Robust Gaussian Graphical Model Estimation with Arbitrary Corruption

Lingxiao Wang [1] Quanquan Gu [1]

Abstract

We study the problem of estimating the high-dimensional Gaussian graphical model where the data are arbitrarily corrupted. We propose a robust estimator for the sparse precision matrix in the high- dimensional regime. At the core of our method is a robust covariance matrix estimator, which is based on truncated inner product. We establish the statistical guarantee of our estimator on both estimation error and model selection consistency. In particular, we show that provided that the number of corrupted samples n_2 for each variable satisfies $n_2 \lesssim \sqrt{n}/\sqrt{\log d}$, where n is the sample size and d is the number of variables, the proposed robust precision matrix estimator attains the same statistical rate as the standard estimator for Gaussian graphical models. In addition, we propose a hypothesis testing procedure to assess the uncertainty of our robust estimator. We demonstrate the effectiveness of our method through extensive experiments on both synthetic data and real-world genomic data.

1 Introduction

Gaussian graphical models (GGMs) have attracted increasing attention in recent years, especially in the field of high-dimensional statistical learning. In Gaussian graphical models, a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ follows a multivariate normal distribution $N_d(\mathbf{0}, \boldsymbol{\Sigma}^*)$. It corresponds to the vertex set $V = \{1, \ldots, d\}$ of an undirected graph $G = (V, E)$, where the edge set E describes the conditional independence relationships between nodes $X_1, \ldots, X_d$. It is well-known that the graph G is encoded by the sparsity pattern of the precision matrix $\boldsymbol{\Theta}^* = \boldsymbol{\Sigma}^{*-1}$. More specifically, no edge connects X_i and X_j if and only if $\Theta^*_{ij} = 0$. Consequently, estimation of the precision matrix $\boldsymbol{\Theta}^*$ corresponds to parameter estimation, and specifying the non-zero set of $\boldsymbol{\Theta}^*$ corresponds to graphical model selection (Cox & Wermuth, 1996).

In the high-dimensional settings, where the number of variables d can exceed the number of observations n, a large body of literature has studied the problem of precision matrix estimation in Gaussian graphical models and their variants (Meinshausen & Bühlmann, 2006; Yuan & Lin, 2007; Friedman et al., 2008; Banerjee et al., 2008; Yuan, 2010; Cai et al., 2011; Wang et al., 2016; Xu & Gu, 2016; Xu et al., 2016; 2017). For instance, Meinshausen & Bühlmann (2006) developed a neighborhood pursuit approach for estimating conditional independence relationship separately for each node in the graph. This method estimates the precision matrix by solving a collection of sparse regression problems using Lasso in parallel. Yuan & Lin (2007); Friedman et al. (2008); Banerjee et al. (2008) proposed a ℓ_1 norm regularized Gaussian negative log-likelihood method, which called Graphical Lasso (GLasso), to directly estimate the precision matrix. More recently, Yuan (2010); Cai et al. (2011) proposed the graphical Dantzig selector and CLIME, respectively. Both of these methods can be solved by linear programming and have more favorable theoretical properties than GLasso.

Note that most of the aforementioned methods rely on the assumption that the observations follow a Gaussian distribution. There also exists some work, such as Ravikumar et al. (2011), studied sub-Gaussian data under bounded higher order moments. However, in many real-word applications, the data can follow a heavy-tailed distribution, or may even be corrupted arbitrarily. In such cases, conventional methods yield inaccurate graph estimation even if there are only a few contaminated observations due to the lack of robustness. In order to address this issue, a large body of literature (Liu et al., 2012; Finegold & Drton, 2011; Hirose & Fujisawa, 2015; Sun & Li, 2012; Yang & Lozano, 2015; Balmand & Dalalyan, 2015; Öllerer & Croux, 2015; Loh & Tan, 2015; Chen et al., 2015; Tarr et al., 2016) has focused on providing more robust estimators for precision matrices in the past years. However, most of these estimators were established under some specific contamination models, thus they are not good at dealing with the situation when data are arbitrarily corrupted.

[1] Department of Computer Science, University of Virginia, Charlottesville, Virginia, USA. Correspondence to: Quanquan Gu <qg5w@virginia.edu>.

Proceedings of the 34^th International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017. Copyright 2017 by the author(s).

In this paper, we propose a robust estimator to estimate the precision matrix in high-dimensional GGMs with arbitrarily corrupted data. More specifically, we consider the situation that the corrupted data can appear in any coordinates of the observations. This includes situations that some observations are outliers or data follow some specific contamination models as special cases. The definition of the arbitrary corruption model will be presented in section 3. The key idea of our method is to use a robust covariance matrix estimator, which remains accurate provided a controlled number of arbitrarily corrupted coordinates. Our theory provides not only the spectral norm based estimation error of the proposed estimator, but also the model selection consistency guarantee. More importantly, we show that provided that the number of corrupted samples n_2 for each variable satisfies $n_2 \lesssim \sqrt{n}/\sqrt{\log d}$, where n is the sample size and d is the number of variables, the proposed robust precision matrix estimator attains the same statistical rate as the standard estimator for Gaussian graphical models. Beyond point estimation, we also propose a hypothesis testing procedure to assess the uncertainty of our robust estimator with corrupted observations, and construct the confidence interval for the point estimate. Thorough experiments on both synthetic data and real-world genomic data corroborate the effectiveness of our method.

The remainder of this paper is organized as follows: In Section 2, we discuss some more related work about the robust precision matrix estimation. Section 3 summarizes our proposed estimation method and testing procedure in general and also introduces some necessary backgrounds. Section 4 presents our main results including estimation error bound and inference property. Section 5 provides numerical results, for our method and a number of other methods, of some simulated datasets and a real example on gene expression data. Section 6 concludes with discussion.

Notation Let $\mathbf{A} = [A_{ij}] \in \mathbb{R}^{d\times d}$ be a $d \times d$ matrix and $\mathbf{x} = [x_1, \ldots, x_d]^\top \in \mathbb{R}^d$ be a d-dimensional vector. For $0 < q < \infty$, we define the ℓ_0, ℓ_q and ℓ_∞ vector norms as $\|\mathbf{x}\|_0 = \sum_{i=1}^d \mathbb{1}\{x_i \neq 0\}$, $\|\mathbf{x}\|_q = (\sum_{i=1}^d |x_i|^q)^{\frac{1}{q}}$, $\|\mathbf{x}\|_\infty = \max_{1\le i\le d} |x_i|$, where $\mathbb{1}\{\cdot\}$ represents the indicator function. We use the following notations for the matrix ℓ_q, $\ell_{\max}$, $\ell_{1,1}$ and ℓ_F norms: $\|\mathbf{A}\|_q = \max_{\|\mathbf{x}\|_q=1} \|\mathbf{A}\mathbf{x}\|_q$, $\|\mathbf{A}\|_{\infty,\infty} = \max_{ij} |A_{ij}|$, $\|\mathbf{A}\|_{1,1} = \sum_{i=1}^d \sum_{j=1}^d |A_{ij}|$, $\|\mathbf{A}\|_F = (\sum_{ij} |A_{ij}|^2)^{1/2}$. We use $\mathbf{A}_{*j} = (A_{1j}, \ldots, A_{dj})^\top$ to denote the j-th column vector of $\mathbf{A}$ and $\mathbf{A}_{*\setminus j}$ to denote the submatrix of $\mathbf{A}$ with the j-th column $\mathbf{A}_{*j}$ removed. We also denote by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ the largest and smallest eigenvalues of matrix $\mathbf{A}$, respectively. Furthermore, for a matrix $\mathbf{\Theta}$ and sets of tuples S, S_1, $\mathbf{\Theta}_{S_1,S}$ denotes the set of numbers $(\Theta_{jk})_{j\in S_1, k\in S}$. We define the maximum degree of a graph or row cardinality as $s = \max_{1\le i\le n} |\{j \in V \mid \Theta^*_{ij} \neq 0\}|$, where $V = \{1, \ldots, d\}$ is the vertex set. Finally, for a sequence of random variables X_n, we write $X_n \xrightarrow{d} X$, for some random variable X, if X_n converges in distribution to X.

2 Related Work

In recent years, some attempts have been made toward the robust estimation of high-dimensional GGMs under different corruption models. For example, to deal with heavy tailed distributions, Liu et al. (2012) developed a semiparametric approach called the nonparanormal SKEPTIC. Finegold & Drton (2011) proposed a penalized likelihood approach based on multivariate t-distributions. They also proposed an alternative t-model which requires the use of variational EM or Markov chain Monte Carlo algorithms. Hirose & Fujisawa (2015) introduced a robust estimation procedure for sparse precision matrices based on the penalized negative γ-likelihood function.

In order to address outliers, Sun & Li (2012) proposed a robust estimation of GGMs via a robustified likelihood function with ℓ_1 penalization. In particular, they first use coordinate descent to efficiently estimate the structure of the precision matrix. Then, based on the estimated structure, they re-estimate the parameters of the precision matrix using iterative proportional fitting algorithm to ensure the positive definiteness of their estimator. Yet their method does not have any theoretical guarantee. Yang & Lozano (2015) proposed a trimmed Graphical Lasso method. Specifically, by adding weights to different data points, they improved upon the original graphical Lasso such that it is more robust to outliers. However, they did not provide any model selection consistency guarantee. Balmand & Dalalyan (2015) also studied the problem of robustly estimating the covariance matrix when data are corrupted by outliers. In particular, they proposed to use a modified scaled lasso procedure for covariance matrix estimation and provided the theoretical guarantee of their method.

Another line of related work is Öllerer & Croux (2015); Loh & Tan (2015); Chen et al. (2015); Tarr et al. (2016), which studied the problem of robust precision matrix estimation in high dimensions under the ϵ-contamination model. In particular, under the cell-wise contamination model, Tarr et al. (2016) evaluated the performance of the Glasso and CLIME estimators together with a U-statistic based robust covariance estimator for sparse precision matrix estimation. Under the same contamination model, Öllerer & Croux (2015) provided an analysis for the robustness of these estimators in terms of breakdown behavior. Later on, from the point of statistical consistency, Loh & Tan (2015) established the statistical error bounds for these estimators. However, these methods (Öllerer & Croux, 2015; Loh & Tan, 2015; Tarr et al., 2016) highly depend on the specific cell-wise contamination structure on the data matrix. Recently, inspired by Tukey's depth estimator (Tukey, 1975) for vector estimation, Chen et al. (2015) introduced the concept of matrix depth and proposed a robust covariance matrix estimator using empirical depth function. They showed that their proposed estimator can achieve minimax optimal statistical rate under Huber's ϵ-contamination model. However, it is computationally very expensive to compute the deepest depth of a matrix even in a moderate dimension, which makes such method infeasible in the high-dimensional regime.

All the aforementioned methods are limited to data with heavy tails and outliers. Therefore, they are not suitable to deal with data that are arbitrarily corrupted.

3 Problem Setup and Estimation Method

In this section, we first introduce the setup of our problem, then we present our proposed estimation method and hypothesis testing procedure.

3.1 Problem Setup

Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ be a d-dimensional multivariate Gaussian random vector with zero mean and covariance matrix $\boldsymbol{\Sigma}^*$. It is associated with an undirected graph $G = (V, E)$ with vertex set $V = (1, \ldots, d)$ corresponding to random variables and edge set $E = \{(j,k) \mid j \neq k, \Theta^*_{jk} \neq 0\}$ describing the connections of nodes, where $\boldsymbol{\Theta}^* = \boldsymbol{\Sigma}^{*-1}$ is the precision matrix.

Suppose we have n i.i.d. observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, each of which is drawn from the multivariate Gaussian distribution $N_d(\mathbf{0}, \boldsymbol{\Sigma}^*)$. Let $\mathbf{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n]^\top \in \mathbb{R}^{n\times d}$ be the data matrix and there may exist arbitrary corruption of the data matrix $\mathbf{X}$. More specifically, for each variable/column of data matrix $\mathbf{X}$, we allow at most n_2 coordinates to be arbitrarily corrupted, and we call this kind of corruption model as the arbitrary corruption model. Note that under the arbitrary corruption model, we do not require the corrupted entries lie in the same n_2 rows. Clearly, a special case of the arbitrary corruption model is the outlier model where the corruption appears in n_2 observations. Under the arbitrary corruption model, n_2 is the upper bound on the number of corruptions for each variable, and under the outlier model, n_2 is the upper bound on the number of outliers. Specifically, under the outlier model, the set of row indices $\{1, \ldots, n\}$ of the data matrix $\mathbf{X}$ is divided into two disjoint subsets A and O with $|A| = n_1, |O| = n_2$, and $n = n_1 + n_2$. $\mathbf{X}_A$ denotes samples drawn from the authentic distribution. $\mathbf{X}_O$ denotes samples that are outliers. In general, there is no constraint on the type of corruptions in our setting except an upper bound on the number of corruptions, i.e., n_2. For example, these corruptions could be drawn from other distributions or even be deterministic.

3.2 Estimation Method

Before we introduce our estimation method, we first introduce the truncated inner product which was proposed by Chen et al. (2013). The truncated inner product $\langle \mathbf{u}, \mathbf{v} \rangle_{n_2}$ is defined as follows: given two n-dimensional vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and the truncation number n_2 satisfying $n_2 \leq n$, we first compute the quantity $q_i = u_i v_i$, for $i = 1, \ldots, n$. Then we sort $\{|q_i|\}_{i=1}^n$ and select the smallest $(n - n_2)$ ones. Let Ω be the set of selected indices with cardinality $|\Omega| = n - n_2$, then we have the truncated inner product as $\langle \mathbf{u}, \mathbf{v} \rangle_{n_2} = \sum_{i\in\Omega} q_i$.

The main idea of our estimation method is to use a robust covariance matrix estimator which can mitigate the impact of arbitrary corruptions. More specifically, given a data matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, which is arbitrarily corrupted, we obtain the robust covariance matrix estimator $\widehat{\boldsymbol{\Sigma}}$ through a truncation procedure that each element $\widehat{\Sigma}_{jk}$ is calculated via truncated inner product $\langle \mathbf{X}_{*j}, \mathbf{X}_{*k} \rangle_{n_2}/n_1$. The motivation of this truncation procedure is that the corrupted coordinates with large magnitude may heavily affect the precision of our estimation results, and this simple truncation procedure can reduce such impact. Next, we introduce our robust estimator, which is based on the robust covariance matrix estimator and CLIME:

$$\widehat{\boldsymbol{\Theta}} = \operatorname*{argmin}_{\boldsymbol{\Theta}\in\mathbb{R}^{d\times d}} \|\boldsymbol{\Theta}\|_{1,1} \quad \text{subject to} \quad \|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Theta} - \mathbf{I}\|_{\infty,\infty} \leq \lambda, \tag{3.1}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the robust covariance matrix estimator obtained through truncation, $\lambda > 0$ is a constraint parameter. We refer to (3.1) as Robust CLIME (RCLIME). Note that here we do not consider the Glasso type estimator since it requires the stringent incoherence condition on the covariance matrix to guarantee the model selection consistency. Let $\boldsymbol{\theta}^*_i = \boldsymbol{\Theta}^*_{*i}$ denote the i-th column of $\boldsymbol{\Theta}^*$. To estimate the precision matrix more efficiently, instead of solving (3.1), we can estimate each column of $\boldsymbol{\Theta}^*$ as follows:

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathbb{R}^{d}} \|\boldsymbol{\theta}\|_{1} \quad \text{subject to} \quad \|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\theta} - \mathbf{e}_i\|_{\infty} \leq \lambda, \tag{3.2}$$

for $i = 1, \ldots, d$, and $\mathbf{e}_i \in \mathbb{R}^d$ denotes a column vector that the i-th element is 1 and others are 0. Note that the combined solution $\widehat{\boldsymbol{\Theta}}^1 = [\widehat{\boldsymbol{\theta}}^1_1, \ldots, \widehat{\boldsymbol{\theta}}^1_d]$ of (3.2) is equivalent to the solution of (3.1) (Cai et al., 2011). In addition, since $\widehat{\boldsymbol{\Theta}}^1$ is not symmetric, we need the following symmetrization procedure to get our robust estimator

$$\widehat{\boldsymbol{\Theta}} = \operatorname*{arginf}_{\boldsymbol{\Theta}\in\mathcal{S}^d_{++}} \|\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}^1\|_1, \tag{3.3}$$

where $\mathcal{S}^d_{++} = \{\mathbf{A} \in \mathbb{R}^{d\times d} \mid \mathbf{A} = \mathbf{A}^\top, \mathbf{A} \succ 0\}$ denotes all $d \times d$ symmetric positive definite matrices. The symmetrization procedure in (3.3) can be solved by the projected gradient descent method, and in practice, we can use many simple symmetrization methods, such as the method provided in Cai et al. (2011).

3.3 Hypothesis Test

Based on the proposed robust estimator (3.1), we are interested in testing whether there is an edge between node j and node k in GGMs (Jankova et al., 2015; Neykov et al., 2015; Gu et al., 2015; Xu et al., 2016). More specifically, we want to develop a procedure for the hypothesis test that $H_0 : \Theta^*_{jk} = 0$ versus $H_1 : \Theta^*_{jk} \neq 0$. Let us assume that the k-th column of the precision matrix $\boldsymbol{\Theta}^*$ to be the vector $\boldsymbol{\theta}^*_k = (\alpha^*, \boldsymbol{\gamma}^{*\top})^\top$ where α^* is the j-th element of the vector $\boldsymbol{\theta}^*_k$ and $\boldsymbol{\gamma}^* \in \mathbb{R}^{d-1}$ is the remaining $(d-1)$-dimensional vector. Thus it is equivalent to test the one dimensional component $H_0 : \alpha^* = 0$ versus the non-restricted alternative $H_1 : \alpha^* \neq 0$. In this case, $\boldsymbol{\gamma}^*$ are nuisance parameters. To this end, we first introduce the following estimation equation projected (EEP) along the direction $\widehat{\mathbf{w}}$:

$$\widehat{S}(\boldsymbol{\theta}) = \widehat{\mathbf{w}}^\top\big(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\theta} - \mathbf{e}_k\big), \tag{3.4}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the the robust covariance matrix estimator and $\widehat{\mathbf{w}}$ is the solution of the optimization problem (3.2) with $i = j$. The motivation of projecting the estimation equation to a sparse direction (3.4) is to help us construct a test statistic which has a tractable limiting distribution in the high-dimensional regime. In high-dimensional settings, $\widehat{\boldsymbol{\Sigma}}$ is not positive definite, we cannot solve the equation $\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}} - \mathbf{e}_k = 0$ by taking the inverse of $\widehat{\boldsymbol{\Sigma}}$ directly. Therefore, given the sparsity assumption on $\boldsymbol{\theta}^*$, the estimator in (3.2) can address such ill-posed problem for solving the estimation equation $\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}} - \mathbf{e}_k = 0$ in high-dimensional settings. Furthermore, projecting the estimation equation to a certain direction (3.4) makes the limiting distribution of $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\boldsymbol{\gamma}}^\top)^\top$ in (3.2) tractable. More specifically, if we choose the $\widehat{\mathbf{w}}$ as the projection direction, then due to the fact that $\widehat{\mathbf{w}}$ is a consistent estimator of $\mathbf{w}^* := \boldsymbol{\Theta}^*_{*j}$, the estimator $\widehat{\boldsymbol{\gamma}}$ of the high-dimensional nuisance parameters in (3.2) is asymptotically ignorable along this direction. Therefore, we can solve the projected estimation equation $\widehat{S}(\alpha, \widehat{\boldsymbol{\gamma}}) = 0$ to get an debiased estimator of α^* as follows:

$$\widetilde{\alpha} = \widehat{\alpha} - \frac{\widehat{\mathbf{w}}^\top(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\theta}} - \mathbf{e}_k)}{\widehat{\mathbf{w}}^\top\widehat{\boldsymbol{\Sigma}}_{*j}}, \tag{3.5}$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\boldsymbol{\gamma}}^\top)^\top$ is the estimator of $\boldsymbol{\Theta}^*_{*k}$, and $\widehat{\mathbf{w}}$ is the estimator of $\boldsymbol{\Theta}^*_{*j}$. Thus we define the following test statistic built upon the debiased estimator $\widetilde{\alpha}$

$$\widehat{T}_n = \sqrt{n_1}(\widetilde{\alpha} - \alpha^*)/\widehat{\sigma}, \tag{3.6}$$

where n is the number of observations, n_2 is the upper bound on the number of corruptions, $n_1 = n - n_2$, and $\widehat{\sigma}^2 = \widehat{w}_j\widehat{\theta}_k + \widehat{w}_k\widehat{\theta}_j$, where $\widehat{w}_j, \widehat{\theta}_j$ denote the j-th elements of $\widehat{\mathbf{w}}$ and $\widehat{\boldsymbol{\theta}}$ respectively. Note that $\widehat{\sigma}^2$ is a consistent estimator to $\sigma^2 = w^*_j\theta^*_k + w^*_k\theta^*_j$ under the Gaussian assumption of the data, where w^*_j, θ^*_k are the j-th and k-th columns of $\mathbf{w}^*$ and $\boldsymbol{\theta}^*$ respectively. We will show in the next section that the proposed debiased estimator $\widetilde{\alpha}$ is consistent to α^*, and the test statistic $\widehat{T}_n$ is asymptotically normal $\sqrt{n_1}(\widetilde{\alpha} - \alpha^*)/\widehat{\sigma} \xrightarrow{d} N(0, 1)$ under the null hypothesis. Therefore, our asymptotic level-α test is given by

$$\Psi_n = \begin{cases} 0 \quad (\equiv \text{ accept } H_0) & \text{if } |\widehat{T}_n| \le C_\alpha, \\ 1 \quad (\equiv \text{ reject } H_0) & \text{if } |\widehat{T}_n| > C_\alpha, \end{cases} \tag{3.7}$$

where $C_\alpha = \Phi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution $N(0, 1)$. Furthermore, we can construct asymptotic level-α confidence intervals of α^* as $\widetilde{\alpha} \pm \Phi^{-1}(1-\alpha/2)\widehat{\sigma}/\sqrt{n}$. Note that in practice, although we have no idea about the exact upper bound on the number of corruptions, i.e., n_2, we can use techniques such as cross-validation to choose the best truncation number n_2.

4 Main Results

In this section, we present our main results and discuss connections with some related works. We start by stating some assumptions, which are required in our analysis. We impose an important eigenvalue condition on the population covariance matrix.

Assumption 4.1. There exist a constant $\kappa > 0$ such that

$$0 < 1/\kappa \le \lambda_{\min}(\boldsymbol{\Sigma}^*) \le \lambda_{\max}(\boldsymbol{\Sigma}^*) \le \kappa < \infty.$$

This assumption can exclude singular or nearly singular covariance matrices, thus guarantee the uniqueness of $\boldsymbol{\Theta}^*$.

In this paper, we consider the precision matrix $\boldsymbol{\Theta}^*$ that belongs to a class of matrices $\mathcal{U}(s)$, i.e., $\mathcal{U}(s) = \big\{\boldsymbol{\Omega} \in \mathbb{R}^{d\times d} \mid \boldsymbol{\Omega} \succ 0, \|\boldsymbol{\Omega}\|_1 \le M, \max_{1\le i\le d}\sum_{j=1}^d \mathbb{1}\{\Omega_{ij} \neq 0\} \le s\big\}$, where $\boldsymbol{\Omega} \succ 0$ means $\boldsymbol{\Omega}$ is positive definite and s corresponds to the row cardinality. Note that this sparse precision matrix class has been previously considered in Cai et al. (2011); Liu & Wang (2012); Zhao & Liu (2013). In addition, it immediately implies that $\|\boldsymbol{\Theta}^*_{*j}\|_1 \le \|\boldsymbol{\Theta}^*\|_1 \le M$, where $\boldsymbol{\Theta}^*_{*j}$ is the jth column vector of $\boldsymbol{\Theta}^*$.

Now, we are ready to provide our main results. The first one characterizes the performance of our robust estimator under the arbitrary corruption model. It shows that even if the upper bound on the number of corruptions n_2 scales with $\sqrt{n}$, where n is the number of observations, our robust estimator can still recover the correct support. Note that our results are derived under the arbitrary corruption model. Since the outlier model is a special case of the arbitrary corruption model, our results can directly apply to the outlier case.

Theorem 4.2. Under the arbitrary corruption model, suppose $\boldsymbol{\Theta}^* \in \mathcal{U}(s)$ and Assumption 4.1 is satisfied. In addition, assume the upper bound on the number of corruptions n_2 satisfies $n_2 \le a\sqrt{n}$ for some constant $a \ge 0$. If

$n \geq 4a^2$, and we choose the regularization parameter satisfying $\lambda = CM\kappa^2\big(\sqrt{\log d/n} + n_2 \log d/n\big)$, then, with probability at least $1 - C_1/d$, the estimator $\widehat{\mathbf{\Theta}}$ in (3.1) satisfies

$$\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_2 \leq C_2 M^2 \kappa^2 \bigg(s\sqrt{\frac{\log d}{n}} + \frac{n_2 s \log d}{n} \bigg). \quad (4.1)$$

Furthermore, if the nonzero entries of $\mathbf{\Theta}^*$ satisfy

$$\min_{i \neq j, \Theta^*_{ij} \neq 0} |\Theta^*_{ij}| \geq C_3 M^2 \kappa^2 \bigg(\sqrt{\frac{\log d}{n}} + \frac{n_2 \log d}{n} \bigg),$$

then the Robust CLIME can correctly identify nonzero entries of $\mathbf{\Theta}^*$.

Remark 4.3. According to (4.1), the estimation error of our robust estimator consists of two terms. The first one $O(s\sqrt{\log d/n})$ corresponds to the estimation error without corruptions. The second extra term $O(sn_2 \log d/n)$, which is linear in n_2, is due to the effect of arbitrary corruption. More specifically, if there is no corruption in our data, then the second term becomes zero since $n_2 = 0$. Therefore, the estimation error of our method reduces to $O(s\sqrt{\log d/n})$, which matches the minimax optimal rate for sparse precision matrix estimation without corruptions in terms of spectral norm (Yuan, 2010; Cai et al., 2011; Ravikumar et al., 2011). In addition, (4.1) in Theorem 4.2 indicates that our robust estimator can correctly recover the support of $\mathbf{\Theta}^*$ even if the upper bound on the number of corruptions n_2 scales with $\sqrt{n/\log d}$, where n is the number of observations. In addition, under the outlier model, this estimation result is comparable to the result provided by Yang & Lozano (2015). In their study, they proved that the proposed estimator can successfully recover the true parameter provided that the upper bound of the number of outliers is $O(\sqrt{n})$. However, Yang & Lozano (2015) does not consider the case when the data is arbitrarily corrupted.

Furthermore, if the upper bound on the number of corruptions n_2 satisfies $n_2 \lesssim \sqrt{n}/\sqrt{\log d}$, our robust estimator can achieve the same statistical rate as the standard estimator for Gaussian graphical models. This is summarized in the following corollary.

Corollary 4.4. Under the same conditions of Theorem 4.2, if we further assume that the upper bound on the number of corruptions n_2 satisfies $n_2 \lesssim \sqrt{n}/\sqrt{\log d}$, then for the robust estimator $\widehat{\mathbf{\Theta}}$ in (3.1), we have, with probability at least $1 - C/d$, that

$$\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_2 \leq C_1 M^2 \kappa^2 s\sqrt{\frac{\log d}{n}}.$$

Furthermore, if the nonzero entries of $\mathbf{\Theta}^*$ satisfy

$$\min_{i \neq j, \Theta^*_{ij} \neq 0} |\Theta^*_{ij}| \geq C_2 M^2 \kappa^2 \sqrt{\log d/n},$$

then the Robust CLIME can correctly identify the nonzero entries of $\mathbf{\Theta}^*$.

Remark 4.5. Compared with Theorem 4.2, Corollary 4.4 implies that under a slightly stricter condition on the upper bound of the number of corruptions $n_2 = O(\sqrt{n}/\sqrt{\log d})$, our robust estimator can successfully recover the true parameter $\mathbf{\Theta}^*$ with guaranteed estimation error $O(s\sqrt{\log d/n})$. Note that this error bound exactly recover the spectral norm error bound for the case without corruptions (Yuan, 2010; Cai et al., 2011; Ravikumar et al., 2011), which demonstrates the superiority of our estimator.

Next, we present the asymptotic results of our proposed test statistics in (3.6), which verifies the effectiveness of our testing procedure. Note that we consider the case that the true observations are drawn from a Gaussian distribution.

Theorem 4.6. Suppose Assumption 4.1 is satisfied and $\sqrt{n_1} sM^2\kappa^4(\sqrt{\log d/n_1} + n_2 \log d/n_1)^2 = o(1)$, where $n_1 = n - n_2$. If we choose regularization parameter satisfying $\lambda = CM\kappa^2(\sqrt{\log d/n_1} + n_2 \log d/n_1)$, then the test statistic in (3.6) is asymptotically normal

$$\frac{\sqrt{n_1}(\widetilde{\alpha} - \alpha^*)}{\widehat{\sigma}} \xrightarrow{d} N(0, 1),$$

where $\widehat{\sigma}^2 = \widehat{w}_j \widehat{\theta}_k + \widehat{w}_k \widehat{\theta}_j$, and $\widetilde{\alpha}$ is defined in (3.5).

Remark 4.7. Theorem 4.6 provides us an efficient test for the existence of an edge in GGMs, and gives us an efficient interval estimation of $\alpha^* = \Theta^*_{ij}$. In addition, Theorem 4.6 implies that if the upper bound on the number of corruptions n_2 satisfies $n_2 \lesssim \sqrt{n}/\sqrt{\log d}$ and the quantity M is a constant, then the assumption $\sqrt{n_1} sM^2\kappa^4(\sqrt{\log d/n_1} + n_2 \log d/n_1)^2 = o(1)$ reduces to $s \log d/\sqrt{n} = o(1)$, which gives us the sparsity assumption that $s = O(\sqrt{n} \log d)$. This requirement on sparsity matches the best-known results for edge testing in GGMs (Liu et al., 2013; Ren et al., 2015). More importantly, Theorem 4.6 suggests that even when $n_2 = O(\sqrt{n}/\sqrt{\log d})$ out of n observations of each variable are arbitrarily corrupted, our testing procedure is still efficient.

5 Experiments

In this section, we compare our robust estimator with some existing methods, including trimmed Graphical Lasso (tGLasso) (Yang & Lozano, 2015), t^*-Lasso (tLasso) (Finegold & Drton, 2011), robust ℓ_1 penalized likelihood (RLL) (Sun & Li, 2012), nonparanormal SKEPTIC (Liu et al., 2012), and pairwise based covariance estimator (spearC) (Loh & Tan, 2015) on some synthetic datasets. Our comparisons focus on their performance in both graph recovery and parameter estimation. The implementation of tLasso and RLL is based on the code provided by authors. The implementation of other baseline algorithms is based on **R** package **huge**[1]. We conduct some simulations to investigate the performance of our proposed hypothesis testing procedure. Furthermore, we compare our method with GLasso on a gene expression data.

5.1 Synthetic Data

In our numerical simulations, we consider the following two settings: (i) $n = 100$, $d = 100$; and (ii) $n = 200$, $d = 400$. We generate the true precision matrices based on two graph structures: cluster and band. More specifically, the precision matrices $\mathbf{\Theta}^*$ are generated by **huge** package, and the magnitude of correlations is the default value (0.3) in the huge generator. In order to incorporate corruptions, we generate our observations by the following procedure.

For the arbitrary corruption model, we first generate the n by d data matrix $\mathbf{X}$ from the Gaussian distribution $N_d(\mathbf{0}, \mathbf{\Theta}^{*-1})$. Then, for each column of the data matrix, we let np coordinates be arbitrarily corrupted, where we consider the corruption rate $p = 0.1$ for small number of corruptions and $p = 0.2$ for large number of corruptions. In addition, each corrupted coordinate is generated by normal distributions $N(\mu, \sigma)$ as follows:

$$M_1^A : \mu = 1, \sigma = 1, \quad M_2^A : \mu = 2, \sigma = 1. \tag{5.1}$$

For the outlier model, we use the setup similar to Sun & Li (2012); Yang & Lozano (2015). Specifically, we generate each observation from the mixture model as follows:

$$\begin{aligned} \boldsymbol{X}_i \sim (1-p)N_d(\mathbf{0}, \mathbf{\Theta}^{*-1}) + \frac{p}{2}N_d(\boldsymbol{\mu}, \mathbf{\Theta}'^{-1}) \\ + \frac{p}{2}N_d(-\boldsymbol{\mu}, \mathbf{\Theta}'^{-1}) \quad \text{for} \quad i = 1, \ldots, n, \end{aligned}$$

where we consider the corruption rate $p = 0.1$ for small number of corruptions and $p = 0.2$ for large number of corruptions. Furthermore, each outlier is generated by normal distributions $N_d(\boldsymbol{\mu}, \mathbf{\Theta'}^{-1})$ as follows

$$M_1^O : \boldsymbol{\mu} = (1, \ldots, 1)^\top, \mathbf{\Theta}' = \mathbf{I}_d, \tag{5.2}$$

$$M_2^O : \boldsymbol{\mu} = (2, \ldots, 2)^\top, \mathbf{\Theta}' = \mathbf{I}_d. \tag{5.3}$$

Note that under both corruption models, we set the corruption rate $p \in \{0.1, 0.2\}$. In other words, we choose the number of corruptions to be 10% and 20% of all observations. This is due to the threshold of the number of corruptions $n_2 = O(\sqrt{n})$ suggested in our theorem.

Point Estimation: We choose tuning parameters of each method as follow. For tGLasso, we choose n_2/n from $[0.5, 1]$, which is suggested by Yang & Lozano (2015). For RLL, we choose $\beta \in \{0.005, 0.01, 0.02\}$, which is suggested by Sun & Li (2012). And for Robust CLIME, we choose n_2 around 15 (± 5). Since the performance of t^*-Lasso is similar to t-Lasso, we just show the results of t^*-Lasso. All results we reported are their best performance based on these parameters.

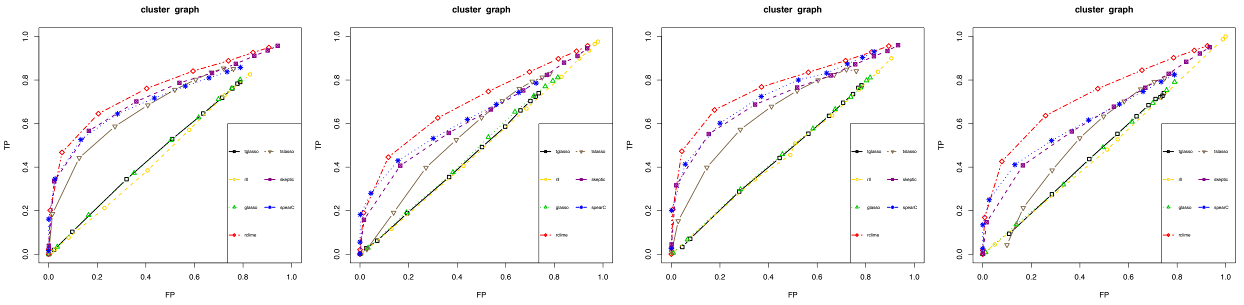
First, we use receiver operating characteristic (ROC) curves to compare the overall performance of our method with others in model selection over the full paths. For the arbitrary corruption model, the ROC curves on cluster graphs averaged over 50 simulations are shown in Figure 1. We can observe that under the arbitrary corruption model, as the number of corruptions increase, the advantage of our approach becomes more significant. For the outlier model, we also observe similar good performance of our method, especially for outliers with large magnitude. Due to space limit, the ROC curves for the outlier model can be found in the longer version of this paper. These results indicate that our method is very competitive in the graph recovery problem with arbitrary corruptions.

Then, we evaluate the performance of our method and some existing approaches in parameter estimation. For model settings mentioned above, we choose the corruption rate $p = 0.1$ for the purpose of comparisons. We generate a dataset as the training sample, and an independent dataset from the same distribution as the test set. We set $n_2/n = 0.9$ for tGLasso, $\beta = 0.01$ for RLL, and $n_2 = np$ for Robust CLIME. We also choose the tuning parameter λ by grid search based on its performance on the training sample and evaluate those estimators on the test set. Here we use Spectral norm error $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_2$ and Frobenius norm error $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_F$ to compare the performance of different methods in parameter estimation. Tables 1 and 2 summarize estimation error results in term of Spectral norm averaged over 50 simulations. These results demonstrate the advantage of our method in parameter estimation. Other comparison results in terms of Frobenius norm error are deferred to the longer version of this paper.

Hypothesis Test: We investigate the finite sample performance of our proposed hypothesis testing procedure through some simulation studies. We use the data generating process similar to Jankova et al. (2015); Neykov et al. (2015), and we consider the case that there are some corruptions in our data. More specifically, for the aforementioned two settings, we consider the band graph structure with band width 1 with the corresponding precision matrix $\mathbf{\Theta}^*$ generated by **R** package **huge**. The magnitude of correlations is the default value in the huge generator. In order to incorporate corruptions, we use the same approach described above to generate observations. Specifically, for the arbitrary corruption model, we generate samples through model M_2^A in (5.1) with $p = 0.1, 0.2$. For the outlier model, we generate samples through model M_2^O in (5.2) with $p = 0.1, 0.2$.

To check the validity of the type I error of our test, we run

[1] http://cran.r-project.org/web/packages/huge



(a) (b) (c) (d)

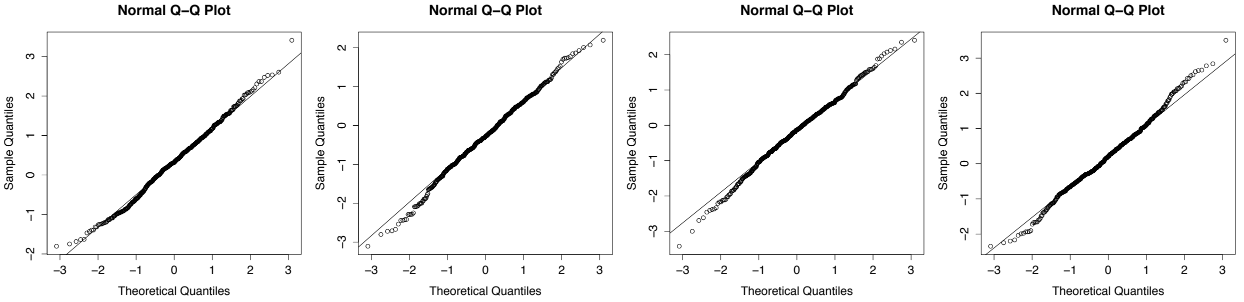
Figure 1. ROC curves of different methods on cluster graphs under the arbitrary corruption model. (a): $d = 400, n = 200, p = 0.1, \mu = 1$; (b): $d = 400, n = 200, p = 0.2, \mu = 1$; (c): $d = 400, n = 200, p = 0.1, \mu = 2$; (d): $d = 400, n = 200, p = 0.2, \mu = 2$

Table 1. Quantitative comparisons of the GLasso, t^*Lasso, tGLasso, RLL, SKEPTIC, spearC and our Robust estimator on the cluster, band graphs in terms of $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_2$ under the outlier model.

Model	d	GLasso	t^*Lasso	tGLasso	RLL	SKEPTIC	spearC	Ours
Band	100	5.803(0.107)	4.180(0.129)	2.755(0.217)	3.639(0.315)	5.579(0.177)	4.071(0.133)	2.471(0.110)
	400	5.886(0.171)	5.755(0.114)	2.939(0.102)	3.739(0.181)	5.891(0.115)	4.481(0.114)	2.877(0.109)
Cluster	100	5.537(0.071)	5.318(0.031)	4.944(0.093)	5.004(0.110)	5.529(0.061)	5.283(0.098)	4.776(0.127)
	400	9.444(0.092)	8.828(0.108)	8.586(0.127)	8.795(0.087)	9.489(0.177)	8.819(0.143)	8.160(0.102)

Table 2. Quantitative comparisons of the GLasso, t^*Lasso, tGLasso, RLL, SKEPTIC, spearC and our Robust estimator on the cluster, band graphs in terms of $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\|_2$ under arbitrary the corruption model.

Model	d	GLasso	t^*Lasso	tGLasso	RLL	SKEPTIC	spearC	Ours
Band	100	4.688(0.096)	3.694(0.258)	4.682(0.118)	4.403(0.107)	4.530(0.134)	4.487(0.159)	3.433(0.171)
	400	4.763(0.127)	3.886(0.219)	4.768(0.095)	4.584(0.183)	4.586(0.107)	4.552(0.128)	3.581(0.144)
Cluster	100	5.022(0.092)	4.447(0.133)	4.906(0.171)	5.496(0.106)	4.907(0.116)	4.850(0.125)	4.111(0.131)
	400	7.133(0.127)	6.798(0.181)	7.408(0.112)	9.050(0.209)	6.840(0.161)	6.804(0.139)	6.428(0.177)



(a) (b) (c) (d)

Figure 2. Q-Q plot of test statistic $\widehat{T}_n$. (a-b): data generated from the outlier model with $d = 100, n = 100$ and $d = 400, n = 200$ respectively; (c,d): data generated from the arbitrary corruption model with $d = 100, n = 100$ and $d = 400, n = 200$ respectively.

500 simulations. The detail of our hypothesis testing procedure is described in Section 3.3. In the two different settings, we set $n_2 = 10$ and $n_2 = 20$ respectively, and we choose the tuning parameters λ by cross-validations. Table 3 summarizes the empirical type I errors of our test in different settings. We can observe that the empirical type I errors are close to the significance level. Figure 2 shows the Q-Q plots of our test statistic $\widehat{T}_n$ in (3.6) based on 500 simulations. These plots corroborate the asymptotic normality of our test statistic. All these results demonstrate the advantage of our hypothesis testing procedure under the arbitrary corruption model.

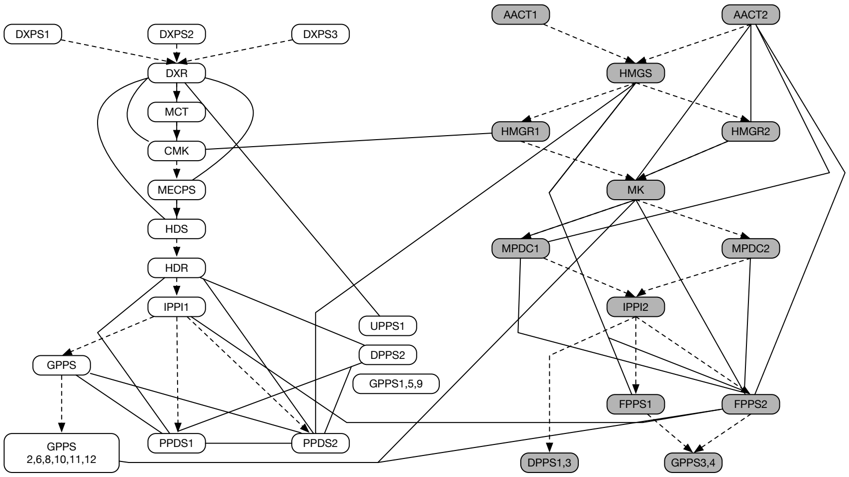


Figure 3. Genetic network identified by Robust CLIME for the gene expression data of *Arabidopsis thaliana*. Solid edges: graph estimated by Robust CLIME. Dotted arrows: known metabolic pathway. Left white figure and Right grey figure correspond to MEP and MAV pathway respectively. Note that the arrow lines correspond to the known metabolic pathway, they do not mean directed networks.

Table 3. Empirical coverage of 95% confidence intervals and Type I error at 0.05 significant level

Corruption model	d	Coverage	Width	Type I error
outlier	100	0.956	0.338	0.044
	400	0.946	0.344	0.054
arbitrary	100	0.950	0.352	0.050
	400	0.942	0.356	0.058

5.2 Gene Expression Data

In this subsection, we use the gene expression data of *Arabidopsis thaliana*, which was analyzed by Wille et al. (2004) and later on by Finegold & Drton (2011); Hirose & Fujisawa (2015), to illustrate the advantage of our method. This data set includes $n = 118$ observations with 39 gene expression levels. For this gene expression dataset, we preprocess it through $\mathbf{R}$ package **limma**[3]. Figure 6 in Appendix illustrates the histogram of some rescaled gene expression data. It shows that some rescaled gene expressions contain some expression levels with extreme large magnitude, which may be outliers. Therefore, we want to apply our method to construct a network among these genes. For Robust CLIME, we set $n_2 = 10$ and adopt 5-fold cross-validation to choose the tuning parameter λ.

The graph estimated by our method is given in Figure 3. The dotted arrows and the solid undirected edges correspond to the known metabolic pathway and the graph estimated by Robust CLIME, respectively. We can see that our approach identifies a similar graph to that obtained by previous analysis of Wille et al. (2004) but with fewer "cross-talk" edges between two pathways. For example, our approach finds the important connection between AACT2 and the group MK, MPDC1, and FFPS2 in MAV pathway. And in MEP path way, it also identifies the connection among DXR, MCT, CMK and MECPS. Other methods such as GLasso tends to estimate more links between two pathways in order to identify these important relationships. These edges between two pathways provided by GLasso might be inaccurate relationships due to the lack of robustness. The graph recovered by GLasso and the graph established by Wille et al. (2004) can be found in the longer version of this paper.

6 Conclusions and Future Work

In this paper, for the Gaussian graphical model estimation with arbitrary corruptions, we proposed a new estimator for high-dimensional precision matrices based on the robust covariance matrix estimator. We not only provide the estimation error bound of our robust estimator, but also propose a hypothesis testing procedure to assess the uncertainty of our robust estimator with corrupted observations, and construct the confidence interval for the point estimate. However, most of the robust high dimensional estimators as well as our proposed estimator are not invariant under the group action (Davies et al., 2005; Draisma et al., 2013), we will study this problem in our future work.

Acknowledgment

We would like to thank the anonymous reviewers for their helpful comments. This research was sponsored in part by the National Science Foundation IIS-1618948 and IIS-1652539. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

[3] Available on http://bioconductor.org/packages/limma

References

Balmand, Samuel and Dalalyan, Arnak. Convex programming approach to robust estimation of a multivariate gaussian model. *arXiv preprint arXiv:1512.04734*, 2015.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. Model selection through sparse maximum likelihood estimation. 9(3):485–516, 2008.

Cai, T., Liu, W., and Luo, X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. 106 (494):594–607, 2011.

Chen, Mengjie, Gao, Chao, and Ren, Zhao. Robust covariance matrix estimation via matrix depth. *arXiv preprint arXiv:1506.00691*, 2015.

Chen, Yudong, Caramanis, Constantine, and Mannor, Shie. Robust high dimensional sparse regression and matching pursuit. *arXiv preprint arXiv:1301.2725*, 2013.

Cox, David Roxbee and Wermuth, Nanny. *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press, 1996.

Davies, P Laurie, Gather, Ursula, et al. Breakdown and groups. *The Annals of Statistics*, 33(3):977–1035, 2005.

Draisma, Jan, Kuhnt, Sonja, Zwiernik, Piotr, et al. Groups acting on gaussian graphical models. *The Annals of Statistics*, 41(4):1944–1969, 2013.

Finegold, Michael and Drton, Mathias. Robust graphical modeling of gene networks using classical and alternative t-distributions. *The Annals of Applied Statistics*, pp. 1057–1080, 2011.

Friedman, J., Hastie, T., and Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Gu, Quanquan, Cao, Yuan, Ning, Yang, and Liu, Han. Local and global inference for high dimensional gaussian copula graphical models. *arXiv preprint arXiv:1502.02347*, 2015.

Hirose, Kei and Fujisawa, Hironori. Robust sparse gaussian graphical modeling. *arXiv preprint arXiv:1508.05571*, 2015.

Jankova, Jana, van de Geer, Sara, et al. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

Liu, Han and Wang, Lie. Tiger: A tuning-insensitive approach for optimally estimating gaussian graphical models. *arXiv preprint arXiv:1209.2437*, 2012.

Liu, Han, Han, Fang, Yuan, Ming, Lafferty, John, and Wasserman, Larry. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, pp. 2293–2326, 2012.

Liu, Weidong et al. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978, 2013.

Loh, Po-Ling and Tan, Xin Lu. High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination. *arXiv preprint arXiv:1509.07229*, 2015.

Meinshausen, N. and Bühlmann, P. High dimensional graphs and variable selection with the lasso. 34(3):1436–1462, 2006.

Neykov, Matey, Ning, Yang, Liu, Jun S, and Liu, Han. A unified theory of confidence regions and testing for high dimensional estimating equations. *arXiv preprint arXiv:1510.08986*, 2015.

Öllerer, Viktoria and Croux, Christophe. Robust high-dimensional precision matrix estimation. In *Modern Nonparametric, Robust and Multivariate Methods*, pp. 325–350. Springer, 2015.

Raskutti, Garvesh, Wainwright, Martin J, and Yu, Bin. Restricted eigenvalue properties for correlated gaussian designs. *The Journal of Machine Learning Research*, 11: 2241–2259, 2010.

Ravikumar, P., Wainwright, M., Raskutti, G., and Yu, B. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. 5:935–980, 2011.

Ren, Zhao, Sun, Tingni, Zhang, Cun-Hui, Zhou, Harrison H, et al. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026, 2015.

Sun, Hokeun and Li, Hongzhe. Robust gaussian graphical modeling via l1 penalization. *Biometrics*, 68(4):1197–1206, 2012.

Tarr, Garth, Müller, Samuel, and Weber, Neville C. Robust estimation of precision matrices under cellwise contamination. *Computational Statistics & Data Analysis*, 93: 404–420, 2016.

Tukey, John W. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, pp. 523–531, 1975.

Vershynin, Roman. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Wang, Lingxiao, Ren, Xiang, and Gu, Quanquan. Precision matrix estimation in high dimensional gaussian graphical models with faster rates. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 177–185, 2016.

Wille, Anja, Zimmermann, Philip, Vranová, Eva, Fürholz, Andreas, Laule, Oliver, Bleuler, Stefan, Hennig, Lars, Prelic, Amela, von Rohr, Peter, Thiele, Lothar, et al. Sparse graphical gaussian modeling of the isoprenoid gene network in arabidopsis thaliana. *Genome Biol*, 5 (11):R92, 2004.

Xu, Pan and Gu, Quanquan. Semiparametric differential graph models. In *Advances in Neural Information Processing Systems*, pp. 1064–1072, 2016.

Xu, Pan, Tian, Lu, and Gu, Quanquan. Communication-efficient distributed estimation and inference for transelliptical graphical models. *arXiv preprint arXiv:1612.09297*, 2016.

Xu, Pan, Zhang, Tingting, and Gu, Quanquan. Efficient algorithm for sparse tensor-variate gaussian graphical models via gradient descent. In *Artificial Intelligence and Statistics*, pp. 923–932, 2017.

Yang, Eunho and Lozano, Aurélie C. Robust gaussian graphical modeling with the trimmed graphical lasso. In *Advances in Neural Information Processing Systems*, pp. 2584–2592, 2015.

Yuan, M. High dimensional inverse covariance matrix estimation via linear programming. 11(8):2261–2286, 2010.

Yuan, M. and Lin, Y. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Zhao, Tuo and Liu, Han. Sparse inverse covariance estimation with calibration. In *Advances in Neural Information Processing Systems*, pp. 2274–2282, 2013.