A General Framework for the Design of Compressive Sensing using Density Evolution

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

2021 IEEE Information Theory Workshop (ITW) | 978-1-6654-0312-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITW48936.2021.9611415

***Abstract*—This paper proposes a general framework to design a sparse sensing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, in a linear measurement system $\mathbf{y} = \mathbf{A}\mathbf{x}^{\natural} + \mathbf{w}$, where $\mathbf{y} \in \mathbb{R}^{n}$, $\mathbf{x}^{\natural} \in \mathbb{R}^{n}$, and $\mathbf{w}$ 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, and provide a framework for the design of matrix A. Particularly, compared to the previous methods, our proposed framework enjoys the following desirable properties: (i) Universality: the design supports both regular sensing and preferential sensing, and incorporates them in a single framework; (ii) Flexibility: the framework can easily adapt the design of A to a signal $\mathbf{x}^{\natural}$ with different underlying structures. As an illustration, we consider the ℓ_1 regularizer, which correspond to Lasso, for both the regular sensing and preferential sensing scheme. Noteworthy, our framework can reproduce the classical result of Lasso, i.e., $m \geq c_0 k \log(n/k)$ (the regular sensing) with regular design after proper distribution approximation, where $c_0 > 0$ is some fixed constant. We also provide numerical experiments to confirm the analytical results and demonstrate the superiority of our framework whenever a preferential treatment of a sub-block of vector $\mathbf{x}^{\natural}$ is required.**

I. Introduction

This paper considers the linear sensing relation, which is written as

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

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

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

where $f(\cdot)$ is the regularizer used to enforce a desired structure for $\widehat{\mathbf{x}}$. To ensure reliable recovery of $\mathbf{x}^{\natural}$, sensing matrix $\mathbf{A}$ needs to satisfy certain conditions, for example, the incoherence in [1], RIP in [2], [3], the *neighborhood stability* in [4], *irrespresentable condition* in [5], etc. However, the above conditions all corresponds for sparse $\mathbf{x}^{\natural}$. For other types of signals, the preferable conditions for $\mathbf{A}$ may not be the same. In this work, we focus on the sparse sensing matrix $\mathbf{A}$, which arises frequently in practical applications where a limited number of sensors are preferred. 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 unequal preference on the quality of signal components.

Related work. At the core of our work is the application of DE in message passing (MP) algorithm. MP [6], [7] has a broad spectrum of applications, ranging from physics to coding theory. When narrowing down to the *compressed sensing* (CS), it has been widely used for signal reconstruction [8]–[16] and analyzing the performance under some specific sensing matrices.

In the context of the sparse sensing matrix, the authors in [17] first proposed a so-called sudocode construction technique and later presented a decoding algorithm based on the MP in [18]. In [19], the non-negative sparse signal $\mathbf{x}^{\natural}$ is considered under the binary sensing matrix. The work in [20] linked the channel encoding with the CS and presented a deterministic way of constructing sensing matrix based on high-girth *low-density parity-check* (LDPC) code. In [9], [11], [21], the authors considered the verification-based decoding and analyzed its performance with DE. In [10], the spatial coupling is first introduced into CS and is evaluated with the decoding scheme adapted from [21]. However, all the above mentioned works focused on the noiseless setting, namely, $\mathbf{w} = \mathbf{0}$. In [12]–[14], the noisy measurement is considered. A sparse sensing matrix based on spatial coupling is analyzed in the large system limit with replica method and DE. They proved its recovery performance to be optimal when m increases at the same rate of n, i.e., $m = O(n)$.

Moreover, in the context of a dense sensing matrix, the analytical tool switches from DE to *state evolution* (SE), which is first proposed in [15], [16]. 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, [22] provided a rigorous proof for the phase transition property by the conditioning technique from Erwin Bolthausen and [23] 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 [6], [14], [24].

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 [25]–[31].

Contributions. Compared to the previous work exploiting MP [9]–[16], [21], our focus is on the sensing matrix design rather than the decoding scheme, which is based on the M-estimator with regularizers. Additionally, our framework enjoys the following benefits:

- **Universibility**. Exploiting the DE, we give a universal framework which supports both the regular sensing and the preferential sensing in recovering the components of the signal. Specifically, the preferential sensing design of sensing matrix equips the compressive sensing method for more accurate (or exact) recovery of the high-priority sub-block of the signal relative to the low-priority sub-block. We emphasize that although we focused on two levels of priority in signal components in this work, one can easily extend the framework to the scenario where multiple levels of preferential treatment on the signal components are needed, by simply incorporating associated equations into the DE.
- **Flexibility**. Generally speaking, previous methods of the sensing matrix design are limited to some specific signal structures. Hence, the analysis and design will not hold accurate for some other signals with special underlying structures. In contrast, our framework can adapt for different signal-structure constraints by simply adjusting degree distributions in DE.

Additionally, as a byproduct of our framework, we give an example of sparse recovery with ℓ_1 regularizers and obtain a closed-form formula for the sensing matrix design. Then, we show that by a proper distribution approximation the classical results, i.e., $m \geq c_0 k \log n$, under the regular sensing matrix design can be reproduced.

II. Problem Description

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

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

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

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

where $f(\mathbf{x})$ is the regularizer used to enforce certain underlying structure for signal $\mathbf{s}$. Our goal is to design a sparse sensing matrix $\mathbf{A}$ which provides preferential treatment for a sub-block of the signal $\mathbf{x}^{\natural}$, for example, this sub-block can be recovered with lower probability of error when comparing with the rest parts of $\mathbf{x}^{\natural}$. Before we proceed, we list our assumptions as

- Measurement system $\mathbf{A}$ is assumed to be sparse. Further, $\mathbf{A}$ is assumed to have entries with with $\mathbb{E}A_{ij} = 0$, and $A_{ij} \in \{0, \pm A^{-1/2}\}$, where an entry $A_{ij} = A^{-1/2}$ $(-A^{1/2})$ implies a positive (negative) relation between the i^{th} sensor and the j^{th} signal component. Having zero as entry implies no relation.
- The regularizer $f(\mathbf{x})$ is assumed to be separable such that $f(\mathbf{x}) = \sum_{i=1}^{n} f(x_i)$.

First we transform (1) to a factor graph [32]. Adopting the viewpoint of Bayesian reasoning, we can reinterpret M-estimator in (2) as the MAP estimator and rewrite it as

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

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

As in [24], 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 $\mathscr{E}$: 1) by placing an edge between the check node of the prior $e^{-f(x_i)}$ and the variable node v_i, and 2) the inclusion of an edge between the variable node v_i and c_j iff A_{ij} is non-zero. Thus, the design of $\mathbf{A}$ transforms to the problem of graph connectivity in $\mathscr{E}$.

III. Sensing Matrix with Regular Sensing

With the aforementioned graphical model, we can view recovering $\mathbf{x}^{\natural}$ as an inference problem, which can be solved via the message-passing algorithm [32]. Adopting the same notations as in [24] as shown in Fig. 1, we denote $m_{i\to a}^{(t)}$ $(\widehat{m}_{a\to i}^{(t)})$ as the message from the variable node v_i to check node c_a (check node c_a to variable node v_i) in the t^{th} round of iteration. Then message-passing algorithm is written as

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

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

where ∂i, ∂a denotes the neighbors connecting with v_i and c_a, respectively, and the symbol $\cong$ refers to equality up to the normalization. In the tth iteration, we recover x_i by maximizing the posterior probability, which reads

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

In the design of matrix $\mathbf{A}$, there are some general desirable properties that we wish to hold (specific requirements will be discussed later): 1) a correct signal reconstruction under the noiseless setting; and 2) minimum number of measurements, or equivalently minimum m. Before proceeding, we first introduce the generating functions $\lambda(\alpha) = \sum_i \lambda_i \alpha^{i-1}$ and $\rho(\alpha) = \sum_i \rho_i \alpha^{i-1}$, which correspond to the distributions of

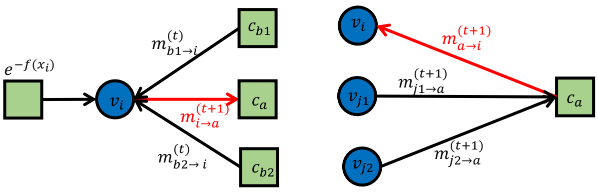

Fig. 1. Illustration of the message-passing algorithm, where the square icon represents the check node while the circle icon represents the variable node.

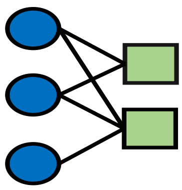

Fig. 2. Illustration of the generating function: $\lambda(\alpha) = \frac{1}{3} + \frac{2\alpha}{3}$ and $\rho(\alpha) = \frac{\alpha}{2} + \frac{\alpha^2}{2}$. The square icon represents check nodes while the circle denotes variable nodes.

degrees 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 functions $\lambda(\alpha)$ and $\rho(\alpha)$ is shown in Fig. 2.

A. Density evolution

We study the reconstruction of $\mathbf{x}^{\natural}$ from $\mathbf{y}$ 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) [32]–[34] that is already proven to be very powerful in analyzing the convergence in sparse graphs resulting from LDPC.

Basically, DE views $m^{(t)}_{i\to a}, \widehat{m}^{(t)}_{a\to i}$ as RVs and tracks the changes of their probability distribution. When the message-passing algorithm converges, we would expect their distributions become more concentrated. However, different from discrete RVs, continuous RVs $m^{(t)}_{i\to a}$ and $\widehat{m}^{(t)}_{a\to i}$ in our case require infinite bits for their precise representation in general, leading to complex formulas for DE. To handle such an issue, we approximate $m^{(t)}_{i\to a}$ and $\widehat{m}^{(t)}_{a\to i}$ as Gaussian RVs, i.e., $m_{i\to a} \sim \mathscr{N}(\mu_{i\to a}, v_{i\to a})$ and $\widehat{m}_{a\to i} \sim \mathscr{N}(\widehat{\mu}_{a\to i}, \widehat{v}_{a\to i})$, respectively. Since the Gaussian distribution is uniquely determined by its mean and variance, we will be able to reduce the DE to finite dimensions. A similar idea has also been used previously in [12], [13], [34].

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

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

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

Then we can show that the DE analysis results in

$$E^{(t+1)} = \mathbb{E}_{\text{prior}(s)}\mathbb{E}_z\Big[h_{\text{mean}}\Big(s + \sum_{i,j}\rho_i\lambda_j z\sqrt{\frac{i}{j}E^{(t)} + \frac{A\sigma^2}{j}}; \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2 + iV^{(t)}}{j}\Big) - s\Big]^2; \quad (5)$$

$$V^{(t+1)} = \mathbb{E}_{\text{prior}(s)}\mathbb{E}_z h_{\text{var}}\Big(s + \sum_{i,j}\rho_i\lambda_j z\sqrt{\frac{i}{j}E^{(t)} + \frac{A\sigma^2}{j}}; \sum_{i,j}\rho_i\lambda_j\frac{A\sigma^2 + iV^{(t)}}{j}\Big), \quad (6)$$

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

$$h_{\text{mean}}(\mu; v) = \lim_{\gamma\to\infty}\frac{\int x_i e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i}{\int e^{-\gamma f(x_i)} e^{-\frac{\gamma(x_i-\mu)^2}{2v}} dx_i}; \quad (7)$$

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

B. Sensing matrix design with regular sensing

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 be 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.

First, we would like a perfect signal reconstruction under the noiseless scenario ($\sigma^2 = 0$), which requires:

- the algorithm converges, i.e., $\lim_{t\to\infty} V^{(t)} = 0$;
- the average error shrinks to zero, i.e., $\lim_{t\to\infty} E^{(t)} = 0$.

Second, we want 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 design criteria as the following optimization problem,

$$\min_{\substack{\boldsymbol{\lambda}\in\Delta_{d_v-1};\\ \boldsymbol{\rho}\in\Delta_{d_c-1}}} \quad \frac{m}{n} = \frac{\sum_{i\ge 2} i\lambda_i}{\sum_{i\ge 2} i\rho_i}, \quad (8)$$

$$\text{s.t.} \quad \lim_{t\to\infty}\left(E^{(t)}, V^{(t)}\right) = (0, 0); \quad (9)$$

$$\lambda_1 = \rho_1 = 0,$$

where Δ_{d-1} denotes the d-dimensional simplex, namely, $\Delta_{d-1} = \left\{\mathbf{z}\in\mathbb{R}^d \mid \sum_i z_i = 1, z_i \ge 0\right\}$, and the constraint in (9) is to avoid one-way message passing.

Generally speaking, we need to run DE numerically to check the requirement (9) 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 (9) to a closed-form equation. As an illustration, we set the prior in (3) to be the Laplacian distribution, i.e., $e^{-\beta|x|}$. In this case, the regularizer $f(\cdot)$ in (2) becomes $\beta\|\cdot\|_1$ and the M-estimator in (2) transforms to Lasso [35].

C. Illustrative example: Laplacian prior

Assuming the signal $\mathbf{x}^\natural$ is k-sparse, i.e., $\|\mathbf{x}^\natural\|_0 \le k$, we would like to recover $\mathbf{x}^\natural$ with the regularizers $\beta\|\cdot\|_1$. Following the approaches in [15], we can show that

$$\begin{aligned} E^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[\text{prox}\left(s+a_1 z\sqrt{E^{(t)}};\beta a_2 V^{(t)}\right)-s\right]^2; \\ V^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathcal{N}(0,1)} \\ &\quad\cdot\left[\beta a_2 V^{(t)}\text{prox}'\left(s+a_1 z\sqrt{E^{(t)}};\beta a_2 V^{(t)}\right)\right], \end{aligned} \tag{10}$$

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

Remark 1. *Compared with SE in [15], our DE tracks both the average variance $V^{(t)}$ and the deviation from mean $E^{(t)}$ while SE only tracks $E^{(t)}$. Assuming $V^{(t)} \propto \sqrt{E^{(t)}}$, our DE equation w.r.t. $E^{(t)}$ in (10) is of similar form of SE.*

Having discussed its relation with SE, we now show that our DE can reproduce the classical results in compressive sensing, namely, $m \ge c_0 k\log(n/k) = O(k\log n)$ (cf. [36]) under the regular sensing matrix design, i.e., when all variable nodes have the same degree d_v and the check nodes have the same degree d_c. Before we proceed, we first approximate the ground-truth distribution with the Laplacian prior. Assuming that the entries of $\mathbf{x}^\natural$ are iid, then each entry is zero with probability $(1-k/n)$ since $\mathbf{x}^\natural \in \mathbb{R}^n$ is k-sparse. 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

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

Simple calculations suggest $\beta = n/(c_0 n\log(n/k))$. Then we conclude the following results

Theorem 2. *Let $\mathbf{x}^\natural$ be a k-sparse signal and assume that β is set to $n/(c_0\log(n/k))$. Then, the necessary conditions for $\lim_{t\to\infty}\left(E^{(t)},V^{(t)}\right)=(0,0)$ in (10) results in $a_1^2 \le n/k$ and $a_2 \le n/\left(c_0 k\log(n/k)\right)$, where a_1 and a_2 are defined as $\sum_{i,j}\rho_i\lambda_j\sqrt{i/j}$ and $\sum_{i,\beta}\rho_i\lambda_j\,(i/j)$, respectively.*

When turning to the regular design, we have $\lambda_{d_v}=1, \rho_{d_c}=1$. Hence, a_1 and a_2 can be written as $\sqrt{n/m}$ and n/m, respectively. Invoking Thm. 2, we can obtain the classical result of the lower bound on the number of measurements $m \ge c_0 k\log(n/k)$.

Having discussed the regular sensing scheme, next we present how to design a compressive sensing matrix that would give different levels of preference in the reconstruction of different parts of the signal $\mathbf{x}^\natural$.

IV. Sensing Matrix With preferential sensing

In certain applications, entries of $\mathbf{x}^\natural$ may have unequal importance from the recovery perspective. One practical application is the case where $\mathbf{x}^\natural$ corresponds to the Fourier coefficients of an image. To better reconstruct the image from its compressive sensing of $\mathbf{x}^\natural$, it is desirable to give a higher priority to the components of $\mathbf{x}^\natural$ that are correspond to low-frequency part in the image than the high-frequency part (cf. the principle of JPEG compression). [1] This section explains as to how we apply our density evolution framework to design the sensing matrix $\mathbf{A}$ such that we can provide unequal preference for different entries of $\mathbf{x}^\natural$. As a result, in recovering the signal, those high priority components will be recovered more accurately than the others.

A. Density evolution

Dividing the entire $\mathbf{x}^\natural$ into the low-priority part $\mathbf{x}_L^\natural$ and high-priority part $\mathbf{x}_H^\natural$, we separately introduce the generating function $\lambda_L(\alpha) = \sum\lambda_i^{(L)}\alpha^{i-1}$ and $\lambda_H(\alpha) = \sum\lambda_i^{(H)}\alpha^{i-1}$ for $\mathbf{x}_L^\natural$ and $\mathbf{x}_H^\natural$, respectively, where $\lambda_i^{(L)}$ ($\lambda_i^{(H)}$) denotes the fraction of nodes with degree i. Similarly, we introduce the generating functions $\rho_L(\alpha) = \sum_i\rho_i^{(L)}\alpha^{i-1}$ and $\rho_H(\alpha) = \sum_i\rho_i^{(H)}\alpha^{i-1}$ for the check nodes connecting the low-priority part $\mathbf{x}_L^\natural$ and high-priority part $\mathbf{x}_H^\natural$, respectively.

Generalizing the analysis of the regular reconstruction, we separately track the average error and variance for $\mathbf{x}_H^\natural$ and $\mathbf{x}_H^\natural$. For the low-priority part $\mathbf{x}_L^\natural$, we define E_L as $\sum_m\sum_{i\in L}\left(\mu_{i\to a}-x_i^\natural\right)^2/(m\cdot|L|)$ and V_L as $\sum_m\sum_{i\in L}v_{i\to a}/(m\cdot|L|)$, where $|L|$ denotes the length of $\mathbf{x}_L^\natural$. Analogously we define E_H and V_H and write the corresponding DE as

$$\begin{aligned} E_L^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[h_{\text{mean}}\left(s+z\cdot b_1^{(t)};b_2^{(t)}\right)-s\right]^2; \\ V_L^{(t+1)} &= \mathbb{E}_{\text{prior}(s)}\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[h_{\text{var}}\left(s+z\cdot b_1^{(t)};b_2^{(t)}\right)\right], \end{aligned} \tag{11}$$

where $b_1(\cdot,\cdot)$ and $b_2(\cdot,\cdot)$ are written as

$$b_1^{(t)} = \sum_{\ell,i,j}\lambda_\ell^{(L)}\rho_i^L\rho_j^{(H)}\sqrt{\frac{A\sigma^2+iE_L^{(t)}+jE_H^{(t)}}{\ell}};$$

$$b_2^{(t)} = \sum_{\ell,i,j}\lambda_\ell^{(L)}\rho_i^L\rho_j^{(H)}\frac{A\sigma^2+iV_L^{(t)}+jV_H^{(t)}}{\ell}.$$

And the definitions of h_{mean} and h_{var} can be found in (7). Switching the index L and H yields the DE w.r.t the pair $\left(E_H^{(t+1)},V_H^{(t+1)}\right)$. Notice we can also put different regularizers $f_L(\cdot)$ and $f_H(\cdot)$ for $\mathbf{x}_L^\natural$ and $\mathbf{x}_H^\natural$. In this case, we need to modify the regularizers $f(\cdot)$ in (7) to $f_L(\cdot)$ and $f_H(\cdot)$, respectively.

Considering the noiseless setting ($\sigma=0$), we require V_L and V_H diminish to zero to ensure the convergence of the message-passing algorithm. However, we do not need the limit of $E_L^{(t)}$ to be zero since we have higher preference for

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

recovering the high-priority part $\mathbf{x}_H^\natural$ rather than the whole signal $\mathbf{x}^\natural$, which leads to the requirement

$$\lim_{t\to\infty}\left(E_H^{(t)}, V_H^{(t)}, V_L^{(t)}\right) = (0,0,0)\,. \quad (12)$$

As an illustration, we revisit the example of ℓ_1 regularizer and show how (12) can be relaxed to be in some closed forms.

Remark 3. *Generally speaking, we need to numerically run the DE update equation in (11) to check the condition (12).*

B. Illustrative example: Laplacian prior

We consider the sparse signal $\mathbf{x}^\natural$ such that the high-priority parts $\mathbf{x}_H^\natural \in \mathbb{R}^{n_H}$ and the low-priority parts $\mathbf{x}_L^\natural \in \mathbb{R}^{n_L}$ are k_H-sparse and k_L-sparse respectively. Additionally, we assume $\frac{k_H}{n_H} \gg \frac{k_L}{n_L}$, which implies that the high-priority part $\mathbf{x}_H^\natural$ contains more information than the low-priority part $\mathbf{x}_L^\natural$. First we require the variance converge to zero, which yields the condition

$$\left[\left(\frac{\beta_H k_H}{n_H}\sum_\ell \frac{\lambda_\ell^{(H)}}{\ell}\right)^2 + \left(\frac{\beta_L k_L}{n_L}\sum_\ell \frac{\lambda_\ell^{(L)}}{\ell}\right)^2\right] \times \left[\left(\sum_i i\rho_i^{(H)}\right)^2 + \left(\sum_i i\rho_i^{(L)}\right)^2\right] \le 1. \quad (13)$$

Then we turn to the behavior of $E_H^{(t)}$. Assuming $E_L^{(t)}$ converges to a fixed non-negative constant $E_L^{(\infty)}$, we would like $E_H^{(t)}$ converge to zero and obtain the condition

$$\frac{k_H}{n_H}\left(\sum_\ell \frac{\lambda_\ell^{(H)}}{\sqrt{\ell}}\right)^2\left[\left(\sum_i \sqrt{i}\rho_i^{(H)}\right)^2 + \left(\sum_i \sqrt{i}\rho_i^{(L)}\right)^2\right] \le 1. \quad (14)$$

Ultimately, the design of the sensing matrix $\mathbf{A}$ reduces to the following optimization problem

$$\min_{\lambda^{(L)},\lambda^{(H)},\rho^{(L)},\rho^{(H)}} \frac{m}{n} = \frac{|L|\sum_i i\lambda_i^{(L)} + |H|\sum_i i\lambda_i^{(H)}}{\sum_i i\left(\rho_i^{(L)} + \rho_i^{(H)}\right)}; \quad (15)$$

$$\text{s.t.}\quad \frac{\sum_i \lambda_i^{(L)} i}{\sum_i \lambda_i^{(H)} i} \times \frac{\sum_i \rho_i^{(H)} i}{\sum_i \rho_i^{(L)} i} = \frac{|H|}{|L|}; \quad (16)$$

$$\text{Condition (13), (14);}$$

$$\lambda_1^{(L)} = \lambda_1^{(H)} = \rho_1^{(L)} = \rho_1^{(H)} = 0.$$

The requirement (16) comes from the degree consistency, namely, $\sum_{i\in H}\mathbb{1}\left((v_i, c_a)\in\mathscr{E}\right) = m\sum_i \rho_i^{(H)} i = |H|\sum_i \lambda_i^{(H)} i$ and $\sum_{i\in L}\mathbb{1}\left((v_i, c_a)\in\mathscr{E}\right) = m\sum_i \rho_i^{(L)} i = |L|\sum_i \lambda_i^{(L)} i$.

V. Numerical Experiments

This section considers the sparse signal and compares the design of preferential sensing with that of regular sensing. Due to the non-convexity nature of (15), we fix the degrees $\{\rho_i^{(H)}\}, \{\rho_i^{(L)}\}$ of the check node to be $\rho_{d_c^H}^{(H)} = 1, \rho_{d_c^L}^{(L)} = 1$, respectively, which means each check node has the same degree. Then we can find the global optimum of $\{\lambda_\ell^{(H)}\}, \{\lambda_\ell^{(L)}\}$ by solving (15), which becomes a convex optimization problem.

Experiment set-up. We fix the check node degrees d_c^H, d_c^L to be 5 and let the maximum variable node degree to be 50. Then we study various setting in which the length n_H of the high-priority part $\mathbf{x}_H^\natural$ to be $\{150, 300\}$ and the corresponding length n_L of the low-priority part $\mathbf{x}_L^\natural$ to be $\{600, 1200\}$. The magnitude of the non-zero entries are fixed to be 1. Then we study the recovery performance with varying $\mathsf{SNR} = \left\|\mathbf{x}^\natural\right\|_2^2/\sigma^2$.

Sensing matrix construction. We design the sensing matrix $\mathbf{A}_{\text{preferential}}$ for preferential sensing via the optimization problem (15). Then we design the sensing matrix $\mathbf{A}_{\text{regular}}$ in (8) which provides regular sensing. (An additional constraint that forces two matrices to have the same number of edges is added in (8) for a fair comparison of performance.) The simulation results are plotted in Fig. 3.

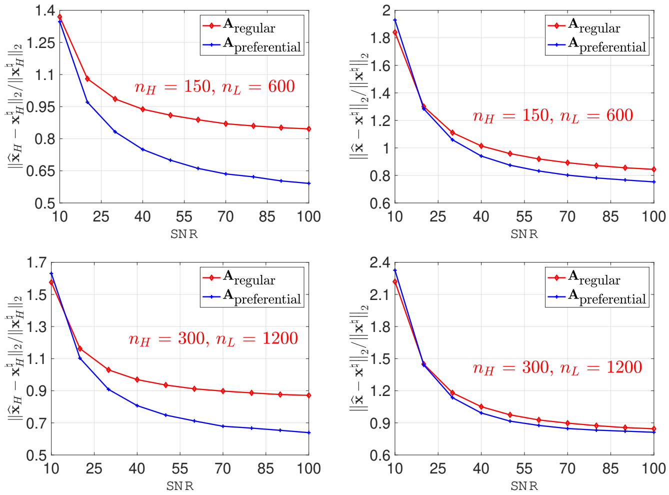

Fig. 3. Comparison of preferential sensing vs regular sensing in terms of $\widehat{\mathbf{x}}_H - \mathbf{x}_H^\natural$ 2. (**Left panel**) We evaluate the reconstruction performance w.r.t the high-priority part $\widehat{\mathbf{x}}_H - \mathbf{x}_H^\natural$ 2/ $\mathbf{x}_H^\natural$ 2. (**Right panel**) We evaluate the reconstruction performance w.r.t the whole signal $\widehat{\mathbf{x}} - \mathbf{x}^\natural$ 2/ $\mathbf{x}^\natural$ 2.

Discussion. Compared with regular sensing, our scheme for preferential sensing reduces the error w.r.t the high-priority part $\mathbf{x}^\natural$ significantly. For example, when $\mathsf{SNR} = 100$ the preferential design reduces the ratio $\|\widehat{\mathbf{x}}_H - \mathbf{x}_H^\natural\|_2/\|\mathbf{x}_H^\natural\|_2$ by approximately 35% compared to the regular reconstruction design while the total loss $\|\widehat{\mathbf{x}} - \mathbf{x}^\natural\|_2$ remains almost the same.

VI. Conclusions

This paper presents a general framework for the sensing matrix design for a linear measurement system. Focusing on a sparse sensing matrix $\mathbf{A}$, we first associated it with a graphical model $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ and transformed the design of $\mathbf{A}$ to the connectivity problem in $\mathscr{G}$. Then we analyzed the impact of the connectivity of the graph on the recovery performance by using the density evolution technique, which is a popular tool in coding theory. Two design strategies are analyzed, namely, regular sensing and preferential sensing. Numerical experiments showed our preferential sensing scheme can greatly reduce the reconstruction error w.r.t the high-priority part with little sacrifice of the total accuracy, which corroborated our theoretical claims.

VII. Acknowledgement

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

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] ——, "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.

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

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] ——, "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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

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

[33] 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.

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

[35] 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.

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