Efficient Inference of Spatially-Varying Gaussian Markov Random Fields With Applications in Gene Regulatory Networks

Visweswaran Ravikumar, Tong Xu, Wajd N. Al-Holou, Salar Fattahi, and Arvind Rao

Abstract—In this paper, we study the problem of inferring spatially-varying Gaussian Markov random fields (SV-GMRF) where the goal is to learn a network of sparse, context-specific GMRFs representing network relationships between genes. An important application of SV-GMRFs is in inference of gene regulatory networks from spatially-resolved transcriptomics datasets. The current work on inference of SV-GMRFs are based on the regularized maximum likelihood estimation (MLE) and suffer from overwhelmingly high computational cost due to their highly nonlinear nature. To alleviate this challenge, we propose a simple and efficient optimization problem in lieu of MLE that comes equipped with strong statistical and computational guarantees. Our proposed optimization problem is extremely efficient in practice: we can solve instances of SV-GMRFs with more than 2 million variables in less than 2 minutes. We apply the developed framework to study how gene regulatory networks in Glioblastoma are spatially rewired within tissue, and identify prominent activity of the transcription factor HES4 and ribosomal proteins as characterizing the gene expression network in the tumor peri-vascular niche that is known to harbor treatment resistant stem cells.

Index Terms—Markov random fields, glioblastoma, spatial transcriptomics.

I. INTRODUCTION

The advent of high throughput sequencing technologies has transformed our understanding of biological systems, and catalyzed the adoption of a systems-level approach to studying biological processes. Networks have emerged as the intuitive framework for reasoning about complex biological systems [1], [2]. Nodes in the network represent individual components, and edges represent direct interactions between them. For example, gene regulatory networks (GRNs) represent the wiring diagram of the cell’s information processing system, with network edges identifying regulatory interactions between different genes. It has become clear that complex diseases like cancer must be understood at the level of this interactome, rather than the classical reductionist approach of studying individual genes [3], [4]. As another example, with billions of neurons and with (imaging) data comprising hundreds of thousands of voxels, the human brain represents one of the most complex physiological networks, whose structure remains a long-standing mystery [5], [6]; see also the survey [7]. The accurate inference of the brain connectivity network will have a far-reaching impact on understanding different neurological disorders [8], [9].

According to the NIH’s BRAIN Initiative, the development of “faster, less expensive, and scalable” technologies is the cornerstone for anatomic reconstruction of neural circuits at realistic scales [10]. Spatially resolved transcriptomics have emerged as a transformative technology in the recent past with immense potential to bolster our understanding of biology [11], [12]. The technology is fast evolving, and allows transcriptome-scale profiling of gene expression at near single-cell resolution in tissues [13]. In studying complex processes such as tumor growth, viewing cancer as a case of evolution within the tissue has provided the groundwork for building a comprehensive theoretical framework to understand tumor diversity [14]. Spatial gradients in exposure to nutrients, oxygen, immune cells and environmental toxicity enforce trade-offs between proliferation and survival strategies in populations of tumor cells [15], [16]. Faced with multiple survival constraints and the need to optimize evolutionary fitness, cancer populations adopt a continuum of transcriptional states adapted to their local tissue micro-environments [17]. This results in significant spatial trends in the gene expression profiles and the underlying regulatory networks of the tumor. Being able to infer these dynamic regulatory networks would provide us with a new lens for understanding tumor heterogeneity. It can help generate hypotheses about context-specific molecular mechanisms important to survival and growth of the tumor, and expose previously unknown vulnerabilities that can be exploited in treatment.
One popular approach to model these problems is based on spatially-varying Markov Random Fields (SV-MRFs) [18]. SV-MRFs are associated with a network of undirected Markov graphs $G_k(V_k, E_k)$, where $V_k$ and $E_k$ are the set of nodes and edges in the graph at location $k$. The node set $V_k$ represents the random variables (e.g. genes) in the model, while the edge set $E_k$ captures the conditional dependency between these variables at location $k$. In the special case of Gaussian Markov Random Fields (GMRFs), the edge set of the Markov graphs can be fully characterized based on the inverse covariance matrix (also known as the precision matrix). In particular, if the entry $(i, j)$ of the precision matrix $\Theta_k$ is zero, then the variables $i$ and $j$ are independent conditioned on the remaining variables.

A widely-used method for the inference of SV-MRFs is based on the so-called regularized maximum-likelihood estimation (MLE). Intuitively, MLE seeks to find a graphical model based on which the observed data is most likely to occur. However, MLE-based methods suffer from major computational challenges that undermine their applicability in large-scale settings. For example, in the Gaussian setting, the MLE requires optimizing over the so-called log-determinant of the inverse covariance matrix, which are known to be intractable in large scales [19]. This drawback is further compounded in the spatially-varying regime, where the precision matrix must be estimated at each spatial location, leading to a dramatic increase in the size of the problem.

### A. Related Work

Recently, many approaches have been proposed for sparse precision matrix estimation in high dimensions. This line of work begins by the inference of a single precision matrix, which can be achieved by $\ell_1$-regularized MLE, also known as Graphical Lasso (GL) [20].

Extending beyond single precision matrix inference, a recent line of research has focused on estimating time-varying MRFs, where the relation among variables may change over time [21]. A common approach for estimating time-varying MRFs is based on kernel methods, where the sample covariance matrix at any given time is a weighted average of the samples over time, and the weights are collected from a predefined kernel [21], [22]. It is recently shown that kernel averaging combined with a tailored $\ell_0$-norm minimization technique can infer time-varying MRF in scales that are not possible via MLE-based approaches [19].

Despite the recent progress in the inference of time-varying MRFs beyond MLE-based methods, the developed techniques cannot be readily extended to spatially-varying instances. The main challenge behind such extension is that interactions among individual MRFs can be drastically more involved in the spatially-varying setting than its time-varying counterpart. Due to this difficulty, most existing results on spatially-varying MRFs are limited to different variants of MLE-based methods such as Fused Graphical Lasso (FGL), Group Graphical Lasso (GGL) [23], and co-hub node joint graphical lasso (CNJGL) [24] which suffer from expensive computational costs. FGL penalizes the pairwise difference of the precision matrices in $\ell_1$-norm, while GGL regularizes the $\ell_2$-norm of the $(i, j)$-th element across all $K$ precision matrices. Moreover, CNJGL assumes that there are a few common nodes that serve as hubs in different networks, and accordingly uses row- column overlap norm penalty to identify these nodes. Guo et al. [25] reparameterized each off-diagonal element as the product of a common factor and difference, then applied separate $\ell_1$-regularization to these two parts. Saegusa and Shojafar [26] proposed to regularize the MLE with a Laplacian-type penalty to exploit the information among different distributions. However, all of the aforementioned techniques are based on MLE, and consequently suffer from a notoriously high computational cost.

To alleviate the computational cost of MLE-based techniques, Lee and Liu [27] proposed to estimate the joint precision matrices based on a constrained $\ell_1$ minimization for inverse matrix estimation (CLIME) technique [28]. Unlike GL, CLIME does not optimize over the complex logdet function and has shown more favorable theoretical properties than GL. Our method is built upon the Elementary Estimator introduced by Yang et al. [29], where the proposed estimator admits a closed-form solution based on soft-thresholding. This method was later extended by Fattahi and Gomez [19] to time-varying setting, showing that it can be solved in near-linear time and memory. Our proposed estimator is also related to FASJEM estimator [30], which uses a similar Elementary Estimator as ours. We will provide an extensive comparison between our technique, FASJEM, and FGL in Section VI.

### B. Our Contributions

To address the aforementioned challenges, we propose a simple estimation method for the inference of spatially-varying GMRFs. Unlike MLE-based methods, our proposed approach is based on a class of simple and computationally efficient optimization methods that come equipped with strong statistical guarantees and are implementable in realistic scales. Our contributions are summarized as follows:

**Computational Guarantee:** Our proposed method reduces to a series of decomposable convex quadratic optimization problems that can be solved efficiently using any off-the-shelf solvers. In addition, the decomposable nature of the proposed optimization problem makes it amenable to parallel and distributed implementation.

**Statistical Guarantees:** In addition to the desirable computational guarantees, we show the statistical consistency of our proposed method—both theoretically and in practice. In particular, we characterize the non-asymptotic consistency of our proposed method, proving that it accurately recovers the underlying graphical model, even in the high-dimensional settings where number of available samples is significantly smaller than the number of unknown parameters. Moreover, it can efficiently reveal the correct sparsity information in the parameters and their differences.

**Application in Inferring Gene Regulatory Networks:** Glioblastoma (GB) is an incurable malignancy of the brain, with a median survival time of only 12-18 months despite therapy with surgical resection, chemotherapy and radiation [31]. Despite aggressive treatment, these tumors inevitably recur and this recurrence is likely due to significant heterogeneity, which has
been highlighted by single cell sequencing studies [32]. Heterogeneous populations of treatment-resistant tumor cells with stem cell-like properties have been identified in GB that have been shown to drive treatment recurrence. Furthermore, these resistant cells often reside within unique microenvironmental niches [33], [34]. The consequence of spatial context in regulating the tumor cell state, stemness properties, and treatment resistance in these tumors is increasingly appreciated [35], [36]. We therefore use our developed statistical framework to understand how the gene networks of GB are rewired as a function of their spatial environment, and identify context-specific upstream regulators of the heterogeneous tumor cell states.

**Organization:** The rest of the paper is organized as follows. In Section II, we formulate the inference of spatially-varying GMRFs and discuss the shortcomings of the existing techniques. Motivated by these shortcomings, we present a new formulation of the problem in Section III. In Section IV, we delineate the statistical guarantees of our proposed formulation, and how to solve it efficiently. Finally, we showcase the performance of our proposed method on synthetically generated as well as the Glioblastoma spatial transcriptomics dataset in Section VI. The proofs of our main theorems are provided in the appendix, available online.

## II. PROBLEM FORMULATION

In this section, we formally define the problem of inferring SV-MRFs. To this goal, we first present some notations that will be used throughout our paper.

**Notations:** The $i$-th element of a vector $v$ or $v_i$ is denoted as $v_i$ or $v_{i,j}$. For a matrix $M$, the notations $M_{i,:}$ and $M_{:,j}$ denote the $i$-th row and $j$-th column, respectively. Moreover, for an index set $S$ and a matrix $M$, the notations $M_{S,:}$ and $M_{:,S}$ refer to a submatrix of $M$ with rows and columns indexed by $S$, respectively. For a matrix $M$ or a vector $v$, the notations $\|M\|_{\ell_q}$ and $\|v\|_q$ correspond to the element-wise $\ell_q$-norm of $M$ and $\ell_q$-norm of $v$, respectively. Moreover, $\|M\|_{q,a}$ and $\|M\|_{q,\max}$ are the induced $q$-norm and the element with the largest absolute value of the matrix $M$, respectively. Moreover, $\|M\|_0$ denotes the total number of nonzero elements in $M$. We use $M \succ 0$ to show that $M$ is positive definite. For a vector $v$ and matrix $M$, the notations $\text{supp}(v)$ and $\text{supp}(M)$ are defined as the index sets of their nonzero elements. Given two sequences $f(n)$ and $g(n)$ indexed by $n$, the notation $f(n) \lesssim g(n)$ implies $f(n) \leq C g(n)$ for some constant $C \leq 1$. Moreover, the notation $f(n) \asymp g(n)$ implies that $f(n) \lesssim g(n)$ and $g(n) \lesssim f(n)$. The sign function $\text{sign}(\cdot)$ is defined as $\text{sign}(x) = x/|x|$ if $x \neq 0$ and $\text{sign}(0) = 0$. Accordingly, when $x$ is a vector, the function $\text{sign}(x)$ is defined as $[\text{sign}(x_1) \ \text{sign}(x_2) \ \ldots \ \text{sign}(x_n)]^\top$.

Consider data samples from $K$ different Gaussian distributions with $d \times d$ covariance matrices $\Sigma_k \in S_d^+$, $k = 1, \ldots, K$ and sparse precision matrices $\tilde{\Theta}_k = \Sigma_k^{-1}$, $k = 1, \ldots, K$. Let $\{x^k_i\}_{i=1}^{n_k}$ be $n_k$ independent samples, each with dimension $d$, drawn from the $k$-th distribution, i.e., $x^k_i \sim N(0, \Sigma_k^k)$, for every $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$. The zero-mean assumption on the distributions is without loss of generality and can be achieved by normalizing the data. Our goal is to estimate the precision matrices $\{\tilde{\Theta}_k\}_k$ given the samples. The most commonly-used method to perform this task is via maximum likelihood estimation (MLE) with an $\ell_1$ regularizer (also known as Graphical Lasso [20]):

$$\tilde{\Theta}_k = \arg\min_{\Theta_k \succ 0} \text{Tr}(\Theta_k \hat{\Sigma}_k) - \log \det(\Theta_k) + \lambda \|\Theta_k\|_{\ell_1}$$

where $\text{Tr}(\cdot)$ is the trace operator and $\hat{\Sigma}_k := \frac{1}{n_k} \sum_{i=1}^{n_k} x^k_i x^k_i\top$ is the sample covariance matrix for distribution $k$. A major drawback of the above estimation method is that it ignores any common structure among different distributions. To address this issue, a common approach is to consider a joint estimation method (also known as joint Graphical Lasso [23]):

$$\{\tilde{\Theta}_k\}_k = \arg\min_{\Theta_k \succ 0} \sum_{k=1}^{K} \left( \text{Tr}(\Theta_k \hat{\Sigma}_k) - \log \det(\Theta_k) + \lambda \|\Theta_k\|_{\ell_1} \right)$$

where $P(\{\Theta_k\}_k)$ is a penalty function that encourages similarity across different precision matrices. A major difficulty in solving joint Graphical Lasso is its computational complexity: in order to obtain an $\epsilon$-accurate solution, typical numerical solvers for (1) have complexity ranging from $O(K d^6 \log(1/\epsilon))$ (via general interior-point methods) [24] to $O(K d^3 / \epsilon)$ (via tailored first-order methods, such as ADMM) [23], [37], [38]. Solvers with such computational complexity fall short of any practical use in the large-scale settings. Indeed, the prohibitive worst-case complexity of methods based on Graphical Lasso is also exemplified in their practical performance; see [19], [39], [40], [41], [42], [43] for different examples.

## III. PROPOSED METHOD

To address the aforementioned issues, we propose the following surrogate optimization problem for estimating sparse precision matrices:

$$\{\tilde{\Theta}_k\}_k = \arg\min_{\{\Theta_k\}_k} \sum_{k=1}^{K} \left( \text{Tr}(\Theta_k \hat{\Sigma}_k) - \log \det(\Theta_k) + \lambda \|\Theta_k\|_{\ell_1} \right)$$

where $P(\{\Theta_k\}_k)$ is a penalty function that encourages similarity across different precision matrices. A major difficulty in solving joint Graphical Lasso is its computational complexity: in order to obtain an $\epsilon$-accurate solution, typical numerical solvers for (1) have complexity ranging from $O(K d^6 \log(1/\epsilon))$ (via general interior-point methods) [24] to $O(K d^3 / \epsilon)$ (via tailored first-order methods, such as ADMM) [23], [37], [38]. Solvers with such computational complexity fall short of any practical use in the large-scale settings. Indeed, the prohibitive worst-case complexity of methods based on Graphical Lasso is also exemplified in their practical performance; see [19], [39], [40], [41], [42], [43] for different examples.

In the above optimization, the backward mapping deviation captures the distance between the estimated precision matrix and the so-called approximate backward mapping which will be described in Section III-A. Moreover, the absolute regularization promotes sparsity in the estimated parameters, whereas spatial regularization encourages common spatial similarities among different parameters. For any given pair $(k, l)$, the weight $W_{kl}^{-1}$ can be interpreted as the “distance” between the $k$-th and $l$-th
MRFs. Accordingly, a large value for $W_{kl}$ encourages similarity between $\Theta_k$ and $\Theta_l$. Two common choices of spatial similarities are sparsity and smoothness:

- **Smoothly-changing GMRF**: In smoothly-changing GMRFs, the adjacent precision matrices vary gradually. In this setting, $q = 2$ can be used as the spatial regularizer in $\text{Elem} - q$ to promote the smoothness in the parameter differences.

- **Sparsely-changing GMRF**: In sparsely-changing GMRFs, the adjacent precision matrices differ only in a few entries. In this setting, $q = 1$ is a natural choice for the spatial regularizer in $\text{Elem} - q$ since it promotes sparsity in the parameter differences.

### A. Approximate Backward Mapping

Our proposed optimization problem is contingent upon the availability of an approximate backward mapping. For a GMRF, the backward mapping is defined as the inverse of the true covariance matrix, i.e., $F^*(\Sigma_k) = \Theta_k^{-1}$ [44]. Based on this definition, a natural surrogate for the backward mapping is $F^*(\hat{\Sigma}_k) = \hat{\Theta}_k^{-1}$, where $\hat{\Theta}_k$ is the sample covariance matrix for distribution $k$. However, in the high-dimensional settings, the number of available samples is significantly smaller than the dimension, and as a result the sample covariance matrix $\hat{\Sigma}_k$ is singular and non-invertible. To alleviate this issue, Yang et al. [29] introduce an approximation of the backward mapping based on soft-thresholding. Consider the operator $\mathcal{S}_T(M) : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$, where $\mathcal{S}_T(M)_{ij} = M_{ij} - \text{sign}(M_{ij}) \min(|M_{ij}|, \nu)$ if $i \neq j$, and $\mathcal{S}_T(M)_{ij} = M_{ij}$ if $i = j$. Given this operator, the approximate backward mapping is defined as $F^*(\hat{\Sigma}_k) = \mathcal{S}_T(\hat{\Sigma}_k)^{-1}$, for every $k = 1, \ldots, K$. An important property of this approximate backward mapping is that it is well-defined even in the high-dimensional setting $n < d$ with an appropriate choice of the threshold $\nu$ [29]. Given this approximate backward mapping, we will show that the estimated precision matrices from $\text{Elem} - q$ are close to their true counterparts with an appropriate choice of parameters.

### B. Decomposability

An important property of $\text{Elem} - q$ is that it naturally decomposes over different coordinates of the precision matrices; for every $(i, j)$ with $1 \leq i \leq j \leq d$, the $ij$-th element of $(\Theta_i)_{k=1}^{K}$ can be obtained by solving the subproblem:

$$\{\hat{\Theta}_{k;ij}\}_{k=1}^{K} = \arg\min_{(\Theta_{k;ij})_{k=1}^{K}} \sum_{k=1}^{K} (\Theta_{k;ij} - [F^*(\hat{\Sigma}_k)]_{ij})^2 + \mu \sum_{k=1}^{K} \Theta_{k;ij} + \sum_{l < k} W_{kl} |\Theta_{k;ij} - \Theta_{l;ij}|^q,$$

$$(\text{Elem} - (i, j, q))$$

Recall that the original problem $\text{Elem} - q$ has $Kd(d + 1)/2$ variables. The above decomposition implies that $\text{Elem} - q$ can be decomposed into $d(d + 1)/2$ smaller subproblems, each with only $K$ variables that can be solved independently in parallel.

This is in stark contrast with the joint Graphical Lasso, which requires a dense coupling among the elements of the precision matrices through the non-decomposable logdet function. Later, we will show how each subproblem can be solved efficiently for different choices of $q$.

### IV. Statistical Guarantees

In this section, we elucidate the statistical properties of $\text{Elem} - q$ for SV-GMRFs with two widely-used spatial structures, namely smoothly-changing and sparsely-changing GMRFs. To this goal, we first need to make two important assumptions on the true precision matrices.

**Assumption 1 (Bounded norm)**: There exist constant numbers $\kappa_1 < \infty$, $\kappa_2 > 0$, and $\kappa_3 < \infty$ such that

$$\|\Theta_i\|_\infty \leq \kappa_1, \quad \inf_{w : \|w\|_1 = 1} \|\Sigma_k w\|_\infty \geq \kappa_2, \quad \|\Sigma_k\|_{\max} \leq \kappa_3$$

for every $k = 1, \ldots, K$.

**Assumption 1** is fairly mild and implies that the true covariance matrices and their inverses have bounded norms.

**Assumption 2 (Weak sparsity)**: Each covariance matrix $\Sigma_k$ satisfies $\max_i \sum_{j=1}^d \|\Sigma_k\|_{ij} \leq s(p)$, for some function $s : [0, 1) \rightarrow \mathbb{R}$ and scalar $0 \leq p < 1$.

Informally, we say “the true covariance matrices are weakly sparse” if $(\Theta_i^*)_{\nu}$ are $(p)$-weakly sparse with $s(p) \ll d$ for some $0 \leq p < 1$. The notion of weak sparsity extends the classical notion of sparsity to dense matrices. Indeed, except for a few special cases, a sparse matrix does not have a sparse inverse. Consequently, a sparse precision matrix may not lead to a sparse covariance matrix. However, a large class of sparse precision matrices have weakly sparse inverses. For instance, if $\Theta_i^*$ has a banded structure with small bandwidth, then it is known that the elements of $\Sigma_k^* = \Theta_i^*-1$ enjoy exponential decay away from the main diagonal elements [45, 46]. Under such circumstances, simple calculation implies that $s(p) \leq C_2 \frac{\kappa_1}{\kappa_2}$ for some constants $C > 0$ and $\rho < 1$. More generally, a similar statement holds for a class of inverse covariance matrices whose support graphs have large average path length [47, 48]; a large class of inverse covariance matrices with row- and column-sparsity structures satisfy this condition. As will be shown later, the weak sparsity parameter $s(p)$ directly controls the sample complexity of our proposed estimator.

Next, we introduce some notations that simplify our subsequent analysis. Let $\pi : \{1, 2, \ldots, K\}^2 \rightarrow \{1, 2, \ldots, K\} (K(1)/2)$ be a fixed, predefined labeling function that assigns a label to each pair $(k, l)$ with $l \geq k$. Let $G$ be a diagonal matrix whose $k$-th diagonal entry is defined as $W_{\pi^{-1}(k)}$. Moreover, let $A \in \mathbb{R}^{K(K-1)/2 \times K}$ be the adjacency matrix defined as $A(\pi(k, l), k) = 1$ and $A(\pi(k, l), l) = -1$, for every $l > k$. Finally, define $\Theta_{ij} = \{\Theta_1;ij \; \Theta_2;ij \; \ldots \; \Theta_K;ij\}^\top$ and $\hat{F}_{ij} = \{[F^*(\hat{\Sigma}_1)]_{ij} \; [F^*(\hat{\Sigma}_2)]_{ij} \; \ldots \; [F^*(\hat{\Sigma}_K)]_{ij}\}$, for every $j \geq i$. It is easy to see that $\|GA\Theta_{ij}\|_q = \sum_{l \geq k} W_{kl} |\Theta_{k;ij} - \Theta_{l;ij}|^q$ for every $j \geq i$, and accordingly, $\text{Elem} - (i,j,q)$ can be written concisely as

$$\hat{\Theta}_{ij} = \arg\min_{\Theta_{ij}} \|\Theta_{ij} - F_{ij}\|_2 + \mu \Omega_{ij} + \gamma \|GA\Theta_{ij}\|_q.$$
Next, we provide sharp statistical guarantees for our proposed method when the precision matrices \( \{ \Theta^*_k \}_{k=1}^K \) change smoothly or sparsely across different distributions.

### A. Smoothly-Changing GMRF

We start with our main assumption on the smoothness of the precision matrices.

**Assumption 3 (Smoothly-changing SV-GMRFs):** There exists a constant \( D \geq 0 \) such that \( \sum_{i \geq k} (\Theta^*_k \Theta^*_i - \Theta^*_i)^2 \leq D^2 \) for every \( (i, j) \).

Informally, we say “SV-GMRF is smoothly-changing” if Assumption 3 is satisfied with a small \( D \). For a smoothly-changing SV-GMRF, it is natural to choose \( q = 2 \) in Elem – \( q \) to promote smoothness in the spatial difference of the precision matrices. Our next theorem characterizes the sample complexity of Elem – \( q \) with \( q = 2 \) for smoothly-changing SV-GMRF. Let

\[
\Theta_{\min} = \min\{ \Theta_{\epsilon,ij} : \Theta_{\epsilon,ij} \neq 0 \}.
\]

**Theorem 1 (Smoothly-changing SV-GMRF):** Consider a smoothly-changing SV-GMRF with parameter \( D \), and weakly-sparse covariance matrices with parameter \( s(\mu) \) for some \( 0 \leq \mu < 1 \). Suppose that the number of samples satisfies

\[
n_k \geq L \log d / \Theta_{\min},
\]

where

\[
L = \max \left\{ \left( \frac{s(\mu)}{\kappa_2} \right)^{2/3}, \left( \frac{\kappa_1 \kappa_3 + D \kappa_2}{\kappa_2} \right)^{2/3} \right\}.
\]

Define \( n_{\min} = \min_k \{ n_k \} \). Moreover, suppose that \( \hat{F}^* (\Sigma_k) = [\Theta^*_k]^{-1} \) \( \Theta_k \approx \kappa_3 \sqrt{\log d / n_{\min}} \). Then, the solution obtained from Elem – \( q \) with \( q = 2 \) and parameters

\[
\gamma \approx \frac{1}{K \| W \|_{\max}} \sqrt{\log d / n_{\min}}, \quad \mu \approx D \sqrt{\log d / n_{\min}},
\]

satisfies the following statements with probability of \( 1 - K d^{-10} \):

- **Sparsistency:** The solution is unique and satisfies \( \supp(\hat{\Theta}_k) = \supp(\Theta^*_k) \) for every \( k \).
- **Estimation error:** The solution satisfies

\[
\| \Theta_k - \Theta^*_k \|_{\max} \leq \left( \frac{\kappa_1 \kappa_3 + D \kappa_2}{\kappa_3} \right)^{1/3}, \quad \text{for every } k.
\]

For smoothly-changing SV-GMRF, the above theorem provides a non-asymptotic guarantee on the estimation error and sparsistency of the estimated precision matrices via Elem – \( q \) with \( q = 2 \), proving that the required number of samples must scale only logarithmically with the dimension \( d \). Moreover, both the estimation error and the required number of samples decrease with a smaller smoothness parameter \( D \); this is expected since a small value of \( D \) implies that the adjacent distributions share more information, and hence, the SV-GMRF is easier to estimate.

### B. Sparsely-Changing GMRF

In sparsely-changing SV-GMRFs, the precision matrices are assumed to change sparsely across different distributions; this is formalized in our next assumption.

**Assumption 4 (Sparsely-changing SV-GMRFs):** There exists a constant \( D_0 \geq 0 \) such that \( \sum_{i \geq k} (\Theta^*_k \Theta^*_i - \Theta^*_i) \leq D_0 \) for every \( (i, j) \).

Similar to the smoothly-changing SV-GMRFs, we say “SV-GMRFs is sparsely-changing” if it satisfies Assumption 4 with a small \( D_0 \). For a sparsely-changing SV-GMRF, it is natural to choose \( q = 1 \) in Elem – \( q \) to promote sparsity in the spatial difference of the precision matrices. To analyze the statistical property of this problem, we first consider (2) with \( q = 1 \) and rewrite it as:

\[
\min \| F^*_\gamma - \Theta_{ij} \|^2 + \mu \| B \Theta_{ij} \|_1, \quad B = \begin{bmatrix} \frac{\gamma}{\mu} G A & I \end{bmatrix}.
\]

The above reformulation is a special case of the generalized Lasso problem introduced by Lee et al. [49]. To show the model selection consistency of the above formulation, we next introduce the notion of irrepresentability.

For any fixed \( (i, j) \), let \( S_B \subseteq \{ 1, 2, ..., K(K+1)/2 \} \) be the support of \( B \Theta^*_{ij} \), i.e., \( \| B \Theta^*_{ij} \| \neq 0 \) for every \( k \in S_B \). Moreover, let \( S_B = \{ 1, 2, ..., K(K+1)/2 \} \setminus S_B \). Evidently, we have \( |S_B| \leq D_0 + S_0 \), where \( D_0 \) is introduced in Assumption 4 and \( S_0 \) is defined as the maximum number of nonzero elements in \( \Theta^*_{ij} \), i.e., \( S_0 = \max_k \{ \| \Theta^*_{ij} \|_0 \} \).

**Assumption 5 (Irrepresentability condition (IC), Lee et al. [49]):** We have

\[
\begin{bmatrix} \Theta^*_{ij} \end{bmatrix}_{S_B} = 0 \quad \text{for every } 0 < \alpha \leq 1, \quad \text{where } B_{S_B}^l = \Theta^*_{ij} \Theta^*_{ij} S_B \}
\]

for some \( 0 < \alpha \leq 1 \), where \( B_{S_B}^l \) is the Moore-Penrose pseudoinverse of a matrix \( B_{S_B} \).

The irrepresentability condition (IC) entails that the rows of \( B \) corresponding to the zero elements of \( B \Theta^*_{ij} \) must be nearly orthogonal to the other rows. Despite the seemingly complicated nature of IC, classical results on Lasso have shown that it is a necessary condition for the exact sparsity recovery, and hence, cannot be relaxed [50, 51]. Later, we show that this condition is satisfied for our problem under a mild condition on the weight matrix \( W \) and parameters \( \mu \) and \( \gamma \).

Another quantity that plays a central role in our derived bounds is the so-called compatibility constant defined as

\[
\kappa_{IC} := \left\| B_{S_B} B_{S_B}^l \right\|_{\infty}^2 + 1.
\]

The compatibility constant \( \kappa_{IC} \) is closely related to IC. In particular, if \( \| B_{S_B} B_{S_B}^l \|_{\infty} \leq 1 - \alpha \) (which is a slightly stronger version of IC), then \( \kappa_{IC} \leq 2 - \alpha \). Similar to IC, we will later show that \( \kappa_{IC} \) remains bounded under a mild condition on the weight matrix \( W \). Finally, we define \( \Delta \Theta_{\min} = \min_{k, i \neq j} \{ \Theta^*_{ij} \Theta^*_{ij} \Theta^*_{ij} ; \Theta^*_{ij} \Theta^*_{ij} \neq 0 \} \).

**Theorem 2 (Sparsely-changing SV-GMRFs):** Consider a sparsely-changing SV-GMRFs with parameter \( D_0 \), and weakly-sparse covariance matrices with parameter \( s(\mu) \) for some \( 0 \leq \mu < 1 \).
p < 1. Suppose that the number of samples satisfies

\[ n_{\text{min}} \gtrsim L \frac{\log d}{\min \Theta_{\text{min}}, \Delta \Theta^2_{\text{min}}} \]

where \( L = \left\{ \left( s(p) \right)^{\frac{1}{2}} \frac{\gamma_2}{\kappa_2} \left( \frac{\kappa_C \kappa_1 \kappa_3}{\nu_2 \alpha} \right)^{\frac{1}{2}} \left( \| W \|_{\max} D_0 + S_0 \right) \right\} \).

Define \( n_{\text{min}} = \min_k \{ n_k \} \). Suppose that \( \tilde{F}^* (\hat{S}_k) = [s p_{\nu_k} (\hat{S}_k)]^{-1} \) with \( \nu_k \approx \kappa_3 \sqrt{\log d / n_k} \). Moreover, suppose that the weight matrix \( W \) and parameters \( \mu \) and \( \gamma \) are chosen such that IC (Assumption 5) is satisfied. Then, the solution obtained from Elem - \( q \) with \( q = 1 \) and parameter \( \mu \simeq \frac{\kappa_C \kappa_1 \kappa_3}{\nu_2 \alpha} \frac{\log d}{n_{\text{min}}} \)

satisfies the following statements with probability \( 1 - K d^{-10} \):

- **Sparsistency:** The solution is unique and satisfies
  \( \text{supp}(\hat{\Theta}_k) = \text{supp}(\Theta^*_k) \) for every \( k \) and \( \text{supp}(\hat{\Theta}_k - \Theta_k) = \text{supp}(\Theta^*_k - \Theta^* k) \) for every \( k > l \).

- **Estimation error:** For every \( (i, j) \), the solution satisfies
  \[ \| \hat{\Theta}_{ij} - \Theta^*_{ij} \| \lesssim \left( \sqrt{\| W \|_{\max} D_0 + S_0} \right) \frac{\kappa_C \kappa_1 \kappa_3}{\nu_2 \alpha} \sqrt{\frac{\log d}{n_{\text{min}}} \sqrt{\log d}} \]

The above theorem characterizes the sample complexity of inferring sparsely-changing SV-GMRFs, showing that the sparsity pattern of the precision matrices and their differences can be recovered exactly, given that the number of samples scale logarithmically with the dimension and the problem satisfies IC. Evidently, our result crucially relies on IC and \( \kappa_C \) being small. This leads to a follow-up question: how restrictive are these conditions in practice? Our next proposition shows that both conditions hold if \( \gamma \) and \( \mu \) are selected such that \( \mu \leq \gamma \leq 2\mu \) and \( W_{kl} \) is the same for every \( k > l \).

**Proposition 1:** Suppose that \( 0 < \mu \leq \gamma \leq 2\mu \) and \( W_{kl} \) is the same for every \( k \) and \( l \). Then, \( 1 \leq \kappa_C \leq 5 \) and IC holds with \( \alpha = \mu / \gamma \).

**Algorithm:** Next, we explain a general algorithm for solving Elem - \( q \). As mentioned before, Elem - \( q \) decomposes over different coordinates \( (i, j) \), where each subproblem can be written as Elem - \( (i, j, q) \). This decomposition leads to a parallelizable algorithm, where each thread solves Elem - \( (i, j, q) \), for a subset of the coordinates \( (i, j) \). This approach is outlined in Algorithm 1.
can be reformulated as a linearly constrained and can be computed in $E_{[53]}^q$ = 1 machines, the total complexity of solving all $q_{JGL}$ fasjem in MATLAB to solve each $q_{(i,j)}$ to $q_{(i,j)}$, $(i-j)_q$, (ii) edge reconnection. To perform Type (i) perturbation, we consider precision matrix by perturbing its parent cluster. We consider two types of perturbations: (i) edge weight perturbation; and (ii) edge reconnection. To perform Type (i) perturbation, we sample a subset of the $M$ modules at the parent cluster, and add a uniform perturbation from the interval $[-0.04, 0.04]$ to the non-zero edges. For Type (ii) perturbation, we replace one of the $M$ modules with a newly simulated one following a power-law degree distribution. Thus, at every cluster, the precision matrix is slightly perturbed relative to its parent, and the precision matrix differences accumulate, which means that the number of different edges of two precision matrices increases with their distance. Fig. 1 illustrates the precision matrices for the two adjacent clusters. Having simulated the precision matrices, at every cluster $k$, we next collect $n_k$ samples from a zero-mean Gaussian distribution with the constructed precision matrix.

We simulate the true precision matrices for $K$ distinct clusters (populations) with varying level of similarity. Within each cluster, we assume that the graph representing the true precision matrix has a disjoint modular structure, with power law degree distribution for nodes within each module. Specifically, we split $d$ genes into $M$ modules, with $d/M$ genes per module generated based on Barabasi-Albert model [56]. Within each cluster, the modules are simulated independently and concatenated to produce a block-diagonal matrix, which is treated as the true precision matrix for the corresponding cluster. In appendix 13.1, available online, we provide more simulations on instances with different number of genes in each module.

In order to simulate population-specific precision matrices, we first generate a random spanning tree over clusters. Starting with the root population, we generate $M$ modules, and in each module, we randomly generate a graph with $d/M$ vertices according to the Barabasi-Albert model. Based on the adjacency matrix of this graph, we select the edge weights uniformly from $[-1, -0.4] \cup [0.4, 1]$. To ensure the positive semi-definiteness of the constructed precision matrix, we use 1.1 times the sum of the absolute values of all off-diagonal elements in each row as the value of the diagonal elements in that row. The simulated modules are used to generate the population-specific block-diagonal precision matrix. We then traverse the spanning tree from the root cluster and, at every new cluster, construct the precision matrix by perturbing its parent cluster. We consider two types of perturbations: (i) edge weight perturbation; and (ii) edge reconnection. To perform Type (i) perturbation, we sample a subset of the $M$ modules at the parent cluster, and add a uniform perturbation from the interval $[-0.04, 0.04]$ to the non-zero edges. For Type (ii) perturbation, we replace one of the $M$ modules with a newly simulated one following a power-law degree distribution. Thus, at every cluster, the precision matrix is slightly perturbed relative to its parent, and the precision matrix differences accumulate, which means that the number of different edges of two precision matrices increases with their distance. Fig. 1 illustrates the precision matrices for the two adjacent clusters. Having simulated the precision matrices, at every cluster $k$, we next collect $n_k$ samples from a zero-mean Gaussian distribution with the constructed precision matrix.

**Algorithm 1:** General Algorithm for Solving $\text{Elem} - q$.

1. **Data:** Data samples $\{x_i^k\}$, parameters $(\mu, \nu_k)$, and weight matrix $W$.
2. **Result:** Solution $(\hat{\Theta}_k)^K_{k=1}$ for $\text{Elem} - q$.
3. Compute the sample covariance matrix $\hat{\Sigma}_k$ for every $k = 1, \ldots, K$.
4. Compute the approximate backward mapping $\hat{F}^*(\hat{\Sigma}_k) = [\exp (\hat{\Sigma}_k)]^{-1}$ for every $k = 1, \ldots, K$.
5. For every $(i, j)$ do
   - Obtain $\hat{\Theta}_{i,j}$ by solving $\text{Elem} - (i, j, q)$.

**VI. NUMERICAL EXPERIMENTS**

In this section, we evaluate the performance of our proposed method on synthetically generated dataset, as well as the Glioblastoma spatial transcriptomics dataset. All experiments are implemented using MATLAB 2021b, and performed with a 3.2GHz 8-Core AMD Ryzen 7 5800H CPU with 16 GB of RAM. We use the function `quadprog` in MATLAB to solve each subproblem. To compare our algorithm with FGL and FASJEM, we used the R packages JGL and fasjem.

**A. Synthetically Generated Dataset**

We use synthetically generated dataset to compare the statistical performance of our proposed method with two other estimators: the fused graphical lasso (FGL) [23] and FASJEM [30].

FGL is an MLE-based approach augmented by a regularizer to promote spatial similarity among different distributions. On the other hand, FASJEM uses the same Elementary Estimator [29] framework as ours while having different regularization term. By comparing the estimated parameters with their true counterparts, we will show that our method outperforms both FGL and FASJEM in recovering the true precision matrices.

**Data Generation:** Our data generation procedure is motivated by ideas proposed by Peng et al. [54] and Lyu et al. [55] to imitate gene expression profiles from a synthetically-generated co-expression network. Our goal is to generate the data synthetically from a known distribution, and then evaluate the performance of the estimated parameters by comparing them to the ground truth.
Fig. 2. Precision, Recall, and F1-score for the estimated precision matrices for different methods with varying sample size. We assume that \( n_k = N \) for every \( k = 1, \ldots, K \). Elem-1 and Elem-2 perform similarly, and they both outperform FGL and FASJEM. The higher value of Recall for FASJEM is due to the underestimation of the regularization parameters that promote sparsity, which in turn leads to a large value of TP.

Fig. 3. Precision, Recall, F1-score for the estimated precision matrices, as well as the runtime of our proposed method with varying dimension. Here, \( K_p \) is total number of parameters to be estimated by our method. Elem-1 and Elem-2 outperform FASJEM in terms of Precision and F1-score, while FASJEM outperforming Elem-1 and Elem-2 in terms of Recall. Similar to the previous experiment, the higher value of Recall for FASJEM is due to the underestimation of the regularization parameters which lead to overly dense precision matrices. Moreover, both Elem-1 and Elem-2 are drastically faster than FASJEM.

**Experiment 1. Varying number of samples:** In our first experiment, we fix \( K = 5, d = 250 \), and \( M = 5 \), and compare the performance of Elem – q with FGL and FASJEM with varying number of samples \( n_k \). We compare the estimation accuracy in terms of Recall = TP/(TP + FN), Precision = TP/(TP + FP), and F1-score = 2(Recall \times Precision)/(Recall + Precision), where TP, FN, and FP correspond to the number of correctly identified nonzero elements (true positive), incorrectly identified zero elements (false negative), and incorrectly identified nonzero elements (false positive), respectively. To fine-tune the weight matrix \( W \) and the parameters \( \mu, \gamma, \nu_k \), we use the distance measure and BIC approach delineated in Section V. Moreover, we use the same BIC approach to fine-tune the parameters of FGL and FASJEM.

Fig. 2 illustrates the performance of different estimation methods. It can be seen that Elem – q with \( q = 1, 2 \) (denoted as Elem-1 and Elem-1) perform almost the same, and they both outperform FASJEM and FGL in terms of the Precision and F1 scores. On the other hand, the Recall score for FASJEM is artificially high due to the underestimation of the regularization parameters via BIC, which in turn leads to overly dense estimation of the precision matrices.

**Experiment 2. Varying dimension:** Next, we analyze the performance of our proposed method for different dimensions \( d \). In particular, we consider a high-dimensional regime where \( d \) is significantly larger than the number of available samples \( n_k \). We fix \( K = 5 \) and set \( n_k = d/2 \). The parameters \( \mu, \gamma, \nu_k \) and the weight matrix \( W \) are tuned as before.

Fig. 3 depicts the Precision, Recall, and F1-score, as well as the runtime of our proposed method with varying dimension. Here, \( K_p \) is total number of parameters to be estimated by our method. Elem-1 and Elem-2 outperform FASJEM in terms of Precision and F1-score, while FASJEM outperforming Elem-1 and Elem-2 in terms of Recall. Similar to the previous experiment, the higher value of Recall for FASJEM is due to the underestimation of the regularization parameters which lead to overly dense precision matrices. Moreover, both Elem-1 and Elem-2 are drastically faster than FASJEM.

**Experiment 3. Varying number of clusters:** Finally, we evaluate the performance of our method with varying number of clusters \( K \). We fix \( d = 500, M = 10 \) and \( n_k = 250 \), and use the same tuned parameters in the previous experiment. Fig. 4 shows the Precision, Recall, and F1 score for our proposed method and FASJEM, as well as their runtime with respect to \( K \). Similar to the previous experiments, both Elem-1 and Elem-2 outperform FASJEM in terms of the estimation accuracy. Moreover, it can
Fig. 4. Precision, Recall, F1-score for the estimated precision matrices, as well as the runtime of our proposed method with varying number of clusters. *Elem-1* and *Elem-2* outperform *FASJEM* in terms of Precision and F1-score, while *FASJEM* outperforming *Elem-1* and *Elem-2* in terms Recall. Similar to the previous experiments, the higher value of Recall for *FASJEM* is due to the underestimation of the regularization parameters which lead to overly dense precision matrices. Moreover, both *Elem-1* and *Elem-2* are drastically faster than *FASJEM*.

Fig. 5. Adjacent tumor sections from a primary GB patient sample are separated into five distinct clusters informed by their expression similarity and spatial proximity.

be seen that in practice, the runtime of *Elem-1* and *Elem-2* scale almost linearly with $K$.

B. Application to Glioblastoma Spatial Transcriptomics

1) **Visium Data Collection and Clustering:** We assayed spatial gene expression profiles using the Visium platform from a primary patient tumor area showing high perfusion signal in diffusion MRI. The Visium slides profile gene expression from a $6.5 \times 6.5$ mm section of tumor tissue. We sampled two adjacent tissue sections from this region, yielding 6158 spots with expression data.

We integrate data from adjacent tissue slices using the reciprocal PCA method in *Seurat* [57]. We use the dimension reduction algorithm *PHATE* [58] to obtain a 3D embedding of the integrated counts data that captures the expression similarity of spots. To cluster the data in a spatially informed manner, we compute two separate distance matrices between spots, from their PHATE embedding and tissue coordinates. We perform upper quantile normalization of the distance matrices based on their 75th quantile to ensure that both expression and spatial distances are in the same scale, and use their sum to define pairwise distances between spots. This dissimilarity matrix is used as input for PAM clustering. Optimal number of clusters ($k = 5$) is identified using the Calinski-Harabasz criterion [59], with the resulting clusters shown in Fig. 5.

To understand biological characteristics of these clusters, and to aid in downstream interpretation of inferred networks, we performed spot deconvolution using the RCTD algorithm [60].

Since spots in the Visium microarray have a resolution of about 60 μm, they could be composed of multiple cell types. We thus used annotated single cell RNASeq dataset from Darmanis et al. [61] to identify compositional differences between the regions. Fig. 6 shows the deconvolution results. We see that the tissue is primarily composed of neoplastic cells. Cluster 4 represents a distinctive peri-vascular niche with significant immune infiltration, and cluster 5 has a high proportion non-neoplastic cells. With this knowledge, we now seek to understand how the gene interaction network vary in different regional microenvironments of this tumor section.

2) **Data Preparation for Network Inference:** We identify the top 2500 genes showing significant spatial trends in their expression determined using the SparkX algorithm [62]. Since gene expression counts data follow a negative-binomial distribution rather than Gaussian, we use the non-paranormal transformation to make the data amenable for analysis using the Gaussian graphical modeling framework [63]. This procedure works by non-parametrically estimating monotone
functions \( f_j \) such that the transformed variables \( f(X) = (f_1(X_1), f_2(X_2), \ldots f_p(X_p)) \) are normally distributed \( f(X) \sim \mathcal{N} (\mu, \Sigma) \). Importantly, the transformation preserves the independence relations between the variables \( X_j \), so the graph structure is not altered. We use the \( R \) package \texttt{huge} [64] for transforming the normalized spot-level counts data.

Inter-cluster similarity constraints for network inference are imposed based on pairwise distances between cluster medoids. Since our clustering is based on proximity in both expression and spatial coordinates, we believe this is a reasonable biological constraint to impose on the network inference algorithm. We transform distances into similarity using the formula

\[
W_{i,j} = \frac{1}{1 + d_{i,j}},
\]

where \( d_{i,j} \) is the sum of expression and spatial distance between medoids for clusters \( i, j \).

We use BIC criterion to identify the optimal sparsity and similarity constraint parameters for the network inference. Network inference showed that of the 2500 genes considered, 1180 have an edge in at least one cluster.

3) Network Structures and Biological Significance: The tissue section shows large variations in number of expressed genes between clusters (visualized in Fig. 13 in the appendix, available online), reflecting biological trends in transcriptional activity across the tumor slide. The number of inferred edges per cluster are respectively 4511, 13785, 446, 8400 and 4534. The spatial region in Cluster 3 has significantly fewer detected genes than other regions explaining the very low connectivity in its network. To compare the extent of similarity between the inferred networks, we use the DeltaCon algorithm [65], a statistically principled and scalable inter-graph similarity function. We see that inferred networks from the different clusters share little similarity, with Clusters 1 and 3 having the maximum pairwise similarity of 0.27. The network in Cluster 4 is maximally different from the other regions. This is as expected, given that this region is compositionally most unique.

Fig. 7 shows the top 15 strongest edges in each cluster from the inferred network, with node sizes scaled by their respective degree and negative interactions shown as dotted lines. We can immediately see that each cluster has distinct underlying regulatory interactions driving their transcriptional states, even if they appear compositionally homogeneous.

Transcription factors are proteins that play a dominant role in regulating gene expression networks of cells and are particularly important in driving tumor growth and evolution [66]. By binding to regulatory regions of target genes, they are responsible for regulating gene expression and thereby controlling cell states. Regulatory interactions involving TFs are therefore of particular interest in understanding gene networks. We highlight these interactions in Fig. 8. Since Cluster 2 has an order of magnitude more edges than other networks, we highlight only the top 100 edges.

Frequently highlighted TFs active across different regions include the AP1 family TFs FOS and JUN, which are known downstream effectors of the Mesenchymal state in Gliomas [67], and other master regulators such as CEBPD [68] and oligodendroglial lineage factors OLIG1 and OLIG2 [69]. We also see significant activity of SOX2 in Cluster 1, a known drivers of stemness features and radiation-resistance in Gliomas [70]. Cluster 3 shows significant activity of Lactotransferrin (LTF), which encodes an iron-binding protein with known innate immune and tumor-suppressive activity [71]. Interestingly, this gene has also been characterized as being an upstream master regulator of different GB subtypes [72], warranting further exploration of this gene in driving tumorigenesis in GB.

The TF network in Cluster 4, which represents the perivascular niche, is most different from the other regions, as expected given its unique microenvironment. This region shows prominent activity of HES4, a known downstream effector of the NOTCH signaling pathway that is known to inhibit cell differentiation and helps maintain the stemness features in Gliomas [73]. HES4 specifically regulates proliferative properties of neural stem cells, and reduces their differentiation. This is a very promising observation, given that the perivascular niche is known to harbor therapy-resistant glioma stem cells whose properties are critically driven through NOTCH signaling [36]. Cluster 5 has dominant activity of NME2, a nucleotide diphosphate kinase enzyme involved in cellular nucleotide metabolism and DNA repair [74]. The NME2 protein has also been identified to be a highly specific Tumor-associated antigen in IDH mutant Gliomas [75]. By studying the regulatory networks in each cluster, we are thus able to infer differential activity of different regions.

Fig. 7. Top 15 strongest edges in each network are shown. Size of the nodes reflects their degree. Dotted lines indicate negative interactions. The nodes are colored by the cluster identity.

Fig. 8. Transcription factor interactions in each cluster are highlighted. For each cluster, we show the top 100 strongest edges involving TFs. The graphs are visualized using the Davidson-Harel layout. Color scheme is same as in Fig. 7.
4) Potential Limitations of Our Approach: In this section, we have strived to highlight how our algorithm for joint network estimation can be paired with the fast-evolving spatial transcriptomics technology to understand molecular drivers of tumor heterogeneity and identify context-specific vulnerabilities of cancer. Even though our algorithm is highly scalable and can be used for transcriptome-wide network inference, the findings depend on sensitivity and resolution of the underlying sequencing technology. Visium data is known to suffer from significant gradients in RNA detection sensitivity [78], as we observe here in our tissue slides that significantly affects the structure of inferred networks. We however believe that with increasing sensitivity and resolution of spatial transcriptomics technologies, our algorithm will be extensively applicable to infer gene regulatory networks driving context-specific tumor evolution. Finally, we provide more experiments and discussions on our biological observations in Appendix 14, available online.

VII. Conclusion

In this work, we study the inference of spatially-varying Gaussian Markov random fields (SV-GMRFs) and its application in gene regulatory networks in Spatially resolved transcriptomics. The existing methods for inferring GMRFs suffer from the so-called curse of dimensionality, which limit their applicability to small-scale and spatially-invariant networks. To address this challenge, we propose a simple and efficient inference framework for inferring SV-GMRFs that comes equipped with strong statistical guarantees. Contrary to the existing MLE-based methods, our proposed method is amenable to parallelization and is based on solving a series of decomposable convex quadratic programs. We show that our proposed method is extremely efficient in practice, and outperforms the existing state-of-the-art techniques—both computationally and statistically. We study the developed framework in the context of inferring gene networks underlying oncogenesis, using Glioblastoma as case study. We characterize gene networks of tumor cells within different microenvironments, and identify drivers of context-dependent tumor adaptation which is an important step towards developing targeted therapies for cancer.

REFERENCES

[1] A.-L. Barabasi and Z. N. Oltvai, “Network biology: Understanding the cell’s functional organization,” Nature Rev. Genet., vol. 5, no. 2, pp. 101–113, 2004.
[2] B. Zhang, Y. Tian, and Z. Zhang, “Network biology in medicine and beyond,” Circulation Cardiovasc. Genet., vol. 7, no. 4, pp. 536–547, 2014.
[3] T. M. Cheng, S. Gulati, R. Agius, and P. A. Bates, “Understanding cancer mechanisms through network dynamics,” Brief. Funct. Genomic., vol. 11, no. 6, pp. 543–560, 2012.
[4] W. Du and O. Elemento, “Cancer systems biology: Embracing complexity to develop better anticancer therapeutic strategies,” Oncogene, vol. 34, no. 25, pp. 3215–3225, 2015.
[5] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: Uses and interpretations,” Neuroimage, vol. 52, no. 3, pp. 1059–1069, 2010.
[6] X. Liu and J. H. Dyun, “Time-varying functional network information extracted from brief instances of spontaneous brain activity,” in Proc. Nature Acad. Sci. USA, vol. 110, no. 11, pp. 4392–4397, 2013.
[7] K. J. Friston, “Functional and effective connectivity: A review,” Brain Connectivity, vol. 1, no. 1, pp. 13–36, 2011.
[8] Y. Du, Z. Fu, and V. D. Calhoun, “Classification and prediction of brain disorders using functional connectivity: Promising but challenging,” Front. Neurosci., vol. 12, 2018, Art. no. 525.
[9] D. S. Bassett and E. T. Bullmore, “Human brain networks in health and disease,” Curr. Opin. Neurol., vol. 22, no. 4, 2009, Art. no. 340.
[10] C. Bargmann et al., “Brain 2025: A scientific vision,” in Brain Research Through Advocating Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, Bethesda, MD, USA: NIH, 2014.
[11] A. Rao, D. Barkley, G. S. França, and I. Yanai, “Exploring tissue architecture using spatial transcriptomics,” Nature, vol. 596, no. 7871, pp. 211–220, 2021.

[12] E. Lein, L. E. Borns, and S. Linnarsson, “The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing,” Science, vol. 358, no. 6359, pp. 64–69, 2017.

[13] M. Asp, J. Bergensträhle, and J. Lundeberg, “Spatially resolved transcriptomes—next generation tools for tissue exploration,” BioEssays, vol. 42, no. 10, 2020, Art. no. 1900221.

[14] L. M. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley, “Cancer as an evolutionary and technological process,” Nature Rev. Cancer, vol. 6, no. 12, pp. 924–935, 2006.

[15] C. Carmona-Fontaine, M. Deforeto, L. Akkari, C. B. Thompson, J. A. Joyce, and J. B. Xavier, “Metabolic origins of spatial organization in the tumor microenvironment,” in Proc. Nature Acad. Sci. USA, vol. 114, no. 11, pp. 2934–2939, 2017.

[16] J. Hauser and U. Alon, “Tumour heterogeneity and the evolutionary trade-offs of cancer,” Nature Rev. Cancer, vol. 20, no. 4, pp. 247–257, 2020.

[17] D. Barkley et al., “Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment,” Nature Genet., vol. 54, no. 8, pp. 1192–1201, 2022.

[18] M. Hanar, “Gaussian random field models for spatial data,” in Handbook of Markov Chain Monte Carlo. Boca Raton, FL: CRC Press, 2011, pp. 449–478.

[19] S. Fattahi and A. Gomez, “Scalable inference of sparsely-changing Gaussian Markov random fields,” in Proc. Adv. Neural Inf. Process. Syst., 2021, pp. 6529–6541.

[20] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” Biostatistics, vol. 9, no. 3, pp. 432–441, 2008.

[21] S. Zhou, J. Lafferty, and L. Wasserman, “Time varying undirected graphs,” Mach. Learn., vol. 80, no. 2, pp. 295–319, 2010.

[22] K. Greinenwald, S. Park, S. Zhou, and A. Giessing, “Time-dependent spatially varying graphical models, with application to brain fMRI data analysis,” in Proc. Int. Conf. Neur. Inf. Process. Syst., 2017, pp. 5834–5842.

[23] P. Danaher, P. Wang, and D. M. Witten, “The joint graphical lasso for inverse covariance estimation across multiple classes,” J. Roy. Statist. Soc. Ser. B Statist. Methodol., vol. 76, no. 2, pp. 373–397, 2014.

[24] K. Mohan, P. London, M. Fazel, D. Witten, and S.-I. Lee, “Node-based learning of multiple Gaussian graphical models,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 445–488, 2014.

[25] J. Guo, E. Levina, G. Michailidis, and J. Zhu, “Joint estimation of multiple graphical models,” Biometrika, vol. 98, pp. 1–15, 2011.

[26] T. A. Saengsa and A. Shojaie, “Joint estimation of precision matrices in heterogeneous populations,” Electron. J. Statist., vol. 10, no. 1, pp. 1341–1392, 2016. [Online]. Available: https://doi.org/10.1214/16-EJS1137

[27] W. Lee and Y. Liu, “Joint estimation of multiple precision matrices with common structures,” J. Mach. Learn. Res., vol. 16, no. 31, pp. 1035–1062, 2015. [Online]. Available: http://jmlr.org/papers/v16/lee15a.html

[28] T. Cai, W. Liu, and X. Luo, “A constrained 1 minimization approach to sparse precision matrix estimation,” J. Amer. Statist. Assoc., vol. 106, no. 494, pp. 594–607, 2011.

[29] E. Yang, A. C. Lozano, and P. K. Ravikumar, “Elementary estimators for inverse covariance estimation across multiple classes,” Electron. J. Statist., vol. 6, no. 491–499, 2012.

[30] D. Kershaw, “Inequalities on the elements of the inverse of a certain tridiagonal matrix,” Math. Comput., vol. 24, pp. 155–158, 1970.

[31] R. Benzi, Michele, and Nader, “Decay bounds and algorithms for approximating functions of sparse matrices,” Electron. Trans. Numer. Anal., vol. 28, pp. 16–39, 2007.

[32] M. Benzi and V. Simoncini, “Decay bounds for functions of hermitian matrices with banded or Kronecker structure,” SIAM J. Matrix Anal. Appl., vol. 36, no. 3, pp. 1263–1282, 2015.

[33] J. D. Lee, Y. Sun, and J. E. Taylor, “On model selection consistency of regularized M-estimators,” Electron. J. Statist., vol. 9, no. 1, pp. 608–642, 2015.

[34] P. Zhao and B. Y. Yu, “On model selection consistency of lasso,” J. Mach. Learn. Res., vol. 7, pp. 2541–2563, 2006.

[35] M. J. Wainwright et al., “Graphical models, exponential families, and variational inference,” Found. Trends Mach. Learn., vol. 1, no. 1/2, pp. 1–305, 2008.

[36] S. Demko, W. F. Moss, and P. W. Smith, “Decay rates for inverses of band matrices,” Math. Comput., vol. 43, no. 168, pp. 491–499, 1984.

[37] K. R. Moon et al., “Visualizing structure and transitions in high-dimensional single-cell data,” Proc. Nat. Acad. Sci. USA, vol. 116, no. 3, pp. 1341–1392, 2016. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.1516336113

[38] Y. Hao et al., “Integrated analysis of multimodal single-cell data,” IEEE Trans. Inf. Theory, vol. 65, no. 5, pp. 2183–2202, May 2019.

[39] R. Foygel and M. Drton, “Extended Bayesian information criteria for sparse regression,” J. Amer. Statist. Assoc., vol. 104, no. 486, pp. 225–237, 2009.

[40] D. Hambardzumyan and G. Bergers, “Glioblastoma: Defining tumor niches,” Trends Cancer, vol. 1, no. 4, pp. 252–265, 2015.

[41] S. Kumar et al., “Intra-tumoral metabolic zonation and resultant phenotypic diversification are dictated by blood vessel proximity,” Cell Metabol., vol. 30, no. 1, pp. 201–211, 2019.

[42] E. Jung et al., “Tumor cell plasticity, heterogeneity, and resistance in crucial microenvironmental niches in glioma,” Nature Commun., vol. 12, no. 1, pp. 1–16, 2021.

[43] D. Hallac, Y. Park, S. Boyd, and J. Leskovec, “Network inference via the time-varying graphical lasso,” in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Sci., 2017, pp. 205–213.

[44] J. Ma and G. Michailidis, “Joint structural estimation of multiple graphical models,” J. Mach. Learn. Res., vol. 17, no. 166, pp. 1–48, 2016. [Online]. Available: https://jmlr.org/papers/v17/15-656.html

[45] S. Fattahi and S. Sojoudi, “Graphical lasso and thresholding: Equivalence and closed-form solutions,” J. Mach. Learn. Res., vol. 20, no. 1, pp. 364–407, 2019.

[46] R. Y. Zhang, S. Fattahi, and S. Sojoudi, “Large-scale sparse inverse covariance estimation via thresholding and max-det matrix completion,” in Proc. Int. Conf. Mach. Learn., 2018, pp. 5766–5775.

[47] S. Fattahi, R. Y. Zhang, and S. Sojoudi, “Sparse inverse covariance estimation for chordal structures,” in Proc. IEEE Eur. Control Conf., 2018, pp. 837–844.

[48] S. Fattahi, R. Y. Zhang, and S. Sojoudi, “Linear-time algorithm for learning large-scale sparse graphical models,” IEEE Access, vol. 7, pp. 6528–6572, 2019.

[49] S. Fattahi and S. Sojoudi, “Closed-form solution and sparsity path for inverse covariance estimation problem,” in Proc. Annu. Amer. Control Conf., 2018, pp. 410–415.

[50] M. J. Wainwright et al., “Graphical models, exponential families, and variational inference,” Found. Trends Mach. Learn., vol. 1, no. 1/2, pp. 1–305, 2008.

[51] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” Biostatistics, vol. 9, no. 3, pp. 432–441, 2008.

[52] J. M. Heddleston, M. Venere, and J. N. Rich, “Deadly teamwork: Neural cancer stem cells and the tumor microenvironment,” Cell Stem Cell, vol. 8, no. 5, pp. 482–485, 2011.
Viswesvaran Ravikumar received the bachelor’s degree in biology from the Indian Institute of Science, Bangalore, and the master’s degree in bioinformatics from the MD Anderson Cancer Center, Houston, USA. He is currently working toward the PhD degree with the Department of Computational Medicine and Bioinformatics, University of Michigan. His research interest is in cancer biology and regulatory genomics.

Tong Xu received the bachelor’s degree in financial engineering from East China Normal University, in 2016, and the master’s degree from the Quantitative Finance and Risk Management program, University of Michigan. He is currently working toward the PhD degree with the Department of Industrial Engineering and Management Sciences, Northwestern University. His research interest is in graphical model inference, Mixed-integer optimization.

Wajid N. Al-Holou is a clinical assistant professor with the Department of Neurological Surgery, the University of Michigan. He specializes in the management of complex brain tumors, primarily focusing on gliomas and metastases to the brain. His research interest focuses primarily on understanding the genetic mechanisms of recurrence in glioblastoma, and in identifying precision medicine therapies to target resistant tumors. He completed his Medical education and 7-year Neurosurgery Residency with the University of Michigan Medical School. Following his residency, he completed a Complex Neurosurgical Oncology Fellowship with the MD Anderson Cancer Center in Houston, TX.

Salar Fattahi received the PhD degree from the University of California, Berkeley. He is an assistant professor with the Department of Industrial and Operations Engineering, University of Michigan. His research focuses on developing efficient and scalable computational methods for structured problems. In particular, he exploits inherent structures in optimization and machine learning problems, such as sparsity, low-rankness, and benign landscape, to develop computational techniques that scale gracefully and enjoy strong guarantees.

Arvind Rao received the PhD degree in electrical engineering and bioinformatics from the University of Michigan, specializing in transcriptional genomics. He is an associate professor with the Department of Computational Medicine and Bioinformatics, the University of Michigan. His group uses image analysis and machine learning methods to link image-derived phenotypes with genetic data, across biological scale (i.e., single cell, tissue, and radiology data). Such methods have found application in radiogenomics, spatial biology and drug repurposing based on phenotypic screens. He was a Lane postdoctoral fellow with Carnegie Mellon University, specializing in bioimage informatics.