A Density Evolution Framework for Recovery of Covariance and Causal Graphs from Compressed Measurements

Muralikrishnna G. Sethuraman and Hang Zhang[§] and Faramarz Fekri
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta

***Abstract*—In this paper, we propose a general framework for designing sensing matrix $\boldsymbol{A} \in \mathbb{R}^{d\times p}$, for estimation of sparse covariance matrix from compressed measurements of the form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{y}, \boldsymbol{n} \in \mathbb{R}^d$, and $\boldsymbol{x} \in \mathbb{R}^p$. By viewing covariance recovery as inference over factor graphs via message passing algorithm, ideas from coding theory, such as *Density Evolution* (DE), are leveraged to construct a framework for the design of the sensing matrix. The proposed framework can handle both (1) regular sensing, i.e., equal importance is given to all entries of the covariance, and (2) preferential sensing, i.e., higher importance is given to a part of the covariance matrix. Through experiments, we show that the sensing matrix designed via density evolution can match the state-of-the-art for covariance recovery in the regular sensing paradigm and attain improved performance in the preferential sensing regime. Additionally, we study the feasibility of causal graph structure recovery using the estimated covariance matrix obtained from the compressed measurements.**

I. Introduction

In this work, we study the feasibility of recovering the covariance matrix and the underlying causal structure of unknown set of variables $\boldsymbol{x} = (x_1, \ldots, x_p)$, by collecting observations through a linear measurement system of the form,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^d$ is of a lower dimension than $\boldsymbol{x} \in \mathbb{R}^p$. The causal semantics of $\boldsymbol{x}$ can be represented using a directed graph where the edges encode the dependencies between the variables. The problem of recovering the causal structure is then equivalent to the graph structure recovery, in other words, recovery of the edge set E.

Graph structure recovery has been a problem of interest in the last few decades within the machine learning community. It is well known that structure recovery is an NP-hard problem, [1], and in general it cannot be uniquely identified [2]. Nevertheless, attempts have been made to recover the structure of the graphical model under various assumptions on the underlying probability distribution governing the system. Additive Noise Models (ANM) have gained a lot of traction in recent years due to their analytic simplicity. Moreover, it has been shown that under some assumptions on the noise distribution, the graph structure can be uniquely identified. Of particular interest is the setting where the noise distribution is Gaussian, for which there exist polynomial time algorithms for structure recovery [3].

However, the aforementioned solutions assume direct access to the observational data which may not be practical in certain applications [4], making it an expensive task to recover the structure of the underlying graph, especially in high dimensions. Our work differs from the existing methods by considering the scenario where the graph structure is recovered from compressed measurements $\boldsymbol{y}$ instead of directly observing $\boldsymbol{x}$. The crux of our approach relies on density evolution analysis of the message-passing algorithm, also known as Belief propagation, min-product, or max-sum. The algorithm was independently developed in different fields in the last century. In 1935, Bethe [5] used it to approximately compute the partition function. [6] developed belief propagation in 1988 to perform exact inference in Acyclic Bayesian Networks.

Related Work. Sparse vector recovery from compressed measurements has been studied quite extensively with several sensing matrices being proposed in the literature [7]–[9]. Over time, Gaussian sensing matrices have become a popular choice for sparse vector recovery. However, [9] showed that the gaussian sensing matrix is not a very good candidate for the recovery of sparse matrices. [10] showed the use of binary matrices, in particular, adjacency of δ-left regular bipartite graph for sparse vector recovery. [11] built upon the work done by [10] and proved that adjacency of δ-left regular bipartite graphs can be used for recovery of sparse matrices.

In the 1960s, [12] proposed a sum-product algorithm to decode *low-density parity check* (LDPC) codes over graphs, which was forgotten for decades and later reinvented along with density evolution to design LDPC codes achieving channel capacity. [13]–[15] analyzed sparse sensing matrices based on spatial coupling using DE for sparse vector recovery. [16] employed density evolution and developed a framework for designing sensing matrices for regular as well as preferential recovery of sparse vectors. For a better understanding of the usage of message passing and density evolution for signal recovery, we refer the readers to [5], [15].

Graph Structure Recovery. Structure recovery methods for directed graphs can broadly be divided into two categories: (1) Independence test-based, and (2) score-based methods.

[§]work performed while at Georgia Tech, now currently at Amazon.

Independence test-based methods [17] typically involves computing the conditional independence between any two nodes in the graph conditioned on all the subsets of the remaining nodes. Whereas, score-based methods rely on a metric to score a candiadte *directed acyclic graph* (DAG) based on how well it explains the data, we refer the reader to [2] for a survey on the graph structure recovery methods.

Contributions This work is the first application of density evolution and message-passing algorithms to design sensing matrices for sparse covariance and graph structure recovery with linear structural equations. We summarize our contributions as follows:

1) We propose a novel approach to optimally design a low dimensional data collection (measurement) scheme from a high dimensional signal that would allow for recovering a sparse covariance matrix from these measurements. We use density evolution-based analysis of the message-passing algorithm to reduce the design procedure into a convex program.
2) We propose two separate design schemes: (i) (*Regular sensing*) equal preference over all the entries of the covariance matrix, and (ii) (*Preferential sensing*) preferential treatment over certain entries of the covariance matrix. We also showcase the feasibility of causal graph recovery from the estimated covariance matrix.

II. Problem Description

In this section, we provide a formal description of our problem statement starting with the notations. All vectors are denoted by lowercase boldface letters, $\boldsymbol{x}$, and matrices by uppercase boldfaced letters, $\boldsymbol{A}$. $\boldsymbol{A}_{*,i}$ denotes the i-th column of the matrix $\boldsymbol{A}$, similarly $\boldsymbol{A}_{j,*}$ denotes the j-th row of $\boldsymbol{A}$. $\|\boldsymbol{A}\|_1 = \sum_{ij} |A_{ij}|$ and $\|\boldsymbol{A}\|_F = \sqrt{\sum_{ij} A_{ij}^2}$ and $\|\boldsymbol{x}\|_p = \left(\sum_i x_i^p\right)^{1/p}$. $\mathrm{v}(\boldsymbol{A})$ denotes the vectorized form of matrix $\boldsymbol{A}$.

Consider a linear measurement system of the form,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{y} \in R^d$ denotes the observations, $\boldsymbol{x} \in R^p$ denotes the unknown vector, $\boldsymbol{A} \in R^{d\times p}$ denotes sensing matrix, and $\boldsymbol{n} \in R^d$ denotes the measurement noise. For the case when $d < p$, we are interested in the problem of recovering the covariance of $\boldsymbol{x}$ from the observations $\boldsymbol{y}$. Our goal is to design a sparse sensing matrix $\boldsymbol{A}$ that is capable of recovering the covariance from compressed measurements and at the same time being able to provide selective preference to a sub-block of the covariance matrix. That is, we would like a sub-block of the covariance matrix to be recovered with a lower probability of error than the rest of the covariance.

A. Covariance Recovery

Under the linear measurement system discussed previously, when the measurement noise is zero, the covariance of the observations $\boldsymbol{y}$ is given by $\boldsymbol{\Sigma}_Y = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T$. We further make the assumption that the covariance of X is a sparse matrix. The covariance recovery can now be posed as the following convex program,

$$\min_{\boldsymbol{\Sigma}} \quad \|\boldsymbol{\Sigma}\|_1 \quad \text{subject to} \quad \boldsymbol{\Sigma}_Y = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T. \tag{P$_1$}$$

Since we only have access to the observed samples of $\boldsymbol{y}$, the true covariance is approximated by the sample covariance, $\boldsymbol{\Sigma}_Y^{(N)} = (1/N)\sum_i \boldsymbol{y}_i\boldsymbol{y}_i^T$. Additionally, upon vectorization, we have $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T = (\boldsymbol{A}\otimes\boldsymbol{A})\mathrm{v}(\boldsymbol{\Sigma})$, where $\otimes$ denotes the Kronecker product. This gives the following equivalent formulation of (P$_1$),

$$\min_{\boldsymbol{\Sigma}} \quad \|\mathrm{v}(\boldsymbol{\Sigma})\|_1 \text{ s.t. } \left\|\mathrm{v}(\boldsymbol{\Sigma}_Y^{(N)}) - (\boldsymbol{A}\otimes\boldsymbol{A})\mathrm{v}(\boldsymbol{\Sigma})\right\|_2^2 \le \kappa. \tag{P$_2$}$$

In this vectorized form, $(\boldsymbol{A}\otimes\boldsymbol{A})$ can be thought of as the new sensing matrix having a Kronecker product structure and $\mathrm{v}(\boldsymbol{\Sigma})$ to be the sparse vector that has to be recovered.

III. Sensing Matrix for Regular Recovery

In this section, we describe the design scheme for the sensing matrix via the *density evolution* methodology. We first focus on the regular sensing setting where all the entries of the covariance matrix are treated equally. For ease of notation let us denote $\boldsymbol{\gamma} = \mathrm{v}(\boldsymbol{\Sigma}_Y)$, $\boldsymbol{\chi} = \mathrm{v}(\boldsymbol{\Sigma})$, and $\boldsymbol{A}^{\otimes} = \boldsymbol{A}\otimes\boldsymbol{A}$. The solution to (P$_2$) can be viewed as the solution to the following *maximum a posteriori* (MAP) estimator

$$\hat{\boldsymbol{\chi}} = \arg\max_{\boldsymbol{\chi}} \exp\left(-\frac{\|\boldsymbol{\gamma} - \boldsymbol{A}^{\otimes}\boldsymbol{\chi}\|_2^2}{2\sigma^2}\right)\exp\big(-f(\boldsymbol{\chi})\big), \tag{3}$$

where $f(\boldsymbol{\chi})$ is the generalized regularizer term. When $f(\boldsymbol{\chi})$ is set to $\|\boldsymbol{\chi}\|_1$ then the MAP estimator is exactly equivalent to (P$_2$). Here, we make a few assumptions on the sensing matrix and the regularizer: (i) The sensing matrix $\boldsymbol{A}$ is sparse with $EA_{ij} = 0$ and $A_{ij} \in \{0, \pm A^{-1/2}\}$, and (ii) The regularizer $f(\boldsymbol{\chi})$ can be decomposed, $f(\boldsymbol{\chi}) = \sum_i f(\chi_i)$.

To develop the design framework, we associate (3) with a factor graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of nodes corresponding to components of $\boldsymbol{\chi}$ (known as *variable nodes*) and components of $\boldsymbol{\gamma}$ (known as *check nodes*). The readers are referred to the appendix for an illustration of the factor graph and some of its structural properties. An edge exists between χ_i and γ_j if $A_{ij}^{\otimes} \neq 0$.

In view of the factor graph, the recovery of $\boldsymbol{\Sigma}_X$ can be thought of as an inference problem over $\mathcal{G}$ which can be solved using the message-passing algorithm. Following the notations of [16], let $m_{i\to a}^{(t)}$ denote the message going from the i-th variable node to the a-th check node at the t-th iteration. Similarly, let $\hat{m}_{a\to i}^{(t)}$ denote the message going from a-th check node to the i-th variable node at the t-th iteration. The message-passing algorithm is then given by

$$m_{i\to a}^{(t+1)}(\chi_i) \cong e^{-f(\chi_i)} \prod_{b\in\partial i\setminus a} \hat{m}_{b\to i}^{(t)}(\chi_i); \tag{4}$$

$$\hat{m}_{a\to i}^{(t+1)}(\chi_i) \cong \int \prod_{j\in\partial a\setminus i} m_{j\to a}^{(t+1)}(\chi_j) e^{-\frac{(\gamma_a - \Sigma_j A_{aj}\chi_j)^2}{2\sigma^2}} d\chi_j, \tag{5}$$

where ∂a, ∂i denote the neighborhood of the a-th check node and the i-th variable node respectively and $\cong$ denotes equality up to a normalization constant. At iteration t, χ_i can be recovered by taking argmax of the product of all the messages coming to the i-th variable node.

We define $\lambda(\alpha)$ and $\rho(\alpha)$ to be the distribution of the number of non-zero entries in the columns and rows of $\boldsymbol{A}$, which would later be used for designing the sensing matrix. The degree distribution of the check nodes and the variable nodes can then be obtained from $\lambda(\alpha)$ and $\rho(\alpha)$, refer to the appendix for more details.

A. Density Evolution

In order to design the sensing matrix we analyze the convergence of the message-passing algorithm on the factor graph. A sensing matrix is "good" if all the messages in the factor graph converge to the right value (the covariance of $\boldsymbol{x}$). To that end, the messages are treated as random variables, and in particular, they are chosen to be Gaussian distributed due to their simplicity. That is, $m_{i\to a}^{(t)} \sim \mathcal{N}(\mu_{i\to a}^{(t)}, v_{i\to a}^{(t)})$ and $\hat{m}_{a\to i}^{(t)} \sim \mathcal{N}(\hat{\mu}_{a\to i}^{(t)}, \hat{v}_{a\to i}^{(t)})$. The convergence of (3) is analyzed by tracking the following two quantities

$$E^{(t)} = \frac{1}{d^2p^2}\sum_{a=1}^{d^2}\sum_{i=1}^{p^2}\Big(\mu_{i\to a}^{(t)} - \chi_i\Big)^2 \tag{6}$$

$$V^{(t)} = \frac{1}{d^2p^2}\sum_{a=1}^{d^2}\sum_{i=1}^{p^2} v_{i\to a}^{(t)}. \tag{7}$$

Where $E^{(t)}$ and $V^{(t)}$ represent the average error and variance of all the messages in the factor graph at iteration t. To enforce sparsity, the regularization function $f(\boldsymbol{\chi})$ is set as $\beta\|\boldsymbol{\chi}\|_1$, this is equivalent to enforcing Laplacian prior on $\boldsymbol{\chi}$. Upon analyzing the density of each message in the factor graph over time(density evolution), (6) and (7) can be reduced to the following form,

$$E^{(t+1)} = \mathbf{E}_{\text{prior}(s)}\mathbf{E}_{z\sim\mathcal{N}(0,1)}\Bigg[\text{prox}\Big(s + a_1 z\sqrt{E^{(t)}}; \beta a_2 V^{(t)}\Big) - s\Bigg]^2 \tag{8}$$

$$V^{(t+1)} = \mathbf{E}_{\text{prior}(s)}\mathbf{E}_{z\sim\mathcal{N}(0,1)}\Bigg[\beta a_2 V^{(t)}\text{prox}'\Big(s + a_1 z\sqrt{E^{(t)}}; \beta a_2 V^{(t)}\Big)\Bigg], \tag{9}$$

where

$$a_1 = \sum_{i,i',j,j'} f^r_{i,i',j,j'}\sqrt{ii'/jj'};$$
$$a_2 = \sum_{i,i',j,j'} f^r_{i,i',j,j'}(ii'/jj'),$$

where $f^r_{i,i',j,j'} = \rho_i\rho_{i'}\lambda_j\lambda_{j'}$. $\text{prox}(a;b)$ denotes the soft-threshold function, and $\text{prox}'(a;b)$ is the derivative of the soft-threshold function with respect to the first argument. For a detailed derivation of these quantities please refer to the supplementary notes.

In designing a "good" sensing matrix we would like to minimize the number of measurements needed to recover $\boldsymbol{\Sigma}$. Additionally, we need the message-passing algorithm to converge, i.e., $V^{(t)} \to 0$ and the average error should shrink to zero, $E^{(t)} \to 0$ as $t \to \infty$. However, enforcing $\lim_{t\to\infty}(E^{(t)}, V^{(t)}) = (0,0)$ is not straightforward as it requires running the DE updates numerically until convergence is achieved. For the case of sparse vector recovery, [16] showed that these requirements can be reduced to two inequality constraints making it easier to check for satisfiability. We extend this to the case of covariance recovery in the form of the following theorem.

Theorem 1: Let $\boldsymbol{\Sigma}$ be k^2-sparse and for $c_0 > 0$, set $\beta = \frac{p^2}{c_0\log(p/k)}$. Then, a sufficient condition for $\lim_{t\to\infty}(E^{(t)}, V^{(t)}) = (0,0)$ results in

$$a_1^2 \le \frac{p^2}{k^2}; \quad a_2 \le \frac{p^2}{2c_0k^2\log(p/k)}.$$

Therefore the design of the sensing matrix can be posed as the following optimization problem,

$$\min_{\substack{\boldsymbol{\lambda}\in\Delta_{d_v};\\ \boldsymbol{\rho}\in\Delta_{d_c}}} \quad \frac{d}{p} = \frac{\sum_{i\ge2} i\lambda_i}{\sum_{j\ge2} j\rho_j} \tag{10}$$

$$\text{s.t} \quad a_1^2 \le \frac{p^2}{k^2} \tag{11}$$

$$a_2 \le \frac{p^2}{2c_0k^2\log(p/k)} \tag{12}$$

$$\lambda_1 = \rho_1 = 0, \tag{13}$$

where Δ_d is a d-dimensional simplex, d_v and d_c denote the maximum column and row degree respectively of sensing matrix $\boldsymbol{A}$. The final constraint (13) is added to avoid one-way message passing.

The above optimization problem tries to find the degree distributions $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ such that d/p (y-dim/x-dim) is minimum while guaranteeing that the message passing algorithm converges in the form of constraints (11) and (12). Hence, ensuring that the inference over the factor graph yields the true covariance of $\boldsymbol{x}$.

Once we have the distributions $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ from solving the above optimization program, we then sample the sensing matrix such that the number of non-zero entries in the rows and columns satisfies the obtained distributions. For every non-zero entry of $\boldsymbol{A}$, $P(A_{ij} = A^{-1/2}) = P(A_{ij} = -A^{-1/2}) = \frac{1}{2}$. With the sensing matrix obtained, (P_2) can be solved using any convex program solver.

IV. Sensing Matrix for Preferential Recovery

In this section, we extend the density evolution based sensing matrix design to the case of preferential recovery of the covariance matrix. That is, we employ the DE framework to construct sensing matrices that provide higher importance to

a sub-block of the covariance matrix. In other words, we treat certain variables as important and try to recover the covariance between the important variables with higher accuracy.

A. Density Evolution

The unknown signal $\boldsymbol{x}$ is divided into two parts $\boldsymbol{x}_H \in \mathbb{R}^{n_H}$ (high priority), and $\boldsymbol{x}_L \in \mathbb{R}^{n_L}$ (low priority), without loss of generality we assume that $\boldsymbol{x} = (\boldsymbol{x}_H, \boldsymbol{x}_L)$. This splits the covariance into four sub-matrices,

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} \boldsymbol{\Sigma}_{HH} & \boldsymbol{\Sigma}_{HL} \\ \boldsymbol{\Sigma}_{LH} & \boldsymbol{\Sigma}_{LL} \end{bmatrix}. \tag{14}$$

In this case, we would like to place higher importance on $\boldsymbol{\Sigma}_{HH}$ and design the sensing matrix in order to recover the higher priority sub-block with higher accuracy than the other components. To that end, we introduce the degree distributions $\lambda_H(\alpha)$ and $\lambda_L(\alpha)$ corresponding to the first n_H columns and the last n_L columns of the sensing matrix respectively. Similarly, $\rho_H(\alpha)$ and $\rho_L(\alpha)$ correspond to the degree distribution of the first n_H rows and the last n_L rows of the sensing matrix.

Generalizing the analysis for regular sensing, the average error and the variance for each sub-matrix of $\boldsymbol{\Sigma}_X$ are separately tracked. For $\boldsymbol{\Sigma}_{HH}$ sub-block, we define

$$E_{HH} = \frac{1}{d^2 \cdot n_H^2} \sum_a \sum_{i \in HH} (\mu_{i \to a} - \chi_i)^2 \tag{15}$$

$$V_{HH} = \frac{1}{d^2 \cdot n_H^2} \sum_a \sum_{i \in HH} v_{t \to a} \tag{16}$$

The average error and variance for LH, HL, and LL are defined in a similar manner. Similar to regular sensing by assuming a Laplacian prior on $\boldsymbol{\chi}$ we then have

$$E_{HH}^{(t+1)} = \mathbf{E}_{\text{prior}(s)} \mathbf{E}_{z \sim \mathcal{N}(0,1)} \Bigg[\text{prox}\Big(s + z b_{HH,1}^{(t)}; b_{HH,2}^{(t)} \Big) - s \Bigg]^2 \tag{17}$$

$$V_{HH}^{(t+1)} = \mathbf{E}_{\text{prior}(s)} \mathbf{E}_{z \sim \mathcal{N}(0,1)} \Bigg[b_{HH,2}^{(t)} \text{prox}'\Big(s + z b_{HH,1}^{(t)}; b_{HH,2}^{(t)} \Big) \Bigg], \tag{18}$$

where $b_{HH,1}^{(t)}$ and $b_{HH,2}^{(t)}$ are defined as follows

$$b_{HH,1}^{(t)} = \sum_{\ell\ell', ii', jj', kk'} f^p_{\ell\ell', ii', jj', kk'} \sqrt{\mathcal{F}}; \tag{19}$$

$$b_{HH,2}^{(t)} = \sum_{\ell\ell', ii', jj', kk'} f^p_{\ell\ell', ii', jj', kk'} \mathcal{F}. \tag{20}$$

$$\mathcal{F} = \frac{A\sigma^2 + ii' V_{HH}^{(t)} + jj' V_{HL}^{(t)} + kk' V_{LL}^{(t)}}{\ell\ell'}. \tag{21}$$

Where $f^p_{\ell\ell', ii', jj', kk'} = \lambda_{H,\ell} \lambda_{H,\ell'} \rho_{H,i} \rho_{H,i'} \rho_{H,j} \rho_{L,j'} \rho_{L,k} \rho_{L,k'}$. In this setting, in order to design a "good" sensing matrix, we would like to ensure that the inference over the factor graph converges to the true covariance for the high-priority portion of the covariance matrix. To that end, the sensing matrix must satisfy the following set of constraints.

R1. We require consistency with respect to the number of non-zero entries in the sensing matrix. Starting with the high-priority part, the number of non-zero entries in the first n_H columns is given by $n_H(\sum_i \lambda_{H,i})$ (counting the non-zeros by column) and $d(\sum_i i\rho_{H,i})$ (counting by rows). Therefore we have the following constraint

$$n_H \Big(\sum_i \lambda_{H,i} \Big) = d \Big(\sum_i \rho_{H,i} \Big).$$

Similarly, the consistency requirement on the low-priority part would yield $n_L(\sum_i \lambda_{L,i}) = d(\sum_i i\rho_{H,i})$.

R2. We require the variances to converge to zero. That is,

$$\lim_{t \to \infty} \Big(V_{HH}^{(t)}, V_{HL}^{(t)}, V_{LL}^{(t)} \Big) = (0, 0, 0)$$

This implies that the message-passing algorithm on the factor graph converges. Here we exclude V_{LH} due to the symmetric nature of the covariance matrix.

R3. Due to the preferential nature of the design we require that the error in the high-priority part of the covariance to be lower than the other sub-matrices. In other words, let $\delta_{E,HH}^{(t)} = E_{HH}^{(t+1)} - E_{HH}^{(t)}$, and we similarly define $\delta_{E,HL}^{(t)}$ and $\delta_{E,LL}^{(t)}$, we want $|\delta_{E,HH}^{(t)}| \leq |\delta_{E,HL}^{(t)}|$ and $|\delta_{E,HH}^{(t)}| \leq |\delta_{E,LL}^{(t)}|$ for all $t \geq T_0$ for some T_0.

Hence, the design of the sensing matrix can be posed as the following convex problem where we minimize the ratio of y-dim over x-dim similar to the regular sensing case.

$$\min_{\substack{\lambda_H \in \Delta_{d_{v_H}}; \\ \lambda_L \in \Delta_{d_{v_L}}; \\ \rho_H \in \Delta_{d_{c_H}}; \\ \rho_L \in \Delta_{d_{c_L}}}} \quad \frac{d}{p} = \frac{n_L \sum_i i\lambda_{L,i} + n_H \sum_i i\lambda_{H,i}}{\sum_j j(\rho_{H,j} + \rho_{L,j})} \tag{22}$$

$$\text{s.t} \quad \frac{\sum_i i\lambda_{L,i}}{\sum_i i\lambda_{H,i}} \times \frac{\sum_i i\rho_{H,i}}{\sum_i i\rho_{L,i}} = \frac{n_H}{n_L}; \tag{23}$$

$$\text{Requirement 2 \& 3;} \tag{24}$$

$$\lambda_{H,1} = \lambda_{L,1} = \rho_{H,1} = \rho_{L,1} = 0, \tag{25}$$

B. Constraint Relaxation for Laplacian Prior

Consider a sparse covariance matrix where the high-priority subpart is k_{HH}-sparse and the low-priority subpart is k_{LL}-sparse, with the added assumption that $k_{HH}/n_h \gg k_{LL}/n_L$. As stated in section III-A, directly enforcing requirements 2 and 3 in equation (24) is not straightforward. Fortunately, by assuming the prior to be Laplacian, requirements 2 and 3 can be relaxed to obtain the following inequalities constraints that are convex in the degree polynomials

$$\left\{ \sqrt{\frac{\beta_{HH} k_{HH}}{n_{HH}}} \sum_\ell \frac{\lambda_{H,\ell}}{\ell} + \sqrt{\frac{\beta_{LL} k_{LL}}{n_{LL}}} \sum_\ell \frac{\lambda_{L,\ell}}{\ell} \right\} \times \left[\Big(\sum_i i\rho_{H,i} \Big)^2 + \Big(\sum_i i\rho_{L,i} \Big)^2 \right] \leq 1. \tag{26}$$

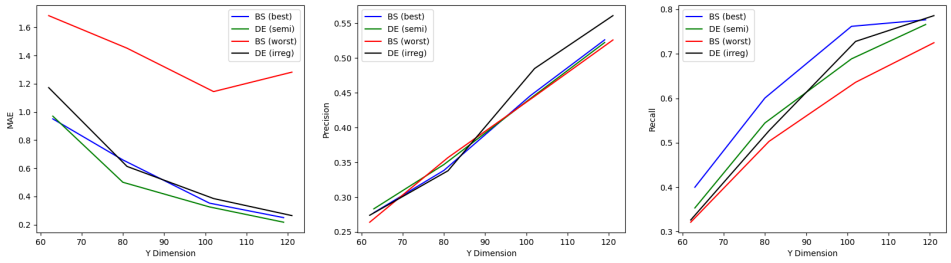

Fig. 1. Performance comparison on regular sensing of covariance matrix

$$\sqrt{\frac{k_{HH}}{n_{HH}}}\left(\sum_{\ell}\frac{\lambda_{H,\ell}}{\sqrt{\ell}}\right) \leq \sqrt{\frac{k_{HL}}{n_{HL}}}\left(\sum_{\ell}\frac{\lambda_{L,\ell}}{\sqrt{\ell}}\right); \tag{27}$$

$$\left(\frac{k_{HH}}{n_{HH}}\right)^{1/4}\left(\sum_{\ell}\frac{\lambda_{H,\ell}}{\sqrt{\ell}}\right) \leq \left(\frac{k_{LL}}{n_{LL}}\right)^{1/4}\left(\sum_{\ell}\frac{\lambda_{L,\ell}}{\sqrt{\ell}}\right). \tag{28}$$

Here, equation (26) corresponds to requirement 2 and equation (28) corresponds to requirement 3. The above inequalities are convex with respect to the degree polynomials and hence can be solved using any convex program solver. The key idea behind the relaxation is to approximate $\delta^{(t)}_{E,HH}, \delta^{(t)}_{V,HH}$, $\delta^{(t)}_{E,HL}, \delta^{(t)}_{V,HL}$ and $\delta^{(t)}_{E,LL}, \delta^{(t)}_{V,LL}$ by its first-order Taylor series approximation and enforcing the operator norm of the Jacobian to be less than one, readers are referred to the supplementary notes for more details.

V. Graph Structure Recovery

In this section, we discuss an application of compressed covariance recovery for causal graph recovery. To that end, the unknown signal is modeled using a structural equation model (SEM) [18], [19] given by,

$$x_i = \boldsymbol{W}^T_{*,i}\boldsymbol{x} + z_i, \quad \forall i = 1, \ldots, p \tag{29}$$

where $\boldsymbol{W}$ denotes the weighted adjacency matrix and z_i corresponds to intrinsic noise in the system. We would also like to learn the weighted adjacency matrix from the compressed measurements $\boldsymbol{y}$ using the recovered covariance of $\boldsymbol{x}$. An equivalent representation of the above SEM is to consider a directed (causal) graph $G = (V, E)$, where $V = \{x_1, \ldots, x_p\}$ with $\boldsymbol{W}$ being its adjacency matrix, i.e., W_{ij} is the weight corresponding to the edge $(x_i, x_j) \in E$. For a given x_j we define *parent set* of x_i, denoted by $Pa(x_i)$, as the set of nodes x_i for which $W_{ij} \neq 0$. This representation allows for a more straightforward causal semantics for the underlying interactions between the variables in the system. From the covariance matrix estimated using the sensing matrix designed via the methodology described in sections III and IV, the algorithm developed by [3] is utilized for recovering the structure of the underlying causal graph.

VI. Experiments

In this section, we present the numerical experiments performed to evaluate covariance and graph recovery. To generate the GBN, we sampled directed graphs from Erdös-Rényi class of random graphs with edge weights set to $\pm 1/2$ with probability $1/2$. We first study the effectiveness of the sensing system for recovery of the entire covariance recovery matrix followed by preferential recovery of the high-priority portion of the covariance matrix. We compare the performance with the current state-of-the-art [11], where the sensing matrix is the adjacency matrix of δ-left-regular bipartite graph. We then evaluate the performance of the sensing system for graph structure recovery.

A. Covariance Recovery

1) Regular Sensing: we consider two different design schemes for constructing the sensing matrix. (i) *Fixed row degree and variable column degree*. In this case, $\rho_i = 1$ when $i = d_c$ and 0 otherwise. We then solve (10) for $\boldsymbol{\lambda}$, and (ii) *Variable row and column degree*. In this case we solve (10) for both $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. In case (i), the resulting optimization program is readily solvable by any convex program solver. For case (ii), we first keep $\boldsymbol{\lambda}$ constant and solve for $\boldsymbol{\rho}$, then using the obtained solution for $\boldsymbol{\rho}$ we solve for $\boldsymbol{\lambda}$.

The recovery performance is evaluated using three metrics, namely, (1) *Maximum Absolute Error* (MAE), (2) *Precision*, and (3) *Recall* of the structure of the covariance.

The two design schemes attain similar performance with respect to all the metrics, as seen in Figure 1. We can also observe that the density evolution based sensing matrices achieve similar performance to that of [11] when δ is tuned. On the other hand, improper assignment of δ results in poor performance compared to the density evolution based design.

2) Preferential Sensing: For the case of preferential sensing, we considered graphs with $p = 200$ nodes, where we choose covariance between the first $n_H = 50$ nodes to be of higher priority. The measurements are then compressed down to $d = 60$ dimensions. The performance of the preferential sensing matrix is compared with that of [11] with respect to the same metrics described in the previous section on the high-priority sub-matrix of the covariance. In this case, we fix the degree distribution of the check nodes and solve for the degree distribution of the variable using the procedure described in section IV-A. As seen from Figure 2, the proposed preferential sensing matrix outperforms the baseline with respect to all the error metrics.

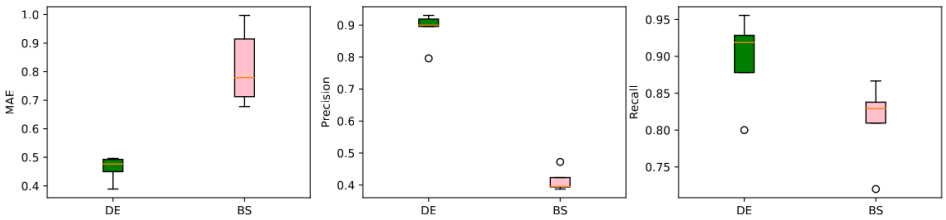

Fig. 2. Performance comparison on preferential sensing of covariance matrix.

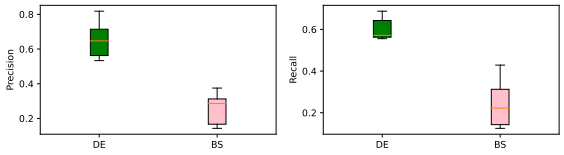

Fig. 3. Performance comparison on graph recovery with $p = 200$ nodes and $n_H = 50$ high priority nodes.

B. Graph Structure Recovery

Using the procedure described in section V, the graph structure was recovered from the covariance matrix. The performance is evaluated using precision and recall as metrics. For preferential recovery, we only consider the edges connecting the high-priority nodes for evaluating the performance. The proposed regular sensing matrix achieves similar performance to that of the baseline, like in the case of covariance recovery and hence we refer the readers to the appendix for details. Figure 3 shows the performance comparison between the proposed preferential sensing matrix and the baseline. As seen from the figure, we see a similar trend to that of covariance recovery, i.e., the preferential sensing system outperforms the baseline with respect to all the metrics.

VII. Conclusion

In this paper, we presented a general framework for collecting lower dimensional samples of the signal generated from a GBN for accurate recovery of the covariance and graph structure under (i) regular and (ii) preferential sensing regimes. We also showcased the feasibility of our approach through numerical simulations. There are several directions that could be of interest in the future. While we restricted our focus to GBNs, exploring other types of additive noise distributions would be an interesting avenue. The supplementary section containing

VIII. Acknowledgments

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

References

[1] M. Chickering, D. Heckerman, and C. Meek, "Large-sample learning of bayesian networks is np-hard," *Journal of Machine Learning Research*, vol. 5, 2004.

[2] R. Guo, L. Cheng, J. Li, P. R. Hahn, and H. Liu, "A survey of learning causality with data: Problems and methods," *ACM Computing Surveys (CSUR)*, vol. 53, no. 4, pp. 1–37, 2020.

[3] A. Ghoshal and J. Honorio, "Learning identifiable gaussian bayesian networks in polynomial time and sample complexity," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/907edb0aa6986220dbffb79a788596ee-Paper.pdf

[4] J. Müller, C. Kuttler, and B. A. Hense, "Sensitivity of the quorum sensing system is achieved by low pass filtering," *Biosystems*, vol. 92, no. 1, pp. 76–81, 2008.

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

[6] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[7] E. Candes, 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.

[8] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of complexity*, vol. 23, no. 4-6, pp. 918–925, 2007.

[9] A. Kaplan, V. Pohl, and D. G. Lee, "On compressive sensing of sparse covariance matrices using deterministic sensing matrices," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4019–4023.

[10] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 798–805.

[11] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. D. Nowak, "Sketching sparse matrices, covariances, and graphs via tensor products," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1373–1388, 2015.

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

[13] 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.

[14] ——, "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.

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

[16] H. Zhang, A. Abdi, and F. Fekri, "a general compressive sensing construct using density evolution," *IEEE Transactions on Signal Processing*, pp. 1–16, 2022.

[17] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, prediction, and search*. MIT press, 2000.

[18] K. A. Bollen, *Structural equations with latent variables*. John Wiley & Sons, 1989, vol. 210.

[19] J. Pearl, *Causality*, 2nd ed. Cambridge University Press, 2009.

The appendix is organized as follows: In Appendix A we discuss some of the structural properties of the factor graph that arise from the Knronecker product structure of the sensing matrix. In Appendix B, we derive the degree distribution of the check nodes and the variable nodes in the factor in terms of the $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. Details regarding the derivation of the density update equations can be found in the supplementary notes, https://muralikgs.github.io/assets/pdf/supplementary_notes_allerton.pdf

Appendix A
Factor Graph

To develop the density evolution framework, we associate (3) with a factor graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of nodes corresponding to components of $\boldsymbol{\chi}$ (variable nodes) and components of $\boldsymbol{\gamma}$ (check nodes). The readers are refered to the appendix for an illustration of the factor graph and some of the structural properties. An edge exists between χ_i and γ_j if $A^{\otimes}_{ij} \neq 0$.

At this point it is important to illustrate some of the key structural properties of the factor graph that arises due to the Kronecker product nature of $\boldsymbol{A}^{\otimes}$, see Figures 4 and 5. The check nodes and the variable nodes consist of d and p blocks respectively, and each check node block contains d nodes and each variable node block contains p nodes. i-th check node block is considered to be connected to j-th variable node block if any node in the i-th check node block is connected to any node in the j-th variable node block. This is true when $A_{ij} \neq 0$. The connection between the nodes in the i-th check node block and j-th variable node block, if it exists, is determined by $\boldsymbol{A}$. That is, within in the blocks, the k-th check node is connected to l-th variable node if $A_{kl} \neq 0$. Figure 5 shows the factor graph for the following sensing matrix,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{30}$$

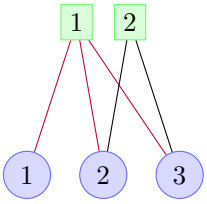

Fig. 4. Illustration of the connections in factor graph corresponding to the Kronecker product when the sensing matrix is given by equation (30). (a) shows the connections at the block level, the number within the node corresponds to the block ID and as seen in (a), the connections at the block level are governed by A.

Appendix B
Degree Distribution of Check nodes and Variable nodes

As described in section 3.1, let $\boldsymbol{\lambda} \in \Delta_{d_v}$, $\boldsymbol{\rho} \in \Delta_{d_c}$ be the degree distributions of columns and rows of $\boldsymbol{A}$. We can divide $\boldsymbol{\gamma}$ and $\boldsymbol{\chi}$ into blocks of size d and p nodes respectively. Each block corresponds to a column of $\boldsymbol{\Sigma}_Y$ and $\boldsymbol{\Sigma}$. Let γ_i denote the i-th block of $\boldsymbol{\gamma}$ and similarly let χ_j denote the j-th block

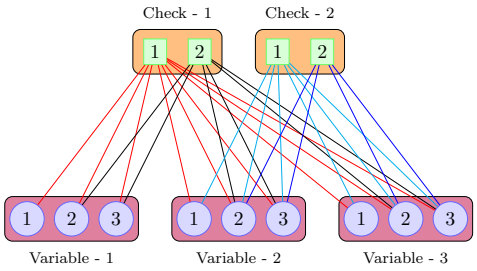

Fig. 5. Illustration of the connections at the node level in the factor graph corresponding to the Kronecker product when the sensing matrix is given by equation (30).

of $\boldsymbol{\chi}$. In the factor graph, 5, blocks γ_i and χ_j are connected if at least one node in γ_i is connected to at least one node in χ_j. The connections at the block level are defined by the sensing matrix $\boldsymbol{A}$. In other words, γ_i and χ_j are connected if $A_{ij} \neq 0$. Figure 6, illustrates the connections at the block level. Let us now focus on the connections between the nodes

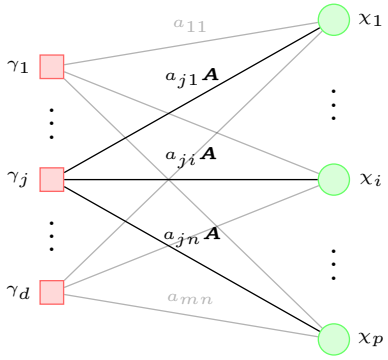

Fig. 6. Illustration of the connections in the factor graph at block level.

in block γ_j and χ_i. We denote $\gamma_j^{(k)}$ to be the k-th node in check node block j and χ_i^{ℓ} to be the ℓ-th node in the variable node block i. The connections between the blocks γ_j and χ_i, if it exists ($A_{ji} \neq 0$), is again characterized by $\boldsymbol{A}$. Figure 7 illustrated the connected between the nodes in a variable node block and a check node block.

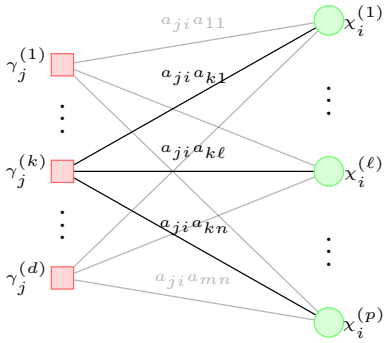

Fig. 7. Illustration of the connections between blocks γ_j and χ_i

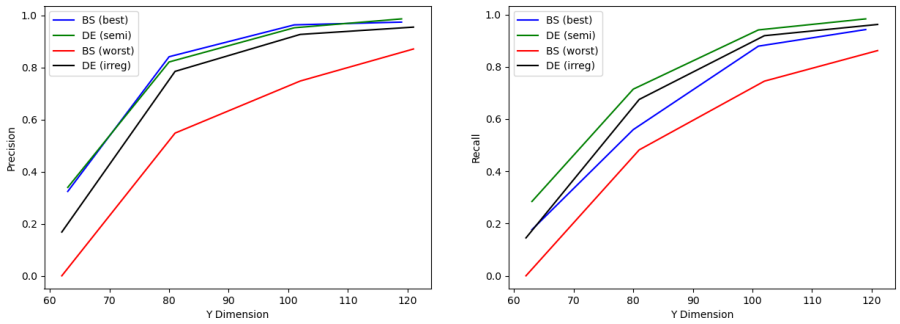

Fig. 8. Comparison of the performance of the proposed sensing system with that of [11] (BS - best, BS - worst lines in the plots). Number of nodes, $p = 200$. The performance is evaluated with respect to *precision* and *recall* of the edges in the graph.

Therefore we now have,

$$\gamma_j^{(k)} = \sum_{i=1}^{p} \sum_{\ell=1}^{p} A_{ji} A_{k\ell} \chi_i^{(\ell)}. \tag{31}$$

Since $\deg(\gamma_j^{(k)})$ would be the number of non-zero terms in the above summation, we then have $\deg(\gamma_j^{(k)}) = \deg(A^{(j)})\deg(A^{(k)})$, where $A^{(j)}$ denotes the j-th row of $\boldsymbol{A}$. Using a similar argument we can also conclude that $\deg(\chi_i^{(\ell)}) = \deg(A_i)\deg(A_\ell)$, where A_i denotes the i-th column of $\boldsymbol{A}$. Since $\deg(A_i) \in \{1, \ldots, d_v\}$ and $\deg(A^{(j)}) \in \{1, \ldots, d_c\}$ we have that $\deg(\gamma_j^{(k)}) \in \{1, \ldots, d_c^2\}$ and $\deg(\chi_i^{(\ell)}) = \{1, \ldots, d_v^2\}$. Therefore we have

$$P\Big(\deg(\gamma_j^{(k)}) = k\Big) = \sum_{j,j':jj'=k} \rho_j \rho_{j'} \tag{32}$$

And,

$$P\Big(\deg(\chi_i^{(\ell)}) = k\Big) = \sum_{i,i':ii'=k} \lambda_i \lambda_{i'} \tag{33}$$

Appendix C
Graph Structure Recovery (Regular Sensing)

Here we compare the performance of the proposed regular sensing matrix on graph structure recovery task with the sensing system proposed by [11]. The sensing systems are evaluated with respect to: (i) MAE, (ii) Precision, and (iii) Recall. We can see from Figure 8 that the relative performance between the two systems is similar to the behavior exhibited on the covariance recovery task. That is, the two sensing systems are at an equal footing when the baseline is tuned.