatially coupled sparse codes on graphs: theory and practice," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 168–176, July 2014.

[41] P. Pascual Cobo, K. Hsieh, and R. Venkataramanan, "Bayes-optimal estimation in generalized linear models via spatial coupling," on arXiv.

[42] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Information and Inference: A Journal of the IMA*, vol. 2, no. 2, pp. 115–144, 2013.

[43] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth, "A unifying tutorial on approximate message passing," *Foundations and Trends in Machine Learning*, vol. 15, no. 4, pp. 335–536, 2022.