A General Compressive Sensing Construct using Density Evolution

Hang Zhang, Afshin Abdi, and Faramarz Fekri
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Abstract*—This paper [1] proposes a general framework to design a sparse sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, for a linear measurement system $\mathbf{y} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$, where $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x}^\natural \in \mathbb{R}^n$, and $\mathbf{w} \in \mathbb{R}^m$ 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 in the design of low density parity check codes, 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 illustrations, we consider the ℓ_1 regularizer, ℓ_2 regularizer, and their linear combination, which corresponds to Lasso regression, ridge regression, and elastic net regression. After proper distribution approximations, we have shown 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 $\boldsymbol{x}_{\mathsf{H}}^\natural$ and low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$. Then, by formulating the sensing system design as a bi-convex optimization problem, we obtain sensing matrices which can provide a preferential treatment for $\boldsymbol{x}_{\mathsf{H}}^\natural$. Numerical experiments with both synthetic data and real-world data are also provided to verify the effectiveness of our design scheme.

I. Introduction

This paper considers a linear sensing relation as

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^\natural + \boldsymbol{w}, \tag{1}$$

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

$$\widehat{\boldsymbol{x}} = \text{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \ \frac{1}{2\sigma^2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x}), \tag{2}$$

where $f(\cdot)$ is the regularizer used to enforce a desired structure for $\widehat{\boldsymbol{x}}$. To ensure reliable recovery of $\boldsymbol{x}^\natural$, sensing matrix $\mathbf{A}$ needs to satisfy certain conditions, e.g., the incoherence in [2], RIP in [3], [4], the *neighborhood stability* in [5], *irrespresentable condition* in [6], etc. Notice that all the above works treat each entry of $\boldsymbol{x}^\natural$ equally. However, in certain applications, entries of $\boldsymbol{x}^\natural$ 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. [2]

In this work, we focus on the sparse sensing matrix $\mathbf{A}$. Leveraging tools from coding theory, namely, *density evolution* (DE), we propose a heuristic but general design framework of $\mathbf{A}$ to meet the requirements of the signal reconstruction such as placing more importance on the accuracy of a 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, Gallagher 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]–[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 $\boldsymbol{x}^\natural$ 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 mentioned works focused on the noiseless setting, i.e., $\boldsymbol{w} = \mathbf{0}$ in (1). In [17]–[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),

[1]Partial preliminary results appeared in 2021 IEEE Information Theory Workshop [1].

[2]An introduction can be found in https://jpeg.org/jpeg/documentation.html.

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 scarifice 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]–[36].

Contributions. Compared to the previous works exploiting MP [14]–[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 $\boldsymbol{x}^\natural \in \mathbb{R}^n$ 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., $m \geq c_0 k \log n$. 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.

II. Problem Description

We begin this section with a formal statement of our problem. Consider the linear measurement system

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^\natural + \boldsymbol{w},$$

where $\boldsymbol{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{x}^\natural \in \mathbb{R}^n$, and $\boldsymbol{w} \in \mathbb{R}^m$, respectively, denote the observations, the sensing matrix, the signal, and the additive sensing noise with its ith entry $w_i \overset{\text{i.i.d}}{\sim} \mathsf{N}(0,\sigma^2)$. We would like to recover $\boldsymbol{x}^\natural$ with the regularized M-estimator, which is written as

$$\widehat{\boldsymbol{x}} = \text{argmin}_{\boldsymbol{x}} \frac{1}{2\sigma^2}\|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x}),$$

where $f(\cdot)$ is the regularizer used to enforce certain underlying structure for signal $\widehat{\boldsymbol{x}}$.

Our goal is to design a sparse sensing matrix $\mathbf{A}$ which provides preferential treatment for a sub-block of the signal $\boldsymbol{x}^\natural$. In other words, the objective is to have a sub-block of the signal to be recovered with lower probability of error when comparing with the rest of $\boldsymbol{x}^\natural$. Before we proceed, we list our two assumptions:

- Measurement system $\mathbf{A}$ is assumed to be sparse. Further, $\mathbf{A}$ is assumed to have entries with $\mathbb{E}A_{ij} = 0$, and $A_{ij} \in \{0, \pm A^{-1/2}\}$, where an entry $A_{ij} = A^{-1/2}$ (or $-A^{1/2}$) implies a positive (negative) relation between the ith sensor and the jth signal component. Having zero as entry implies no relation.
- The regularizer $f(\boldsymbol{x})$ is assumed to be separable such that $f(\boldsymbol{x}) = \sum_{i=1}^n f_i(x_i)$. If it is not mentioned specifically, we assume all functions $f_i(\cdot)$ are the same.

First we transform (1) to a factor graph [37]. Adopting the viewpoint of Bayesian reasoning, we can reinterpret M-estimator in (2) as the *maximum a posteriori* (MAP) estimator and rewrite it as

$$\widehat{x} = \text{argmax}_x \exp\left(-\frac{\|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2}{2\sigma^2}\right) \times \exp\left(-f(\boldsymbol{x})\right).$$

The first term $\exp\left(-\frac{\|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2}{2\sigma^2}\right)$ is viewed as the probability $\mathbb{P}(\boldsymbol{y}|\boldsymbol{x})$ while the second term $\exp(-f(\boldsymbol{x}))$ is regarded as the prior imposed on $\boldsymbol{x}$. Notice the term $e^{-f(\cdot)}$ may not necessarily be the true prior on $\boldsymbol{x}^\natural$.

As in [29], we associate (2) with a factor graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ denotes the node set and $\mathscr{E}$ is the edge set. First we discuss set $\mathscr{V}$, which consists of two types of nodes: variable nodes and check nodes. We represent each entry x_i as a variable node v_i and each entry y_a as a check node c_a. Additionally, we construct a check node corresponds to each prior $e^{-f(x_i)}$. Then we construct the edge set $\mathscr{E}$ by: (i) placing an edge between the check node of the prior $e^{-f(x_i)}$ and the variable node v_i, and (ii) introducing an edge between the variable node v_i and c_j iff A_{ij} is non-zero. Thus, the design of $\mathbf{A}$ is transformed to the problem of graph connectivity in $\mathscr{E}$. Before to proceed, we list the notations used in this work.

Notations. We denote $c, c^{'}, c_0 > 0$ as some fixed positive constants. For two arbitrary real numbers a, b, we denote $a \lesssim b$ when there exists some positive constant $c_0 > 0$ such that $a \leq c_0 b$. Similarly, we define the notation $a \gtrsim b$. If $a \lesssim b$ and $a \gtrsim b$ hold simultaneously, we denote as $a \asymp b$. We have $a \propto b$ when a is proportional to b. For two distributions d_1 and d_2, we denote $d_1 \cong d_2$ if they are equal up to some normalization.

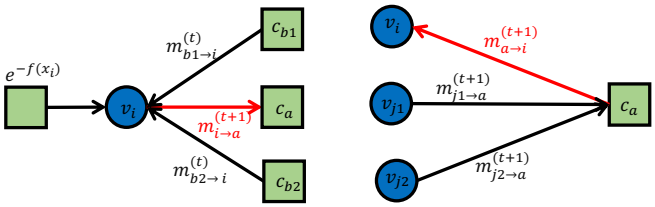

Fig. 1. Illustration of the message-passing algorithm, where the square icons represent the check nodes while the circle icons represent the variable nodes.

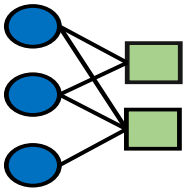

Fig. 2. Illustration of the generating polynomials: $\lambda(\alpha) = \frac{1}{3} + \frac{2\alpha}{3}$ and $\rho(\alpha) = \frac{\alpha}{2} + \frac{\alpha^2}{2}$. The square icons represent the check nodes while the circle icons represent the variable nodes.

III. Sensing Matrix for Regular Sensing

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

$$m_{i\to a}^{(t+1)}(x_i) \cong e^{-f(x_i)} \prod_{b\in\partial i\backslash a} \widehat{m}_{b\to i}^{(t)}(x_i); \tag{3}$$

$$\widehat{m}_{a\to i}^{(t+1)}(x_i) \cong \int \prod_{j\in\partial a\backslash i} m_{j\to a}^{(t+1)}(x_i)\cdot e^{-\frac{\left(y_a-\sum_{j=1}^n A_{aj}x_j\right)^2}{2\sigma^2}} dx_j, \tag{4}$$

where ∂i and ∂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 tth iteration, we recover x_i by maximizing the posterior probability

$$\widehat{x}_i^{(t)} = \text{argmax}_{x_i}\, \mathbb{P}(x_i|\boldsymbol{y}) \approx \text{argmax}_{x_i}\, e^{-f(x_i)} \prod_{a\in\partial i} \widehat{m}_{a\to i}^{(t)}(x_i). \tag{5}$$

In the design of matrix $\mathbf{A}$, 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 $\lambda(\alpha) = \sum_i \lambda_i \alpha^{i-1}$ 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 λ_i as the fraction of variable nodes with degree i, and similarly we define ρ_i for the check nodes. An illustration of the generating polynomials $\lambda(\alpha)$ and $\rho(\alpha)$ is shown in Figure 2.

A. Density evolution

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

Basically, DE views $m_{i\to a}^{(t)}$ and $\widehat{m}_{a\to i}^{(t)}$ 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 $m_{i\to a}^{(t)}$ and $\widehat{m}_{a\to i}^{(t)}$ 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 $m_{i\to a}^{(t)}$ and $\widehat{m}_{a\to i}^{(t)}$ as Gaussian RVs, i.e., $m_{i\to a} \sim \mathsf{N}(\mu_{i\to a}, v_{i\to a})$ and $\widehat{m}_{a\to i} \sim \mathsf{N}(\widehat{\mu}_{a\to i}, \widehat{v}_{a\to i})$, 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 average of the variance, respectively, and are defined as

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

$$V^{(t)} = \frac{1}{m\cdot n}\sum_{i=1}^n\sum_{a=1}^m v_{i\to a}^{(t)}.$$

Then we can show that the DE analysis yields

$$E^{(t+1)} = \mathbb{E}_{\text{prior}(s)}\mathbb{E}_z\Bigg[h_{\text{mean}}\Bigg(s + \sum_{i,j}\rho_i\lambda_j z\sqrt{\frac{i}{j}E^{(t)} + \frac{A\sigma^2}{j}}; \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2 + iV^{(t)}}{j}\Bigg) - s\Bigg]^2; \tag{6}$$

$$V^{(t+1)} = \mathbb{E}_{\text{prior}(s)}\mathbb{E}_z h_{\text{var}}\Bigg(s + \sum_{i,j}\rho_i\lambda_j z\sqrt{\frac{i}{j}E^{(t)} + \frac{A\sigma^2}{j}}; \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2 + iV^{(t)}}{j}\Bigg), \tag{7}$$

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

$$h_{\text{mean}}(\mu; v) = \lim_{\gamma\to\infty}\frac{\int x_i e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i}{\int e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i}; \tag{8}$$

$$h_{\text{var}}(\mu; v) = \lim_{\gamma\to\infty}\frac{\gamma\int x_i^2 e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i}{\int e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i} - \left(h_{\text{mean}}(\mu; v)\right)^2.$$

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 $\mathscr{G} = (\mathscr{V}, \mathscr{E})$,

or equivalently the sensing matrix $\mathbf{A}$, by setting A_{ij} as $\mathbb{P}(A_{ij} = A^{-1/2}) = \mathbb{P}(A_{ij} = -A^{-1/2}) = \frac{1}{2}$, if there is an edge $(v_i, c_j) \in \mathscr{E}$; 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 ($\sigma^2 = 0$). [3] 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 $n\left(\sum_i i\lambda_i\right) = m\left(\sum_i i\rho_i\right) = \sum_{i,j} \mathbb{1}((v_i, c_j) \in \mathscr{E})$, we formulate the above two design criteria as the following optimization problem

$$\min_{\substack{\boldsymbol{\lambda}\in\Delta_{\mathsf{dv}_{\max}-1};\\ \boldsymbol{\rho}\in\Delta_{\mathsf{dc}_{\max}-1}}} \quad \frac{m}{n} = \frac{\sum_{i\geq 2} i\lambda_i}{\sum_{i\geq 2} i\rho_i}, \tag{9}$$

$$\text{s.t.} \quad \lim_{t\to\infty} (E^{(t)}, V^{(t)}) = (0,0); \tag{10}$$

$$\lambda_1 = \rho_1 = 0, \tag{11}$$

where Δ_{d-1} denotes the d-dimensional simplex, namely, $\Delta_{d-1} \triangleq \{\boldsymbol{z} \in \mathbb{R}^d \mid \sum_i z_i = 1, z_i \geq 0\}$. 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., $e^{-\beta|x|}$, 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., $e^{\beta|x|^2}$, 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 ℓ_1, ℓ_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 $\lim_{t\to\infty}(E^{(t)}, V^{(t)}) = (0,0)$ to some closed-form expressions.

Example 1 (Regular sensing with ℓ_1 regularizer)**.** *Assuming the signal $\boldsymbol{x}^\natural$ is k-sparse, i.e., $\left\|\boldsymbol{x}^\natural\right\|_0 \leq k$, we would like to recover $\boldsymbol{x}^\natural$ 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*

$$\begin{aligned} E^{(t+1)} &= \mathbb{E}_{\text{prior}(s),z\sim\mathsf{N}(0,1)}\left[\mathsf{prox}\left(s + a_1 z\sqrt{E^{(t)}}; \beta a_2 V^{(t)}\right) - s\right]^2; \\ V^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)} \\ &\quad \cdot\left[\beta a_2 V^{(t)}\mathsf{prox}'\left(s + a_1 z\sqrt{E^{(t)}}; \beta a_2 V^{(t)}\right)\right], \end{aligned} \tag{12}$$

where a_1 is defined as $\sum_{i,j}\rho_i\lambda_j\sqrt{i/j}$, and a_2 is defined as $\sum_{i,j}\rho_i\lambda_j\,(i/j)$. Further, operator $\mathsf{prox}(a;b)$ is the soft-thresholding estimator defined as $\text{sign}(a)\max(|a|-b,0)$, and operator $\mathsf{prox}'(a;b)$ is the derivative w.r.t. the first argument.

Remark 1. *Unlike SE that only tracks $E^{(t)}$ [20], 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 dv *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 $\boldsymbol{x}^\natural$ are iid and $\boldsymbol{x}^\natural \in \mathbb{R}^n$ is k-sparse, each entry becomes zero with probability $(1-k/n)$. Hence we set β 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 $\lim_{t\to\infty}(E^{(t)}, V^{(t)}) = (0,0)$ under the noiseless setting (i.e., $\sigma = 0$), we conclude the following*

Theorem 1. *Let $\boldsymbol{x}^\natural$ be a k-sparse signal and assume that β 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) $\sum_{i,j}\rho_i\lambda_j\sqrt{i/j} \leq c_1'\sqrt{n/k}$ and (ii) $\sum_{i,j}\rho_i\lambda_j\,(i/j) \leq \frac{c_2' n}{k\log(n/k)}$, where $c_1', c_2' > 0$ are some positive constants.*

Remark 2. *Consider the settings in Theorem 1 and assume (i) $\sum_{i,j}\rho_i\lambda_j\sqrt{i/j} \leq c_1'\sqrt{n/k}$ and (ii) $\sum_{i,j}\rho_i\lambda_j\,(i/j) \leq \frac{c_2' n}{k\log(n/k)}$. Then, there exists positive constants $\varepsilon > 0$ and $0 < \gamma < 1$ such that $\{E^{(t)}, V^{(t)}\}$ 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 deferred to Section A.*

Example 2 (Regular sensing with ℓ_2 regularizer)**.** *In addition*

[3] We consider the noiseless setting only for the purpose of deriving the minimum sensor number m. This does not affect the application of our designed sensing matrices under the noisy setting. Actually, this logic is also used in the classical papers, i.e., [17], [20], [40].

to the Laplacian prior, we also considered the Gaussian prior, i.e., $e^{-\|\boldsymbol{x}\|_2^2}$, which makes the M-estimator in (2) the ridge regression [44]. Assuming the ground-truth $\boldsymbol{x}^\natural$ to be Gaussian distributed with zero mean and unit variance, we would like to recover the signal $\boldsymbol{x}^\natural$ with the regularizer $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$.

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

$$E^{(t+1)} = \frac{a_1^2 E^{(t)} + a_2^2 (V^{(t)})^2}{\left(1 + a_2 V^{(t)}\right)^2};$$
$$V^{(t+1)} = \frac{a_2 V^{(t)}}{1 + a_2 V^{(t)}}, \tag{13}$$

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

Stage II: Distribution approximation. *We can skip this stage as the ground-truth prior of $\boldsymbol{x}^\natural$, namely, $\mathsf{N}(0,1)$, 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, $E^{(t)} \le e^{-c_0(t-T)} E^{(T)}$ and $V^{(t)} \le e^{-c_1(t-T)} V^{(T)}$ whenever $t \ge T$. 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 Figure 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 $m \ge n$. This criteria is consistent with the previous finding: no savings can be achieved provided that $\boldsymbol{x}^\natural$ resides within the whole linear space $\mathbb{R}^n$.

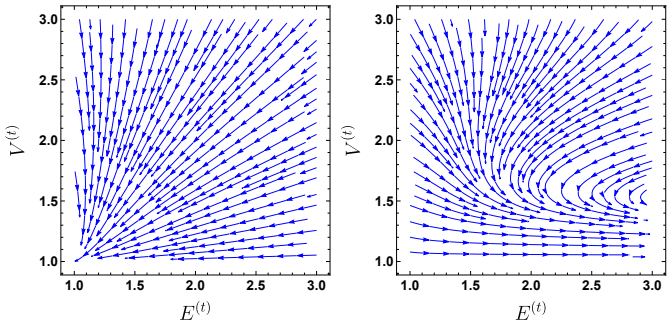

Fig. 3. Illustration of DE in (13). **Left panel**: $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} < 1$. **Right panel**: $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} > 1$. Notice that the left panel has a fix-point $(0,0)$ while the right panel is with non-zero fix-point.

Example 3 (Regular sensing with elastic net regularizer)**.** *We revisit the sparse signal setting where $\boldsymbol{x}^\natural$ is assumed to be k-sparse. Instead of ℓ_1 regularization, we consider the elastic net regularization for signal reconstruction, where $f(\boldsymbol{x})$ is written as $\beta_1\|\boldsymbol{x}\|_1 + \beta_2\|\boldsymbol{x}\|_2^2$ ($\beta_1, \beta_2 > 0$). For the ease of analysis, we pick $\beta_1 = \beta_2$ and write it as β.*

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*

$$E^{(t+1)} = \mathbb{E}_{\text{prior}(s)} \mathbb{E}_{z\sim\mathsf{N}(0,1)} \left[\text{prox}\left(\frac{s + a_1 z\sqrt{E^{(t)}}}{1 + 2\beta\cdot a_2 V^{(t)}}; \frac{\beta\cdot a_2 V^{(t)}}{1 + 2\beta\cdot a_2 V^{(t)}}\right) - s\right]^2;$$
$$V^{(t+1)} = \frac{\beta\cdot a_2 V^{(t)}}{1 + 2\beta\cdot a_2 V^{(t)}} \cdot \mathbb{E}_{\text{prior}(s)} \mathbb{E}_{z\sim\mathsf{N}(0,1)} \text{prox}'\left(\frac{s + a_1 z\sqrt{E^{(t)}}}{1 + 2\beta\cdot a_2 V^{(t)}}; \frac{\beta\cdot a_2 V^{(t)}}{1 + 2\beta\cdot a_2 V^{(t)}}\right). \tag{14}$$

Stage II: Distribution approximation. *Following the same procedure as in Example 1, we let*

$$\frac{1}{Z}\int_{|\alpha|\le c_0} e^{-\beta|\alpha| - \beta|\alpha|^2} d\alpha = 1 - \frac{k}{n}, \tag{15}$$

where Z is the normalization constant defined as $Z \triangleq \int_{-\infty}^{\infty} \exp(-\beta|\alpha| - \beta|\alpha|^2) d\alpha$, and c_0 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

$$\frac{\text{erfc}(\sqrt{\beta}(1+2c)/2)}{\text{erfc}(\sqrt{\beta}/2)} = \frac{k}{n}, \tag{16}$$

*where $\text{erfc}(\cdot)$ is the **complementary error function** defined as $2/\sqrt{\pi} \cdot \int_{(\cdot)}^{\infty} e^{-\alpha^2} d\alpha$. 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 $\text{erfc}(\alpha) \approx \frac{e^{-\alpha^2}}{\alpha\sqrt{\pi}}(1 - \frac{1}{2\alpha^2} + \cdots + (-1)^n \frac{(2n-1)!!}{(2\alpha^2)^n} + \cdots)$ when $\alpha \to \infty$ (Page 584 in [45]), we can rewrite (16) as*

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

which leads to $\beta \approx c_0 \log c_1 n/k$ (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 $\boldsymbol{x}^\natural$ is k-sparse and set β as $c_0 \cdot \log c_1 n/k$. Then the necessary conditions for $\lim_{t\to\infty}$ associated with the DE equation in (14) are (i) $\sum_{i,j} \rho_i \lambda_j \sqrt{i/j} \le c_1' \sqrt{n/k}$ and (ii) $\sum_{i,j} \rho_i \lambda_j (i/j) \le \frac{c_2'' n}{k \log(n/k)}$, 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., $E^{(t)} \le \gamma^t E^{(0)}$ and $V^{(t)} \le \gamma^t V^{(0)}$, $0 < \gamma < 1$.

Having studied the ℓ_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 $\mathbf{A}$ such that we can provide preferential treatment for different entries of $\boldsymbol{x}^\natural$. For example, the high priority components will be recovered more accurately than the low priority parts of $\boldsymbol{x}^\natural$.

A. Density evolution

Dividing the entire $\boldsymbol{x}^\natural$ into the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural \in \mathbb{R}^{n_{\mathsf{H}}}$ and low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural \in \mathbb{R}^{n_{\mathsf{L}}}$, we separately introduce the generating polynomials $\lambda_{\mathsf{H}}(\alpha) = \sum \lambda_{\mathsf{H},i}\alpha^{i-1}$ and $\lambda_{\mathsf{L}}(\alpha) = \sum \lambda_{\mathsf{L},i}\alpha^{i-1}$ for the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$ and the low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$, respectively. Note that $\lambda_{\mathsf{H},i}$ (and likewise $\lambda_{\mathsf{L},i}$) denotes the fraction of variable nodes corresponding to high-priority part (low-priority part) with degree i. Similarly, we introduce the generating polynomials $\rho_{\mathsf{H}}(\alpha) = \sum_i \rho_{\mathsf{H},i}\alpha^{i-1}$ and $\rho_{\mathsf{L}}(\alpha) = \sum_i \rho_{\mathsf{L},i}\alpha^{i-1}$ for the edges of the check nodes connecting to the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$ and to the low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$, respectively.

Generalizing the analysis of the regular sensing, we separately track the average error and variance for $\boldsymbol{x}_{\mathsf{H}}^\natural$ and $\boldsymbol{x}_{\mathsf{L}}^\natural$. For the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$, we define E_{H} as $\sum_m \sum_{i\in\mathsf{H}}(\mu_{i\to a} - x_i^\natural)^2/(m\cdot n_{\mathsf{H}})$ and V_{H} as $\sum_m \sum_{i\in\mathsf{H}} v_{i\to a}/(m\cdot n_{\mathsf{H}})$, where n_{H} denotes the length of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$. Analogously we define E_{L} and V_{L} for the low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$. We then write the corresponding DE as

$$\begin{aligned} E_{\mathsf{H}}^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[h_{\text{mean}}\left(s + z\cdot b_{\mathsf{H},1}^{(t)}; b_{\mathsf{H},2}^{(t)}\right) - s\right]^2; \\ V_{\mathsf{H}}^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[h_{\text{var}}\left(s + z\cdot b_{\mathsf{H},1}^{(t)}; b_{\mathsf{H},2}^{(t)}\right)\right], \end{aligned} \tag{17}$$

where $b_{\mathsf{H},1}^{(t)}$ and $b_{\mathsf{H},2}^{(t)}$ are defined as

$$\begin{aligned} b_{\mathsf{H},1}^{(t)} &= \sum_{\ell,i,j} \lambda_{\mathsf{H},\ell}\rho_{\mathsf{L},i}\rho_{\mathsf{H},j}\sqrt{\frac{A\sigma^2 + iE_{\mathsf{L}}^{(t)} + jE_{\mathsf{H}}^{(t)}}{\ell}}; \\ b_{\mathsf{H},2}^{(t)} &= \sum_{\ell,i,j} \lambda_{\mathsf{H},\ell}\rho_{\mathsf{L},i}\rho_{\mathsf{H},j}\frac{A\sigma^2 + iV_{\mathsf{L}}^{(t)} + jV_{\mathsf{H}}^{(t)}}{\ell}. \end{aligned}$$

The definitions of h_{mean} and h_{var} are as in (8). Switching the index H with L yields the DE w.r.t. the pair $(E_{\mathsf{L}}^{(t+1)}, V_{\mathsf{L}}^{(t+1)})$. Notice we can also put different regularizers $f_{\mathsf{H}}(\cdot)$ and $f_{\mathsf{L}}(\cdot)$ for $\boldsymbol{x}_{\mathsf{H}}^\natural$ and $\boldsymbol{x}_{\mathsf{L}}^\natural$. In this case, we need to modify the regularizers $f(\cdot)$ in (8) to $f_{\mathsf{H}}(\cdot)$ and $f_{\mathsf{L}}(\cdot)$, 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:
Consistency requirement w.r.t. edge number. Consider the total number of edges incident with the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$, $\sum_{i\in\mathsf{H}} \mathbb{1}\left((v_i, c_a)\in\mathscr{E}\right)$. From the viewpoint of the variable nodes, we can compute this number as $n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}\right)$. Likewise, from the viewpoint of the check nodes, the total number of edges is obtained as $\sum_{i\in\mathsf{H}} \mathbb{1}\left((v_i, c_a)\in\mathscr{E}\right) = m\left(\sum_i i\rho_{\mathsf{H},i}\right)$. Since the edge number should be the same with either of the above two counting methods, we obtain

$$\sum_{i\in\mathsf{H}} \mathbb{1}\left[(v_i, c_a)\in\mathscr{E}\right] = n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}\right) = m\left(\sum_i i\rho_{\mathsf{H},i}\right).$$

Similarly, the consistency requirement for the edges connecting to the low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$ would give $\sum_{i\in\mathsf{L}} \mathbb{1}\left((v_i, c_a)\in\mathscr{E}\right) = m(\sum_i i\rho_{\mathsf{L},i}) = n_{\mathsf{L}}(\sum_i i\lambda_{\mathsf{L},i})$.

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 $E_{\mathsf{H}}^{(t)}$ in the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$ to be zero. Therefore, the requirements can be summarized as

  Requirement 1. *In the noiseless setting, i.e., $\sigma = 0$, we require the quantities $E_{\mathsf{H}}^{(t)}, V_{\mathsf{H}}^{(t)}$, and $V_{\mathsf{L}}^{(t)}$ in* (17) *converge to zero*

  $$\lim_{t\to\infty}\left(E_{\mathsf{H}}^{(t)}, V_{\mathsf{H}}^{(t)}, V_{\mathsf{L}}^{(t)}\right) = (0,0,0), \tag{18}$$

  which implies the MP converges and the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$ can be perfectly reconstructed.

  Notice that no constraint is placed on the average error $E_{\mathsf{L}}^{(t)}$ for the low-priority part $\boldsymbol{x}_{\mathsf{L}}^\natural$, since it is given a lower priority in reconstruction.
- **Preferential sensing for the noisy measurement.** Different from the noiseless setting, the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$ cannot be perfectly reconstructed in the presence of measurement noise, i.e., $\lim_{t\to\infty} E_{\mathsf{H}}^{(t)} > 0$. Instead we consider the difference across iterations, namely, $\delta_{E,\mathsf{H}}^{(t)} = E_{\mathsf{H}}^{(t+1)} - E_{\mathsf{H}}^{(t)}$ and $\delta_{E,\mathsf{L}}^{(t)} = E_{\mathsf{L}}^{(t+1)} - E_{\mathsf{L}}^{(t)}$, which corresponds to the convergence rate. To provide an extra protection for the high-priority part $\boldsymbol{x}_{\mathsf{H}}^\natural$, we would like $\delta_{\mathsf{H}}^{(t)}$ to decrease at a faster rate. Hence, the following requirement:

  Requirement 2. *There exits a positive constant T_0 such that the average error $E_{\mathsf{H}}^{(t)}$ converges faster than $E_{\mathsf{L}}^{(t)}$ whenever $t \geq T_0$, i.e., $|\delta_{E,\mathsf{H}}^{(t)}| \leq |\delta_{E,\mathsf{L}}^{(t)}|$.*

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

$$\min_{\substack{\boldsymbol{\lambda}_{\mathsf{L}}\in\Delta_{\mathsf{dv}_{\mathsf{L}}-1},\\ \boldsymbol{\lambda}_{\mathsf{H}}\in\Delta_{\mathsf{dv}_{\mathsf{H}}-1},\\ \boldsymbol{\rho}_{\mathsf{L}}\in\Delta_{\mathsf{dc}_{\mathsf{H}}-1},\\ \boldsymbol{\rho}_{\mathsf{H}}\in\Delta_{\mathsf{dc}_{\mathsf{L}}-1}}} \frac{m}{n} = \frac{n_{\mathsf{L}}\left(\sum_i i\lambda_{\mathsf{L},i}\right) + n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}\right)}{\sum_i i\left(\rho_{\mathsf{L},i} + \rho_{\mathsf{H},i}\right)}; \tag{19}$$

$$\text{s.t.} \quad \frac{\sum_i i\lambda_{\mathsf{L},i}}{\sum_i i\lambda_{\mathsf{H},i}} \times \frac{\sum_i i\rho_{\mathsf{H},i}}{\sum_i i\rho_{\mathsf{L},i}} = \frac{n_{\mathsf{H}}}{n_{\mathsf{L}}}; \tag{20}$$

$$\text{Requirement (1) and (2);} \tag{21}$$

$$\lambda_{\mathsf{L},1} = \lambda_{\mathsf{H},1} = \rho_{\mathsf{L},1} = \rho_{\mathsf{H},1} = 0, \tag{22}$$

where Δ_{d-1} denotes the d-dimensional simplex, and the parameters dv_H and dc_L denote the maximum degree w.r.t. the variable nodes corresponding to the high-priority part $\boldsymbol{x}_\mathsf{H}^\natural$ and low-priority part $\boldsymbol{x}_\mathsf{L}^\natural$, respectively. Similarly we define the maximum degree dc_H and dc_L 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 ℓ_1 regularizer)**.** *Consider a sparse signal $\boldsymbol{x}^\natural$ whose high-priority part $\boldsymbol{x}_\mathsf{H}^\natural \in \mathbb{R}^{n_\mathsf{H}}$ and the low-priority part $\boldsymbol{x}_\mathsf{L}^\natural \in \mathbb{R}^{n_\mathsf{L}}$ are k_H-sparse and k_L-sparse, respectively. In addition, we assume $k_\mathsf{H}/n_\mathsf{H} \gg k_\mathsf{L}/n_\mathsf{L}$, implying that the high-priority part $\boldsymbol{x}_\mathsf{H}^\natural$ contains more data.* [4]

Ideally, we need to numerically run the DE update equation in (17) *to check whether the requirement in* (21) *holds or not, which can 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., $\lim_{t\to\infty}(V_\mathsf{H}^{(t)}, V_\mathsf{L}^{(t)}) = (0,0)$. 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. $V_\mathsf{H}^{(t)}$ and $V_\mathsf{L}^{(t)}$; and (ii) we require that the average variance $(V_\mathsf{H}^{(t)}, V_\mathsf{L}^{(t)})$ to converge in the region where the magnitudes of $V_\mathsf{H}^{(t)}$ and $V_\mathsf{L}^{(t)}$ are sufficiently small.*

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

$$\begin{bmatrix}\delta_{V,\mathsf{H}}^{(t+1)} \\ \delta_{V,\mathsf{L}}^{(t+1)}\end{bmatrix} = \mathbf{L}_V^{(t)} \begin{bmatrix}\delta_{V,\mathsf{H}}^{(t)} \\ \delta_{V,\mathsf{L}}^{(t)}\end{bmatrix}$$

via the Taylor-expansion. Imposing the convergence constraints on $V_\mathsf{H}^{(t)}$ and $V_\mathsf{L}^{(t)}$, i.e., $\lim_{t\to\infty}\left(\delta_{V,\mathsf{H}}^{(t)}, \delta_{V,\mathsf{L}}^{(t)}\right) = (0,0)$, yields the condition $\inf_t \|\mathbf{L}_V^{(t)}\|_{\mathsf{OP}} \le 1$. That is

$$\left[\left(\frac{\beta_\mathsf{H} k_\mathsf{H}}{n_\mathsf{H}}\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)^2 + \left(\frac{\beta_\mathsf{L} k_\mathsf{L}}{n_\mathsf{L}}\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\ell}\right)^2\right] \times \left[\left(\sum_i i\rho_{\mathsf{H},i}\right)^2 + \left(\sum_i i\rho_{\mathsf{L},i}\right)^2\right] \le 1. \tag{23}$$

[4]The high-priority part $\boldsymbol{x}_\mathsf{H}^\natural$ may still receive extra protection even if $k_\mathsf{H}/n_\mathsf{H} \le k_\mathsf{L}/n_\mathsf{L}$. One numerical experiment is attached in the Appendix.

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

$$\frac{k_\mathsf{H}}{n_\mathsf{H}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)^2 \left[\left(\sum_i \sqrt{i}\rho_{\mathsf{H},i}\right)^2 + \left(\sum_i \sqrt{i}\rho_{\mathsf{L},i}\right)^2\right] \le 1. \tag{24}$$

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 $\delta_{E,\mathsf{H}}^{(t)} = E_\mathsf{H}^{(t+1)} - E_\mathsf{H}^{(t)}$ and $\delta_{E,\mathsf{L}}^{(t)} = E_\mathsf{L}^{(t+1)} - E_\mathsf{L}^{(t)}$. Using the Requirement 2, we perform the Taylor expansion w.r.t. the difference $\delta_{E,\mathsf{H}}^{(t)}$ and $\delta_{E,\mathsf{L}}^{(t)}$, and obtain the linear equation*

$$\begin{bmatrix}\delta_{E,\mathsf{H}}^{(t+1)} \\ \delta_{E,\mathsf{L}}^{(t+1)}\end{bmatrix} = \begin{bmatrix}L_{E,11} & L_{E,12} \\ L_{E,21} & L_{E,22}\end{bmatrix} \begin{bmatrix}\delta_{E,\mathsf{H}}^{(t)} \\ \delta_{E,\mathsf{L}}^{(t)}\end{bmatrix}.$$

To ensure the reduction of $\delta_{E,\mathsf{H}}^{(t)}$ at a faster rate than $\delta_{E,\mathsf{L}}^{(t)}$, we would require $L_{E,11} \le L_{E,21}$ and $L_{E,12} \le L_{E,22}$. This results in

$$\frac{k_\mathsf{H}}{n_\mathsf{H}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)^2 \le \frac{k_\mathsf{L}}{n_\mathsf{L}}\left(\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\sqrt{\ell}}\right)^2, \tag{25}$$

which completes the relaxation.

Example 5 (Preferential sensing with ℓ_2 regularizer)**.** *We revisit the Gaussian setting where $\boldsymbol{x}^\natural \in \mathbb{R}^{n_\mathsf{H}+n_\mathsf{L}}$ can be divided into two disjoint parts: the high-priority part $\boldsymbol{x}_\mathsf{H}^\natural \in \mathbb{R}^{n_\mathsf{H}}$ and the low-priority part $\boldsymbol{x}_\mathsf{L}^\natural \in \mathbb{R}^{n_\mathsf{L}}$. Their priors are assumed to be $e^{-\beta_\mathsf{H}\|\cdot\|_2^2}$ and $e^{-\beta_\mathsf{L}\|\cdot\|_2^2}$; 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 $V_\mathsf{H}^{(t)}$ and $V_\mathsf{L}^{(t)}$ yields*

$$\left[\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)^2 + \left(\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\ell}\right)^2\right] \times \left[\left(\sum_i \rho_{\mathsf{H},i} i\right)^2 + \left(\sum_i \rho_{\mathsf{L},i} i\right)^2\right] \le 1. \tag{26}$$

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

$$\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)^4 \left(\sum_i \rho_{\mathsf{L},i}\sqrt{i}\right)^2 \times \left[\left(\sum_i \rho_{\mathsf{H},i} i\right)^2 \left(\sum_j \frac{\rho_{\mathsf{L},j}}{\sqrt{j}}\right)^2 + \left(\sum_j \rho_{\mathsf{L},j}\sqrt{j}\right)^2\right] \le 1. \tag{27}$$

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*

$$\sum_{\ell} \frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}} \leq \sum_{\ell} \frac{\lambda_{\mathsf{L},\ell}}{\sqrt{\ell}}. \tag{28}$$

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 $\boldsymbol{x}^\natural$ as in Example 4. Instead of ℓ_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 has 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 $\{\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}\}$ as the solution in the tth iteration. Then we conclude that $\{\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}\}$ yields a monotonic non-increasing sequence such that (i) it satisfies*

$$\begin{aligned} &\frac{n_{\mathsf{L}}\left(\sum_i i\lambda_{\mathsf{L},i}^{(t+1)}\right) + n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}^{(t+1)}\right)}{\sum_i i\left(\rho_{\mathsf{L},i}^{(t+1)} + \rho_{\mathsf{H},i}^{(t+1)}\right)} \\ \leq\ &\frac{n_{\mathsf{L}}\left(\sum_i i\lambda_{\mathsf{L},i}^{(t)}\right) + n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}^{(t)}\right)}{\sum_i i\left(\rho_{\mathsf{L},i}^{(t)} + \rho_{\mathsf{H},i}^{(t)}\right)}; \end{aligned}$$

and, (ii) has finite limit, i.e.,

$$\lim_{t\to\infty} \frac{n_{\mathsf{L}}\left(\sum_i i\lambda_{\mathsf{L},i}^{(t)}\right) + n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}^{(t)}\right)}{\sum_i i\left(\rho_{\mathsf{L},i}^{(t)} + \rho_{\mathsf{H},i}^{(t)}\right)} < \infty.$$

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 $e^{-f(\boldsymbol{x})}$, which belongs to the exponential family distributions. In this subsection, we generalize it to arbitrary distributions $\widehat{\text{prior}}(\boldsymbol{x})$. One example of the non-exponential distribution is sparse Gaussian, i.e., $(k/n)\cdot e^{-(x-\mu)^2/2\sigma^2} + (1-k/n)\,\delta(x)$, which is used to model k-sparse signals. With the generalized prior, the MP in (3) is modified to

$$\begin{aligned} m_{i\to a}^{(t+1)}(x_i) &\cong \widehat{\text{prior}}(x_i) \prod_{b\in\partial i\setminus a} \widehat{m}_{b\to i}^{(t)}(x_i); \\ \widehat{m}_{a\to i}^{(t+1)}(x_i) &\cong \int \prod_{j\in\partial a\setminus i} m_{j\to a}^{(t+1)}(x_i) \times e^{-\frac{\left(y_a - \sum_{j=1}^n A_{aj}x_j\right)^2}{2\sigma^2}} dx_j, \end{aligned} \tag{29}$$

and the decoding step at each iteration becomes

$$\widehat{x}_i^{(t)} = \text{argmax}_{x_i}\mathbb{P}(x_i|\boldsymbol{y}) \approx \text{argmax}_{x_i}\widehat{\text{prior}}(x_i)\cdot\prod_{a\in\partial i}\widehat{m}_{a\to i}^{(t)}(x_i). \tag{30}$$

Moreover, the functions $h_{\text{mean}}(\cdot;\cdot)$ and $h_{\text{var}}(\cdot;\cdot)$ in (6) are modified to $\widehat{h}_{\text{mean}}(\cdot;\cdot)$ and $\widehat{h}_{\text{var}}(\cdot;\cdot)$ as

$$\begin{aligned} \widehat{h}_{\text{mean}}(\mu;v) &= \lim_{\gamma\to\infty} \frac{\int x_i\cdot e^{\gamma\log\widehat{\text{prior}}(x_i)}\cdot e^{-\frac{\gamma(x_i-\mu)^2}{2v}}dx_i}{\int e^{\gamma\log\widehat{\text{prior}}(x_i)}\cdot e^{-\frac{\gamma(x_i-\mu)^2}{2v}}dx_i}; \\ \widehat{h}_{\text{var}}(\mu;v) &= \lim_{\gamma\to\infty} \frac{\gamma\int x_i^2\cdot e^{\gamma\log\widehat{\text{prior}}(x_i)}\cdot e^{-\frac{\gamma(x_i-\mu)^2}{2v}}dx_i}{\int e^{\gamma\log\widehat{\text{prior}}(x_i)}\cdot e^{-\frac{\gamma(x_i-\mu)^2}{2v}}dx_i} \\ &\quad - \left(\widehat{h}_{\text{mean}}(\mu;v)\right)^2. \end{aligned}$$

Afterwards, we can design the sensing matrix with the same procedure as in (9) and (19).

B. MMSE decoder

Notice that both (5) and (30) reconstruct the signal by minimizing the error probability $\mathbb{P}\left(\widehat{\boldsymbol{x}}\neq\boldsymbol{x}^\natural\right)$, which can be regarded as a MAP decoder. This subsection considers MMSE decoder, which is to minimize the ℓ_2 error, i.e., $\left\|\widehat{\boldsymbol{x}}-\boldsymbol{x}^\natural\right\|_2$. The message-passing procedure stays the same as (29) while the decoding procedure needs to be modified to

$$\widehat{x}_i^{(t)} = \int x_i\mathbb{P}(x_i|\boldsymbol{y})dx_i \approx \int\left(x_i\cdot\widehat{\text{prior}}(x_i)\cdot\prod_{a\in\partial i}\widehat{m}_{a\to i}^{(t)}(x_i)\right)dx_i.$$

Moreover, the functions $h_{\text{mean}}(\cdot;\cdot)$ and $h_{\text{var}}(\cdot;\cdot)$ in the DE in (6) are modified to $\widetilde{h}_{mean}(\cdot;\cdot)$ and $\widetilde{h}_{\text{var}}(\cdot;\cdot)$ as

$$\begin{aligned} \widetilde{h}_{\text{mean}}(\mu;v) &= \frac{\int x_i\cdot\widehat{\text{prior}}(x_i)\cdot e^{-\frac{(x_i-\mu)^2}{2v}}dx_i}{\int\widehat{\text{prior}}(x_i)\cdot e^{-\frac{(x_i-\mu)^2}{2v}}dx_i}; \\ \widetilde{h}_{\text{var}}(\mu;v) &= \frac{\int x_i^2\cdot\widehat{\text{prior}}(x_i)\cdot e^{-\frac{(x_i-\mu)^2}{2v}}dx_i}{\int\widehat{\text{prior}}(x_i)\cdot e^{-\frac{(x_i-\mu)^2}{2v}}dx_i} - \left(\widetilde{h}_{\text{mean}}(\mu;v)\right)^2. \end{aligned}$$

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 $\{\rho_{\mathsf{H},i}\}$ and $\{\rho_{\mathsf{L},i}\}$ of the check nodes to $\rho_{\mathsf{H},\mathsf{dc_H}} = 1$ and $\rho_{\mathsf{L},\mathsf{dc_L}} = 1$, respectively. Therefore, each check node has $\mathsf{dc_H}$ edges connecting to the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ and $\mathsf{dc_L}$ edges connecting to the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$.

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 ℓ_1 regularizer and elastic net regularizer.

We evaluate two types of sensing matrices for the preferential sensing, namely, $\mathbf{A}_{\text{preferential}}^{(\text{init})}$ and $\mathbf{A}_{\text{preferential}}^{(\text{final})}$, which correspond to the distributions $\{\boldsymbol{\lambda}_{\mathsf{H}}\}$ and $\{\boldsymbol{\lambda}_{\mathsf{L}}\}$ in the initialization phase and at the final outcome of Algorithm 1. As the baseline, we design the sensing matrix $\mathbf{A}_{\text{regular}}$ via (9) which provides regular sensing with an additional constraint which enforces equal edge number with $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ 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 $\mathsf{SNR} = \|\boldsymbol{x}^{\natural}\|_2^2/\sigma^2$. We separately evaluate the signal recovery performance via the partial and full reconstruction error, which corresponds to the error of the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$ and that of the whole signal $\left\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\right\|_2$, respectively.

*1) **Experiments with sparse signal:*** We consider the case where $\boldsymbol{x}^{\natural}$ is a $(k_{\mathsf{H}} + k_{\mathsf{L}})$-sparse signal. We fix the check node degrees $\mathsf{dc_H}$ and $\mathsf{dc_L}$ as 5 and let the maximum variable node degree $\mathsf{dv_{max}}$ 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 $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ and n_{L} of the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$ 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., ℓ_1 regularizer ($\|\cdot\|_1$) [2], [4] and elastic net regularizer ($\|\cdot\|_1 + \|\cdot\|_2^2$) [46]; (ii) **greedy methods**, e.g., *orthogonal 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 Figure 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.

Algorithm 1 Design of Sensing Matrix for Preferential Sensing.

- **Input**: maximum variable node degree $\mathsf{dv_{max}}$, check node degree $\mathsf{dc_H}$ and $\mathsf{dc_L}$, signal lengths n_{H} and n_{L}, sparsity numbers k_{H} and k_{L}, and iteration number T.
- **Initialization**: set $\beta_{\mathsf{H}} \asymp \log\left(\frac{n_{\mathsf{H}}}{k_{\mathsf{L}}}\right), \beta_{\mathsf{L}} \asymp \log\left(\frac{n_{\mathsf{L}}}{k_{\mathsf{L}}}\right)$. Then we initialize $\{\lambda_{\mathsf{H},i}\}$ and $\{\lambda_{\mathsf{L},i}\}$ as

$$\begin{aligned}
&\min_{\substack{\boldsymbol{\lambda}_{\mathsf{H}} \in \Delta_{\mathsf{dv_{max}}-1}, \\ \boldsymbol{\lambda}_{\mathsf{L}} \in \Delta_{\mathsf{dv_{max}}-1}}} \sum_i i\lambda_{\mathsf{H},i}, \\
\text{s.t.}\ \ & n_{\mathsf{H}}\mathsf{dc_L}\left(\sum_i i\lambda_{\mathsf{H},i}\right) = n_{\mathsf{L}}\mathsf{dc_H}\left(\sum_i i\lambda_{\mathsf{L},i}\right); \\
& \left(\frac{\beta_{\mathsf{H}}k_{\mathsf{H}}}{n_{\mathsf{H}}}\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)^2 + \left(\frac{\beta_{\mathsf{L}}k_{\mathsf{L}}}{n_{\mathsf{L}}}\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\ell}\right)^2 \\
& \leq \frac{1}{(\mathsf{dc_H})^2 + (\mathsf{dc_L})^2}; \\
& \sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}} \leq \frac{\sqrt{n_{\mathsf{H}}}}{\sqrt{k_{\mathsf{H}}}\sqrt{\mathsf{dc_H} + \mathsf{dc_L}}}; \\
& \lambda_{\mathsf{L},1} = \lambda_{\mathsf{H},1} = 0,
\end{aligned}$$

  which is equivalent to (19) without the Requirement 2.
- **Iterative Update:** denote $\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}$ (or $\boldsymbol{\lambda}_{\mathsf{L}}^{(t)}$) as the updated version of $\boldsymbol{\lambda}_{\mathsf{H}}$ (or $\boldsymbol{\lambda}_{\mathsf{L}}$) at the tth iteration.
- **For time $t = 1$ to T:** update $\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}$ and $\boldsymbol{\lambda}_{\mathsf{L}}^{(t)}$ by alternating minimization of (19) with Requirement 1 and Requirement 2 being replaced by (23), (24), and (25).
  1) **Update** $\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}$ with $\boldsymbol{\lambda}_{\mathsf{L}}$ being fixed to be $\boldsymbol{\lambda}_{\mathsf{L}}^{(t-1)}$;
  2) **Update** $\boldsymbol{\lambda}_{\mathsf{L}}^{(t)}$ with $\boldsymbol{\lambda}_{\mathsf{H}}$ being fixed to be $\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}$.
- **Output**: degree distribution $\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}$ and $\boldsymbol{\lambda}_{\mathsf{L}}^{(t)}$.

- **Optimization methods**. In the low-SNR regime, we observe that elastic net regularizer has a better performance than ℓ_1 regularizer. However, elastic net regularizer's performance falls behind ℓ_1 regularizer when entering the high-SNR regime. One intuitive explanation is that the extra $\|\cdot\|_2^2$ term in the elastic net regularizer promotes reconstructed signal with lower energy, i.e., $\|\widehat{\boldsymbol{x}}\|_2^2$.
- **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 ℓ_1 regularizer.

In summary, we conclude that our designed sensing matrices, both $\mathbf{A}_{\text{preferential}}^{(\text{init})}$ and $\mathbf{A}_{\text{preferential}}^{(\text{final})}$, have improved performance under all signal reconstruction methods. For the ease of conducting experiments, we will stick to the ℓ_1 regularizer in the following context as it has the best overall performance. The obtained conclusion should remain valid for other methods.

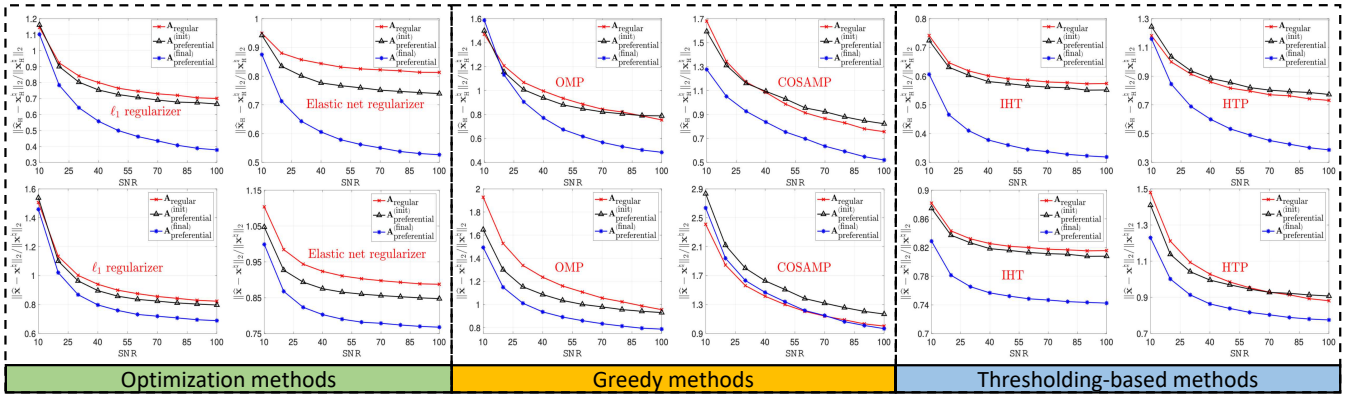

Fig. 4. Comparison of preferential sensing vs regular sensing under different signal reconstruction methods. The length n_{H} of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ is set as 100; while the length n_{L} of the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$ is set as 400. (**Top**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2/\|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$. (**Bottom**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

*b) **Impact of sparsity number:*** We fix the length n_{H} of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ as 100 and the length n_{L} of the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$ as 400. The simulation results are plotted in Figure 5.

Discussion. First, we investigate the recovery performance w.r.t. the high priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$. Using the sensing matrix $\mathbf{A}_{\text{regular}}$ (regular sensing) as the baseline, we conclude that our sensing matrix $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ (preferential sensing) achieves better performance when the signal is more sparse. Consider the case when $\mathsf{SNR} = 100$. When $k_{\mathsf{H}} = k_{\mathsf{L}} = 10$, the ratio $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2/\|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$ for $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ is approximately 0.35 while that of the $\mathbf{A}_{\text{regular}}$ 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 $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$ 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 $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ achieves significant improvements in comparison to its initialized version $\mathbf{A}_{\text{preferential}}^{(\text{init})}$.

*c) **Impact of signal length:*** We also studied various settings in which the length n_{H} of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ is set to $\{150, 200, 250, 300\}$ and the corresponding length n_{L} of the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$ is set to $\{600, 800, 1000, 1200\}$. The simulation results are plotted in Figure 6.

Discussion. Compared to regular sensing, our sensing matrix $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ can reduce the error in the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ significantly. For example, when $\mathsf{SNR} = 100$, the ratio $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2/\|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$ reduces between $40\% \sim 60\%$ with the sensing matrix $\mathbf{A}_{\text{preferential}}^{(\text{final})}$. Meanwhile, w.r.t. the whole signal $\boldsymbol{x}^{\natural}$, the ratio $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$ decreases with a smaller magnitude.

*d) **Miscellaneous numerical experiments I:*** In addition, we evaluate our designed sensing matrices when the condition $k_{\mathsf{H}}/n_{\mathsf{H}} \geq k_{\mathsf{L}}/n_{\mathsf{L}}$ is violated, or equivalently, we let $k_{\mathsf{H}}/n_{\mathsf{H}} \leq k_{\mathsf{L}}/n_{\mathsf{L}}$. We set the pair $(n_{\mathsf{H}}, k_{\mathsf{H}})$ as $(400, 10)$ and $(n_{\mathsf{L}}, k_{\mathsf{L}})$ as $(100, 10)$, respectively. The numerical experiment is put in Figure 7.

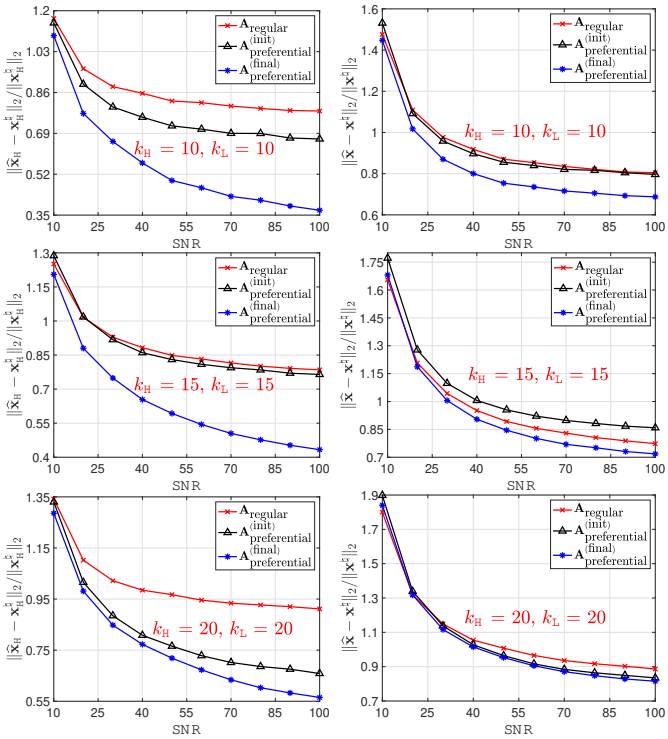

Fig. 5. Comparison of preferential sensing vs regular sensing. The length n_{H} of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ is set as 100; while the length n_{L} of the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$ is set as 400. (**Left panel**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2/\|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$. (**Right panel**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

Discussion. We conclude that our designed scheme may still bring performance improvement even if the requirement $k_{\mathsf{H}}/n_{\mathsf{H}} \gg k_{\mathsf{L}}/n_{\mathsf{L}}$ is violated. However, we observe a performance degradation of $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ when compared with $\mathbf{A}_{\text{preferential}}^{(\text{init})}$. This suggests that Requirement 2 may backfire in protecting the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ if $k_{\mathsf{H}}/n_{\mathsf{H}} \leq k_{\mathsf{L}}/n_{\mathsf{L}}$.

*e) **Miscellaneous numerical experiments II:*** In addition to our proposed scheme, another method for preferential sensing

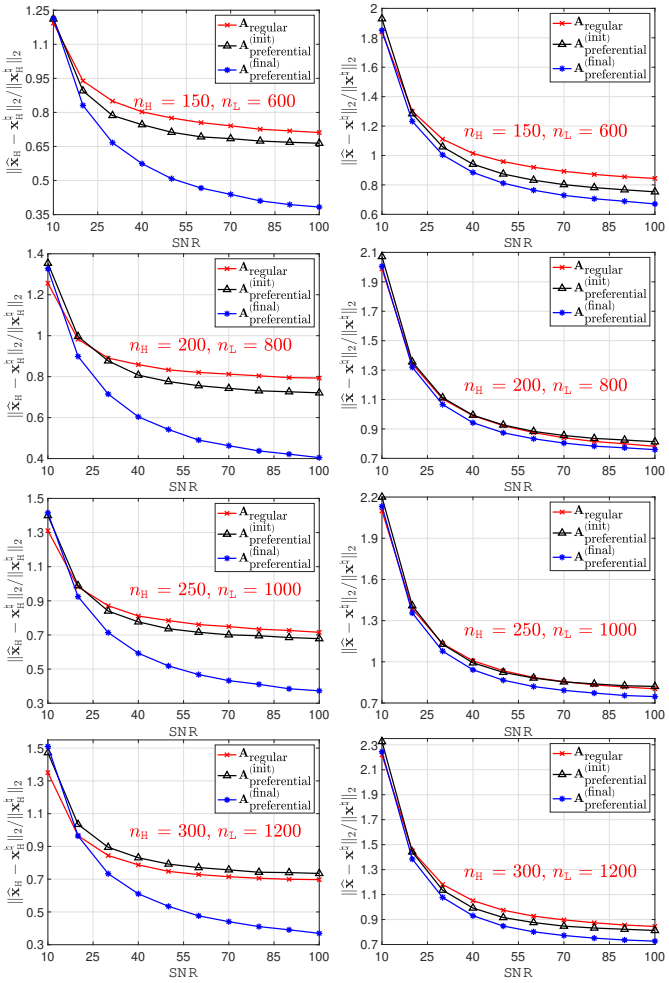

Fig. 6. Comparison of preferential sensing vs regular sensing. Both the sparsity number k_{H} and k_{L} are set as 15. (**Left panel**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2/\|\boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2$. (**Right panel**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

is the *decouple* sensing matrix design (i.e., we separately design the sensing matrices $\mathbf{A}_{\mathsf{H}}$ and $\mathbf{A}_{\mathsf{L}}$ for the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ and low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$; and then stack them together). Then, we compare our designed scheme and the decoupled design scheme, which is denoted as $\mathbf{A}_{\text{decouple}}$. For a fair comparison, we enforce the sensor number m and corresponding edge numbers connecting $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ and $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ to be same. However, there is no inference between the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ and low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$, or equivalently, no check node connecting to $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ and $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ simultaneously. The simulation result is put in Figure 8, from which we conclude that our proposed scheme yields a better reconstructed signal.

*2) **Experiments with Gaussian signal:*** We consider the Gaussian signal such that the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ and low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ follow Gaussian priors $e^{-\beta_{\mathsf{H}}(\cdot)^2}$ and $e^{-\beta_{\mathsf{L}}(\cdot)^2}$, respectively. The simulation results are in Figure 9.

Discussion. We conclude that our design scheme $\mathbf{A}^{(\text{final})}_{\text{preferential}}$ can greatly reduce the reconstruction error in the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ while the total reconstruction error $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2$ stays almost the same, which verifies the effectiveness of our design scheme in giving preferential protection of $\boldsymbol{x}^{\natural}_{\mathsf{H}}$.

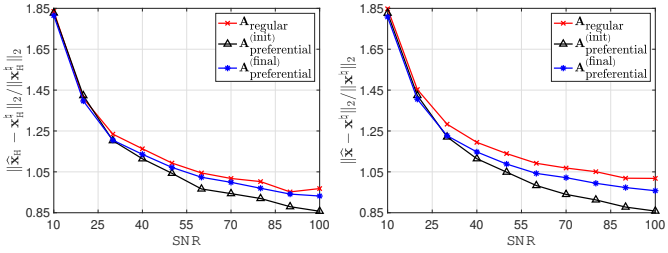

Fig. 7. Comparison of preferential sensing vs regular sensing ($k_{\mathsf{H}}/n_{\mathsf{H}} \leq k_{\mathsf{L}}/n_{\mathsf{L}}$). The length n_{H} of the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ is set as 400 and k_{H} is set as 10; while the length n_{L} of the low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ is set as 100 and k_{L} is set as 10. (**Left panel**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2/\|\boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2$. (**Right panel**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

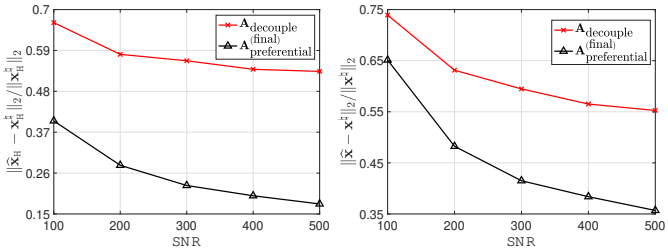

Fig. 8. Comparison of our proposed scheme vs decoupled design scheme for preferential sensing. The length n_{H} of the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ is set as 400 and k_{H} is set as 10; while the length n_{L} of the low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ is set as 100 and k_{L} is set as 10. (**Left panel**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2/\|\boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2$. (**Right panel**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

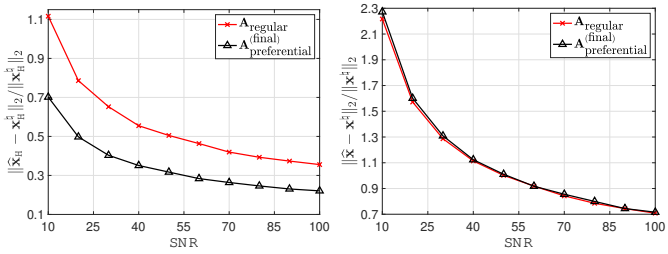

Fig. 9. Comparison of preferential sensing vs regular sensing. The length n_{H} of the high-priority part $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ is set as 100; while the length n_{L} of the low-priority part $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ is set as 400. The prior for $\boldsymbol{x}^{\natural}_{\mathsf{H}}$ is $\propto e^{-x^2/20}$ while the prior for $\boldsymbol{x}^{\natural}_{\mathsf{L}}$ is $\propto e^{-x^2/2}$. (**Left panel**) We evaluate the reconstruction performance w.r.t. the high-priority part $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2/\|\boldsymbol{x}^{\natural}_{\mathsf{H}}\|_2$. (**Right panel**) We evaluate the reconstruction performance w.r.t. the whole signal $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2/\|\boldsymbol{x}^{\natural}\|_2$.

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 $\mathscr{H}(\cdot)$, which generates four submatrices 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 $\boldsymbol{x}_{\mathsf{H}}^{\natural}$; while the horizontal detail coefficients, vertical detail coefficients, and diagonal detail coefficients are regarded as the low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$. Hence we can write the sensing relation in (1) as

$$\boldsymbol{y} = \mathbf{A}\mathscr{H}(\text{Image}) + \boldsymbol{w}, \tag{31}$$

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

*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

$$r_{H,(\cdot)} \triangleq \frac{\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2}{\|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2};$$
$$r_{W,(\cdot)} \triangleq \frac{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2}{\|\boldsymbol{x}^{\natural}\|_2},$$

which correspond to the ℓ_2 error in the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ and the entire signal $\boldsymbol{x}^{\natural}$, respectively. We use the sensing matrix $\mathbf{A}_{\text{regular}}$ as the benchmark. In addition, we omit the results of $\mathbf{A}_{\text{preferential}}^{(\text{init})}$, 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 Figure 10. From Table I and Figure 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 Figure 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 $\lambda_{\mathsf{H}}(\cdot)$ and $\lambda_{\mathsf{L}}(\cdot)$ of the variable nodes for the sensing matrix $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ are obtained as*

$$\begin{aligned}
&\lambda_{\mathsf{H}}(\alpha)\\
&= 0.0057856\alpha + 0.025915\alpha^2 + 0.36394\alpha^3 + 0.35183\alpha^4\\
&+ 0.10333\alpha^5 + 0.04134\alpha^6 + 0.021619\alpha^7 + 0.013508\alpha^8\\
&+ 0.0094374\alpha^9 + 0.0070906\alpha^{10} + 0.0056\alpha^{11}\\
&+ 0.0045851\alpha^{12} + 0.0038574\alpha^{13} + 0.0033145\alpha^{14}\\
&+ 0.0028963\alpha^{15} + .0025659\alpha^{16} + 0.0022992\alpha^{17}\\
&+ 0.0020801\alpha^{18} + 0.0018973\alpha^{19} + 0.0017428\alpha^{20}\\
&+ 0.0016109\alpha^{21} + 0.001497\alpha^{22} + 0.001398\alpha^{23}\\
&+ 0.0013111\alpha^{24} + 0.0012344\alpha^{25} + 0.0011662\alpha^{26}\\
&+ 0.0011053\alpha^{27} + 0.0010506\alpha^{28} + 0.0010013\alpha^{29}\\
&+ 0.0009565\alpha^{30} + 0.00091576\alpha^{31} + 0.00087852\alpha^{32}\\
&+ 0.00084436\alpha^{33} + 0.00081292\alpha^{34} + 0.00078388\alpha^{35}\\
&+ 0.00075697\alpha^{36} + 0.00073197\alpha^{37} + 0.00070867\alpha^{38}\\
&+ 0.00068691\alpha^{39} + 0.00066652\alpha^{40} + 0.00064738\alpha^{41}\\
&+ 0.00062937\alpha^{42} + 0.00061238\alpha^{43} + 0.00059633\alpha^{44}\\
&+ 0.00058114\alpha^{45} + 0.00056673\alpha^{46} + 0.00055304\alpha^{47}\\
&+ 0.00054001\alpha^{48} + 0.0005276\alpha^{49};\\
&\lambda_{\mathsf{L}}(\alpha) = \alpha.
\end{aligned}$$

The check node degrees dc_{H} and dc_{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 $\rho(\alpha) = \alpha^7$, respectively.

VII. Conclusions

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 $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ and transformed the design of $\mathbf{A}$ to the connectivity problem in $\mathscr{G}$. 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

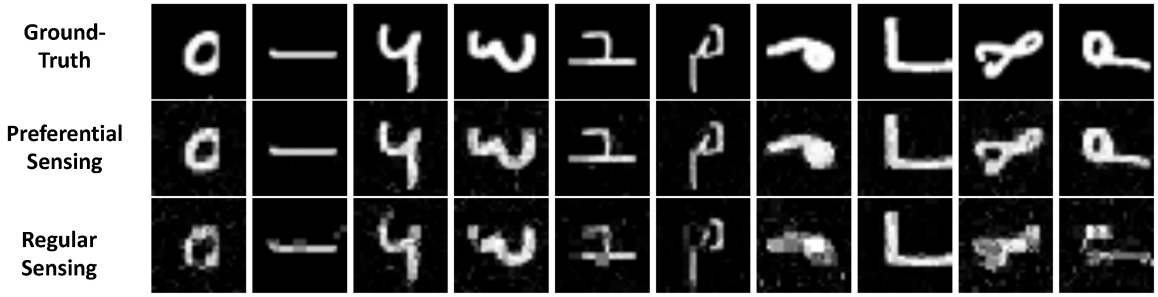

Fig. 10. The performance comparison between the sensing matrix for preferential sensing $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ and sensing matrix for regular sensing $\mathbf{A}_{\text{regular}}$. (**Top**) The ground-truth images. (**Middle**) The reconstructed images with the sensing matrix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

	Training Set				Testing Set			
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**	0.4233	**0.42216**	0.52556
3	**0.24613**	0.43677	**0.42514**	0.53163	**0.26446**	0.46766	**0.43534**	0.56189
4	**0.28331**	0.44377	**0.44623**	0.53791	**0.30092**	0.4445	**0.45804**	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	**0.25503**	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	**0.43116**	0.62785

TABLE I

THE INDEX $i = p$ CORRESPONDS TO THE SENSING MATRIX $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ FOR THE PREFERENTIAL SENSING; WHILE THE INDEX $i = r$ CORRESPONDS TO THE SENSING MATRIX $\mathbf{A}_{\text{regular}}$ FOR THE REGULAR SENSING. WE DEFINE THE RATIO $r_{H,(i)}$ $(i = \{p, r\})$ AS THE ERROR W.R.T. THE HIGH PRIORITY PART, NAMELY, $\|\widehat{\boldsymbol{x}}_{\mathsf{H}} - \boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2 / \|\boldsymbol{x}_{\mathsf{H}}^{\natural}\|_2$. SIMILARLY WE DEFINE THE RATIO $r_{W,(i)}$ $(i = \{p, r\})$ AS THE RATIO W.R.T. THE WHOLE SIGNAL, NAMELY, $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{\natural}\|_2 / \|\boldsymbol{x}^{\natural}\|_2$. MOREOVER, WE PUT THE RESULTS CORRESPONDING TO THE SENSING MATRIX $\mathbf{A}_{\text{preferential}}^{(\text{final})}$ IN THE BOLD FONT.

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

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

$$V^{(t+1)} = \underbrace{\mathbb{E}_{\text{prior}(s),z\sim\mathsf{N}(0,1)}\left[\beta a_2 V^{(t)} \mathsf{prox}'\left(s + a_1 z\sqrt{E^{(t)}}; \beta a_2 V^{(t)}\right)\right]}_{\triangleq \Psi_V\left(E^{(t)}; V^{(t)}\right)}.$$

The derivation of the necessary conditions for $\lim_{t\to\infty}\left(E^{(t)}, V^{(t)}\right) = (0,0)$ 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

$$\begin{aligned}
\delta_E^{(t+1)} =& \ \Psi_E\left(E^{(t+1)}, V^{(t+1)}\right) - \Psi_E\left(E^{(t)}, V^{(t)}\right) \\
=& \left(\frac{\partial \Psi_E(E,V)}{\partial E}\Big|_{E=E^{(t)},V=V^{(t)}}\right)\cdot \delta_E^{(t)} \\
&+ \left(\frac{\partial \Psi_E(E,V)}{\partial V}\Big|_{E=E^{(t)},V=V^{(t)}}\right)\cdot \delta_V^{(t)} \\
&+ O\left[(\delta_E^{(t)})^2\right] + O\left[(\delta_V^{(t)})^2\right]. \qquad (32)
\end{aligned}$$

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 $\left(\frac{\partial \Psi_E(E,V)}{\partial E}\right)^{(t)}\Big|_{E=E^{(t)},V=V^{(t)}}$, $\frac{\partial \Psi_E(E,V)}{\partial V}\Big|_{E=E^{(t)},V=V^{(t)}}$, $\frac{\partial \Psi_V(E,V)}{\partial E}\Big|_{E=E^{(t)},V=V^{(t)}}$, and $\frac{\partial \Psi_V(E,V)}{\partial V}\Big|_{E=E^{(t)},V=V^{(t)}}$ as $\left(\frac{\partial \Psi_E(E,V)}{\partial E}\right)^{(t)}$, $\left(\frac{\partial \Psi_E(E,V)}{\partial V}\right)^{(t)}$, $\left(\frac{\partial \Psi_V(E,V)}{\partial E}\right)^{(t)}$, and $\left(\frac{\partial \Psi_V(E,V)}{\partial V}\right)^{(t)}$, respectively. We obtain the linear equation

Fig. 11. (**Left**) Ground-truth image. (**Middle**) Reconstructed image via sensing matrix $\mathbf{A}^{(\text{final})}_{\text{preferential}}$ for preferential sensing. (**Right**) Reconstructed image via sensing matrix $\mathbf{A}_{\text{regular}}$ for regular sensing.

$$\begin{bmatrix}\delta_E^{(t+1)}\\ \delta_V^{(t+1)}\end{bmatrix} = \underbrace{\begin{bmatrix}\left(\frac{\partial\Psi_E(E,V)}{\partial E}\right)^{(t)} & \left(\frac{\partial\Psi_E(E,V)}{\partial V}\right)^{(t)}\\ \left(\frac{\partial\Psi_V(E,V)}{\partial E}\right)^{(t)} & \left(\frac{\partial\Psi_V(E,V)}{\partial V}\right)^{(t)}\end{bmatrix}}_{\triangleq\mathbf{L}^{(t)}}\begin{bmatrix}\delta_E^{(t)}\\ \delta_V^{(t)}\end{bmatrix},$$

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

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

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

$$\begin{aligned}
&\frac{\partial\Psi_E\left(E,V\right)}{\partial E}\Big|_{E=E^{(t)},V=V^{(t)}}\\
=\;& a_1^2\cdot\mathbb{E}_{\text{prior}(s)}\left[\Phi\left(-\frac{s+a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)+\Phi\left(\frac{s-a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)\right]\\
\overset{①}{=}\;& \frac{a_1^2k}{n}\left[\Phi\left(-\frac{c_0+a_2V^{(t)}}{a_2\sqrt{E^{(t)}}}\right)+\Phi\left(\frac{c_0-a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)\right]\\
&+2a_1^2\left(1-\frac{k}{n}\right)\Phi\left(-\frac{a_2V^{(t)}}{a_1E^{(t)}}\right)\\
\overset{②}{\to}\;& \frac{ka_1^2}{n}+2a_1^2\left(1-\frac{k}{n}\right)\Phi\left(-\frac{a_2V^{(t)}}{\sqrt{a_1E^{(t)}}}\right)\overset{③}{\geq}\frac{ka_1^2}{n};\\
&\frac{\partial\Psi_V\left(E,V\right)}{\partial V}\Big|_{E=E^{(t)},V=V^{(t)}}\\
=\;& \beta a_2\cdot\mathbb{E}_{\text{prior}(s)}\left[\Phi\left(-\frac{s+a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)+\Phi\left(\frac{s-a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)\right]\\
\overset{④}{=}\;& \frac{\beta a_2k}{n}\left[\Phi\left(-\frac{c_0+a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)+\Phi\left(\frac{c_0-a_2V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)\right]\\
&+2\beta a_2\left(1-\frac{k}{n}\right)\Phi\left(-\frac{a_2V^{(t)}}{a_1E^{(t)}}\right)\\
\overset{⑤}{\to}\;& \frac{k\beta a_2}{n}+2\beta a_2\left(1-\frac{k}{n}\right)\Phi\left(-\frac{a_2V^{(t)}}{\sqrt{a_1E^{(t)}}}\right)\overset{⑥}{\geq}\frac{k\beta a_2}{n},
\end{aligned}\tag{33}$$

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

$$\lim_{E^{(t)}\to 0}\Phi\left(-\frac{c_0+a_2V^{(t)}}{\sqrt{a_1E^{(t)}}}\right)+\Phi\left(\frac{c_0-a_2V^{(t)}}{\sqrt{a_1E^{(t)}}}\right)=1,$$

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 $\mathbb{E}_z h_{\text{mean}}(\cdot;\cdot)$ and $\mathbb{E}_z h_{\text{var}}(\text{mean};\text{var})$ are written as

$$\begin{aligned}
\mathbb{E}_z h_{\text{mean}}\left(s+a_1z\sqrt{E^{(t)}};a_2V^{(t)}\right) &= \frac{a_1^2E^{(t)}+a_2^2\left(V^{(t)}\right)^2s^2}{\left(1+a_2V^{(t)}\right)^2};\\
\mathbb{E}_z h_{\text{var}}\left(s+a_1z\sqrt{E^{(t)}};a_2V^{(t)}\right) &= \frac{a_2V^{(t)}}{1+a_2V^{(t)}},
\end{aligned}$$

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

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

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

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

$$\begin{aligned}
E^{(t+1)} &\leq a_1^2E^{(t)}+\frac{a_2V_SV^{(t)}}{2}\\
&\overset{①}{\leq} a_1^2E^{(t)}+\frac{a_2V_S}{2}\left(\frac{a_2}{1+a_2}\right)^tV^{(0)},
\end{aligned}\tag{34}$$

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

$$\begin{aligned}
\widetilde{E}^{(t+1)} &= \frac{E^{(t+1)}}{a_1^{2(t+1)}}\leq\frac{E^{(t)}}{a_1^{2t}}+\frac{a_2V_SV^{(0)}}{2a_1^2}\left(\frac{a_2}{(1+a_2)a_1^2}\right)^t\\
&= \widetilde{E}^{(t)}+\frac{a_2V_SV^{(0)}}{2a_1^2}\left(\frac{a_2}{(1+a_2)a_1^2}\right)^t,
\end{aligned}$$

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

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

Iterating over all such inequalities, we obtain the equation

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

which leads to

$$E^{(t+1)} \leq a_1^{2t}E^{(1)} + \underbrace{\frac{a_2 V_S V^{(0)}}{2a_1^2}\cdot\frac{a_2}{1+a_2}\frac{a_1^{2t}-\left(\frac{a_2}{1+a_2}\right)^t}{1-\frac{a_2}{(1+a_2)a_1^2}}}_{\vartheta}. \quad (35)$$

Since $a_1 < 1$ and $a_2/(1+a_2) < 1$, we have the second term ϑ 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 $a_1^{2t}E^{(1)}$ 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 $h_{\text{mean}}(\cdot;\cdot)$ and $h_{\text{var}}(\cdot;\cdot)$. Recalling their definitions in (8), our goal is to study the probability distribution $\exp\left[-\gamma\left(\beta|x| + \beta x^2 + {}^{(x-\mu)^2}/_{2v}\right)\right]$. Denote $\widetilde{\mu}$ and $\widetilde{v}$ as ${}^{\mu}/_{1+2\beta v}$ and $\frac{v}{1+2\beta v}$, respectively, we can show that the above distribution is equivalent to $\exp\left(-\frac{\gamma(x-\widetilde{\mu})^2}{2\widetilde{v}} - \gamma\beta|x|\right)$, which is of the similar form associated with the ℓ_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 $\mathbf{L}^{(t)}$ as

$$\mathbf{L}^{(t)} \triangleq \begin{bmatrix} \left(\frac{\partial\Psi_E(E,V)}{\partial E}\right)^{(t)} & \left(\frac{\partial\Psi_E(E,V)}{\partial V}\right)^{(t)} \\ \left(\frac{\partial\Psi_V(E,V)}{\partial E}\right)^{(t)} & \left(\frac{\partial\Psi_V(E,V)}{\partial V}\right)^{(t)} \end{bmatrix},$$

and require $\inf_t \|\mathbf{L}^{(t)}\|_{\text{OP}} \leq 1$. With some standard calculations, we have

$$\|\mathbf{L}^{(t)}\|_{\text{OP}} = \max\left[\left(\frac{\partial\Psi_E(E,V)}{\partial E}\right)^{(t)}, \left(\frac{\partial\Psi_V(E,V)}{\partial V}\right)^{(t)}\right]. \quad (36)$$

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

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

where ϑ_1 and ϑ_2 are defined as

$$\vartheta_1 \triangleq \mathbb{E}_{\text{prior}(s)}\left[\Phi\left(\frac{s-a_2\beta V^{(t)}}{a_1\sqrt{E^{(t)}}}\right) + \Phi\left(-\frac{s+a_2\beta V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)\right];$$

$$\vartheta_2 \triangleq \mathbb{E}_{\text{prior}(s)}\, s\cdot\left(\exp\left[-\left(\tfrac{s+a_2\beta V^{(t)}}{a_1\sqrt{E^{(t)}}}\right)^2/2\right] - \exp\left[-\left(\tfrac{a_2\beta V^{(t)}-s}{a_1\sqrt{E^{(t)}}}\right)^2/2\right]\right),$$

respectively. Plugging $\text{prior}(s) = {}^{k}/_{n}\cdot\mathbb{1}(c_0) + (1-{}^{k}/_{n})\mathbb{1}(0)$ into ϑ_1 and ϑ_2 then yields

$$\lim_{(E^{(t)},V^{(t)})\to(0,0)}\frac{\partial\Psi_E(E,V)}{\partial E}\Big|_{E=E^{(t)},V=V^{(t)}} \geq \frac{k\cdot a_1^2}{n}. \quad (37)$$

As for ${}^{\partial\Psi_V}/_{\partial V}$, we have

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

which yields

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

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

Appendix D
Proof of Proposition 3

Without loss of generality, we assume the updating order is $\{\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}\}$. For the notation simplicity, we denote the solution of (19) as $\text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)})$. Easily, we can verify (19) is a convex optimization problem w.r.t. $\boldsymbol{\lambda}_{\mathsf{H}}$ with fixed $\boldsymbol{\lambda}_{\mathsf{L}}, \boldsymbol{\rho}_{\mathsf{H}}$, and $\boldsymbol{\rho}_{\mathsf{L}}$. This results in

$$\text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}) \geq \text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t+1)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}).$$

Iterating the above procedure w.r.t. $\boldsymbol{\lambda}_{\mathsf{L}}, \boldsymbol{\rho}_{\mathsf{H}}$, and $\boldsymbol{\rho}_{\mathsf{L}}$, we obtain

$$\text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t+1)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}) \geq \text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t+1)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t+1)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}) \geq \cdots \geq \text{OPT}(\boldsymbol{\lambda}_{\mathsf{H}}^{(t+1)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t+1)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t+1)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t+1)}),$$

which completes the proof such that $\left\{\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}\right\}$ constitutes a monotonic non-increasing sequence. Combing with the fact such that (19) is non-negative, we can show that $\left\{\boldsymbol{\lambda}_{\mathsf{H}}^{(t)}, \boldsymbol{\lambda}_{\mathsf{L}}^{(t)}, \boldsymbol{\rho}_{\mathsf{H}}^{(t)}, \boldsymbol{\rho}_{\mathsf{L}}^{(t)}\right\}$ has a finite limit, i.e., $\lim_{t\to\infty}\frac{n_{\mathsf{L}}\left(\sum_i i\lambda_{\mathsf{L},i}^{(t)}\right)+n_{\mathsf{H}}\left(\sum_i i\lambda_{\mathsf{H},i}^{(t)}\right)}{\sum_i i\left(\rho_{\mathsf{L},i}^{(t)}+\rho_{\mathsf{H},i}^{(t)}\right)} < \infty$.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CCF-2007807 and ECCS-2027195.

References

[1] H. Zhang, A. Abdi, and F. Fekri, "A general framework for the design of compressive sensing using density evolution," in *IEEE Information Theory Workshop (ITW'21)*.

[2] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on information theory*, vol. 52, no. 1, pp. 6–18, 2005.

[3] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] N. Meinshausen, P. Bühlmann *et al.*, "High-dimensional graphs and variable selection with the lasso," *The annals of statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.

[6] P. Zhao and B. Yu, "On model selection consistency of lasso," *Journal of Machine learning research*, vol. 7, no. Nov, pp. 2541–2563, 2006.

[7] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[8] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[9] R. Gallager, "Low-density parity-check codes," *IRE Transactions on information theory*, vol. 8, no. 1, pp. 21–28, 1962.

[10] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Transactions on communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[11] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of pearl's" belief propagation" algorithm," *IEEE Journal on selected areas in communications*, vol. 16, no. 2, pp. 140–152, 1998.

[12] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on information theory*, vol. 47, no. 2, pp. 599–618, 2001.

[13] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Compressed sensing reconstruction via belief propagation," *preprint*, vol. 14, 2006.

[14] F. Zhang and H. D. Pfister, "Verification decoding of high-rate ldpc codes with applications in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5042–5058, 2012.

[15] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2010, pp. 347–353.

[16] Y. Eftekhari, A. Heidarzadeh, A. H. Banihashemi, and I. Lambadaris, "Density evolution analysis of node-based verification-based algorithms in compressed sensing," *IEEE transactions on information theory*, vol. 58, no. 10, pp. 6616–6645, 2012.

[17] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, no. 2, p. 021005, 2012.

[18] ——, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[19] L. Zdeborová and F. Krzakala, "Statistical physics of inference: Thresholds and algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, 2016.

[20] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[21] M. A. Maleki, *Approximate message passing algorithms for compressed sensing*. Stanford University, 2010.

[22] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Sudocodes fast measurement and reconstruction of sparse signals," in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 2804–2808.

[23] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269–280, 2009.

[24] V. Chandar, D. Shah, and G. W. Wornell, "A simple message-passing algorithm for compressed sensing," in *2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 1968–1972.

[25] A. G. Dimakis, R. Smarandache, and P. O. Vontobel, "Ldpc codes for compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3093–3114, 2012.

[26] M. G. Luby and M. Mitzenmacher, "Verification-based decoding for packet-based low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 120–127, 2005.

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

[28] D. Donoho and A. Montanari, "High dimensional robust m-estimation: Asymptotic variance via approximate message passing," *Probability Theory and Related Fields*, vol. 166, no. 3-4, pp. 935–969, 2016.

[29] A. Montanari, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.

[30] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *2007 IEEE Information Theory Workshop*. IEEE, 2007, pp. 414–419.

[31] ——, "Further results on performance analysis for compressive sensing using expander graphs," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*. IEEE, 2007, pp. 621–625.

[32] M. A. Khajehnejad, A. G. Dimakis, and B. Hassibi, "Nonnegative compressed sensing with minimal perturbed expanders," in *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*. IEEE, 2009, pp. 696–701.

[33] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4299–4308, 2009.

[34] W. Lu, K. Kpalma, and J. Ronsin, "Sparse binary matrices of ldpc codes for compressed sensing," in *Data compression conference (DCC)*, 2012, pp. 10–pages.

[35] J. Zhang, G. Han, and Y. Fang, "Deterministic construction of compressed sensing matrices from protograph ldpc codes," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1960–1964, 2015.

[36] A. Mousavi, G. Dasarathy, and R. G. Baraniuk, "Deepcodec: Adaptive sensing and recovery via deep convolutional neural networks," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2017, pp. 744–744.

[37] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[38] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE transactions on information theory*, vol. 47, no. 2, pp. 619–637, 2001.

[39] S.-Y. Chung, "On the construction of some capacity-approaching coding schemes," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[40] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[42] A. E. Hoerl and R. W. Kennard, "Ridge regression: applications to nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970.

[43] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.

[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[45] G. B. Arfken and H. J. Weber, "Mathematical methods for physicists," 1999.

[46] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[47] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[48] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[49] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[50] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, vol. 1, no. 3.

[51] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

E Discussion of the DE for both regular and irregular designs

First, we explain the physical meaning of the quantities $E^{(t)}$ and $V^{(t)}$, which track the average error and the average variance at the tth iteration, respectively. Since the physical meaning of $V^{(t)}$ can be easily obtained, we focus on the explanation of $E^{(t)}$. For the convenience of the analysis, we rewrite the MAP estimator as

$$\widehat{\boldsymbol{x}} = \text{argmax}_{\boldsymbol{x}} \exp\left(-\frac{\gamma\|\boldsymbol{y}-\mathbf{A}\boldsymbol{x}\|_2^2}{2\sigma^2}\right)\cdot\exp\left(-\gamma f(\boldsymbol{x})\right),$$

where $\gamma > 0$ is a redundant positive constant. Then we restate the message-passing algorithm, which is used to solve the MAP estimator, as

$$\widehat{m}_{a\to i}^{(t+1)}(x_i) \cong \int\prod_{j\in\partial a\backslash i} m_{j\to a}^{(t)}(x_i)\times e^{-\frac{\gamma\left(y_a-\sum_{j=1}^n A_{aj}x_j\right)^2}{2\sigma^2}}dx_j$$
$$m_{i\to a}^{(t+1)}(x_i) \cong e^{-\gamma f(x_i)}\prod_{b\in\partial i\backslash a}\widehat{m}_{b\to i}^{(t+1)}(x_i).$$

The MAP estimator of $\widehat{x}_i$ is hence written as

$$\widehat{x}_i = \text{argmax}_{x_i}\mathbb{P}\left(x_i|\boldsymbol{y}\right) \approx \text{argmax}_{x_i} e^{-\gamma f(x_i)}\prod_{a\in\partial i}\widehat{m}_{a\to i}^{(t)}(x_i).$$

Notice that $\widehat{x}_i$ can be rewritten as the mean w.r.t. the probability measure $e^{-\gamma f(x_i)}\prod_{a\in\partial i}\widehat{m}_{a\to i}^{(t)}$, namely,

$$\widehat{x}_i \approx \int_{x_i} x_i e^{-\gamma f(x_i)}\prod_{a\in\partial i}\widehat{m}_{a\to i}^{(t)}(x_i)dx_i,$$

by letting $\gamma\to\infty$. Since the mean $\mu_{i\to a}$ is computed as

$$\mu_{i\to a} = \int_{x_i} x_i e^{-\gamma f(x_i)}\prod_{b\in\partial i\backslash a}\widehat{m}_{b\to i}^{(t)}(x_i)dx_i,$$

which is close to $\widehat{x}_i$, we obtain the approximation $m^{-1}\sum_{a=1}^m(\mu_{i\to a}-x_i^\natural)^2$ as $(\widehat{x}_i-x_i^\natural)^2$. We then conclude

$$E^{(t)} = \frac{1}{mn}\sum_{i=1}^n\sum_{a=1}^m\left(\mu_{i\to a}-x_i^\natural\right)^2 \approx \frac{1}{n}\sum_{i=1}^n\left(\widehat{x}_i-x_i^\natural\right)^2,$$

which is approximately the average of error at the tth iteration. Having discussed the physical meaning of the quantities $E^{(t)}$ and $V^{(t)}$, we turn to the derivation of the DE equation.

A. Supporting Lemmas

We begin the derivation with the following lemma, which is stated as

Lemma 1. *Consider the message flow $\widehat{m}_{a\to i}^{(t+1)}$ from the check node a to the variable node i and approximate it as a Gaussian RV with mean $\widehat{\mu}_{a\to i}^{(t+1)}$ and variance $\widehat{v}_{a\to i}^{(t+1)}$, i.e., $\widehat{m}_{a\to i}^{(t+1)} \sim \mathsf{N}\left(\widehat{\mu}_{a\to i}^{(t+1)},\widehat{v}_{a\to i}^{(t+1)}\right)$. Then, we can obtain the following update equation at the $(t+1)$th iteration*

$$\widehat{\mu}_{a\to i}^{(t+1)} = x_i + A\sum_{j\in\partial a\backslash i}A_{ai}A_{aj}\left(x_j-\ \mu_{j\to a}^{(t)}\right) + AA_{ai}w_a;$$
$$\widehat{v}_{a\to i}^{(t+1)} = A\sigma^2 + |\partial a|V^{(t)},$$

where $|\partial a|$ denotes the degree of the check node a.

Proof. Consider the message flow $\widehat{m}_{a\to i}^{(t+1)}$ from check-node to variable node at the $(t+1)$th iteration

$$\widehat{m}_{a\to i}^{(t+1)} = \frac{1}{Z_{a\to i}^t}\int\prod_{j\in\partial a\backslash i}m_{j\to a}^{(t)}(x_j) \times\ \exp\left(-\frac{\gamma\left(y_a-\sum_{j=1}A_{aj}x_j\right)^2}{2\sigma^2}\right)dx_j. \tag{39}$$

Approximate the message flow $m_{j\to a}^{(t+1)}$ as a Gaussian RV with mean $\mu_{j\to a}^{(t+1)}$ and variance $v_{j\to a}^{(t+1)}$. Plugging into (39) yields

$$\widehat{m}_{a\to i}^{(t+1)} = \frac{1}{Z_{a\to i}^t}\int\prod_{j\in\partial a\backslash i}\exp\left[-\frac{\gamma\left(x_j-\mu_{j\to a}^{(t)}\right)^2}{2v_{j\to a}^{(t+1)}}\right] \times\ \exp\left(-\frac{\gamma\left(y_a-\sum_{j=1}A_{aj}x_j\right)^2}{2\sigma^2}\right)dx_j. \tag{40}$$

The direct calculation of the above integral involves the cross terms such as $A_{aj_1}A_{aj_2}x_{j_1}x_{j_2}$ $(j_1\neq j_2)$, which can be cumbersome. To handle this issue, we adopt the trick in [1], [18], whose basic idea is to introduce a redundant variable ω and exploit the relation

$$e^{-\frac{t^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}\int e^{-\frac{\omega^2}{2\sigma^2}+\frac{it\omega}{\sigma^2}}d\omega,$$

where t is an arbitrary number. As such, we can transform (40) to

$$\widehat{m}_{a\to i}^{(t+1)} \cong \int d\omega\prod_{j\in\partial a\backslash i}dx_j\cdot\exp\left[-\frac{\gamma\left(x_j-\mu_{j\to a}^{(t)}\right)^2}{2v_{j\to a}^{(t+1)}}\right] \times\exp\left[-\frac{i\omega\gamma\left(y_a-\sum_{j=1}A_{aj}x_j\right)}{\sigma^2}\right]\cdot\exp\left[-\frac{\gamma\omega^2}{2\sigma^2}\right],$$

which diminishes the cross term $x_{j_1}x_{j_2}$ $(j_1\neq j_2)$. Rearranging the terms for each x_j, we can iteratively perform the integral such that

$$\int dx_j\cdot\exp\left(-\frac{\gamma\left(x_j-\mu_{j\to a}^{(t)}\right)^2}{2v_{j\to a}^{(t)}}+\frac{i\omega\gamma A_{aj}x_j}{\sigma^2}\right)$$
$$= \sqrt{\frac{2\pi v_{j\to a}^{(t)}}{\gamma}}\cdot\exp\left[-\frac{\gamma(\widehat{\mu}_{j\to a}^{(t)})^2}{2\widehat{v}_{j\to a}^{(t)}}+\frac{v_{j\to a}^{(t)}\left(\frac{\gamma\mu_{j\to a}^{(t)}}{v_{j\to a}^{(t)}}+\frac{i\gamma\omega A_{aj}}{\sigma^2}\right)^2}{2\gamma}\right]$$
$$= \sqrt{\frac{2\pi v_{j\to a}^{(t)}}{\gamma}}\cdot\exp\left(-\frac{\gamma\omega^2A_{aj}^2v_{j\to a}^{(t)}}{2\sigma^4}+\frac{i\gamma\omega A_{aj}\mu_{j\to a}^{(t)}}{\sigma^2}\right).$$

With some algebraic manipulations, we can compute its mean $\widehat{\mu}_{a\to i}^{(t+1)}$ and its variance $\widehat{v}_{a\to i}^{(t+1)}$ as

$$\widehat{\mu}_{a\to i}^{(t+1)} = \frac{A_{ai}\left(y_a-\sum_{j\in\partial a\backslash i}A_{aj}\mu_{j\to a}^{(t)}\right)}{A_{ai}^2};$$
$$\widehat{v}_{a\to i}^{(t+1)} = \frac{\sigma^2+\sum_{j\in\partial a\backslash i}A_{aj}^2v_{j\to a}^{(t)}}{A_{ai}^2}.$$

The following analysis focuses on how to approximate these two values. We begin by the discussion w.r.t. the variance $\widehat{v}_{a\to i}^{(t+1)}$. Note we have

$$\widehat{v}_{a\to i}^{(t+1)} \overset{①}{\approx} A\sigma^2 + \sum_{j\in\partial a\setminus i} v_{j\to a}^{(t)},$$

where in ① we use $A_{ai}^2 \approx \mathbb{E}\left(A_{ai}^2 | A_{ai}\neq 0\right) = A^{-1}$ for $i\in\partial a$. As for the sum $\sum_{j\in\partial a\setminus i} v_{j\to a}^{(t)}$, we can view it to be randomly sampled from the set of variances $\left\{v_{j\to a}^{(t)}\right\}$ and approximate it as

$$\sum_{j\in\partial a\setminus i} v_{j\to a}^{(t)} \approx \left(|\partial a|-1\right) V^{(t)} \approx |\partial a| V^{(t)}.$$

Notice that the variance is closely related with the check node degree $|\partial a|$. Having obtained the variance $\widehat{v}_{a\to i}^{(t+1)}$, we turn to the mean $\widehat{\mu}_{a\to i}^{(t+1)}$, which is computed as

$$\begin{aligned}
\widehat{\mu}_{a\to i}^{(t+1)} =& \frac{A_{ai}\left(y_a - \sum_{j\in\partial a\setminus i} A_{aj}\mu_{j\to a}^{(t)}\right)}{A_{ai}^2} \\
\overset{②}{\approx}& AA_{ai}\left(A_{ai}x_i + \sum_{j\in\partial a\setminus i} A_{aj}\left(x_j - \mu_{j\to a}^{(t)}\right) + w_a\right) \\
\overset{③}{\approx}& x_i + A\sum_{j\in\partial a\setminus i} A_{ai}A_{aj}\left(x_j - \mu_{j\to a}^{(t)}\right) + AA_{ai}w_a,
\end{aligned}$$

where in ② and ③ we use the approximation $A_{ai}^2\approx A^{-1}$ for $i\in\partial a$.

□

B. Derivation of DE

We study the message flow $m_{i\to a}^{(t+1)}$ from the variable node i to the check node a

$$m_{i\to a}^{(t+1)} \cong e^{-\gamma f(x_i)} \prod_{b\in\partial i\setminus a} e^{-\frac{\gamma\left(x_i-\widehat{\mu}_{b\to i}^{(t+1)}\right)^2}{2\widehat{v}_{b\to i}^{(t+1)}}}.$$

To begin with, we study the product $\prod_{b\in\partial i\setminus a}\exp\Bigg(-\frac{\gamma\left(x_i-\widehat{\mu}_{b\to i}^{(t+1)}\right)^2}{2\widehat{v}_{b\to i}^{(t)}}\Bigg)$. Its variance $\widetilde{v}_{i\to a}^{(t+1)}$ is approximately computed as

$$\frac{\gamma}{\widetilde{v}_{i\to a}^{(t+1)}} \approx \sum_{b\in\partial i\setminus a}\frac{\gamma}{\widehat{v}_{b\to i}^{(t+1)}},$$

which yields

$$\widetilde{v}_{i\to a}^{(t+1)} = \left(\frac{|\partial i|-1}{A\sigma^2+|\partial a|V^{(t)}}\right)^{-1} \approx \frac{A\sigma^2+|\partial a|V^{(t)}}{|\partial i|}.$$

Further, the mean $\widetilde{\mu}_{i\to a}^{(t+1)}$ is calculated as

$$\begin{aligned}
\widetilde{\mu}_{i\to a}^{(t+1)} =& \left(\sum_{b\in\partial i\setminus a}\frac{\widehat{\mu}_{b\to i}^{(t+1)}}{\widehat{v}_{b\to i}^{(t+1)}}\right) / \left(\sum_{b\in\partial i\setminus a}\frac{1}{\widehat{v}_{b\to i}^{(t+1)}}\right)^{-1} \\
\overset{①}{=}& \frac{A\sigma^2+|\partial a|V^{(t)}}{|\partial i|} \\
&\times\left(\sum_{b\in\partial i\setminus a}\frac{x_i + A\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right) + AA_{bi}w_b}{A\sigma^2+|\partial a|V^{(t)}}\right) \\
\approx& x_i + \frac{A}{|\partial i|}\left[\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right) + \sum_{b\in\partial i\setminus a}A_{bi}w_b\right],
\end{aligned}$$

where in ① we invoke Lemma 1. We then approximate the term $\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right) + \sum_{b\in\partial i\setminus a}A_{bi}w_b$ as a Gaussian RV with its mean being calculated as

$$\mathbb{E}\left[\sum_{b\in\partial i\setminus a}\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right) + \sum_{b\in\partial i\setminus a}A_{bi}w_b\right] = 0,$$

and its variance as

$$\begin{aligned}
&\mathbb{E}\left[\sum_{b\in\partial i\setminus a}\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right) + \sum_{b\in\partial i\setminus a}A_{bi}w_b\right]^2 \\
=& \mathbb{E}\left[\sum_{b\in\partial i\setminus a}\sum_{j\in\partial b\setminus i}A_{bi}A_{bj}\left(x_j - \mu_{j\to b}^{(t)}\right)\right]^2 + \mathbb{E}\left[\sum_{b\in\partial i\setminus a}A_{bi}w_b\right]^2 \\
\approx& A^{-2}|\partial i|\sum_{j\in\partial a\setminus i}\left(x_j-\mu_{j\to b}^{(t)}\right)^2 + A^{-1}\sigma^2|\partial i| \\
\overset{②}{\approx}& |\partial i|\left(A^{-2}|\partial a|E^{(t)} + A^{-1}\sigma^2\right).
\end{aligned}$$

In ② we assume the term $\left(x_j-\mu_{j\to b}^{(t)}\right)^2$ is randomly sampled among all possible pairs (i,a). Hence for the fixed degree $|\partial i|$ and $|\partial a|$, we can approximate the mean $\widetilde{\mu}_{i\to a}^{(t+1)}$ as a Gaussian RV with mean $x_i + z\sqrt{\left(A\sigma^2+|\partial a|E^{(t)}\right)/|\partial i|}$ and variance $\left(A\sigma^2+|\partial a|V^{(t)}\right)/|\partial i|$, namely,

$$x\sim\mathsf{N}\left(x_i + z\sqrt{\frac{A\sigma^2+|\partial a|E^{(t)}}{|\partial i|}}, \frac{A\sigma^2+|\partial a|V^{(t)}}{|\partial i|}\right),$$

where z is a standard normal RV. Recalling that the distribution of the degrees of the variable node i and check node a satisfies $\mathbb{P}(|\partial i|=\alpha)=\lambda_\alpha$ and $\mathbb{P}(|\partial a|=\beta)=\rho_\beta$, we can approximate the distribution of the product $\prod_{b\in\partial i\setminus a}\exp\left[-\gamma\left(x_i-\widehat{\mu}_{b\to i}^{(t)}\right)^2/(2\widehat{v}_{b\to i}^{(t)})\right]$ as the mixture Gaussian $\sum_{i,j}\rho_i\lambda_j\mathsf{N}\left(z\sqrt{\frac{iE^{(t)}+A\sigma^2}{j}}, \frac{A\sigma^2+iV^{(t)}}{j}\right)$ [5] and further approximate it as a single Gaussian RV with mean $x_i + \sum_{i,j}\rho_i\lambda_j z\sqrt{\frac{iE^{(t)}+A\sigma^2}{j}}$ and variance $\sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2+iV^{(t)}}{j}$. Invoking the definitions of $h_{\text{mean}}(\cdot;\cdot)$ and $h_{\text{var}}(\cdot;\cdot)$ as in (8), we then approximate the mean $\mu_{i\to a}^{(t+1)}$ and the variance $v_{i\to a}^{(t+1)}$ as

$$\begin{aligned}
\mu_{i\to a}^{(t+1)} \approx h_{\text{mean}}\Bigg(&x_i + z\sum_{i,j}\rho_i\lambda_j\sqrt{\frac{iE^{(t)}+A\sigma^2}{j}}; \\
&\sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2+iV^{(t)}}{j}\Bigg);
\end{aligned}$$

[5] One hidden assumption is that there is no-local loops in the graphical model we constructed, which is widely used in the previous work [7].

$$v_{i\to a}^{(t+1)} \approx h_{\text{var}}\bigg(x_i + z\sum_{i,j}\rho_i\lambda_j\sqrt{\frac{iE^{(t)}+A\sigma^2}{j}};\ \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2+iV^{(t)}}{j}\bigg).$$

Then, the DE w.r.t. the average error $E^{(t+1)}$ is derived as

$$\begin{aligned}E^{(t+1)} &= \frac{1}{mn}\sum_{a=1}^{m}\sum_{i=1}^{n}\left(\mu_{i\to a}^{(t+1)} - x_i^{\natural}\right)^2\\ &\approx \mathbb{E}_{\text{prior}(s)}\mathbb{E}_z\bigg[h_{\text{mean}}\bigg(x_i^{\natural} + z\sum_{i,j}\rho_i\lambda_j\sqrt{\frac{iE^{(t)}+A\sigma^2}{j}};\ \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2+iV^{(t)}}{j}\bigg) - x_i^{\natural}\bigg]^2.\end{aligned}$$

Following a similar method, we obtain the DE w.r.t. the average variance $V^{(t+1)}$ as stated in (6). This completes the proof.

C. Derivation of DE for Irregular Design

Different from the regular design, we separately track the average error and average variance w.r.t. the high-priority part and low-priority part. Then we define four quantities, namely, $E_{\mathsf{L}}^{(t)}, E_{\mathsf{H}}^{(t)}$, $V_{\mathsf{L}}^{(t)}$, and $V_{\mathsf{H}}^{(t)}$, which are written as

$$\begin{aligned}E_{\mathsf{L}}^{(t)} &= \frac{1}{mn_{\mathsf{L}}}\sum_{a=1}^{m}\sum_{i\in\mathsf{L}}\left(\mu_{i\to a}^{(t)} - x_i^{\natural}\right)^2;\\ E_{\mathsf{H}}^{(t)} &= \frac{1}{mn_{\mathsf{H}}}\sum_{a=1}^{m}\sum_{i\in\mathsf{H}}\left(\mu_{i\to a}^{(t)} - x_i^{\natural}\right)^2;\\ V_{\mathsf{L}}^{(t)} &= \frac{1}{mn_{\mathsf{L}}}\sum_{a=1}^{m}\sum_{i\in\mathsf{L}}v_{i\to a}^{(t)};\\ V_{\mathsf{H}}^{(t)} &= \frac{1}{mn_{\mathsf{H}}}\sum_{a=1}^{m}\sum_{i\in\mathsf{H}}v_{i\to a}^{(t)},\end{aligned}$$

where n_{H} and n_{L} denote the length of the high-priority part $\boldsymbol{x}_{\mathsf{H}}^{\natural}$ and low-priority part $\boldsymbol{x}_{\mathsf{L}}^{\natural}$, respectively. Following the same procedure as above then yields the proof of (17). The derivation details are omitted for the clarify of presentation.

F Discussion of Subsection IV-C

We start the discussion by outlining the DE equation w.r.t. $E_{\mathsf{H}}^{(t)}, E_{\mathsf{L}}^{(t)}, V_{\mathsf{H}}^{(t)}$, and $V_{\mathsf{L}}^{(t)}$

$$\begin{aligned}E_{\mathsf{H}}^{(t+1)} &= \underbrace{\mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[\mathsf{prox}\left(s + z\cdot b_{\mathsf{H},1}^{(t)};\beta_{\mathsf{H}}b_{\mathsf{H},2}^{(t)}\right) - s\right]^2}_{\triangleq\Psi_{E,\mathsf{H}}\left(E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)};\\ E_{\mathsf{L}}^{(t+1)} &= \underbrace{\mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[\mathsf{prox}\left(s + z\cdot b_{\mathsf{L},1}^{(t)};\beta_{\mathsf{L}}b_{\mathsf{L},2}^{(t)}\right) - s\right]^2}_{\triangleq\Psi_{E,\mathsf{L}}\left(E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)};\\ V_{\mathsf{H}}^{(t+1)} &= \underbrace{\mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[\beta_{\mathsf{H}}b_{\mathsf{H},2}\cdot\mathsf{prox}'\left(s + z\cdot b_{\mathsf{H},1}^{(t)};\beta_{\mathsf{H}}b_{\mathsf{H},2}^{(t)}\right)\right]}_{\triangleq\Psi_{V,\mathsf{H}}\left(E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)};\\ V_{\mathsf{L}}^{(t+1)} &= \underbrace{\mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathsf{N}(0,1)}\left[\beta_{\mathsf{L}}b_{\mathsf{L},2}\cdot\mathsf{prox}'\left(s + z\cdot b_{\mathsf{L},1}^{(t)};\beta_{\mathsf{L}}b_{\mathsf{L},2}^{(t)}\right)\right]}_{\triangleq\Psi_{V,\mathsf{L}}\left(E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)},\end{aligned}$$

where notation $\mathsf{prox}(a;b)$ is the soft-thresholding estimator defined as $\text{sign}(a)\max(|a|-b,0)$, notation $\mathsf{prox}'(a;b)$ is the derivative w.r.t. the first argument, and the notations $b_{\mathsf{H},1}^{(t)}, b_{\mathsf{H},2}^{(t)}, b_{\mathsf{L},1}^{(t)}$, and $b_{\mathsf{L},2}^{(t)}$ are defined as

$$\begin{aligned}b_{\mathsf{H},1}^{(t)} &= \sum_{\ell,i,j}\lambda_{\mathsf{H},\ell}\rho_{\mathsf{H},i}\rho_{\mathsf{L},j}\sqrt{\frac{A\sigma^2+iE_{\mathsf{H}}^{(t)}+jE_{\mathsf{L}}^{(t)}}{\ell}};\\ b_{\mathsf{H},2}^{(t)} &= \sum_{\ell,i,j}\lambda_{\mathsf{H},\ell}\rho_{\mathsf{H},i}\rho_{\mathsf{L},j}\frac{A\sigma^2+iV_{\mathsf{H}}^{(t)}+jV_{\mathsf{L}}^{(t)}}{\ell};\\ b_{\mathsf{L},1}^{(t)} &= \sum_{\ell,i,j}\lambda_{\mathsf{L},\ell}\rho_{\mathsf{L},i}\rho_{\mathsf{H},j}\sqrt{\frac{A\sigma^2+iE_{\mathsf{L}}^{(t)}+jE_{\mathsf{H}}^{(t)}}{\ell}};\\ b_{\mathsf{L},2}^{(t)} &= \sum_{\ell,i,j}\lambda_{\mathsf{L},\ell}\rho_{\mathsf{L},i}\rho_{\mathsf{H},j}\frac{A\sigma^2+iV_{\mathsf{L}}^{(t)}+jV_{\mathsf{H}}^{(t)}}{\ell}.\end{aligned}$$

Similar to the proof in Section A, we define the differences across iterations as

$$\begin{aligned}\delta_{E,\mathsf{H}}^{(t)} &\triangleq E_{\mathsf{H}}^{(t+1)} - E_{\mathsf{H}}^{(t)};\quad \delta_{E,\mathsf{L}}^{(t)} \triangleq E_{\mathsf{L}}^{(t+1)} - E_{\mathsf{L}}^{(t)};\\ \delta_{V,\mathsf{H}}^{(t)} &\triangleq V_{\mathsf{H}}^{(t+1)} - V_{\mathsf{H}}^{(t)};\quad \delta_{V,\mathsf{L}}^{(t)} \triangleq V_{\mathsf{H}}^{(t+1)} - V_{\mathsf{H}}^{(t)}.\end{aligned}$$

A. Discussion of (23)

This subsection follows the same logic as in Section A. We first relax the Requirement 1 w.r.t. the average variance $V_{\mathsf{H}}^{(t)}$ and $V_{\mathsf{L}}^{(t)}$. Performing the Taylor-expansion, we obtain

$$\begin{aligned}&\delta_{V,\mathsf{H}}^{(t+1)}\\ &= \Psi_{V,\mathsf{H}}\left(V_{\mathsf{H}}^{(t+1)},V_{\mathsf{L}}^{(t+1)},E_{\mathsf{H}}^{(t+1)},E_{\mathsf{L}}^{(t+1)}\right)\\ &- \Psi_{V,\mathsf{H}}\left(V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)},E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)}\right)\\ &= \left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\Big|_{E_{\mathsf{H}}=E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}=E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}=V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}=V_{\mathsf{L}}^{(t)}}\right)\delta_{E,\mathsf{H}}^{(t)}\\ &+ \left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\Big|_{E_{\mathsf{H}}=E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}=E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}=V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}=V_{\mathsf{L}}^{(t)}}\right)\delta_{E,\mathsf{L}}^{(t)}\\ &+ \left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\Big|_{E_{\mathsf{H}}=E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}=E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}=V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}=V_{\mathsf{L}}^{(t)}}\right)\delta_{V,\mathsf{H}}^{(t)}\\ &+ \left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\Big|_{E_{\mathsf{H}}=E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}=E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}=V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}=V_{\mathsf{L}}^{(t)}}\right)\delta_{V,\mathsf{L}}^{(t)}\\ &+ O\left[(\delta_{V,\mathsf{H}}^{(t)})^2\right] + O\left[(\delta_{V,\mathsf{L}}^{(t)})^2\right].\end{aligned}\tag{41}$$

Following the same logic in Section A, our derivation consists of two parts:

- **Part I**. We verify that $(0,0)$ is a fixed point of the DE equation w.r.t. $V_{\mathsf{H}}^{(t)}$ and $V_{\mathsf{L}}^{(t)}$;
- **Part II**. We show the DE equation w.r.t. $V_{\mathsf{H}}^{(t)}$ and $V_{\mathsf{L}}^{(t)}$ converges within the proximity of the origin points.

Our following derivation focuses on showing that DE converges, or equivalently, $\lim_{t\to\infty}\left(\delta_{V,\mathsf{H}}^{(t)},\delta_{V,\mathsf{L}}^{(t)}\right) = (0,0)$, as the second part can be easily verified. We consider the region where $V_{\mathsf{H}}^{(t)}, V_{\mathsf{L}}^{(t)}, \delta_{V,\mathsf{H}}^{(t)}$, and $\delta_{V,\mathsf{L}}^{(t)}$ are sufficiently small and hence can safely omit the quadratic terms in (41). Exploiting the fact that $\partial\Psi_{V,\mathsf{H}}/\partial E_{\mathsf{H}} = 0$ and $\partial\Psi_{V,\mathsf{H}}/\partial E_{\mathsf{L}} = 0$, we obtain the linear relation

$$\begin{bmatrix}\delta_{V,\mathsf{H}}^{(t+1)}\\ \delta_{V,\mathsf{L}}^{(t+1)}\end{bmatrix} = \underbrace{\begin{bmatrix}\left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)} & \left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{L}}}\right)^{(t)}\\ \left(\frac{\partial\Psi_{V,\mathsf{L}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)} & \left(\frac{\partial\Psi_{V,\mathsf{L}}(\cdot)}{\partial V_{\mathsf{L}}}\right)^{(t)}\end{bmatrix}}_{\mathbf{L}_V^{(t)}} \begin{bmatrix}\delta_{V,\mathsf{H}}^{(t)}\\ \delta_{V,\mathsf{L}}^{(t)}\end{bmatrix},$$

where the notation $\left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)}$ is an abbreviation for the gradient

$$\left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)} = \frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\Big|_{E_{\mathsf{H}}=E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}=E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}=V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}=V_{\mathsf{L}}^{(t)}}.$$

Similarly we define the notations $\left(\partial\Psi_{V,\mathsf{H}}(\cdot)/\partial V_{\mathsf{L}}\right)^{(t)}$, $\left(\partial\Psi_{V,\mathsf{L}}(\cdot)/\partial V_{\mathsf{H}}\right)^{(t)}$, and $\left(\partial\Psi_{V,\mathsf{L}}(\cdot)/\partial V_{\mathsf{L}}\right)^{(t)}$. Then we require $\inf_t \|\mathbf{L}_V^{(t)}\|_{\mathsf{OP}} \le 1$. Otherwise, the values of $\delta_{V,\mathsf{H}}^{(t)}$ and $\delta_{V,\mathsf{L}}^{(t)}$ will keep increasing and stay away from zero. We then lower bound the gradients $\left(\partial\Psi_{V,\mathsf{H}}(\cdot)/\partial V_{\mathsf{H}}\right)^{(t)}$ and $\left(\partial\Psi_{V,\mathsf{H}}(\cdot)/\partial V_{\mathsf{L}}\right)^{(t)}$ as

$$\begin{aligned}
&\left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)} \overset{①}{=} \beta_{\mathsf{H}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{H},i}\right)\\
&\times\left[2\left(1-\frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\right)\Phi\left(-\frac{\beta_{\mathsf{H}}b_{\mathsf{H},2}^{(t)}}{b_{\mathsf{H},1}^{(t)}}\right)+\frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\right]\\
&\overset{②}{\ge} \frac{k_{\mathsf{H}}\beta_{\mathsf{H}}}{n_{\mathsf{H}}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{H},i}\right);\\
&\left(\frac{\partial\Psi_{V,\mathsf{H}}(\cdot)}{\partial V_{\mathsf{L}}}\right)^{(t)} \overset{③}{=} \beta_{\mathsf{H}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{L},i}\right)\\
&\times\left[2\left(1-\frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\right)\Phi\left(-\frac{\beta_{\mathsf{H}}b_{\mathsf{H},2}^{(t)}}{b_{\mathsf{H},1}^{(t)}}\right)+\frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\right]\\
&\overset{④}{\ge} \frac{k_{\mathsf{H}}\beta_{\mathsf{H}}}{n_{\mathsf{H}}}\left(\sum_\ell \frac{\lambda_{\mathsf{H},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{L},i}\right),
\end{aligned}$$

where $\Phi(\cdot)=(2\pi)^{-1/2}\int_{-\infty}^{(\cdot)} e^{-z^2/2}dz$ is the CDF of the standard normal RV z, i.e., $z\sim\mathsf{N}(0,1)$. In ① and ③, we follow the same computation procedure as in (33), and in ② and ④ we drop the non-negative terms $\Phi(\cdot)$. Following a similar procedure, we lower bound the gradients $\left(\partial\Psi_{V,\mathsf{L}}(\cdot)/\partial V_{\mathsf{H}}\right)^{(t)}$ and $\left(\partial\Psi_{V,\mathsf{L}}(\cdot)/\partial V_{\mathsf{L}}\right)^{(t)}$ as

$$\begin{aligned}
\left(\frac{\partial\Psi_{V,\mathsf{L}}(\cdot)}{\partial V_{\mathsf{H}}}\right)^{(t)} &\ge \frac{k_{\mathsf{L}}\beta_{\mathsf{L}}}{n_{\mathsf{L}}}\left(\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{H},i}\right);\\
\left(\frac{\partial\Psi_{V,\mathsf{L}}(\cdot)}{\partial V_{\mathsf{L}}}\right)^{(t)} &\ge \frac{k_{\mathsf{L}}\beta_{\mathsf{L}}}{n_{\mathsf{L}}}\left(\sum_\ell \frac{\lambda_{\mathsf{L},\ell}}{\ell}\right)\cdot\left(\sum_i i\rho_{\mathsf{L},i}\right),
\end{aligned}$$

and conclude the discussion.

B. Discussion of (24)

This subsection relaxes the requirement $\lim_{t\to\infty}E_{\mathsf{H}}^{(t)}=0$, which consists of two parts:

- **Part I.** we consider the necessary conditions such that DE equation w.r.t. $E_{\mathsf{H}}^{(t)}$ converges;
- **Part II**. We verify that 0 is a fixed point of DE w.r.t. $E_{\mathsf{H}}^{(t)}$ given that $\lim_{t\to\infty}\left(V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)=(0,0)$.

Since the second part can be easily verified, we focus on the first part. We consider the region where $E_{\mathsf{H}}^{(t)}$ and $\delta_{E,\mathsf{H}}^{(t)}$ are all sufficiently small and require $\delta_{E,\mathsf{H}}^{(t)}$ to converge to zero. Via the Taylor expansion, we obtain the following linear equation

$$\delta_{E,H}^{(t+1)} \approx \left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)}\delta_{E,\mathsf{H}}^{(t)} + \left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)}\delta_{E,\mathsf{L}}^{(t)}, \tag{42}$$

where $\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)}$ denotes the gradient $\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}$ at the point $\left(E_{\mathsf{H}}^{(t)},E_{\mathsf{L}}^{(t)},V_{\mathsf{H}}^{(t)},V_{\mathsf{L}}^{(t)}\right)$. Enforcing the variable $\delta_{E,\mathsf{H}}^{(t)}$ to converge to zero, we require

$$\inf_t\left[\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)}\right]^2+\left[\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)}\right]^2\le 1.$$

Then our goal becomes lower-bounding the gradients, which are written as

$$\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)} \ge \frac{k_{\mathsf{H}}b_{\mathsf{H},1}^{(t)}}{n_{\mathsf{H}}}\left(\sum_\ell\frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)\left(\sum_{i,j}\frac{i\rho_{\mathsf{H},i}\rho_{\mathsf{L},j}}{\sqrt{iE_{\mathsf{H}}^{(t)}+jE_{\mathsf{L}}^{(t)}}}\right); \tag{43}$$

$$\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)} \ge \frac{k_{\mathsf{H}}b_{\mathsf{H},1}^{(t)}}{n_{\mathsf{H}}}\left(\sum_\ell\frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)\left(\sum_{i,j}\frac{j\rho_{\mathsf{H},i}\rho_{\mathsf{L},j}}{\sqrt{iE_{\mathsf{H}}^{(t)}+jE_{\mathsf{L}}^{(t)}}}\right). \tag{44}$$

Taking the limit $E_{\mathsf{H}}^{(t)}\to 0$, we can conclude the relaxation by simplifying (43) and (44) as

$$\begin{aligned}
\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)} &\ge \frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\left(\sum_\ell\frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)^2\left(\sum_i\sqrt{i}\rho_{\mathsf{H},i}\right)^2;\\
\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)} &\ge \frac{k_{\mathsf{H}}}{n_{\mathsf{H}}}\left(\sum_\ell\frac{\lambda_{\mathsf{H},\ell}}{\sqrt{\ell}}\right)^2\left(\sum_i\sqrt{i}\rho_{\mathsf{L},i}\right)^2.
\end{aligned}$$

C. Discussion of (25)

The basic idea is to linearize the DE update equation with Taylor expansion and enforce the difference $\delta_{V,\mathsf{H}}^{(t)}$ to decrease at a faster rate than $\delta_{V,\mathsf{L}}^{(t)}$:

$$\begin{aligned}
\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)} &\le \left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)};\\
\left(\frac{\Psi_{E,\mathsf{H}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)} &\le \left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)}.
\end{aligned} \tag{45}$$

Following the same logic as (43) and (44), we can lower-bound the gradients $\left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)}$ and $\left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)}$ as

$$\begin{aligned}
\left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{H}}}\right)^{(t)} &\ge \frac{k_{\mathsf{L}}}{n_{\mathsf{L}}}\left(\sum_\ell\frac{\lambda_{\mathsf{L},\ell}}{\sqrt{\ell}}\right)^2\left(\sum_i\sqrt{i}\rho_{\mathsf{H},i}\right)^2;\\
\left(\frac{\Psi_{E,\mathsf{L}}(\cdot)}{\partial E_{\mathsf{L}}}\right)^{(t)} &\ge \frac{k_{\mathsf{L}}}{n_{\mathsf{L}}}\left(\sum_\ell\frac{\lambda_{\mathsf{L},\ell}}{\sqrt{\ell}}\right)^2\left(\sum_i\sqrt{i}\rho_{\mathsf{L},i}\right)^2.
\end{aligned}$$

Combining with (45) will then yield the Requirement 2.

REFERENCES

[1] H. Nishimori, *Statistical physics of spin glasses and information processing: an introduction*. Clarendon Press, 2001, no. 111.

[2] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[3] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.