A General Compressive Sensing Construct Using Density Evolution

Hang Zhang ©, Afshin Abdi ©, and Faramarz Fekri ©, Fellow, IEEE

Abstract—This paper proposes a general framework to design a sparse sensing matrix $A \in \mathbb{R}^{m \times n}$, for a linear measurement system $y = Ax + w$, where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $w \in \mathbb{R}^m$ denote the measurements, the signal with certain structures, and the measurement noise, respectively. By viewing the signal reconstruction from the measurements as a message passing algorithm over a graphical model, we leverage tools from coding theory, namely the density evolution technique, and provide a framework for the design of matrix $A$. Two design schemes for the sensing matrix, namely, (i) a regular sensing and (ii) a preferential sensing, are proposed and are incorporated into a single framework. As illustration, we consider the Lasso regression, ridge regression, and elastic net regression, and show that our framework can reproduce the classical results on the minimum sensor number, i.e., $m$. In the preferential sensing scenario, we consider the case in which the whole signal is divided into two disjoint parts, namely, high-priority part $x^p$ and low-priority part $x^l$. Then, by formulating the sensing system design as a bi-convex optimization problem, we obtain sensing matrices which can provide a preferential treatment for $x^p$. Numerical experiments with both synthetic data and real-world data are also provided to verify the effectiveness of our design scheme.

Index Terms—Compressed sensing, graphical models.

I. INTRODUCTION

This paper considers a linear sensing relation as

$$y = Ax + w,$$

where $y \in \mathbb{R}^m$ denotes the measurements, $A \in \mathbb{R}^{m \times n}$ is the sensing matrix, $x \in \mathbb{R}^n$ is the signal to be reconstructed, and $w \in \mathbb{R}^m$ is the measurement noise with iid Gaussian distribution $N(0, \sigma^2)$. To reconstruct $x$ from $y$, one widely used method is the regularized M-estimator

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \frac{||y - Ax||^2}{2\sigma^2} + f(x),$$

where $f(\cdot)$ is the regularizer used to enforce a desired structure for $\hat{x}$. To ensure reliable recovery of $x$, sensing matrix $A$ needs to satisfy certain conditions, e.g., the incoherence in [2], RIP in [3, 4], the neighborhood stability in [5], irrepresentable condition in [6], etc. Notice that all the above works treat each entry of $x$ equally. However, in certain applications, entries of $x$ may have unequal importance from the recovery perspective. One practical application is the image compression, i.e., JPEG compression, where coefficients corresponding to the high-frequency part are more critical than the rest of coefficients. In this work, we focus on the sparse sensing matrix $A$. Leveraging tools from coding theory, namely, density evolution (DE), we propose a heuristic but general design framework of $A$ to meet the requirements of the signal reconstruction such as placing more importance on the accuracy of certain components of the signal. At the core of our work is the application of DE in message passing (MP) algorithm, which is also referred to as belief propagation, or sum-product, or min-sum algorithm. These different names are due to its broad spectrum of applications and its constant rediscovery in different fields. In physics, this algorithm existed no later than 1935, when Bethe used a free-energy functional to approximate the partition function (cf. [7]). In the probabilistic inference, Pearl developed it in 1988 for acyclic Bayesian networks and showed it leads to the exact inference [8]. The most interesting thing is its discovery in the coding theory. In early 1960s, Gallager proposed sum-product algorithm to decode low density parity check (LDPC) codes over graphs [9]. However, Gallager work was almost forgotten and was rediscovered again in 90s [10, 11]. Later [12] equipped it with DE and used it for the design of LDPC codes for capacity achieving over certain channels. When narrowing down to the compressed sensing (CS), MP has been widely used for signal reconstruction [13, 14, 15, 16, 17, 18, 19, 20, 21] and analyzing the performance under some specific sensing matrices. The following briefly discusses the related work in the sensing matrix.

Related work: In the context of the sparse sensing matrix, the authors in [22] first proposed a so-called sudocode construction technique and later presented a decoding algorithm based on the MP in [23]. In [24], the non-negative sparse signal $x$ is considered under the binary sensing matrix. The work in [25] linked the channel encoding with the CS and presented a deterministic way of constructing sensing matrix based on a high-girth LDPC code. In [14, 16, 26], the authors considered the verification-based decoding and analyzed its performance with DE. In [15], the spatial coupling is first introduced into CS and is evaluated with the decoding scheme adapted from [26]. However, all the above

1An introduction can be found in https://jpeg.org/jpeg/documentation.html.
mentioned works focused on the noiseless setting, i.e., \( w = 0 \) in (1). In [17], [18], [19], the noisy measurement is considered. A sparse sensing matrix based on spatial coupling is analyzed in the large system limit with replica method and DE. They proved its recovery performance to be optimal when \( m \) increases at the same rate of \( n \), i.e., \( m = O(n) \).

Moreover, in the context of a dense sensing matrix, the analytical tool switches from DE to state evolution (SE), which is first proposed in [20], [21]. Together with SE comes the approximate message passing (AMP) decoding scheme. The empirical experiments suggest AMP has better scalability when compared with \( L_1 \) construction scheme without much sacrifice in the performance. Additionally, an exact phase transition formula can be obtained from SE, which predicts the performance of AMP to a good extent. Later, [27] provided a rigorous proof for the phase transition property by the conditioning technique from Erwin Bolthausen and [28] extended AMP to general M-estimation.

Note that the above mentioned related works are not exhaustive due to their large volume. For a better understanding of the MP algorithm, the DE, and their application to the compressive sensing, we refer the interested readers to [7], [19], [29]. In addition to the work based on MP, there are other works based on LDPC codes or graphical models. Since they are not closely related to ours, we only mention their names without further discussion [30], [31], [32], [33], [34], [35], [36].

Contributions: Compared to the previous works exploiting MP [14], [15], [16], [17], [18], [19], [20], [21], [26], our focus is on the sensing matrix design rather than the decoding scheme, which is based on the M-estimator with regularizer. Exploiting the DE, we propose a universal framework which supports both the regular sensing and the preferential sensing for recovering the signal. A detailed description of our contributions comes as follows.

- Regular Sensing: We consider the sparse signal setting and Gaussian signal setting. For the sparse signal setting, we consider a \( k \)-sparse signal \( x^k \in \mathbb{R}^n \) and associate it with a prior distribution such that each entry is zero with probability \( 1 - k/n \). For both the \( L_1 \) regularization and elastic net regularization, we can reproduce the classical results in CS, i.e., \( m \geq c k \log n \). For the Gaussian signal setting, we consider the Gaussian prior and show the minimum sensor number \( m \) should be the same order of the signal length \( n \).

- Preferential Sensing: We revisit the sparse signal setting and Gaussian signal setting; and design the sensing matrix that would result in more accurate (or exact) recovery of the high-priority sub-block of the signal relative to the low-priority sub-block. Numerical experiments confirm the effectiveness of our framework: the reconstruction error in the high-priority sub-block can be reduced significantly.

In addition, we should emphasize that although we only consider three types of regularizations, our framework can easily be extended to other priors.

Organization: In Section II, we formally state our problem and construct the graphical model. In Section III, we focus on the regular sensing and propose the density evolution framework. In Section IV, the framework is further extended to the preferential sensing. Generalizations are presented in Section V, simulation results are put in Section VI, and conclusions are drawn in Section VII.
As an inference problem, which can be solved via the message-passing algorithm. The square icons represent the check nodes while the circle icons represent the variable nodes.

III. SENSING MATRIX FOR REGULAR SENSING

With the aforementioned graphical model, we can view recovering $x^2$ as an inference problem, which can be solved via the message-passing algorithm [37]. Adopting the same notations as in [29] as shown in Fig. 1, we denote $m^{(t)}_{t \rightarrow a}$ as the message from the variable node $v_i$ to check node $c_a$ at the $t$th round of iteration. Likewise, we denote $\hat{m}^{(t)}_{t \rightarrow i}$ as the message from the check node $c_a$ to variable node $v_i$. Then message-passing algorithm is written as

$$m^{(t+1)}_{t \rightarrow a}(x_i) \cong e^{-f(x_i)} \prod_{b \in \partial \setminus a} m^{(t)}_{b \rightarrow t}(x_i);$$

$$\hat{m}^{(t+1)}_{a \rightarrow i}(x_i) \cong \int \prod_{j \in \partial \setminus i} m^{(t+1)}_{j \rightarrow a}(x_i) \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{j=1}^{n} A_{a,j} x_j \right)^2} dx_j,$$

where $\partial t$ and $\partial a$ denote the neighbors connecting with $v_i$ and $c_a$, respectively, and the symbol $\cong$ refers to the equality up to the normalization. At the $t$th iteration, we recover $x_i$ by maximizing the posterior probability

$$\hat{x}_i^{(t)} = \arg\max_{x_i} \mathbb{P}(x_i | y) \cong \arg\max_{x_i} e^{-f(x_i)} \prod_{a \in \partial i} \hat{m}^{(t)}_{a \rightarrow i}(x_i).$$

In the design of matrix $A$, there are some general desirable properties that we wish to hold (specific requirements will be discussed later): (i) a correct signal reconstruction under the noiseless setting; and (ii) minimum number of measurements, or equivalently minimum $m$. Before proceeding, we first introduce the generating polynomials $\lambda(\alpha) = \sum_i \lambda_i \alpha^{-i}$ and $\rho(\alpha) = \sum_i \rho_i \alpha^{-i-1}$, which correspond to the degree distributions for variable nodes and check nodes, respectively. We denote the coefficient $\lambda_i$ as the fraction of variable nodes with degree $i$, and similarly we define $\rho_i$ for the check nodes. An illustration of the generating polynomials $\lambda(\alpha)$ and $\rho(\alpha)$ is shown in Fig. 2.

A. Density Evolution

To design the matrix $A$, we study the reconstruction of $x^2$ from $y$ via the convergence analysis of the message-passing over the factor graph. Due to the parsimonious setting of $A$, we have $\varnothing$ to be sparse and propose to borrow a tool known as density evolution (DE) [37, 38, 39] that is already proven to be very powerful in analyzing the convergence in sparse graphs resulting from LDPC.

Basically, DE views $m^{(t)}_{t \rightarrow a}$ and $\hat{m}^{(t)}_{a \rightarrow i}$ as RVs and tracks the changes of their probability distribution. When the message-passing algorithm converges, we would expect their distributions to become more concentrated. However, different from discrete RVs, continuous RVs $m^{(t)}_{t \rightarrow a}$ and $\hat{m}^{(t)}_{a \rightarrow i}$ in our case require infinite bits for their precise representation in general, leading to complex formulas for DE. To handle such an issue, we approximate $m^{(t)}_{t \rightarrow a}$ and $\hat{m}^{(t)}_{a \rightarrow i}$ as Gaussian RVs, i.e., $m^{(t)}_{t \rightarrow a} \sim N(\mu_{t \rightarrow a}, \sigma_{t \rightarrow a})$ and $\hat{m}^{(t)}_{a \rightarrow i} \sim N(\mu_{a \rightarrow i}, \sigma_{a \rightarrow i})$, respectively. Since the Gaussian distribution is uniquely determined by its mean and variance, we will be able to reduce the DE to finite dimensions as in [17, 18, 39].

In our work, the DE tracks two quantities $E^{(t)}$ and $V^{(t)}$, which denote the deviation from the mean and the average of the variance, respectively, and are defined as

$$E^{(t)} = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{a=1}^{m} \left( m^{(t)}_{t \rightarrow a} - x_i^2 \right)^2;$$

$$V^{(t)} = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{a=1}^{m} \left( \hat{m}^{(t)}_{a \rightarrow i} \right)^2.$$

Then we can show that the DE analysis yields

$$E^{(t+1)} = \mathbb{E}_{\operatorname{prior}(s)} \mathbb{E}_{z} \left[ h_{\text{mean}} \left( s + \sum_{i,j} \rho_i \lambda_j z \sqrt{\frac{i}{j}} E^{(t)} + \frac{A \sigma^2}{j} \right) \right];$$

$$\sum_{i,j} \rho_i \lambda_j A \sigma^2 + i V^{(t)} \right) - s \right] - 2 \right];$$

$$V^{(t+1)} = \mathbb{E}_{\operatorname{prior}(s)} \mathbb{E}_{h_{\text{var}}} \left( s + \sum_{i,j} \rho_i \lambda_j z \sqrt{\frac{i}{j}} E^{(t)} + \frac{A \sigma^2}{j} \right);$$

$$\sum_{i,j} \rho_i \lambda_j A \sigma^2 + i V^{(t)} \right) \right),$$

where $\operatorname{prior}(\cdot)$ denotes the true prior on the entries of $x^2$, and $z$ is a standard normal RV $N(0, 1)$. The functions $h_{\text{mean}}(\cdot)$ and $h_{\text{var}}(\cdot)$ are to approximate the mean $\mu_{t \rightarrow a}$ and variance $\sigma_{t \rightarrow a}$, which are given by

$$h_{\text{mean}}(\mu; v) = \lim_{\gamma \rightarrow \infty} \int x_i e^{-\gamma f(x_i)} e^{-\gamma (x_i - \mu)^2} dx_i;$$

$$h_{\text{var}}(\mu; v) = \lim_{\gamma \rightarrow \infty} \int x_i^2 e^{-\gamma f(x_i)} e^{-\gamma (x_i - \mu)^2} dx_i;$$

$$- \left( h_{\text{mean}}(\mu; v) \right)^2.$$

---

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
For detailed explanations and the proof, we refer interested readers to the supplementary material.

B. Sensing Matrix Design

Once the values of polynomial coefficients \( \{ \lambda_i \}_i \) and \( \{ \rho_i \}_i \) are determined, we can construct a random graph \( G = (V, \delta) \), or equivalently the sensing matrix \( A \), by setting \( A_{ij} = \text{P}(A_{ij} = A^{-1/2}) = \text{P}(A_{ij} = -A^{-1/2}) = \frac{1}{2} \), if there is an edge \((v_i, c_j) \in \delta\); otherwise we set \( A_{ij} \) to zero. Hence the sensing matrix construction reduces to obtaining the feasible values of \( \{ \lambda_i \}_i \) and \( \{ \rho_i \}_i \) while satisfying certain properties for the signal reconstruction as discussed in the following.

Our first requirement is to have a perfect signal reconstruction under the noiseless scenario (\( \sigma^2 = 0 \)). This implies that

- the algorithm must converge, i.e., \( \lim_{t \to \infty} V(t) = 0 \);
- the average error should shrink to zero, i.e., \( \lim_{t \to \infty} E(t) = 0 \).

Second, we wish to minimize the number of measurements. Using the fact that \( n(\sum_i \lambda_i) = m(\sum_i \rho_i) = \sum_i z_i(v_i, c_j) \in \delta \), we formulate the above two design criteria as the following optimization problem

\[
\min_{\lambda \in \Delta_{d-1}, \rho \in \Delta_{d-1}} \frac{m}{n} = \sum_{i \geq 2} \hat{\alpha}_i,
\]

\[\text{s.t. } \lim_{t \to \infty} \left( E(t), V(t) \right) = (0, 0); \]

\[\lambda_i = 1, \rho_i = 0, \]

where \( \Delta_{d-1} \) denotes the \( d \)-dimensional simplex, namely, \( \Delta_{d-1} \ni z \in \mathbb{R}^d \mid \sum_i z_i = 1, z_i \geq 0 \). The constraint in (11) is to avoid one-way message passing as in [12], [39].

Generally speaking, we need to run DE numerically to check the requirement (10) for every possible values of \( \{ \lambda_i \}_i \) and \( \{ \rho_i \}_i \). However, for certain choices of regularizers \( f(\cdot) \), we can reduce the requirement (10) to some closed-form equations. For example, if we set the prior in (3) to be a Laplacian distribution, i.e., \( e^{-\beta|z|} \), then the regularizer \( f(\cdot) \) in (2) becomes \( \beta \| \cdot \|_1 \) and the M-estimator in (2) transforms to Lasso [41]; if we set the prior to be Gaussian distribution, i.e., \( e^{\beta|z|^2} \), the M-estimator in (2) transforms to ridge regression [42]. More discussions come as follows.

C. Examples of Regular Sensing for Various Priors

This subsection considers some specific priors and illustrates our design schemes of the corresponding sensing matrices. Roughly speaking, our design scheme is divided into 3 stages: (i) DE analysis; (ii) distribution approximation; and (iii) convergence criteria derivation. In the following context, we will study the \( \ell_1 \), \( \ell_2 \), and elastic net regularization; and show how to apply our proposed design scheme. For other types of regularizations, we can follow a similar procedure and simplify the requirement \( \lim_{t \to \infty} \left( E(t), V(t) \right) = (0, 0) \) to some closed-form expressions.

Example 1 (Regular sensing with \( \ell_1 \) regularizer): Assuming the signal \( z^x \) is \( k \)-sparse, i.e., \( \| z^x \|_0 \leq k \), we would like to recover \( z^x \) with the regularizers \( \beta \| \cdot \|_1 \), which corresponds to the Laplacian prior.

**Stage I: DE analysis.** Following the approaches in [20] in the noiseless case, we can show that

\[
E^{(t+1)} = \mathbb{E}_{\text{prior}(s), z \sim \mathcal{N}(0,1)} \left[ \text{prox} \left( s + a_1 z \sqrt{E^{(t)}}; \beta a_2 V^{(t)} \right) - s \right]^2 ;
\]

\[
V^{(t+1)} = \mathbb{E}_{\text{prior}(s) \mathbb{E}_{z \sim \mathcal{N}(0,1)}} \left[ \beta a_2 V^{(t)} \text{prox} \left( s + a_1 z \sqrt{E^{(t)}}; \beta a_2 V^{(t)} \right) \right],
\]

(12)

where \( a_1 \) is defined as \( \sum_{i,j} \rho_i \lambda_j / \sqrt{|j|} \) and \( a_2 \) is defined as \( \sum_{i,j} \rho_i \lambda_j (i/j) \). Further, operator \( \text{prox}(a; b) \) is the soft-thresholding estimator defined as \( \text{sign}(a) \max(|a| - b, 0) \), and operator \( \text{prox}(a; b) \) is the derivative w.r.t. the first argument.

**Remark 1:** Unlike SE that only tracks \( E(t) \), our DE takes into account both the average variance \( V(t) \) and the deviation from mean \( E(t) \). Assuming \( V(t) \propto \sqrt{E(t)} \), our DE equation w.r.t. \( E(t) \) in (12) reduces to a similar form as SE.

Having discussed its relation with SE, we now show that our DE can reproduce the classical results in compressive sensing, namely, \( m \geq c_0 k \log(n/k) = O(k \log n) \) (cf. [43]) under the regular sensing matrix design, i.e., when all variable nodes have the same degree \( d \) and the check nodes have the same degree \( dc \).

**Stage II: Distribution approximation.** We approximate the ground-truth distribution with the Laplacian prior. Assuming that the entries of \( x^z \) are iid and \( x^z \in \mathbb{R}^n \) is \( k \)-sparse, each entry becomes zero with probability \((1 - k/n)\). Hence we set \( \beta \) such that the probability mass within the region \([-c_0, c_0] \) (where \( c_0 \) is some small positive constant) with the Laplacian prior is equal to \( 1 - k/n \). That is

\[
\frac{\beta}{2} \int_{|\alpha| \leq c_0} e^{-\beta|\alpha|} d\alpha = 1 - \frac{k}{n},
\]

which results in \( \beta = c_0 \log(n/k) \).

**Stage III: Convergence criteria derivation.** Enforcing the criteria \( \lim_{t \to \infty} \left( E(t), V(t) \right) = (0, 0) \) under the noiseless setting (i.e., \( \sigma = 0 \)), we conclude the following

**Theorem 1:** Let \( x^z \) be a \( k \)-sparse signal and assume that \( \beta \) is set to \( c_0 \log(n/k) \). Then, the necessary conditions for \( \lim_{t \to \infty} \left( E(t), V(t) \right) = (0, 0) \) associated with the DE equation in (12) are

\( (i) \sum_{i,j} \rho_i \lambda_j \sqrt{|j|} \leq c_1' \sqrt{n}/k \)

and

\( (ii) \sum_{i,j} \rho_i \lambda_j (i/j) \leq c_2' \sqrt{n} / k \log(n/k) \),

where \( c_1', c_2' > 0 \) are some positive constants.

**Remark 2:** Consider the settings in Theorem 1 and assume \( (i) \sum_{i,j} \rho_i \lambda_j \sqrt{|j|} \leq c_1' \sqrt{n}/k \) and \( (ii) \sum_{i,j} \rho_i \lambda_j (i/j) \leq c_2' \sqrt{n} / k \log(n/k) \). Then, there exists positive constants \( \varepsilon > 0 \) and \( \gamma < 1 \) such that \( \left\{ E(t), V(t) \right\} \) generated by (12) decrease exponentially provided the initial point \( (E(0)^2 + V(0)^2) \leq \varepsilon \), i.e., \( E(t) \leq \gamma^t E(0) \) and \( V(t) \leq \gamma^t V(0) \).

When turning to the regular design, namely, all variable nodes are with the degree \( dv \) and likewise all check nodes are with degree \( dc \), we can write \( a_1 \) and \( a_2 \) as \( \sqrt{n/m} \) and \( n/m \), respectively. Invoking Theorem 1 will then yield the classical
result of the lower bound on the number of measurements $m \geq c_0 k \log(n/k)$. The technical details are referred to Section A.

Example 2 (Regular sensing with $\ell_2$ regularizer): In addition to the Laplacian prior, we also considered the Gaussian prior, i.e., $e^{-\|x\|^2}$, which makes the M-estimator in (2) the ridge regression [44]. Assuming the ground-truth $x^s$ to be Gaussian distributed with zero mean and unit variance, we would like to recover the signal $x^s$ with the regularizer $f(x) = \|x\|^2$.

Stage I: DE analysis. Following a similar procedure, we obtain the following DE equation

$$E^{(t+1)} = \frac{a_1^2 E^{(t)} + a_2^2 (V^{(t)})^2}{(1 + a_2 V^{(t)})^2};$$

$$V^{(t+1)} = \frac{a_2 V^{(t)}}{1 + a_2 V^{(t)}},$$

(13)

where $a_1, a_2$ are defined the same as above, i.e., $a_1 = \sum_{i,j} \rho_i \lambda_j \sqrt{i/j}$ and $a_2 = \sum_{i,j} \rho_i \lambda_j (i/j)$.

Stage II: Distribution approximation. We can skip this stage as the ground-truth prior of $x^s$, namely, $N(0, 1)$, is used for the regularization.

Stage III: Convergence criteria derivation. Same as the above example, we let $\lim_{t \to \infty} E^{(t)}, V^{(t)} = (0, 0)$ when $\sigma = 0$ and obtain the following theorem.

Theorem 2: Provided that $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} < 1$, we have the average error $E^{(T)}$ and variance $V^{(T)}$ in (13) decrease exponentially after some iteration index $T$, that is, $E^{(t)} \leq e^{-c_0 (t-T)} E^{(T)}$ and $V^{(t)} \leq e^{-c_1 (t-T)} V^{(T)}$ whenever $t \geq T$. Here $c_0, c_1 > 0$ are some fixed constants.

Its proof is referred to Section B. To verify Theorem 2, we plot the trajectories of DE in (13), which is put in Fig. 3. Depending on whether $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j}$ is less than one or not, we find $E^{(t)}, V^{(t)}$ can converge to different fixed points.

With some standard algebraic manipulations, we can reduce Theorem 2 to the criteria $m \geq n$. This criteria is consistent with the previous finding: no savings can be achieved provided that $x^s$ resides within the whole linear space $\mathbb{R}^n$.

Example 3 (Regular sensing with elastic net regularizer): We revisit the sparse signal setting where $x^s$ is assumed to be $k$-sparse. Instead of $\ell_1$ regularization, we consider the elastic net regularization for signal reconstruction, where $f(x)$ is written as $\beta_1 \|x\|_1 + \beta_2 \|x\|^2_2$ ($\beta_1, \beta_2 > 0$). For the ease of analysis, we pick $\beta_1 = \beta_2$ and write it as $\beta$.

Stage I: DE analysis. Denote $a_1$ and $a_2$ as $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j}$ and $\sum_{i,j} \rho_i \lambda_j (i/j)$, respectively, we can write the corresponding DE equation as

$$E^{(t+1)} = E^{(t)} E_{\text{prior}(s)} \sum_{n=0}^{\infty} \left( s + a_1 z \sqrt{E^{(t)}} \right) \left( 1 + 2 \beta a_2 V^{(t)} \right)^{-\gamma} ;$$

$$V^{(t+1)} = \frac{\beta \cdot a_2 V^{(t)}}{1 + 2 \beta \cdot a_2 V^{(t)}} \cdot E_{\text{prior}(s)} \sum_{n=0}^{\infty} \left( s + a_1 z \sqrt{E^{(t)}} \right) \left( 1 + 2 \beta a_2 V^{(t)} \right)^{-\gamma} ;$$

(14)

where $Z$ is the normalization constant defined as $Z \triangleq \int_{-\infty}^{\infty} e^{-\beta |x| - \beta |x|^2} dx$, and $c_D$ is some small positive constant. Its physical meaning is that the two distributions have the same probability mass around zero. A detailed calculation suggests (15) is equivalent to

$$\frac{\text{erfc}(\sqrt{\beta(1+2c)/2})}{\text{erfc}(\sqrt{\beta/2})} = \frac{k}{n},$$

(16)

where $\text{erfc}(\cdot)$ is the complementary error function defined as $2/\sqrt{\pi} \int_{-\infty}^{\infty} e^{-x^2} dx$. Due to the complicated nature of $\text{erfc}(\cdot)$, generally speaking, we cannot express the solution of (16) in a closed form. For the benefits of our analysis, we instead consider its asymptotic behavior when $k \ll n$, or equivalently, $\beta \gg 1$. Exploiting the relation $\text{erfc}(\alpha) \approx \frac{e^{-\alpha^2}}{\sqrt{\pi} \alpha} (1 + \frac{1}{2} + \cdots + (-1)^{n-1} \frac{(2n-1)!}{(2\pi)^n} + \cdots)$ when $\alpha \to \infty$ (Page 584 in [45]), we can rewrite (16) as

$$\frac{e^{-c(t+1)\beta}}{1 + 2e^{-c(t+1)\beta}} \approx \frac{k}{n},$$

which leads to $\beta \approx c_D \log(c_D n/k)$ (Parameters $c_0, c_1 > 0$ are some positive constants associated with $c$).

Stage III: Convergence criteria derivation. For the DE equation in (14), we follow a similar procedure as in Example 1 and obtain the following necessary condition for $\lim_{t \to \infty} E^{(t)}, V^{(t)} = (0, 0)$.

Theorem 3: Consider the sparse signal setting where $x^s$ is $k$-sparse and set $\beta$ as $c_D \log(c_D n/k)$. Then the necessary conditions for $\lim_{t \to \infty} E^{(t)}, V^{(t)}$ associated with the DE equation in (14) are (i) $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} \leq c_1 \sqrt{n/k}$ and (ii) $\sum_{i,j} \rho_i \lambda_j (i/j) \leq \frac{c_0 n}{k \log(n/k)}$, where $c_1, c_2 > 0$ are some positive constants.

We notice that the necessary conditions in Theorem 3 are almost the same as that in Theorem 1. The only differences lies in the positive constants $c_1, c_1'$, $c_2$, and $c_2'$. In addition, we can show that $E^{(t)}, V^{(t)}$ generated by (14) decreases exponentially given the conditions in Theorem 3 and the initial point $(E^{(0)}, V^{(0)})$ is close to the origin point, i.e., $E^{(t)} \leq \gamma t E^{(0)}$ and $V^{(t)} \leq \gamma V^{(0)}$, $0 < \gamma < 1$. 

Fig. 3. Illustration of DE in (13). Left panel: $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} < 1$. Right panel: $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} > 1$. Notice that the left panel has a fix-point $(0, 0)$ while the right panel is with non-zero fix-point.
Having studied the \( \ell_1 \) regression, ridge regression, and elastic net regression, we have shown that the requirement \( \lim_{t \to \infty} (E^{(t)}, V^{(t)}) = (0, 0) \) in our design framework can reproduce the classical results.

IV. SENSING MATRIX FOR PREFERENTIAL SENSING

Having discussed the regular sensing scheme, this section explains as to how we apply our DE framework to design the sensing matrix \( A \) such that we can provide preferential treatment for different entries of \( x^\natural \). For example, the high priority components will be recovered more accurately than the low priority parts of \( x^\natural \).

A. Density Evolution

Dividing the entire \( x^\natural \) into the high-priority part \( x^\natural_H \in \mathbb{R}^{m_H} \) and low-priority part \( x^\natural_L \in \mathbb{R}^{m_L} \), we separately introduce the generating polynomials \( \lambda_H(\alpha) = \sum_{i} \lambda_{H,i} \alpha^i \) and \( \lambda_L(\alpha) = \sum_{i} \lambda_{L,i} \alpha^i \) for the high-priority part \( x^\natural_H \) and the low-priority part \( x^\natural_L \), respectively. Note that \( \lambda_{H,i} \) (and likewise \( \lambda_{L,i} \)) denotes the fraction of variable nodes corresponding to the high-priority part (low-priority part) with degree \( i \). Similarly, we introduce the generating polynomials \( \rho_H(\alpha) = \sum_{i} \rho_{H,i} \alpha^i \) and \( \rho_L(\alpha) = \sum_{i} \rho_{L,i} \alpha^i \) for the edges of the check nodes connecting to the high-priority part \( x^\natural_H \) and to the low-priority part \( x^\natural_L \), respectively.

Generalizing the analysis of the regular sensing, we separately track the average error and variance for \( x^\natural_H \) and \( x^\natural_L \). For the high-priority part \( x^\natural_H \), we define \( E^H \) as \( \sum_{n_H} \sum_{i \in H} (\mu_{i,n_H} - x^\natural_i)^2/(m \cdot n_H) \) and \( V^H \) as \( \sum_{m} \sum_{i \in H} (x^\natural_{i,n_H} + \delta^t_{H,i})/(m \cdot n_H) \), where \( n_H \) denotes the length of the high-priority part \( x^\natural_H \). Analogously, we define \( E^L \) and \( V^L \) for the low-priority part \( x^\natural_L \). Then we write the corresponding DE as

\[
E^{H(t+1)}_H = \mathbb{E}_{\text{prior}(s)} \mathbb{E}_{z \sim N(0,1)} \left[ h_{\text{mean}} \left( s + z \cdot b^{H(t+1)}_{H,1}, b^{H(t+1)}_{H,2} \right) - s \right]^2 ; \\
V^{H(t+1)}_L = \mathbb{E}_{\text{prior}(s)} \mathbb{E}_{z \sim N(0,1)} \left[ h_{\text{var}} \left( s + z \cdot b^{H(t+1)}_{H,1}, b^{H(t+1)}_{H,2} \right) \right],
\]

(17)

where \( b^{H(t+1)}_{H,1} \) and \( b^{H(t+1)}_{H,2} \) are defined as

\[
b^{H(t+1)}_{H,1} = \sum_{\ell,i,j} \lambda_{H,\ell,i} \rho_{L,j} \rho_{H,j} \left( \frac{\mathbb{A} s^2 + i E^{H(t)}_H + j E^{H(t)}_H}{\ell} \right) ; \\
b^{H(t+1)}_{H,2} = \sum_{\ell,i,j} \lambda_{H,\ell,i} \rho_{L,j} \rho_{H,j} \left( \frac{\mathbb{A} s^2 + i V^{H(t)}_L + j V^{H(t)}_L}{\ell} \right).
\]

The definitions of \( h_{\text{mean}} \) and \( h_{\text{var}} \) are as in (8). Switching the index \( H \) with \( L \) yields the DE w.r.t. the pair \( (E^{L(t+1)}, V^{L(t+1)}) \). Notice we can also put different regularizers \( f_H(\cdot) \) and \( f_L(\cdot) \) for \( x^\natural_H \) and \( x^\natural_L \). In this case, we need to modify the regularizers \( f(\cdot) \) in (8) to \( f_H(\cdot) \) and \( f_L(\cdot) \), respectively.

Remark 3:

B. Sensing Matrix Design

In addition to the constraints used in (9), the sensing matrix for preferential sensing must satisfy the following constraint:

\[
\text{Consistency requirement w.r.t. edge number: Consider the total number of edges incident with the high-priority part } x^\natural_H \text{ and the low-priority part } x^\natural_L \text{ in } (17) \text{. From the viewpoint of the variable nodes, we can compute this number as } n_H \left( \sum_i \lambda_{H,i} \right) \text{. Likewise, from the viewpoint of the check nodes, the total number of edges is obtained as } n_L \left( \sum_i \rho_{L,i} \right) \text{. Since the edge number should be the same with either of the above two counting methods, we obtain } \\
\sum_{i \in H} \left( \mu_{i,n_H} - x^\natural_i \right) = n_H \left( \sum_i i \lambda_{H,i} \right) = m \left( \sum_i i \rho_{H,i} \right).
\]

Similarly, the consistency requirement for the edges connecting to the low-priority part \( x^\natural_L \) would give \( \sum_{i \in L} \left( \mu_{i,n_L} - x^\natural_i \right) = n_L \left( \sum_i i \lambda_{L,i} \right) = m \left( \sum_i i \rho_{L,i} \right) \).

Moreover, we may have additional constraints depending on the measurement noise:

- **Preferential sensing for the noiseless measurement:** In the noiseless setting (\( \sigma = 0 \)), we require \( V_H \) and \( V_L \) to diminish to zero to ensure the convergence of the MP algorithm.

Besides, we require the average error \( E^{H(t)}_H \) in the high-priority part \( x^\natural_H \) to be zero. Therefore, the requirements can be summarized as

**Requirement 1.** In the noiseless setting, i.e., \( \sigma = 0 \), we require the quantities \( E^{H(t)}_H, V^{H(t)}_H, \) and \( V^{H(t)}_L \) in (17) to converge to zero

\[
\lim_{t \to \infty} (E^{H(t)}_H, V^{H(t)}_H, V^{H(t)}_L) = (0, 0, 0),
\]

(18)

which implies the MP converges and the high-priority part \( x^\natural_H \) can be perfectly reconstructed.

Notice that no constraint is placed on the average error \( E^{H(t)}_L \) for the low-priority part \( x^\natural_L \), since it is given a lower priority in reconstruction.

- **Preferential sensing for the noisy measurement:** Different from the noiseless setting, the high-priority part \( x^\natural_H \) cannot be perfectly reconstructed in the presence of measurement noise, i.e., \( \lim_{t \to \infty} E^{H(t)}_H > 0 \). Instead we consider the difference across iterations, namely, \( \delta^{H(t)}_E = E^{H(t+1)}_H - E^{H(t)}_H \) and \( \delta^{L(t)}_E = L^{L(t+1)}_L - L^{L(t)}_L \), which corresponds to the convergence rate. To provide an extra protection for the high-priority part \( x^\natural_H \), we would like \( \delta^{H(t)}_E \) to decrease at a faster rate. Hence, the following requirement:

**Requirement 2:** There exists a positive constant \( T_0 \) such that the average error \( E^{H(t)}_H \) converges faster than \( E^{L(t)}_L \) whenever \( t \geq T_0 \), i.e., \( \delta^{H(t)}_E \leq \delta^{L(t)}_E \).

Apart from the above constraints, we also require \( \lambda_{H,1} = \lambda_{H,1} = \rho_{L,1} = \rho_{L,1} = 0 \) to avoid one-way message passing [12, 37, 39]. Summarizing the above discussion, the design of the sensing matrix \( A \) for minimum number of measurements \( m \) reduces to the following optimization problem

\[
\min_{\lambda_{H,1} \in \Delta_{\lambda_{H,1}}, \rho_{L,1} \in \Delta_{\rho_{L,1}}, \rho_{L} \in \Delta_{\rho_{L}}, \rho_{H} \in \Delta_{\rho_{H}}} \left( \frac{m}{n} \right) \sum_{i} i \lambda_{H,i} + \frac{1}{n_H} \sum_{i} i \lambda_{H,i} ;
\]

(19)

subject to

\[
\sum_{i} i \lambda_{H,i} \frac{n_L}{n} + \sum_{i} i \rho_{H,i} \frac{n_L}{n} = n_H \frac{n}{n_L} ;
\]

(20)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Requirement (1) and (2):

\[
\lambda_{H,1} = \lambda_{H,1} = \rho_{H,1} = \rho_{H,1} = 0, \tag{21}
\]

where \(\Delta_{L-1}\) denotes the \(d\)-dimensional simplex, and the parameters \(d_{VH}\) and \(d_{VL}\) denote the maximum degree w.r.t. the variable nodes corresponding to the high-priority part \(x^H_\ell\) and low-priority part \(x^L_\ell\), respectively. Similarly we define the maximum degree \(d_{VH}\) and \(d_{VL}\) w.r.t. the check nodes.

The difficulties of the optimization problem in (19) come from two-fold: (i) requirements from DE; and (ii) non-convex nature of (19).

C. Example of Preferential Sensing for Various Priors

We will revisit the previous examples and show how to simplify the optimization problem in (19). Similar to the procedure in Subsection III-C, our relaxation procedure consists of three stages. Here, we focus on Stage III as the first two stages are exactly the same as that in Subsection III-C.

Example 4 (Preferential sensing with \(\ell_1\) regularizer): Consider a sparse signal \(x^s\) whose high-priority part \(x^H_\ell \in \mathbb{R}^{n_H}\) and the low-priority part \(x^L_\ell \in \mathbb{R}^{n_L}\) are \(k_H\)-sparse and \(k_L\)-sparse, respectively. In addition, we assume \(k_H/n_H \gg k_L/n_L\), implying that the high-priority part \(x^H_\ell\) contains more data.\(^3\)

Ideally, we need to numerically run the DE update equation in (17) to check whether the requirements in (21) are met, which may be computationally prohibitive. In practice, we would relax these conditions to arrive at some closed forms. The following outlines our relaxation strategy with all technical details being deferred to the supplementary material.

Relaxation of Requirement 1: First we require the variance to converge to zero, i.e., \(\lim_{t \to \infty} \text{Var}(V_{H,1}^{(t)}, V_{L,1}^{(t)}) = (0, 0)\). The derivation of its necessary condition consists of two parts: (i) we require the point \((0,0)\) to be a fixed point of the DE equation w.r.t. \(V_{H,1}^{(t)}\) and \(V_{L,1}^{(t)}\); and (ii) we require that the average variance \(\text{Var}(V_{H,1}^{(t)}, V_{L,1}^{(t)})\) converges in the region where the magnitudes of \(V_{H,1}^{(t)}\) and \(V_{L,1}^{(t)}\) are sufficiently small.

The main technical challenge lies in investigating the convergence of \((V_{H,1}^{(t)}, V_{L,1}^{(t)})\). Define the difference \(\delta_{V_{H,1}^{(t)}}\) and \(\delta_{V_{L,1}^{(t)}}\) across iterations as \(\delta_{V_{H,1}^{(t)}} = V_{H,1}^{(t+1)} - V_{H,1}^{(t)}\) and \(\delta_{V_{L,1}^{(t)}} = V_{L,1}^{(t+1)} - V_{L,1}^{(t)}\), respectively. Then, we obtain a linear equation

\[
\begin{bmatrix}
\delta_{V_{H,1}^{(t+1)}} \\
\delta_{V_{L,1}^{(t+1)}}
\end{bmatrix} = \begin{bmatrix} L_{V_{H,1}^{(t)}} & L_{V_{L,1}^{(t)}} \end{bmatrix} \begin{bmatrix}
\delta_{V_{H,1}^{(t)}} \\
\delta_{V_{L,1}^{(t)}}
\end{bmatrix}.
\]

via the Taylor-expansion. Imposing the convergence constraints on \(V_{H,1}^{(t)}\) and \(V_{L,1}^{(t)}\), i.e., \(\lim_{t \to \infty} \text{Var}(V_{H,1}^{(t)}, V_{L,1}^{(t)}) = (0, 0)\), yields the condition \(\inf \|L_{V_{H,1}^{(t)}}\|_{op} \leq 1\). That is

\[
\left[ \frac{(\beta_H k_H/n_H \sum_{\ell} \lambda_{H,\ell})^2}{\ell} + \frac{(\beta_L k_L/n_L \sum_{\ell} \lambda_{L,\ell})^2}{\ell} \right] \times \left[ \left( \sum_i \rho_{H,i} \right)^2 + \left( \sum_i \rho_{L,i} \right)^2 \right] \leq 1. \tag{23}
\]

\(^3\)The high-priority part \(x^H_\ell\) may still receive extra protection even if \(k_H/n_H \leq k_L/n_L\). One numerical experiment is attached in the Appendix.

Then we turn to the behavior of \(E_{H,1}^{(t)}\) assuming \(E_{L,1}^{(t)}\) converges to a fixed non-negative constant \(E_{L,1}^{(\infty)}\), we would like \(E_{H,1}^{(t)}\) to converge to zero. Following the same strategy as above, we obtain the following condition

\[
\frac{k_H}{n_H} \left( \sum_{\ell} \frac{\lambda_{H,\ell}}{\sqrt{\ell}} \right)^2 \left[ \left( \sum_i \sqrt{i} \rho_{H,i} \right)^2 + \left( \sum_i \sqrt{i} \rho_{L,i} \right)^2 \right] \leq 1. \tag{24}
\]

A formal statement is summarized as

**Proposition 1**: Consider the setting in Example 4, then the necessary conditions for Requirement 1 are given by (23) and (24).

The technical details are put in the supplementary material.

Relaxation of Requirement 2: First we define the difference across iterations as \(\delta_{E_{H,1}^{(t)}} = E_{H,1}^{(t)} - E_{H,1}^{(t+1)}\) and \(\delta_{E_{L,1}^{(t)}} = E_{L,1}^{(t)} - E_{L,1}^{(t+1)}\). Using the Requirement 2, we perform the Taylor expansion w.r.t. the difference \(\delta_{E_{H,1}^{(t)}}\) and \(\delta_{E_{L,1}^{(t)}}\), and obtain the linear equation

\[
\begin{bmatrix}
\delta_{E_{H,1}^{(t)}} \\
\delta_{E_{L,1}^{(t)}}
\end{bmatrix} = \begin{bmatrix} L_{E_{H,1}^{(t+1)}} & L_{E_{L,1}^{(t+1)}} \end{bmatrix} \begin{bmatrix}
\delta_{E_{H,1}^{(t+1)}} \\
\delta_{E_{L,1}^{(t+1)}}
\end{bmatrix}.
\]

To ensure the reduction of \(\delta_{E_{H,1}^{(t)}}\) at a faster rate than \(\delta_{E_{L,1}^{(t)}}\), we would require \(L_{E,1,11} \leq L_{E,21}\) and \(L_{E,1,2} \leq L_{E,2,2}\). This results in

\[
\frac{k_H}{n_H} \left( \sum_{\ell} \frac{\lambda_{H,\ell}}{\sqrt{\ell}} \right)^2 \leq \frac{k_L}{n_L} \left( \sum_{\ell} \frac{\lambda_{L,\ell}}{\sqrt{\ell}} \right)^2, \tag{25}
\]

which completes the relaxation.

**Example 5 (Preferential sensing with \(\ell_2\) regularizer)**: We revisit the Gaussian setting where \(x^s \in \mathbb{R}^{n_H+n_L}\) can be divided into two disjoint parts: the high-priority part \(x^H_\ell \in \mathbb{R}^{n_H}\) and the low-priority part \(x^L_\ell \in \mathbb{R}^{n_L}\). Their priors are assumed to be \(e^{-\beta_H \| \cdot \|_2^2}\) and \(e^{-\beta_L \| \cdot \|_2^2}\); and the corresponding regularizers are picked as \(\beta_H \| \cdot \|_2^2\) and \(\beta_L \| \cdot \|_2^2\), respectively. Then we conclude

**Relaxation of Requirement 1**: Imposing the convergence constraints on \(V_{H,1}^{(t)}\) and \(V_{L,1}^{(t)}\) yields

\[
\left[ \left( \sum_i \sqrt{i} \rho_{H,i} \right)^2 + \left( \sum_i \sqrt{i} \rho_{L,i} \right)^2 \right] \times \left[ \left( \sum_i \rho_{H,i} \right)^2 + \left( \sum_i \rho_{L,i} \right)^2 \right] \leq 1. \tag{26}
\]

As for the necessary condition for \(\lim_{t \to \infty} E_{H,1}^{(t)} = 0\), we have

\[
\left( \sum_i \rho_{H,i} \right)^2 \left( \sum_j \rho_{L,j} \sqrt{j} \right)^2 + \left( \sum_i \rho_{H,i} \sqrt{i} \right)^2 \left( \sum_j \rho_{L,j} \sqrt{j} \right)^2 \leq 1. \tag{27}
\]

The formal statement is summarized as
Proposition 2: Consider the setting in Example 5, then the necessary conditions in Requirement 1 are given by (26) and (27).

Relaxation of Requirement 2: We obtain the relaxed condition reading as
\[ \sum_{\ell} \lambda_{H,\ell} \sqrt{\ell} \leq C, \]
(28)

Since its derivation is almost the same at that of Example 4, we omit the technical details for the conciseness of presentation.

Example 6 (Preferential sensing with elastic net regularizer): We revisit the setting of sparse signal \( x \) as in Example 4. Instead of \( \ell_1 \) regularizer, we adopt the elastic net regularizer for the signal recovery. Following the same procedure as that in Example 4 and Example 5, we obtain the relaxations of Requirement 1 and 2, which are in the same form of (23), (24), and (25). This is consistent with our findings in the regular sensing setting.

Summarizing the above discussions, we have shown how to transform the constraints in (21) to the closed-forms. Afterwards, we can perform alternating minimization method to solve (19). We can show that the alternating minimization method can reach the local optimal. A formal statement is given as

Proposition 3: Relaxing the constraints in (21) with the above procedure as in Example 4, Example 5, and Example 6, we perform alternating minimization in (19) and denote \( \{\lambda^{(t)}_{H,\ell}, \lambda^{(t)}_{L,\ell}, \rho^{(t)}_{H,i}, \rho^{(t)}_{L,i}\} \) as the solution in the \( t \)th iteration. Then we conclude that \( \{\lambda^{(t)}_{H,\ell}, \lambda^{(t)}_{L,\ell}, \rho^{(t)}_{H,i}, \rho^{(t)}_{L,i}\} \) yields a monotonic non-increasing sequence such that (i) it satisfies
\[ nL \left( \sum_{\ell} \lambda^{(t+1)}_{L,\ell} \right) + nH \left( \sum_{\ell} \lambda^{(t+1)}_{H,\ell} \right) \leq nL \left( \sum_{\ell} \lambda^{(t)}_{L,\ell} \right) + nH \left( \sum_{\ell} \lambda^{(t)}_{H,\ell} \right); \]
and, (ii) has finite limit, i.e.,
\[ \lim_{t \to \infty} \frac{nL \left( \sum_{\ell} \lambda^{(t)}_{L,\ell} \right) + nH \left( \sum_{\ell} \lambda^{(t)}_{H,\ell} \right)}{\sum_{\ell} \rho^{(t)}_{H,i}} < \infty. \]

The technical proof is referred to the Appendix for the interested readers.

V. POTENTIAL GENERALIZATIONS

This section discusses two possible generalizations, i.e., non-exponential family priors and reconstruction via a minimum mean square error (MMSE) decoder. The design principles of the sensing matrix are exactly the same as (9) and (19) except that the DE equations need to be modified.

A. Non-Exponential Priors

Previous sections assume the prior to be \( e^{-f(x)} \), which belongs to the exponential family distributions. In this subsection, we generalize it to arbitrary distributions \( \text{prior}(x) \). One example of the non-exponential distribution is sparse Gaussian, i.e., \((k/n) \cdot e^{-((x \cdot n)^2)/2\sigma^2} + (1 - k/n) \delta(x)\), which is used to model \( k \)-sparse signals. With the generalized prior, the MP in (3) is modified to
\[ m_{i \to a}^{(t+1)}(x_i) \equiv \text{prior}(x_i) \prod_{b \in \partial_i \setminus a} \hat{m}_{b \to i}^{(t)}(x_i); \]
and the decoding step at each iteration becomes
\[ \hat{x}_i^{(t)} = \arg\max_{x_i} \text{P}(x_i | y) \approx \arg\max_{x_i} \text{prior}(x_i) \cdot \prod_{a \in \partial_i} \hat{m}_{a \to i}^{(t)}(x_i). \]
Moreover, the functions \( h_{\text{mean}}(\cdot; \cdot) \) and \( h_{\text{var}}(\cdot; \cdot) \) in (6) are modified to \( \tilde{h}_{\text{mean}}(\cdot; \cdot) \) and \( \tilde{h}_{\text{var}}(\cdot; \cdot) \) as
\[ \tilde{h}_{\text{mean}}(\mu; v) = \gamma \int_{-v}^{v} e^{\gamma \log \text{prior}(x_i)} \cdot e^{-\frac{(x_i \cdot v)^2}{2v^2}} dx_i; \]
\[ \tilde{h}_{\text{var}}(\mu; v) = \gamma \int_{-v}^{v} e^{\gamma \log \text{prior}(x_i)} \cdot e^{-\frac{(x_i \cdot v)^2}{2v^2}} dx_i. \]

B. MMSE Decoder

Notice that both (5) and (30) reconstruct the signal by minimizing the error probability \( \text{P}(\hat{x} \neq x) \), which can be regarded as a MAP decoder. This subsection considers MMSE decoder, which is to minimize the \( \ell_2 \) error, i.e., \( ||\hat{x} - x||_2^2 \). The message-passing procedure stays the same as (29) while the decoding procedure needs to be modified to
\[ \hat{x}_i^{(t)} = \int x_i \text{P}(x_i | y) dx_i \approx \int x_i \cdot \hat{m}_{i \to a}^{(t)}(x_i) dx_i. \]
Moreover, the functions \( h_{\text{mean}}(\cdot; \cdot) \) and \( h_{\text{var}}(\cdot; \cdot) \) in the DE in (6) are modified to \( \tilde{h}_{\text{mean}}(\cdot; \cdot) \) and \( \tilde{h}_{\text{var}}(\cdot; \cdot) \) as
\[ \tilde{h}_{\text{mean}}(\mu; v) = \int_{-v}^{v} \hat{m}_{i \to a}^{(t)}(x_i) \cdot e^{-\frac{(x_i \cdot v)^2}{2v^2}} dx_i; \]
\[ \tilde{h}_{\text{var}}(\mu; v) = \int_{-v}^{v} \hat{m}_{i \to a}^{(t)}(x_i) \cdot e^{-\frac{(x_i \cdot v)^2}{2v^2}} dx_i - \left( \tilde{h}_{\text{mean}}(\mu; v) \right)^2. \]

Having discussed two potential directions of generalization, next we will present the numerical experiments.

VI. NUMERICAL EXPERIMENTS

This section presents the numerical experiments using both synthetic data and real-world data. We consider the sparse signal and compare the design of preferential sensing with that of the
regular sensing. For the simplicity of the code design and the construction of the corresponding sensing matrix, we fix the degrees \( \rho_{H,i} \) and \( \rho_{L,i} \) of the check nodes to \( \rho_{H,dc} = 1 \) and \( \rho_{L,dc} = 1 \), respectively. Therefore, each check node has \( dc_H \) edges connecting to the high-priority part \( x_H^i \) and \( dc_L \) edges connecting to the low-priority part \( x_L^i \).

### A. Sensing Matrix Construction

**Sensing matrix design for sparse signal:** First, we consider the sparse signal setting. We construct the sensing matrix with the algorithm being illustrated in Algorithm 1, which applies to both \( \ell_1 \) regularizer and elastic net regularizer.

We evaluate two types of sensing matrices for the preferential sensing, namely, \( A^{(init)}_{\text{preferential}} \) and \( A^{(final)}_{\text{preferential}} \), which correspond to the distributions \( \{ \lambda_H \} \) and \( \{ \lambda_L \} \) in the initialization phase and at the final outcome of Algorithm 1. As the baseline, we design the sensing matrix \( A_{\text{regular}} \) via (9) which provides regular sensing with an additional constraint which enforces equal edge number with \( A^{(final)}_{\text{preferential}} \) for a fair comparison.

**Sensing matrix design for Gaussian signal:** In addition to the sparse signal, we design the preferential sensing matrix for Gaussian signals. The matrix design algorithm is in the same spirit as Algorithm 1. The only difference is that we replace (23), (24), and (25) with (26), (27), and (28), respectively. Its presentation is omitted due to its similarities to Algorithm 1.

### B. Experiments With Synthetic Data

We study the recovery performance with varying SNR = \( \| x^* \|^2_2 / \sigma^2 \). We separately evaluate the signal recovery performance via the partial and full reconstruction error, which corresponds to the error of the high-priority part \( \| x_H - x_H^* \|_2 \) and that of the whole signal \( \| x - x^* \|_2 \), respectively.

1) **Experiments With Sparse Signal:** We consider the case where \( x^* \) is a \( (k_H + k_L) \)-sparse signal. We fix the check node degrees \( dc_H \) and \( dc_L \) as 5 and let the maximum variable node degree \( dv_{\text{max}} \) as 50. The magnitude of the non-zero entries is set to 1.

   a) **Evaluation under different signal reconstruction methods:** We fix the length \( n_H \) of the high-priority part \( x_H^i \) and \( n_L \) of the low-priority part \( x_L^i \) as 100 and 400, respectively. The corresponding sparsity number \( k_H \) and \( k_L \) are picked as 10 and 10, respectively.

   We consider 3 types of methods: (i) **optimization methods**, e.g., \( \ell_1 \) regularizer (\( \| \cdot \|_1 \)) [2], [4] and elastic net regularizer (\( \| \cdot \|_1 + \| \cdot \|_2^2 \)) [46]; (ii) **greedy methods**, e.g., orthogonal

![Fig. 4. Comparison of preferential sensing vs regular sensing under different signal reconstruction methods. The length \( \lambda_H \) of the high-priority part \( x_H^i \) is set as 100; while the length \( n_L \) of the low-priority part \( x_L^i \) is set as 400. (Top) We evaluate the reconstruction performance w.r.t. the high-priority part \( \| x_H - x_H^* \|_2 / \| x_H^* \|_2 \). (Bottom) We evaluate the reconstruction performance w.r.t. the whole signal \( \| x - x^* \|_2 / \| x^* \|_2 \).](image)
matching pursuit (OMP) [47] and compressive sampling matching pursuit (COSAMP) [48]; and (iii) thresholding-based methods, e.g., iterative hard thresholding (IHT) [49] and hard thresholding pursuit (HTP) [43]. A brief introduction of these algorithms is referred to Chapter 3 in [50]. The simulation results are shown in Fig. 4. Discussion. We show that our design scheme can reduce reconstruction errors with various signal reconstruction methods, despite that our design scheme is rooted in the optimization methods. In addition, we find that different signal reconstruction methods will lead to different errors. A detailed discussion comes as follows.

- **Optimization methods**: In the low-SNR regime, we observe that elastic net regularizer has a better performance than $\ell_1$ regularizer. However, elastic net regularizer’s performance falls behind $\ell_1$ regularizer when entering the high-SNR regime. One intuitive explanation is that the extra $\|\cdots\|^2$ term in the elastic net regularizer promotes reconstructed signal with lower energy, i.e., $\|\hat{x}\|^2_2$.

- **Greedy methods**: We observe that OMP has a slightly better performance than COSAMP. However, their performance are relatively worse than the optimization methods.

- **Thresholding-based methods**: We observe that IHT has the best performance in the low-SNR regime among all signal reconstruction methods. However, its performance is rather steady with varying SNR and is surpassed by other methods when SNR increases. As for HTP, we find that it has a similar performance of $\ell_1$ regularizer.

In summary, we conclude that our designed sensing matrices, both $A^{(\text{init})}_{\text{preferential}}$ and $A^{(\text{final})}_{\text{preferential}}$, have improved performance under all signal reconstruction methods. For the ease of conducting experiments, we will stick to the $\ell_1$ regularizer in the following context as it has the best overall performance. The obtained conclusion should remain valid for other methods.

b) **Impact of sparsity number**: We fix the length $n_H$ of the high-priority part $x^*_H$ as 100 and the length $n_L$ of the low-priority part $x^*_L$ as 400. The simulation results are plotted in Fig. 5.

Discussion: First, we investigate the recovery performance w.r.t. the high priority part $x^*_H$. Using the sensing matrix $A_{\text{regular}}$ (regular sensing) as the baseline, we conclude that our sensing matrix $A^{(\text{final})}_{\text{preferential}}$ (preferential sensing) achieves better performance when the signal is more sparse. Consider the case when SNR = 100. When $k_H = k_L = 10$, the ratio $\|\hat{x}_H - x^*_H\|_2/\|x^*_H\|_2$ for $A^{(\text{final})}_{\text{preferential}}$ is approximately 0.35 while that of the $A_{\text{regular}}$ is 0.86. When the sparsity number $k_H$ and $k_L$ increase to 15, the improvement is approximately $(0.85 - 0.4)/0.85 \approx 53\%$. When the sparsity number $k_H$ and $k_L$ increase to 20, the corresponding improvement further decreases to $(0.95 - 0.55)/0.95 \approx 42\%$.

When turning to the reconstruction error $\|\hat{x} - x^2\|_2/\|x^2\|_2$ w.r.t. the whole signal, we notice a similar phenomenon, i.e., a sparser signal contributes to better performance. Additionally, we notice the sensing matrix $A^{(\text{final})}_{\text{preferential}}$ achieves significant improvements in comparison to its initialized version $A^{(\text{init})}_{\text{preferential}}$.

c) **Impact of signal length**: We also studied various settings in which the length $n_H$ of the high-priority part $x^*_H$ is set to $\{150, 200, 250, 300\}$ and the corresponding length $n_L$ of the low-priority part $x^*_L$ is set to $\{600, 800, 1000, 1200\}$. The simulation results are plotted in Fig. 6. Discussion. Compared to regular sensing, our sensing matrix $A^{(\text{final})}_{\text{preferential}}$ can reduce the error in the high-priority part $x^*_H$ significantly. For example, when SNR = 100, the ratio $\|\hat{x}_H - x^*_H\|_2/\|x^*_H\|_2$ reduces between $40\% - 60\%$ with the sensing matrix $A^{(\text{final})}_{\text{preferential}}$. Meanwhile, w.r.t. the whole signal $x^2$, the ratio $\|\hat{x} - x^2\|_2/\|x^2\|_2$ decreases with a smaller magnitude.

d) **Miscellaneous numerical experiments I**: In addition, we evaluate our designed sensing matrices when the condition $k_H/n_H \geq k_L/n_L$ is violated, or equivalently, we let $k_H/n_H \leq k_L/n_L$. We set the pair $(n_H, k_H)$ as $(400, 10)$ and $(n_L, k_L)$ as $(100, 10)$, respectively. The numerical experiment is put in Fig. 7.

Discussion: We conclude that our designed scheme may still bring performance improvement even if the requirement $k_H/n_H \gg k_L/n_L$ is violated. However, we observe a performance degradation of $A^{(\text{final})}_{\text{preferential}}$ when compared with $A^{(\text{init})}_{\text{preferential}}$. This suggests that Requirement 2 may backfire in protecting the high-priority part $x^*_H$ if $k_H/n_H \leq k_L/n_L$.

e) **Miscellaneous numerical experiments II**: In addition to our proposed scheme, another method for preferential sensing is the *decouple* sensing matrix design (i.e., we separately design the sensing matrices $A_H$ and $A_L$ for the high-priority part $x^*_H$ and low-priority part $x^*_L$; and then stack them together). Then, we compare our designed scheme and the decoupled design scheme, which is denoted as $A_{\text{decouple}}$. For a fair comparison, we enforce the sensor number $m$ and corresponding edge numbers connecting $x^*_H$ and $x^*_L$ to be same. However, there is no interference between the high-priority part $x^*_H$ and low-priority part $x^*_L$, or equivalently, no check node connecting to $x^*_H$ and $x^*_L$.
We evaluate the reconstruction performance w.r.t. the whole signal $\|\hat{x} - x^h\|_2 / \|x^h\|_2$. The prior for $x^h$ is $\propto e^{-\beta H(x)}$ and the prior for $x^l$ is almost the same, which verifies the effectiveness of our design scheme in giving preferential protection of $x^h$.}

### C. Experiments With Real-World Data

This subsection evaluates our designed sensing matrices with the real-world data. We compare the performance of sensing matrices for images using (i) MNIST dataset [51], which consists of 10000 images in the testing set and 60000 images in the training set; and (ii) Lena image. Here we formulate the image representations as sparse signals.

To obtain a sparse representation for each image, we perform a 2D Haar transform $\mathcal{H}(\cdot)$, which generates four sub-matrices being called as the approximation coefficients (at the coarsest level), horizontal detail coefficients, vertical detail coefficients, and diagonal detail coefficients. The approximation coefficients are at the coarsest level and are treated as the high-priority part $x^h$; while the horizontal detail coefficients, vertical detail coefficients, and diagonal detail coefficients are regarded as the low-priority part $x^l$. Hence we can write the sensing relation in (1) as

$$ y = A \mathcal{H}(\text{Image}) + w, \quad \text{(31)} $$

where Image denotes the input image, $\mathcal{H}(\cdot)$ denotes the vectorization of the coefficients and is viewed as the sparse ground-truth signal, and $w$ denotes the sensing noise. The sensing matrix $A$ is designed such that the approximation coefficients of $\mathcal{H}(\text{Image})$ can be better reconstructed.
The degree distributions

\[ \| \mathbf{x}_H^{(\text{init})} \|_2 \] and the entire signal \( \| \mathbf{x} \|_2 \), respectively. We use the sensing matrix \( \mathbf{A}_\text{regular} \) as the benchmark. In addition, we omit the results of \( \mathbf{A}_\text{regular}^{(\text{init})} \) since the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) has better performance.

The results are listed in Table I. A subset of the reconstructed images are shown in Fig. 10. From Table I and Fig. 10, we conclude that our sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) for the preferential sensing can better preserve the images when comparing with the sensing matrix \( \mathbf{A}_\text{regular} \) for the regular sensing.

---

### Table I

| Digit | \( r_{H,(p)} \) | \( r_{H,(r)} \) | \( r_{W,(p)} \) | \( r_{W,(r)} \) | \( r_{H,(p)} \) | \( r_{H,(r)} \) | \( r_{W,(p)} \) | \( r_{W,(r)} \) |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0     | 0.28315        | 0.5154         | 0.44818        | 0.60131        | 0.30292        | 0.45749        | 0.46283        | 0.56486        |
| 1     | 0.16746        | 0.33751        | 0.29332        | 0.41599        | 0.1511         | 0.45264        | 0.2659        | 0.51864        |
| 2     | 0.26303        | 0.50365        | 0.42984        | 0.59959        | 0.24896        | 0.42216        | 0.52556        |
| 3     | 0.24613        | 0.43677        | 0.42514        | 0.53163        | 0.26446        | 0.46766        | 0.43534        | 0.56189        |
| 4     | 0.28351        | 0.44377        | 0.44623        | 0.53791        | 0.30092        | 0.44545        | 0.48504        | 0.53749        |
| 5     | 0.28405        | 0.53511        | 0.45727        | 0.6198         | 0.27258        | 0.47044        | 0.44382        | 0.56622        |
| 6     | 0.28801        | 0.39436        | 0.45053        | 0.51701        | 0.27084        | 0.5086         | 0.44134        | 0.59534        |
| 7     | 0.25953        | 0.41621        | 0.41809        | 0.52896        | 0.27266        | 0.51329        | 0.41693        | 0.5783         |
| 8     | 0.31263        | 0.51918        | 0.47618        | 0.61492        | 0.32731        | 0.48163        | 0.48699        | 0.5837         |
| 9     | 0.30171        | 0.54394        | 0.45241        | 0.61799        | 0.27385        | 0.55313        | 0.6277        | 0.62785        |

We define the ratio \( r_{H,(p)} \) (\( i = (p,r) \)) as the error W.R.T. the high priority part, namely, \( \| \mathbf{x}_H - \mathbf{x}_H^{(\text{init})} \|_2 / \| \mathbf{x}_H \|_2 \). Similarly we define the ratio \( r_{W,(p)} \) (\( i = (p,r) \)) as the ratio W.R.T. the whole signal. Moreover, we put the results corresponding to the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) in the bold font.

---

Fig. 10. The performance comparison between the sensing matrix for preferential sensing \( \mathbf{A}_\text{preferential}^{(\text{final})} \) and sensing matrix for regular sensing \( \mathbf{A}_\text{regular} \). (Top) The ground-truth images. (Middle) The reconstructed images with the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \). (Bottom) The reconstructed images with the sensing matrix \( \mathbf{A}_\text{regular} \).

1) Experiments With MNIST: We set the images from MNIST as the input, which consists of 10000 images in the testing set and 60000 images in the training set with each image being of dimension 28 × 28.

The whole datasets can be divided into 10 categories with each category representing a digit from zero to nine. For each digit, we design one unique sensing matrix. The lengths \( n_H \) and \( n_L \) are set to \( (28/2)^2 = 196 \) and \( 3 \times (28/2)^2 = 588 \), respectively. The sparsity coefficients \( k_H \) and \( k_L \) varied among different digits.

**Discussion:** To evaluate the performance, we define ratios \( r_{H,(\cdot)} \) and \( r_{W,(\cdot)} \) as

\[
\begin{align*}
r_{H,(\cdot)} & \triangleq \frac{\| \mathbf{x}_H - \mathbf{x}_H^{(\text{init})} \|_2}{\| \mathbf{x}_H \|_2}, \\
r_{W,(\cdot)} & \triangleq \frac{\| \mathbf{x} - \mathbf{x}_H \|_2}{\| \mathbf{x} \|_2},
\end{align*}
\]

which correspond to the \( \ell_2 \) error in the high-priority part \( \mathbf{x}_H^{(\text{init})} \) and the entire signal \( \mathbf{x} \), respectively. We use the sensing matrix \( \mathbf{A}_\text{regular} \) as the benchmark. In addition, we omit the results of \( \mathbf{A}_\text{init}^{(\text{init})} \) since the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) has better performance.

The results are listed in Table I. A subset of the reconstructed images are shown in Fig. 10. From Table I and Fig. 10, we conclude that our sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) for the preferential sensing can better preserve the images when comparing with the sensing matrix \( \mathbf{A}_\text{regular} \) for the regular sensing.

2) Experiments With Lena Image: We evaluate the benefits of using \( \mathbf{A}_\text{preferential}^{(\text{final})} \) for the Lena image with dimension 512 × 512. Notice that the sensing matrix would have been prohibitively large if we used the whole image as the input. To put more specifically, we would need a matrix with the width \( 512^2 = 262144 \). To handle such issue, we divide the whole images into a set of sub-blocks with dimensions 32 × 32 and design one sensing matrix with the width \( 32^2 = 1024 \). For each sub-block, we first obtain a sparse representation with a 2D Haar transform and then reconstruct the signal in (31).

**Discussion:** The comparison of results is plotted in Fig. 11, from which we conclude that the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) has much better performance in image reconstruction in comparison with the sensing matrix \( \mathbf{A}_\text{regular} \). The ratios \( r_{H,(p)} \) and \( r_{H,(r)} \) are computed as 0.0446 and 0.3029, respectively; while the ratio \( r_{W,(p)} \) and \( r_{W,(r)} \) are computed as 0.0709 and 0.3144, respectively.

**Remark 4:** The degree distributions \( \lambda_H(\cdot) \) and \( \lambda_L(\cdot) \) of the variable nodes for the sensing matrix \( \mathbf{A}_\text{preferential}^{(\text{final})} \) are obtained as

\[
\begin{align*}
\lambda_H(\alpha) & = 0.0057856\alpha + 0.025915\alpha^2 + 0.36394\alpha^3 + 0.35183\alpha^4 \\
& + 0.10333\alpha^5 + 0.04134\alpha^6 + 0.021619\alpha^7 + 0.013508\alpha^8 \\
& + 0.0093743\alpha^9 + 0.0070966\alpha^{10} + 0.00560\alpha^{11} \\
& + 0.0045851\alpha^{12} + 0.0038574\alpha^{13} + 0.0033145\alpha^{14}.
\end{align*}
\]
\[ +0.0028963\alpha_{15} + 0.0025659\alpha_{16} + 0.0022992\alpha_{17} \\
+0.0020801\alpha_{18} + 0.0018973\alpha_{19} + 0.0017428\alpha_{20} \\
+0.0016109\alpha_{21} + 0.001497\alpha_{22} + 0.001398\alpha_{23} \\
+0.0013111\alpha_{24} + 0.0012344\alpha_{25} + 0.0011662\alpha_{26} \\
+0.0011053\alpha_{27} + 0.0010013\alpha_{28} + 0.00091576\alpha_{29} \\
+0.0009565\alpha_{30} + 0.00091576\alpha_{31} + 0.00084436\alpha_{32} \\
+0.00084436\alpha_{33} + 0.00081292\alpha_{34} + 0.00073888\alpha_{35} \\
+0.00075697\alpha_{36} + 0.00073197\alpha_{37} + 0.00070867\alpha_{38} \\
+0.00068691\alpha_{39} + 0.0006652\alpha_{40} + 0.00064738\alpha_{41} \\
+0.00062937\alpha_{42} + 0.00061238\alpha_{43} + 0.00059633\alpha_{44} \\
+0.00058114\alpha_{45} + 0.00056673\alpha_{46} + 0.00055304\alpha_{47} \\
+0.00054001\alpha_{48} + 0.0005276\alpha_{49}; \]

\[ \lambda_L(\alpha) = \alpha. \]

The check node degrees \( d_{c_H} \) and \( d_{c_L} \) are both set as 4. Meanwhile, the sensing matrix \( \mathbf{A}_{\text{regular}} \) designed in (9) is a regular sensing matrix whose variable node and check node degree distributions are given by \( \lambda(\alpha) = \alpha^2 \) and \( \rho(\alpha) = \alpha^2 \), respectively.

VII. CONCLUSION

This paper presented a general framework of the sensing matrix design for a linear measurement system. Focusing on a sparse sensing matrix \( \mathbf{A} \), we associated it with a graphical model \( G = (\mathcal{V}, \mathcal{E}) \) and transformed the design of \( \mathbf{A} \) to the connectivity problem in \( \mathcal{G} \). With the density evolution technique, we proposed two design strategies, i.e., regular sensing and preferential sensing. In the regular sensing scenario, all entries of the signal are recovered with equal accuracy; while in the preferential sensing scenario, the entries in the high-priority sub-block are recovered more accurately (or exactly) relative to the entries in the low-priority sub-block. We then analyzed the impact of the connectivity of the graph on the recovery performance. For the regular sensing, our framework can reproduce the classical results for both the sparse signals and Gaussian signals. For the preferential sensing, our framework can lead to a significant reduction of the reconstruction error in the high-priority part. Numerical experiments with both synthetic data and real-world data are presented to corroborate our claims.

**APPENDIX A**

**PROOF OF THEOREM 1**

**Proof:** We begin the proof by restating the DE equation w.r.t. \( E^{(t+1)} \) and \( V^{(t+1)} \) as

\[
E^{(t+1)} = \mathbb{E}_{\text{prior}(s), z \sim N(0,1)} \left[ \text{prox} \left( s + a_1 z \sqrt{E^{(t)}}; \beta a_2 V^{(t)} \right) - s \right]^2; \\
= \Psi_E(E^{(t)}, V^{(t)})
\]

\[
V^{(t+1)} = \mathbb{E}_{\text{prior}(s), z \sim N(0,1)} \left[ \beta a_2 V^{(t)} \text{prox} \left( s + a_1 z \sqrt{E^{(t)}}; \beta a_2 V^{(t)} \right) \right]; \\
= \Psi_V(E^{(t)}, V^{(t)})
\]

The derivation of the necessary conditions for \( \lim_{t \to \infty} (E^{(t)}, V^{(t)}) = (0, 0) \) consists of two parts:

- **Part I:** We verify that \( (0, 0) \) is a fixed-point of the DE equation;
- **Part II:** We consider the necessary condition such that DE equation converges within the proximity of the origin points, i.e., \( E^{(t)} \) and \( V^{(t)} \) are close to zero.

Since Part I can be easily verified, we put our major focus on Part II. Define the difference across iterations as \( \delta_E^{(t)} = E^{(t+1)} - E^{(t)} \) and \( \delta_V^{(t)} = V^{(t+1)} - V^{(t)} \), we would like to show \( \lim_{t \to \infty} (\delta_E^{(t)}, \delta_V^{(t)}) = (0, 0) \). With Taylor expansion, we obtain

\[
\delta_E^{(t+1)} = \Psi_E \left( E^{(t+1)}, V^{(t+1)} \right) - \Psi_E \left( E^{(t)}, V^{(t)} \right) \\
= \left( \frac{\partial \Psi_E(E, V)}{\partial E} \right)_{E=E^{(t)}, V=V^{(t)}} \cdot \delta_E^{(t)} \\
+ \left( \frac{\partial \Psi_E(E, V)}{\partial V} \right)_{E=E^{(t)}, V=V^{(t)}} \cdot \delta_V^{(t)} \\
+ O \left( \delta_E^{(t)} \right)^2 + O \left( \delta_V^{(t)} \right)^2. \tag{32}
\]

Consider the region where \( \delta_E^{(t)} \) and \( \delta_V^{(t)} \) are sufficiently small, we require \( \delta_E^{(t)} \) and \( \delta_V^{(t)} \) to converge to zero. Notice the quadratic terms in (32) can be safely omitted in this region. Denote the gradients \( \frac{\partial \Psi_E(E, V)}{\partial E} \), \( \frac{\partial \Psi_E(E, V)}{\partial E} \), \( \frac{\partial \Psi_E(E, V)}{\partial V} \), and \( \frac{\partial \Psi_E(E, V)}{\partial V} \), then we have

\[
\frac{\partial \Psi_E(E, V)}{\partial E} \bigg|_{E=E^{(t)}, V=V^{(t)}}, \quad \frac{\partial \Psi_E(E, V)}{\partial V} \bigg|_{E=E^{(t)}, V=V^{(t)}}, \quad \frac{\partial \Psi_E(E, V)}{\partial E} \bigg|_{E=E^{(t)}, V=V^{(t)}}, \quad \frac{\partial \Psi_E(E, V)}{\partial V} \bigg|_{E=E^{(t)}, V=V^{(t)}}
\]
\[ \left( \frac{\partial \Psi}{\partial E}(E,V) \right)^{(t)} \text{ and } \left( \frac{\partial \Psi}{\partial V}(E,V) \right)^{(t)} \text{ respectively. We obtain the linear equation} \]

\[ \left[ \begin{array}{c}
\delta^{(t+1)} \\
\delta V^{(t+1)}
\end{array} \right] = \left[ \begin{array}{c}
\left( \frac{\partial \Psi}{\partial E}(E,V) \right)^{(t)} \\
\left( \frac{\partial \Psi}{\partial V}(E,V) \right)^{(t)}
\end{array} \right] \left[ \begin{array}{c}
\delta E^{(t)} \\
\delta V^{(t)}
\end{array} \right], \]

and would require the lower bound of the operator norm of the matrix \( \mathbf{L}^{(t)} \) to be no greater than 1, i.e., \( \inf_{\mathbf{L}^{(t)}} \|[\mathbf{L}^{(t)}]\|_{\text{op}} \leq 1 \), since otherwise the values of \( \delta E^{(t)} \) and \( \delta V^{(t)} \) will keep increasing. Exploiting the fact \( \frac{\partial \Psi}{\partial E} = 0 \), we conclude

\[ \|\mathbf{L}^{(t)}\|_{\text{op}} = \max \left[ \left( \frac{\partial \Psi}{\partial E}(E,V) \right)^{(t)}, \left( \frac{\partial \Psi}{\partial V}(E,V) \right)^{(t)} \right]. \]

The proof is then concluded by computing the lower bounds of the gradients \( \frac{\partial \Psi}{\partial E}(E,V) \) and \( \frac{\partial \Psi}{\partial V}(E,V) \) as

\[ \frac{\partial \Psi}{\partial E}(E,V) \bigg|_{E=E_{(t)}, V=V_{(t)}} = a^2 E_{(t)} \cdot \mathbb{E}_{\text{prior}}(s) \left[ \Phi \left( \frac{s + a_2 V(t)}{\sqrt{a_1 V(t)}} \right) + \Phi \left( \frac{s - a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \right] \]

\[ \geq \frac{a^2 k}{n} \left[ \Phi \left( \frac{c_0 + a_2 V(t)}{\sqrt{a_1 V(t)}} \right) + \Phi \left( \frac{c_0 - a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \right] + 2a^2 \left( \frac{1 - \beta}{n} \right) \Phi \left( \frac{-a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \geq \frac{a^2 k}{n} ; \]

\[ \frac{\partial \Psi}{\partial V}(E,V) \bigg|_{E=E_{(t)}, V=V_{(t)}} = \beta a_2 E_{(t)} \cdot \mathbb{E}_{\text{prior}}(s) \left[ \Phi \left( \frac{s + a_2 V(t)}{\sqrt{a_1 V(t)}} \right) + \Phi \left( \frac{s - a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \right] \]

\[ \geq \frac{\beta a_2 k}{n} \left[ \Phi \left( \frac{c_0 + a_2 V(t)}{\sqrt{a_1 V(t)}} \right) + \Phi \left( \frac{c_0 - a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \right] + 2\beta a_2 \left( \frac{1 - \beta}{n} \right) \Phi \left( \frac{-a_2 V(t)}{\sqrt{a_1 V(t)}} \right) \geq \frac{\beta a_2 k}{n} , \]

where \( \Phi(\cdot) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-z^2/2} dz \) is the CDF of the standard normal RV \( z \), namely, \( z \sim N(0,1) \). In (1) and (4) we use the prior distribution \( \text{prior}(s) = \frac{k}{n} \cdot 1(c_0) + (1 - k/n) 1(0) \). Further, in (2) and (5) we use the fact

\[ \lim_{E_{(t)} \to 0} \Phi \left( \frac{c_0 + a_2 V(t)}{\sqrt{a_1 E(t)}} \right) + \Phi \left( \frac{c_0 - a_2 V(t)}{\sqrt{a_1 E(t)}} \right) = 1, \]

since \( c_0 \neq 0 \). Finally, in (3) and (6) we omit the non-negative terms \( \Phi(\cdot) \).

**APPENDIX B**

**PROOF OF THEOREM 2**

We begin the proof by restating that the functions \( \mathbb{E}_{\vartheta} h_{\text{mean}}(\cdot, \cdot) \) and \( \mathbb{E}_{\vartheta} h_{\text{var}}(\text{mean; var}) \) are written as

\[ \mathbb{E}_{\vartheta} h_{\text{mean}} \left( s + a_1 z \sqrt{E(t)}; a_2 V(t) \right) = \frac{a_2^2 E(t) + a_2^2 (V(t))^2}{(1 + a_2 V(t))^2} ; \]

\[ \mathbb{E}_{\vartheta} h_{\text{var}} \left( s + a_1 z \sqrt{E(t)}; a_2 V(t) \right) = \frac{a_2^2 V(t)}{1 + a_2 V(t)} , \]

which can be easily verified. Then, we prove that \( V(t) \) decreases exponentially since \( a_2 > 0 \) and hence for an arbitrary time index \( T_1 \) the relation

\[ V(t) \leq \frac{a_2}{1 + a_2} \left( t - T_1 \right) V(T_1) = e^{-c_1 (t - T_1)} V(T_1) \]

holds for \( t \geq T_1 \), where \( c_1 \) is defined as \( \log(1 + a_2) > 0 \).

Afterwards, we study the behavior of \( E(t) \). Denote \( V_S \) as \( \mathbb{E}_{\text{prior}}(s^2) \), we have

\[ E(t+1) \leq a_1^2 E(t) + \frac{a_2 V_S (V(t))}{2} \leq a_1^2 E(t) + \frac{a_2 V_S (V(t))}{2} \left( \frac{a_2}{1 + a_2} \right)^t V(0), \]

where in (1) we use the relation \( V(t) \leq (a_2/(1 + a_2))^t V(0) \). Define a new sequence \( \bar{E}(t) = E(t)/a_1^2 \), we can transform (34) to

\[ \bar{E}(t+1) = \bar{E}(t) + \frac{a_2 V_S (V(t))}{2a_1^2} \left( \frac{a_2}{1 + a_2} \right)^t ; \]

after rearranging the terms. Due to the time-variation, we also have the relation

\[ \bar{E}(t) \leq \bar{E}(t-1) + \frac{a_2 V_S (V(t))}{2a_1^2} \left( \frac{a_2}{1 + a_2} \right)^t \cdot \left( \frac{1}{(1 + a_2) a_1^t} \right)^{t-1} . \]

Iterating over all such inequalities, we obtain the equation

\[ \bar{E}(t+1) \leq \bar{E}(1) + \frac{a_2 V_S (V(0))}{2a_1^2} \frac{a_2}{1 + a_2} \left( \frac{1}{(1 + a_2) a_1^t} \right)^t \cdot \left( \frac{1}{1 - \frac{a_2}{(1 + a_2) a_1^t}} \right)^t , \]

which leads to

\[ E(t+1) \leq a_1^2 E(1) + \frac{a_2 V_S (V(0))}{2a_1^2} \left( \frac{a_2}{1 + a_2} - \frac{a_2}{(1 + a_2) a_1^t} \right)^t \cdot \left( \frac{1}{1 - \frac{a_2}{(1 + a_2) a_1^t}} \right)^t . \]

Since \( a_1 < 1 \) and \( a_2/(1 + a_2) < 1 \), we have the second term \( \vartheta \) in (35) to be negligible as \( t \) goes to infinity. Hence we can choose a sufficiently large \( T \) such that for \( t \geq T \), we have \( E(t+1) \) is approximately equal to \( a_1^2 E(1) \) and conclude the exponential decay of \( E(t) \).
**APPENDIX C**

**PROOF OF THEOREM 3**

To begin with, we briefly discuss how to derive the DE equation for the elastic net regularization, to put more specifically how to compute the corresponding functions $\hat{h}_{\text{mean}}(\cdot)$ and $\hat{h}_{\text{var}}(\cdot)$. Recalling their definitions in (8), our goal is to study the probability distribution $\exp[-\gamma(\beta |x| + \beta x^2 + (x - \mu)^2/2\nu)]$. Denote $\bar{\mu}$ and $\bar{v}$ as $\mu/1 + 2\beta v$ and $\nu/1 + 2\beta v$, respectively, we can show that the above distribution is equivalent to $\exp(-\gamma(\bar{x}^2/\bar{v}) - \beta |\bar{x}|)$, which is of the similar form associated with the $\ell_1$ regularizer. Following the same procedure then yields the corresponding DE equation. For the notation simplicity, we denote the DE equation as $(E(t+1), V(t+1)) = (\Psi_E(E(t); V(t)), \Psi_V(E(t); V(t)))$.

Then, we study the necessary conditions of $\lim_{t \to \infty} (E(t), V(t)) = (0, 0)$. Following the same procedure as in Section A, we define matrix $L(t)$ as

$$
L(t) = \begin{bmatrix}
\frac{\partial \Psi_E(E, V)}{\partial E} |_{E=E(t), V=V(t)} \\
\frac{\partial \Psi_E(E, V)}{\partial V} |_{E=E(t), V=V(t)} \\
\frac{\partial \Psi_V(E, V)}{\partial E} |_{E=E(t), V=V(t)} \\
\frac{\partial \Psi_V(E, V)}{\partial V} |_{E=E(t), V=V(t)}
\end{bmatrix},
$$

and require $\inf_{t} \|L(t)\|_{\text{op}} \leq 1$. With some standard calculations, we have

$$
\|L(t)\|_{\text{op}} = \max \left[ \left( \frac{\partial \Psi_E(E, V)}{\partial E} |_{E=E(t), V=V(t)} \right)^2, \left( \frac{\partial \Psi_V(E, V)}{\partial V} |_{E=E(t), V=V(t)} \right)^2 \right].
$$

(36)

We conclude the proof by computing the lower bounds of $\frac{\partial \Psi_E(E, V)}{\partial E}$ and $\frac{\partial \Psi_V(E, V)}{\partial V}$ around $(0, 0)$, which proceeds as follows. Following the same procedure as in Section A, we have

$$
\frac{\partial \Psi_E(E, V)}{\partial E} |_{E=E(t), V=V(t)} = \frac{a_2^2 \cdot \vartheta_1}{1 + 2a_2^2 V(t)^2} + \frac{a_1 \cdot \vartheta_2}{1 + 2a_2^2 V(t)^2 \sqrt{2\pi E(t)}},
$$

where $\vartheta_1$ and $\vartheta_2$ are defined as

$$
\vartheta_1 \triangleq E_{\text{prior}(s)} \left[ \Phi \left( \frac{s - a_2^2 V(t)}{a_1 \sqrt{E(t)}} \right) + \Phi \left( \frac{s + a_2^2 V(t)}{a_1 \sqrt{E(t)}} \right) \right],
$$

$$
\vartheta_2 \triangleq E_{\text{prior}(s)} s \cdot \left( \exp \left[ -\frac{(s + a_2^2 V(t))^2}{2a_1^2 E(t)} \right] - \exp \left[ -\frac{(s - a_2^2 V(t))^2}{2a_1^2 E(t)} \right] \right),
$$

respectively. Plugging $\text{prior}(s) = k/n \cdot 1(0) + (1 - k/n)1(\xi_0)$ into $\vartheta_1$ and $\vartheta_2$ then yields

$$
\lim_{(E(t), V(t)) \to (0, 0)} \frac{\partial \Psi_E(E, V)}{\partial E} |_{E=E(t), V=V(t)} \geq \frac{k \cdot a_2^2}{n}, \quad (37)
$$

As for $\frac{\partial \Psi_V}{\partial V}$, we have

$$
\frac{\partial \Psi_V}{\partial V} |_{E=E(t), V=V(t)} = \frac{\beta \cdot a_2 \vartheta_1}{1 + 2\beta \cdot a_2^2 V(t)^2},
$$

which yields

$$
\lim_{(E(t), V(t)) \to (0, 0)} \frac{\partial \Psi_V(E, V)}{\partial V} |_{E=E(t), V=V(t)} \geq \frac{k \beta a_2^2}{n}. \quad (38)
$$

Thus, we complete the proof by combing (36), (37), and (38); and letting $\inf_t \|L(t)\|_{\text{op}} \leq 1$.

**APPENDIX D**

**PROOF OF PROPOSITION 3**

Without loss of generality, we assume the updating order is $\{\lambda_H^{(t)}, \lambda_L^{(t)}, \rho_H^{(t)}, \rho_L^{(t)}\}$. For the notation simplicity, we denote the DE equation as (19) as $\text{OPT}(\lambda_H^{(t)}, \lambda_L^{(t)}, \rho_H^{(t)}, \rho_L^{(t)})$. Easily, we can verify (19) is a convex optimization problem w.r.t. $\lambda_L$, $\rho_H$, and $\rho_L$. This results in

$$
\text{OPT}(\lambda_H^{(t)}, \lambda_L^{(t)}, \rho_H^{(t)}, \rho_L^{(t)}) \geq \text{OPT}(\lambda_H^{(t+1)}, \lambda_L^{(t+1)}, \rho_H^{(t+1)}, \rho_L^{(t+1)}),
$$

which completes the proof such that $\{\lambda_H^{(t)}, \lambda_L^{(t)}, \rho_H^{(t)}, \rho_L^{(t)}\}$ constitutes a monotonic non-increasing sequence. Combining with the fact such that (19) is non-negative, we can show that $\{\lambda_H^{(t)}, \lambda_L^{(t)}, \rho_H^{(t)}, \rho_L^{(t)}\}$ has a finite limit, i.e.,

$$
\lim_{t \to \infty} \sum_{i=1}^{n}(\lambda_H^{(t)} + \lambda_L^{(t)} + \rho_H^{(t)} + \rho_L^{(t)}) < \infty.
$$

**REFERENCES**

[1] H. Zhang, A. Abdi, and F. Fekri, “A general framework for the design of compressive sensing using density evolution,” in Proc. IEEE Inf. Theory Workshop, pp. 1–6.

[2] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Sparse recovery from overcomplete representations in the presence of noise,” IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[3] E. J. Candes, J. K. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] M. Mezard and A. Montanari, Information, Physics, and Computation. Oxford, U.K.: Oxford Univ. Press, 2009.

[5] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Amsterdam, The Netherlands: Elsevier, 2014.

[6] M. Heinrich and M. Moeller, “Low-density parity-check codes,” IRE Trans. Inf. Theory, vol. 8, no. 1, pp. 21–28, 1962.

[7] R. Gallager, “Low-density parity-check codes,” IRE Trans. Inf. Theory, vol. 8, no. 1, pp. 21–28, 1962.

[8] C. Berrou and A. Glavieux, “Near optimum error correcting coding: Turbo-codes,” IEEE Trans. Commun., vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[9] J. F. Cheng, “Turbo decoding as an instance of Pearl’s “belief propagation” algorithm,” IEEE J. Sel. Areas Commun., vol. 16, no. 2, pp. 140–152, Feb. 1998.

[10] J. T. Richardson and R. L. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[11] S. Sarvotham, D. Baron, and R. G. Baraniuk, “Compressed sensing reconstruction via belief propagation,” preprint, 2006.
F. Zhang and H. D. Pfister, “Verification decoding of high-rate LDPC codes with applications in compressed sensing,” IEEE Trans. Inf. Theory, vol. 58, no. 8, pp. 4042–4058, Aug. 2012.

S. Kudekar and H. D. Pfister, “The effect of spatial coupling on compressive sensing,” in Proc. 48th Annu. Allerton Conf. Commun., Control, Comput., 2010, pp. 347–353.

Y. Eftekhari, A. Heidarzadeh, A. H. Banishahemi, and I. Lambadaris, “Density evolution analysis of node-based verification-based algorithms in compressed sensing,” IEEE Trans. Inf. Theory, vol. 58, no. 10, pp. 6616–6645, Oct. 2012.

F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Statistical-physics-based reconstruction in compressed sensing,” Phys. Rev. X, vol. 2, no. 2, 2012, Art. no. 021005.

F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices,” J. Stat. Mechanics: Theory Experiment, vol. 2012, no. 08, 2012, Art. no. P08009.

L. Zdeborová and F. Krzakala, “Statistical physics of inference: Thresholds and algorithms,” Adv. Phys., vol. 65, no. 5, pp. 453–552, 2016.

D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” Proc. Natl. Acad. Sci., vol. 106, no. 45, pp. 18911–18919, 2009.

M. A. Maleki, Approximate Message Passing Algorithms for Compressed Sensing, Redwood, CA, USA: Stanford Univ., 2010.

S. Sarvotham, D. Baron, and R. G. Baraniuk, “Sudocodes fast measurement and reconstruction of sparse signals,” in Proc. IEEE Int. Symp. Information Theory, pp. 2804–2808, 2008.

D. Baron, S. Sarvotham, and R. G. Baraniuk, “Bayesian compressive sensing via belief propagation,” IEEE Trans. Signal Process., vol. 58, no. 1, pp. 269–280, Jan. 2010.

V. Chandar, D. Shah, and G. W. Wornell, “A simple message-passing algorithm for compressed sensing,” in Proc. IEEE Int. Symp. Information Theory, 2010, pp. 1968–1972.

A. G. Dimakis, R. Smarandache, and P. O. Vontobel, “LDPC codes for compressed sensing,” IEEE Trans. Inf. Theory, vol. 58, no. 5, pp. 3093–3114, May 2012.

M. G. Luby and M. Mitzenmacher, “Verification-based decoding for packet-based low-density parity-check codes,” IEEE Trans. Inf. Theory, vol. 51, no. 1, pp. 120–127, Jan. 2005.

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.

D. Donoho and A. Montanari, “High dimensional robust m-estimation: Asymptotic variance via approximate message passing,” Probability Theory Related Fields, vol. 166, no. 3–4, pp. 953–969, 2016.

A. Montanari, “Graphical models concepts in compressed sensing,” in Compressed Sensing Theory and Applications, Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 394–438.

W. Xu and B. Hassibi, “Efficient compressive sensing with deterministic guarantees using expander graphs,” in Proc. IEEE Inf. Theory Workshop, 2007, pp. 414–419.

W. Xu and B. Hassibi, “Further results on performance analysis for compressive sensing using expander graphs,” in Proc. Conf. Rec. 41st Asilomar Conf. Signals, Syst. Comput., 2007, pp. 621–625.

M. A. Khajehnejad, A. G. Dimakis, and B. Hassibi, “Nonnegative compressed sensing with minimal perturbed expanders,” in Proc. IEEE 13th Digit. Signal Process. Workshop 5th IEEE Signal Process. Educ. Workshop, 2009, pp. 696–701.

S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, “Efficient and robust compressed sensing using optimized expander graphs,” IEEE Trans. Inf. Theory, vol. 55, no. 9, pp. 4299–4308, Sep. 2009.

W. Lu, K. Kpalma, and J. Ronnias, “Sparse binary matrices of LDPC codes for compressed sensing,” in Proc. Data Compression Conf., 2012, pp. 1–10.

J. Zhang, G. Han, and Y. Fang, “Deterministic construction of compressed sensing matrices from projective LDPC codes,” IEEE Signal Process. Lett., vol. 22, no. 11, pp. 1960–1964, Nov. 2015.

A. Mousavi, G. Dasarathy, and R. G. Baraniuk, “Deepcodec: Adaptive sensing and recovery via deep convolutional neural networks,” in Proc. 55th Annu. Allerton Conf. Commun., Control, Comput., 2017, pp. 744–744.

T. Richardson and R. Urbanke, Modern Coding Theory, Cambridge, U.K.: Cambridge Univ. Press, 2008.

T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, “Design of capacity-achieving irregular low-density parity-check codes,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 619–637, Feb. 2001.

S.-Y. Chung, “On the construction of some capacity-approaching coding schemes,” Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2000.

S. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” Found. Comput. Math., vol. 12, no. 6, pp. 805–849, 2012.

R. Tibshirani, “Regression shrinkage and selection via the Lasso,” J. Roy. Stat. Soc.: Ser. B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

A. E. Hoerl and R. W. Kennard, “Ridge regression: Applications to nonorthogonal problems,” Technometrics, vol. 12, no. 1, pp. 69–82, 1970.

S. Foucart, “Hard thresholding pursuit: An algorithm for compressive sensing,” SIAM J. Numer. Anal., vol. 49, no. 6, pp. 2543–2563, 2011.

T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, (Springer Series in Statistics Series). New York, NY, USA: Springer New York Inc., 2001.

G. B. Arfken and H. J. Weber, “Mathematical methods for physicists,” Academic Press, 1999.

H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.), vol. 67, no. 2, pp. 301–320, 2005.

J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, 2009.

T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” Appl. Comput. Harmon. Anal., vol. 27, no. 3, pp. 265–274, 2009.

S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing (Ser. Applied and Numerical Harmonic Analysis). New York: Springer, 2013.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

Hang Zhang received the Ph.D. degree in ECE from the Georgia Institute of Technology, Atlanta, GA, USA, in 2021. He was with Baidu Research and later moved to Amazon, Inc. His research interests include high-dimensional statistics, cost-accuracy trade-offs, robust, and efficient algorithms. He was the recipient of the outstanding research award in Georgia Tech Center for Signal and Information Processing.

Afshin Abdli received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2020. Since then, he has been with Qualcomm Technologies, Inc., as an ML Research Engineer. His research interests include distributed machine learning, model compression, and applications of machine learning in coding and wireless/5G communication.