

Research Article

Learning Latent Variable Gaussian Graphical Model for Biomolecular Network with Low Sample Complexity

Yanbo Wang, Quan Liu, and Bo Yuan

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Bo Yuan; boyuan@sjtu.edu.cn

Received 2 July 2016; Accepted 5 September 2016

Academic Editor: Ezequiel López-Rubio

Copyright © 2016 Yanbo Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Learning a Gaussian graphical model with latent variables is ill posed when there is insufficient sample complexity, thus having to be appropriately regularized. A common choice is convex ℓ_1 plus nuclear norm to regularize the searching process. However, the best estimator performance is not always achieved with these additive convex regularizations, especially when the sample complexity is low. In this paper, we consider a concave additive regularization which does not require the strong irrepresentable condition. We use concave regularization to correct the intrinsic estimation biases from Lasso and nuclear penalty as well. We establish the proximity operators for our concave regularizations, respectively, which induces sparsity and low rankness. In addition, we extend our method to also allow the decomposition of fused structure-sparsity plus low rankness, providing a powerful tool for models with temporal information. Specifically, we develop a nontrivial modified alternating direction method of multipliers with at least local convergence. Finally, we use both synthetic and real data to validate the excellence of our method. In the application of reconstructing two-stage cancer networks, “the Warburg effect” can be revealed directly.

1. Introduction

Learning a graphical model from high-dimensional but partial observations is ill posed, leading to infinitely numerous solutions. A possible approach to address this underdetermined problem is to impose a low complexity solution with a low-dimensional structure (geometry), such as the sparse vector [1], the low-rank matrix [2], and their combinations (the sparse and low-rank decomposition) [3].

One feasible way to learn such a graphical model is to capture any conditional independence between each pair of the variables with a sparsity prior. Under the assumption of multivariate normal distribution, this reconstruction can be simplified as an inverse of the covariance matrix through a penalized optimization plus a sparsity-induced regularization (Gaussian graphical model) as [4]

$$\widehat{\Omega} = \arg \min_{\Omega \geq 0} \text{Tr}(S\Omega) - \log \det \Omega + \lambda |\Omega|_1, \quad (1)$$

where S is the covariance matrix of the data, $\Omega = S^{-1}$ represents its inverse, and λ denotes the tuning parameter for the sparsity-induced regularization.

Unfortunately, there is the possibility that a few of the variables are hidden or unobserved, thus requiring a latent model. Imagine a complex network with a few latent variables, each densely interacting with multiple observed variables. Thus, the sparsity assumption will not hold because of this latent structure (Figure 1). For instance, transcriptional factors (proteins) which regulate RNA transcriptions are not directly observed from whole-genome expressions (genechip or microarray). Therefore, an additive regularization (sparse plus low-rank recovery) has been developed to decompose the sparse interactions among the observed variables (sparsity) from a few latent variables (low rankness), that is, latent variables Gaussian graphical model (LVGGM) [5] as

$$\begin{aligned} [\widehat{\Omega}_X, \widehat{L}] = \arg \min_{\Omega_X - L > 0, L \geq 0} & \text{Tr}(S_X(\Omega_X - L)) \\ & - \log \det(\Omega_X - L) + \lambda_1 |\Omega_X|_1 + \lambda_2 \|L\|_*, \end{aligned} \quad (2)$$

where S_X is the covariance matrix of the observed data, Ω_X is the inverse of S , and $\Omega_X - L$ is the surpl component of Ω_X , corresponding to the latent variables L with low rank. λ_1 and λ_2 are the tuning parameters for sparsity and low rankness,

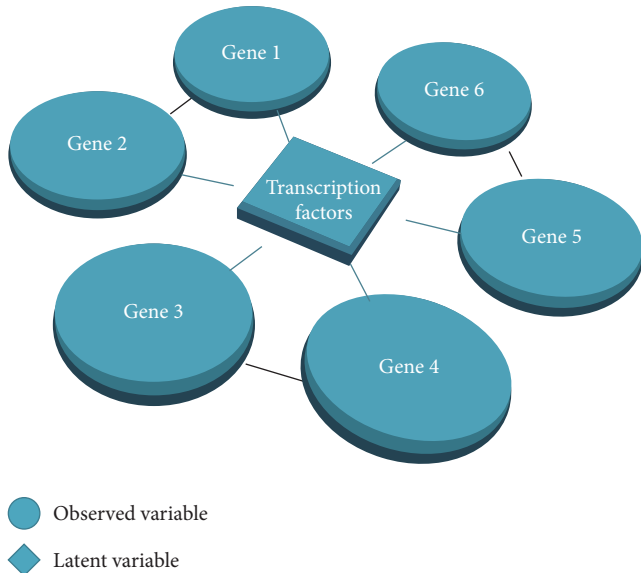


FIGURE 1: A network with a few latent variables. The edges between a pair of nodes represent that the two nodes are independent condition on remaining network.

respectively. The details of (1) and (2) will be depicted in Section 2.

An ultimate approach for such a sparse and low-rank recovery would be ℓ_0 (base pursuit) plus the selection of exact ranks and thus a NP-hard problem [6]. A common relaxation is the use of ℓ_1 plus the nuclear norm. An alternative relaxation is the use of concave penalties (Table 1, such as MCP, log or exponential type penalty, and SCAD), which have been verified to be very effective including biomolecular network reconstruction and image denoising [7, 8]. However, this concave approach so far has only been applied in sparse estimation problems.

The overall goal of this paper is therefore to develop a computational framework for a concave regularization with additive sparse and low-rank constraints [9] because of our desire to encode latent variables in a Gaussian graphical model but with insufficient samples, a very important issue in gene interaction network with latent regulatory factors. Below we want to justify why we chose a concave approach.

We chose an additive concave regularization because it does not require the strong irrepresentable condition, particularly when sample complexities are not sufficient [10]. This irrepresentable condition is highly relevant for biological observations ($p \gg n$ and often with very limited n). In contrast, a strong irrepresentable condition is necessary for Lasso in order for it to be selection consistent (requiring more sample complexities).

We chose a concave regularization also because of its parameter-estimation consistency, able to correct the intrinsic estimation biases from Lasso [11]. Here the bias is defined as the inevitable shrinkage introduced by Lasso, linearly expanded along with the tuning parameter λ . This bias issue has been noted during a regression setting [12] and a sparse precision matrix estimation [7].

Finally, we chose a concave regularization because an error bound has already been established for sparse least square problems [10]. C.-H. Zhang and T. Zhang established a bound by imposing an appropriate ℓ_2 regularity condition such that a family of column-normalized matrices can guarantee a desirable estimation under an appropriate sparsity assumption [10], leading to the error bound that is no worse than Lasso. This general result holds for the entire concave regularization family including the bridge penalties (ℓ_q , $q < 1$). Note that both of our regularizations belong to the family of bridge penalties.

As aforementioned we have intentionally selected a regularization scheme from the general concave regularization family because of its established theoretical supremacy (improve the variable selection accuracy and gain the oracle properties by reducing the bias of Lasso) [12, 13]. Our motivation is to provide the community with at least a computational alternative when some additive concave penalties are critically needed for some niches of timely applications (really demanding the oracle property), particularly when observations are limited (such as gene expression arrays).

Overall, the first contribution of our effort is having provided a novel bridge-nuclear penalty to induce a low-rank structure as well as a bridge-fused penalty to induce a fused-structural sparsity [14]. Note that we have explicitly derived the proximity operators [15] for these penalties (concave), respectively, an essential and important step towards a gradient-based optimization [16]. The structural sparsity here is used to join multiple graphical models to compare their differences (evolution of network structures, different stages and types of tumorigenesis, and other network comparison problems).

Our second contribution is to provide a modified alternative direction multiplier method (modified ADMM) [17] to numerically optimize these concave estimators (via their proximity operators), leading to at least some local solutions. We chose ADMM as the optimization method because we can prove its local convergence. We note the convergence of ADMM when applied to convex LVGGM as has been proved by [18].

Our third contribution is having provided a vigorous proof for the local convergence of this ADMM based on a framework for analyzing linear constrained optimization algorithms [19]. We use variational inequality to derive the contraction property in each iteration, which guarantees the monotonic convergence to a stationary point. Our experiments using both synthetic and real data indicated better performances compared to the classical convex regularizations. Overall we have developed an unified computational approach for additive concave regularization.

2. Latent Gaussian Graphical Model with Additive Concave Regularization

2.1. Notation. We defined our notations as follows. For n -dimensional vector $x \in \mathcal{R}^n$, for $q > 0$, we define the ℓ_q norm $\|x\|_q = (\sum_{i=1}^n |x_i|^q)^{1/q}$. Here we note the ℓ_q norm is a quasinorm for $0 < q < 1$. $I_p \in \mathcal{R}^{p \times p}$ represents the

TABLE 1: Examples of concave penalties $R(t)$.

Penalty	$R(t)$	$\frac{d}{dt}R(t)$
ℓ_0	$I(t > 0)$	None
Bridge ($0 < \alpha < 1$)	$ t ^\alpha$	$\alpha t ^{\alpha-1} \cdot \text{sign}(t)$
Capped- ℓ_1	$\min\left(\frac{\gamma}{2}, t \right)$	$I\left(t \leq \frac{\gamma}{2}\right) \cdot \text{sign}(t)$
MCP	$\int_0^{ t } \left(1 - \frac{x}{\gamma}\right)_+ dx$	$\left(1 - \frac{ t }{\gamma}\right)_+ \cdot \text{sign}(t)$
SCAD	$\int_0^{ t } 1 \wedge \left(1 - \frac{x-1}{\gamma-1}\right)_+ dx$	$1 \wedge \left(1 - \frac{ t -1}{\gamma-1}\right)_+ \cdot \text{sign}(t)$
Log-type penalty	$\frac{1}{\log(\gamma+1)} \log(\gamma t +1)$	$\frac{1}{\log(\gamma+1)} \frac{\gamma}{\gamma t +1}$
Exponential-type penalty	$\frac{1}{1 - \exp(-\gamma)} (1 - \exp(-\gamma t))$	$\frac{\gamma}{1 - \exp(-\gamma)} \exp(-\gamma t)$

We note $I(\cdot)$ is a $\{0, 1\}$ value indicator function.

identity matrix. For rectangular matrices M, N in $\mathcal{R}^{p \times q}$, the spectral norm $\|M\|$ is the largest singular value, $\|M\| = \sup_{x \in \mathcal{R}^q} (\|Mx\|/\|x\|)$, and the nuclear norm $\|M\|_*$ is the sum of the singular values. The Frobenius norm $\|M\|_F$ is the l_2 norm of the singular values; $\|M\|_F = (\text{Tr}(M^T M))^{1/2}$. The l_∞ norm is defined by $\|M\|_\infty = \max_{i,j} M_{ij}$. The Hadamard product of $M \circ N$ is defined as $[M \circ N]_{ij} = [M]_{ij} [N]_{ij}$. The function $\text{sign}(x)$ extracts the sign of a real number x .

2.2. Gaussian Graphical Model with Concave Regularization.

In this section, we briefly review the related works on Gaussian graphical model (GGM) and latent variable Gaussian graphical model (LVGGM).

A GGM also known as a Gaussian-Markov random field-based method is defined with respect to a graph $G(V, E)$. The set of nodes V consists of p individual variables (features) with n observations $X = (x_1, x_2, \dots, x_p)' \in \mathcal{R}^{p \times n}$ under the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. The edges represent the conditional independencies among the variables, where the edges $e_{ij} \notin E$, if and only if x_i and x_j are independent, conditioned on the remaining variables. With the Gaussian assumption, this conditional independence for any pairs of nodes x_i and x_j is equivalent to their partial correlation ρ_{ij} being zero, so that $e_{ij} \notin E \Leftrightarrow \rho_{ij} = \Sigma_{ij}^{-1} / \sqrt{\Sigma_{ii}^{-1} \Sigma_{jj}^{-1}} = 0$. It is important to note that such an equivalence requires that the random variables X be sampled from the families of distributions characterized with a semigroup property, including multivariate Gaussian, elliptical, multivariate hypergeometric, multivariate negative hypergeometric, multinomial, and Dirichlet distributions [20].

Therefore, the GGM can be learned by estimating its precision matrix, the inverse of the covariance matrix, in which each off-diagonal element represents its conditional independence. This estimation can be formulated as a sparsity-regularized optimization problem (1) with the sparsity assumption about the network structure. Note that there are a number of alternatives to formulate this sparsity-induced regularization. For instances, convex regularizations

with the group l_2 norm have been used to estimate the precision matrix [4, 21]. These regularization schemes have been extensively discussed, whose bounds have been estimated and even analytically proved [22, 23].

A more recent effort to achieve better statistical properties, including the oracle properties and unbiased estimation, has been developed by Fan et al., using a folder concave-based regularization [7] as follows:

$$\begin{aligned} \widehat{\Omega} = \arg \min_{\Omega \geq 0} & \text{Tr}(\Omega) - \log \det \Omega \\ & + \lambda \sum_{i=1}^p \sum_{j=1}^p \text{SCAD}_{\lambda, a}(|\omega_{ij}|). \end{aligned} \quad (3)$$

The SCAD is symmetric and a quadratic spline on $[0, \infty)$, whose first-order derivative is given by $\text{SCAD}'_{\lambda, a}(x) = \lambda \{I(|x| \leq \lambda) + ((a\lambda - |x|)_+ / (a-1)\lambda) I(|x| > \lambda)\}$ with a suggested value of $a = 3.7$. Model (3) has some admirable performances in practice but still a few limitations. At least it is difficult to be extended to include possible complex and additive regularization schemes, such as the combination of sparsity with low rankness or structural sparsity. Still the most challenging is nonconvex nature for all concave-based approaches, possibly leading to many local solutions. Thus, one of the goals of this paper is to extend Fan et al.'s work to more complex and even additive regularization schemes.

2.3. Latent Gaussian Graphical Model with Additive Concave Regularization.

A Gaussian graphical model can be incorporated with a few hidden variables (such as the latent Gaussian Graphical Model). Let $X \in \mathcal{R}^p$ be the observed variables and $Z \in \mathcal{R}^r$ ($r \ll p$) the latent ones. Here typically we assume the number for those latent variables is small (low rankness). Thus, (X, Z) can be jointly sampled from a multivariate normal distribution. Suppose its covariance matrix is $\Sigma_{(X, Z)} = \begin{pmatrix} \Sigma_X & \Sigma_{XZ} \\ \Sigma_{ZX} & \Sigma_Z \end{pmatrix}$ and its corresponding precision matrix is $\Omega_{(X, Z)} = \begin{pmatrix} \Omega_X & \Omega_{XZ} \\ \Omega_{ZX} & \Omega_Z \end{pmatrix}$. Hence $\Sigma_X^{-1} = \Omega_X - \Omega_{XZ} \Omega_Z^{-1} \Omega_{ZX}$ is called the Surl component of Ω_X , where Ω_X is a sparse matrix and $\Omega_{XZ} \Omega_Z^{-1} \Omega_{ZX}$ is denoted as L , being a low-rank matrix due to $r \ll p$. Chandrasekaran introduced a regularized

optimization with multiple additive terms, named as the latent variable graphical model (2) [5], and the error bound of LVGGM is proofed by Meng et al. [24]. Here λ_1 and λ_2 are two tuning parameters, which are often hard to choose. An empirical way is to use certain information criteria or conduct K -fold cross-validations.

We propose a novel bridge-nuclear penalty to induce a low-rank structure as follows.

Definition 1. For a matrix $M \in \mathcal{R}^{n \times n}$, suppose its singular value decomposition is $M = U\Sigma V^T$, with $U, \Sigma, V \in \mathcal{R}^{n \times n}$ the 1/2 quasinuclear norm of M being defined as $\|M\|_{\clubsuit} = (\sum_{i=1}^n (\sigma_i)^{1/2})^2$.

By employing the 1/2 quasinorm (bridge penalty) plus the 1/2 quasinuclear norm (bridge-nuclear penalty), we formulate our latent Gaussian graphical model with additive concave penalties as

$$\begin{aligned} [\widehat{\Omega}_X, \widehat{L}] = \arg \min_{\Omega_X - L > 0, L \geq 0} & \text{Tr}(S(\Omega_X - L)) \\ & - \log \det(\Omega_X - L) + \lambda_1 \|\Omega_X\|_{1/2}^{1/2} \\ & + \lambda_2 \|L\|_{\clubsuit}^{1/2}. \end{aligned} \quad (4)$$

Note again, our concave LVGGM does not need strong irrepresentable condition and thus can be applied to low sample complexity.

2.4. Joint Multiple Latent Gaussian Graphical Model with Additive Concave Regularization. To demonstrate the applicability of our additive approach, we purposely include a fused-structural sparsity, being used together with the aforementioned sparsity and low rankness. Here a total of three penalties are additively combined, potentially useful to model a network comparison problem with latent variables (joint latent variable Gaussian Graphical Model, JLVGGM). The evolution of biological and some other networks (such as social network) often has some invariant portion over the progression, which can thus be captured by our fused regularization over k individual snapshots. We consider this a potentially very useful approach to model a regulatory network in biology, since many gene interactions will remain invariantly imposed by their functional constraints (house-keeping, etc.).

The most commonly used constraint so far assumes that network evolution is gradual and local, representing mainly sporadic and minor structural changes, with most of the

systems remaining intact. For instance, Guo et al. developed a joint Gaussian graphical model to learn multiple snapshots, assuming their biological networks being only partially changed [25]. Recently, Danaher et al. [26] used a fused-lasso scheme to model multiple stages of tumorigenesis.

As an extension to Danaher's fused graphical model with latent variables, also as an example to demonstrate our additive strategy with a structural sparsity, we formulate a joint model with latent variables. Suppose $X_1 \in \mathcal{R}^{p \times n_1}, X_2 \in \mathcal{R}^{p \times n_2}, \dots, X_k \in \mathcal{R}^{p \times n_k}$ are independent and identically distributed from $\mathcal{N}(\mu_k, \Sigma_k)$. We formulate our joint model with latent variables as

$$\begin{aligned} [\widehat{\Omega}_{X_i}, \widehat{L}_i] = \arg \min_{\Omega_{X_i} - L_i > 0, L_i \geq 0} & \sum_{i=1}^k \left\{ \text{Tr}(S_i(\Omega_{X_i} - L_i)) \right. \\ & - \log \det(\Omega_{X_i} - L_i) \left. \right\} + \sum_{i=1}^k \left\{ \lambda_{1i} \|\Omega_{X_i}\|_{1/2}^{1/2} \right. \\ & \left. + \lambda_{2i} \|L_i\|_{\clubsuit}^{1/2} \right\} + \sum_{i < j}^{k-1} \lambda_{3i} \|\Omega_{X_i} - \Omega_{X_j}\|_{1/2}^{1/2}, \end{aligned} \quad (5)$$

where $\lambda_{3i}, i = 1, 2, \dots, k-1$ represent the similarity measure among the temporal networks according to the tuning parameters.

3. Modified Alternating Direction Method of Multipliers

In this section, we want to establish the algorithm and its convergence of the modified alternating direction method of multipliers (ADMM). We applied this numerical method to our graphical model with latent variables. First we derived the proximity operators individually for $\ell_{1/2}$, the 1/2 quasinuclear norm $\|\cdot\|_{\clubsuit}$ (bridge-nuclear penalty), and the fused $\ell_{1/2}$. With these proximity operators, we design a gradient-based but nonsmooth optimization based on alternating direction method of multipliers.

3.1. Proximity Operator

Theorem 2. *The proximity operator of $\|x\|_{1/2}^{1/2}$ is the global solution for the following problem:*

$$\text{Prox}_{\lambda \|x\|_{1/2}^{1/2}}(y) = \arg \min_x \left\{ \frac{1}{2} \|x - y\|_2^2 + \lambda \|x\|_{1/2}^{1/2} \right\}. \quad (6)$$

One has

$$\text{Prox}_{\lambda \|x\|_{1/2}^{1/2}}(y) = \begin{cases} \frac{2}{3}y \left(1 + \cos \left(\frac{2\pi}{3} - \frac{2}{3} \arccos \left(\frac{\lambda}{4} \left(\frac{|y|}{3} \right)^{-3/2} \right) \right) \right), & y > \frac{3}{2} \sqrt[3]{\lambda^2} \\ 0, & |y| \leq \frac{3}{2} \sqrt[3]{\lambda^2} \\ \frac{2}{3}y \left(1 + \cos \left(\frac{2\pi}{3} - \frac{2}{3} \arccos \left(\frac{\lambda}{4} \left(\frac{|y|}{3} \right)^{-3/2} \right) \right) \right), & y < -\frac{3}{2} \sqrt[3]{\lambda^2}. \end{cases} \quad (7)$$

Proof. Since the subdifferential of a nonconvex function is not well defined, we resolve the optimization problem (6) as follows.

If $\text{Prox}_{\lambda\|x\|^{1/2}}(y) = 0$ and $x > 0$, it implies that $(1/2)x^2 - xy + \lambda\sqrt{x} \geq 0$, which yields $y \leq x/2 + \lambda/\sqrt{x} \leq (3/2)\sqrt[3]{\lambda^2}$. By symmetry, we have $y \geq -(3/2)\sqrt[3]{\lambda^2}$ in the case $\text{Prox}_{\lambda\|x\|^{1/2}}(y) = 0$ and $x < 0$.

Otherwise, when $\text{Prox}_{\lambda\|x\|^{1/2}}(y) \neq 0$, the global solution of (6) is one of the three roots for the following algebraic equation: $(1/\sqrt{|x|})(\sqrt{|x|}^3 - y\sqrt{|x|} + \lambda(\text{sign}(x)/2)) = 0$, which is derived by taking the derivative on both sides of (6). With similar calculations like [27], the equation

$$\sqrt{|x|}^3 - y\sqrt{|x|} + \lambda \frac{\text{sign}(x)}{2} = 0 \quad (8)$$

has three roots in a compact trigonometric form as

$$\sqrt{|x|} = 2\sqrt{\frac{y}{3}} \cos \frac{\pi + 2k\pi - \arccos\left((\lambda/4)(y/3)^{-3/2}\right)}{3}, \quad (9)$$

$$k = 0, 1, 2,$$

under the conditions $x > 0$ and $y > (3/2)\sqrt[3]{\lambda^2/2}$. It is validated that (6) will reach a global minimum, when $k = 0$. Since x and y are either positive or negative simultaneously, we have

$$\sqrt{|x|} = 2\sqrt{\frac{-y}{3}} \cos \frac{\pi - \arccos\left((\lambda/4)(-y/3)^{-3/2}\right)}{3}, \quad (10)$$

$$y > \frac{3}{2}\sqrt[3]{\frac{\lambda^2}{2}}.$$

Similarly we have

$$\sqrt{|x|} = 2\sqrt{\frac{-y}{3}} \cos \frac{\pi - \arccos\left((\lambda/4)(-y/3)^{-3/2}\right)}{3}, \quad (11)$$

$$y < -\frac{3}{2}\sqrt[3]{\frac{\lambda^2}{2}}.$$

By taking a square root on both sides of (10) and (11) within the domain $[(-\infty, -(3/2)\sqrt[3]{\lambda^2}) \cap (-\infty, -(3/2)\sqrt[3]{\lambda^2/2})] \cup [(3/2)\sqrt[3]{\lambda^2}, +\infty) \cap ((3/2)\sqrt[3]{\lambda^2/2}, +\infty)]$, the statement in Theorem 2 is proven. \square

Theorem 3. Assuming $X, Y \in \mathcal{R}^{n \times n}$, the proximity operator of the 1/2 quasinuclear norm is given as the global minimum of

$$\text{Prox}_{\lambda\|X\|^{1/2}}(Y) = \arg \min_X \left\{ \frac{1}{2} \|X - Y\|_F^2 + \lambda \|X\|^{1/2} \right\}, \quad (12)$$

where $d(X), d(Y) \in \mathcal{R}^{n \times 1}$ represent the single value of X and Y , respectively, in nonincreasing order. U_Y and V_Y are the left and right orthogonal matrices with the singular value decomposition of $Y = U_Y \text{Diag}(d(Y)) V_Y'$. We have

$$\text{Prox}_{\lambda\|X\|^{1/2}}(Y) = U_Y \text{Diag}([t_1, t_2, \dots, t_n]) V_Y', \quad (13)$$

with $t_i = (2/3)\sigma_i(1 + \cos(2\pi/3 - (2/3)\arccos((\lambda/4)(|\sigma_i|/3)^{-3/2}))$), if $\sigma_i > (3/2)\sqrt[3]{\lambda^2}$ and $t_i = 0$, for $\sigma_i \leq (3/2)\sqrt[3]{\lambda^2}$ else.

Proof. Through von Neumann's trace inequality [28], we have

$$\begin{aligned} \|X - Y\|_F^2 &= \text{Tr}(X - Y)(X - Y)' \\ &= \text{Tr}(XX') - 2\text{Tr}(XY') + \text{Tr}(YY') \\ &= \sum_{i=1}^n d_i^2(X) - 2\text{Tr}(XY') + \sum_{i=1}^n d_i^2(Y) \\ &\geq \sum_{i=1}^n (d_i^2(X) - 2d_i(X)d_i(Y) + d_i^2(Y)). \end{aligned} \quad (14)$$

The equality holds if and only if $Y = U_Y \text{Diag}(d(X)) V_Y^T$. Then the optimization is reduced to

$$\begin{aligned} \arg \min_{d_i(X) > 0} \frac{1}{2} d_i^2(X) - d_i(X)d_i(Y) + \frac{1}{2} d_i^2(Y) \\ + \lambda \sqrt{d_i(X)}. \end{aligned} \quad (15)$$

We note here (15) is just a special case of (6). This completes the proof of Theorem 3. We note this regularization would induce a low-rank approximation of Y due to the threshold of t_i . \square

Theorem 4. The proximity operator of fused $\ell_{1/2}$ regularization is given as the global minimum of

$$\begin{aligned} \text{Prox}_{\lambda\|X_{ij}^{(2)} - X_{ij}^{(1)}\|^{1/2}}(A_{ij}^{(1)}, A_{ij}^{(2)}) \\ = \arg \min_{X_{ij}^{(1)}, X_{ij}^{(2)}} \frac{1}{2} \sum_{k=1,2} \|X_{ij}^{(k)} - A_{ij}^{(k)}\|_F^2 \\ + \lambda \|X_{ij}^{(2)} - X_{ij}^{(1)}\|_{1/2}. \end{aligned} \quad (16)$$

One has

$$\text{Prox}_{\lambda \|X_{ij}^{(2)} - X_{ij}^{(1)}\|^{1/2}}(A_{ij}^{(1)}, A_{ij}^{(2)}) = \begin{cases} \left(\frac{A_{ij}^{(1)} + A_{ij}^{(2)} - K_{ij}}{2}, \frac{A_{ij}^{(1)} + A_{ij}^{(2)} + K_{ij}}{2} \right), & A_{ij}^{(1)} - A_{ij}^{(2)} > \frac{3}{2} \sqrt[3]{\lambda^2}, \\ \left(\frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{2}, \frac{A_{ij}^{(1)} + A_{ij}^{(2)}}{2} \right), & |A_{ij}^{(1)} - A_{ij}^{(2)}| \leq \frac{3}{2} \sqrt[3]{\lambda^2}, \\ \left(\frac{A_{ij}^{(1)} + A_{ij}^{(2)} + K_{ij}}{2}, \frac{A_{ij}^{(1)} + A_{ij}^{(2)} - K_{ij}}{2} \right), & A_{ij}^{(1)} - A_{ij}^{(2)} < -\frac{3}{2} \sqrt[3]{\lambda^2}, \end{cases} \quad (17)$$

with $K_{ij} = (2/3)(A_{ij}^{(1)} - A_{ij}^{(2)})(1 + \cos(2\pi/3 - (2/3) \arccos((\lambda/4)(|A_{ij}^{(1)} - A_{ij}^{(2)}|/3)^{-3/2})))$.

Proof. If $X_{ij}^{(1)} = X_{ij}^{(2)}$, this yields $X_{ij}^{(1)} = X_{ij}^{(2)} = (A_{ij}^{(1)} + A_{ij}^{(2)})/2$. If $X_{ij}^{(1)} < X_{ij}^{(2)}$; by taking derivation, we have

$$\begin{aligned} X_{ij}^{(1)} - A_{ij}^{(1)} - \lambda \frac{\text{sign}(X_{ij}^{(2)} - X_{ij}^{(1)})}{2\sqrt{|X_{ij}^{(2)} - X_{ij}^{(1)}|}} &= 0, \\ X_{ij}^{(2)} - A_{ij}^{(2)} + \lambda \frac{\text{sign}(X_{ij}^{(2)} - X_{ij}^{(1)})}{2\sqrt{|X_{ij}^{(2)} - X_{ij}^{(1)}|}} &= 0. \end{aligned} \quad (18)$$

Denoting $X_{ij}^{(2)} - X_{ij}^{(1)} = K_{ij} > 0$, we have

$$K_{ij} + (A_{ij}^{(1)} - A_{ij}^{(2)}) + \frac{\lambda \text{sign}(K_{ij})}{\sqrt{|K_{ij}|}} = 0. \quad (19)$$

We note here that (19) is just the same form as (8).

Therefore, we have $X_{ij}^{(2)} = (A_{ij}^{(1)} + A_{ij}^{(2)})/2 + K_{ij}/2$ and $X_{ij}^{(1)} = (A_{ij}^{(1)} + A_{ij}^{(2)})/2 - K_{ij}/2$. By taking a similar calculate when $X_{ij}^{(1)} > X_{ij}^{(2)}$, we obtain our proposition. \square

4. Modified Alternating Direction Method of Multipliers

In this section, we provide an algorithm to approach the local solution of additive concave regularization problem (2). The Lagrangian of (2) is

$$\begin{aligned} \mathcal{L} = & \text{Tr}(SX) - \log \det X - \langle Y, (\Omega_X - L - X) \rangle \\ & + \lambda_1 \|\Omega_X\|_{1/2}^{1/2} + \lambda_2 \|L\|_{\clubsuit}^{1/2} + \frac{\rho}{2} \|\Omega_X - L - X\|_F^2, \end{aligned} \quad (20)$$

where $Y \in \mathcal{R}^{p \times p}$ represents the corresponding dual variable. We propose a modified ADMM discretization to optimize the Lagrangian as Algorithm 5 with at least local convergence.

Before we prove the convergence result, we need to prove the following contraction property which is the key for the proof of the convergence of general ADMM [29].

Algorithm 5 (modified ADMM).

- (1) Initialize X, Ω_X, L, Y parameters $\rho, \lambda_1, \lambda_2, \alpha_1 = 1$
- for $i = 1, 2, \dots, k$ until convergence
 - (2) $\omega^{(k)} = (X^{(k)}, \Omega_X^{(k)}, L^{(k)}, Y^{(k)})^T$
 - (3)
 - (a) $\widehat{X}^{(k+1)} = \arg \min_{X>0} \text{Tr}(SX) - \log \det X - \langle Y^{(k)}, \Omega_X^{(k)} - L^{(k)} - X \rangle + (\rho/2) \|\Omega_X^{(k)} - L^{(k)} - X\|_F^2;$
 - (b) $\widehat{\Omega}_X^{(k+1)} = \arg \min_{\Omega_X} \lambda_1 \|\Omega_X\|_{1/2}^{1/2} - \langle Y^{(k)}, \Omega_X - L^{(k)} - X^{(k)} \rangle + (\rho/2) \|\Omega_X - L^{(k)} - X^{(k)}\|_F^2;$
 - (c) $\widehat{L}^{(k+1)} = \arg \min_{L>0} \lambda_2 \|L\|_{\clubsuit}^{1/2} - \langle Y^{(k)}, \Omega_X^{(k)} - L - X^{(k)} \rangle + (\rho/2) \|\Omega_X^{(k)} - L - X^{(k)}\|_F^2;$
 - (4) $\widehat{Y}^{(k+1)} = Y^{(k)} - \rho(\widehat{\Omega}_X^{(k+1)} - \widehat{L}^{(k+1)} - \widehat{X}^{(k+1)})$, $\widehat{\omega} = (\widehat{X}^{(k+1)}, \widehat{\Omega}_X^{(k+1)}, \widehat{L}^{(k+1)}, \widehat{Y}^{(k+1)})^T$
 - (5) $\alpha_{k+1} = (1 + \sqrt{1 + 4\alpha_k^2})/2$ and $\gamma = (\alpha_k - 1)/\alpha_{k+1}$.
 - (6) $\omega^{(k+1)} = \omega^{(k)} - \gamma(\omega^{(k)} - \widehat{\omega})$.

end

Theorem 6. Assume that $\omega^* = (X^*, \Omega_X^*, L^*, Y^*)^T$ is a global optimal solution of (2) and $\omega^{(k)} = (X^{(k)}, \Omega_X^{(k)}, L^{(k)}, Y^{(k)})^T$ is optimized by Algorithm 5. Thus the contradicted property holds; that is,

$$\begin{aligned} \|\omega^{(k+1)} - \omega^*\|_F^2 &\leq \|\omega^{(k)} - \omega^*\|_F^2 \\ &- \frac{(\gamma - 2 + \sqrt{3})}{\gamma} \|\omega^{(k)} - \omega^{(k+1)}\|_F^2. \end{aligned} \quad (21)$$

Proof. For any $X, \Omega_X, L, Y \in \mathcal{R}^{P \times P}$, the optimality conditions (20) are

$$\begin{aligned}
& \text{Tr } SX - \text{Tr } S\widehat{X}^{(k+1)} - \log \det X + \log \det \widehat{X}^{(k+1)} \\
& - \langle Y^{(k)}, \Omega_X^{(k)} - L^{(k)} - X \rangle \\
& + \langle Y^{(k)}, \Omega_X^{(k)} - L^{(k)} - \widehat{X}^{(k+1)} \rangle \\
& - \rho \langle X - \widehat{X}^{(k+1)}, (\Omega_X^{(k)} - L^{(k)} - \widehat{X}^{(k+1)}) \rangle \geq 0, \\
& \lambda_1 \|\Omega_X\|_{1/2}^{1/2} - \lambda_1 \|\widehat{\Omega}_X^{(k+1)}\|_{1/2}^{1/2} \\
& - \langle Y^{(k)}, \Omega_X - L^{(k)} - X^{(k)} \rangle \\
& + \langle Y^{(k)}, \widehat{\Omega}_X^{(k+1)} - L^{(k)} - X^{(k)} \rangle \\
& + \rho \langle \Omega_X - \widehat{\Omega}_X^{(k+1)}, (\widehat{\Omega}_X^{(k+1)} - L^{(k)} - X^{(k)}) \rangle \geq 0, \\
& \lambda_1 \|L\|_{1/2}^{1/2} - \lambda_1 \|\widehat{L}^{(k+1)}\|_{1/2}^{1/2} - \langle Y^{(k)}, \Omega_X^{(k)} - L - X^{(k)} \rangle \\
& + \langle Y^{(k)}, \Omega_X^{(k)} - \widehat{L}^{(k+1)} - X^{(k)} \rangle \\
& - \rho \langle L - \widehat{L}^{(k+1)}, (\Omega_X^{(k)} - \widehat{L}^{(k+1)} - X^{(k)}) \rangle \geq 0.
\end{aligned} \tag{22}$$

By denoting $T = \rho(\widehat{\Omega}_X^{(k+1)} - \Omega_X^{(k)}) - \rho(\widehat{L}^{(k+1)} - L^{(k)}) - \rho(\widehat{X}^{(k+1)} - X^{(k)})$,

$$\begin{aligned}
F(\widehat{\omega}) &= \begin{pmatrix} \widehat{Y}^{(k+1)} \\ -\widehat{Y}^{(k+1)} \\ \widehat{Y}^{(k+1)} \\ \rho^2 (\widehat{\Omega}_X^{(k+1)} - \widehat{X}^{(k+1)} - \widehat{L}^{(k+1)}) \end{pmatrix}, \\
\mathcal{O}_1 &= \begin{pmatrix} \rho (\widehat{X}^{(k+1)} - X^{(k)}) \\ \rho (\widehat{\Omega}_X^{(k+1)} - \Omega_X^{(k)}) \\ \rho (\widehat{L}^{(k+1)} - L^{(k)}) \\ \rho (\widehat{Y}^{(k+1)} - Y^{(k)}) \end{pmatrix} + \begin{pmatrix} -T \\ T \\ -T \\ 0 \end{pmatrix},
\end{aligned} \tag{23}$$

we have

$$\begin{aligned}
& \text{Tr}(SX) - \text{Tr}(S\widehat{X}^{(k+1)}) + \lambda_1 \|\Omega_X\|_{1/2}^{1/2} - \lambda_1 \|\widehat{\Omega}_X^{(k+1)}\|_{1/2}^{1/2} \\
& + \lambda_2 \|L\|_{1/2}^{1/2} - \lambda_2 \|\widehat{L}^{(k+1)}\|_{1/2}^{1/2} \\
& + \langle \omega - \widehat{\omega}, F(\widehat{\omega}) + \mathcal{O}_1 \rangle \geq 0.
\end{aligned} \tag{24}$$

Without loss of generality, we let $\omega = (X, \Omega_X, L, Y)^T = \omega^*$, which yields $\rho \langle \widehat{\omega} - \omega^*, \omega^{(k)} - \widehat{\omega} \rangle \geq \langle \omega^* - \widehat{\omega}, F(\widehat{\omega}) + \mathcal{O}_1 \rangle$.

Using the optimal conditional (variational inequalities) [30], it follows that

$$\langle \omega^* - \widehat{\omega}, F(\omega^*) \rangle \geq 0. \tag{25}$$

Since $F(\omega)$ is monotone, we have $\langle \omega^* - \widehat{\omega}, F(\widehat{\omega}) \rangle \geq \langle \omega^* - \widehat{\omega}, F(\omega^*) \rangle$. Therefore, we have $\rho \langle \omega^{(k)} - \omega^*, \omega^{(k)} - \widehat{\omega} \rangle \geq \rho \|\widehat{\omega} - \omega^{(k)}\|_F^2 + \delta$, with $\delta = \langle Y^{(k)} - \widehat{Y}^{(k+1)}, \rho(X^{(k)} - \widehat{X}^{(k+1)}) + \rho(L^{(k)} - \widehat{L}^{(k+1)}) - \rho(\Omega_X^{(k)} - \widehat{\Omega}_X^{(k+1)}) \rangle$. We denote

$$H = \begin{pmatrix} I_p & 0 & 0 & \frac{1}{2}I_p \\ 0 & -I_p & 0 & \frac{1}{2}I_p \\ 0 & 0 & I_p & \frac{1}{2}I_p \\ \frac{1}{2}I_p & \frac{1}{2}I_p & \frac{1}{2}I_p & \frac{1}{2}I_p \end{pmatrix}; \tag{26}$$

thus the smallest eigenvalue is $(2 - \sqrt{3})/2$. We obtain

$$\begin{aligned}
\rho \|\widehat{\omega} - \omega^{(k)}\|_F^2 + \delta &= \rho \left\langle \begin{pmatrix} X^{(k)} - X^{(k+1)} \\ \Omega_X^{(k)} - \Omega_X^{(k+1)} \\ L^{(k)} - L^{(k+1)} \\ Y^{(k)} - Y^{(k+1)} \end{pmatrix}, \right. \\
& H \begin{pmatrix} X^{(k)} - X^{(k+1)} \\ \Omega_X^{(k)} - \Omega_X^{(k+1)} \\ L^{(k)} - L^{(k+1)} \\ Y^{(k)} - Y^{(k+1)} \end{pmatrix} \left. \right\rangle = \text{Tr} \left((\omega^{(k)} - \widehat{\omega})^T \right. \\
& \cdot H (\omega^{(k)} - \widehat{\omega}) \left. \right) \geq \frac{2 - \sqrt{3}}{2} \rho \|\omega^{(k)} - \widehat{\omega}\|_F^2.
\end{aligned} \tag{27}$$

In summary, we have

$$\begin{aligned}
& \|\widehat{\omega}^{(k+1)} - \omega^*\|_F^2 = \|\omega^{(k)} - \gamma(\omega^{(k)} - \widehat{\omega}) - \omega^*\|_F^2 \\
& = \|\omega^{(k)} - \omega^*\|_F^2 + \gamma^2 \|\omega^{(k)} - \widehat{\omega}\|_F^2 \\
& \quad - 2\gamma \langle (\omega^{(k)} - \omega^*), (\omega^{(k)} - \widehat{\omega}) \rangle \\
& \leq \|\omega^{(k)} - \omega^*\|_F^2 \\
& \quad - \gamma(\gamma - 2 + \sqrt{3}) \|\omega^{(k)} - \widehat{\omega}\|_F^2 \\
& \leq \|\omega^{(k)} - \omega^*\|_F^2 \\
& \quad - \frac{(\gamma - 2 + \sqrt{3})}{\gamma} \|\omega^{(k)} - \omega^{(k+1)}\|_F^2.
\end{aligned} \tag{28}$$

□

Theorem 7. The sequence $\omega^{(k+1)}$ produced by Algorithm 5 from a given initial value converges monotonically to an optimal solution to problem.

Proof. From Theorem 6, we can easily get the following:

- (1) $\|\omega^{(k)} - \omega^{(k+1)}\|_F \rightarrow 0$.
- (2) $\{\omega^{(k)}\}$ lies in a compact region.
- (3) $\|\omega^{(k)} - \omega^*\|_F$ is monotonically nonincreasing and thus converging.

It follows that $\omega^{(k+1)}$ is a Cauchy sequence and thus has a limit point $\bar{\omega}$. Therefore we have $\omega^{(k+1)}$ converging monotonically to $\bar{\omega}$. \square

5. Numerical Evaluation

In this paper, we focus on the effective local convergence of ill-posed graphical model problems with additive concave regularization. The specific aim is to develop an additive regularization combining both sparsity and low rankness, essential for our latent graphical model. Again we chose a concave approach (instead of the popular convex methods) because of its presumptive advantage (the oracle property) especially when the data complexity is not sufficient ($p \gg n$). It is thus important to first develop a robust numerical method to at least obtain a consistent local solution. We believe our methods are highly applicable to situations where the number of variables far exceeds their observations (large p small n problem, typical for most biological observations).

Below we use both artificial and real data to establish the effectiveness of our methods. We demonstrate that our nonconvex regularization can reduce the biases, as illustrated in its applications to the selected biological and financial problems. In addition, we assess the performance of our concave models comparing with the corresponding convex ones.

5.1. Artificial Data. Our concave graphical model is supposed to perform better as an asymptotically unbiased estimator. We generate our artificial data according to the following steps.

We generate 5% nonzero entries uniformly from a 50×50 matrix $\widehat{\Omega}$. Each nonzero element in this $\widehat{\Omega}_{ij}$ is sampled from a Gaussian distribution of $\mathcal{N}(0, \sigma^2)$. We construct a symmetric precision matrix as $\Omega = \widehat{\Omega} + \widehat{\Omega}^T$. To guarantee the positive definiteness of this Ω , we update the matrix $\Omega = \Omega + 1.1I_{50}$ iteratively, such that all of its eigenvalues be greater than zero. We randomly choose 45 columns and their corresponding rows as a subblock matrix Ω_O form Ω ; thus the remaining 10% variables are latent. Finally, we take the inverse of the precision matrix $\Sigma_O = \Omega_O^{-1}$ as its original covariance matrix to sample a Gaussian distribution.

We divide our experiments into two groups with the increasing sample complexity: $p = (1/2)N$, N random samples from the multivariate Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with $\sigma^2 = 3$. Important to note is that in order to test the effectiveness of our method for an ill-posed problem ($p \gg n$), we purposely only increase the number of observations while keeping their variances a constant.

We perform the following comparisons:

- (1) convex Gaussian graphical model with latent variables (LVGGM) [5];
- (2) concave Gaussian graphical model with latent variables (concave LVGGM).

We use this example to validate that our method has better performance than the classic LVGGM. Firstly, for parameters estimation consistence as Figure 2 shows, our concave LVGGM estimate parameters more accurately than the classic model in [5], even with a local solution. We want to note that our method performs better with the larger observed data in Figure 2(c) but is significantly superior to the classical method with inadequate data in Figure 2(a). This is very important with biological application usually with very insufficient observation.

Finally, to assess the consistence of model selection for our methods, we use the Matthews correlation coefficient (MCC $\in [-1, 1]$) to quantify recovery merits, where

$$\begin{aligned} \text{MCC} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \end{aligned} \quad (29)$$

with TP, TN, FP, and FN being the numbers of true positives, true negatives, false positives, and false negatives. The larger the MCC is, the better the reconstruction model performs. Figures 2(b) and 2(d), depicting LVGGM and concave LVGGM, respectively, are both model selection consistent.

5.2. The Biomolecular Network of Medulloblastoma. To demonstrate that our method performs consistent in both artificial data and real application, we apply our method to human medulloblastoma. Medulloblastoma is the most common form of childhood brain tumors. This cancer has at least four subgroups, including the WNT subgroups and the sonic hedgehog (SHH), plus the subgroups 3 and 4, with the molecular etiology remaining elusive for the latter two. In China, patients with medulloblastomas are still largely treated with universal and aggressive procedures, combining radical surgery, radiation, and chemotherapy, which might in fact fail the subgroups 3 and 4 (poor prognosis with unknown reasons), or probably have overtreated the WNT subgroup (more differentiated). Thus, it is important to develop a methodology to clearly distinguish and identify the key molecular markers and their interactions at system level, representing each of the subgroups decisively. We hope our results will yield a set of genes in some given structures which can be used as the signatures (biomarkers) to guide diagnosis, treatment, and prognosis in future.

The gene expression data are publicly available from the National Center for Biotechnology Information (NCBI), consisting of 73 individual cancer samples (8, 10, 16, and 39 samples for WNT, SHH, subgroup 3, and subgroup 4, resp.), each labeled with a specific subgroup (accession number: GDS4296) [31]. Firstly, We select 1146 genes out of 54676, whose expressions show significantly larger variance across

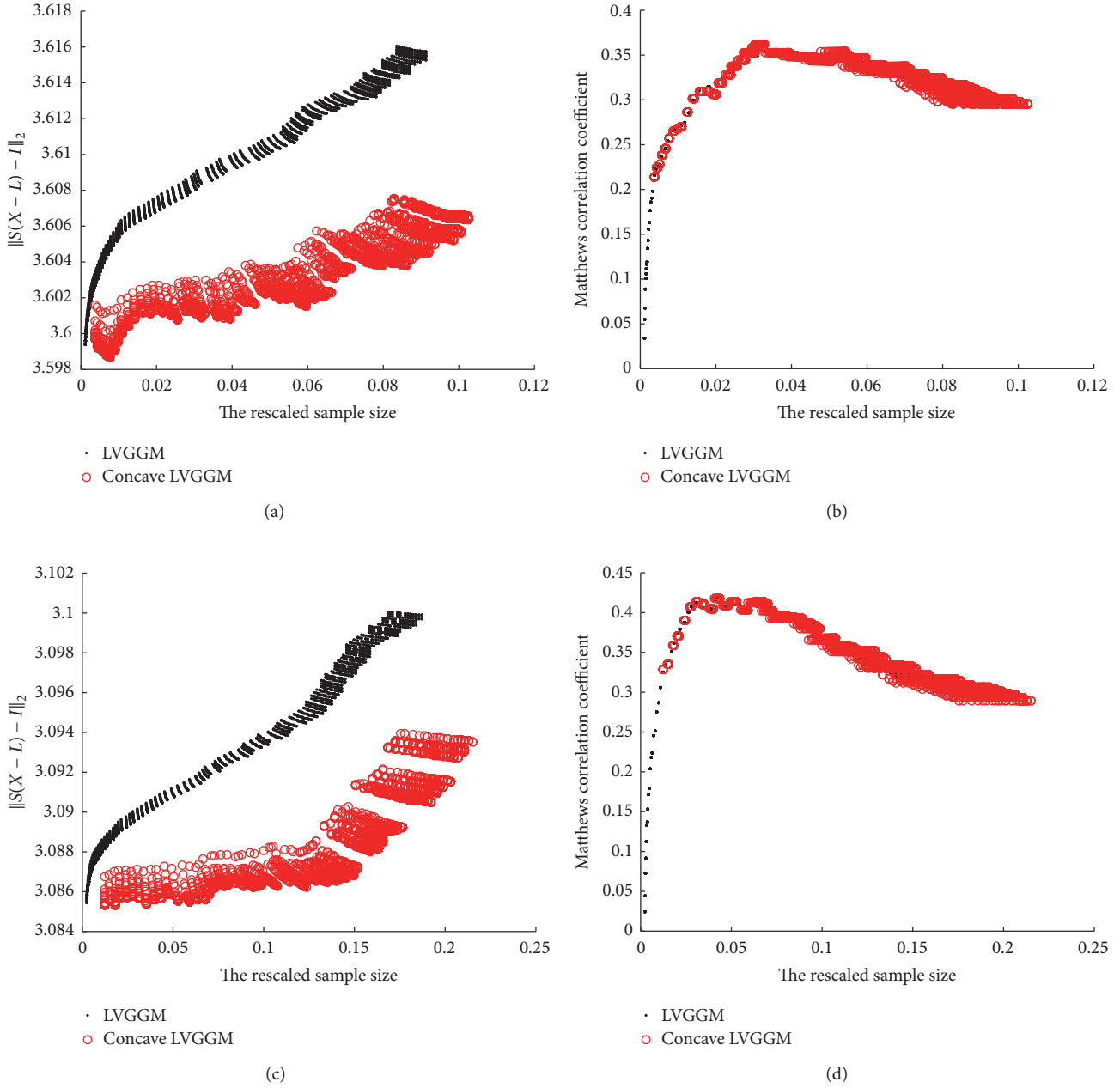


FIGURE 2: Simulation chains for the graphical model with latent variables. The estimation error $\|S(X-L) - I\|_2$ versus the rescaled sample size $n/(s \log(p) + r \log(p))$ ((a) and (c)), where s is the number of nonzero entries of X and r is the rank of L . The MCC rate versus the rescaled sample size ((b) and (d)). We note the sample complexity is $n = (1/2)p$ for (a) and (b) and $n = p$ for (c) and (d).

both of the cancer samples. In order to tune our parameters we use Bayesian information criterion (BIC):

$$\text{BIC} = -2(-\log \det(X-L) + \text{Tr}(S(X-L))) + \gamma k \log n \quad (30)$$

to select the tuning parameters λ_1, λ_2 with a constant parameter γ . Here k presents the nonzeros numbers of matrix X , while n denotes the total sample numbers. We use BIC because it is intrinsically incline to identify the “true”

model while being asymptotically consistent in selecting such a model. We did consider another criterion such as Akaike information criterion (AIC) as it tends to explain the data, thus suffering the risk of overfitting. Since the BIC does not have a closed form with respect to λ_1 and λ_2 , we carry out a grid-based screening (Figure 3).

As demonstrated in Figure 4(a), the famous WNT/beta-catenin pathway is recovered, including WNT16 and XIST. The WNT pathway is required for basic developmental processes, such as cell-fate specification, progenitor-cell

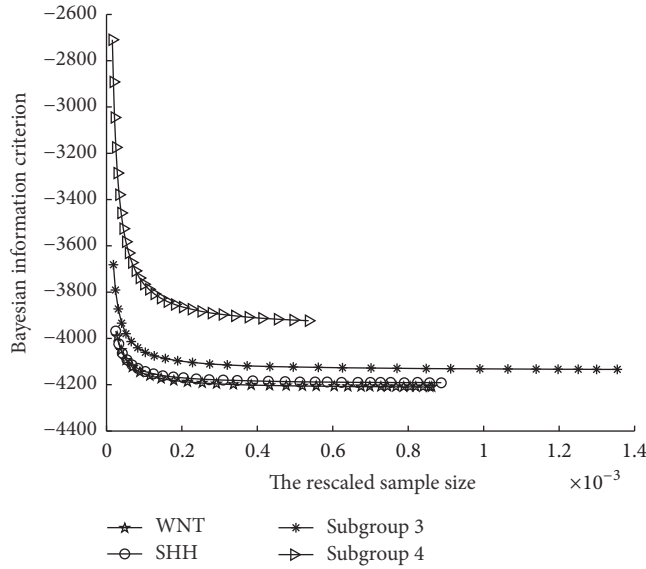


FIGURE 3: BIC scores. The BIC score for our concave LVGGM over the rescaled sample size $n/(s \log(p) + r \log(p))$. We use a grid-based search for the local minimums for (λ_1, λ_2) ; the parameters are $(\lambda_1, \lambda_2) = (0.0005, 0.001), (0.00038966, 0.001), (0.0002931, 0.001), (0.00047241, 0.001)$ for subgroups WNT, SHH, 3, and 4, respectively.

proliferation, and the control of asymmetric cell division. Identified here is the canonical pathway, which inhibits the beta-catenin degradation complex. Our results thus strongly support the etiological role of the canonical WNT signaling in the pathogenesis of this subgroup of tumors [32]. As illustrated in Figure 4(b), part of the SHH signaling pathway is identified, including oncogene MAGEA, which is associated with shorter survival of tumor cells [33]. Since SHH subgroup medulloblastomas develops from cerebellar granule neuron progenitors, which are supposed to guide axon growth into muscles, it is our speculation that cells of this subgroup of tumors might interact with muscle cells sometime during early development. Interestingly, many modules identified here including TAC1 and TTR, are implicated in some sensory-related activities during early development for subgroup 3 (Figure 4(c)). Consistent with our observations, those sensory genes are overexpressed for this subgroup of tumors according to the independent studies [32]. It is clearly distinguishable from the other three, particularly when combined with the sensory basis (see above). However, note that the genes of subgroup 4 are more randomly distributed than the other three, without the apparent modular structures. Interestingly, a number of important oncogenes and tumor suppressor genes are found such as FOXG1 (Figure 4(d)).

5.3. *The Structural Changes on Gene Regulatory Networks Occurring during the Progression of Human Lung Cancers.* To indicate our concave additive regularizations can be used to model a structured sparsity (fused sparsity). We perform the following experiment to identify the structural changes on

gene regulatory networks occurring during the progression of human lung cancers. The lung cancer dataset contains 22283 microarray-derived gene expression measurements from large airway epithelial cells sampled from 97 patients with lung cancer and 90 controls [34]. The data are publicly available from the Gene Expression Omnibus [35]; its accession number is GDS2771.

We first selected the 36 most important biological modules (CDK4, CDK6, CDK2AP2, BCL2, XIAP, BAX, CASP8, CASP9, FOXA2, HNF4A, EBP, HGF, NFKB2, STAT3, IL6, IL10, HIF1AN, MYC, GSK3B, TP53, PTEN, RBI, MDM2, PDGFRA, SOX2, PIK3CA, FGFR1, IGF1R, EPHA2, MET, EGFR, DDR2, KEAP1, KRAS, and AKT1) out of the whole genome for lung cancer according to [36]. The goal of this experiment is to find out if our method can consistently detect key structural changes on the level of gene regulatory networks associated with lung cancer. It is well established that many protein-factors are directly or indirectly engaged in the control of transcripts. Also important to note is the availability of gene annotation information, which is still limited mainly for some of the key and frequently expressed genes. We take the notion that the network structure will largely determine the key aspects of biological functionality [37].

We want to know if there are any key differences between normal and cancer cells. We use the 36 key biological modules to reconstruct a two-stage gene regulatory network by our Concave JLVGGM. To approach statistical significance, we do a random permutation to the tuning parameter λ_{11} for 1000 times. According to the second law of thermodynamics, it implies that the cancer network becomes unstable during the progress of tumorigenesis, consistent with the notion that instability is a common phenotype of cancer cells. The tuning parameters are taken as $\lambda_{12} = 0.8\lambda_{11}$. An interesting phenomenon we observed during our modeling effort is as follows: we found that the abnormal metabolism and apoptosis may be closely related to the progression of lung cancers according to Figure 5. As “the Warburg effect” suggests, most proliferation of cancer cells relies on aerobic glycolysis. Thus, the abnormal metabolism we detected in the cancer cells here is intriguing and significant. It is widely known that apoptosis plays a critical role in cancer and particularly lung cancer. However, based on our structural results (undirected graph), we do not know how these changes would be causally related.

6. Conclusion

We have developed a concave regularization approach for LVGGM for low sample complexity as well as for reducing the biases. Computational method based on proximity operators is provided with at least local convergence. Our methodology establishes its practical value as demonstrated by the numerical results (see Figure 2). Finally, as a future direction, we plan to assess a fully Bayesian interpretation of concave LVGGM. This may be helpful for an advisable choice of the tuning parameter as displayed in the framework of [38].

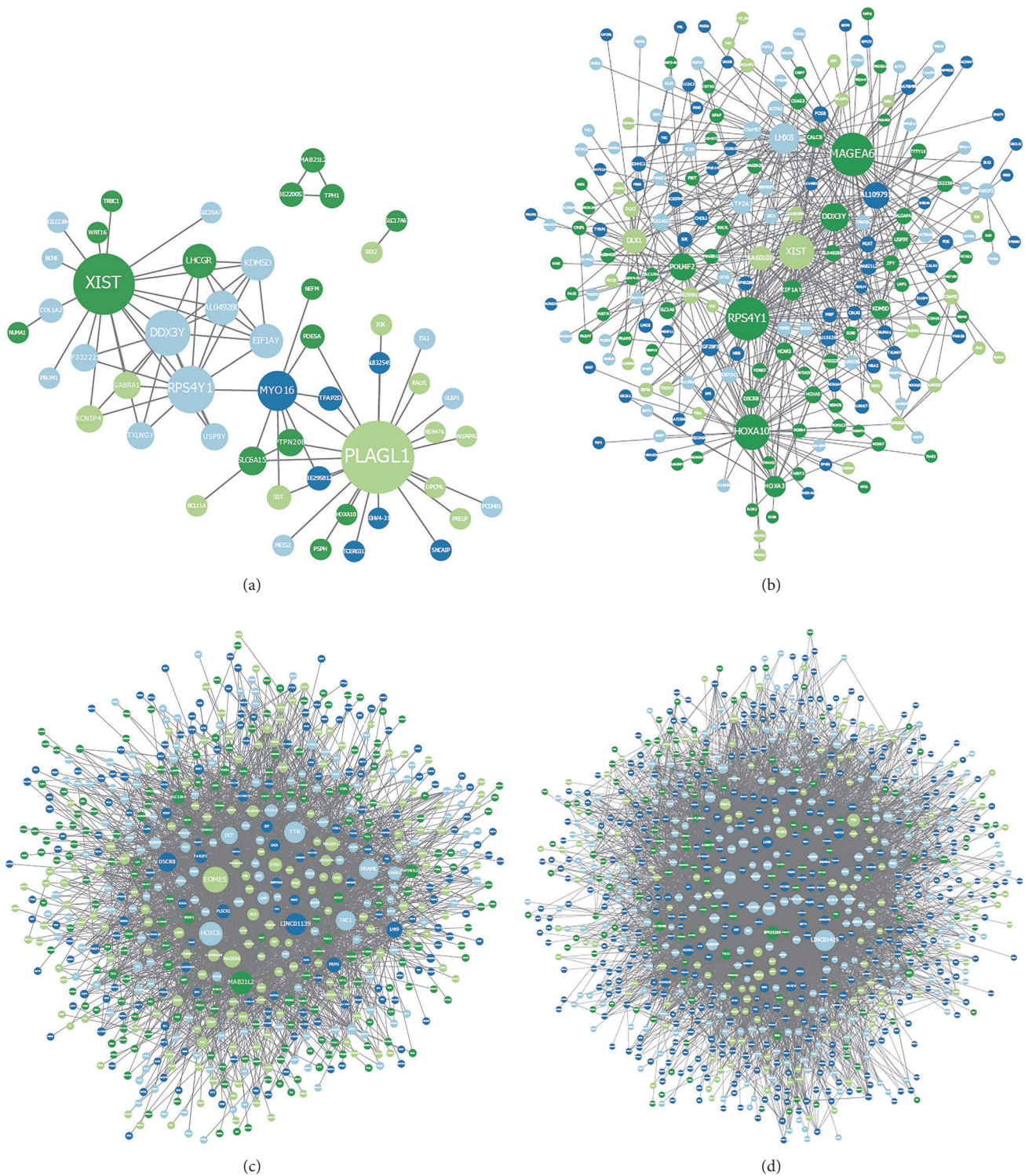


FIGURE 4: Topological for the biomolecular network of medulloblastomas. The sizes of nodes correspond to the degrees of their interactions. The colors represent similar classes according to the pairwise Pearson correlations of gene expressions. Demonstrated here is the absolute essentiality of using conditional independence to interpret the apparent correlations between gene expressions. (a) represents the WNT subgroup; (b) the SHH subgroup; (c) the subgroup 3; and (d) the subgroup 4. The regularization parameters are chosen to optimize BIC score.

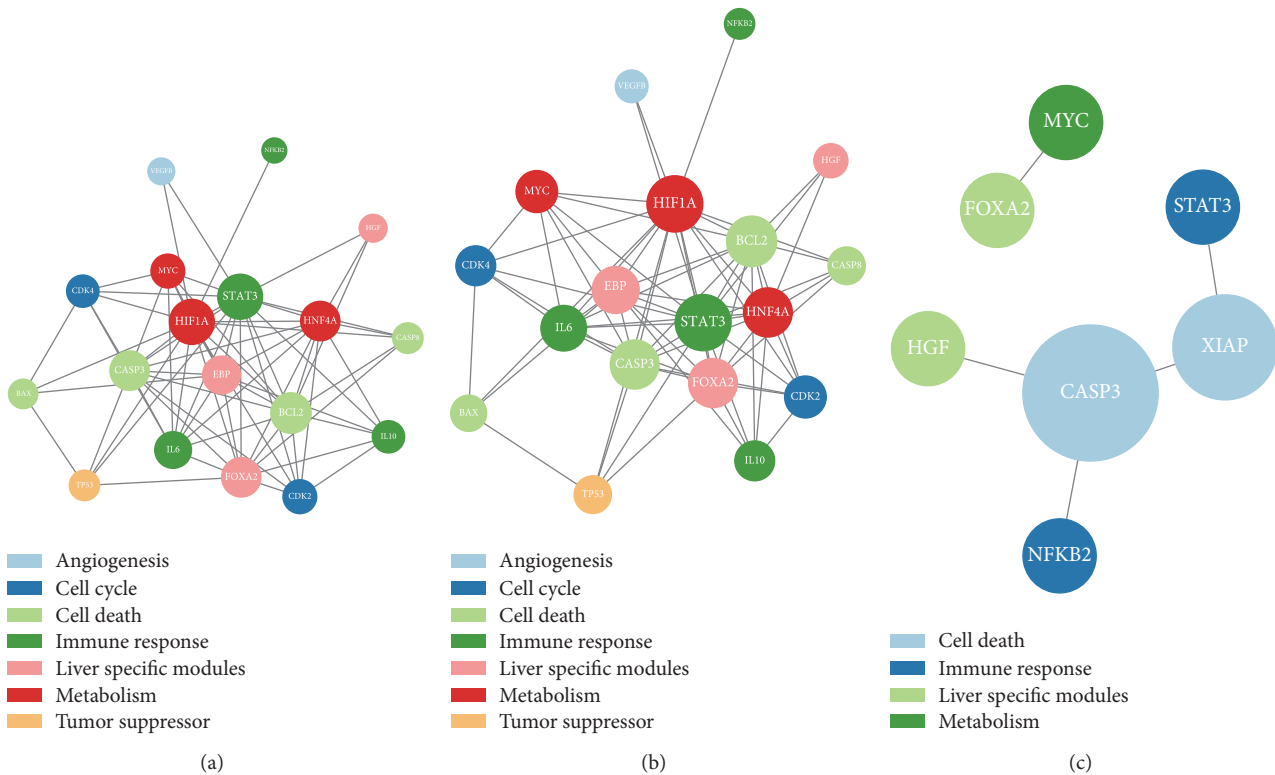


FIGURE 5: Gene regulatory networks occurring during the progression of human lung cancers. The network reconstructed in 75% confidence (the edges emerge at least 750 times) under 1000 times random permutations. The normal stage (a) and the cancer stage (b). Changes about edges between normal stage and cancer stage (c). The parameters are $\lambda_{11} = 0.003$, $\lambda_{21} = 0.006$, $\lambda_{31} = 0.001$, $\lambda_{12} = 0.8\lambda_{11}$, and $\lambda_{21} = \lambda_{22}$. The regularization parameters are manually chosen to induce high sparsity for better visualization and highlighting the dominating edges.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors acknowledge the support of the fund of the National Natural Science Foundation of China (Grant no. 14Z103010221).

References

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, article 11, 2011.
- [4] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.
- [5] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *The Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.
- [6] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–124, 2007.
- [7] J. Fan, Y. Feng, and Y. Wu, "Network exploration via the adaptive lasso and SCAD penalties," *The Annals of Applied Statistics*, vol. 3, no. 2, pp. 521–541, 2009.
- [8] Q. Xie, Q. Zhao, D. Meng et al., "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 1692–1700, 2016.
- [9] E. Yang and P. Ravikumar, "Dirty statistical models," in *Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS '13)*, pp. 611–619, December 2013.
- [10] C.-H. Zhang and T. Zhang, "A general theory of concave regularization for high-dimensional sparse estimation problems," *Statistical Science*, vol. 27, no. 4, pp. 576–593, 2012.
- [11] C.-H. Zhang and J. Huang, "The sparsity and bias of the LASSO selection in high-dimensional linear regression," *The Annals of Statistics*, vol. 36, no. 4, pp. 1567–1594, 2008.
- [12] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [13] J. Fan, L. Xue, and H. Zou, "Strong oracle optimality of folded concave penalized estimation," *The Annals of Statistics*, vol. 42, no. 3, pp. 819–849, 2014.
- [14] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 67, no. 1, pp. 91–108, 2005.

- [15] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.
- [16] Y. Nesterov and A. Nemirovski, "On first-order algorithms for l_1 /nuclear norm minimization," *Acta Numerica*, vol. 22, pp. 509–575, 2013.
- [17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [18] S. Ma, L. Xue, and H. Zou, "Alternating direction methods for latent variable Gaussian graphical model selection," *Neural Computation*, vol. 25, no. 8, pp. 2172–2198, 2013.
- [19] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.
- [20] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian and New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.
- [21] M. Yuan and Y. Lin, "Model selection and estimation in the Gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.
- [22] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing l_1 -penalized log-determinant divergence," *Electronic Journal of Statistics*, vol. 5, pp. 935–980, 2011.
- [23] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.
- [24] Z. Meng, B. Eriksson, and A. O. Hero III, "Learning latent variable Gaussian graphical models," <https://arxiv.org/abs/1406.2721>.
- [25] J. Guo, E. Levina, G. Michailidis, and J. Zhu, "Joint estimation of multiple graphical models," *Biometrika*, vol. 98, no. 1, pp. 1–15, 2011.
- [26] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 2, pp. 373–397, 2014.
- [27] Z. Xu, X. Chang, F. Xu, and H. Zhang, " $l_1/2$ regularization: a thresholding representation theory and a fast solver," *IEEE Transactions on neural networks and learning systems*, vol. 23, no. 7, pp. 1013–1027, 2012.
- [28] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.
- [29] E. X. Fang, B. He, H. Liu, and X. Yuan, "Generalized alternating direction method of multipliers: new theoretical insights and applications," *Mathematical Programming Computation*, vol. 7, no. 2, pp. 149–187, 2015.
- [30] B.-S. He, "Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities," *Computational Optimization and Applications*, vol. 42, no. 2, pp. 195–212, 2009.
- [31] G. Robinson, M. Parker, T. A. Kranenburg et al., "Novel mutations target distinct subgroups of medulloblastoma," *Nature*, vol. 487, no. 7409, pp. 43–48, 2012.
- [32] P. A. Northcott, A. Korshunov, H. Witt et al., "Medulloblastoma comprises four distinct molecular variants," *Journal of Clinical Oncology*, vol. 29, no. 11, pp. 1408–1414, 2011.
- [33] A. O. Gure, R. Chua, B. Williamson et al., "Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer," *Clinical Cancer Research*, vol. 11, no. 22, pp. 8055–8062, 2005.
- [34] A. Spira, J. E. Beane, V. Shah et al., "Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer," *Nature Medicine*, vol. 13, no. 3, pp. 361–366, 2007.
- [35] T. Barrett, T. O. Suzek, D. B. Troup et al., "NCBI GEO: mining millions of expression profiles—database and tools," *Nucleic Acids Research*, vol. 33, pp. D562–D566, 2005.
- [36] G. Wang, X. Zhu, L. Hood, and P. Ao, "From Phage lambda to human cancer: endogenous molecular-cellular network hypothesis," *Quantitative Biology*, vol. 1, no. 1, pp. 32–49, 2013.
- [37] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles, "Metabolic network structure determines key aspects of functionality and regulation," *Nature*, vol. 420, no. 6912, pp. 190–193, 2002.
- [38] N. G. Polson and J. G. Scott, "Local shrinkage rules, Lévy processes and regularized regression," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 74, no. 2, pp. 287–311, 2012.