The Sample Complexity of Differential Analysis for Networks that Obey Conservation Laws

Jiajun Cheng, Anirudh Rayas, Rajasekhar Anguluri, and Gautam Dasarathy

Department of Electrical, Computer, & Energy Engineering, Arizona State University

Abstract

Networked systems that obey conservation laws are common in many domains such as power grids, biological systems, and social networks. These systems are described by so-called balance equations that link injected flows and node potentials, ensuring that the flow at each node is balanced. For example, electric networks follow Kirchhoff's laws, while social networks model group consensus. Understanding the structure of these networks based on node potential data has become an important research topic. In this work, we focus on the problem of differential network analysis for systems that obey conservation laws. That is, instead of the structure of a network, we focus on estimating the structural differences between two networks from their node potential data. We propose a method that uses a high-dimensional estimator to directly identify these structural changes. We provide theoretical guarantees and test our method on both synthetic networks and benchmark power network data to validate its performance. The results show that our method works well but also highlight some gaps between the theoretical guarantees and experimental outcomes. Addressing these gaps is an important step for improving future methods.

1 Introduction

Consider a graph $\mathcal{G} = ([p], E)$ and let $X \in \mathbb{R}^p$ be a p-dimensional vector of *injected flows* at the vertices $[p]$. Let $Y \in \mathbb{R}^p$ be the vector of the *vertex potentials*. The flows and potentials are said to satisfy a *conservation law* with respect to the graph $\mathcal{G}$ if they obey the relationship $Y = (B^*)^{-1}X$, where B^* is an invertible Laplacian matrix associated with G [1]. At each vertex, the flows directly counteract the injections. Such an equation is called a *balance equation* between potential and inject flows. For instance, electric networks obey Kirchoff's laws and social networks reflect a consensus of views. In many real-world applications, learning the unknown structure of a network is essential for understanding and managing complex systems. This task often involves inferring the structure of graphs that obey conservation laws based on node potentials, a topic that has garnered considerable attention in recent years (see, for instance, [1, 2, 3]).

Network systems in many practical applications are not static; their structure evolves over time. Thus, one often aims to understand the difference in structures of two networks. This problem of *differential network analysis* has become crucial for several tasks across a wide range of fields, including biomedicine, social networks, and cyber-physical systems [4, 5, 6, 7]. Given the importance of this problem, we are motivated to study the estimation of the *difference matrix* $\Delta^* = B_2^* - B_1^*$ between the structures of two networks that obey conservation laws, where B_1^* and B_2^* are invertible Laplacian matrix that represent the structure of the corresponding networks. In [8], we proposed an estimator that directly estimates the *difference matrix* based on node potential data. Additionally, we establish the convexity of the estimator and demonstrate its superior performance in experiments. In this paper, we mainly focus on the statistical guarantees of this estimator. Building on the *primal-dual witness* framework (see, e.g.,) [9], we quantify the sample requirement necessary for accurately recovering the true sparse network changes. Specifically, we guarantee that an element-wise ℓ_∞ error bound of order $\mathcal{O}(\sqrt{\frac{p\log(p)}{n}})$ holds with high probability, with a sample complexity satisfying $\Omega(pd^2\log(p))$, where d represents the maximum degree of any row or column in

Δ^*. Finally, we note a gap between our theorem and the results obtained in the experiments. The challenge arise from the *square root perturbation bounds*, which we discuss later in the paper. Addressing this issue could lead to more accurate statistical guarantees. Given the importance of this problem, we are motivated to study the estimation of the *difference matrix* $\Delta^* = B_2^* - B_1^*$ between the structures of two networks that obey conservation laws, where B_1^* and B_2^* are invertible Laplacian matrix that represent the structure of the corresponding networks. In [8], we proposed an estimator that directly estimates the *difference matrix* based on node potential data. Additionally, we establish the convexity of the estimator and demonstrate its superior performance in experiments. In this paper, we mainly focus on the statistical guarantees of this estimator. Building on the *primal-dual witness* framework (see, e.g., [9]) we quantify the sample requirement necessary for accurately recovering the true sparse network changes. Specifically, we guarantee that an element-wise ℓ_∞ error bound of order $\mathcal{O}(\sqrt{\frac{p\log(p)}{n}})$ holds with high probability, with a sample complexity satisfying $\Omega(pd^2\log(p))$, where d represents the maximum degree of any row or column in Δ^*. Finally, we note a gap between our theorem and the scaling law suggested by experiments. The difference arises from our sub-optimal *square root perturbation bounds*, which we discuss later in the paper. Improving on this could lead to more accurate (and significantly better) statistical guarantees.

2 Preliminaries and Background

Consider two graphs, $\mathcal{G}_1 = ([p], E_1)$ and $\mathcal{G}_2 = ([p], E_2)$, with identical node sets $[p]$ but different edge sets E_1 and E_2. These graphs are associated with corresponding (invertible) Laplacian matrices B_1^* and B_2^*, respectively. Let $X_1 \sim \mathcal{N}(0, \Sigma_{X_1})$ and $X_2 \sim \mathcal{N}(0, \Sigma_{X_2})$ be random injection vectors associated with each graph, where the covariances Σ_{X_1} and Σ_{X_2} are assumed to be known. These injection vectors can be interpreted as electric flows governed by Kirchhoff's law or as traffic flows in a transportation network. The goal is to estimate the *difference matrix* Δ^* by observing node potentials $Y_i = (B_i^*)^{-1}X_i$ for $i \in \{1, 2\}$. First, notice that $Y_i \sim \mathcal{N}(0, {\Theta^*}_i^{-1})$, where $\Theta_i^* = B_i^*\Sigma_{X_i}^{-1}B_i^*$, $i \in \{1, 2\}$. Letting $M_{X_i} \succ 0$ denote the unique square root [10] of Σ_{X_i}. We next need to develop an expression for Δ^* as a function of Θ_1^* and Θ_2^*, we can define $\widetilde{Y}_i = M_{X_i}Y_i$ and set $\widetilde{\Theta}_i^* = (\mathrm{Cov}[\widetilde{Y}_i])^{-1}$, this expression follows by direct substitution. Now notice that with these definitions, we have $\Delta^* = M_{X_2}(\widetilde{\Theta}_2^*)^{\frac{1}{2}}M_{X_2} - M_{X_1}(\widetilde{\Theta}_1^*)^{\frac{1}{2}}M_{X_1}$. In [8], we use this insight to design a regularized estimator

$$\hat{\Delta} \in \mathrm{argmin}_{\Delta\in\mathbb{R}^{p\times p}} L(\Delta) + \lambda_n\|\Delta\|_{1,\mathrm{off}}, \tag{1}$$

where $\lambda_n \geq 0$, $L(\Delta) = \frac{1}{4}\left(\langle\widehat{\Psi}_1\Delta, \widehat{\Psi}_2\rangle + \langle\widehat{\Psi}_2\Delta, \widehat{\Psi}_1\rangle\right) - \langle\Delta, \widehat{\Psi}_1 - \widehat{\Psi}_2\rangle$, and $\widehat{\Psi}_i = M_{X_i}^{-1}\widetilde{S}_i^{\frac{1}{2}}M_{X_i}^{-1}$ which is an estimate of $\widetilde{\Psi}_i = M_{X_i}^{-1}(\widetilde{\Theta}_i^*)^{-\frac{1}{2}}M_{X_i}^{-1}$; $\widetilde{S}_i^{\frac{1}{2}}$. $\|\Delta\|_{1,\mathrm{off}}$ represents the ℓ_1-norm applied to the off-diagonal elements of Δ. This estimator is an ℓ_1 regularized variant of the D-trace loss [11]. While the above problem setup assumes Gaussian injection vectors, our methods hold for more general distributions. Toward characterizing the performance in these cases, we begin by defining tail conditions that dictate the achievable sample complexity of the problem.

3 Tail condition

We utilize the concept of tail condition from [9] to describe the distribution. This concept plays an important role in the central lemma of this paper.

Definition 1. Based on Ravikumar (Tail condition, [9]), the random vector Y satisfies the tail condition $T(f, \upsilon_*)$ if there exists a constant $\upsilon_* > 0$ and a function $f : \mathbb{N} \times (0, \infty) \to (0, \infty)$ such that for any $k, l \in [p]$ and $\delta \in (0, 1/\upsilon_*)$:

$$\mathbb{P}\left[|(\widetilde{S}_i)_{kl} - (\widetilde{\Theta}_i^{*\,-1})_{kl}| \geq \delta\right] \leq \frac{1}{f(n, \delta)} \tag{2}$$

$f(n, \delta)$ is monotonically increasing with respect to n or δ when the other variable is fixed. Two examples of tail functions f satisfy this monotone property are an exponential-type function, meaning that $f(n, \delta) = exp(cn\delta^a)$,

if $c, a > 0$, and a polynomial-type tail function meaning that $f(n, \delta) = cn^m\delta^{2m}$, for some positive integer m and scalar $c > 0$. The following inverse functions are required to show sample complexity,

$$n_f(\delta, r) := \max\{n : f(n, \delta) \leq p^\eta\} \quad \text{and} \quad \delta_f(n, r) := \max\{\delta : f(n, \delta) \leq p^\eta\}.$$

Since function $f(n, \delta)$ has the monotonicity property, the above functions are well defined and we know that if $n > n_f(\delta, p^\eta)$, this implies that $\delta \geq \delta_f(n, p^\eta)$.

3.1 Sub-Gaussian and Polynomial tail

Similar to the [6]. We study the case of sub-Gaussian and polynomial tail. We define the following

Definition 2 (Sub-Gaussian random variable)**.** A mean-zero random vector $Z \in \mathbb{R}^p$ with covariance matrix Σ is called sub-Gaussian if there exists a constant $\sigma \in (0, \infty)$ such that

$$\mathbb{E}\left[\exp\left(tZ_i(\Sigma_{ii})^{-1/2}\right)\right] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right),$$

for all $t \in \mathbb{R}$ and $i = 1, \ldots, p$, where Σ_{ii} is the (i, i) element of Σ.

Based on the [9], if $\tilde{Y}_i$ is *sub-Gaussian* with parameter σ. We have $v_* = \left\{\max_i(\widetilde{\Theta}^*)_{ii}^{-1} \cdot 8\left(1 + 4\sigma^2\right)\right\}^{-1}$, and

$$f(n, \delta) = \frac{1}{4}\exp\left(-c_* n\delta^2\right), \quad \textit{with} \quad c_* = [128(1 + 4\delta^2)^2 \max_i(\widetilde{\Theta}^*)_{ii}^{-1})^2]^{-1} \tag{3}$$

and the inverse functions take the form

$$\delta_f(p^\eta, n) = \sqrt{\frac{\log(4/p^\eta)}{c_* n}}, \quad \textit{and} \quad n_f(p^\eta, \delta) = \frac{\log(4/p^\eta)}{c_*\delta^2}. \tag{4}$$

Beyond sub-Gaussian distributions, we also provide guarantees for cases where the injection vectors X_1 and X_2 follow other distributions, including those with polynomial tails

Definition 3 (Polynomial tail)**.** A random vector $Z \in \mathbb{R}^p$ is said to have a polynomial tail if there exists a positive integer m and scalar $K_m \in \mathbb{R}$ such that

$$\mathbb{E}\left[\exp\left(tZ_i(\Sigma_{ii})^{-1/2}\right)^{4m}\right] \leq K_m,$$

for all $t \in \mathbb{R}$ and $i = 1, \ldots, p$.

If $\tilde{Y}_i$ is *polynomial tail* with parameter K_{1m} and K_{2m} respectively,

$$f(n, \delta) = c_k n^m \delta^{2m} \quad \textit{where} \quad c_* = \frac{1}{(2m^2 + 12m^2(\max_i(\widetilde{\Theta}^*)_{ii}^{-1}))^2 m(K_m + 1))}, \tag{5}$$

and the inverse tail functions take the form,

$$\delta_f(n, p^\eta) = \left(\frac{p^\eta}{c_*}\right)^{\frac{1}{2m}} \frac{1}{\sqrt{n}}, \quad \textit{and} \quad n_f(\delta, p^\eta) = \left(\frac{p^\eta}{c_*}\right)^{\frac{1}{m}} \frac{1}{\delta^2} \tag{6}$$

4 Statement of Main result

Our first result provides a theoretical analysis of the performance of estimator (1) when Y_1 and Y_2 follow a Sub-Gaussian distribution. In our setup, we assume that $\widetilde{Y}_i$ follow a Gaussian distribution. However, we work with sub-Gaussian distributions, which are a generalization of Gaussian distributions and include many common practical distributions. Additionally, we also extend this analysis to cases where Y_1 and Y_2 are non-Gaussian. Our results indicate that estimator (1) will reliably capture the sparsity structure of Δ^* and is close to the Δ^* with high probability as sample size n grows at a rate proportional to $pd^2 \log p$. Since our approach an ℓ_1 regularized variant of the D-trace loss function, the main result might appear similar to the [6]. However, In section 5.2.1, we discuss a novel challenge, which we refer to as the *Square Root Perturbation Bounds*. This challenge leads us to theorems that are not encompassed by previous work and opens a valuable avenue for future work.

We assume the true network difference Δ^* is sparse, that is, let $S = \{(i,j) : \Delta^*_{i,j} \neq 0\}$ be the support of Δ^* and $s = |S|$, $s < p$, d is the maximum number of non-zero entries across all rows in Δ^*, and $\max\{\|\widetilde{\Psi}_1\|_\infty, \|\widetilde{\Psi}_2\|_\infty\} \leq M$. We also define$\|A\|_1 = \|\text{vec}(A)\|_1$, $\|A\|_\infty = \max_{i,j}|A_{ij}|$ to denote the element-wise norm and in addition $\|A\|_{1,\infty} \triangleq \max_{i=1,\ldots,p} \sum_{j=1}^{p} |A_{ij}|$, and $\kappa_\Gamma = \|{\Gamma^*}^{-1}_{S,S}\|_{1,\infty}$. Denote $\hat{\Gamma} = \frac{\hat{\Psi}_1 \otimes \hat{\Psi}_2 + \hat{\Psi}_2 \otimes \hat{\Psi}_1}{2}$ and $\Gamma^* = \frac{\tilde{\Psi}_1 \otimes \tilde{\Psi}_2 + \tilde{\Psi}_2 \otimes \tilde{\Psi}_1}{2}$. This is also the Hessian matrix for Δ of the D-trace loss function (**??**). We can consider that Γ could be used to describe the relationships between two sets of variables in this case could be for understanding the overall interaction between the two sets. For two subsets $T_1, T_2 \subseteq \{1,\ldots,p\} \times \{1,\ldots,p\}$, the submatrix of X, denoted by $X_{T_1T_2}$, consists of rows and columns indexed by T_1 and T_2, respectively. We begin with the required assumption.

Assumption 1. *We assume the following irrepresentability condition*

$$\max_{e \in S^c} \|\Gamma^*_{e,S}(\Gamma^*_{S,S})^{-1}\|_1 \leq 1 - \alpha. \tag{7}$$

Where there are some $\alpha \in (0,1]$

The Hessian $\Gamma^*_{(j,k),(l,m)}$ also shows the covariance of the random variable linked to each edge of the graph. This assumption is very similar to the one in [9], it imposes control on the influences that non-edge terms (indexed by S^c), can have on the edge-based terms (indexed by S).

The result of the converge rate is stated in terms of the tail function f, and its inverse n_f and δ_f. and The choices of λ is specified in terms of a user-defined parameter $\eta > 2$. larger choices of η yield a faster rate of convergence but lead to more stringent requirements on sample size.

Theorem 1. Let $\widetilde{Y}_1$ and $\widetilde{Y}_2$ be the node potential vector. Suppose that $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are sub-Gaussian with parameter σ_1 and σ_2, respectively. Under the irrepresentability condition (1) for some $\eta > 2$ and with a sample size for both $\widetilde{Y}_1$ and $\widetilde{Y}_2$ that is lower bounded as $n \geq p\widetilde{C}_0(\eta \log(p) + \log(4))$, Then, with probability larger than $1 - 1/p^{\eta-2}$, for some $\eta > 2$:

(a) $\widehat{\Delta}$ recovers the sparsity structure of Δ^*; that is, $\hat{\Delta}_{S^C}$=0.

(b) $\widehat{\Delta}$ satisfies the element-wise ℓ_∞ bound $\|\widehat{\Delta} - \Delta^*\|_\infty \leq \widetilde{C}_1\sqrt{p}\left[\frac{\eta \log(p) + \log(4)}{n}\right]^{\frac{1}{2}}$.

Where $\widetilde{C}_0$ and $\widetilde{C}_1$ are constant depending on κ_Γ, M, α, which we assume they remain constant as a function of n, p, and d. Specifically, for constant $\widetilde{C}_0$, can be simplified into $\mathcal{O}(d^2)$ (see Appendix A.2 for their definitions and detailed proof). Thus, we can have element-wise ℓ_∞ bound $\|\hat{\Delta} - \Delta^*\|_\infty \in \mathcal{O}(\sqrt{\frac{p\log(p)}{n}})$ with sample size $n = \Omega(pd^2 \log(p))$ holds with high probability. For the other quantities that are involved in the theorem statement, κ_Γ and M measure the size of the Hessian (${\Gamma^*}^{-1}$). Finally, both $\widetilde{C}_0$ and $\widetilde{C}_1$ also depend on the irrepresentability 1, which growing as the parameter α approaches 0.

Beyond the sub-Gaussian distribution, We also provide guarantees for other injection distributions with polynomial tails, making our approach applicable to a wider range of practical problems.

Corollary 1. Under the same condition and notation in Theorem 1 with probability greater than $1 - \frac{2}{p^{\tau-2}}$, When κ_Γ, M, α remain constant and the bound can be summarized as sample size $n = \Omega(pd^2 \log(p))$ can guarantee have the following rate in Frobenius and spectral norm for Exponential-type tails,

$$\|\hat{\Delta} - \Delta^*\|_F = \mathcal{O}\left(\sqrt{s+p}\sqrt{p}\left[\frac{\eta \log(p) + \log(4)}{n}\right]^{\frac{1}{2}}\right), \|\hat{\Delta} - \Delta^*\|_{op} = \mathcal{O}\left(\min\{\sqrt{s+p}, d\}\sqrt{p}\left[\frac{\eta \log(p) + \log(4)}{n}\right]^{\frac{1}{2}}\right).$$

Proof. The Frobenius and operator norm calculations follow directly from standard matrix norm inequalities applied to the element-wise ℓ_∞ bound in part (b) of Theorem 1. Details proof are provided in Appendix A.2.3. □

Theorem 2. Under the same assumption of Theorem 1. Suppose that $\widetilde{Y}_1$ and $\widetilde{Y}_2$ have bounded moment as in Definition 3 and with a sample size that is lower bounded as $n \geq p\widetilde{C}_{P_0}^2 \left(p^{\eta/m}\right)$, Then, with probability larger than $1 - 2/p^{\eta-2}$, for some $\eta > 2$:

(a) $\hat{\Delta}$ recovers the sparsity structure of Δ^*; that is, $\hat{\Delta}_{S^C} = 0$.

(b) $\hat{\Delta}$ satisfies the element-wise ℓ_∞ bound $\|\hat{\Delta} - \Delta^*\|_\infty \leq \widetilde{C}_{P_1}\sqrt{p}\left[\frac{p^{\eta/m}}{n}\right]^{\frac{1}{2}}$.

Similar to Theorem 1, $\widetilde{C}_{P_0}$ and $\widetilde{C}_{P_1}$ are constant depending on κ_Γ, M, α, which we assume they remain constant as a function of n, p, and d. We can have element-wise ℓ_∞ bound $\|\hat{\Delta} - \Delta^*\|_\infty \in \mathcal{O}(\sqrt{p}\sqrt{\frac{p^{\eta/m}}{n}})$ with sample size $n = \Omega(pd^2 p^{\eta/m})$ holds with high probability.

Corollary 2. Under the assumption in the Theorem 2, we have the following rate in Frobenius and spectral norm for Polynomial-type tails

$$\|\hat{\Delta} - \Delta^*\|_F = \mathcal{O}\left(\sqrt{s+p}\sqrt{p}\left[\frac{p^{\eta/m}}{n}\right]^{\frac{1}{2}}\right), \|\hat{\Delta} - \Delta^*\|_{op} = \mathcal{O}\left(\min\{\sqrt{s+p}, d\}\sqrt{p}\left[\frac{p^{\eta/m}}{n}\right]^{\frac{1}{2}}\right).$$

Proof. Similar to the previous proof, detailed proof are provided in Appendix A.2.3. □

5 Proofs Outline of Main Result

Our proof builds on the *primal-dual witness* framework established in [9]. The method involves constructing a *primal-dual witness* pair $(\widetilde{\Delta}, \widetilde{Z})$ that satisfies the optimality conditions of our problem (1). If the construction succeeds, then $\widetilde{\Delta} = \hat{\Delta}$. Thus, the main direction of our proof is to demonstrate that this construction succeeds with high probability.

While developing our proof, we encountered an interesting challenge: the *square root perturbation bound* problem. We address this issue to derive statistical guarantees for our convex estimator in (1). First, we construct the *primal-dual witness method* and then provide a sequence of lemmas establishing the necessary conditions for the method to hold. Finally, we show that all conditions are satisfied under the assumptions specified in Theorems 1 and 2.

5.1 Primal-Dual witness method

We construct a *Primal-dual witness* pair $(\tilde{\Delta}, \tilde{Z})$ as following. The $\tilde{\Delta}$ is the solution to the restrictive problem:

$$\widetilde{\Delta} = \underset{\Delta \in \mathbb{R}^{p\times p}, \Delta_{S^c}=0}{\arg\min} \; L(\Delta) + \lambda_n \|\Delta\|_{1,\text{off}} \tag{8}$$

We note that the sub-differential of $\|.\|_{1,\text{off}}$ with respect of Δ contains set of matrices $Z \in \mathcal{R}^{p\times p}$ such that

$$Z_{ij} = \begin{cases} 0, & \text{if } i = j \\ \text{sign}(\Delta_{ij}), & \text{if } i \neq j \text{ and } \Delta_{ij} \neq 0 \\ \in [-1, +1], & \text{if } i \neq j \text{ and } \Delta_{ij} = 0 \end{cases} \tag{9}$$

The $\widetilde{\Delta}$ in the sub-differential of $\|\Delta\|_{1,\text{off}}$ is chosen so it satisfies the optimality condition of (8). This can be done by setting $\widetilde{Z}_{S^c}$ as

$$\widetilde{Z}_{S^c} = -\frac{1}{\lambda_n}\{\frac{1}{2}(\widehat{\Psi}_1\widetilde{\Delta}\widehat{\Psi}_2 + \widehat{\Psi}_2\widetilde{\Delta}\widehat{\Psi}_1) - \widehat{\Psi}_1 + \widehat{\Psi}_2\}_{S^c}$$

Finally, we need to establish the *strict dual feasibility*

$$|\widetilde{Z}_{ij}| < 1, \text{For all}(i, j)$$

5.2 Supporting Results and Lemmas

The proof of theorem 1 and 2 requires a sequence of lemmas. We further define the following notations.

$$\epsilon = \|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty\|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty + \|\widetilde{\Psi}_1\|_\infty\|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty + \|\widetilde{\Psi}_2\|_\infty\|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty,$$

$$\tilde{\epsilon} = \|\widehat{\Psi}_1 - \widehat{\Psi}_2 - \widetilde{\Psi}_1 + \widetilde{\Psi}_2\|_\infty, \Delta_\Gamma = \hat{\Gamma} - \Gamma^*.$$

Let $R(\Delta_\Gamma)$ denote the difference of gradient $\nabla g(\widehat{\Gamma}_{S,S})$ from its first-order Taylor expansion around $\Gamma^*_{S,S}$, we obtain the remainder takes the form

$$R(\Delta_\Gamma) = \widehat{\Gamma}^{-1}_{S,S} - \Gamma^{*\,-1}_{S,S} + \Gamma^{*\,-1}_{S,S}(\Delta_\Gamma)_{S,S}\Gamma^{*\,-1}_{S,S}.$$

In this section, we first provide conditions for the *strict dual feasibility condition* to hold. Then, we control the remainder term $R(\Delta_\Gamma)$, providing element-wise error bounds on $\hat{\Delta}$ and Δ^* in terms of ϵ. Finally, we show that with an appropriate choice of λ_n and under the sample size bounded as stated in Theorem 1 and 2, all conditions for the *strict dual feasibility condition* are satisfied, and the proof of the final result can be done by applying the *Square Root Perturbation Bounds*.

Lemma 1 (Conditions for strict dual feasibility and error bound)**.** Let the regularization parameter $\lambda_n > 0$, and assume that Assumption 1 holds. If the following conditions are satisfied:

(i) $\hat{\Delta}_{S^C} = 0$, if

$$\max_{e\in S^c}\|\hat{\Gamma}_{e,S}(\hat{\Gamma}_{S,S})^{-1}\|_1 \leq 1 - \alpha/2, \quad \|\hat{\Gamma}_{e,S}(\hat{\Gamma}_{S,S})^{-1} - \Gamma^*_{e,S}(\Gamma^*_{S,S})^{-1}\|_1 \leq \frac{\alpha\lambda_n}{8M}, \tag{10}$$

$$\tilde{\epsilon} \leq \frac{\alpha\lambda_n}{2(4-\alpha)}; \tag{11}$$

(ii) $\hat{\Delta}_{S^C} = 0$, if

$$\epsilon < \frac{1}{6s\kappa_\Gamma}, \tag{12}$$

$$3s\epsilon(\kappa_\Gamma + 2sM^2\kappa_\Gamma^2) \leq 0.5\alpha\min(1, 0.25\lambda_n M^{-1}), \tag{13}$$

$$\tilde{\epsilon} \leq \frac{\alpha\lambda_n}{2(4-\alpha)}; \tag{14}$$

(iii) *Under the conditions in* ***(ii)****, we also have*

$$\|\hat{\Delta} - \Delta^*\|_\infty < (\tilde{\epsilon} + \lambda_n)\kappa_\Gamma + 3(\tilde{\epsilon} + 2M + \lambda_n)s\epsilon\kappa_\Gamma^2. \tag{15}$$

Proof. Since the proofs are very similar to the [6],we only provide a high-level picture of our proofs. It is divided into three parts corresponding to **(i)**, **(ii)**, and **(iii)**. Part **(i)** establishes the condition for the *primal-dual witness* to hold. Part **(ii)** builds the connection between **(i)** and **(iii)**, while part **(iii)** provides the error bound necessary for our central lemma, which will be discussed later.

For part **(i)**, we establish that the *Primal-Dual Witness* holds under the conditions stated in Lemma 1. Specifically, we follow the proof of Lemma A1 in the Supplementary Material of [6] to verify that the inequality below implies the two inequalities in (10):

$$\max_{e\in S^c}\left\|\widehat{\Gamma}_{e,S}\left(\widehat{\Gamma}_{S,S}\right)^{-1}-\Gamma^*_{e,S}\left(\Gamma^*_{S,S}\right)^{-1}\right\|_1\leq 0.5\alpha\min\left(1,0.25\lambda_n M^{-1}\right). \tag{16}$$

The second inequality in (10) follows directly from (16). To prove the first inequality, we combine the fact that $\alpha=1-\max_{e\in S^c}\|\Gamma^*_{e,S}(\Gamma^*_{S,S})^{-1}\|_1$ and apply the triangle inequality.

For part **(ii)**, we can see that the right-hand side of (16) is equivalent to the right-hand side of (13). Therefore, it suffices to show that the left-hand side of (16) is smaller than the left-hand side of (13). This follows the same reasoning as in the proof of Lemma A1 in [6], with adjustments to match our notation. Specifically, Σ^*_X and Σ^*_Y in [6] are replaced with $\widetilde{\Psi}_1$ and $\widetilde{\Psi}_2$, respectively, while $\widehat{\Sigma}_X$ and $\widehat{\Sigma}_Y$ are replaced with $\widehat{\Psi}_1$ and $\widehat{\Psi}_2$. Apart from this notational change, the argument remains consistent with the original proof. We will later verify all conditions in part **(ii)** with a sample size lower-bounded as specified in Theorems 1 and 2 within the proof of the central Lemma.

For part **(iii)**, using the inequalities from **(iii)** and Lemma 2, which is stated right after, we obtain the inequality (15). See the proofs of Lemmas A1 and A2 in [6] for further details.

□

Lemma 2. (control of remainder) Combine condition 12 from Lemma 1 and we get $\|{\Gamma^*}^{-1}_{S,S}\|_{1,\infty}\|(\Delta_\Gamma)_{S,S}\|_{1,\infty}<\frac{1}{3}$. Then,

$$\|R(\Delta_\Gamma)\|_\infty\leq 6d\epsilon^2\kappa_\Gamma^3,\|R(\Delta_\Gamma)\|_{1,\infty}\leq 6d^2\epsilon^2\kappa_\Gamma^3 \tag{17}$$

Since $R(\Delta_\Gamma)=\hat{\Gamma}^{-1}_{S,S}-{\Gamma^*}^{-1}_{S,S}+{\Gamma^*}^{-1}_{S,S}(\Delta_\Gamma)_{S,S}{\Gamma^*}^{-1}_{S,S}$. We have

$$\|\hat{\Gamma}^{-1}_{S,S}-{\Gamma^*}^{-1}_{S,S}\|_\infty\leq 6d\epsilon^2\kappa_\Gamma^3+2\epsilon\kappa_\Gamma^2 \tag{18}$$

$$\|\hat{\Gamma}^{-1}_{S,S}-{\Gamma^*}^{-1}_{S,S}\|_{1,\infty}\leq 6d^2\epsilon^2\kappa_\Gamma^3+2d\epsilon\kappa_\Gamma^2 \tag{19}$$

Proof. The proof of this follows the same approach as Lemmas A2 in the Supplementary Material of [6]. □

5.2.1 Square Root Perturbation Bounds

The estimator (1) use $\widehat{\Psi}_i$ as proxy for the true $\widetilde{\Psi}_i$. The first step is to obtain bounds on the differences $\widehat{\Psi}_i-\widetilde{\Psi}_i$, which can be written as $M^{-1}_{X_i}\left(\widetilde{S}_i^{\frac{1}{2}}-\left(\widetilde{\Theta}^*_i\right)^{-\frac{1}{2}}\right)M^{-1}_{X_i}$. To bound the differences $\widetilde{S}_i^{\frac{1}{2}}-\left(\widetilde{\Theta}^*_i\right)^{-\frac{1}{2}}$, we need to identify alternative expressions that bound the behavior of the square root function.

Lemma 3. (Generalized Power-Stormer Inequalities [12]) Let A and B be positive operators and let $n\geq 1$. If f is an operator monotone function. Then,

$$\|A^{\frac{1}{n}}-B^{\frac{1}{n}}\|_{op}\leq\|A-B\|_{op}^{\frac{1}{n}}$$

From Lemma 3 we can obtain the following chain of inequality,

$$\|\sqrt{S_i}-\sqrt{{\Theta^*_i}^{-1}}\|_\infty\leq\|\sqrt{S_i}-\sqrt{{\Theta^*_i}^{-1}}\|_{op}\leq\sqrt{\|S_i-{\Theta^*_i}^{-1}\|_{op}}\leq\sqrt{p}\|S_i-{\Theta^*_i}^{-1}\|_\infty$$

Instead of direct calculation on the $\widetilde{S}_i^{\frac{1}{2}} - (\widetilde{\Theta}_i^*)^{-\frac{1}{2}}$, now we can use $\sqrt{p}\|\widetilde{S}_i - \widetilde{\Theta}_i^{*-1}\|_\infty$ to have bounds on the differences $\widetilde{S} - (\widetilde{\Theta}_i^*)^{-1}$. The final result of the Theorem 1 and 2 is based on the the control of noise $\widehat{\Psi}_i - \widetilde{\Psi}_i$. We have inequality:

$$\|\widehat{\Psi}_i - \widetilde{\Psi}_i\|_\infty = \|M_{X_i}^{-1}\|_\infty \|(\widetilde{S}_i)^{\frac{1}{2}} - (\widetilde{\Theta}_i)^{*-\frac{1}{2}}\|_\infty \|M_{X_i}^{-1}\|_\infty \leq \|M_{X_i}^{-1}\|_\infty \sqrt{p} \|\widetilde{S}_i - (\widetilde{\Theta}_i^*)^{-1}\|_\infty \|M_{X_i}^{-1}\|_\infty \tag{20}$$

Lemma 4. (Control of noise term) Let $\widetilde{\delta}_{f_i} = \|M_{X_i}^{-1}\|_\infty \sqrt{p}\delta_{f_i}(n, p^\eta)\|M_{X_i}^{-1}\|_\infty$. For some $\eta > 2$, We have

$$\mathbb{P}\left(\|\widehat{\Psi}_i - \widetilde{\Psi}_i\|_\infty \leq \widetilde{\delta}_{f_i}(n, p^\eta)\right) \geq 1 - \frac{1}{p^{\eta-2}},$$

Proof. For any $0 < \delta < 1/v_*$ and $r \geq 1$, if $n > n_f(\delta, r)$, we have $f(n, \delta) > r$ and thus $\delta_f(n, r) < \delta$, since $f(n, \delta)$ is monotonically increasing in δ. Thus $\mathbb{P}\{\left|(\widetilde{S}_i)_{kl} - (\widetilde{\Theta}_i^*)_{kl}^{-1}\right| \geq \delta_f(n, r)\} \leq 1/f(n, \delta_f(n, r)) = r^{-1}$, applying union bound across all entries, we have $\mathbb{P}\{\|(\widetilde{S}_i) - (\widetilde{\Theta}_i^*)^{-1}\|_\infty < \delta_f(n, r)\} > 1 - p^2 r^{-1}$. Finally, apply the inequalities (20), and set $r = p^\tau$. We obtain

$$\mathbb{P}\left(\|M_{X_i}^{-1}\|_\infty \sqrt{p}\, \|\widetilde{S}_i - (\widetilde{\Theta}_i^*)^{-1}\|_\infty \, \|M_{X_i}^{-1}\|_\infty \leq \|M_{X_i}^{-1}\|_\infty \sqrt{p}\, \delta_f(n, p^\eta)\, \|M_{X_i}^{-1}\|_\infty\right) \geq 1 - \frac{1}{p^{\eta-2}}. \tag{21}$$

□

Note that by applying the union bound across both matrices, we obtain a combined probability bound of $1 - \frac{2}{p^{\eta-2}}$. We now need to offer guarantees for support recovery and bounds for the infinity norm of $\widehat{\Delta}$ for distribution that comply with the tail condition $T(f, v_*)$ as outlined in the Definition 1.

Lemma 5. (Central Lemma) let $\widehat{\Delta}$ be the unique solution of the problem 1 with

$$\lambda_n = \max\{\frac{2(4-\alpha)(\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2})}{\alpha}, \frac{24dM(\kappa_\Gamma + dM^2\kappa_\Gamma^2)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1})}{\alpha}\}.$$

For a distribution satisfying the assumption 1 and the tail condition $T(f, v_*)$ in 1. Then if the sample size is lower bounded as

$$n > n_f\left(\min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{4-\alpha}\right\}, p^\eta\right)$$

then with probability greater than $1 - \frac{1}{p^{\eta-2}}$, the $\widehat{\Delta}$ recovers the sparsity structure of Δ^*, and satisfies the ℓ_∞ bound

$$\|\widehat{\Delta} - \Delta^*\|_\infty \leq (\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2} + \lambda_n)\kappa_\Gamma + 3d\kappa_\Gamma^2(\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2} + \lambda_n + 2M)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1}) \tag{22}$$

Proof Sketch. *We show that the Primal-Dual witness construction succeeds (see proof of Lemma 1 and 2) with the probability and information stated in this Lemma, This is equivalent to demonstrating that the inequalities in Lemma 1 holds with the required probability. We have show that* $\mathbb{P}\left(\|\widehat{\Psi}_i - \widetilde{\Psi}_i\|_\infty \leq \widetilde{\delta}_{f_i}(n, p^\eta)\right) \geq 1 - \frac{1}{p^{\eta-2}}$, *in Lemma 4. Given that event* $\|\widehat{\Psi}_i - \widetilde{\Psi}_i\|_\infty \leq \widetilde{\delta}_{f_i}(n, p^\eta)$ *holds, we then demonstrate that the inequalities in 1 are satisfied. We provide full proof in Appendix A.1.*

The result in Theorem 1 and 2 follow by combining those sequence of lemmas. Substituting inverse functions of *Sub-Gaussian* and *Polynomial tail* in 1 to 22, we obtain δ_{f_1} and δ_{f_2}. Finally, the main result is proved by applying some algebraic manipulations to (22).

6 Numerical Simulations

It should be noted that while the sample complexity scales as $pd^2 \log p$, one might have hoped it would scale as $d^2 \log p$. In fact, this is what the experimental results in [8] appear to suggest. In this section, we perform additional simulations on various networks to evaluate the performance of the proposed estimator. The experiments include four synthetic random networks of size $p = 30$ and three benchmark networks. We assess the edge recovery performance using the F-score.

$$\text{F-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \in [0, 1],$$

where TP (true positives) is the number of correctly detected edges, FP (false positives) is the number of non-existent edges that were detected, and FN (false negatives) is the number of actual edges that were not detected. We also evaluate the worst-case error using $\|\widehat{\Delta} - \Delta^*\|_\infty$.

6.1 Syntehtic Networks

The synthetic networks examined include Erdős-Rényi graphs, characterized by randomly connected nodes; Small-World graphs, generated using the Watts-Strogatz model; Scale-Free graphs, constructed through the Barabási-Albert model; and structured Grid Graphs. For each network, an adjacency matrix A is constructed, and the corresponding Laplacian matrix $L = D - A$ is computed, where D is the degree matrix. The Laplacian is then regularized as $B_1 = L + I_p$, ensuring positive definiteness, where I_p denotes the p-dimensional identity matrix. To model changes in edge structure, a difference matrix Δ^* is applied such that $\Delta^* + B_1 = B_2$. The support of Δ^* is determined by a binary sparsity pattern, with a sparsity factor of 0.90. This means 90% of the entries in Δ^* are zeros, ensuring the *difference matrix* is highly sparse. Sample sizes are scaled with the network properties according to $N \propto d^2 \log p$, where d is the maximum degree of the network.

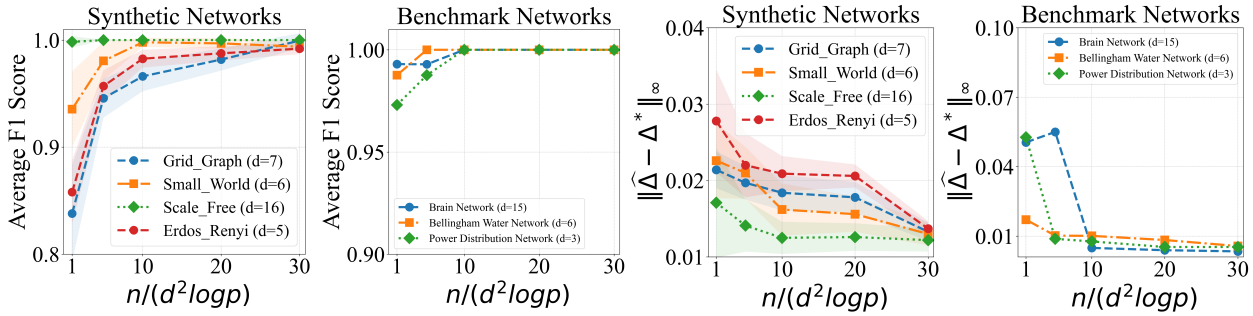

Figure 1: Estimation performance for synthetic networks. Solid curves represent the mean performance, while shaded areas indicate standard deviations.

6.2 Benchmark networks

We focus on three distinct networks in this analysis: **Power Distribution Network:** The IEEE 33-bus power distribution system consists of 33 buses and 32 branches (edges), with a maximum degree of $d = 3$. The raw dataset is publicly accessible at [1]. **Water Distribution Network:** The ground truth adjacency matrix, featuring 121 nodes and 162 edges with a maximum degree of $d = 6$, is created by processing the raw dataset described in [13] using the WNTR simulator[2]. The dataset is available at [3]. **Brain Network:** This study utilizes a publicly available benchmark connectivity matrix[4], with details on its construction outlined in [14]. The ground truth adjacency matrix, denoted as A, is a 90×90 matrix, where each row and column represent a unique region of interest (ROI) in the brain.

7 Discussion and Future Work

We plot the performance of the estimator as it improves with the rescaled sample size $n/d^2\log(p)$ instead of $\mathcal{O}(pd^2\log(p))$. This scaling is sublinear in the dimension p. so that for sub-Gaussian random variables, the method can succeed in $n < p$ case. Although all simulation results demonstrate that the estimator improves with the rescaled sample size $n/d^2\log(p)$, our theoretical analysis has only established a sample complexity of $\mathcal{O}(pd^2\log(p))$. Closing this gap between theory and simulation remains a key objective for our future research. The primary challenge in achieving a tighter bound lies in deriving a concentration inequality for the square root of a matrix.

References

[1] Anirudh Rayas, Rajasekhar Anguluri, and Gautam Dasarathy. Learning the structure of large networked systems obeying conservation laws. In *Advances in Neural Information Processing Systems*, volume 35, pages 14637–14650. Curran Associates, Inc., 2022.

[2] Rajasekhar Anguluri, Gautam Dasarathy, Oliver Kosut, and Lalitha Sankar. Grid topology identification with hidden nodes via structured norm minimization. *IEEE Control Systems Letters*, 6:1244–1249, 2021.

[3] Deepjyoti Deka, Saurav Talukdar, Michael Chertkov, and Murti V Salapaka. Graphical models in meshed distribution grids: Topology estimation, change detection & limitations. *IEEE Transactions on Smart Grid*, 11(5):4299–4310, 2020.

[4] Ali Shojaie. Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2), 2021.

[5] Sen Na, Mladen Kolar, and Oluwasanmi Koyejo. Estimating differential latent variable graphical models with applications to brain connectivity. *Biometrika*, 108(2):425–442, 2021.

[6] Huili Yuan, Ruibin Xi, Chong Chen, and Minghua Deng. Differential network analysis via lasso penalized D-trace loss. *Biometrika*, 104(4):755–770, 2017.

[7] Gautam Dasarathy, Parikshit Shah, and Richard G Baraniuk. Sketched covariance testing: A compression-statistics tradeoff. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 676–680. IEEE, 2017.

[8] Anirudh Rayas, Rajasekhar Anguluri, Jiajun Cheng, and Gautam Dasarathy. Differential analysis for networks obeying conservation laws. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[9] Pradeep Ravikumar, Martin J Wainwright, Garvesh Raskutti, and Bin Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 22: 935–980, 2011.

[10] Rajendra Bhatia. Positive definite matrices. Princeton University Press, 2009.

[11] Teng Zhang and Hui Zou. Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika*, 101(1):103–120, 2014.

[12] John Phillips. On the uniform continuity of operator functions and generalized powers-stormer inequalities. 1987. URL http://hdl.handle.net/1828/1506. Available since 2009-08-14T17:19:22Z.

[13] Erika Hernadez, Steven Hoagland, and Lindell Ormsbee. Water distribution database for research applications. In *World Environmental and Water Resources Congress*, pages 465–474, 2016.

[14] Antonín Škoch, Barbora Rehák Bučková, Jan Mareš, Jaroslav Tintěra, Pavel Sanda, Lucia Jajcay, Jiří Horáček, Filip Španiel, and Jaroslav Hlinka. Human brain structural connectivity matrices–ready for modelling. *Scientific Data*, 9(1), 2022.

A Appendix

For clarity, we recall the following definitions.

$$\epsilon = \|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty \|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty + \|\widetilde{\Psi}_1\|_\infty \|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty + \|\widetilde{\Psi}_2\|_\infty \|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty,$$

$$\tilde{\epsilon} = \|\widehat{\Psi}_1 - \widehat{\Psi}_2 - \widetilde{\Psi}_1 + \widetilde{\Psi}_2\|_\infty.$$

A.1 Proof of Lemma 5

Let

$$\bar{\delta} = \min\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{4-\alpha}\},$$

$$\lambda_n = \max\{\frac{2(4-\alpha)(\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2})}{\alpha}, \frac{24dM(\kappa_\Gamma + dM^2\kappa_\Gamma^2)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1})}{\alpha}\}.$$

For $n > n_f(\bar{\delta}, p^\eta)$, the following hold with probability at least $1 - 1/p^{\eta-2}$, we have

$$\|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty \le \widetilde{\delta}_{f_1}(n_1, p^\eta) = \widetilde{\delta}_{f_1} < \bar{\delta},$$

$$\|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty \le \widetilde{\delta}_{f_2}(n_2, p^\eta) = \widetilde{\delta}_{f_2} < \bar{\delta}.$$

we now prove our *Central Lemma* in the following steps. First we verify all conditions in part (ii) of Lemma 1. Since $\epsilon \le \widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1} < \bar{\delta}^2 + 2M\bar{\delta}$ and $\bar{\delta} = -M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}$. condition (12) can be easily verified. From $\delta_{f_1} < \bar{\delta} \le -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}$, we get $\epsilon \le \delta_{f_1}\delta_{f_2} + M\delta_{f_2} + M\delta_{f_1} \le \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}$, which implies that $3d\epsilon(\kappa_\Gamma + dM^2\kappa_\Gamma^2) \le \alpha/8 < \alpha/2$. Since $\widetilde{\delta}_{f_1} = \widetilde{\delta}_{f_1}(n_1, p^\eta) < \bar{\delta} \le \frac{\alpha M}{4-\alpha}$, $\widetilde{\delta}_{f_2} = \widetilde{\delta}_{f_2}(n_2, p^\eta) < \bar{\delta} \le \frac{\alpha M}{4-\alpha}$, now we have $\frac{2(4-\alpha)(\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2})}{\alpha} \le 4M$. Then, Recall the condition (13) we have definition of λ_n, $0.25M^{-1}\lambda_n \le 1$. Now we have,

$$\begin{aligned} 3d\epsilon(\kappa_\Gamma + dM^2\kappa_\Gamma^2) &\le 3d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1}) \\ &= 8^{-1}\alpha M^{-1}\frac{24dM(\kappa_\Gamma + dM^2\kappa_\Gamma^2)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1})}{\alpha} \\ &\le 8^{-1}\alpha M^{-1}\lambda_n. \end{aligned}$$

Combining the above results, condition (13) is verified. The condition (14) can be verified as following,

$$\tilde{\epsilon} \le \|\widehat{\Psi}_1 - \widetilde{\Psi}_1\|_\infty + \|\widehat{\Psi}_2 - \widetilde{\Psi}_2\|_\infty \le (\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2}) \le \frac{\alpha\lambda_n}{2(4-\alpha)}.$$

A.2 Proof of Theorem 1 and Theorem 2

Then, by part (iii) of the Lemma 1, We have

$$\begin{aligned} \|\widehat{\Delta} - \Delta^*\|_\infty &< (\hat{\epsilon} + \lambda_n)\kappa_\Gamma + 3(\hat{\epsilon} + 2M + \lambda_n)d\epsilon\kappa_\Gamma^2 \\ &\le (\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2} + \lambda_n)\kappa_\Gamma + 3d\kappa_\Gamma^2(\widetilde{\delta}_{f_1} + \widetilde{\delta}_{f_2} + \lambda_n + 2M)(\widetilde{\delta}_{f_1}\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_2} + M\widetilde{\delta}_{f_1}). \end{aligned} \tag{23}$$

A.2.1 Theorem 1

Suppose that $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are sub-Gaussian,

Part(a): From Lemma 5, we show that if $n > n_f(\delta, p^\eta)$, then $\widehat{\Delta}$ recovers the sparsity structure of Δ^*. we use the tail condition for sub-gaussian distributions and inverse functions from 1 we have

$$\tilde{n}_f(\delta, p^\eta) = \min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{16}\right\}^{-2} p\widetilde{C}^2(\eta\log(p)+\log(4))$$

Where $\widetilde{C} = 128[1 + 4\max(\sigma_1^2, \sigma_2^2)\max(\max_i((\Theta^{*}{}_1^{-1})_{ii}), \max_i((\Theta^{*}{}_2^{-1})_{ii}))^2]^{\frac{1}{4}}\max(\|M_{X_1}^{-1}\|_\infty^2, \|M_{X_2}^{-1}\|_\infty^2)$. We can conclude that if

$$n \geq \min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{16}\right\}^{-2} p\widetilde{C}^2(\eta\log(p) + \log(4)),$$

the estimator $\widehat{\Delta}$ recovers the sparsity structure of Δ^*.

$\widetilde{C} = 128\Big(1 + 4\max(\sigma_1^2, \sigma_2^2)\max\{\max_i\{(\widetilde{\Theta}_1^{*^{-1}})_{ii}\}, \max_i\{(\widetilde{\Theta}_2^{*^{-1}})_{ii}\}\}^2\Big)^{\frac{1}{4}}\max\{\|M_{X_1}^{-1}\|_\infty^2, \|M_{X_2}^{-1}\|_\infty^2\}$.
if we assume that M, κ_Γ, and α are constant as functions of n, p, d, the above expression can be simplified into $n = \Omega(d^2 p\log(p))$.

Part(b) : Compute $\widetilde{\delta}_{f_i}$ based on the definition 1, we obtain

$$\widetilde{\delta}_{f_1} \leq [128(1 + 4\sigma_1^2)\max_i((\widetilde{\Theta}_1^{*^{-1}})_{ii})^2]^{\frac{1}{4}}\|M_{X_1}^{-1}\|_\infty^2 p^{\frac{1}{2}}\left\{\frac{\eta log(p) + log(4)}{n}\right\}^{\frac{1}{2}}$$

$$\widetilde{\delta}_{f_2} \leq [128(1 + 4\sigma_2^2)\max_i((\widetilde{\Theta}_2^{*^{-1}})_{ii})^2]^{\frac{1}{4}}\|M_{X_2}^{-1}\|_\infty^2 p^{\frac{1}{2}}\left\{\frac{\eta log(p) + log(4)}{n}\right\}^{\frac{1}{2}}$$

The main result in Theorem 1 can now be proved by substituting the above inequalities into the right-hand side of equation 23 with some algebra manipulation. Finally we obtain

$$\begin{aligned}\|\widehat{\Delta} - \Delta^*\|_\infty \leq\ & \left[\kappa_\Gamma + 3d\kappa_\Gamma^2 M^2(\frac{\alpha^2}{256} + \frac{\alpha}{8})\right]\left[2\widetilde{C} + \max\left\{\frac{8\widetilde{C}}{\alpha}, \left(\frac{6}{16} + \frac{32}{\alpha}\right)dM^2(\kappa_\Gamma + dM^2\kappa_\Gamma^2)\right\}\right] \\ & + \left[\frac{6}{16}dM^2\kappa_\Gamma^2(\alpha + 32)\widetilde{C}\right]\sqrt{p}\left[\frac{\eta\log(p) + \log(4)}{n}\right]^{\frac{1}{2}}.\end{aligned}$$

Let $\widetilde{C}_0$ be defined as

$$\min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{16}\right\}^{-2} p\widetilde{C}^2,$$

and $\widetilde{C}_1$ be defined as

$$\left[\kappa_\Gamma + 3d\kappa_\Gamma^2 M^2(\frac{\alpha^2}{256} + \frac{\alpha}{8})\right]\left[2\widetilde{C} + \max\left\{\frac{8\widetilde{C}}{\alpha}, \left(\frac{6}{16} + \frac{32}{\alpha}\right)dM^2(\kappa_\Gamma + dM^2\kappa_\Gamma^2)\right\}\right] + \left[\frac{6}{16}dM^2\kappa_\Gamma^2(\alpha + 32)\widetilde{C}\right].$$

This completes the proof of Theorem 1.

A.2.2 Theorem 2

Suppose that $\widetilde{Y}_1$ and $\widetilde{Y}_2$ are polynomial-tailed distributions, The proof is very similar to the proof of Theorem 1. Hence, we only provide high-level explanation.
Part(a): Similar to the Theorem 1, now we use polynoimal type tail bound to compute $n_f(\delta, p^\eta)$, we obtain

$$n > \min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{16}\right\}^{-2} p\widetilde{C}_P^2\left(p^{\eta/m}\right)$$

where $\widetilde{C}_P = \max(\|M_{X_1}^{-1}\|_\infty^2, \|M_{X_2}^{-1}\|_\infty^2)\{2m[m*(\max(K_{1m}, K_{2m})+1)]^{\frac{1}{2m}}(\max(\max_i((\Theta^{*}{}_1^{-1})_{ii}), \max_i(\Theta^{*}{}_2^{-1}{}_{ii}))\}^{\frac{1}{2}}$.
Part(b): Similary to the proof of theorem 1, based on the definition 1. we obtain $\widetilde{\delta}_{f_i}$ for polynomial type tail,

$$\widetilde{\delta}_{f_1} \leq \max(\|M_{X_1}^{-1}\|_\infty^2, \|M_{X_2}^{-1}\|_\infty^2)\{2m[m*(K_{1m}+1)]^{\frac{1}{2m}}(\max_i((\Theta^{*}{}_1^{-1})_{ii})\}^{\frac{1}{2}} p^{\frac{1}{2}}\left\{\frac{p^{\eta/m}}{n}\right\}^{\frac{1}{2}},$$

$$\widetilde{\delta}_{f_2} \leq \max(\|M_{X_1}^{-1}\|_\infty^2, \|M_{X_2}^{-1}\|_\infty^2)\{2m[m*(K_{2m}+1)]^{\frac{1}{2m}}(\max_i((\Theta^{*}{}_2^{-1})_{ii})\}^{\frac{1}{2}} p^{\frac{1}{2}}\left\{\frac{p^{\eta/m}}{n}\right\}^{\frac{1}{2}}.$$

The main result in Theorem 1 can now be proved by substituting the above inequalities into the right-hand side of equation 23 with some algebra manipulation

$$\|\hat{\Delta} - \Delta^*\|_\infty \leq \left[\kappa_\Gamma + 3d\kappa_\Gamma^2 M^2(\frac{\alpha^2}{256} + \frac{\alpha}{8})\right]\left[2\widetilde{C}_P + \max\left\{\frac{8\widetilde{C}_P}{\alpha}, \left(\frac{6}{16} + \frac{32}{\alpha}\right)dM^2(\kappa_\Gamma + dM^2\kappa_\Gamma^2)\right\}\right]$$
$$+ \left[\frac{6}{16}dM^2\kappa_\Gamma^2(\alpha+32)\widetilde{C}_P\right]\sqrt{p}\left[\frac{p^{\eta/m}}{n}\right]^{\frac{1}{2}}.$$

Similar to the previous proof, let $\widetilde{C}_{P_0}$ be defined as

$$\min\left\{-M + \sqrt{M^2 + (6d\kappa_\Gamma)^{-1}}, -M + \sqrt{M^2 + \frac{\alpha}{24d(\kappa_\Gamma + dM^2\kappa_\Gamma^2)}}, \frac{\alpha M}{16}\right\}^{-2}\widetilde{C}_P^2,$$

and $\widetilde{C}_{P_1}$ be defined as

$$\left[\kappa_\Gamma + 3d\kappa_\Gamma^2 M^2(\frac{\alpha^2}{256} + \frac{\alpha}{8})\right]\left[2\widetilde{C}_P + \max\left\{\frac{8\widetilde{C}_P}{\alpha}, \left(\frac{6}{16} + \frac{32}{\alpha}\right)dM^2(\kappa_\Gamma + dM^2\kappa_\Gamma^2)\right\}\right] + \left[\frac{6}{16}dM^2\kappa_\Gamma^2(\alpha+32)\widetilde{C}_P\right].$$

This completes the proof of Theorem 2.

A.2.3 Proof of Corollary 1 and Corollary 2

Consider the following inequality:

$$\|\hat{\Delta} - \Delta^*\|_F^2 = \sum_{i,j}(\hat{\Delta}_{ij} - B_{ij}^*)^2 = \sum_i(\hat{\Delta}_{ii} - \Delta_{ii}^*)^2 + \sum_{i\neq j}(\hat{\Delta}_{ij} - \Delta_{ij}^*)^2 \tag{24}$$

$$\leq p\|\hat{\Delta} - \Delta^*\|_\infty^2 + s\|\hat{\Delta} - \Delta^*\|_\infty^2 \tag{25}$$

$$= (s+p)\|\hat{\Delta} - \Delta^*\|_\infty^2, \tag{26}$$

We now show the spectral norm.

$$\|\hat{\Delta} - \Delta^*\|_2 \leq d\|\hat{\Delta} - \Delta^*\|_\infty \tag{27}$$

and that

$$\|\hat{\Delta} - \Delta^*\|_2 \leq \|\hat{\Delta} - \Delta^*\|_F \leq \sqrt{s+p}\|\hat{\Delta} - \Delta^*\|_\infty. \tag{28}$$

We obtain

$$\|\hat{\Delta} - \Delta^*\|_2 \leq \min\{\sqrt{s+p}, d\}\|\hat{\Delta} - \Delta^*\|_\infty. \tag{29}$$